ANALYSIS OF SERIAL TURBO CODES OVER ABELIAN GROUPS FOR SYMMETRIC CHANNELS

FEDERICA GARIN* AND FABIO FAGNANI†

Abstract. In this paper we study serial turbo interconnections of Abelian group codes, to be used on symmetric channels. Particular attention is devoted to AWGN channel with input restricted to m-PSK constellation, with corresponding group structure $\mathbb{Z}_m$.

We establish the exact asymptotic decay of the average symbol and word error probabilities when the interleaver length goes to infinity (interleaver gain). Moreover, we give a detailed characterization of the distance parameter characterizing the behavior for the signal-to-noise ratio going to infinity (effective free distance). Some of our results are new also in the binary context: in particular, the lower bound to the error probability decay.

Key words. serial interconnected codes, turbo codes, trellis-coded modulation, average error probability, codes over groups, effective distance

AMS subject classifications. 94B10, 94B12, 94B70, 90C27

1. Introduction.

Turbo codes have made their appearance in 1993 with the pioneering work [5] (see also [6]). Since then, there has been a huge effort to understand their great performance, by a significant number of researchers in the field, e.g. [2, 3, 9, 10, 12, 14, 29, 35, 37, 39, 38, 40]. In the meanwhile, a lot of variations and extensions have been proposed with respect to the original parallel concatenated scheme in [6]. A particularly relevant scheme is the serially concatenated version, first proposed in [4] and furtherly studied in [29]. Other significant variations concern the possibility to work with a more restricted class of coupling interleavers [7, 11] in order to achieve a higher convergence abscissa in the iterative decoding algorithm.

In the theoretical analysis of these coding schemes we can find two main lines: on the one hand, they have been studied in combination with the suboptimal iterative decoding algorithm (see e.g. [35, 10, 14]), and on the other hand, they have been studied in a more classical setting, considering optimal maximum likelihood (ML) decoding. This second approach separates the analysis of the coding schemes from the use of the suboptimal iterative decoding (see e.g. [2, 3, 4, 9, 12, 29, 37, 39, 40]). The results in [3, 4, 29] provide upper bounds for the decay of the averaged bit error rate $\overline{P_b(e)}$, of type $CN^{-\alpha}$ where α is a parameter, called interleaver gain, only depending on the constituent codes of the turbo scheme. A further more refined analysis has been done for the situation when the Bhattacharyya parameter γ of the channel goes to zero, showing that the above constant C could be estimated as $C = K\gamma^{d^*}$ where d^* is a sort of generalized minimum distance of the scheme and K is a universal constant. The main contribution of this line of research has been to discover these two parameters α and d^* and their relation to the constituent encoders, which makes them interesting from a design point of view.

Another significant extension of the theory of turbo codes has been the study of schemes for non-binary alphabets, in order to use them over channels with a non-binary input alphabet.

*F. Garin (corresponding author) was with Dipartimento di Matematica, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy. She is now with Department of Information Engineering (DEI), Università di Padova, Via Gradenigo 6/a, 35131 Padova, Italy (garin@dei.unipd.it).

†F. Fagnani is with Dipartimento di Matematica, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy (fabio.fagnani@polito.it).

Mostly, the construction of such non-binary turbo codes has followed the so called 'pragmatic design' approach: turbo codes are designed for the binary case and optimized independently from the external high-order input constellation. Joint optimization is very little, usually regarding the mapping of the coded bits into the constellation signals. Performance evaluation is typically obtained by simulation. As a matter of fact, almost all schemes presented in the literature belong to this class (see e.g. [31], [36],[21]).

On the contrary, in the so called 'analytical design' turbo codes are designed and optimized by taking into account the chosen non-binary channel. This is the same philosophy of classical trellis coded modulation (TCM) schemes. Pioneering works in this direction are [23, 34, 16]. This approach exploits the symmetries of some important families of codes, such as additive Gaussian channels with geometrically uniform input constellation, by using a suitable algebraic structure matched to the channel. There are two fundamental difficulties in the theoretical developments of such turbo schemes, new with respect to the binary case:

- in general, the relevant distance affecting error probability and its estimations is not Hamming distance, e.g. for Gaussian channels Euclidean distance matters;
- in order to obtain bit or word error rate to be independent from the transmitted information word, one needs to carefully choose the algebraic structure of the constituent encoders. For example, the so-called bit-geometrically-uniform encoders [23, 16] ensure a sort of isometry in the scheme between the input Hamming distance and the output Euclidean distance; these encoders give a bit error rate independent from the information word, but this independence is only reached in the average sense and not for a single realization of the interleaver and, moreover, it forces the use of encoders on non-Abelian groups, creating a lot of algebraic technical problems in the treatment of the convolutional codes.

This paper also follows the 'analytical design' approach, but taking a different road with respect to previous works. Here, we study serial turbo interconnections on an Abelian group (typically $\mathbb{Z}_m$). The constituent encoders are chosen to be homomorphic and this does not guarantee the bit error rate $P_b(e)$ (or its average version $\overline{P_b(e)}$) to be independent from the transmitted information word. However, if we consider the symbol error rate $P_s(e)$, where the symbol is to be intended as a suitable aggregation of the input bits in order to form the input group of the encoder (typically $\mathbb{Z}_m$ again), this independence is ensured. Of course $P_s(e)$ and $P_b(e)$ are strictly linked to each other, so that estimations on the former lead to estimations on the latter. Moreover, the choice to work with Abelian groups allows us to use a much richer and flexible theory at the level of convolutional codes, widely studied in the past [30, 17, 18, 19].

In this paper we study the average symbol and word error rate $\overline{P_s(e)}$, $\overline{P_w(e)}$ for such schemes on symmetric channels and under ML decoding. We determine the asymptotic decay $N^{-\alpha}$ when the interleaver length N goes to infinity. Our results extend and complete the bounds given in [4, 29]. The extension is two-fold: we pass from binary codes to schemes over generic Abelian groups, and, we move from the uniform interleaver assumption by also considering structured families of interleavers. Moreover, we provide a lower bound which was missing even for the binary case. Technical proofs are inspired by [15] which analyzes parallel interconnection schemes for binary codes.

Finally, we carry on a detailed investigation of the asymptotic analysis when the

signal-to-noise ratio (SNR) goes to infinity. Similarly to results in [4] for the binary case, we find that the leading term of both word and symbol error probabilities (for $N \to +\infty$) depends on the channel's SNR through a sort of distance parameter q^* which is a function of the constituent encoders. However, differently from the binary case, the exact characterization of q^* is given as the solution of a combinatorial optimization problem involving both constituent encoders.

Here, we give a brief outline of this paper. Sections 2 and 3 contain some basic notation and results on block and convolutional encoders over Abelian groups: special attention is devoted to homomorphic state-space realizations, which play a fundamental role in all our estimations. In Section 4 we introduce the serially concatenated schemes and the specific examples we want to focus on. Section 5 contains the fundamental original theoretical results of the paper: Theorem 5.4 describes the asymptotic estimations for word and symbol error probabilities when $N \to +\infty$ and, furtherly, when the SNR goes to infinity. In particular, the expression for the interleaver gain α is derived and the combinatorial optimization problem describing the effective distance q^* is formulated. All proofs are given in Section 6. Section 7 gives a detailed account of the computation of α and d^* for all our examples. An appendix containing some results on the algebraic properties of encoders over $\mathbb{Z}_m$ completes the paper.

2. Homomorphic block encoders for symmetric channels.

2.1. Notation. Given a set Ω and $A \subseteq \Omega$, the symbol $\mathbb{1}_A : \Omega \to \{0,1\}$ will denote the indicator function of A, i.e. $\mathbb{1}_A(x) = 1$ if and only if $x \in A$. $|A|$ will denote the cardinality of A. We will denote by $\mathbb{N}$ the set of non-negative integers, and by $\mathbb{N}^*$ the set of positive integers. Throughout this paper, vectors will always be column vectors, and will be denoted by boldface letters. Given a vector $\boldsymbol{w} \in \mathbb{N}^k$, we let $|\boldsymbol{w}| = \sum_j \boldsymbol{w}_j$. We will denote by $\boldsymbol{e}_j$ a vector of the appropriate length (clear by the context or explicitly stated) made by all zeros except a one in position j. Given two sets A, B, B^A will denote vectors with entries in B, having length $|A|$ and components indexed by elements of A instead of integers $1, \dots, |A|$. By log and exp we will denote logarithm and exponential with respect to the same base $b > 1$. Given groups G and H, $\mathrm{Hom}(G, H)$ will denote the group of all homomorphisms from G to H, while $\mathrm{Aut}(G)$ will be the group of automorphisms of G.

2.2. Symmetric channels. A memoryless channel is described by: an input alphabet $\mathcal{X}$ (which we will always assume is finite), an output alphabet $\mathcal{Y}$, endowed with a σ-algebra $\mathcal{B} \subseteq 2^{\mathcal{Y}}$ and a probability measure μ; a family of transition probability densities $W(\cdot|x)$ on $\mathcal{Y}$, indexed by the inputs $x \in \mathcal{X}$. Such a channel will be denoted by $(\mathcal{X}, \mathcal{Y}, W)$. In most applications, either $\mathcal{Y}$ is finite, and μ is the counting measure, so $W(\cdot|x)$ are simply probability vectors, or $\mathcal{Y} = \mathbb{R}^n$ and μ is the Lebesgue measure.

To give a formal definition of symmetric memoryless channels, we need to recall some definitions of group actions. Given a group $(G, +)$ with neutral element 0, and given a set A, G acts on A if for every $g \in G$ there exists a map $a \mapsto ga$ from A to A, such that $(h + g)a = h(ga)$ for all $h, g \in G$, and $a \in A$, and $0a = a$ for all $a \in A$. For finite A, the group action of G on A is said to be (simply) transitive if for every $a, b \in A$, there exists a (unique) element $g \in G$ such that $ga = b$. If G acts simply transitively on A, G and A are in bijection, through the map $\theta : G \to A$ defined by $\theta(g) = ga_0$ for an arbitrary but fixed $a_0 \in A$.

Given a probability space $\mathcal{Y}$, with σ-algebra $\mathcal{B}$ and probability measure μ, we say that a group G acts isometrically on $\mathcal{Y}$ if there exists an action of G on $\mathcal{Y}$ consisting

of measurable bijections such that $\mu(gA) = \mu(A) \forall A \in \mathcal{B}, \forall g \in G$. If $\mathcal{Y}$ is finite, then all group actions on $\mathcal{Y}$ are isometric. If $\mathcal{Y} = \mathbb{R}^n$, then an action is isometric when all maps $y \mapsto gy$ are isometries of $\mathbb{R}^n$.

Given a group G, a memoryless channel $(\mathcal{X}, \mathcal{Y}, W)$ is called G-symmetric if:

1. G acts simply transitively on $\mathcal{X}$;
2. G acts isometrically on $\mathcal{Y}$;
3. $W(y|x) = W(gy|gx)$ for every $g \in G$, $x \in \mathcal{X}$, $y \in \mathcal{Y}$.

In this case, the bijection $\theta : G \to \mathcal{X}$ defined by $\theta(g) = gx_0$ for some fixed $x_0 \in \mathcal{X}$ is called an isometric labelling.

The most common examples of G-symmetric channels are the following.

- **Binary-input output-symmetric channels.** $\mathbb{Z}_2$-symmetric channels are known in the coding literature as binary-input output-symmetric (BIOS) channels. Well-known examples are binary symmetric channel (BSC), binary erasure channel (BEC), and binary-input AWGN (BIAWGN) channel.
- **Geometrically uniform AWGN channels.** A n–dimensional constellation is a finite subset $S \subset \mathbb{R}^n$ that spans $\mathbb{R}^n$; we denote with $\Gamma(S)$ its symmetry group, i.e. the group of the Euclidean isometries of $\mathbb{R}^n$ mapping S into S itself. A constellation S is said to be geometrically uniform (GU) with generating group G if G is a subgroup of $\Gamma(S)$ whose action on S is simply transitive. The simplest example of GU constellation is the 1-dimensional antipodal constellation $\{-1, 1\}$ (a.k.a 2-points Pulse Amplitude Modulation, 2–PAM). A 2-dimensional example is the m-PSK constellation

$$S = \{e^{\frac{2\pi i l}{m}} \ : \ l = 0, \dots m-1\} \subseteq \mathbb{C} \simeq \mathbb{R}^2$$

  which always has the generating group $\mathbb{Z}_m$ (seen as rotations of angles multiple of $2\pi/m$) and for even m also has the non-Abelian generating group $D_{m/2}$. For a complete theory of GU constellations and generating groups, see [20] and [32].
  Given a GU constellation $S \subset \mathbb{R}^q$ with generating group G, define the *S–AWGN channel* as the memoryless channel $(S, \mathbb{R}^n, W)$ where the family W of n-dimensional transition densities is given by $W(y|x) = N(y - x)$ for any $x \in S$; $N(\cdot)$ is the density of a n-dimensional diagonal Gaussian r.v. $N(y) = (2\pi\sigma^2)^{-n/2} e^{-||y||^2/(2\sigma^2)}$.
  Other examples of G-symmetric channels can be obtained from the S-AWGN by suitable symmetric quantizations of the channel output, e.g. quantizing with respect to the Voronoi regions of the same constellation S.
- **m-ary symmetric channels.** This is a simple generalization of the BSC: $\mathcal{X} = \mathcal{Y} = \{0, 1, \dots, m-1\}$ and $W(y|x) = \epsilon/(m-1)$ if $y \neq x$, $W(y|x) = 1 - \epsilon$ if $y = x$. This channel is G-symmetric for any group G with $|G| = m$; in particular, for $G = \mathbb{Z}_m$.

Now fix a G-symmetric channel $(\mathcal{X}, \mathcal{Y}, W)$, and denote by W_n, μ_n and $\boldsymbol{\theta}$ the natural extensions of W, μ and θ to multiple uses of the channel. Define the pairwise equivocation probability of a word $\boldsymbol{c} \in G^n$, $\mathrm{P}(\mathbf{0} \to \boldsymbol{c})$, to be the probability that, for some fixed decoding rule, the decoder will prefer $\boldsymbol{c}$ to $\mathbf{0}$, given that $\boldsymbol{\theta}(\mathbf{0})$ was transmitted. In this paper, we consider maximum likelihood decoding, with the choice

to break ties uniformly at random, so that

$$\begin{aligned}\mathrm{P}(\mathbf{0}\to\boldsymbol{c}) &= \int_{\mathcal{Y}^n} W_n(\cdot|\boldsymbol{\theta}(\mathbf{0}))\mathbb{1}_{\{W_n(\cdot|\boldsymbol{\theta}(\boldsymbol{c}))>W_n(\cdot|\boldsymbol{\theta}(\mathbf{0}))\}}\,\mathrm{d}\mu_n \\ &\quad + \tfrac{1}{2}\int_{\mathcal{Y}^n} W_n(\cdot|\boldsymbol{\theta}(\mathbf{0}))\mathbb{1}_{\{W_n(\cdot|\boldsymbol{\theta}(\boldsymbol{c}))=W_n(\cdot|\boldsymbol{\theta}(\mathbf{0}))\}}\,\mathrm{d}\mu_n\,.\end{aligned}$$

Note that $\mathrm{P}(\mathbf{0}\to\boldsymbol{c})$ depends only on the number of occurrences in $\boldsymbol{c}$ of each non-zero element of G. We will call the type of $\boldsymbol{c}$, denoted $\boldsymbol{w}_{\mathrm{T}}(\boldsymbol{c})\in\mathbb{N}^{G\setminus\{0\}}$, a counter of such occurrences: $(\boldsymbol{w}_{\mathrm{T}}(\boldsymbol{c}))_g=\sum_{i=1}^n \mathbb{1}_{\{g\}}(\boldsymbol{c}_i)$. Given $\boldsymbol{w}\in\mathbb{N}^{G\setminus\{0\}}$, we will use the notation $Q(\boldsymbol{w})$ to denote $\mathrm{P}(\mathbf{0}\to\boldsymbol{c})$ for any $\boldsymbol{c}$ with $\boldsymbol{w}_{\mathrm{T}}(\boldsymbol{c})=\boldsymbol{w}$.

In the sequel, we will use the following two estimations of $\mathrm{P}(\mathbf{0}\to\boldsymbol{c})$, both involving $\mathrm{w}_{\mathrm{H}}(\boldsymbol{c})$, the Hamming weight of $\boldsymbol{c}$ (number or non-zero elements). The well-known Bhattacharyya upper bound is:

$$\begin{aligned}\mathrm{P}(\mathbf{0}\to\boldsymbol{c}) &\le \int_{\mathcal{Y}^n} W_n(\cdot|\boldsymbol{\theta}(\mathbf{0}))\mathbb{1}_{\{W_n(\cdot|\boldsymbol{\theta}(\boldsymbol{c}))\ge W_n(\cdot|\boldsymbol{\theta}(\mathbf{0}))\}}\,\mathrm{d}\mu_n \\ &\le \prod_{i=1}^n \int_{\mathcal{Y}} \sqrt{W(\cdot|\theta(0))W(\cdot|\theta(c_i))}\,\mathrm{d}\mu \\ &\le \gamma^{\mathrm{w}_{\mathrm{H}}(\boldsymbol{c})}\end{aligned}$$

where γ is the (worse) Bhattacharyya noise parameter of the channel defined as:

$$\gamma=\max_{g\neq 0}\int_{\mathcal{Y}}\sqrt{W(\cdot|\theta(0))W(\cdot|\theta(g))}\,\mathrm{d}\mu$$

A lower bound for pairwise equivocation probability is easily obtained:

$$\begin{aligned}\mathrm{P}(\mathbf{0}\to\boldsymbol{c}) &\ge \int_{\mathcal{Y}^n} W_n(\cdot|\boldsymbol{\theta}(\mathbf{0}))\mathbb{1}_{\{W_n(\cdot|\boldsymbol{\theta}(\boldsymbol{c}))>W_n(\cdot|\boldsymbol{\theta}(\mathbf{0}))\}}\,\mathrm{d}\mu_n \\ &\ge \prod_{i=1}^n \int_{\mathcal{Y}} W(\cdot|\theta(0))\mathbb{1}_{\{W(\cdot|\theta(c_i))>W(\cdot|\theta(0))\}}\,\mathrm{d}\mu \\ &\ge p^{\mathrm{w}_{\mathrm{H}}(\boldsymbol{c})}\end{aligned}$$

where p is the (worse) equivocation probability of the channel, defined as:

$$p=\min_{g\neq 0}\int_{\mathcal{Y}} W(\cdot|\theta(0))\mathbb{1}_{\{W(\cdot|\theta(c_i))>W(\cdot|\theta(0))\}}\,\mathrm{d}\mu$$

Let's see what these definitions give in the examples of G-symmetric channels we have presented.

- **BIOS channels.** The names Bhattacharyya parameter and equivocation probability for γ and p are mostly used only in this context, where there is only one non-zero $g\in G$ and so there is no maximization (resp. minimization) in the definition of γ (resp. p).
  For BSC with cross-over probability ϵ, if $\mathrm{w}_{\mathrm{H}}(\boldsymbol{c})=w$, $\mathrm{P}(\mathbf{0}\to\boldsymbol{c})=Q(w)$ with

$$Q(w)=\begin{cases}\sum_{r=\lceil w/2\rceil}^{w}\binom{w}{r}\epsilon^r(1-\epsilon)^{w-r} & \text{if } w \text{ is odd}\\ \sum_{r=1+w/2}^{w}\binom{w}{r}\epsilon^r(1-\epsilon)^{w-r}+\frac{1}{2}\binom{w}{w/2}\epsilon^{w/2}(1-\epsilon)^{w/2} & \text{if } w \text{ is even.}\end{cases}$$

The Bhattacharyya parameter is $\gamma = 2\sqrt{\epsilon(1-\epsilon)}$, while the equivocation probability is $p = \epsilon$.
For BEC with erasure probability ϵ, the only terms in $\mathrm{P}(\mathbf{0} \to \boldsymbol{c})$ come from breaking ties: if $\mathrm{w_H}(\boldsymbol{c}) = w$, $\mathrm{P}(\mathbf{0} \to \boldsymbol{c}) = Q(w) = \frac{1}{2}\epsilon^w$. Here $\gamma = \epsilon$, while $p = 0$.
For BIAWGN channel, see below.

- **S-AWGN channels.** Here, with $S \subset \mathbb{R}^d$ and codewords of length n,

$$\mathrm{P}(\mathbf{0} \to \boldsymbol{c}) = \int_{\mathbb{R}^{dn}} W_n(\cdot|\boldsymbol{\theta}(\mathbf{0}))\mathbb{1}_{\{W_n(\cdot|\boldsymbol{\theta}(\boldsymbol{c}))>W_n(\cdot|\boldsymbol{\theta}(\mathbf{0}))\}}\,\mathrm{d}\mu_n = \tfrac{1}{2}\operatorname{erfc}\left(\tfrac{\|\boldsymbol{\theta}(\boldsymbol{c})-\boldsymbol{\theta}(\mathbf{0})\|}{2\sqrt{2\sigma^2}}\right)$$

where $\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{+\infty} e^{-t^2}\mathrm{d}t$ and $\|\cdot\|$ denotes Euclidean norm.
It is well-known that all points of a geometrically uniform constellation lie on a sphere and it is usually assumed that constellations have barycenter in the origin, so the radius of the sphere, squared, is the signal energy per transmitted symbol E_s. This remark allows to find explicit dependence of $\mathrm{P}(\mathbf{0} \to \boldsymbol{c})$ on the SNR:

$$\mathrm{P}(\mathbf{0} \to \boldsymbol{c}) = \tfrac{1}{2}\operatorname{erfc}\left(\tfrac{\|\boldsymbol{\theta}(\boldsymbol{c})-\boldsymbol{\theta}(\mathbf{0})\|}{2\sqrt{E_s}}\sqrt{\tfrac{E_s}{N_0}}\right)$$

where E_s/N_0 is the signal-to-noise ratio per transmitted symbol.
Clearly, having fixed S and θ, $\mathrm{P}(\mathbf{0} \to \boldsymbol{c})$ is a function of $\boldsymbol{w}_{\mathrm{T}}(\boldsymbol{c})$ only: if $\boldsymbol{w}_{\mathrm{T}}(\boldsymbol{c}) = \boldsymbol{h}$, one can write $\|\boldsymbol{\theta}(\boldsymbol{c}) - \boldsymbol{\theta}(\mathbf{0})\| = \sqrt{\sum_g \boldsymbol{h}_g \|\theta(g) - \theta(0)\|^2}$. However, the expression for $\mathrm{P}(\mathbf{0} \to \boldsymbol{c})$ suggests to define a different weight, which better captures the geometry of S; it is a re-scaled squared Euclidean distance from zero, which we will call Euclidean weight: $\mathrm{w_E}(\boldsymbol{c}) = \|\boldsymbol{\theta}(\boldsymbol{c}) - \boldsymbol{\theta}(\mathbf{0})\|^2/4E_s$.
One can also compute $\gamma = \max_{g\neq 0} e^{-\|\theta(g)-\theta(0)\|^2/(8\sigma^2)} = \left(e^{-E_s/N_0}\right)^{\min_{g\neq 0}\mathrm{w_E}(g)}$.
Finally, $p = \min_{g\neq 0}\frac{1}{2}\operatorname{erfc}\left(\frac{\|\theta(g)-\theta(0)\|}{2\sqrt{2\sigma^2}}\right) = \frac{1}{2}\operatorname{erfc}\sqrt{\max_{g\neq 0}\mathrm{w_E}(g)\frac{E_s}{N_0}}$.

- **m-ary symmetric channels.** A symmetric channel with alphabet size m and error probability ϵ has

$$\mathrm{P}(\mathbf{0} \to \boldsymbol{c}) = \sum_{s=1}^{w}\tbinom{w}{s}[(1-\tfrac{1}{m-1})\epsilon]^{w-s}\sum_{r=\lfloor s/2\rfloor+1}^{s}\tbinom{s}{r}\left(\tfrac{\epsilon}{m-1}\right)^r(1-\epsilon)^{s-r}$$
$$+\tfrac{1}{2}\sum_{s=0}^{\lfloor w/2\rfloor}\tbinom{w}{2s}[(1-\tfrac{1}{m-1})\epsilon]^{w-2s}\tbinom{2s}{s}\left(\tfrac{\epsilon(1-\epsilon)}{m-1}\right)^s$$

In this case, $\gamma = \sqrt{\frac{\epsilon}{m-1}}\left(2\sqrt{1-\epsilon} + (m-2)\sqrt{\frac{\epsilon}{m-1}}\right)$ and $p = \frac{\epsilon}{m-1}$.

2.3. Weights. In this paper, we will deal with different kinds of weight: we have already seen Hamming, type and Euclidean weight, in relation with the channel's noise, and we will need other weights related to the coding schemes. In this section we present a general definition and we establish some useful properties.

Definition 2.1. *A weight on an Abelian group Z consists of a positive integer ρ and of a map $\boldsymbol{w} : Z \to \mathbb{N}^\rho$ satisfying the following properties:*

1. $\boldsymbol{w}(0) = \mathbf{0}$;
2. $|\boldsymbol{w}(z^1 + z^2)| \leq |\boldsymbol{w}(z^1)| + |\boldsymbol{w}(z^2)|$ *for every* $z^1, z^2 \in Z$;
3. $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_\rho\} \subseteq \boldsymbol{w}(Z)$ *(here* $\boldsymbol{e}_j \in \mathbb{N}^\rho$*).*

A few considerations on the above definition:

- Item 2 says that summation in Z can not create any extra weight;

- Item 3 is a simple minimality assumption which ensures that the full semigroup structure of $\mathbb{N}^\rho$ is used.

Whenever we have a weight $\boldsymbol{w}$ we will consider its natural extension to vectors by componentwise sum $\boldsymbol{w} : Z^N \to \mathbb{N}^\rho$, defined by $\boldsymbol{w}(\boldsymbol{z}) = \sum_j \boldsymbol{w}(z_j)$.

Given $\boldsymbol{h} \in \mathbb{N}^\rho$, we will use the following notation: $Z^N_{\boldsymbol{h}} = \{\boldsymbol{z} \in Z^N \,:\, \boldsymbol{w}(\boldsymbol{z}) = \boldsymbol{h}\}$. Moreover, if $\boldsymbol{h} \in \mathbb{N}^\rho$ we will use the notation

$$\binom{N}{\boldsymbol{h}} = \begin{cases} \frac{N!}{\boldsymbol{h}_1! \cdots \boldsymbol{h}_\rho! (N-|\boldsymbol{h}|)!} & \text{if } |\boldsymbol{h}| \le N \\ 0 & \text{otherwise} \end{cases}$$

The following result will be useful later

LEMMA 2.2. *Suppose $\boldsymbol{w}$ is a weight on Z. For every $\boldsymbol{h} \in \mathbb{N}^\rho$ we have that*

$$\binom{N}{\boldsymbol{h}} \le |Z^N_{\boldsymbol{h}}| \le \big(|Z|\, N\big)^{|\boldsymbol{h}|}$$

Proof. For $i = 1, \dots, \rho$, let $\eta_i \in Z$ be such that $\boldsymbol{w}(\eta_i) = \boldsymbol{e}_i \in \mathbb{N}^\rho$ (they surely exist by point 3 of definition of weight). The lower bound is trivially true if $|\boldsymbol{h}| > N$. Otherwise, consider the words in Z^N with support cardinality $|\boldsymbol{h}|$ made by exactly $\boldsymbol{h}_j$ times η_j, for $j = 1, \dots, \rho$: there are $\binom{N}{\boldsymbol{h}}$ such words, and all of them have invariants weight vector $\boldsymbol{h}$.

The upper bound is clearly true if $\boldsymbol{h} = \boldsymbol{0}$. Therefore assume that $\boldsymbol{h} \neq \boldsymbol{0}$ For any $\boldsymbol{z} \in Z^N_{\boldsymbol{h}}$ consider the subset $\mathcal{J}$ of indices $j \in \{1, \dots, N\}$ for which $\boldsymbol{z}_j \neq 0$. Clearly, $1 \le |\mathcal{J}| \le |\boldsymbol{h}|$. It thus follows that the number of elements in $Z^N_{\boldsymbol{h}}$ can be upper bound considering all possible subsets $\mathcal{J}$ of cardinality $1 \le |\mathcal{J}| \le |h|$ and all the possible elements of Z in the positions in $\mathcal{J}$. In other words

$$|Z^N_{\boldsymbol{h}}| \le \sum_{j=1}^{|\boldsymbol{h}|} \binom{N}{j} |Z|^j \le |Z|^{|\boldsymbol{h}|} \sum_{j=1}^{|\boldsymbol{h}|} \binom{N}{j}$$

It is now sufficient to use the inequality $\sum_{j=1}^{|\boldsymbol{h}|} \binom{N}{j} \le N^{|\boldsymbol{h}|}$ to obtain the result. ∎

Hamming weight and type weight are examples of weights always available on any set Z:

- **Hamming weight:** $\rho = 1$, $\mathrm{w_H}(z) = 1 - \mathbb{1}_{\{0\}}(z)$;
- **Type weight:** $\rho = |Z| - 1$, or better $\boldsymbol{w}_{\mathrm{T}}(z) \in \mathbb{N}^{Z \setminus \{0\}}$, because we prefer indexing the components of $\boldsymbol{w}_{\mathrm{T}}(z)$ directly by the elements in $Z \setminus \{0\}$ instead of by integers $1, \dots, |Z| - 1$; define $\boldsymbol{w}_{\mathrm{T}}(z)_a = \mathbb{1}_{\{a\}}(z)$ for every $a \in Z \setminus \{0\}$.

Clearly, $|\boldsymbol{w}_{\mathrm{T}}(\boldsymbol{z})| = \mathrm{w_H}(\boldsymbol{z})$ and moreover, for any weight $\boldsymbol{w}$ on Z, it necessary holds

$$\mathrm{w_H}(\boldsymbol{z}) \le |\boldsymbol{w}(\boldsymbol{z})| \le \mathrm{w_{max}} \mathrm{w_H}(\boldsymbol{z})\,.$$

where $\mathrm{w_{max}} = \max_{z \in Z} |\boldsymbol{w}(z)|$.

On Abelian groups, it will be particularly important to consider the weights compatible with the algebraic structure, as defined below.

DEFINITION 2.3. *Given an Abelian group U, a distance* d *on U is called compatible with the group structure of U if* $\mathrm{d}(u, v) = \mathrm{d}(u + w, v + w)$ *for all $u, v, w \in U$.*

A weight $\boldsymbol{w} : U \to \mathbb{N}$ is called compatible with the group U if there exists a distance d *compatible with U such that, for all $u \in U$, $\boldsymbol{w}(u) = \mathrm{d}(u, 0)$.*

Notice that if d is a compatible distance on U, the natural extension (by componentwise summation) on U^k remains compatible: it will be denoted with the same symbol d, as well as the associated weight w. Also notice that Hamming and type weights are always compatible with any fixed group U.

2.4. Homomorphic block encoders. Word and symbol error rate. We fix a finite Abelian group Γ and we consider transmission on a memoryless Γ-symmetric channel. Given another finite Abelian group U, we define a block encoder of rate k/n, over Γ with inputs in U, to be any injective group homomorphism $\phi : U^k \to \Gamma^n$; we define the corresponding code to be the image of the encoder.

We let ξ to be a r.v. uniformly distributed on U^k (ξ is the word to be sent) and independent from the channel noise. We let $\hat{\xi}$ to be the ML estimate of ξ from the received word y. In this setting, we can clearly define the word error probability of our code in the usual way:

$$P_w(e|\boldsymbol{u}) = \mathbb{P}(\hat{\xi} \neq \boldsymbol{u}|\xi = \boldsymbol{u}) \,, \qquad P_w(e) = \mathbb{P}(\hat{\xi} \neq \xi) = \frac{1}{|U|^k} \sum_{\boldsymbol{u}\in U^k} P_w(e|\boldsymbol{u}) \,.$$

Our assumptions ensure that the Uniform Error Property holds, i.e. the word error probability does not depend on which word has been sent and, in particular, we can assume that the all-zero word has been sent: $P_w(e) = P_w(e|\mathbf{0})$.

Another interesting property of a code (or, more precisely, of an encoder) is its bit error rate. In our abstract setting, it is more convenient to consider a symbol error rate, where symbols can be, for example, the elements of U or, as we will see, also something 'smaller'. We propose the following definition.

Given a distance d compatible with U and such that $\mathrm{d}(u, 0) \neq 0$ for all $u \neq 0$, we define a *symbol error rate* with respect to d as

$$P_s(e|\boldsymbol{u}) = \sum_{\hat{\boldsymbol{u}}\in U^k} \frac{\mathrm{d}(\hat{\boldsymbol{u}}, \boldsymbol{u})}{k\rho_U} \mathbb{P}(\hat{\xi} = \hat{\boldsymbol{u}}|\xi = \boldsymbol{u})$$

where ρ_U is the diameter of U with respect to d. Moreover, we put

$$P_s(e) = \frac{1}{|U|^k} \sum_{\boldsymbol{u}\in U^k} P_s(e|\boldsymbol{u}) \,.$$

The compatibility of the distance with U, together with the previous assumptions, ensures that also for $P_s(e)$ the Uniform Error Property holds true:

$$P_s(e) = P_s(e|\mathbf{0}) = \sum_{\hat{\boldsymbol{u}}\in U^k} \frac{\boldsymbol{w}(\hat{\boldsymbol{u}})}{k\rho_U} \mathbb{P}(\hat{\xi} = \hat{\boldsymbol{u}}|\xi = \mathbf{0}) \,,$$

where $\boldsymbol{w}$ is the weight associated with the distance d. In this case, $\rho_U = \max_{u\in U} \boldsymbol{w}(u)$, and we have the inequality

$$P_s(e) \geq \frac{1}{N} \frac{\min\limits_{u\in U, u\neq 0} \boldsymbol{w}(u)}{\max\limits_{u\in U} \boldsymbol{w}(u)} P_w(e)$$

When d and $\boldsymbol{w}$ are Hamming distance and weight respectively, the above definition simply gives the usual Symbol Error Rate, where symbols are elements in U, and if $U = \mathbb{Z}_2$ this is the classical Bit Error Rate. When $U = \mathbb{Z}_2^a$, in addition to the Symbol Error Rate, we can find also the Bit Error Rate taking as distance the number of different bits (Hamming weight in $(\mathbb{Z}_2^a)^k$ identified with $\mathbb{Z}_2^{ak}$) instead of the number of different symbols.

3. Convolutional encoders over Abelian groups. In this section we will recall some basic facts of the theory of convolutional codes over Abelian groups which will be needed for the sequel. Further details can be found in [30, 17, 18, 19] and the reference therein.

3.1. State maps and error events. Let U and Y be two finite Abelian groups. Consider the spaces of sequences $U^{\mathbb{N}}$ and $Y^{\mathbb{N}}$, respectively, both equipped with the componentwise group structure. Convolutional codes will be for us homomorphic maps $\phi : U^{\mathbb{N}} \to Y^{\mathbb{N}}$ satisfying some properties which are introduced below. In coding theory the only maps between sequence spaces which are really relevant are those which admit a realization through finite state maps.

A *(homomorphic) state map* η from $U^{\mathbb{N}}$ to $Y^{\mathbb{N}}$ consists in another Abelian group X and in four homomorphisms

$$\begin{array}{ll} F : X \to X\,, & L : U \to X \\ R : X \to Y\,, & S : U \to Y \end{array}$$

X is called the *state space* of the state map and if X is finite, then the state map is said to be a finite state map. A state map is formally denoted by the quadruple $\eta = (F, L, R, S)$. A finite state map can be pictorially described by a trellis, in the usual way: at each time step, we draw vertices corresponding to the elements of X, then we draw an edge from vertex x at time t to vertex x' at time $t+1$, with input tag u and output label y if and only if $x' = Fx + Lu$ and $y = Rx + Su$.

Given a homomorphic state map η and a state $x \in X$, we can define a map $\eta_x : U^{\mathbb{N}} \to Y^{\mathbb{N}}$ mapping $\mathbf{u} \in U^{\mathbb{N}}$ into $\mathbf{y} = \eta_x(\mathbf{u})$ computed recursively starting from the initial condition $x_0 = x$, as follows:

$$\left\{ \begin{array}{rcll} x_{t+1} & = & Fx_t + Lu_t & \\ y_t & = & Rx_t + Su_t & \end{array} \right. \qquad \forall t \in \mathbb{N}. \tag{3.1}$$

Explicitly, we can write

$$y_t = RF^t x + R\sum_{j=1}^{t} F^j L u_{t-j} + Su_t\,. \tag{3.2}$$

Notice that η_0 is a homomorphism.

A homomorphic map $\phi : U^{\mathbb{N}} \to Y^{\mathbb{N}}$ is said to be a *convolutional encoder* if there exists a homomorphic finite state map $\eta = (F, L, R, S)$ such that $\phi = \eta_0$. In this case η is said to be a *state space realization* of ϕ. Given a convolutional encoder $\phi : U^{\mathbb{N}} \to Y^{\mathbb{N}}$, there may exist many homomorphic finite state maps realizing ϕ. A state map η is said to be a minimal realization of ϕ if it has the state space with the minimal number of states among the possible realizations of ϕ. An important consequence of minimality ([13, p. 48] or [27, p. 192]) are the following properties:

- **Observability:** Let $\mathbf{u}', \mathbf{u}'' \in U^{\mathbb{N}}$ and $\mathbf{x}', \mathbf{x}'' \in X^{\mathbb{N}}$ be such that both pairs $(\mathbf{u}', \mathbf{x}')$ and $(\mathbf{u}', \mathbf{x}')$ satisfy the first relation of (3.1). Let $\mathbf{y}', \mathbf{y}''$ be the corresponding output sequences. Then, if ${\mathbf{u}'}_t = {\mathbf{u}''}_t$ and ${\mathbf{y}'}_t = {\mathbf{y}''}_t$ for all $t = 0, \ldots, |X| - 1$, necessarily, it must hold ${\mathbf{x}'}_0 = {\mathbf{x}''}_0$.
- **Reachability:** For any x, $x' \in X$ there exist $t \leq |X| - 1$, $\mathbf{u} \in U^{\mathbb{N}}$ and $\mathbf{x} \in X^{\mathbb{N}}$ satisfying the first relation in (3.1) such that $\mathbf{x}_0 = x$ and $\mathbf{x}_t = x'$. The smallest t for which this condition holds for any x, $x' \in X$ is called the *reachability index* of η and denoted by ν.

From now on, whenever we are considering a convolutional encoder $\phi : U^{\mathbb{N}} \to Y^{\mathbb{N}}$, we will assume that an underlying minimal state space representation η has been fixed once and for all: in particular, to any given $\boldsymbol{u} \in U^{\mathbb{N}}$, and initial state x we can unambiguously associate a state sequence $\boldsymbol{x} \in X^{\mathbb{N}}$. Whenever the initial state x is not explicitly mentioned, we assume that $x = 0$. Notice moreover that x_t only depends on $\boldsymbol{u}$ up to time $t-1$.

We now define the key concept of error event.

DEFINITION 3.1. *Let $\boldsymbol{u} \in U^{\mathbb{N}}$ be an input sequence with associated state sequence $\boldsymbol{x}$. $\boldsymbol{u}$ is said to be an input error event for ϕ if there exist $t_1 \leq t_2$ such that:*

(i) *$u_t = 0$ for all $t < t_1$ and $t > t_2$;*
(ii) *$x_t = 0$ for all $t \leq t_1$ and $t > t_2$;*
(iii) *$x_t \neq 0$ for all $t \in]t_1, t_2]$.*

The corresponding codeword $\boldsymbol{y} = \phi(\boldsymbol{u})$ is said to be an error event. We call $[t_1, t_2]$ the active window and $t_2 - t_1$ the length of the (input) error event and we denote it by $l(\boldsymbol{u})$ or by $l(\boldsymbol{y})$.

The following property shows that the length of an error event cannot grow unbounded. We omit the proof since it is a straightforward generalization of Lemma 20 in [15] (binary case) using the observability property of the minimal realization.

PROPOSITION 3.2. *Given a convolutional encoder $\phi : U^{\mathbb{N}} \to Y^{\mathbb{N}}$, there exists a constant $L > 0$ such that any error event $\boldsymbol{u}$ has length $l(\boldsymbol{u}) \leq L\,(w_H(\boldsymbol{u}) + w_H(\phi(\boldsymbol{u}))$.*

The *support* of a sequence $\boldsymbol{u} \in U^{\mathbb{N}}$ is defined by

$$\mathrm{supp}(\boldsymbol{u}) = \{t \in \mathbb{N} \ : \ u_t \neq 0\}\,.$$

$\boldsymbol{u}$ is said to have *finite support* if its support has finite cardinality. Notice that the cardinality of the support of a sequence coincides with Hamming weight.

3.2. Laurent series formalism. In many situations the description of a convolutional encoder through a state representation or the corresponding trellis is sufficient and quite appropriate. As in the classical binary case there are also more algebraic but equivalent ways to describe convolutional codes which, on the other hand, turn out to be quite useful in investigating concepts like recursiveness, non-catastrophicity etc. This is what we are going to do next.

Given a group U, we consider the group of *Laurent series*

$$U((D)) = \left\{ \sum u_k D^k \ : \ u_k \in U\,, \exists\, k_0 \in \mathbb{Z}\ u_k = 0\ \forall k < k_0 \right\}\,.$$

Inside $U((D))$ there are two relevant subgroups: the polynomials $U[D]$ and the usual formal power series $U[[D]]$.

Relation (3.2), for $x = 0$, can be interpreted as a multiplicative operator (product being defined in the Cauchy style) from $U((D))$ to $G((D))$ with the multiplicative symbol given by

$$\phi(D) = \sum_{j=1}^{\infty} (RF^j L) D^j + S \in \mathrm{Hom}(U, Y)[[D]]\,. \tag{3.3}$$

$\phi(D)$ is called the *transfer function* associated with ϕ. Conversely, given a generic $\phi(D) \in \mathrm{Hom}(U, Y)[[D]]$, we can ask if it is the transfer function of a convolutional encoder. The answer is that this is true if and only if $\phi(D)$ is rational. Rationality is defined similarly to the field case. Consider the ring $\mathbb{Z}((D))$ of Laurent series with coefficients in $\mathbb{Z}$. The invertible elements in $\mathbb{Z}((D))$ are those Laurent series

whose trailing coefficient is equal to 1 or -1: we denote this subset with the symbol $\mathbb{Z}((D))^*$. Given any Abelian group U, $U((D))$ is naturally a $\mathbb{Z}((D))$–module. We define the submodule of *rational elements* of $U((D))$ as

$$U(D) = \{u(D) \in U((D)) \ : \ \exists p(D) \in \mathbb{Z}[D] \cap \mathbb{Z}((D))^* \, , \ p(D)u(D) \in U[D]\} \, .$$

Notice that rational Laurent series can always be represented as a fraction $u(D) = p(D)^{-1}v(D)$ for some suitable polynomials $p(D) \in \mathbb{Z}[D] \cap \mathbb{Z}((D))^*$ and $v(D) \in U[D]$. The assumption on $p(D)$ is exactly to make sure that $1/p(D)$ is a meaningful element of $\mathbb{Z}((D))$. It can be proven that $\phi(D) = \sum_{k=0}^{\infty} \phi_k D^k \in \mathrm{Hom}(U, Y)[[D]]$ is the transfer function of a convolutional encoder if and only if it is rational (see Proposition 5.2 in [17]). Rationality has a useful characterization at the level of the underlying sequence ϕ_k: it is equivalent to the fact that ϕ_k is periodic for sufficiently large k. A special type of convolutional encoders are the polynomial ones, namely those for which $\phi(D) \in \mathrm{Hom}(U, Y)[D]$.

In the sequel we will often 'confuse' the group sequence $U^{\mathbb{N}}$ with the formal power series $U[[D]]$ through the one-to-one correspondence

$$(u_t)_{t \in \mathbb{N}} \leftrightarrow \sum_t u_t D^t \, .$$

In particular $u_0 D^{t_0}$ will often be used to denote the sequence $\boldsymbol{u}$ which is equal to u_0 at time t_0 and equal to 0 otherwise. Notice that finite support sequences are in this way represented by polynomials in D and polynomial encoders transform polynomials into polynomials.

3.3. Properties of convolutional encoders. In this section we describe how some classical properties can be generalized to our setting; we will need them in analyzing our concatenated schemes. Some further properties, specific for the case when the input and output groups are free $\mathbb{Z}_m$–modules, will be given in the appendix.

3.3.1. Non-catastrophicity. The classical definition of non-catastrophic encoders is the following.

Definition 3.3. *A convolutional encoder $\phi : U^{\mathbb{N}} \to V^{\mathbb{N}}$ is* non-catastrophic *if, for all $\boldsymbol{u} \in U^{\mathbb{N}}$, $\mathrm{w_H}(\phi(\boldsymbol{u})) < \infty$ implies that $\mathrm{w_H}(\boldsymbol{u}) < \infty$.*

An useful remark is that systematic encoders are surely non-catastrophic. Also, non-catastrophic encoders have the following nice characterization (direct consequence of [19, Coroll. 1, p. 41])

Proposition 3.4. *Let $\phi : U^{\mathbb{N}} \to V^{\mathbb{N}}$ be a convolutional encoder. The following conditions are equivalent:*

1. *ϕ is non-catastrophic;*
2. *ϕ admits a polynomial left inverse;*
3. *there exists $\mu > 0$ suchthat, for all $\boldsymbol{u} \in U^{\mathbb{N}}$ it holds $\mathrm{w_H}(\boldsymbol{u}) \leq \mu \, \mathrm{w_H}(\phi(\boldsymbol{u}))$.*

Notice that condition 2 gives a practical tool for testing if an encoder is non-catastrophic, and it also shows that non-catastrophicity is a property stronger than injectivity. Instead condition 3 is a sort of continuity reformulation.

3.3.2. Recursiveness. Binary convolutional encoders are defined to be recursive when no input word with Hamming weight one can give a finite-weight output; this property can be easily generalized to our setting.

Definition 3.5. *Given a weight $\boldsymbol{w} : U \to \mathbb{N}^{\rho}$, a convolutional encoder $\phi : U^{\mathbb{N}} \to Y^{\mathbb{N}}$ is $\boldsymbol{w}$–recursive if, for all $\boldsymbol{u} \in U^{\mathbb{N}}$ such that $|\boldsymbol{w}(\boldsymbol{u})| = 1$, it holds $\mathrm{w_H}(\phi(\boldsymbol{u})) = +\infty$.*

When $\boldsymbol{w}$ is the Hamming weight, this is the usual definition of recursiveness.

See the appendix for a characterization of recursive encoders on free $\mathbb{Z}_m$–modules which allows to easily test for recursiveness.

3.3.3. Small input-weight codewords. All convolutional encoders, including the recursive ones, admit non-zero finite support input sequences whose image also has finite support. This fact is obvious from the rationality property. Indeed, if the transfer function $\phi(D)$ is of type $\phi(D) = p(D)^{-1}\phi'(D)$ where $p(d) \in \mathbb{Z}((D))^* \cap \mathbb{Z}[D]$ and $\phi'(D) \in \mathrm{Hom}(U, Y)[D]$, we can observe that any polynomial input of type $u(D) = p(D)v(D)$ for some $v(D) \in U[D]$ is transformed into another polynomial $\phi(D)u(D) = \phi'(D)v(D)$.

We now present a sharper result which shows how to construct input sequences with support of cardinality 2, whose image has finite support: this will be useful later on.

PROPOSITION 3.6. *Let $\phi : U^{\mathbb{N}} \to Y^{\mathbb{N}}$ be e a convolutional encoder and let $u_1, \dots, u_r \in U$ be such that $\sum u_i = 0$. We can find time instants $t_1, \dots t_r$ such that given $\boldsymbol{u} = \sum u_j D^{t_j}$ we have that $\phi(\boldsymbol{u})$ has finite support.*

Proof. Consider the transfer function $\phi(D) = \sum_k \phi_k D^k$. By rationality we know that there exists $k_0 \in \mathbb{N}$ and $T \in \mathbb{N}$ such that $\phi_k = \phi_{k+T}$ for every $k \geq k_0$. Consider now the input sequence $\boldsymbol{u} = \sum_j u_j D^{(j-1)T}$. We have that

$$\phi(\boldsymbol{u})_t = \sum_{j=1}^{r} \phi_{t-(j-1)T} u_j$$

Notice that if we choose $t \geq k_0 + (r-1)T$, we easily obtain that $\phi_{t-(j-1)T} = \phi_t$ for every j so that, $\phi(\boldsymbol{u})_t = 0$. This proves the result. □

From the above result we obtain as an immediate corollary the following property, well-known at least for the binary case.

PROPOSITION 3.7. *Given a recursive convolutional encoder $\phi : U^{\mathbb{N}} \to Y^{\mathbb{N}}$, there exists $\delta \in \mathbb{N}$ such that, for any $u \in U$ the input sequence $\boldsymbol{u} = u - uD^{\delta}$ is an error event.*

3.3.4. Free distance. In the classical analysis by Benedetto et al. [4], an essential design parameter is the free distance of the outer encoder. When the concatenating group is not the group of all permutations (the classical uniform interleaver), we have to consider a slightly different parameter: instead of taking the minimum Hamming weight among non-zero outer codewords, we minimize some other suitable weight.

DEFINITION 3.8. *Given a convolutional encoder $\phi : U^{\mathbb{N}} \to Y^{\mathbb{N}}$ and a weight $\boldsymbol{w} : Y^{\mathbb{N}} \to \mathbb{N}^{\rho}$, we define the $\boldsymbol{w}$–free distance of ϕ to be*

$$d_f(\phi, \boldsymbol{w}) = \min\{|\boldsymbol{w}(\boldsymbol{c})| : \boldsymbol{c} = \phi(\boldsymbol{u}), \boldsymbol{u} \in U^{\mathbb{N}}, \boldsymbol{u} \neq \boldsymbol{0}\}$$

The classical free distance is the $\mathrm{w_H}$–free distance.

3.4. Terminated convolutional encoders. Suppose $\phi : U^{\mathbb{N}} \to Y^{\mathbb{N}}$ is a convolutional encoder with minimal state space X. We now define the terminated block codes associated with ϕ as follows.

Fix $N \in \mathbb{N}^*$. Given a vector $\boldsymbol{u} = (u_0, \dots, u_{N-1}) \in U^N$, let x_N be the corresponding state at time N. Because of the reachability condition it is possible to find input elements $\tilde{u}_N, \dots, \tilde{u}_{N+\nu-1}$ such that the state at time $N + \nu - 1$ is equal to 0.

This input string may not be unique and we assume we have fixed a specific one as a function of the terminal state x_N we had reached in such a way that the mapping

$$x_N \mapsto (\tilde{u}_N, \dots, \tilde{u}_{N+\nu-1})$$

is a homomorphism. It is a straightforward algebraic verification that this is always indeed possible. Given $\boldsymbol{u} = (u_0, \dots, u_{N-1}) \in U^N$ we now consider the associated input sequence $\tilde{\boldsymbol{u}} = (u_0, \dots, u_{N-1}, \tilde{u}_N, \dots, \tilde{u}_{N+\nu-1}, 0, 0, \dots)$. We then define the *N-terminated block encoder* as

$$\phi^N : U^N \to Y^{N+\nu}, \quad \phi^N(\boldsymbol{u}) = \phi(\tilde{\boldsymbol{u}})|_{[0,N+\nu-1]}.$$

For the assumptions made, ϕ^N is also a homomorphism. $\mathcal{C}^N = \operatorname{Im}\phi^N$ is called the *N-block code* associated to ϕ^N.

An input vector $\boldsymbol{u} \in U^N$ is an input error event for ϕ^N if $\tilde{\boldsymbol{u}} \in U^{\mathbb{N}}$ is an input error event for ϕ. In this case $\boldsymbol{c} = \phi^N(\boldsymbol{u})$ is called an error event for ϕ^N. Suppose the active window of $\tilde{\boldsymbol{u}}$ is equal to $[t_1, t_2]$. Then, the (input) error event is said to be *regular* if $t_2 \leq N$, otherwise it is called *terminated*. For a terminated error event, we call $N - t_1$ its length.

Notice that any codeword $\boldsymbol{c} \in \mathcal{C}^N$ can be written as $\boldsymbol{c} = \sum_{j=1}^{n+1} \boldsymbol{c}_j$ where $\boldsymbol{c}_j$ are regular error events for $j = 1, \dots, n$ and $\boldsymbol{c}_{n+1}$ is either zero or a terminating error event and the active windows of all these events are disjoint. We will use the notation $n(\boldsymbol{c})$ to denote the number of regular error events in the above decomposition of $\boldsymbol{c}$. Also notice that the above decomposition is unique, up to a permutation of the regular error events.

Some codewords have a decomposition in error events which is the same up to shifts of their error events, and for this reason share many important properties. More formally, we propose the following definition:

- two error events $\boldsymbol{c} = \phi^N(\boldsymbol{u})$ and $\boldsymbol{c}' = \phi^{N'}(\boldsymbol{u}')$ (notice that possibly $N \neq N'$) are said to be shift equivalent if the corresponding extended inputs $\tilde{\boldsymbol{u}}$, $\tilde{\boldsymbol{u}}' \in U^{\mathbb{N}}$ differ only by a shift.
- two codewords $\boldsymbol{c} = \phi^N(\boldsymbol{u})$ and $\boldsymbol{c}' = \phi^{N'}(\boldsymbol{u}')$ are said to be shift equivalent if there exist error event decompositions $\boldsymbol{c} = \sum_{j=1}^{n+1} \boldsymbol{c}_j$ and $\boldsymbol{c}' = \sum_{i=1}^{n'+1} \boldsymbol{c}'_i$ such that $\boldsymbol{c}_i$ and $\boldsymbol{c}'_i$ are shift equivalent for all i.

Notice that, given two shift equivalent codewords $\boldsymbol{c}$ and $\boldsymbol{c}'$, clearly $n(\boldsymbol{c}) = n(\boldsymbol{c}')$ and moreover, given a weight $\boldsymbol{w}$ on the alphabet Y, $\boldsymbol{w}(\boldsymbol{c}) = \boldsymbol{w}(\boldsymbol{c}')$.

REMARK 3.9. *Now we want to underline a property which is somehow similar to an inclusion of $\mathcal{C}^N$ in $\mathcal{C}^{N'}$ for $N \leq N'$ (while strictly speaking an inclusion cannot occur, as the two codes are subsets of different spaces). If $N \leq N'$, for all $\boldsymbol{c} \in \mathcal{C}^N$ we can construct $\boldsymbol{c}' \in \mathcal{C}^{N'}$ such that $\boldsymbol{c}$ and $\boldsymbol{c}'$ are shift equivalent, by properly adding zeros.*

3.5. Enumerating functions and growth estimates. A fundamental concept for all encoders is the so called weight enumerating function, since it is well known to play a basic role in all performance evaluations. While in the binary case, there is only one possible weight to be considered, namely the Hamming one, in our setting many choices are possible and we will need to consider different possibilities in later sections. We start defining the basic one based on the Hamming weights in the input and output groups.

DEFINITION 3.10. *Given a convolutional encoder $\phi : U^{\mathbb{N}} \to Y^{\mathbb{N}}$, consider its associated N-terminated block encoder $\phi^N : U^N \to Y^{N+\nu}$. Define its input/output*

support enumerating coefficients as:

$$\Lambda^N_{w,d,n} = |\{\boldsymbol{u} \in U^N : \mathrm{w_H}(\boldsymbol{u}) = w,\, \mathrm{w_H}(\phi^N(\boldsymbol{u})) = d,\, n(\phi^N(\boldsymbol{u})) = n\}|$$

In some cases, we will need to replace the Hamming weight with other possible weights in the input and in the output. We will use the notation $A^N_{\boldsymbol{w},\boldsymbol{d},n}$ to denote enumerating coefficients relative to some specified input weight $\boldsymbol{w}$ and output weight $\boldsymbol{d}$.

The following proposition gives a growth estimation for input/output support enumerating coefficients: this will allow us to have general bounds (even if quite loose) on all the different weight enumerators. We omit the proof, which is a straightforward generalization of Proposition 10 in [15].

PROPOSITION 3.11. *There exist two positive constants a and b such that*

$$\Lambda^N_{w,d,n} \leq \binom{N+n}{n} a^w b^d \,.$$

4. Serial ensembles.

4.1. Serial interconnections. We now precisely define the serial interconnected schemes we are going to consider. We start with a Γ-symmetric channel. We also fix the *input* Abelian group U. All encoders we will consider will be driven by words on U and will output symbols in Γ. The interconnection will take place through a third Abelian group, say Y called the *interconnection* group. We now fix two convolutional encoders denoted, respectively, the *outer* and *inner* encoder:

$$\phi_o : U^{\mathbb{N}} \to (Y^r)^{\mathbb{N}}\,, \quad \phi_i : (Y^s)^{\mathbb{N}} \to (\Gamma^l)^{\mathbb{N}}\,.$$

Denote by ν_o and ν_i the reachability indices of ϕ_o and ϕ_i respectively, and define the set $\mathcal{N} = \{N \in \mathbb{N}^* \,:\, s | r(N+\nu_o)\}$. Consider now the terminations, for $N \in \mathcal{N}$:

$$\phi^N_o : U^N \to Y^{r(N+\nu_o)}\,, \quad \phi^N_i : Y^{sM_N} \to \Gamma^{l(M_N+\nu_i)}\,,$$

where $sM_N = r(N+\nu_o)$. We now fix, for every $N \in \mathbb{N}^*$, a subgroup $G_N \subseteq \mathrm{Aut}(Y^{r(N+\nu_o)})$. The triple $(\phi_o, \phi_i, (G_N)_{N\in\mathcal{N}})$ is said to be a *serial interconnected ensemble*. The asymptotic rate of the serial interconnected ensemble above is defined by the product

$$R = \frac{\log|U|}{r}\frac{s}{l} \text{ bits per channel use}\,.$$

To the serial interconnected ensemble above we can associate a random sequence of encoders and codes as follows. Define Π_N to be a r.v. uniformly distributed over G_N and consider the corresponding homomorphic encoder $\Phi^N = \phi^N_i \circ \Pi_N \circ \phi^N_o$ and group code $\mathcal{C}^N = \mathrm{Im}(\Phi^N)$: they are called, respectively, the *random encoder* and the *random code* associated with the given ensemble.

The following picture describes the above construction:

$$\xrightarrow{U^N} \boxed{\phi^N_o} \xrightarrow{Y^{r(N+\nu_o)}} \boxed{\pi_N} \xrightarrow{Y^{sM_N}} \boxed{\phi^N_i} \xrightarrow{\Gamma^{l(M_N+\nu_i)}}$$

In the sequel we will denote by $\mathbb{P}$ and $\mathbb{E}$ probability and expected value, respectively, made with respect to the probabilistic space underlying the sequence Π_N. We will also use the notation $\overline{P_w(e)}$ and $\overline{P_s(e)}$, respectively, for the average word and symbol error probabilities.

4.2. Regular ensembles. Our aim is to give asymptotic results for $\overline{P_w(e)}$ and $\overline{P_s(e)}$ when $N \to \infty$, keeping fixed the constituent encoders. To do so, we need to make further assumptions on the groups G_N: roughly, we need to enforce some compatibility among the groups as N varies and that the number of invariants of the group action does not grow with N. Following [15] we propose the following definition.

DEFINITION 4.1. *The sequence of groups G_N (and the corresponding ensemble) is said to be regular if there exists a weight $\boldsymbol{w}_G : Y \to \mathbb{N}^\rho$ such that, for every N and for all $\boldsymbol{y}$, $\boldsymbol{z}, \in Y^{r(N+\nu_0)}$, it holds*

$$\boldsymbol{w}_G(\boldsymbol{y}) = \boldsymbol{w}_G(\boldsymbol{z}) \Leftrightarrow \quad \exists \sigma \in G_N \ : \sigma \boldsymbol{y} = \boldsymbol{z}$$

$\boldsymbol{w}_G(\boldsymbol{y})$ will be called the invariants weight vector of $\boldsymbol{y} \in Y^{r(N+\nu_0)}$.

Property of regularity simply says that all actions of the groups G_N on the sets $Y^{r(N+\nu_0)}$ can be described through a finite (constant) family of invariants: the ρ components of the weight $\boldsymbol{w}_G$. We will use the notation $Y^L_{\boldsymbol{h}} = \{\boldsymbol{x} \in Y^L : \boldsymbol{w}_G(\boldsymbol{x}) = \boldsymbol{h}\}$. Moreover we denote by $G_N(\boldsymbol{y}, \boldsymbol{z})$ the subset of elements in G_N mapping $\boldsymbol{y}$ to $\boldsymbol{z}$. Using standard results on group actions (the class formula) [26], we can show that

REMARK 4.2.

$$\frac{|G_N(\boldsymbol{u}, \boldsymbol{v})|}{|G_N|} = \begin{cases} 0 & \text{if } \boldsymbol{w}_G(\boldsymbol{u}) \neq \boldsymbol{w}_G(\boldsymbol{v}) \\ 1/|Y^{r(N+\nu_0)}_{\boldsymbol{h}}| & \text{if } \boldsymbol{w}_G(\boldsymbol{u}) = \boldsymbol{w}_G(\boldsymbol{v}) = \boldsymbol{h} \end{cases}$$

This technical result will be needed later.

LEMMA 4.3. *Assume that $\boldsymbol{y}, \boldsymbol{z} \in Y^{r(N+\nu_o)}$ are such that for every index $i \in \{0, \dots, N + \nu_o - 1\}$, $\boldsymbol{y}_i \neq \boldsymbol{0}$ yields $\boldsymbol{z}_i = \boldsymbol{0}$. Then, given any $\sigma \in G_N$ and given any i we have that $(\sigma \boldsymbol{y})_i \neq \boldsymbol{0} \ \Rightarrow \ (\sigma \boldsymbol{z})_i \neq (\sigma \boldsymbol{y})_i$.*

Proof. Notice that $|\boldsymbol{w}_G(\sigma \boldsymbol{y} + \sigma(-\boldsymbol{z}))| = |\boldsymbol{w}_G(\boldsymbol{y} - \boldsymbol{z})| = |\boldsymbol{w}_G(\boldsymbol{y})| + |\boldsymbol{w}_G(-\boldsymbol{z})|$. On the other hand, if $\sigma \boldsymbol{y}$ and $\sigma \boldsymbol{z}$ were equal in an index where they are not equal to $\boldsymbol{0}$, by point 2 of Definition 2.1, we would have

$$|\boldsymbol{w}_G(\sigma \boldsymbol{y} + \sigma(-\boldsymbol{z}))| < |\boldsymbol{w}_G(\sigma \boldsymbol{y})| + |\boldsymbol{w}_G(\sigma(-\boldsymbol{z}))| = |\boldsymbol{w}_G(\boldsymbol{y})| + |\boldsymbol{w}_G(-\boldsymbol{z})| .$$

This ends the proof. □

We now present two fundamental examples of regular actions.

1. *Symbol permutation* In this case we simply take $G_N = S_{r(N+\nu_0)}$ the full symmetric group acting on $Y^{r(N+\nu_0)}$ by standard permutation. In this case the invariant weight is the type weight: $\rho = |Y| - 1$ and $\boldsymbol{w}_G(y) \in \mathbb{N}^\rho$ by $(\boldsymbol{w}_G(y))_a = \mathbb{1}_a(y)$ as a varies in $Y \setminus \{0\}$. Notice that

$$\left| Y^{r(N+\nu_o)}_{\boldsymbol{h}} \right| = \binom{r(N + \nu_o)}{\boldsymbol{h}} .$$

2. *Separate channels symbol permutation* Assume $Y = Y_1 \times Y_2$ and assume $G_{N,1}$ and $G_{N,2}$ are sequences of groups acting regularly on $Y_1^{r(N+\nu_0)}$ and $Y_2^{r(N+\nu_0)}$ respectively with invariant weights $\boldsymbol{w}^1_G : Y_1 \to \mathbb{N}^{\rho_1}$ and $\boldsymbol{w}^2_G : Y_2 \to \mathbb{N}^{\rho_2}$ respectively. Then, we can consider a regular action given by $G_N = G_{N,1} \times G_{N,2}$ acting componentwise on $Y_1 \times Y_2$. Its invariant weight is given by $\boldsymbol{w}_G : Y_1 \times Y_2 \to \mathbb{N}^{\rho_1 + \rho_2}$, $\boldsymbol{w}_G(y_1, y_2) = (\boldsymbol{w}^1_G(y_1), \boldsymbol{w}^2_G(y_2))$. Notice that in this case

$$\left| Y^{r(N+\nu_o)}_{\boldsymbol{h}_1, \boldsymbol{h}_2} \right| = \binom{r(N + \nu_o)}{\boldsymbol{h}_1} \binom{r(N + \nu_o)}{\boldsymbol{h}_2} .$$

4.3. Examples of serial ensembles. Below we assume $\Gamma = \mathbb{Z}_m$.

- *Repeat-Convolute codes.* We choose $U = Y = \mathbb{Z}_m$ and $\phi_o = \mathrm{Rep}_r : \mathbb{Z}_m^{\mathbb{N}} \to (\mathbb{Z}_m^r)^{\mathbb{N}}$ to be the r-repetition encoder $\mathrm{Rep}_r(\boldsymbol{u})_t = (u_t, \dots u_t)$. We let $\phi_i : (\mathbb{Z}_m^s)^{\mathbb{N}} \to (\mathbb{Z}_m^s)^{\mathbb{N}}$ be a rate-1 non-catastrophic convolutional encoder. Finally we choose for the coupling interleavers the symbol permutation groups $G_N = S_{rN}$. The corresponding invariant weight is thus the type weight $\boldsymbol{w}_{\mathrm{T}} : \mathbb{Z}_m^r \to \mathbb{N}^{m-1}$, where $(\boldsymbol{w}_{\mathrm{T}}(\boldsymbol{y}))_j$ is the number of elements equal to j in the vector $\boldsymbol{y} \in \mathbb{Z}_m^r$. The rate of the scheme is

$$R = \frac{\log m}{r} \text{ bits/ch. use.}$$

  For this ensemble, we will need the assumption that ϕ_i is $\boldsymbol{w}$-recursive, which is the same as asking it is $\mathrm{w_H}$-recursive.
- *Structured LDPC codes.* We choose $U = \mathbb{Z}_m$, $Y = \mathbb{Z}_m^c \times \mathbb{Z}_m$ and the systematic encoder $\phi_o : \mathbb{Z}_m^{\mathbb{N}} \to (\mathbb{Z}_m^c \times \mathbb{Z}_m)^{\mathbb{N}}$, $\phi_o(\boldsymbol{u}) = (\mathrm{Rep}_c(\boldsymbol{u}), \boldsymbol{u})$. Instead ϕ_i is itself the serial interconnection of two encoders. We consider $\mathrm{Sum}_d : (\mathbb{Z}_m^d)^{\mathbb{N}} \to \mathbb{Z}_m^{\mathbb{N}}$ defined by $\mathrm{Sum}_d(\boldsymbol{y}) = (y_1 + \cdots + y_d, y_{d+1} + \cdots + y_{2d}, \dots)$, and we take a $\mathrm{w_H}$-recursive non-catastrophic rate-1 convolutional encoder $\psi : \mathbb{Z}_m^{\mathbb{N}} \to \mathbb{Z}_m^{\mathbb{N}}$. Finally we take $\phi_i : (\mathbb{Z}_m^d \times \mathbb{Z}_m)^{\mathbb{N}} \to (\mathbb{Z}_m \times \mathbb{Z}_m)^{\mathbb{N}}$ defined by

$$\phi_i(\boldsymbol{y}^1, \boldsymbol{y}^2) = ((\psi \circ \mathrm{Sum}_d)(\boldsymbol{y}^1), \boldsymbol{y}^2).$$

  When taking the truncated versions of these encoders, we must make sure to have suitable lengths, so we take: $\phi_o^N : \mathbb{Z}_m^{dN} \to \mathbb{Z}_m^{cdN} \times \mathbb{Z}_m^{dN}$ and $\phi_i^N : \mathbb{Z}_m^{cdN} \times \mathbb{Z}_m^{dN} \to \mathbb{Z}_m^{cN+\nu_\psi} \times \mathbb{Z}_m^{dN} = \Gamma^{(c+d)N+\nu_\psi}$. So, the design rate of the serial encoder $\phi^N = \phi_i^N \circ \Pi_N \circ \phi_o^N$ is $R = \log m \, \frac{d}{c+d}$.
  As interconnection group, we choose the separated channels symbol permutation $G_N = S_{cdN} \times S_{dN}$.
  This family of codes is a generalization of Repeat-Convolute codes: the additional summator Sum_d is the same as the grouping factor introduced in Irregular Repeat Accumulate codes [28].
  We can easily construct a parity-check matrix for the code $\mathcal{C}^N = \mathrm{Im}(\Phi^N) \subseteq \mathbb{Z}_m^{(c+d)N+\nu_\psi}$, which is sparse and has a structured and a random part, so that we have a structured LDPC ensemble, generalizing staircase LDPC codes. In fact, notice that

$$\begin{aligned}(\boldsymbol{c}^1, \boldsymbol{c}^2) \in \mathcal{C}^N &\Leftrightarrow \boldsymbol{c}^1 = \psi^N \circ \mathrm{Sum}_d \circ \pi_N^1 \circ \mathrm{Rep}_c \circ (\pi_N^2)^{-1}(\boldsymbol{c}^2) \\ &\Leftrightarrow (\psi^N)^{-1}(\boldsymbol{c}^1) = \mathrm{Sum}_d \circ \pi_N^1 \circ \mathrm{Rep}_c \circ (\pi_N^2)^{-1}(c^2)\end{aligned}$$

  It is clear that the permutation π_N^2 does not play any essential role: we needed it only to fit this scheme in our assumptions, but we can take it out without changing the performance of the scheme.
  Note that the non-catastrophicity of ϕ_i is needed to make the syndrome matrix 'low density' , i.e. with a number of non-zero elements per row and per column which is small and does not grow with N. More precisely, the matrix $H_2 = \mathrm{Sum}_d \pi_N^1 \mathrm{Rep}_d$ is a random low density matrix with entries in $\{0, 1\}$, depending only on c, d and π_N, with at most c elements equal to 1 in each column and at most d on any row. Instead $H_1 = (\psi^N)^{-1}$ depends on the choice of ψ, and is also low density, having a number of non-zero elements per row and per column at most equal to the degree of the polynomial $\psi^{-1}(D)$.

5. Main result: interleaver gain. The well-known analysis by Benedetto et al. [4] showed an interleaver gain, in the sense that average error probability is asymptotically vanishing when the interleaver length grows. Their result holds true under the assumption that both constituent encoders are systematic recursive convolutional encoders and that the free distance of the outer encoder was $d_f^o \geq 2$ to ensure $\overline{P_b(e)} \to 0$ and $d_f^o \geq 3$ to have also $\overline{P_w(e)} \to 0$.

In this section, we will comment on how the classical assumptions on the constituent encoders can be adapted to our setting, and we will state our results on the interleaver gain. All the proofs will be given in Section 6. Proofs' techniques are inspired by [15] but they cannot be seen as consequences of the results in [15]: the fact of working with serial interconnections and in a non-binary situation makes the derivation quite different and the combinatorics more involved .

From now on, we will be always considering a regular serial ensemble (see Definition 4.1), with outer encoder $\phi_o : U^{\mathbb{N}} \to (Y^r)^{\mathbb{N}}$ and inner encoder $\phi_i : (Y^s)^{\mathbb{N}} \to (\Gamma^l)^{\mathbb{N}}$, and with a family of interconnection groups (G_N) with invariants weight vector $\boldsymbol{w}_G$. The symbol error probability will be with respect to a fixed weight on the input group U, denoted $\boldsymbol{w}_{\text{in}}$, with the requirement that $\boldsymbol{w}_{\text{in}}(u) \neq 0$ for all $u \neq 0$.

First of all, we have to generalize the assumptions about the constituent encoders introduced in [4].

When considering one single convolutional encoder, non-catastrophicity is usually needed to ensure good asymptotic properties. However, when dealing with a concatenated scheme the assumption that all constituent encoders are non-catastrophic can be slightly weakened, as it was already recognized for example in [15] and in [25] (in the latter, the authors consider serial schemes where the inner encoder is heavily punctured and becomes not injective). The essential assumption is that the overall scheme is non-catastrophic, and this can be obtained by asking classical non-catastrophicity of the outer encoder and a weaker property of the inner encoder: ϕ_i must be non-catastrophic when restricted to the inputs he will actually receive, i.e. the permuted outer codewords.

When we are dealing with ensembles of concatenated codes, each code of the ensemble must be non-catastrophic, in the sense specified above. This leads to the following definition.

DEFINITION 5.1. *A regular serial ensemble with constituent encoders $\phi_o : U^{\mathbb{N}} \to (Y^r)^{\mathbb{N}}$ and $\phi_i : (Y^s)^{\mathbb{N}} \to (\Gamma^l)^{\mathbb{N}}$ and regular group family (G_N) is* concatenatedly non-catastrophic *if there exist two positive constants μ_o and μ_i such that, for all $N \in \mathbb{N}^*$ and for all $\boldsymbol{u} \in U^N$:*

1. $\mathrm{w_H}(\boldsymbol{u}) \leq \mu_o \left|\boldsymbol{w}_G\left(\phi_o^N(\boldsymbol{u})\right)\right|$*;*
2. *for all $\pi \in G_N$, $\left|\boldsymbol{w}_G\left(\phi_o^N(\boldsymbol{u})\right)\right| \leq \mu_i \mathrm{w_H}\left(\phi_i^N \circ \pi \circ \phi_o^N(\boldsymbol{u})\right)$.*

Notice that 1 is equivalent to the non-catastrophicity of ϕ_o (see Prop. 3.4). In the examples introduced in previous section, we have an example where both encoders are non-catastrophic (Repeat-Convolute), and an example where only concatenated non-catastrophicity is true (Structured LDPC). In fact, in this second example, non-catastrophicity of ψ ensures sparsity of the parity-check matrix, but due to non-injectivity of Sum_d the inner encoder is indeed catastrophic; overall non-catastrophicity of the concatenated scheme is ensured by the systematic branch.

About the other classical assumptions on the constituent encoders ($d_f^o \geq 3$ and recursiveness of ϕ_i), clearly they must be re-stated considering the suitable connecting weight $\boldsymbol{w}_G$ instead of Hamming weight, using the definitions introduced in Sect. 3.

However, we will comment later in this section why these assumptions are sufficient and not necessary to obtain some interleaver gain, and how they can be weakened.

Now we will introduce some useful definitions, and then state the interleaver gain result, which will answer to the question: 'Is the average error probability asymptotically vanishing when the interleaver length grows to infinity? And if so, how fast is the decay?'. From now on, we will always assume that we are considering a concatenatedly non-catastrophic ensemble.

Let $\mathcal{C}_o^N = \phi_o^N(U^N)) \subseteq Y^{r(N+\nu_o)}$ be the outer block code, and let

$$H = \{\boldsymbol{w}_G(\boldsymbol{c}) : \boldsymbol{c} \in \mathcal{C}_o^N \text{ for some } N\,,\ \boldsymbol{c} \neq \boldsymbol{0}\}\,.$$

Notice that, with this notation, (2) in Definition 5.1 is equivalent to

$$\forall N \in \mathbb{N}^*,\ \forall \boldsymbol{h} \in H,\ \forall \boldsymbol{c} \in Y^{r(N+\nu_o)} \text{ such that } \boldsymbol{w}_G(\boldsymbol{c}) = \boldsymbol{h},\ |\boldsymbol{h}| \leq \mu_i |\boldsymbol{w}_{\mathrm{T}}\left(\phi_i^N(\boldsymbol{c})\right)|$$

Given $\boldsymbol{h} \in H$, we look at the decomposition of codewords in error events, as defined in Sections 3.1 and 3.4, and we define:

- $n_o(\boldsymbol{h}) = \max\{n(\boldsymbol{c}) \text{ such that } \exists N, \exists \boldsymbol{c} \in \mathcal{C}_o^N : \boldsymbol{w}_G(\boldsymbol{c}) = \boldsymbol{h}\}$
- $n_i(\boldsymbol{h}) = \max\{n(\boldsymbol{x}) \text{ s.t. } \exists N, \exists \boldsymbol{u} \in Y^{r(N+\nu_o)} : \boldsymbol{x} = \phi_i^N(\boldsymbol{u}), \boldsymbol{w}_G(\boldsymbol{u}) = \boldsymbol{h}\}$
- $f(\boldsymbol{h}) = 1 + |\boldsymbol{h}| - n_o(\boldsymbol{h}) - n_i(\boldsymbol{h})$

REMARK 5.2. *Both maxima in the above definition are well defined, since we clearly have $n(\boldsymbol{c}) \leq |\boldsymbol{h}|$, $n(\boldsymbol{x}) \leq |\boldsymbol{h}|$. Moreover, notice that, because of Remark 3.9 in Sect. 3.4, the sequence of sets $\{n(\boldsymbol{c})$ such that $\exists \boldsymbol{c} \in \mathcal{C}_o^N : \boldsymbol{w}_G(\boldsymbol{c}) = \boldsymbol{h}\}$ is increasing in N and so there exists $\bar{N}(\boldsymbol{h})$ such that*

$$n_o(\boldsymbol{h}) = \max\{n(\boldsymbol{c}) \text{ such that } \exists \boldsymbol{c} \in \mathcal{C}_o^{\bar{N}(\boldsymbol{h})} : \boldsymbol{w}_G(\boldsymbol{c}) = \boldsymbol{h}\}$$

An analogous statement holds true for $n_i(\boldsymbol{h})$.

It is also clear that for what $n_o(\boldsymbol{h})$ is concerned, maximum can always be obtained with a codeword which only admits regular error events, while this is not necessarily true for $n_i(\boldsymbol{h})$.

Finally, we define:

$$\alpha = \inf\{f(\boldsymbol{h}),\ \boldsymbol{h} \in H\}\,. \tag{5.1}$$

Notice that the function f takes values in $\mathbb{Z}$ and H is non-empty, so either $\alpha = -\infty$ or $\alpha = \min\{f(\boldsymbol{h}),\ \boldsymbol{h} \in H\}$.

Our main result (formally stated in Theorem 5.4) is that, for sufficiently good channels, if $\alpha \geq 1$,

$$\overline{P_s(e)} \asymp N^{-\alpha} \quad \text{and} \quad \overline{P_w(e)} \asymp N^{-\alpha+1} \quad \text{for } N \to \infty\,.$$

We will give sufficient conditions ensuring that $\alpha \geq 1$, enforcing some properties on the constituent encoders (Propositions 5.5 and 5.7).

In addition to the decay of $\overline{P_s(e)}$ and $\overline{P_w(e)}$ asymptotically in N, we want to underline the dependence of the error probability on the channel's signal-to-noise ratio, following the steps of Benedetto et al. [4] and looking for an analogous of the classical effective free distance.

We define the set of the vectors $\boldsymbol{h}$ minimizing $f(\boldsymbol{h})$:

$$\mathcal{H} = \{\boldsymbol{h} \in H : f(\boldsymbol{h}) = \alpha\}$$

and we define the effective distance

$$q^* = \max_{\boldsymbol{h}\in\mathcal{H}}\{q^*(\boldsymbol{h})\}$$

where

$$q^*(\boldsymbol{h}) = \max\{\mathrm{P}(\mathbf{0}\to\boldsymbol{x})\colon \exists N, \exists \boldsymbol{u}\in Y^{r(N+\nu_o)} : \boldsymbol{x}=\phi_i^N(\boldsymbol{u}), \boldsymbol{w}_G(\boldsymbol{u})=\boldsymbol{h}, n(\boldsymbol{x})=n_i(\boldsymbol{h})\}\,.$$

REMARK 5.3. *We can prove that maxima in the definitions of $q^*(\boldsymbol{h})$ and q^* are well-defined. In principle the number of $\boldsymbol{x}$ involved in the maximum defining $q^*(\boldsymbol{h})$ is infinite. However, we can always restrict the search to a finite set, in the following way. As a first step, we can find an upper bound on the values of $\mathrm{P}(\mathbf{0}\to\boldsymbol{x})$ to consider, trivially by computing $\bar{q}=\mathrm{P}(\mathbf{0}\to\bar{\boldsymbol{x}})$ for one admissible $\bar{\boldsymbol{x}}$. Then we restrict our search to the set:*

$$\bar{X} = \{\boldsymbol{x}\colon \mathrm{P}(\mathbf{0}\to\boldsymbol{x})\geq\bar{q} \text{ and } \exists N, \exists \boldsymbol{u}\in Y^{r(N+\nu_o)} : \boldsymbol{x}=\phi_i^N(\boldsymbol{u}), \boldsymbol{w}_G(\boldsymbol{u})=\boldsymbol{h}, n(\boldsymbol{x})=n_i(\boldsymbol{h})\}.$$

Now note that $\mathrm{P}(\mathbf{0}\to\boldsymbol{x})\geq\bar{q}$ implies $\gamma^{\mathrm{w_H}(\boldsymbol{x})}\geq\bar{q}$, i.e. $\mathrm{w_H}(\boldsymbol{x})\leq\log\bar{q}/\log\gamma$. Now, by Prop. 3.2, we can bound the length of all error events in the decomposition of $\boldsymbol{x}\in\bar{X}$. This implies that, up to shift equivalence, the family of all possible error events appearing in $\boldsymbol{x}\in\bar{X}$ is finite. Therefore, also $\bar{X}$, up to shift equivalence, is finite. The same argument also applies to q^*.

Later, we will also see that under suitable assumptions $\mathcal{H}$ is a finite set (Prop. 5.6).

Using the definition of q^*, we can state the interleaver gain result in a stronger way that underlines, additionally to the decay with N, also the dependence on the channel.

THEOREM 5.4. *Consider a regular and concatenatedly non-catastrophic serial ensemble $(\phi_o,\phi_i,(G_N)_{N\in\mathcal{N}})$, corresponding to the encoding scheme*

$$\xrightarrow{U^N}\boxed{\phi_o^N}\xrightarrow{Y^{r(N+\nu_o)}}\boxed{\pi_N}\xrightarrow{Y^{sM_N}}\boxed{\phi_i^N}\xrightarrow{\Gamma^{l(M_N+\nu_i)}}$$

If $\alpha\geq 1$, there exist positive constants c, c_1, c_2 and γ_0, depending only on ϕ_o,ϕ_i and (G_N), such that, for all Γ-symmetric channels with Bhattacharyya parameter $\gamma<\gamma_0$:

$$c\,q^*N^{-\alpha}\leq\overline{P_w(e)}\leq c_1 q^* c_2^{(\log q^*/\log\gamma)}N^{-\alpha}+O(N^{-\alpha-1})$$

Moreover, for a given input weight $\boldsymbol{w}_{\mathrm{in}}$ (compatible with U and satisfying $\boldsymbol{w}_{\mathrm{in}}(u)\neq 0$ for all $u\neq 0$),

$$c\frac{\boldsymbol{w}_{\mathrm{in}}^{\max}}{\boldsymbol{w}_{\mathrm{in}}^{\min}}q^*N^{-\alpha+1}\leq\overline{P_s(e)}\leq c_1\frac{\boldsymbol{w}_{\mathrm{in}}^{\max}}{\boldsymbol{w}_{\mathrm{in}}^{\min}}q^*c_2^{(\log q^*/\log\gamma)}N^{-\alpha+1}+O(N^{-\alpha})$$

where $\boldsymbol{w}_{\mathrm{in}}^{\max}=\max_{u\in U}\boldsymbol{w}_{\mathrm{in}}(u)$ and $\boldsymbol{w}_{\mathrm{in}}^{\min}=\min_{u\in U\setminus\{0\}}\boldsymbol{w}_{\mathrm{in}}(u)$.

The terms q^* in the lower bound and $q^*c_2^{(\log q^*/\log\gamma)}$ in the upper bound describe the behaviour of $\overline{P_s(e)}$ and $\overline{P_w(e)}$ with respect to the channel's noise. Note that $q^*=\mathrm{P}(\mathbf{0}\to\boldsymbol{c})$ for some word $\boldsymbol{c}$, so that, with the notation $w^*=\mathrm{w_H}(\boldsymbol{c})$, $q^*\leq\gamma^{w^*}$ and $\frac{\log q^*}{\log\gamma}\leq w^*$. Theorem 5.4 holds true for any Γ-symmetric channel with $\gamma\leq\gamma_0$. However, it is particularly relevant for families of channels where $\gamma\to 0$ by preserving the geometry of the channel, thus having constant w^* for all the family: in this case, q^* fully describes the decay of error probability when $\gamma\to 0$ (i.e. SNR grows to infinity). For example, one can consider BSC with crossover probability ϵ and let $\epsilon\to 0$,

or an S-AWGN channel, fixing the shape of the constellation and letting $E_s/N_0 \to \infty$.

Now we will show how the free distance of ϕ_o and the recursiveness of ϕ_i come into the picture. First of all, we generalize the classical assumptions in the most natural way, simply replacing Hamming weight with the interconnection weight $\boldsymbol{w}_G$: this will ensure that there is an interleaving gain (namely that $\alpha \geq 1$). From now on, let's denote by d_f^o the $\boldsymbol{w}_G$–free distance of ϕ_o.

PROPOSITION 5.5. *Assume that $d_f^o \geq 2$ and ϕ_i is $\boldsymbol{w}_G$–recursive. Then*

$$\lfloor (d_f^o + 1)/2 \rfloor \leq \alpha \leq d_f^o \,.$$

In particular, $\alpha \geq 1$ and if $d_f^o \geq 3$ then $\alpha \geq 2$.

In some particular cases we will give tighter upper bounds on α (see Sect. 7).

The strong assumptions used in Prop. 5.5 have another interesting consequence:

PROPOSITION 5.6. *If $d_f^o \geq 3$ and ϕ_i is $\boldsymbol{w}_G$–recursive, then $\mathcal{H}$ is a finite set.*

However, these assumptions are not necessary to obtain an interleaver gain. For example, in the case of parallel concatenations with multiple branches, these assumptions would mean that all constituent encoders are recursive, while it is known that there is an interleaver gain, even if smaller, also when only some of them are recursive [15]. Also, a relaxation of the classical assumptions will allow us to give results about very interesting examples, such as the heavily punctured serial schemes considered in [25], or the class of structured LDPC interpreted as serial schemes that we introduced in Sect. 4.3. Thus, we are interested in a generalization of Prop. 5.5.

PROPOSITION 5.7. *Assume that the interconnection weight has the structure $\boldsymbol{w}_G = (\boldsymbol{w}_1, \boldsymbol{w}_2) : Y^{sM} \to \mathbb{N}^{\rho_1} \times \mathbb{N}^{\rho_2}$ (possibly $\rho_2 = 0$, but $\rho_1 \geq 1$); denote by $d_{f,1}^o$ the $\boldsymbol{w}_1$–free distance of ϕ_o. Assume that $d_{f,1}^o \geq 2$ and ϕ_i is $\boldsymbol{w}_1$–recursive. Then,*

$$\lfloor (d_{f,1}^o + 1)/2 \rfloor \leq \alpha \leq d_f^o \,.$$

In particular, $\alpha \geq 1$, and if $d_{f,1}^o \geq 3$ then $\alpha \geq 2$.

Notice that Prop. 5.5 is a particular case of Prop. 5.7, where $\rho_2 = 0$ and so $d_f^o = d_{f,1}^o$.

6. Proofs of the main results. In this section, we prove our main results, i.e. Theorem 5.4 and Proposition 5.7. We prove the upper bound for $\overline{P_s(e)}$ and the lower bound for $\overline{P_w(e)}$; the whole result stated in Theorem 5.4 is then obtained by the simple remark

$$\overline{P_s(e)} \geq \frac{1}{N} \frac{\boldsymbol{w}_{\text{in}}^{\min}}{\boldsymbol{w}_{\text{in}}^{\max}} \overline{P_w(e)} \,.$$

6.1. Upper bound. This proof is based on the union-Bhattacharyya bound (see e.g. [29]) and on estimations of the weight enumerating coefficients of the constituent encoders.

We will consider only the case when the symbol error rate $P_s(e)$ is defined with respect to Hamming input weight ($\boldsymbol{w}_{\text{in}} = \mathrm{w_H}$); however this will give results true for every other compatible weight, up to a positive constant factor.

The well-known union bound gives

$$\overline{P_s(e)} \leq \sum_w \sum_{\boldsymbol{d}} \frac{w}{N} \overline{A_{w,d}}^N Q(\boldsymbol{d}) \tag{6.1}$$

where $\overline{A_{w,\boldsymbol{d}}}^N$ is the average number of codewords of a serial ensemble with input Hamming weight w and output type weight $\boldsymbol{d}$.

The standard technique (see [29, 4]) is to express $\overline{A_{w,\boldsymbol{d}}}^N$ as a function of suitable enumerating coefficients of the constituent encoders. Here, we need:

- $A^{o,N}_{w,\boldsymbol{h}}$ the number of codewords of ϕ_o^N with input Hamming weight w and output invariants weight vector $\boldsymbol{h}$;
- $A^{i,N}_{\boldsymbol{h},\boldsymbol{d}}$ the number of codewords of ϕ_i^N with input invariants weight vector $\boldsymbol{h}$ and output type weight $\boldsymbol{d}$.

PROPOSITION 6.1. $$\overline{A_{w,\boldsymbol{d}}}^N = \sum_{\boldsymbol{h}\in H} \frac{A^{o,N}_{w,\boldsymbol{h}} A^{i,N}_{\boldsymbol{h},\boldsymbol{d}}}{|Y_{\boldsymbol{h}}^{r(N+\nu_o)}|}.$$

Proof.

$$\overline{A_{w,\boldsymbol{d}}}^N = \sum_{\boldsymbol{u}:\mathrm{w_H}(\boldsymbol{u})=w} \; \sum_{\boldsymbol{v}:\boldsymbol{w}_\mathrm{T}(\phi_i^N(\boldsymbol{v}))=\boldsymbol{d}} \mathbb{P}(\Pi_N(\phi_o^N(\boldsymbol{u})) = \boldsymbol{v})$$

By Remark 4.2,

$$\mathbb{P}(\Pi_N(\phi_o^N(\boldsymbol{u})) = \boldsymbol{v}) = \frac{|G_N(\phi_o^N(\boldsymbol{u}), \boldsymbol{v})|}{|G_N|} = \begin{cases} 0 & \text{if } \boldsymbol{w}_G(\phi_o^N(\boldsymbol{u})) \neq \boldsymbol{w}_G(\boldsymbol{v}) \\ \frac{1}{|Y_{\boldsymbol{h}}^{r(N+\nu_o)}|} & \text{if } \boldsymbol{w}_G(\phi_o^N(\boldsymbol{u})) = \boldsymbol{w}_G(\boldsymbol{v}) = \boldsymbol{h} \end{cases}$$

Substituting in the expression above, we obtain the thesis. ∎

By Lemma 2.2, we know that $|Y_{\boldsymbol{h}}^{r(N+\nu_o)}| \geq \binom{r(N+\nu_o)}{\boldsymbol{h}}$. Thus, by the inequality (6.1) and Prop. 6.1 we have

$$\overline{P_s(e)} \leq \sum_{w,\boldsymbol{h},\boldsymbol{d}} \frac{w}{N} \frac{1}{\binom{r(N+\nu_o)}{\boldsymbol{h}}} A^{o,N}_{w,\boldsymbol{h}} A^{i,N}_{\boldsymbol{h},\boldsymbol{d}} Q(\boldsymbol{d}) \tag{6.2}$$

We have some inequalities involving the indexes $w, \boldsymbol{h}, \boldsymbol{d}$ which are necessary conditions to have non-zero $A^{o,N}_{w,\boldsymbol{h}} A^{i,N}_{\boldsymbol{h},\boldsymbol{d}}$. They are listed in Definition 6.2 and Prop. 6.3.

DEFINITION 6.2. *Let $I \subseteq \mathbb{N}^* \times H \times \mathbb{N}^{\Gamma\setminus\{0\}}$ be the set of triples $(w, \boldsymbol{h}, \boldsymbol{d})$ satisfying the following conditions:*

- $1 \leq w \leq N$;
- $|\boldsymbol{d}| \leq l\,(M_N + \nu_i)$;
- $w \leq \mu_o |\boldsymbol{h}|$ *and* $|\boldsymbol{h}| \leq \mu_i |\boldsymbol{d}|$ *(μ_o and μ_i as in Def. 5.1).*

PROPOSITION 6.3. *If $(w, \boldsymbol{h}, \boldsymbol{d}) \notin I$, then $A^{o,N}_{w,\boldsymbol{h}} A^{i,N}_{\boldsymbol{h},\boldsymbol{d}} = 0$.*

Proof. The first two inequalities are trivial remarks about the length of the input and code words and the definition of free distance; the last one is the concatenated non-catastrophicity of the ensemble (see Def. 5.1). ∎

Now, we need to estimate the product $A^{o,N}_{w,\boldsymbol{h}} A^{i,N}_{\boldsymbol{h},\boldsymbol{d}}$ when it is non-zero. We start with the following inequalities deriving from Prop. 3.11.

PROPOSITION 6.4. *There exist some positive constants a_o, a_i, b_o, b_i such that, for every $(w, \boldsymbol{h}, \boldsymbol{d}) \in I$:*

1. $A^{o,N}_{w,\boldsymbol{h}} \leq \sum_{n_o=1}^{n_o(\boldsymbol{h})} \binom{N+n_o}{n_o} a_o^w b_o^{|\boldsymbol{h}|}$
2. $A^{i,N}_{\boldsymbol{h},\boldsymbol{d}} \leq \sum_{n_i=0}^{n_i^{\max}} \binom{N+n_i}{n_i} a_i^{|\boldsymbol{h}|} b_i^{|\boldsymbol{d}|}$, *where* $n_i^{\max} = \begin{cases} n_i(\boldsymbol{h}) & \text{if } Q(\boldsymbol{d}) \leq q^*(\boldsymbol{h}) \\ n_i(\boldsymbol{h}) - 1 & \text{if } Q(\boldsymbol{d}) > q^*(\boldsymbol{h}) \end{cases}$

Proof. Let $\mathrm{w}_{\max} = \max\{|\boldsymbol{w}_G(v)| : v \in Y^r\}$, so that $|\boldsymbol{w}_G(\boldsymbol{v})|/\mathrm{w}_{\max} \le \mathrm{w}_{\mathrm{H}}(\boldsymbol{v}) \le |\boldsymbol{w}_G(\boldsymbol{v})|$ for all $\boldsymbol{v} \in Y^{r(N+\nu_o)}$. Then:

$$A^{o,N}_{w,\boldsymbol{h}} \le \sum_{n_o=1}^{n_o(\boldsymbol{h})} \sum_{h'=\lfloor |\boldsymbol{h}|/\mathrm{w}_{\max} \rfloor}^{|\boldsymbol{h}|} \Lambda^{o,N}_{w,h',n_o}$$

where $\Lambda^{o,N}_{w,h',n_o}$ is the input/output support enumerating coefficient of ϕ_o^N, as defined in Sect. 3.5. The conclusion now follows in a straightforward way from Prop. 3.11.

The proof for the inner encoder is similar, but we want to exploit the fact that, by definition of $q^*(\boldsymbol{h})$, there are no codewords of input weight $\boldsymbol{h}$, output weight $\boldsymbol{d}$ such that $Q(\boldsymbol{d}) > q^*(\boldsymbol{h})$ having $n_i(\boldsymbol{h})$ error events in their decomposition. To do so, we need to define $A^{i,N}_{\boldsymbol{h},\boldsymbol{d},n}$ to be the number of codewords of ϕ_i^N with input invariants weight vector $\boldsymbol{h}$, output type weight $\boldsymbol{d}$, and n error events in the decomposition, so that

$$A^{i,N}_{\boldsymbol{h},\boldsymbol{d}} = \sum_{n_i=0}^{n_i(\boldsymbol{h})} A^{i,N}_{\boldsymbol{h},\boldsymbol{d},\boldsymbol{n}_i} = \sum_{n_i=0}^{n_i^{\max}} A^{i,N}_{\boldsymbol{h},\boldsymbol{d},n_i}$$

because $A^{i,N}_{\boldsymbol{h},\boldsymbol{d},n_i(\boldsymbol{h})} = 0$ if $Q(\boldsymbol{d}) > q^*(\boldsymbol{h})$. Then we conclude the proof as for the outer encoder, with $\hat{\mathrm{w}}_{\max} = \max\{|\boldsymbol{w}_{\mathrm{T}}(g)| : g \in \Gamma^l\}$:

$$A^{i,N}_{\boldsymbol{h},\boldsymbol{d}} \le \sum_{n_i=1}^{n_i^{\max}} \sum_{h'=\left\lfloor \frac{|\boldsymbol{h}|}{\mathrm{w}_{\max}} \right\rfloor}^{|\boldsymbol{h}|} \sum_{d'=\left\lfloor \frac{|\boldsymbol{d}|}{\hat{\mathrm{w}}_{\max}} \right\rfloor}^{|\boldsymbol{d}|} \Lambda^{i,N}_{h',d',n_i}\,.$$

□

We now prove the following combinatorial inequality.

PROPOSITION 6.5. *There exists a constant $C > 0$ such that, for all $\boldsymbol{h} \in H$ with $|\boldsymbol{h}| \le \mathrm{w}_{\max}\, r(N+\nu_o)$ (where $\mathrm{w}_{\max} = \max\{|\boldsymbol{w}_G(v)| : v \in Y^r\}$):*

$$\frac{1}{\binom{r(N+\nu_o)}{\boldsymbol{h}}} \sum_{n_o=1}^{n_o(\boldsymbol{h})} \sum_{n_i=0}^{n_i^{\max}} \binom{N+n_o}{n_o} \binom{N+n_i}{n_i} \le \begin{cases} C^{|\boldsymbol{h}|}\, \frac{|\boldsymbol{h}|^{f(\boldsymbol{h})-1}}{N^{f(\boldsymbol{h})-1}} & \text{if } n_i^{\max} = n_i(\boldsymbol{h}) \\ C^{|\boldsymbol{h}|}\, \frac{|\boldsymbol{h}|^{f(\boldsymbol{h})}}{N^{f(\boldsymbol{h})}} & \text{if } n_i^{\max} = n_i(\boldsymbol{h}) - 1 \end{cases}$$

Proof. First, we have $\binom{r(N+\nu_o)}{\boldsymbol{h}} \ge \left[\frac{r(N+\nu_o)}{e|\boldsymbol{h}|}\right]^{|\boldsymbol{h}|}$.

This gives $\frac{1}{\binom{r(N+\nu_o)}{\boldsymbol{h}}} \le C^{|\boldsymbol{h}|} \left[\frac{|\boldsymbol{h}|}{N}\right]^{|\boldsymbol{h}|}$ for some constant $C > 0$.

For the other terms, we use the following simple combinatorial inequalities:

- $\binom{N+n}{n} \le \left[\frac{s}{N}\right]^{s-n} \binom{N+s}{s}$ for all $n \ge 0$, $s, N \ge 1$ satisfying $s \ge n$;
- there exists a constant $c > 0$ such that $\binom{N+n}{n} \le c \left[\frac{N+n}{N}\right]^N \left[\frac{N+n}{n}\right]^n$ for all $n, N \ge 1$;
- $\left[\frac{N+n}{N}\right]^N \le e^n$ for all $n \ge 0$, $N \ge 1$.

As $n_o(\boldsymbol{h}) \le |\boldsymbol{h}|$, these inequalities give

$$\begin{aligned} \sum_{n_o=1}^{n_o(\boldsymbol{h})} \binom{N+n_o}{n_o} &\le \sum_{n_o=1}^{n_o(\boldsymbol{h})} \left[\frac{|\boldsymbol{h}|}{N}\right]^{|\boldsymbol{h}|-n_o} c e^{|\boldsymbol{h}|} \left[\frac{N+|\boldsymbol{h}|}{|\boldsymbol{h}|}\right]^{|\boldsymbol{h}|} \\ &\le c_o^{|\boldsymbol{h}|} \sum_{n_o=1}^{n_o(\boldsymbol{h})} |\boldsymbol{h}|^{-n_o} N^{n_o} \\ &\le C_o^{|\boldsymbol{h}|} |\boldsymbol{h}|^{-n_o(\boldsymbol{h})} N^{n_o(\boldsymbol{h})} \end{aligned}$$

(for some positive constants c_o and C_o). The second inequality is true thanks to the assumption that $|\boldsymbol{h}| \leq \mathrm{w}_{\max} r(N+\nu_o)$. Similar estimation can be obtained for the summation concerning the inner encoder and this yields the result. □

If we substitute the estimations given by Propositions 6.4 and 6.5 into the expression (6.2) and we use Prop. 6.3, we get, for some positive constants C_1, C_2, C_3:

$$\overline{P_s(e)} \leq \sum_{\substack{(w,\boldsymbol{h},\boldsymbol{d})\in I:\\ Q(\boldsymbol{d})\leq q^*(\boldsymbol{h})}} \frac{|\boldsymbol{h}|^{f(\boldsymbol{h})-1}}{N^{f(\boldsymbol{h})}} C_1^w C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d}) + \sum_{\substack{(w,\boldsymbol{h},\boldsymbol{d})\in I:\\ Q(\boldsymbol{d})> q^*(\boldsymbol{h})}} \frac{|\boldsymbol{h}|^{f(\boldsymbol{h})}}{N^{f(\boldsymbol{h})+1}} C_1^w C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d}) \tag{6.3}$$

Now, we split the first summation into two terms, separating $\boldsymbol{h} \in \mathcal{H}$ from $\boldsymbol{h} \notin \mathcal{H}$. Define:

- $I_\alpha = \{(w,\boldsymbol{h},\boldsymbol{d}) \in I :\ f(\boldsymbol{h}) = \alpha,\ Q(\boldsymbol{d}) \leq q^*(\boldsymbol{h})\}$,
- $I_> = \{(w,\boldsymbol{h},\boldsymbol{d}) \in I :\ f(\boldsymbol{h}) > \alpha,\ Q(\boldsymbol{d}) \leq q^*(\boldsymbol{h})\}$,
- $I_* = \{(w,\boldsymbol{h},\boldsymbol{d}) \in I :\ Q(\boldsymbol{d}) > q^*(\boldsymbol{h})\}$.

Eq. (6.3) can be re-written as follows:

$$\begin{aligned}
\overline{P_s(e)} \leq{}& \frac{1}{N^\alpha} \sum_{(w,\boldsymbol{h},\boldsymbol{d})\in I_\alpha} |\boldsymbol{h}|^{\alpha-1} C_1^w C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d}) \\
&+ \frac{1}{N^{\alpha+1}} \sum_{(w,\boldsymbol{h},\boldsymbol{d})\in I_>} C_1^w \left(\frac{|\boldsymbol{h}|}{N}\right)^{f(\boldsymbol{h})-\alpha-1} |\boldsymbol{h}|^\alpha C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d}) \\
&+ \frac{1}{N^{\alpha+1}} \sum_{(w,\boldsymbol{h},\boldsymbol{d})\in I_*} C_1^w \left(\frac{|\boldsymbol{h}|}{N}\right)^{f(\boldsymbol{h})-\alpha} |\boldsymbol{h}|^\alpha C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d})
\end{aligned}$$

In the following, we will show that the first summation is bounded by the expression $c\, q^* \exp(\log q^* / \log \gamma)$ (Prop. 6.6), and that the second and the third summations are bounded by a constant $c'(\gamma)$ (Prop. 6.7), thus ending the proof of the upper bound.

PROPOSITION 6.6. *There exist some positive constants γ_0 and c such that, for any BIOS channel with $\gamma < \gamma_0$,*

$$\sum_{(w,\boldsymbol{h},\boldsymbol{d})\in I_\alpha} |\boldsymbol{h}|^{\alpha-1} C_1^w C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d}) \leq c\, q^* \exp(\log q^* / \log \gamma)$$

Proof. Recall that $(w,\boldsymbol{h},\boldsymbol{d}) \in I_\alpha$ implies that $w \leq \mu_o |\boldsymbol{h}|$, $|\boldsymbol{h}| \leq \mu_i |\boldsymbol{d}|$ and $Q(\boldsymbol{d}) \leq q^*(\boldsymbol{h}) \leq q^*$. So:

$$\begin{aligned}
&\sum_{(w,\boldsymbol{h},\boldsymbol{d})\in I_\alpha} |\boldsymbol{h}|^{\alpha-1} C_1^w C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d}) \\
&\qquad \leq \sum_{\substack{\boldsymbol{d}\in\mathbb{N}^{\Gamma\setminus\{0\}}\\ Q(\boldsymbol{d})\leq q^*}} \sum_{\substack{\boldsymbol{h}\in\mathbb{N}^\rho\\ |\boldsymbol{h}|\leq \mu_i|\boldsymbol{d}|}} |\boldsymbol{h}|^{\alpha-1} C_2^{|\boldsymbol{h}|} \Big(\sum_{w\leq \mu_o|\boldsymbol{h}|} C_1^w \Big) C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d}) \\
&\qquad \leq \sum_{\substack{\boldsymbol{d}\in\mathbb{N}^{\Gamma\setminus\{0\}}\\ Q(\boldsymbol{d})\leq q^*}} \sum_{\substack{\boldsymbol{h}\in\mathbb{N}^\rho\\ |\boldsymbol{h}|\leq \mu_i|\boldsymbol{d}|}} |\boldsymbol{h}|^{\alpha-1} C_2^{|\boldsymbol{h}|} \mu_o |\boldsymbol{h}| C_1^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d}) \\
&\qquad \leq \sum_{\substack{\boldsymbol{d}\in\mathbb{N}^{\Gamma\setminus\{0\}}\\ Q(\boldsymbol{d})\leq q^*}} K^{|\boldsymbol{d}|} Q(\boldsymbol{d}) \qquad \text{(for some suitable } K > 0).
\end{aligned}$$

Now, we split the summation, recalling the Bhattacharyya bound $Q(\boldsymbol{d}) \leq \gamma^{|\boldsymbol{d}|}$:

$$\sum_{\substack{\boldsymbol{d}\in\mathbb{N}^{\Gamma\setminus\{0\}}\\ Q(\boldsymbol{d})\leq q^*}} K^{|\boldsymbol{d}|}Q(\boldsymbol{d}) = \sum_{\substack{\boldsymbol{d}\in\mathbb{N}^{\Gamma\setminus\{0\}}\\ Q(\boldsymbol{d})\leq q^*, \gamma^{|\boldsymbol{d}|}>q^*}} K^{|\boldsymbol{d}|}q^* + \sum_{\substack{\boldsymbol{d}\in\mathbb{N}^{\Gamma\setminus\{0\}}\\ \gamma^{|\boldsymbol{d}|}\leq q^*}} K^{|\boldsymbol{d}|}\gamma^{|\boldsymbol{d}|}$$

Now let's find a bound for the number of $\boldsymbol{d}$'s involved in the first summation: $\gamma^{|\boldsymbol{d}|} > q^*$ means $|\boldsymbol{d}| < \log q^*/\log\gamma$ and so there are less than $(|\Gamma|-1)^{\log q^*/\log\gamma}$ type weights satisfying this inequality:

$$\sum_{\substack{\boldsymbol{d}\in\mathbb{N}^{\Gamma\setminus\{0\}}\\ Q(\boldsymbol{d})\leq q^*, \gamma^{|\boldsymbol{d}|}\geq q^*}} K^{|\boldsymbol{d}|}q^* \leq ((|\Gamma|-1)K)^{\log q^*/\log\gamma} q^*$$

For the second term, note that, for $\gamma < 1/K$, the series is convergent, and bounded by a constant times its first term, which has $|\boldsymbol{d}| = \log q^*/\log\gamma$, i.e.

$$\sum_{\substack{\boldsymbol{d}\in\mathbb{N}^{\Gamma\setminus\{0\}}\\ \gamma^{|\boldsymbol{d}|}\leq q^*}} K^{|\boldsymbol{d}|}\gamma^{|\boldsymbol{d}|} \leq CK^{(\log q^*/\log\gamma)}q^*$$

□

PROPOSITION 6.7. *There exists a constant $\gamma_0 > 0$, depending on ϕ_o, ϕ_i and (G_N) and there exists $c'(\gamma) > 0$ depending only on γ such that, for all $\gamma < \gamma_0$*

$$\sum_{(w,\boldsymbol{h},\boldsymbol{d})\in I_>} C_1^w \left(\frac{|\boldsymbol{h}|}{N}\right)^{f(\boldsymbol{h})-\alpha-1} |\boldsymbol{h}|^\alpha C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d})$$
$$+ \sum_{(w,\boldsymbol{h},\boldsymbol{d})\in I_*} C_1^w \left(\frac{|\boldsymbol{h}|}{N}\right)^{f(\boldsymbol{h})-\alpha} |\boldsymbol{h}|^\alpha C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d}) \;\leq\; c'(\gamma)$$

Proof. Notice that, for $(w,\boldsymbol{h},\boldsymbol{d}) \in I_>$, $0 \leq f(\boldsymbol{h}) - \alpha - 1 \leq |\boldsymbol{h}| \leq cN$, where the first inequality holds true because $\boldsymbol{h} \in H\setminus\mathcal{H}$, the second immediately follows from the definitions of $f(\boldsymbol{h})$ and α, the third is true, for a suitable $c > 1$, because $|\boldsymbol{h}| \leq r(N+\nu_o)$ for all $\boldsymbol{h} \in H$. These inequalities imply that $(|\boldsymbol{h}|/N)^{f(\boldsymbol{h})-\alpha-1} \leq c^{|\boldsymbol{h}|}$. Analogously, for all $(w,\boldsymbol{h},d) \in I_*$, $0 \leq f(\boldsymbol{h}) - \alpha \leq |\boldsymbol{h}| \leq cN$ and so $(|\boldsymbol{h}|/N)^{f(\boldsymbol{h})-\alpha} \leq c^{|\boldsymbol{h}|}$.

This gives:

$$\sum_{(w,\boldsymbol{h},\boldsymbol{d})\in I_>} C_1^w \left(\frac{|\boldsymbol{h}|}{N}\right)^{f(\boldsymbol{h})-\alpha-1} |\boldsymbol{h}|^\alpha C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d})$$
$$+ \sum_{(w,\boldsymbol{h},\boldsymbol{d})\in I_*} C_1^w \left(\frac{|\boldsymbol{h}|}{N}\right)^{f(\boldsymbol{h})-\alpha} |\boldsymbol{h}|^\alpha C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d})$$
$$\leq \sum_{(w,\boldsymbol{h},\boldsymbol{d})\in I} C_1^w c^{|\boldsymbol{h}|} C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d})$$

Noticing also that $\sum_{w\leq\mu_o|\boldsymbol{h}|} C_1^w \leq \mu_o|\boldsymbol{h}|C_1^{|\boldsymbol{h}|}$, we have:

$$\sum_{(w,\boldsymbol{h},\boldsymbol{d})\in I} C_1^w c^{|\boldsymbol{h}|} C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d}) \leq \sum_{\boldsymbol{d}\in\mathbb{N}^{\Gamma\setminus\{0\}}} \sum_{\boldsymbol{h}:|\boldsymbol{h}|\leq\mu_i|\boldsymbol{d}|} \mu_o|\boldsymbol{h}|C_1^{|\boldsymbol{h}|} c^{|\boldsymbol{h}|} |\boldsymbol{h}|^\alpha C_2^{|\boldsymbol{h}|} C_3^{|\boldsymbol{d}|} Q(\boldsymbol{d})$$

Now notice that, for some $K > 1$,

$$\sum_{\boldsymbol{h}:|\boldsymbol{h}|\leq\mu_i|\boldsymbol{d}|} \mu_o|\boldsymbol{h}|C_1^{|\boldsymbol{h}|}c^{|\boldsymbol{h}|}|\boldsymbol{h}|^{\alpha}C_2^{|\boldsymbol{h}|} \leq K^{|\boldsymbol{d}|}$$

Finally, we use the Bhattacharyya bound.

$$\sum_{\boldsymbol{d}\in\mathbb{N}^{\Gamma}\setminus\{0\}} (KC)_3^{|\boldsymbol{d}|}Q(\boldsymbol{d}) \leq \sum_{d\in\mathbb{N}} \sum_{\boldsymbol{d}\in\mathbb{N}^{\Gamma}\setminus\{0\}:|\boldsymbol{d}|=d} (KC_3\gamma)^d = c'(\gamma) < \infty$$

if γ is sufficiently small to ensure convergence. □

6.2. Lower bound. The lower bound is based on the following simple remark involving the equivocation probability.

REMARK 6.8. *If $\boldsymbol{c} \in \mathcal{C}^N$, then $P_w(e) \geq \mathrm{P}(\boldsymbol{0} \to \boldsymbol{c})$. So, if one defines $Q_{\max}(\pi_N) = \max\{\mathrm{P}(\boldsymbol{0} \to \boldsymbol{c}), \boldsymbol{c} \in \phi_i^N \circ \pi_N \circ \phi_o^N(U^N)\}$, for any q,*

$$\overline{P_w(e)} \geq q\,\mathbb{P}\big(Q_{\max}(\Pi_N) \geq q\big)\,.$$

We focus our attention on the value $q = q^*$, and we find the following lower bound to $\mathbb{P}\big(Q_{\max}(\Pi_N) \geq q\big)$, thus ending the proof of the lower bound in Theorem 5.4.

PROPOSITION 6.9. *If $\alpha \geq 1$, there exists a constant $C > 0$ such that*

$$\mathbb{P}\big(Q_{\max}(\Pi_N) \geq q\big) \geq CN^{-\alpha+1}\,.$$

In the remainder of this section, we will prove Prop. 6.9. To do so, we need to define some particular codewords which are essential for the bound. We start fixing once and for all the following objects:

1. A weight vector $\boldsymbol{h} \in H$ such that $q^*(\boldsymbol{h}) = q^*$.
2. An outer codeword $\boldsymbol{c}^* \in \phi_o^N(U^N)$ such that $\boldsymbol{w}_G(\boldsymbol{c}^*) = \boldsymbol{h}$ and $n(\boldsymbol{c}) = n_o(\boldsymbol{h})$. Let $n_o = n_o(\boldsymbol{h})$ and let $\boldsymbol{c}^* = \boldsymbol{c}_1^* + \ldots + \boldsymbol{c}_{n_o+1}^*$ be an error event decomposition of $\boldsymbol{c}^*$ (see Sect. 3.4). Denote by l_k the length of $\boldsymbol{c}_k^*$ and define $l_{\max} = \max\{l_1, \ldots, l_{n_o}\}$. If $\alpha = 1$, we need a slightly different definition: $l_{\max} = \max\{l_1, \ldots, l_{n_o}, L\}$ for a suitable constant L which will be chosen in Lemma 6.12.
3. An input word $\boldsymbol{u}^*$ for the inner encoder, such that $\boldsymbol{w}_G(\boldsymbol{u}^*) = \boldsymbol{h}$ and such that $\boldsymbol{x}^* = \phi_i^N(\boldsymbol{u}^*)$ has equivocation $\mathrm{P}(\boldsymbol{0} \to \boldsymbol{x}^*) = q^*(\boldsymbol{h})$ and $n(\boldsymbol{x}^*) = n_i(\boldsymbol{h})$. Let $n_i = n_i(\boldsymbol{h})$ and let $\boldsymbol{x}^* = \boldsymbol{x}_1^* + \ldots + \boldsymbol{x}_{n_i+1}^*$ be an error event decomposition of $\boldsymbol{x}^*$. Denote by $\boldsymbol{u}_k^*$ the input error event corresponding to $\boldsymbol{x}_k^*$ and by λ_k its length (with $\lambda_{n_i+1} = 0$ if there is no terminating event). Define $\lambda_{\max} = \max\{\lambda_1, \ldots, \lambda_{n_i}\}$. Again, for $\alpha = 1$ we need a different definition $\lambda_{\max} = \max\{\lambda_1, \ldots, \lambda_{n_i}, \Lambda\}$ for a suitable constant Λ described in Lemma 6.12.

Notice that $\boldsymbol{c}^*$ can be chosen in such a way that it doesn't have any terminating event and that it does not depend on N, while this may not be possible for $\boldsymbol{u}^*$. However, we can assume that the error events $\boldsymbol{x}_k^*$ and their inputs $\boldsymbol{u}_k^*$ remain the same apart from some possible translations (see Remark 3.9). Also remember that $n_o \geq 1$, $n_i \geq 0$.

Now, we select a sufficiently big set of shift equivalent words for both $\boldsymbol{c}^*$ and $\boldsymbol{x}^*$, choosing many positions for the error events of $\boldsymbol{c}^*$ and for the input error events of $\boldsymbol{u}^*$, across all the time axis $[0, N + \nu_o - 1]$ for $\boldsymbol{c}^*$ and $[0, M_N - 1]$ for $\boldsymbol{u}^*$.

Let's start with $\boldsymbol{c}^*$. Define $\mathcal{A} = [0, \left\lfloor \frac{N}{n_o l_{\max}} \right\rfloor - 1]$. Given $\boldsymbol{a} \in \mathcal{A}^{n_o}$, we define $\boldsymbol{c}_{\boldsymbol{a}}^*$ to be the outer codeword which, for every $k = 1, \ldots, n_o$, contains exactly one

shifted copy of the error event $\boldsymbol{c}_k^*$ starting at time $a_k l_{\max} + (k-1)|\mathcal{A}|l_{\max}$. Clearly by construction all error events in $\boldsymbol{c}_{\boldsymbol{a}}^*$ have disjoint support.

In the same way, we consider the inner input word $\boldsymbol{u}^*$. For $n_i \geq 1$, define $\mathcal{B} = [0, \left\lfloor \frac{M_N - \lambda_{n_i+1}}{n_i \lambda_{\max}} \right\rfloor - 1]$. Given $\boldsymbol{b} \in \mathcal{B}^{n_i}$ we define $\boldsymbol{u}_{\boldsymbol{b}}^*$ to be the inner input word which, for every $k = 1, \dots, n_i$, contains exactly one translated copy of the input error event $\boldsymbol{u}_k^*$ starting at time $b_k \lambda_{\max} + (k-1)|\mathcal{B}|\lambda_{\max}$, while the terminating event $\boldsymbol{u}_{n_i+1}^*$ (if there is one) remains fixed in its position in the interval $[M_N - \lambda_{n_i+1} - 1, M_N - 1]$. Let $\boldsymbol{x}_{\boldsymbol{b}}^*$ be the output $\boldsymbol{x}_{\boldsymbol{b}}^* = \phi_i^N(\boldsymbol{u}_{\boldsymbol{b}}^*)$.

Given $\boldsymbol{a} \in \mathcal{A}^{n_o}$ and $\boldsymbol{b} \in \mathcal{B}^{n_i}$, if $n_i \geq 1$ we define the event

$$E_{\boldsymbol{a},\boldsymbol{b}} = \{\Pi_N(\boldsymbol{c}_{\boldsymbol{a}}^*) = \boldsymbol{u}_{\boldsymbol{b}}^*\} = G_N(\boldsymbol{c}_{\boldsymbol{a}}^*, \boldsymbol{u}_{\boldsymbol{b}}^*)$$

and we also define $E_{\boldsymbol{a}} = \bigcup_{\boldsymbol{b} \in \mathcal{B}} E_{\boldsymbol{a},\boldsymbol{b}}$ (notice that this is an union of disjoint events). If $n_i = 0$, we simply let $E_{\boldsymbol{a}} = \{\Pi_N(\boldsymbol{c}_{\boldsymbol{a}}^*) = \boldsymbol{u}^*\} = G_N(\boldsymbol{c}_{\boldsymbol{a}}^*, \boldsymbol{u}^*)$.

REMARK 6.10. *Clearly $\pi_N \in E_{\boldsymbol{a},\boldsymbol{b}}$ implies $Q_{\max}(\pi_N) \geq \mathrm{P}(\mathbf{0} \to \boldsymbol{x}_{\boldsymbol{b}}^*) = q^*$. Hence,*

$$\mathbb{P}\big(Q_{\max}(\Pi_N) \geq q^*\big) \geq \mathbb{P}\Big(\bigcup_{\boldsymbol{a} \in \mathcal{A}^{n_o}} E_{\boldsymbol{a}}\Big).$$

Our aim is now to estimate this last probability, using:

$$\mathbb{P}\Big(\bigcup_{\boldsymbol{a} \in \mathcal{A}^{n_o}} E_{\boldsymbol{a}}\Big) \geq \sum_{\boldsymbol{a} \in \mathcal{A}^{n_o}} \mathbb{P}(E_{\boldsymbol{a}}) - \sum_{\substack{\boldsymbol{a},\boldsymbol{a}' \in \mathcal{A}^{n_o} \\ \boldsymbol{a} \neq \boldsymbol{a}'}} \mathbb{P}(E_{\boldsymbol{a}} \cap E_{\boldsymbol{a}'})$$

We will prove a lower bound for the first term (Lemma 6.11) and an upper bound for the second term (Lemma 6.12).

LEMMA 6.11. *With the convention $|\mathcal{B}| = 1$ if $n_i = 0$,*

$$\sum_{\boldsymbol{a} \in \mathcal{A}^{n_o}} \mathbb{P}(E_{\boldsymbol{a}}) \geq |\mathcal{A}|^{n_o} |\mathcal{B}|^{n_i} \frac{1}{[|Y| r(N + \nu_o)]^{|\boldsymbol{h}|}}$$

Proof. Clearly $\mathbb{P}(E_{\boldsymbol{a}}) = \dfrac{|E_{\boldsymbol{a}}|}{|G_N|} = \dfrac{|\mathcal{B}|^{n_i} |G_N(\boldsymbol{c}^*, \boldsymbol{u}^*)|}{|G_N|}$.
By Remark 4.2 and Lemma 2.2, $\dfrac{|G_N(\boldsymbol{c}^*, \boldsymbol{u}^*)|}{|G_N|} = \dfrac{1}{|Y_{\boldsymbol{h}}^{r(N+\nu_o)}|} \geq \dfrac{1}{[|Y| r(N + \nu_o)]^{|\boldsymbol{h}|}}$ ∎

LEMMA 6.12. *If $\alpha \geq 2$:*

$$\sum_{\substack{\boldsymbol{a},\boldsymbol{a}' \in \mathcal{A}^{n_o} \\ \boldsymbol{a} \neq \boldsymbol{a}'}} \mathbb{P}(E_{\boldsymbol{a}} \cap E_{\boldsymbol{a}'}) \leq |\mathcal{A}|^{2n_o} |\mathcal{B}|^{2n_i} \left(\frac{2e|\boldsymbol{h}|}{r(N + \nu_o)} \right)^{2|\boldsymbol{h}|}$$

while if $\alpha = 1$:

$$\sum_{\substack{\boldsymbol{a},\boldsymbol{a}' \in \mathcal{A}^{n_o} \\ \boldsymbol{a} \neq \boldsymbol{a}'}} \mathbb{P}(E_{\boldsymbol{a}} \cap E_{\boldsymbol{a}'}) \leq \frac{|\mathcal{A}|^{n_o} |\mathcal{B}|^{n_i}}{[r(N + \nu_o)]^{|\boldsymbol{h}|}} C$$

for some constant $0 < C < \frac{1}{|Y|^{|\boldsymbol{h}|}}$.

Proof. We have

$$E_{\boldsymbol{a}} \cap E_{\boldsymbol{a}'} = \bigcup_{\boldsymbol{b},\boldsymbol{b}' \in \mathcal{B}^{n_i}} (E_{\boldsymbol{a},\boldsymbol{b}} \cap E_{\boldsymbol{a}',\boldsymbol{b}'}) .$$

Consequently, by the union bound,

$$\sum_{\substack{\boldsymbol{a},\boldsymbol{a}'\in\mathcal{A}^{n_o}\\ \boldsymbol{a}\neq\boldsymbol{a}'}} \mathbb{P}(E_{\boldsymbol{a}}\cap E_{\boldsymbol{a}'}) \leq \sum_{\substack{\boldsymbol{a},\boldsymbol{a}'\in\mathcal{A}^{n_o}\\ \boldsymbol{a}\neq\boldsymbol{a}'}} \sum_{\boldsymbol{b},\boldsymbol{b}'\in\mathcal{B}^{n_i}} \mathbb{P}(E_{\boldsymbol{a},\boldsymbol{b}}\cap E_{\boldsymbol{a}',\boldsymbol{b}'})$$

Now we need to deal with $\mathbb{P}(E_{\boldsymbol{a},\boldsymbol{b}}\cap E_{\boldsymbol{a}',\boldsymbol{b}'})$. First of all note that if there is an incomplete error event in $\boldsymbol{u}^*$, surely $\mathbb{P}(E_{\boldsymbol{a},\boldsymbol{b}}\cap E_{\boldsymbol{a}',\boldsymbol{b}'})=0$ for all $\boldsymbol{a}\neq\boldsymbol{a}'$ and for all $\boldsymbol{b},\boldsymbol{b}'$. Now consider the case of only regular events. Fix any $\boldsymbol{a}\neq\boldsymbol{a}'$ and $\boldsymbol{b},\boldsymbol{b}'$ such that there exists $\pi\in E_{\boldsymbol{a},\boldsymbol{b}}\cap E_{\boldsymbol{a}',\boldsymbol{b}'}$. By the definition of $\boldsymbol{c}^*_{\boldsymbol{a}}$ and $\boldsymbol{c}^*_{\boldsymbol{a}'}$, we can find outer codewords $\tilde{\boldsymbol{c}}^*$, $\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}}$, $\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}'}$ (possibly $\tilde{\boldsymbol{c}}^*=\boldsymbol{0}$) having disjoint supports, each consisting of some of the error events $\boldsymbol{c}^*_k$, such that $\boldsymbol{c}^*_{\boldsymbol{a}}=\tilde{\boldsymbol{c}}^*+\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}}$ and $\boldsymbol{c}^*_{\boldsymbol{a}'}=\tilde{\boldsymbol{c}}^*+\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}'}$. More precisely, letting $\tilde{n}_o=\mathrm{d_H}(\boldsymbol{a},\boldsymbol{a}')$, i.e. the number of i's such that $\boldsymbol{a}_i\neq\boldsymbol{a}'_i$, $\tilde{\boldsymbol{c}}^*$ consists of $n_o-\tilde{n}_o$ error events, and $\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}}$ consists of $\tilde{n}_o$ error events and is shift equivalent to $\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}'}$. Clearly, $\boldsymbol{w}_G(\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}})=\boldsymbol{w}_G(\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}'})=\boldsymbol{h}-\boldsymbol{w}_G(\tilde{\boldsymbol{c}}^*)$.

Similarly, we can find inner input words $\tilde{\boldsymbol{u}}^*$, $\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}}$, $\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}'}$ (possibly $\tilde{\boldsymbol{u}}^*=\boldsymbol{0}$ or $\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}}=\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}'}=\boldsymbol{0}$) having disjoint supports, each consisting of some of the input error events $\boldsymbol{u}^*_k$, such that $\boldsymbol{u}^*_{\boldsymbol{b}}=\tilde{\boldsymbol{u}}^*+\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}}$ and $\boldsymbol{u}^*_{\boldsymbol{b}'}=\tilde{\boldsymbol{u}}^*+\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}'}$. Letting $\tilde{n}_i=\mathrm{d_H}(\boldsymbol{b},\boldsymbol{b}')$, $\tilde{\boldsymbol{u}}^*$ has $n_i-\tilde{n}_i$ error events and $\tilde{u}^*_{\boldsymbol{b}}$ has $\tilde{n}_i$ error events and is shift equivalent to $\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}'}$. Clearly, $\boldsymbol{w}_G(\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}})=\boldsymbol{w}_G(\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}'})=\boldsymbol{h}-\boldsymbol{w}_G(\tilde{\boldsymbol{u}}^*)$.

As a consequence of Lemma 4.3, if $\pi\in E_{\boldsymbol{a},\boldsymbol{b}}\cap E_{\boldsymbol{a}',\boldsymbol{b}'}$, then $\pi(\tilde{\boldsymbol{c}}^*)=\tilde{\boldsymbol{u}}^*$, $\pi(\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}})=\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}}$ and $\pi(\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}'})=\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}'}$. This implies that $\boldsymbol{w}_G(\tilde{\boldsymbol{u}}^*)=\boldsymbol{w}_G(\tilde{\boldsymbol{c}}^*)$ and that $\boldsymbol{w}_G(\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}})=\boldsymbol{w}_G(\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}'})=\boldsymbol{w}_G(\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}})=\boldsymbol{w}_G(\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}'})=\boldsymbol{h}-\boldsymbol{w}_G(\tilde{\boldsymbol{c}}^*)$. We will use the notation $\tilde{\boldsymbol{h}}=\boldsymbol{w}_G(\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}})$. Note that if $\mathbb{P}\big(E_{\boldsymbol{a},\boldsymbol{b}}\cap E_{\boldsymbol{a}',\boldsymbol{b}'}\big)\neq 0$ and $(\boldsymbol{a},\boldsymbol{b})\neq(\boldsymbol{a}',\boldsymbol{b}')$ then surely both $\boldsymbol{a}\neq\boldsymbol{a}'$ and $\boldsymbol{b}\neq\boldsymbol{b}'$. Also note that

$$\mathbb{P}\big(E_{\boldsymbol{a},\boldsymbol{b}}\cap E_{\boldsymbol{a}',\boldsymbol{b}'}\big)\leq\mathbb{P}\big(\Pi_N(\tilde{\boldsymbol{c}}^*+\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}}+\tilde{\boldsymbol{c}}^*_{\boldsymbol{a}'})=\tilde{\boldsymbol{u}}^*+\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}}+\tilde{\boldsymbol{u}}^*_{\boldsymbol{b}'}\big)=\frac{1}{Y^{r(N+\nu_o)}_{\boldsymbol{h}+\tilde{\boldsymbol{h}}}}$$

In the simple case when $\tilde{\boldsymbol{h}}=\boldsymbol{h}$, this gives

$$\mathbb{P}(E_{\boldsymbol{a},\boldsymbol{b}}\cap E_{\boldsymbol{a}',\boldsymbol{b}'})\leq\frac{1}{|Y^{r(N+\nu_o)}_{2\boldsymbol{h}}|}\leq\left(\frac{2e|\boldsymbol{h}|}{r(N+\nu_o)}\right)^{2|\boldsymbol{h}|}$$

where the last line comes from Lemma 2.2.

Now notice that $\tilde{\boldsymbol{h}}\in H$ and $\boldsymbol{h}-\tilde{\boldsymbol{h}}\in H\cup\{\boldsymbol{0}\}$, so that

$$1+|\tilde{\boldsymbol{h}}|-\tilde{n}_o-\tilde{n}_i\geq f(\tilde{\boldsymbol{h}})\geq\alpha=1+|\boldsymbol{h}|-n_o-n_i \tag{6.4}$$

and, if $\tilde{\boldsymbol{h}}\neq\boldsymbol{h}$,

$$1+|\boldsymbol{h}-\tilde{\boldsymbol{h}}|-(n_o-\tilde{n}_o)-(n_i-\tilde{n}_i)\geq f(\boldsymbol{h}-\tilde{\boldsymbol{h}})\geq\alpha=1+|\boldsymbol{h}|-n_o-n_i\,. \tag{6.5}$$

Equations (6.4) and (6.5) together are possible only in the case when $\alpha=1$. So, for $\alpha\geq 2$, surely $\boldsymbol{h}=\tilde{\boldsymbol{h}}$, and this ends the proof:

$$\sum_{\boldsymbol{a},\boldsymbol{a}'\in\mathcal{A}^{n_o},\boldsymbol{a}\neq\boldsymbol{a}'}\ \sum_{\boldsymbol{b},\boldsymbol{b}'\in\mathcal{B}^{n_i},\boldsymbol{b}\neq\boldsymbol{b}'}\mathbb{P}(E_{\boldsymbol{a},\boldsymbol{b}}\cap E_{\boldsymbol{a}',\boldsymbol{b}'})\leq|\mathcal{A}|^{2n_o}\,|\mathcal{B}|^{2n_i}\,\left(\frac{2e|\boldsymbol{h}|}{r(N+\nu_o)}\right)^{2|\boldsymbol{h}|}\,.$$

For $\alpha = 1$, instead, note that in this case $\tilde{n}_0 + \tilde{n}_i = |\tilde{\boldsymbol{h}}|$ and $n_0 + n_i = |\boldsymbol{h}|$. We can estimate:

$$\begin{aligned}
\sum_{\substack{\boldsymbol{a},\boldsymbol{a}'\in\mathcal{A}^{n_o}\\ \boldsymbol{a}\neq\boldsymbol{a}'}} \mathbb{P}(E_{\boldsymbol{a}}\cap E_{\boldsymbol{a}'}) &= \sum_{1\le\tilde{n}_o\le n_o}\ \sum_{\substack{\boldsymbol{a},\boldsymbol{a}'\in\mathcal{A}^{n_o}\\ 1\le \mathrm{d_H}(\boldsymbol{a},\boldsymbol{a}')\le\tilde{n}_o}}\ \sum_{1\le\tilde{n}_i\le n_i}\ \sum_{\substack{\boldsymbol{b},\boldsymbol{b}'\in\mathcal{B}^{n_i}\\ 1\le \mathrm{d_H}(\boldsymbol{b},\boldsymbol{b}')\le\tilde{n}_i}} \mathbb{P}(E_{\boldsymbol{a},\boldsymbol{b}}\cap E_{\boldsymbol{a}',\boldsymbol{b}'})\\
&\le \sum_{1\le\tilde{n}_o\le n_o}\binom{n_o}{\tilde{n}_o}|\mathcal{A}|^{n_o+\tilde{n}_o}\sum_{1\le\tilde{n}_i\le n_i}\binom{n_i}{\tilde{n}_i}|\mathcal{B}|^{n_i+\tilde{n}_i}\left(\frac{e(|\boldsymbol{h}|+\tilde{n}_o+\tilde{n}_i)}{r(N+\nu_o)}\right)^{|\boldsymbol{h}|+\tilde{n}_o+\tilde{n}_i}\\
&\le \frac{|\mathcal{A}|^{n_o}\,|\mathcal{B}|^{n_i}}{[r(N+\nu_o)]^{|\boldsymbol{h}|}}2e|\boldsymbol{h}|\left[\left(1+\frac{|\mathcal{A}|2e|\boldsymbol{h}|}{r(N+\nu_o)}\right)^{n_o}-1\right]\left[\left(1+\frac{|\mathcal{B}|2e|\boldsymbol{h}|}{r(N+\nu_o)}\right)^{n_i}-1\right]\\
&\le \frac{|\mathcal{A}|^{n_o}\,|\mathcal{B}|^{n_i}}{[r(N+\nu_o)]^{|\boldsymbol{h}|}}C
\end{aligned}$$

for some constant $C > 0$. Choosing suitably large constants L and Λ in the definition of $l_{\max}$ and $\lambda_{\max}$ ensures that $|\mathcal{A}|$ and $|\mathcal{B}|$ are small enough to have $C < \frac{1}{|Y|^{|\boldsymbol{h}|}}$, which will be useful later. □

Now we can conclude the proof of Prop. 6.9. Using Lemmas 6.11 and 6.12, for $\alpha \ge 2$ we get

$$\sum_{\boldsymbol{a}\in\mathcal{A}^{n_o}}\mathbb{P}(E_{\boldsymbol{a}})-\sum_{\substack{\boldsymbol{a},\boldsymbol{a}'\in\mathcal{A}^{n_o}\\ \boldsymbol{a}\neq\boldsymbol{a}'}}\mathbb{P}(E_{\boldsymbol{a}}\cap E_{\boldsymbol{a}'}) \ge \frac{|\mathcal{A}|^{n_o}\,|\mathcal{B}|^{n_i}}{[r(N+\nu_o)]^{|\boldsymbol{h}|}}\left(\frac{1}{|Y|^{|\boldsymbol{h}|}}-|\mathcal{A}|^{n_o}\,|\mathcal{B}|^{n_i}\frac{(2e|\boldsymbol{h}|)^{2|\boldsymbol{h}|}}{[r(N+\nu_o)]^{|\boldsymbol{h}|}}\right)$$

For $N\to\infty$, as $|\mathcal{A}| \asymp N$ and $|\mathcal{B}| \asymp N$, we have $\frac{|\mathcal{A}|^{n_o}\,|\mathcal{B}|}{[r(N+\nu_o)]^{|\boldsymbol{h}|}} \asymp N^{-\alpha+1}$. We conclude the proof by noticing that $|\mathcal{A}|^{n_o}\,|\mathcal{B}|^{n_i}\frac{(2e|\boldsymbol{h}|)^{2|\boldsymbol{h}|}}{[r(N+\nu_o)]^{|\boldsymbol{h}|}} \asymp N^{-\alpha+1}\to 0$.

For $\alpha = 1$, Lemmas 6.11 and 6.12 give

$$\sum_{\boldsymbol{a}\in\mathcal{A}^{n_o}}\mathbb{P}(E_{\boldsymbol{a}})-\sum_{\substack{\boldsymbol{a},\boldsymbol{a}'\in\mathcal{A}^{n_o}\\ \boldsymbol{a}\neq\boldsymbol{a}'}}\mathbb{P}(E_{\boldsymbol{a}}\cap E_{\boldsymbol{a}'}) \ge \frac{|\mathcal{A}|^{n_o}\,|\mathcal{B}|^{n_i}}{[r(N+\nu_o)]^{|\boldsymbol{h}|}}\left(\frac{1}{|Y|^{|\boldsymbol{h}|}}-C\right)$$

for some $0 < C < \frac{1}{|Y|^{|\boldsymbol{h}|}}$. Then the equality $n_o + n_i = |\boldsymbol{h}|$ ensures that the right-hand side is bounded from below by a strictly positive constant.

6.3. Proof of Propositions 5.5, 5.6 and 5.7. We prove here Prop. 5.6 and Prop. 5.7; clearly the latter one also implies the weaker Prop. 5.5, which can be obtained as a special case just taking $\rho_2 = 0$ so that $\boldsymbol{w}_G = \boldsymbol{w}_1$ and $d_f^o = d_{f,1}^o$.

LEMMA 6.13. *Under the same assumptions as in Prop. 5.7, for all* $\boldsymbol{h} = (\boldsymbol{h}_1, \boldsymbol{h}_2) \in H \subseteq \mathbb{N}^{\rho_1}\times\mathbb{N}^{\rho_2}$*:*

- $1 \le n_o(\boldsymbol{h}) \le \lfloor |\boldsymbol{h}_1|/d_{f,1}^o \rfloor$*;*
- $0 \le n_i(\boldsymbol{h}) \le \lfloor |\boldsymbol{h}_1|/2 \rfloor + |\boldsymbol{h}_2|$*;*
- $1 + |\boldsymbol{h}_1| - \lfloor |\boldsymbol{h}_1|/d_{f,1}^o \rfloor - \lfloor |\boldsymbol{h}_1|/2 \rfloor \le f(\boldsymbol{h}) \le |\boldsymbol{h}|$

Proof. The upper bounds for $n_o(\boldsymbol{h})$ and $n_i(\boldsymbol{h})$ are an immediate consequence of the definition of $d_{f,1}^o$ and of the $\boldsymbol{w}_1$–recursiveness of ϕ_i. For the lower bounds, see Remark 5.2. Then, the estimations for $f(\boldsymbol{h})$ directly follow. □

The definitions of d_f^o and H now clearly imply that $\alpha \le d_f^o$, while the lower bound for α comes from the following property. We omit its easy proof, already sketched in [4].

PROPOSITION 6.14. *Given any constant $c \geq 2$, the function $g : \mathbb{N} \to \mathbb{N}$ defined by $g_c(h) = 1 + h - \lfloor h/c \rfloor - \lfloor h/2 \rfloor$ has $\min_{h \geq c} g_c(h) = \lfloor (c+1)/2 \rfloor$ and*

$$\arg\min_{h \geq c} g_c(h) = \begin{cases} 2\mathbb{N}^* & \text{if } c = 2; \\ \{c,\ c+1,\ 2c\} & \text{if } c = 3; \\ \{c,\ c+1\} & \text{if } c \text{ is odd, } c \geq 5; \\ \{c\} & \text{if } c \text{ is even, } c \geq 4. \end{cases}$$

The lower bound for $f(\boldsymbol{h})$ in Lemma 6.13 can be re-written as $f(\boldsymbol{h}) \geq g_{d^o_{f,1}}(|\boldsymbol{h}_1|)$. For all $\boldsymbol{h} \in H$, clearly $|\boldsymbol{h}_1| \geq d^o_{f,1}$ and so, by Prop. 6.14,

$$\alpha = \min_{\boldsymbol{h} \in H} f(\boldsymbol{h}) \geq \lfloor (d^o_{f,1} + 1)/2 \rfloor .$$

Clearly this lower bound for α immediately means that $d^o_{f,1} \geq 2$ gives $\alpha \geq 1$ and $d^o_{f,1} \geq 3$ gives $\alpha \geq 2$.

Finally we prove that $\mathcal{H}$ is a finite set, under the assumption that $\rho_2 = 0$ and $d^o_f \geq 3$. For any $\boldsymbol{h} \in \mathcal{H}$, i.e. such that $f(\boldsymbol{h}) = \alpha$, by Lemma 6.13 we get:

$$\alpha = f(\boldsymbol{h}) \geq 1 + |\boldsymbol{h}| - \lfloor |\boldsymbol{h}|/d^o_f \rfloor - \lfloor |\boldsymbol{h}|/2 \rfloor \geq 1 + |\boldsymbol{h}| \left(\tfrac{1}{2} - \tfrac{1}{d^o_f} \right)$$

which gives $|\boldsymbol{h}| \leq (\alpha - 1) \frac{2d^o_f}{d^o_f - 2}$, ending the proof.

7. Examples. In this section we consider particular cases, where we can characterize α and q^* exactly or we can give tighter bounds than the general ones. We will particularly focus on the relevant examples introduced in Sect. 4.3. Throughout this section, we will consider $\Gamma = \mathbb{Z}_m$; in some cases we will restrict our attention to m-PSK–AWGN channels.

7.1. Classical free $\mathbb{Z}_m$ serial scheme. We call this scheme classical, because it is the simplest and the most natural generalization of the classical binary serial concatenations introduced in [4].

In the general scheme, take $U = \mathbb{Z}_m^k$, $Y = \mathbb{Z}_m$, $\Gamma = \mathbb{Z}_m$, $G_N = S_{r(N+\nu_o)}$ and consider constituent encoders which are rational matrices $\phi_o \in \mathbb{Z}_m(D)^{k \times r}$ and $\phi_i \in \mathbb{Z}_m(D)^{s \times l}$. See the appendix for properties of convolutional encoders in this particular setting.

Consider symbol error probability with respect to Hamming weight on $\mathbb{Z}_m$ (extended componentwise). Take as interconnection group $G_N = S_{r(N+\nu_o)}$, i.e. all the permutations moving around the elements of $\mathbb{Z}_m$. Clearly the invariant weight $\boldsymbol{w}_G$ will be the type weight $\boldsymbol{w}_{\mathrm{T}}$ on $\mathbb{Z}_m$(extended componentwise). Notice that in this scheme, we can think 'symbols' in the most intuitive way, i.e. to be the elements of $\mathbb{Z}_m$, both in input, in the interconnection and at the output. Clearly, if one takes $m = 2$, symbols are just bits, type weight and Euclidean weight are equal to Hamming weight and so we get the classical binary schemes introduced in [4].

For this ensemble, we have an explicit expression for α if m is a power of 2, and tight bounds for α for general m; we also have simple examples showing that, without more information about the constituent encoders, nothing tighter than these bounds can be found.

Let $m = p_1^{\alpha_1} \dots p_l^{\alpha_l}$ be the prime factors decomposition of m and let $\phi_{j,o} : \mathbb{Z}_{p_j}^{k\mathbb{N}} \to \mathbb{Z}_{p_j}^{r\mathbb{N}}$ be obtained by taking the restriction of ϕ_o to inputs in $\frac{m}{p_j}\mathbb{Z}_m^k$ and then identifying

$\frac{m}{p_j}\mathbb{Z}_m$ with $\mathbb{Z}_{p_j}$ through the natural fields isomorphism. With this notation, the following bounds for α hold true.

PROPOSITION 7.1. *For the classical free $\mathbb{Z}_m$ ensemble,*

$$\lfloor (d_f^o+1)/2 \rfloor \leq \alpha \leq d_f^o - \lfloor d_f^o/p_{\min} \rfloor$$

where $p_{\min} = \min\{p_j : d_f^o = d_f(\phi_{j,o})\}$.

Proof. We already have the bound $\lfloor (d_f^o+1)/2 \rfloor \leq \alpha \leq d_f^o$ (Prop. 5.5), so we just need to prove the tighter upper bound.

Notice that, by Prop. A.2, $\mathcal{P} := \{p_j : d_f^o = d_f(\phi_{j,o})\} \neq \emptyset$. We want to prove that $\alpha \leq d_f^o - \lfloor d_f^o/p \rfloor$ for all $p \in \mathcal{P}$. So, fix any $p \in \mathcal{P}$; consider a word $\boldsymbol{c} \in \frac{m}{p}\phi_o^N(\mathbb{Z}_m^{kN})$ such that $\mathrm{w_H}(\boldsymbol{c}) = d_f^o$; let $\boldsymbol{h} = \boldsymbol{w}_{\mathrm{T}}(\boldsymbol{c})$.

Let $a_1, \ldots, a_{d_f^o} \in \frac{m}{p}\mathbb{Z}_m \setminus \{0\}$ be the non-zero symbols of the word $\boldsymbol{c}$ (possibly with the same symbol repeated many times). Consider $a_1, \ldots, a_p$: by Lemma A.3 (applied to $\mathbb{Z}_p$), there exist indexes $\{j_1, \ldots, j_n\} \subseteq \{1, \ldots, p\}$ such that $a_{j_1} + \ldots, a_{j_n} = 0 \mod m$. Then, by Prop. 3.6, there exist distinct times $t_1, \ldots, t_n$ such that $\phi_i(a_{j_1}D^{t_1} + \ldots + a_{j_n}D^{t_n})$ has finite support, i.e. is formed by some (at least one) error events. By applying the same argument to $a_{p+1}, \ldots, a_{2p}$ and so on up to $a_{(\lfloor d_f^o/p \rfloor - 1)p+1}, \ldots, a_{\lfloor d_f^o/p \rfloor p}$, we obtain that $n_i(\boldsymbol{h}) \geq \lfloor d_f^o/p \rfloor$. Clearly $n_o(\boldsymbol{h}) = 1$ and so we can conclude: $\alpha \leq f(\boldsymbol{h}) \leq 1 + d_f^o - 1 - \lfloor d_f^o/p \rfloor$. ∎

From Prop. 7.1, together with Prop. A.2 (see the appendix),we get the exact value of α for the case when m is a power of 2:

COROLLARY 7.2. *For the classical free $\mathbb{Z}_m$ ensemble, if m is a power of 2,*

$$\alpha = \left\lfloor (d_f^o+1)/2 \right\rfloor .$$

It is more difficult to get an explicit formula for q^*. We can just notice that whenever $\alpha = \lfloor (d_f^o+1)/2 \rfloor$ (so in particular for the classical free $\mathbb{Z}_m$ ensemble when m is a power of 2), the inequalities $n_o(\boldsymbol{h}) \leq \lfloor |\boldsymbol{h}|/d_f^o \rfloor$ and $n_i(\boldsymbol{h}) \leq \lfloor |\boldsymbol{h}|/2 \rfloor$ (Lemma 6.13) and Prop. 6.14, make simpler the description of $\mathcal{H}$. When $m = 2$, the description of $\mathcal{H}$ gets even simpler, because the type weight is a scalar, equal to the Hamming weight, and for all $h \in H$, $n_i(h) = \lfloor h/2 \rfloor$. This allows to find the following explicit formula $q^* = Q(d^*)$ in the binary classical ensemble (d^* was already described by Benedetto et al. [4], but here our result is more precise for odd values of d_f^o). Define $d_{f,2}^i$ and $d_{f,3}^i$ to be the minimum output Hamming weight of a regular error event of the inner encoder constrained to input Hamming weight 2 and 3 respectively ($d_{f,3} = +\infty$ if such an event does not exist). Also define $d_{1,\mathrm{term}}^i$ to be the minimum output Hamming weight of a terminated error event with input Hamming weight 1.

PROPOSITION 7.3. *For the binary classical ensemble, $q^* = Q(d^*)$, where:*

- *if d_f^o is even, $d^* = \frac{1}{2} d_f^o d_{f,2}^i$;*
- *if d_f^o is odd ($d_f^o \geq 5$),*

$$d^* = \begin{cases} \frac{d_f^o-3}{2}\, d_{f,2}^i + \min\left\{ d_{f,2}^i + d_{1,term}^i, d_{f,3}^i, 2d_{f,2}^i \right\} & \text{if } d_f^o + 1 \in H \\ \frac{d_f^o-3}{2}\, d_{f,2}^i + \min\left\{ d_{f,2}^i + d_{1,term}^i, d_{f,3}^i \right\} & \text{if } d_f^o + 1 \notin H \end{cases}$$

- *if $d_f^o = 3$,*

$$d^* = \begin{cases} \min\left\{ d_{f,2}^i + d_{1,term}^i, d_{f,3}^i, 2d_{f,2}^i \right\} & \text{if } 4 \in H \\ \min\left\{ d_{f,2}^i + d_{1,term}^i, d_{f,3}^i, 3d_{f,2}^i \right\} & \text{if } 4 \notin H \end{cases}$$

Now we give three examples with $\Gamma = \mathbb{Z}_m$ and with $G_N = S_{r(N+\nu_o)}$. For the computation of q^*, we consider the specific case of the m-PSK–AWGN channel, for which we can find an explicit expression. Remember that, for S–AWGN channel, given a type $\boldsymbol{d}$, $Q(\boldsymbol{d}) = \frac{1}{2}\operatorname{erfc}\sqrt{(\sum_g d_g \mathrm{w_E}(g))E_s/N_0}$. Note that when S is m-PSK, $\mathrm{w_E}(g) = \sin^2(g\pi/m)$.

EXAMPLE 7.4. *[**Repeat-Accumulate codes**] The encoders are $\phi_o = \mathrm{Rep}_r$ (with $r \geq 2$) and $\phi_i = \frac{1}{1-D}$. We assume that the termination rule for the accumulator is the one that always brings to the zero state in one trellis step (using the input $-a$ if we are in state a).*

We obtain $\alpha = \min\{r-1,\, r - \lfloor r/p \rfloor\}$, where p is the smallest prime divisor of m, and $q^* = \frac{1}{2}\operatorname{erfc}\sqrt{d^* E_s/N_0}$ where:

- *for $m = 2$, $d^* = \lfloor (r+1)/2 \rfloor$;*
- *for even $m \geq 4$,*

$$d^* = \begin{cases} r\mathrm{w_E}(1) & \text{if } r = 2 \text{ or } r = 3 \\ \lfloor (r+1)/2 \rfloor & \text{if } r \geq 4 \end{cases}$$

- *for odd $m \geq 3$, let p be the smallest prime divisor of m and let $n = m/p$. Define*

$$d_*(r,m) = \left\lfloor \tfrac{r}{p} \right\rfloor \sum_{i=1}^{p-1} \mathrm{w_E}(in) + \min_{1 \leq j \leq p-1} \sum_{i=1}^{r \bmod p} \mathrm{w_E}(ijn \bmod p)$$

Then:

$$d^* = \begin{cases} r\mathrm{w_E}(1) & \text{if } r < p \\ d_*(r,m) & \text{if } r \geq 2p \\ \min\{r\mathrm{w_E}(1),\, d_*(r,m)\} & \text{if } p \leq r < 2p \end{cases}$$

Sketch of how to get this result:

- *$m = 2$:*
  We can use the explicit expressions we have for α and d^* in the binary case. For Rep_r, $d_f^o = r$ and $d_f^o + 1 \notin \mathcal{H}$; for the accumulator, $d_{f,2}^i = d_{1,term}^i = 1$ and $d_{f,3}^i = +\infty$.
- *even $m \geq 4$:*
  Notice that $d_f(\frac{m}{p_i}\mathrm{Rep}_r) = r$ for all prime $p_i | m$, so that, if $2|m$, by Prop. 7.1 $\alpha = \lfloor (r+1)/2 \rfloor$. Then compute:
  - *$H = (r\mathbb{N})^{m-1} \setminus \{\mathbf{0}\}$.*
  - *If $r \geq 4$, $\mathcal{H} = \{\boldsymbol{h} \in H : |\boldsymbol{h}| = r, n_o(\boldsymbol{h}) = 1, n_i(\boldsymbol{h}) = \lfloor |\boldsymbol{h}|/2 \rfloor\} = \{\boldsymbol{k}\}$ where $\boldsymbol{k}_{m/2} = r$ and $\boldsymbol{k}_i = 0$ for all $i \neq m/2$. This gives $d^* = \lfloor (r+1)/2 \rfloor \mathrm{w_E}(m/2)$. Notice that $\mathrm{w_E}(m/2) = 1$.*
  - *If $r = 3$, we have $\mathcal{H} = \{\boldsymbol{k}, 2\boldsymbol{k}, \boldsymbol{k}^{(1)}, \ldots, \boldsymbol{k}^{(m/2-1)}\}$ with $\boldsymbol{k}$ as above and $\boldsymbol{k}^{(j)}$ defined by $\boldsymbol{k}_j^{(j)} = \boldsymbol{k}_{m-j}^{(j)} = r$ and $\boldsymbol{k}_i^{(j)} = 0$ for all $i \notin \{j, m-j\}$. Then:*

$$d^* = \min\{\left\lfloor \tfrac{r+1}{2} \right\rfloor \mathrm{w_E}\left(\tfrac{m}{2}\right), r\mathrm{w_E}(1), \ldots, r\mathrm{w_E}(m)\} = \min\{2\mathrm{w_E}\left(\tfrac{m}{2}\right), 3\mathrm{w_E}(1)\}$$

Then $m \geq 4$ implies $\mathrm{w_E}(1) \leq 1/2 = \mathrm{w_E}(m/2)/2$, so $d^* = 3\mathrm{w_E}(1)$.

  - *If $r = 2$, $\mathcal{H} = \{\boldsymbol{h} \in H : \boldsymbol{h}_i = \boldsymbol{h}_{m-i} \forall i = 1, \dots, m/2 - 1\}$ and, with the same reasonings as above, we find again: $d^* = \lfloor (r+1)/2 \rfloor$ for $m = 2$ and $d^* = d^*((2, 0, \dots, 0, 2)) = 2\mathrm{w_E}(\phi_i(1 - D)) = 2\mathrm{w_E}(1)$ for $m \geq 4$.*
- *odd $m \geq 3$:*
  - *$H = (r\mathbb{N})^{m-1} \setminus \{\mathbf{0}\}$.*
  - *Compute:*

$$\min_{\boldsymbol{h} \in H : |\boldsymbol{h}| = kr} f(\boldsymbol{h}) = \begin{cases} 1 + kr - k - \frac{k}{2} r & \text{if } k \text{ even} \\ 1 + kr - k - \frac{k-1}{2} r - \lfloor r/p \rfloor & \text{if } k \text{ odd} \end{cases}$$

  Notice that both expressions are non decreasing in k, and increasing in k if $r \geq 3$, so that $\alpha = \min\{r - 1, r - \lfloor r/p \rfloor\}$
  - *if $r = 2$, $\alpha = 1$ and $\mathcal{H} = \{\boldsymbol{h} \in H : \boldsymbol{h}_i = \boldsymbol{h}_{m-i} \forall i = 1, \dots, m/2 - 1\}$, so that $d^* = r\mathrm{w_E}(1)$, obtained for $\boldsymbol{h} = r\boldsymbol{e}_1 + r\boldsymbol{e}_{-1}$;*
  - *if $2 < r < p$, $\alpha = r - 1$ and $\mathcal{H} = \{\boldsymbol{k}^{(1)}, \dots, \boldsymbol{k}^{((m-1)/2)}\}$, the $\boldsymbol{k}^j$'s defined as for even m. So again $d^* = r\mathrm{w_E}(1)$.*
  - *if $r \geq 2p$, $\alpha = r - \lfloor r/p \rfloor$ and $\mathcal{H} = \{r\boldsymbol{e}_{m/p}, r\boldsymbol{e}_{2m/p}, \dots, r\boldsymbol{e}_{(p-1)m/p}\}$, from which the expression for d^* easily follows.*
  - *if $p \leq r < 2p$, $\alpha = r - 1 = r - \lfloor r/p \rfloor$, and $\mathcal{H}$ is the union of the set $\mathcal{H}$ computed for $r < p$ and the one computed for $r \geq 2p$; thus, d^* is the minimum of the two values obtained before.*

The Repeat-Accumulate code on $\mathbb{Z}_3$ ($r \geq 3$) is an example where the upper bound $\alpha \leq d_f^o - \lfloor d_f^o/3 \rfloor$ is reached with equality. Now, we show another simple Repeat-Convolute code on $\mathbb{Z}_3$ such that the lower bound $\alpha \geq \lfloor (d_f^o + 1)/2 \rfloor$ is reached with equality, showing that the bounds in Prop. 7.1 are the best possible for general m.

EXAMPLE 7.5. *Consider $m = 3$ and $\phi_o = \mathrm{Rep}_r$ ($r \geq 2$) and $\phi_i = 1/(1 + D) = \sum_{t \geq 0} D^{2t} + 2D^{2t+1}$, with the termination rule that always brings to the zero state in one trellis step (i.e. if at time t the codeword has $c_t = a$, we terminate using the input $u_{t+1} = -a$ if t is even, a if t is odd).*

Then, as for the Repeat-Accumulate, $H = r\mathbb{N}^2 \setminus \{(0, 0)\}$ and given $\boldsymbol{h} = (rh_1, rh_2)$ we have $n_o(\boldsymbol{h}) = h_1 + h_2$. But now, when we look at the inner encoder to compute $n_i(\boldsymbol{h})$, we find $n_i(\boldsymbol{h}) = \lfloor |\boldsymbol{h}|/2 \rfloor$, because all the following inputs produce a complete error event of ϕ_i: $\boldsymbol{u} = D^t + D^{t+1}$, $\boldsymbol{u} = 2D^t + 2D^{t+1}$ $\boldsymbol{u} = D^t + 2D^{t+2}$ and $\boldsymbol{u} = 2D^t + D^{t+2}$.

As a consequence,

$$\alpha = \lfloor (r+1)/2 \rfloor$$

Let's compute $q^* = \frac{1}{2} \operatorname{erfc} \sqrt{d^* E_s / N_0}$ for this example. First of all we need $\mathcal{H}$:

$$\mathcal{H} = \begin{cases} H & \text{if } r = 2 \\ \{(r, 0),\ (0, r),\ (r, r),\ (2r, 0),\ (0, 2r)\} & \text{if } r = 3 \\ \{(r, 0),\ (0, r)\} & \text{if } r > 3 \end{cases}$$

Now consider that: $\mathrm{w_E}(\phi_i(D^t + D^{t+1})) = \mathrm{w_E}(D^t) = \mathrm{w_E}(1) = 3/4$; analogously $\mathrm{w_E}(\phi_i(2D^t + 2D^{t+1})) = \mathrm{w_E}(2) = 3/4$; while $\mathrm{w_E}(\phi_i(D^t + 2D^{t+2})) = \mathrm{w_E}(D^t + 2D^{t+1}) = \mathrm{w_E}(1) + \mathrm{w_E}(2) = 3/2$ and the same for $\mathrm{w_E}(\phi_i(2D^t + D^{t+2})) = 3/2$. Assuming that termination is always done in one single step, we also have that $\mathrm{w_E}(\phi_i^N(D^{N-1})) = \mathrm{w_E}(\phi_i^N(2D^{N-1})) = 3/4$. Finally, we get $d^* = \frac{3}{4} \left\lfloor \frac{r+1}{2} \right\rfloor$.

Notice that for Repeat-Accumulate codes (Example 7.4) $\alpha = \lfloor (r+1)/2 \rfloor$ for all even m. This is true for all Repeat-Convolute codes, by Prop. 7.1 together with the

remark that $d_f(\frac{m}{p_i}\mathrm{Rep}_r) = r$ for all prime $p_i|m$. However, for a general outer encoder ϕ_o this is not true: the assumption that m is a power of two is essential in Coroll. 7.2, as shown by the following example.

EXAMPLE 7.6. *Let $m = 6$. Consider ϕ_o which is the following slight variation of a Repeat code: $\phi_o = [1, 1, 1, 1, 3]^T$. Let the inner encoder be the Accumulator $\phi_i = \frac{1}{1-D}$. For $p_1 = 2$ we have $\phi_{1,o} = [1, 1, 1, 1, 1]^T$, which has $d_f(\phi_{1,o}) = 5$, while for $p_2 = 3$ we have $\phi_{2,o} = [1, 1, 1, 1, 0]^T$, which has $d_f(\phi_{2,o}) = 4$, and so $d_f^o = d_f(\phi_{2,o}) = 4$. The bounds given in Prop. 7.1 give us $2 \le \alpha \le 3$ and now we will show that $\alpha = 3$. Notice that $f(\boldsymbol{h}) = \alpha$ implies that $1 + |\boldsymbol{h}| - \frac{|\boldsymbol{h}|}{d_f^o} - \frac{|\boldsymbol{h}|}{2} \le \alpha \le 3$ and then $|\boldsymbol{h}| \le 8$, so $\mathcal{H} \subseteq \{\boldsymbol{h} \in H : |\boldsymbol{h}| \le 8\}$. There are seven elements in H with $|\boldsymbol{h}| \le 8$. By computing $f(\boldsymbol{h})$ for the all of them, we get $\alpha = 3$ and $\mathcal{H} = \{(0,4,0,0,0), (0,0,5,0,0), (0,0,0,4,0)\}$ and finally, for 6-PSK–AWGN channel, $q^* = \frac{1}{2}\operatorname{erfc}\sqrt{d^* E_s/N_0}$ with $d^* = 2\mathrm{w_E}(2) + \mathrm{w_E}(4) = 9/4$, reached when $\boldsymbol{h} = (0,4,0,0,0)$ and $\boldsymbol{h} = (0,0,0,4,0)$.*

7.2. Subgroups of permutations for the $\mathbb{Z}_m$ scheme. In the previous section, we have considered $\mathbb{Z}_m$-schemes

$$\xrightarrow{\mathbb{Z}_m^{kN}} \boxed{\phi_o^N} \xrightarrow{\mathbb{Z}_m^{r(N+\nu_o)}} \boxed{\pi_N} \xrightarrow{\mathbb{Z}_m^{sM_N}} \boxed{\phi_i^N} \xrightarrow{\mathbb{Z}_m^{l(M_N+\nu_i)}}$$

by taking $U = \mathbb{Z}_m^k$, $Y = \mathbb{Z}_m$, $\Gamma = \mathbb{Z}_m$ in the general serial scheme. However, we can also obtain some $\mathbb{Z}_m$ schemes by taking $Y = \mathbb{Z}_m^a$. Then, if we consider on $\mathbb{Z}_m^a$ a weight given by the componentwise extension of the type weight on $\mathbb{Z}_m$, we get again the same scheme as above. However, in this case we can also consider permutations moving around not single elements of $\mathbb{Z}_m$, but only the vectors in $\mathbb{Z}_m^a$, so that the invariant weight is the type weight on $\mathbb{Z}_m^a$. Or, on the contrary, we can consider a 'separate channels permutation': the invariant weight is $\boldsymbol{w} \in (\mathbb{N}^{m-1})^a$ given by the type weight on each separate component of $\mathbb{Z}_m^a$.

Even though these schemes are quite similar to the classical one, differing only for a restriction of the permutations to a subgroup of $S_{r(N+\nu_o)}$, Prop. A.2 and Coroll. 7.2 do not hold true. We give here a simple example, for the binary case $m = 2$, and for the 'separate channels' permutation, where $\alpha > \lfloor (d_f^o + 1)/2 \rfloor$.

EXAMPLE 7.7. *Consider the following outer and inner binary encoders:*

$$\phi_o = \left[1, \tfrac{1}{1+D+D^3}\right]^T \qquad \phi_i = \begin{bmatrix} \frac{1}{1+D} & 0 \\ 0 & \frac{1}{1+D} \end{bmatrix}$$

and consider the 'separate channels permutation' ensemble (here $m = 2$ and $a = 2$ and so $\boldsymbol{w} \in \mathbb{N}^2$ is the Hamming weight of the two streams). The outer encoder has free distance $d_f^o = 4$ and all the words $\boldsymbol{c}$ of the outer code such that $\mathrm{d_H}(\boldsymbol{c}) = d_f^o$ are obtained when input is $1 + D + D^3$ or its shifts and have $\boldsymbol{w}(\boldsymbol{c}) = (3,1)$. The inner encoder is simply the rate-1 Accumulator, but acting separately on the two input streams.

We claim that for this scheme $\alpha = 3 > \lfloor (d_f^o + 1)/2 \rfloor = 2$. In fact, we know that $\alpha \ge \lfloor (d_f^o + 1)/2 \rfloor = 2$, where equality could be reached only if there was $\boldsymbol{h} \in H$ such that $|\boldsymbol{h}| = 4$, $n_o(|\boldsymbol{h}|) = 1$, $n_i(\boldsymbol{h}) = 2$, but this is not possible, as the only $\boldsymbol{h} \in H$ such that $|\boldsymbol{h}| = 4$ is $\boldsymbol{h} = (3,1)$, which has $n_o(\boldsymbol{h}) = 1$ but $n_i(\boldsymbol{h}) = 1$, giving $f(\boldsymbol{h}) = 3$ and so $\alpha = 3$.

By an exhaustive listing of all small-weight codewords, we can also find $\mathcal{H}$, noting that $\boldsymbol{h} \in \mathcal{H}$ implies $|\boldsymbol{h}| \le 8$, and then we can find $q^* = Q(3)$.

REMARK 7.8. *The 'separate channels' ensemble is particularly interesting because it allows to include in our generalized serial concatenations also traditional parallel turbo codes (as it was already noticed e.g. in [1]): a turbo code with b parallel branches, each with an encoder ψ_j of rate k_j/n_j, can always be seen as Repeat-Convolute scheme, where $\phi_o = \mathrm{Rep}_r$ and $r = \sum k_j$, the interleaver acts separately on the b streams of $k_j \times N$ bits and ϕ_i is a block diagonal matrix, where the blocks are the ψ_j's.*

7.3. Structured LDPC ensemble. For a description of these schemes, see Section 4.3. Here we give some statements about the parameters α and d^*.

First of all, we have the following tight bounds for α.

PROPOSITION 7.9. *For the structured LDPC ensemble,*

$$\lfloor (c+1)/2 \rfloor \leq \alpha \leq c - \lfloor c/p_{\min} \rfloor$$

where $p_{\min} = \min\{p_j \geq 2 : p_j | m\}$.

Proof. By Prop. 5.7, we have $\alpha \geq \lfloor (d^o_{f,1} + 1)/2 \rfloor = \lfloor (c+1)/2 \rfloor$. The proof of the upper bound is similar to the proof of Prop. 7.1. Notice that here $p_{\min}$ is computed considering all prime factors of m because the outer encoder is a simple repetition code. □

In particular, this Proposition implies that for all even m the interleaver gain is:

$$\alpha = \lfloor (c+1)/2 \rfloor .$$

In the binary case ($m = 2$), we can also characterize q^*. In fact, we can easily describe $\mathcal{H}$:

$$\mathcal{H} = \begin{cases} \{(2w, w) : w \in \mathbb{N}^*\} & \text{if } c = 2 \\ \{(3,1),\ (6,2)\} & \text{if } c = 3 \\ \{(c,1)\} & \text{if } c \geq 4 \end{cases}$$

and then compute $q^* = Q(d^*)$:

$$d^* = \begin{cases} 1 & \text{if } c \text{ is even} \\ 2 & \text{if } c = 3 \\ 1 + \min\{d_{1,\mathrm{term}}(\psi), d_{f,3}(\psi)\} & \text{if } c \text{ is odd, } c \geq 5 \end{cases}$$

where $d_{1,\mathrm{term}}(\psi), d_{f,3}(\psi)$ are defined as $d^i_{1,\mathrm{term}}, d^i_{f,3}$ in Prop. 7.3 but referring here to ψ instead of ϕ_i. If the inner encoder is truncated instead of terminated, $d^* = 2$ for all odd c.

Notice that the choice of ψ has almost no influence on d^*. This happens because pairs of bits which are repetition of a same information bit can be permuted by some interleaver in such a way that they are summed up by Sum_d, producing a zero output. The value of d^* is given by this worse case scenario. This remark suggests to consider interleavers with a better spread, enforcing the fact that 1's coming from the same error event of Rep_c cannot end up in positions where they would be summed up by Sum_d. However, the analysis of such a smaller ensemble, with a set of interleavers which is not a group, is beyond the scope of this paper. See [24] for the results that can be obtained and a sketch of the techniques used; more details on structured LDPC codes will be given in a paper in preparation.

For general m there is no explicit simple characterization of q^*, and neither there is one for all even m. We can just notice that, on m-PSK–AWGN channels, by the same argument used for $m = 2$, if both m and c are even then $q^* = \frac{1}{2}\operatorname{erfc}\sqrt{d^* E_s/N_0}$ with $d^* \leq \mathrm{w_E}(m/2) = 1$. This upper bound is achieved, for example, simply taking $\psi = 1/(1-D)$. To see that this upper bound is not always achieved, take for example $m = 6$ and $\psi = 1/(1+D)$: we have that $\psi(1+D) = 1$ which is an error event of Euclidean weight $\mathrm{w_E}(1) = 1/4$, so that for $c = 2$ or $c = 4$ we have $d^* = \frac{1}{4} + \frac{c}{2}\frac{1}{4} < 1$.

8. Conclusion. In this paper we have studied the average ML performance of a wide class of generalized serial turbo schemes, coupling two convolutional encoders over groups through an interleaver respecting the group structure; these codes are designed to be used on symmetric channels. A particularly relevant example is the case when the convolutional codes are modules on $\mathbb{Z}_m$, the interleaver is a permutation and the channel is AWGN with m–PSK input constellation.

We have obtained the exact asymptotic decay of the symbol and word error probability when the interleaver length goes to infinity and also the behavior when the SNR goes to infinity. The performance is characterized by the interleaver gain α and the effective distance q^*, which are defined as the solution of an optimization problem and in general jointly depend on both constituent encoders, differently from the binary case. To make clear the meaning of these parameters, we have explicitly computed them in some examples encompassing most of the relevant scenarios.

This work is a first attempt to give a rigorous analysis of generalized serial turbo schemes for non-binary modulations. It leaves many open questions among which the most natural ones are, in our opinion, the following:

- Our analysis provides design parameters. The next research step would be to extensively search for pairs of constituent encoders being 'good' with respect to these parameters, and to confirm the validity of the approach by a significant simulation analysis, as it has already been done for the binary turbo codes. However, here the optimization of the constituent encoders under some fixed complexity bound (e.g. the size of the state spaces) seems a quite challenging problem, due to the complexity of the combinatorial optimization problem.
- The ensemble analysis in this paper is limited to the average behavior. The study of the average properties is usually only the first step in the understanding of a family of codes. The next step is to find the typical performance. For the binary classical serial turbo ensemble, the typical error probability turns out to be much better than the average, with sub-exponential vs. polynomial decay, but depending on the same design parameters [8]. This is analogous of the well known behaviour of ML-decoded LDPC codes (see [22], [33]): for the (c, d)-regular LDPC ensemble, with $c \geq 3$, the average error probability is known to decrease to zero as $N^{1-c/2}$ for even c and N^{2-c} for odd c, while the error probability of a typical code goes to zero exponentially fast. In both cases, the average is strongly affected by a small fraction of very bad codes, which can be expurgated. The proofs proposed in [8] heavily rely on properties of binary codes, and so the generalization to our setting is not straightforward.

Appendix. Properties of free $\mathbb{Z}_m$ convolutional encoders.

In this Section, we consider convolutional encoders $\phi : \mathbb{Z}_m^{k\mathbb{N}} \to \mathbb{Z}_m^{n\mathbb{N}}$ which can be represented as matrices $\phi \in \mathbb{Z}_m^{k\times n}(D) \simeq \mathbb{Z}_m(D)^{k\times n}$. We will call them free $\mathbb{Z}_m$ convolutional encoders. They are the most straightforward generalization of classical

binary convolutional encoders, and they have some interesting properties. Let us start with a simple algebraic remark: we know that ϕ can be represented as $\phi = p(D)^{-1}q(D)$ for some $p(D) \in \mathbb{Z}[D] \cap \mathbb{Z}((D))^*$ and $q(D) \in \mathbb{Z}_m[D]^{k\times n}$. Since all the algebraic structures involved are also $\mathbb{Z}_m$-modules, it turns out that we can as well assume that $p(D) \in \mathbb{Z}_m[D] \cap \mathbb{Z}_m((D))^*$ which in practice means that $p(D)$ has all coefficients in $\mathbb{Z}_m$ and the trailing coefficient is in $\mathbb{Z}_m^*$.

In Sect. 3.3.2, we gave a general definition of recursiveness. In the binary case (for simplicity consider scalar input, i.e. $\phi : \mathbb{Z}_2^{\mathbb{N}} \to \mathbb{Z}_2^{n\mathbb{N}}$), there are well-known characterizations of $\mathrm{w_H}$–recursive encoders: ϕ is recursive when its shift-register state representation has a feedback, or equivalently if $\phi = \frac{1}{q(D)}[p_1(D), \dots, p_n(D)]$, with $\gcd\{q, p_1, \dots, p_n\} = 1$ has non-trivial denominator, i.e. $q(D) \neq D^h$. This latter characterization allows to check very easily if an encoder is recursive and we will now generalize it to recursiveness of free $\mathbb{Z}_m$ encoders with respect to Hamming or equivalently to type weight in $\mathbb{Z}_m$ (not the Hamming weight in $\mathbb{Z}_m^k$).

First of all, without loss of generality we can restrict ourselves to considering scalar encoders $\phi : \mathbb{Z}_m^{\mathbb{N}} \to \mathbb{Z}_m^{\mathbb{N}}$: if not so, notice that $\phi : \mathbb{Z}_m^{k\mathbb{N}} \to \mathbb{Z}_m^{n\mathbb{N}}$ is $\boldsymbol{w}$–recursive ($\boldsymbol{w}$ being the Hamming or the type weight in $\mathbb{Z}_m$) if and only if each column of its matrix has at least one entry which is a scalar $\boldsymbol{w}$–recursive encoder.

Then, if m is a prime (so that $\mathbb{Z}_m$ is a field), $\phi = p(D)/q(D)$ with $p(D), q(D) \in \mathbb{Z}_m[D]$ and $\gcd(p, q) = 1$ is $\boldsymbol{w}$–recursive if and only if $q(D) \neq D^t$: as in the binary case, we can identify recursive encoders at a glance, just looking at their denominator.

If m is not a prime, let $m = p_1^{\alpha_1} \dots p_l^{\alpha_l}$ be its prime factors decomposition and let $\phi_i : \mathbb{Z}_{p_i}^{\mathbb{N}} \to \mathbb{Z}_{p_i}^{\mathbb{N}}$ be obtained by taking the restriction of ϕ to inputs in $\frac{m}{p_i}\mathbb{Z}_m$ and then identifying $\frac{m}{p_i}\mathbb{Z}_m$ with $\mathbb{Z}_{p_i}$ through the natural fields isomorphism mapping.

PROPOSITION A.1. *ϕ is $\boldsymbol{w}$–recursive if and only if $\phi_1, \dots, \phi_l$ are $\boldsymbol{w}$–recursive.*

Proof. The first implication is trivial. Conversely, knowing that $\phi_1, \dots, \phi_l$ are recursive, we want to show that $\mathrm{w_H}(\phi(D)u) = +\infty$ for any $u \in \mathbb{Z}_m \setminus \{0\}$. Since $u < m$, there exists $i \in \{1, \dots, l\}$ and $r \in \mathbb{N}$ such that $p_i^r | u$, $p_i^{r+1} \nmid u$, and $p_i^{r+1} | m$. Consider $\tilde{u} = (p_i^{-1}m)(p_i^{-r}u) = (p_i^{-r-1}m)u$. Clearly, $\tilde{u} \neq 0$ and, by the assumptions made, $\mathrm{w_H}(\phi(D)\tilde{u}) = \infty$. This clearly implies that also $\mathrm{w_H}(\phi(D)u) = \infty$. ∎

The characterization given by Prop. A.1 is helpful because the encoders $\phi_1, \dots, \phi_l$ can be obtained very easily from ϕ: writing $\phi(D) = p(D)/q(D)$ with $p(D), q(D) \in \mathbb{Z}_m[D]$, we have $\phi_j(D) = \tilde{p}(D)/\tilde{q}(D)$ where $\tilde{p}(D), \tilde{q}(D)$ are polynomials in $\mathbb{Z}_{p_j}$ obtained multiplying each coefficient of $p(D)$ (resp. $q(D)$) by m/p_j (modulo m) and then identifying the corresponding element in $\mathbb{Z}_{p_j}$.

For example, the encoder $\phi : \mathbb{Z}_8^{\mathbb{N}} \to \mathbb{Z}_8^{\mathbb{N}}$ defined by $\phi(D) = \frac{1+3D^2}{1+7D}$ is not recursive. We cannot tell it simply looking at the denominator, which is non-trivial. We can see it using the definition given in Sect. 3.3.2: notice that $\phi(D) = (1+3D^2)\sum_{t\geq 0} D^t$ and then input $u(D) = 2$ produces output $2\phi(D) = 2 + 2D \in \mathbb{Z}_8[D]$. We can also check the recursiveness of ϕ using Prop. A.1: as $m = 8$ has only one prime divisor $p_1 = 2$, we need to check only one encoder $\phi_1 = \frac{1+D^2}{1+D} = 1 + D$ which clearly isn't recursive.

By the same technique of looking at the encoders $\phi_1, \dots, \phi_l$ defined above, we can obtain a characterization of the free distance of ϕ with respect to Hamming or type weight in $\mathbb{Z}_m$. This characterization is not interesting under a computational point of view, as the computation of the free distance of encoders over fields or rings does not have a different complexity, but it is essential to find tight bounds for the interleaver gain of free $\mathbb{Z}_m$ serial schemes (Prop. 7.1 and Coroll. 7.2).

PROPOSITION A.2. *Let d_f be the $\boldsymbol{w}$–free distance of ϕ and $d_f(\phi_j)$ be the $\boldsymbol{w}$–free distance of ϕ_j, where $\boldsymbol{w}$ is Hamming or type weight in $\mathbb{Z}_m$ and $\mathbb{Z}_{p_j}$ respectively. Then:*

$$d_f = \min_{j=1,\dots,l} \{d_f(\phi_j)\}$$

Proof. Clearly, for all $j = 1, \dots, l$, $d_f(\phi_j) = d_f(\frac{m}{p_j}\phi) \geq d_f$. Now we will prove that there exists at least one j such that $d_f(\phi_j) = d_f$.

Let $\mathcal{C} = \phi\left(\mathbb{Z}_m^k((D))\right)$ and let $\boldsymbol{x} \in \mathcal{C}$ be a codeword such that $\mathrm{w_H}(\boldsymbol{x}) = d_f$. The key remark is that $\mathrm{w_H}(\boldsymbol{x}) = d_f$ implies that all non-zero symbols (i.e. elements of $\mathbb{Z}_m$) of $\boldsymbol{x}$ have the same annihilator. In fact, $\mathrm{w_H}(\boldsymbol{x}) = d_f$ means that $\forall \boldsymbol{y} \in \mathcal{C} \setminus \{\boldsymbol{0}\}$ $\mathrm{w_H}(\boldsymbol{y}) \geq \mathrm{w_H}(\boldsymbol{x})$, which implies that

$$\nexists a \in \mathbb{Z}_m \text{ such that } 0 < \mathrm{w_H}(a\boldsymbol{x}) < \mathrm{w_H}(\boldsymbol{x})$$

and so, for all $a \in \mathbb{Z}_m$, either $a\boldsymbol{x} = \boldsymbol{0}$ or $\mathrm{w_H}(a\boldsymbol{x}) = d_f$ i.e. $a\boldsymbol{x}_i \neq 0$ for all $\boldsymbol{x}_i \neq 0$.

This remark implies that there exists $d|m$ (possibly $d = 1$) and there exists p_j a prime factor of m such that $\mathrm{w_H}(d\boldsymbol{x}) = \mathrm{w_H}(\boldsymbol{x})$ and $p_j\, d\, \boldsymbol{x} = \boldsymbol{0}$. Now, choosing $\boldsymbol{c} = d\,\boldsymbol{x}$ we have a codeword $\boldsymbol{c} \in \frac{m}{p_j}\mathcal{C}$ such that $\mathrm{w_H}(\boldsymbol{c}) = d_f$, so that we can conclude: $d_f(\phi_j) = d_f(\frac{m}{p_j}\phi) = d_f$. □

Finally, when proving Prop. 7.1 we need also the following simple lemma, even though it is just a property of $\mathbb{Z}_m$ and not of convolutional codes.

LEMMA A.3. *Given $a_1, \dots, a_m \in \mathbb{Z}_m \setminus \{0\}$, there exist indexes $\{i_1, \dots, i_n\} \subseteq \{1, \dots, m\}$ such that $a_{i_1} + \ldots + a_{i_n} = 0 \mod m$.*

Proof. By contradiction, assume that $\sum_{i\in\mathcal{I}} a_i \neq 0 \mod m$ for all non-empty $\mathcal{I} \subseteq \{1, \dots, m\}$. Then, in particular, $\sum_{i=1}^{n} a_i \neq 0 \mod m$ for all $n = 2, \dots, m$ and so $a_1 \notin \{-a_2, -a_2 - a_3, \dots, -\sum_{j=2}^{m} a_i\}$, which, being a set of $m-1$ distinct non-zero elements of $\mathbb{Z}_m$ is $\mathbb{Z}_m \setminus \{0\}$ itself, contradicting $a_1 \in \mathbb{Z}_m \setminus \{0\}$. □

REFERENCES

[1] L. BAZZI, M. MAHDIAN, AND D. SPIELMAN, *The minimum distance of turbo-like codes*, submitted to IEEE Trans. Inform. Theory (2003), available online: `http://math.mit.edu/~spielman/PAPERS/mindist.pdf`

[2] S. BENEDETTO AND G. MONTORSI, *Unveiling turbo codes: some results on parallel concatenated coding schemes*, IEEE Trans. Inform. Theory, 42 (1996), pp. 409–428.

[3] S. BENEDETTO AND G. MONTORSI, *Design of parallel concatenated convolutional codes*, IEEE Trans. Communications, 44 (1996), pp. 591–600.

[4] S. BENEDETTO, D. DIVSALAR, G. MONTORSI, AND F. POLLARA, *Serial concatenation of interleaved codes: performance analysis, design and iterative decoding*, IEEE Trans. Inform. Theory, 44 (1998), pp. 909–926.

[5] C. BERROU, A. GLAVIEUX, AND P. THITIMAJSHIMA, *Near Shannon Limit Error-Correctiong Coding and Decoding: Turbo Codes*, Proc. IEEE Int. Conf. Communications (1993), pp. 1064–1070.

[6] C. BERROU AND A. GLAVIEUX, *Near optimum error correcting coding and decoding: turbo-codes*, IEEE Trans. Communications, 44 (1996), pp. 1261–1271.

[7] C. BERROU, M. JÉZÉQUEL, C. DOUILLARD, AND S. KEROUÉDAN, *The advantages of non-binary turbo codes*, Proc. Inform. Theory Workshop (2001), pp. 61–63.

[8] G. COMO, F. FAGNANI, AND F. GARIN, *ML performances of serial turbo codes do not concentrate*, Proc. Intern. Symp. Turbo Codes (2006).

[9] D. DIVSALAR, *A simple tight bound on error probability of block codes with application to turbo codes*, JPL TDA Progress Report, 42-139 (1999), pp. 1–35.

[10] D. DIVSALAR, S. DOLINAR, AND F. POLLARA, *Iterative turbo decoder analysis based on density evolution*, IEEE J. Sel. Areas Communications, 19 (2001), pp. 891–907.

[11] C. DOUILLARD AND C. BERROU, *Turbo codes with rate-$m/m+1$ constituent convolutional codes*, IEEE Trans. Communications, 53 (2005), pp. 1630–1638.

[12] T. M. Duman and M. Salehi, *New performance bounds of turbo codes*, IEEE Trans. Communications, 46 (1998), pp. 717–723.
[13] S. Eilenberg, *Automata, machines, and languages. Vol A.*, Academic Press, 1974.
[14] H. El-Gamal and A. R. Hammons, *Analyzing the turbo decoder using the Gaussian approximation*, IEEE Trans. Inform. Theory, 47 (2001), pp. 671–686.
[15] F. Fagnani, *Performance of parallel concatenated coding schemes*, IEEE Trans. Inform. Theory, 54 (2008), pp. 1521–1535.
[16] F. Fagnani, B. Scanavino, S. Zampieri, and R. Garello, *Some results on combined parallel concatenated schemes with trellis-coded modulation*, Proc. Int. Symp. Inform. Theory (2002), p. 444.
[17] F. Fagnani and S. Zampieri, *Convolutional codes over finite Abelian groups: some basic results*, in Codes, systems and graphical models, B. Marcus and J. Rosenthal, eds, IMA Volumes in Mathematics and its applications, vol. 123, pp. 327–346, 2001.
[18] F. Fagnani and S. Zampieri, *System-theoretic properties of convolutional codes over rings*, IEEE Trans. Inform. Theory, 47 (2001), pp. 2256–2274.
[19] F. Fagnani and S. Zampieri, *Minimal and systematic convolutional codes over finite Abelian groups*, Linear Algebra Appl., 378 (2004), pp. 31–59.
[20] G. D. Forney, *Geometrically uniform codes*, IEEE Trans. Inform. Theory, 37 (1991), pp. 1241–1260.
[21] C. Fragouli, R.D. Wesel, *Turbo-Encoder design for symbol-interleaved parallel concatenated Trellis-Coded Modulation*, IEEE Trans. Communications, 49 (2001), pp. 425–435.
[22] R. G. Gallager, *Low Density Parity Check Codes*, Cambridge, MA: MIT Press, 1963.
[23] R. Garello, G. Montorsi, S. Benedetto, D. Divsalar, and F. Pollara, *Labelings and encoders with the uniform bit error property with applications to serially concatenated trellis codes*, IEEE Trans. Inform. Theory, 48 (2002), pp. 123–136.
[24] F. Garin, G. Como, and F. Fagnani, *Staircase and other structured linear-time encodable LDPC codes: analysis and design*, Proc. Int. Symp. Inform. Theory (2007), pp. 1226–1230.
[25] A. Graell i Amat, G. Montorsi, and F. Vatta, *Analysis and design of rate compatible serial concatenated convolutional codes*, Proc. Int. Symp. Inform. Theory (2005), pp. 607-611.
[26] T. W. Hungerford, *Algebra*, Springer-Verlag, 1974.
[27] R. E. Kalman, P. L. Falb, and M.A. Arbib, *Topics in mathematical system theory*, McGraw Hill, 1969.
[28] H. Jin, A. Khandekar, and R. J. McEliece, *Irregular Repeat-Accumulate Codes*, Proc. Intern. Symp. Turbo Codes (2000).
[29] H. Jin and R. J. McEliece, *Coding theorems for turbo code ensembles*, IEEE Trans. Inform. Theory, 48 (2002), pp. 1451–1461.
[30] R. Johannesson, Z.-X. Wan, and E. Wittenmark, *Some structural properties of convolutional codes over rings*, IEEE Trans. Inform. Theory, 44 (1998), pp. 839–845.
[31] S. Le Goff, A. Glavieux, and C. Berrou, *Turbo-codes and high spectral efficiency modulation*, Proc. IEEE Int. Conf. Communications (1994), pp. 645–649.
[32] H.-A. Loeliger, *Signal sets matched to groups*, IEEE Trans. Inform. Theory, 37 (1991), pp. 1675–1682.
[33] D. J. C. MacKay, *Good Error Correcting Codes Based On Very Sparse Matrices*, IEEE Trans. Inform. Theory, 45 (1999), pp. 399–431.
[34] H. Ogiwara, A. Mizutome, and K. Koike, *Performance evaluation of parallel concatenated Trellis-Coded Modulation*, IEICE Trans. Fundamentals, E84-A (2001), pp. 2410–2417.
[35] T. Richardson, *The geometry of turbo-decoding dynamics*, IEEE Trans. Inform. Theory, 46 (2000), pp. 9–23.
[36] P. Robertson and T. Wörz, *Novel bandwidth efficient coding scheme employing turbo-codes*, Proc. IEEE Int. Conf. Communications (1996), pp. 962–967.
[37] I. Sason and S. Shamai (Shitz) *Improved upper bounds on the ML decoding error probability of parallel and serially concatenated turbo codes via their ensemble distance spectrum*, IEEE Trans. Inform. Theory, 46 (2000), pp. 24–47.
[38] I. Sason, E. Telatar, and R. Urbanke, *The asymptotic input-output weight distributions and thresholds of convolutional and turbo-like encoders*, IEEE Trans. Inform. Theory, 48 (2002), pp. 3052–3061.
[39] S. Shamai (Shitz) and I. Sason, *Variations on the Gallager bounds, connections and applications*, IEEE Trans. Inform. Theory, 48 (2002), pp- 3029–3051.
[40] E. Telatar and R. Urbanke, *On the ensemble performance of turbo codes*, Proc. IEEE Int. Symp. Inform. Theory (1997), pp. 105–105.
[41] D. V. Truhachev, M. Lentmaier, and K. S. Zigangirov, *Some results concerning design and decoding of turbo-codes*, Probl. Inf. Transm., 37 (2001), pp. 190–205.