Average ML Asymptotic Performances of Different Serial Turbo Ensembles

Fabio Fagnani
Dipartimento di Matematica
Politecnico di Torino
C.so Duca degli Abruzzi 24
10129 Torino, Italy
Email: fabio.fagnani@polito.it

Roberto Garello
Dipartimento di Elettronica
Politecnico di Torino
C.so Duca degli Abruzzi 24
10129 Torino, Italy
Email: roberto.garello@polito.it

Federica Garin
Dipartimento di Matematica
Politecnico di Torino
C.so Duca degli Abruzzi 24
10129 Torino, Italy
Email: federica.garin@polito.it

***Abstract*— In this paper we study the ML error probability of serially concatenated schemes averaged over different interleaver ensembles. We prove asymptotic results when the interleaver length goes to infinity: differently from the parallel case, the choice of the ensemble can change the decreasing speed of error probability.**

I. Introduction

The uniform interleaver is a standard technique introduced by Benedetto and Montorsi [1] in order to analyze the performances of turbo codes: as the complexity of the analysis of a single concatenated code is too big, they fixed the component encoders and the interleaver length N and considered an interleaver uniformly drawn from the set of all permutations of N elements. By this approach, they studied the average error probability of parallel turbo codes, and particularly they provided asymptotic results when interleaver length N goes to infinity in the form of an upper bound to the average bit error probability: $\overline{P_b(e)} \leq CN^{-1}$ for some positive constant C only depending on the component encoders and on the channel signal-to-noise ratio (SNR). This bound is tight, as shown in [3], in the sense that[1] $\overline{P_b(e)} \asymp N^{-1}$.

An analogous study has been carried out also for serial turbo codes, (see [2], [6] and [5]); the result is that average bit and word error probabilities satisfy

$$\overline{P_b(e)} \asymp N^{-\lfloor (d_f^o+1)/2 \rfloor} \text{ and } \overline{P_w(e)} \asymp N^{-\lfloor (d_f^o-1)/2 \rfloor} \quad (1)$$

where d_f^o is the free distance (minimum Hamming weight) of the outer encoder.

Naturally the question arises if we can find some permutations which give better performances than the average. A way to study this problem is to do the same kind of average-based analysis, but on a subset of all interleavers. In order to re-use the same techniques, we need to look at subgroups of the permutation group (and we also need that the number of invariants of the subgroup does not grow with the interleaver length, see [3] for a precise definition of 'regular permutation group families').

[1]The symbol $a_n \asymp n^{-\alpha}$, for $n \to \infty$, means that $c_1/n^\alpha \leq a_n \leq c_2/n^\alpha$ for some positive constants c_1, c_2; in this paper, the constants can depend on the component encoders and on the channel SNR.

In the case of parallel concatenations, the study in [3] shows that for any choice of a regular family of permutation subgroups the average error probability has the same asymptotic behavior $\overline{P_b(e)} \asymp N^{-1}$. Some further study has then been done to find out, for particular subgroups, how the coefficient of the main term changes (see [4] for this in a more general setting).

The interesting fact is that, for serial concatenations instead, the exponent of N can change when we consider different subgroups, and there are indeed examples in which the speed of convergence is better than the one in (1).

All the results we have cited in this introduction, as well as all the results we will give in this paper, have to be considered coding theorems in the same sense as in [6]: there is an implicit assumption that the channel is memoryless binary-input symmetric-output (e.g. BSC or BIAWGN), whose noiseness is described by the Bhattacharyya parameter γ, and we state that there exists a threshold $\gamma_0 > 0$ such that, for any fixed $\gamma < \gamma_0$, our statements hold true. So, our statements are true for sufficiently good channels (small γ, high SNR), but without requiring that $\gamma \to 0$; we make no attempt here to give tight estimations of the threshold γ_0. Another implicit assumption is that decoding is maximum likelihood (ML).

II. Ensembles description

We fix an outer and an inner encoder ϕ^o and ϕ^i, which are both non-catastrophic convolutional encoders, with rates k/m and m/n respectively ($1 \leq k < m \leq n$) and controllability indexes (constraint lengths) ν^o and ν^i. We will need some more assumptions on the component encoders, which we will introduce later because they vary slightly from one ensemble to another.

We fix a parameter $N \in \mathbb{N}$, and we terminate ϕ^o after N trellis steps and ϕ^i after $M := N + \nu_o$ trellis steps, obtaining $\phi_N^o : \mathbb{Z}_2^{kN} \to \mathbb{Z}_2^{mM}$ and $\phi_N^i : \mathbb{Z}_2^{mM} \to \mathbb{Z}_2^{n(M+\nu^i)}$. We will consider a serially concatenated scheme, with as interleaver a permutation π_N of mM bits:

$$\xrightarrow{kN \text{ bits}} \boxed{\phi_N^o} \xrightarrow{mM \text{ bits}} \boxed{\pi_N} \xrightarrow{mM \text{ bits}} \boxed{\phi_N^i} \xrightarrow{n(M+\nu^i) \text{ bits}}$$

We will consider three different ensembles. In each of them, the interleaver is a random variable Π_N uniformly distributed

over a subgroup G_N of S_{mM} (the group of all the permutations of mM elements).

(E1) *Bit permutation*: $G_N = S_{mM}$. This case is the classical uniform interleaver: $\pi \in G_N$ is a permutation acting on all the mM bits of a codeword $\boldsymbol{c} \in \mathbb{Z}_2^{mM}$.

(E2) *Separate channels bit permutation*: $G_N = (S_M)^m$. Here we consider the output of the outer encoder as m separate channels, i.e. we read an outer codeword as an element $\boldsymbol{c} \in (\mathbb{Z}_2^M)^m$. Inside each channel, bits are randomly permuted, but different channels are never mixed, i.e. $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m) \in G_N$ acts on $\boldsymbol{c} = (c_1, \ldots, c_m) \in (\mathbb{Z}_2^M)^m$ in the following way: $\boldsymbol{\pi}(\boldsymbol{c}) = (\pi_1(c_1), \ldots, \pi_m(c_m))$.

(E3) *Symbol permutation*: $G_N = S_M$. Here we permute the symbols of the output of ϕ_N^o (i.e. the elements of $\mathbb{Z}_2^m$ coming out from the convolutional encoder at any trellis step). We can still read outer codewords as elements $(c_1, \ldots, c_m) \in (\mathbb{Z}_2^M)^m$ and consider the action $\pi(c_1, \ldots, c_m) = (\pi(c_1), \ldots, \pi(c_m))$, or equivalently we read the outer codewords as elements $(c_1, \ldots, c_M) \in (\mathbb{Z}_2^m)^M$, and a permutation $\pi \in G_N$ is a permutation acting on M symbols.

Notice that in the classical case (E1) the action of G_N keeps invariant the Hamming weight of the word, while in the case (E2) it keeps invariant the m Hamming weights of the m components, and in case (E3), for any symbol in $\mathbb{Z}_2^m$, the number of its appearances in the word does not change.

We will write $\mathrm{w_H}(\boldsymbol{c})$ to denote the Hamming weight (number of ones) of a codeword $\boldsymbol{c}$. Given $\boldsymbol{c} \in (\mathbb{Z}_2^M)^m$ we will write $\mathrm{w}_m(\boldsymbol{c})$ for the Hamming weights of its m components: $\mathrm{w}_m(c_1, \ldots, c_m) = (\mathrm{w_H}(c_1), \ldots, \mathrm{w_H}(c_m))$. Given $\boldsymbol{c} \in (\mathbb{Z}_2^m)^M$, $\mathrm{w_s}(\boldsymbol{u}) \in \mathbb{N}^{2^m-1}$ will be a vector indicating, for each non-zero element of $\mathbb{Z}_2^m$, the number of its appearances in $\boldsymbol{c}$. Given $\boldsymbol{h} = (h_1, \ldots, h_a) \in \mathbb{N}^a$, we will write $|\boldsymbol{h}| = h_1 + \ldots + h_a$. So, $|\mathrm{w_H}(\boldsymbol{u})| = \mathrm{w_H}(\boldsymbol{u})$ and $|\mathrm{w}_m(\boldsymbol{u})| = \mathrm{w_H}(\boldsymbol{u})$, while $|\mathrm{w_s}(\boldsymbol{u})|$ is the number of non-zero symbols in $\boldsymbol{u}$ (a sort of Hamming weight defined on symbols instead of bits).

We will write $\rho(\boldsymbol{u}) = \mathrm{w_H}(\boldsymbol{u})$ in case (E1), $\rho(\boldsymbol{u}) = \mathrm{w}_m(\boldsymbol{u})$ in case (E2) and $\rho(\boldsymbol{u}) = \mathrm{w_s}(\boldsymbol{u})$ in case (E3) and we will call $\rho(\boldsymbol{u})$ the vector weight of $\boldsymbol{u}$. With these notations, the above remark on the invariants of the action of G_N can be restated more precisely: $\rho(\boldsymbol{u}) = \rho(\boldsymbol{v}) \Leftrightarrow \exists \pi \in G_N : \boldsymbol{v} = \pi(\boldsymbol{u})$.

A very important parameter in the analysis of average performances of serial ensembles is d_f^o, the free distance of the outer encoder. Anyhow, for our third ensemble, the most relevant weight is not the Hamming weight, but the number of non-zero symbols, and hence we define the *symbol free distance* of the outer encoder as $d_s^o := \min_{\boldsymbol{u}}\{|\mathrm{w_s}(\phi^o(\boldsymbol{u}))|\}$. Clearly $d_s^o \leq d_f^o$. We will write d^o to mean d_f^o for (E1) and (E2) and d_s^o for (E3).

We state now the assumptions we need on ϕ^o and ϕ^i, in addition to non-catastrophicity (the reasons will be clear later):

- ϕ^o must have $d^o \geq 3$;
- ϕ^i must be recursive in the following sense:
$$|\rho(\boldsymbol{u})| = 1 \Rightarrow \mathrm{w_H}(\phi^i(\boldsymbol{u})) = \infty\,.$$

Notice that what we are requiring is standard recursiveness of ϕ^i for (E1) and (E2), while for (E3) we ask a more restrictive assumption, a recursiveness with respect to symbols: no input with only one non-zero symbol (regardless to the Hamming weight of this symbol) can produce a finite weight output.

III. Average performances

For the classical ensemble (E1), it is well-known that

$$\overline{P_b(e)} \asymp N^{-\alpha} \text{ and } \overline{P_w(e)} \asymp N^{-\alpha+1} \tag{2}$$

with $\alpha = \lfloor (d_f^o+1)/2 \rfloor$ (see [2],[6], [5]). We claim that also for (E2) and (E3) there exists $\alpha > 0$ such that Eq. (2) holds true; in this section, we will characterize this exponent. A sketch of the proofs is presented in next section, while in section V there is an example of component encoders for which ensemble (E2) performs better than the others.

Let $\mathcal{C}^{o,N} = \phi^{o,N}(\mathbb{Z}_2^{kN}) \subseteq \mathbb{Z}_2^{mM}$ be the outer block code and let $H = \{\rho(\boldsymbol{c}) : \boldsymbol{c} \in \mathcal{C}^{o,N} \text{ for some } N\}$. Following [1], we define an error event of a convolutional encoder as a codeword whose corresponding trellis state sequence, for some $t_1 < t_2$, is zero for all $t \leq t_1$ and $t > t_2$, and is non-zero for all $t_1 < t \leq t_2$. We call $[t_1, t_2]$ the support and $t_2 - t_1$ the length of the event. When an encoder is terminated after N trellis steps, an error event is said to be regular if $t_2 \leq N$, otherwise it is called terminated. For a terminated event, we call $N - t_1$ its length. Any codeword $\boldsymbol{c}$ of a terminated convolutional encoder is the sum of some $n(\boldsymbol{c})$ regular error events, plus possibly a terminated one, all with non-overlapping supports. We define:

- $n_o(\boldsymbol{h}) = \max\{n(\boldsymbol{c}) : \boldsymbol{c} \in \mathcal{C}^{o,N}, \rho(\boldsymbol{c}) = \boldsymbol{h}\}$
- $n_i(\boldsymbol{h}) = \max\{n(\boldsymbol{x}) : \boldsymbol{x} = \phi^{i,N}(\boldsymbol{u}), \rho(\boldsymbol{u}) = \boldsymbol{h}\}$
- $f(\boldsymbol{h}) = 1 + |\boldsymbol{h}| - n_o(\boldsymbol{h}) - n_i(\boldsymbol{h})$

It is clear that for any given $\boldsymbol{h}$, the above definitions do not depend on N, if N is chosen sufficiently large, fact that we will always assume from now on. 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 $n_i(\boldsymbol{h})$. Notice that, for all $\boldsymbol{h} \in H$, $1 \leq n_o(\boldsymbol{h}) \leq \lfloor |\boldsymbol{h}|/d^o \rfloor$ and $0 \leq n_i(\boldsymbol{h}) \leq \lfloor |\boldsymbol{h}|/2 \rfloor$ and then

$$\lfloor (d^o+1)/2 \rfloor \leq 1 + |\boldsymbol{h}| - \lfloor |\boldsymbol{h}|/d^o \rfloor - \lfloor |\boldsymbol{h}|/2 \rfloor \leq f(\boldsymbol{h}) \leq |\boldsymbol{h}| \tag{3}$$

We define:

$$\alpha = \min\{f(\boldsymbol{h}),\, \boldsymbol{h} \in H\}\,. \tag{4}$$

Our main result is that, for such α, (2) holds true.

As clearly there exists $\boldsymbol{h} \in H$ with $|\boldsymbol{h}| = d^o$, (3) gives

$$\lfloor (d^o + 1)/2 \rfloor \leq \alpha \leq d^o \tag{5}$$

For (E1), the well-known equality $\alpha = \lfloor (d^o + 1)/2 \rfloor$ comes from the following property of convolutional encoders, which implies that $n_i(d_f^o) = \lfloor d_f^o/2 \rfloor$:

Proposition 1: There exists $\delta \in \mathbb{N}$ such that, for any $j \in \{1, \ldots, m\}$, input $(1 + e_j D^\delta)$ (i.e. a word made of all zeros except two e_j spaced δ apart, where e_j is a vector in $\mathbb{Z}_2^m$ with all zeros except a one in position j) produces as output of ϕ^i a regular error event. □

If we call α_j the exponent corresponding to ensemble (Ej), inequality (5) and the above expression for α_1 imply that $\alpha_2 \leq \alpha_1$; an example where strict inequality holds true is given in section V. On the contrary, no general ordering holds true for α_3 and α_1 or α_2.

Now, we define the set of the vectors $\boldsymbol{h}$ minimizing $f(\boldsymbol{h})$: $\mathcal{H} = \{\boldsymbol{h} \in H : f(\boldsymbol{h}) = \alpha\}$. Inequalities (3) and (5) imply that $\mathcal{H}$ is a finite set. In fact, if $f(\boldsymbol{h}) = \alpha$,
$d^o \geq f(\boldsymbol{h}) \geq 1 + |\boldsymbol{h}| - \lfloor |\boldsymbol{h}|/d^o \rfloor - \lfloor |\boldsymbol{h}|/2 \rfloor \geq 1 + |\boldsymbol{h}| \left(\frac{1}{2} - \frac{1}{d^o}\right)$
which gives $|\boldsymbol{h}| \leq \lfloor 2d^o(d^o-1)/(d^o-2) \rfloor \leq 4d^o$.

We define:

- $d^*(\boldsymbol{h}) = \min\{\mathrm{w_H}(\boldsymbol{x}) : \boldsymbol{x} = \phi^i(\boldsymbol{u}), \rho(\boldsymbol{u}) = \boldsymbol{h}, n(\boldsymbol{x}) = n_i(\boldsymbol{h})\}$
- $d^* = \min\{d^*(\boldsymbol{h}) : \boldsymbol{h} \in \mathcal{H}\}$.

With the above definitions, we can state our main result in a form that underlines, additionally to the exponent of N, also the dependence of the coefficient on the channel SNR:

Theorem 1: There exist positive constants c_1, c_2 and c_3 (c_1, c_2 depending only on the component encoders, not on the SNR, c_3 possibly depending on the SNR) such that:

- $c_1 p^{d^*} N^{-\alpha} \leq \overline{P_b(e)} \leq c_2 \gamma^{d^*} N^{-\alpha} + c_3 N^{-\alpha-1}$;
- $kc_1 p^{d^*} N^{-\alpha+1} \leq \overline{P_w(e)} \leq kc_2 \gamma^{d^*} N^{-\alpha+1} + kc_3 N^{-\alpha}$;

where p is the equivocation probability and γ is the Bhattacharyya noise parameter of the channel. □

For a definition of p and γ see e.g. [3] and [6]. For the BIAWGN channel, $p = 1/2 \operatorname{erfc} \sqrt{E_s/N_0}$ and $\gamma = e^{-E_s/N_0}$; note that they exhibit a quite similar dependence on the SNR per transmitted bit E_s/N_0.

IV. Sketch of the proofs

In this section we give the outline of our proofs. We prove the upper bound for $\overline{P_b(e)}$ and the lower bound for $\overline{P_w(e)}$; then Thm. 1 follows by the trivial remark $\overline{P_b(e)} \geq \frac{1}{kN} \overline{P_w(e)}$.

A. Upper bound

This proof is based on the union-Bhattacharyya bound (see e.g. [6]) and on estimations of the weight enumerating coefficients of the component encoders.

The union-Bhattacharyya bound gives:

$$\overline{P_b(e)} \leq \sum_w \sum_d \frac{w}{kN} \overline{A_{w,d}}^N \gamma^d \tag{6}$$

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

Recalling which are the invariants under the action of our permutation groups, we can express $\overline{A_{w,d}}^N$ as a function of proper enumerating coefficients of the component encoders: we define $A^{o,N}_{w,\boldsymbol{h}}$ to be the number of codewords of ϕ^o_N with input Hamming weight w and output vector weight $\boldsymbol{h}$, $A^{i,N}_{\boldsymbol{h},d}$ to be the number of codewords of ϕ^i_N with input vector weight $\boldsymbol{h}$ and output Hamming weight d.

Proposition 2: $\overline{A_{w,d}}^N = \sum_{\boldsymbol{h} \in H} \frac{1}{\mathcal{M}_{\boldsymbol{h}}} A^{o,N}_{w,\boldsymbol{h}} A^{i,N}_{\boldsymbol{h},d}$, where:

$\mathcal{M}_{\boldsymbol{h}} = \binom{mM}{h}$ for (E1), $\mathcal{M}_{\boldsymbol{h}} = \binom{M}{h_1}\binom{M}{h_2}\cdots\binom{M}{h_m}$ for (E2) and $\mathcal{M}_{\boldsymbol{h}} = \binom{M}{\boldsymbol{h}}$ for (E3). □

Notice that, for all our three ensembles,

$$\left(\frac{M}{e|\boldsymbol{h}|}\right)^{|\boldsymbol{h}|} \leq \mathcal{M}_{\boldsymbol{h},N} \leq (mM)^{|\boldsymbol{h}|}. \tag{7}$$

Thus, by (6) and Prop. 2 we have

$$\overline{P_b(e)} \leq \sum_{w,\boldsymbol{h},d} \frac{w}{kN} \left[\frac{e|\boldsymbol{h}|}{M}\right]^{|\boldsymbol{h}|} A^{o,N}_{w,\boldsymbol{h}} A^{i,N}_{\boldsymbol{h},d} \gamma^d \tag{8}$$

Remark 1: The indexes of the summations clearly must satisfy $w \leq kN$, $\boldsymbol{h} \in H$ (which implies $d^o \leq |\boldsymbol{h}| \leq mM$) and $d \leq n(M + \nu_i)$. Moreover, the three indexes are related to each other. A first remark is that, for each $\boldsymbol{h}$, the output weight must be $d \geq d^*(\boldsymbol{h})$. □

Other relevant relations involving w, $\boldsymbol{h}$ and d are given by the following property, which follows from non-catastrophicity.

Proposition 3: There exist two positive constants μ_o and μ_i (trivially equal to 1 when the encoders are systematic) such that the summation in (8) is only over $w \leq \mu_o |\boldsymbol{h}|$ and over $\boldsymbol{h}$ satisfying $|\boldsymbol{h}| \leq \mu_i d$. □

We will use the following estimations of the enumerating coefficients (see [3] for an idea of the proof):

Proposition 4: For some positive constants a_o, a_i, b_o, b_i:

- $A^{o,N}_{w,\boldsymbol{h}} \leq \sum_{n_o=1}^{n_o(\boldsymbol{h})} \binom{N+n_o}{n_o} a_o^w b_o^{|\boldsymbol{h}|}$
- $A^{i,N}_{\boldsymbol{h},d} \leq \sum_{n_i=0}^{n_i(\boldsymbol{h})} \binom{N+n_i}{n_i} a_i^{|\boldsymbol{h}|} b_i^d$ □

Then, by estimations of the binomial coefficients, we prove the following inequality: for some positive constant C,

$$\left[\frac{e|\boldsymbol{h}|}{M}\right]^{|\boldsymbol{h}|} \sum_{n_o=1}^{n_o(\boldsymbol{h})} \sum_{n_i=0}^{n_i(\boldsymbol{h})} \binom{N+n_o}{n_o}\binom{N+n_i}{n_i} \leq C^{|\boldsymbol{h}|} \frac{|\boldsymbol{h}|^{f(\boldsymbol{h})-1}}{N^{f(\boldsymbol{h})-1}}$$

Finally, substituting all the estimations into (8), we get, for some positive constants C_1, C_2, C_3:

$$\overline{P_b(e)} \leq \sum_{w,\boldsymbol{h},d} \frac{|\boldsymbol{h}|^{f(\boldsymbol{h})-1}}{N^{f(\boldsymbol{h})}} C_1^w C_2^{|\boldsymbol{h}|} C_3^d \gamma^d \tag{9}$$

where the indexes of the summation satisfy the inequalities described in Remark 1 and Prop. 3.

Now, we split the summation into two terms, separating $\boldsymbol{h} \in \mathcal{H}$ from $\boldsymbol{h} \notin \mathcal{H}$. In the following we show that the first term is bounded by $c\gamma^{d^*} N^{-\alpha}$, while the second is bounded by $c'(\gamma) N^{-\alpha-1}$, thus ending the proof.

For the first term, remember that $\mathcal{H}$ is finite and so also when $N \to \infty$ the summation over $\boldsymbol{h}$ has a finite number of terms. As $w \leq \mu_o |\boldsymbol{h}|$, also the summation over w has finitely many terms. Finally notice that, for each $\boldsymbol{h} \in \mathcal{H}$,

$$\sum_{d=d^*(\boldsymbol{h})}^{n(M+\nu_i)} (C_3\gamma)^d \leq \sum_{d=d^*}^{+\infty} (C_3\gamma)^d \leq C_4 (C_3\gamma)^{d^*}$$

(the last inequality holds true, for some $C_4 > 0$, if $\gamma < 1/C_3$).

The second term is equal to

$$\frac{1}{N^{\alpha+1}} \sum_{d,\boldsymbol{h}\notin\mathcal{H},w} C_1^w \left(\frac{|\boldsymbol{h}|}{N}\right)^{f(\boldsymbol{h})-\alpha-1} |\boldsymbol{h}|^\alpha C_2^{|\boldsymbol{h}|} C_3^d \gamma^d \tag{10}$$

Note that $\boldsymbol{h} \in H \backslash \mathcal{H}$ implies $f(\boldsymbol{h}) - \alpha - 1 \geq 0$ and so, as surely $|\boldsymbol{h}| \leq mM < 2mN$, and trivially $f(\boldsymbol{h}) - \alpha - 1 < |\boldsymbol{h}|$, we have $(|\boldsymbol{h}|/N)^{f(\boldsymbol{h})-\alpha-1} < (2m)^{|\boldsymbol{h}|}$. As $\sum_{w \leq \mu_o|\boldsymbol{h}|} C_1^w \leq \mu_o |\boldsymbol{h}| C_1^{|\boldsymbol{h}|}$, the second term (10) is bounded by

$$\frac{1}{N^{\alpha+1}} \sum_{d,\boldsymbol{h}} \mu_o |\boldsymbol{h}| C_1^{|\boldsymbol{h}|} (2m)^{|\boldsymbol{h}|} |\boldsymbol{h}|^\alpha C_2^{|\boldsymbol{h}|} C_3^d \gamma^d$$

For each d, the summation over $\boldsymbol{h}$ is for $|\boldsymbol{h}| \leq \mu_i d$ and, for a constant $K > 1$, $\sum_{\boldsymbol{h}:|\boldsymbol{h}|\leq\mu_i d} \mu_o |\boldsymbol{h}| C_1^{|\boldsymbol{h}|} (2m)^{|\boldsymbol{h}|} |\boldsymbol{h}|^\alpha C_2^{|\boldsymbol{h}|} \leq K^d$.

Finally, $\sum_{d\in\mathbb{N}}(KC_3\gamma)^d = c'(\gamma) < \infty$ if $\gamma < 1/(KC_3)$.

B. Lower bound

The main idea is the same used in [3] for parallel ensembles:

- given $\pi_N \in G_N$, if the minimum distance $d_f(\pi_N)$ of the serially concatenated scheme with interleaver π_N satisfies $d_f(\pi_N) \leq d$, then the word error probability of such scheme, $P_w(e|\pi_N)$, satisfies $P_w(e|\pi_N) \geq p^d$;
- for all $d > 0$, $\overline{P_w(e)} \geq p^d\,\mathbb{P}(d_f(\Pi_N) \leq d)$, where $\mathbb{P}$ is the probability defined over the ensemble we are considering.

For some fixed values of d, we can find a lower bound to $\mathbb{P}(d_f(\Pi_N) \leq d)$. This is the result we will prove:

Proposition 5: For any $\boldsymbol{h} \in H$, there exists $C > 0$ such that $\mathbb{P}(d_f(\Pi_N) \leq d^*(\boldsymbol{h})) \geq CN^{|\boldsymbol{h}|-n_o(\boldsymbol{h})-n_i(\boldsymbol{h})}$. □

If we choose $\boldsymbol{h} \in \mathcal{H}$ such that $d^*(\boldsymbol{h}) = d^*$, this result gives $\mathbb{P}(d_f(\Pi_N) \leq d^*) \geq CN^{-\alpha+1}$ and so $\overline{P_w(e)} \geq p^{d^*}CN^{-\alpha+1}$.

We now give the outline of the proof of Prop. 5. We fix once and for all the following objects:

1) A weight vector $\boldsymbol{h} \in H$.
2) An outer codeword $c^* \in \mathcal{C}^{o,N}$ such that $\rho(c^*) = \boldsymbol{h}$, consisting of $n_o = n_o(\boldsymbol{h})$ regular error events $c_1^*, \ldots, c_{n_o}^*$. Denote by l_k the length of c_k^*.
3) An input codeword u^* for the inner encoder, such that $\rho(u^*) = \boldsymbol{h}$ and such that $x^* = \phi^{i,N}(u^*)$ has weight $\mathrm{w_H}(x^*) = d^*(\boldsymbol{h})$ and is the concatenation of $n_i = n_i(\boldsymbol{h})$ regular error events $x_1^*, \ldots, x_{n_i}^*$ plus a possible terminated one $x_{n_i+1}^*$. Denote by u_k^* the input corresponding to x_k^* and by λ_k its length.

Notice that c^* can be chosen not to depend on N, while this may not be possible for u^*. However, we can assume that the error events x_k^* and their inputs u_k^* remain the same apart from some possible translations.

We now select a number of possible recombinations for both c^* and u^* where their respective error events are permuted and moved across the entire time axis $[0, M-1]$.

Let us start with c^*. First we fix n_o consecutive intervals having length respectively $M_k = \left\lfloor \frac{N}{n_o l_k} \right\rfloor l_k$ for $k = 1, \ldots, n_o$. Consider now $\mathcal{A} = \prod_{k=1}^{n_o}[0, M_k/l_k - 1]$. Given $\boldsymbol{a} \in \mathcal{A}$ we define $c_{\boldsymbol{a}}^*$ to be the outer codeword which, for every $k = 1, \ldots, n_o$, contains exactly one translated copy of the error event c_k^* starting at time $a_k l_k + \sum_{j<k} M_j$.

A similar job is done on the inner input word u^*. In this case we split the time axis into $n_i + 1$ consecutive intervals having length respectively $Q_k = \left\lfloor \frac{M-\lambda_{n_i+1}}{n_i\lambda_k} \right\rfloor \lambda_k$ for $k = 1, \ldots, n_i$ and a last one of length $Q_{n_i+1} \geq \lambda_{n_i+1}$. Consider now $\mathcal{B} = \prod_{k=1}^{n_i}[0, Q_k/\lambda_k - 1]$. Given $\boldsymbol{b} \in \mathcal{B}$ we let $u_{\boldsymbol{b}}^*$ be the inner input word which, for every $k = 1, \ldots, n_i$, contains exactly one translated copy of the input error event u_k^* starting at time $b_k\lambda_k + \sum_{j<k} Q_j$, while the terminated one remains fixed in its position in the interval $[M - Q_{n_i+1}, M]$. Let $x_{\boldsymbol{b}}^*$ be the output $x_{\boldsymbol{b}}^* = \phi^{i,N}(u_{\boldsymbol{b}}^*)$.

Given $\boldsymbol{a} \in \mathcal{A}$ and $\boldsymbol{b} \in \mathcal{B}$, we define the events

$$E_{\boldsymbol{a},\boldsymbol{b}} = \{\Pi_N(c_{\boldsymbol{a}}^*) = u_{\boldsymbol{b}}^*\} \quad \text{and} \quad E_{\boldsymbol{a}} = \bigcup_{\boldsymbol{b}\in\mathcal{B}} E_{\boldsymbol{a},\boldsymbol{b}}\,.$$

Notice that $E_{\boldsymbol{a}}$ is an union of disjoint events. Clearly $\pi_N \in E_{\boldsymbol{a},\boldsymbol{b}}$ implies $d_f(\pi_N) \leq \mathrm{w_H}(x_{\boldsymbol{b}}^*) = d^*(\boldsymbol{h})$, so:

$$\mathbb{P}\left(d_f(\Pi_N) \leq d^*(\boldsymbol{h})\right) \geq \mathbb{P}\left(\cup_{\boldsymbol{a}\in\mathcal{A}} E_{\boldsymbol{a}}\right)\,.$$

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

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

We will prove a lower bound for the first term and an upper bound for the second term, thus ending the proof of Prop. 5. In both parts, we will use the following remark: $\forall \boldsymbol{u}, \boldsymbol{v} \in \mathbb{Z}_2^{mM}$,

$$\mathbb{P}(\Pi_N(\boldsymbol{u}) = \boldsymbol{v}) = \begin{cases} 1/\mathcal{M}_{\boldsymbol{h}} & \text{if } \rho(\boldsymbol{u}) = \rho(\boldsymbol{v}) = \boldsymbol{h} \\ 0 & \text{if } \rho(\boldsymbol{u}) \neq \rho(\boldsymbol{v}) \end{cases} \tag{11}$$

Consider the first term: by Eq. (11) and (7), we have

$$\sum_{\boldsymbol{a}\in\mathcal{A}} \mathbb{P}(E_{\boldsymbol{a}}) = \sum_{\boldsymbol{a}\in\mathcal{A}}\sum_{\boldsymbol{b}\in\mathcal{B}} \mathbb{P}(E_{\boldsymbol{a},\boldsymbol{b}}) = \frac{\#\mathcal{A}\#\mathcal{B}}{\mathcal{M}_{\boldsymbol{h}}} \asymp N^{n_o+n_i-|\boldsymbol{h}|}\,.$$

It remains to estimate the second term. We have:

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

Our aim is to give a bound to the number of pairs $(\boldsymbol{a},\boldsymbol{b}), (\boldsymbol{a}',\boldsymbol{b}')$ such that $E_{\boldsymbol{a},\boldsymbol{b}} \cap E_{\boldsymbol{a}',\boldsymbol{b}'} \neq \emptyset$ and a bound to $\mathbb{P}(E_{\boldsymbol{a},\boldsymbol{b}} \cap E_{\boldsymbol{a}',\boldsymbol{b}'})$.

Assume we have fixed pairs $(\boldsymbol{a},\boldsymbol{b}), (\boldsymbol{a}',\boldsymbol{b}') \in \mathcal{A}\times\mathcal{B}$ with $\boldsymbol{a} \neq \boldsymbol{a}'$ and such that $E_{\boldsymbol{a},\boldsymbol{b}} \cap E_{\boldsymbol{a}',\boldsymbol{b}'} \neq \emptyset$. This immediately implies that $\boldsymbol{b} \neq \boldsymbol{b}'$. Let $\mathrm{d_H}(\boldsymbol{a},\boldsymbol{a}')$ be the number of components of the vectors $\boldsymbol{a}$ and $\boldsymbol{a}'$ which are different (the Hamming weight of $\boldsymbol{a} - \boldsymbol{a}'$) and analogously define $\mathrm{d_H}(\boldsymbol{b},\boldsymbol{b}')$. We have $1 \leq \mathrm{d_H}(\boldsymbol{a},\boldsymbol{a}') \leq n_o$ and $1 \leq \mathrm{d_H}(\boldsymbol{b},\boldsymbol{b}') \leq n_i$.

By the definition of $c_{\boldsymbol{a}}^*$ and $c_{\boldsymbol{a}'}^*$, we can find outer codewords $\tilde{c}^*$, $\tilde{c}_{\boldsymbol{a}}^*$, $\tilde{c}_{\boldsymbol{a}'}^*$ (possibly $\tilde{c}^* = 0$) having disjoint supports, each consisting of some of the error events c_k^*, such that $c_{\boldsymbol{a}}^* = \tilde{c}^* + \tilde{c}_{\boldsymbol{a}}^*$ and $c_{\boldsymbol{a}'}^* = \tilde{c}^* + \tilde{c}_{\boldsymbol{a}'}^*$. More precisely, letting $\tilde{n}_o = \mathrm{d_H}(\boldsymbol{a},\boldsymbol{a}')$, $\tilde{c}^*$ consists of $n_o - \tilde{n}_o$ error events, and $\tilde{c}_{\boldsymbol{a}}^*$, $\tilde{c}_{\boldsymbol{a}'}^*$ consists of the same $\tilde{n}_o$ error events, only shifted in different positions. Clearly, $\rho(\tilde{c}_{\boldsymbol{a}}^*) = \rho(\tilde{c}_{\boldsymbol{a}'}^*) = \boldsymbol{h} - \rho(\tilde{c}^*)$. Also notice that $\tilde{n}_o \leq \lfloor|\rho(\tilde{c}_{\boldsymbol{a}'}^*)|/d^o\rfloor$ and that $|\rho(\tilde{c}_{\boldsymbol{a}'}^*)| \geq d^o$ because $\tilde{n}_o \geq 1$.

Similarly, we can find inner input words $\tilde{u}^*$, $\tilde{u}_{\boldsymbol{b}}^*$, $\tilde{u}_{\boldsymbol{b}'}^*$ (possibly $\tilde{u}^* = 0$) having disjoint supports, each consisting of some of the input error events u_k^*, such that $u_{\boldsymbol{b}}^* = \tilde{u}^* + \tilde{u}_{\boldsymbol{b}}^*$ and $u_{\boldsymbol{b}'}^* = \tilde{u}^* + \tilde{u}_{\boldsymbol{b}'}^*$. Letting $\tilde{n}_i = \mathrm{d_H}(\boldsymbol{b},\boldsymbol{b}')$, $\tilde{u}^*$ has $n_i - \tilde{n}_i$ error events and $\tilde{u}_{\boldsymbol{b}}^*$, $\tilde{u}_{\boldsymbol{b}'}^*$ are made by the same $\tilde{n}_i$ error events shifted in different positions. Clearly, $\rho(\tilde{u}_{\boldsymbol{b}}^*) = \rho(\tilde{u}_{\boldsymbol{b}'}^*) = \boldsymbol{h} - \rho(\tilde{u}^*)$. By ρ–recursiveness of ϕ^i, $\tilde{n}_i \leq \lfloor|\rho(\tilde{u}_{\boldsymbol{b}'}^*)|/2\rfloor$.

The fundamental remark is that if $\pi_N \in E_{\boldsymbol{a},\boldsymbol{b}} \cap E_{\boldsymbol{a}',\boldsymbol{b}'}$ then $\pi_N(\tilde{c}^*) = \tilde{u}^*$, $\pi_N(\tilde{c}_{\boldsymbol{a}}^*) = \tilde{u}_{\boldsymbol{b}}^*$ and $\pi_N(\tilde{c}_{\boldsymbol{a}'}^*) = \tilde{u}_{\boldsymbol{b}'}^*$.

This implies that $\rho(\tilde{u}^*) = \rho(\tilde{c}^*)$ and that $\rho(\tilde{u}_{\boldsymbol{b}}^*) = \rho(\tilde{u}_{\boldsymbol{b}'}^*) = \rho(\tilde{c}_{\boldsymbol{a}}^*) = \rho(\tilde{c}_{\boldsymbol{a}'}^*) = \boldsymbol{h} - \rho(\tilde{c}^*)$. We will use the notation $\tilde{\boldsymbol{h}} = \rho(\tilde{u}_{\boldsymbol{b}}^*)$. We have: $|\tilde{\boldsymbol{h}}| \leq |\boldsymbol{h}|$, $|\tilde{\boldsymbol{h}}| \geq d^o$, $1 \leq \tilde{n}_o \leq \lfloor|\tilde{\boldsymbol{h}}|/d^o\rfloor$ and $1 \leq \tilde{n}_i \leq \lfloor|\tilde{\boldsymbol{h}}|/2\rfloor$.

Now we define $\tilde{c} = \tilde{c}^* + \tilde{c}_{\boldsymbol{a}}^* + \tilde{c}_{\boldsymbol{a}'}^*$ and $\tilde{u} = \tilde{u}^* + \tilde{u}_{\boldsymbol{b}}^* + \tilde{u}_{\boldsymbol{b}'}^*$. By construction, $\tilde{c}$ is made by $n_o + \mathrm{d_H}(\boldsymbol{a},\boldsymbol{a}')$ regular error events,

and $\tilde{u}$ gives $n_i + \mathrm{d_H}(\boldsymbol{b}, \boldsymbol{b}')$ regular error events; their invariants weight is $\rho(\tilde{c}) = \rho(\tilde{u}) = \boldsymbol{h} + \tilde{\boldsymbol{h}}$. The above fundamental remark also implies that $E_{\boldsymbol{a},\boldsymbol{b}} \cap E_{\boldsymbol{a}',\boldsymbol{b}'} \subseteq \{\Pi_N(\tilde{c}) = \tilde{u}\}$. Hence, by (7) and (11), $\mathbb{P}(E_{\boldsymbol{a},\boldsymbol{b}} \cap E_{\boldsymbol{a}',\boldsymbol{b}'}) \leq 1/\mathcal{M}_{\boldsymbol{h}+\tilde{\boldsymbol{h}}} \asymp N^{-(|\boldsymbol{h}|+|\tilde{\boldsymbol{h}}|)}$.

We can now end the proof:

$$\begin{aligned}
&\sum_{\substack{\boldsymbol{a},\boldsymbol{a}' \in \mathcal{A} \\ \boldsymbol{a}' \neq a}} \sum_{\boldsymbol{b},\boldsymbol{b}' \in \mathcal{B}} \mathbb{P}(E_{\boldsymbol{a},\boldsymbol{b}} \cap E_{\boldsymbol{a}',\boldsymbol{b}'}) \\
&= \sum_{\substack{\tilde{\boldsymbol{h}}: \\ d^o \leq |\tilde{\boldsymbol{h}}| \leq |\boldsymbol{h}|}} \sum_{\substack{\boldsymbol{a},\boldsymbol{a}' \in \mathcal{A}:\, \boldsymbol{a} \neq \boldsymbol{a}', \\ \mathrm{d_H}(\boldsymbol{a},\boldsymbol{a}') \leq \lfloor |\tilde{\boldsymbol{h}}|/d^o \rfloor}} \sum_{\substack{\boldsymbol{b},\boldsymbol{b}' \in \mathcal{B}:\, \boldsymbol{b} \neq \boldsymbol{b}', \\ \mathrm{d_H}(\boldsymbol{b},\boldsymbol{b}') \leq \lfloor |\tilde{\boldsymbol{h}}|/2 \rfloor}} \mathbb{P}(E_{\boldsymbol{a},\boldsymbol{b}} \cap E_{\boldsymbol{a}',\boldsymbol{b}'}) \\
&\leq \sum_{\tilde{\boldsymbol{h}}: d^o \leq |\tilde{\boldsymbol{h}}| \leq |\boldsymbol{h}|} c\, N^{n_o + \lfloor |\tilde{\boldsymbol{h}}|/d^o \rfloor + n_i + \lfloor |\tilde{\boldsymbol{h}}|/2 \rfloor - |\boldsymbol{h}| - |\tilde{\boldsymbol{h}}|} \\
&\leq C(\boldsymbol{h})\, N^{n_o + n_i - |\boldsymbol{h}| - \varepsilon}
\end{aligned}$$

where $\varepsilon = \min\{f(\tilde{\boldsymbol{h}}) : d^o \leq |\tilde{\boldsymbol{h}}| \leq |\boldsymbol{h}|\}$ and $C(\boldsymbol{h}) > 0$. As $\boldsymbol{h} \in H$, $\varepsilon \geq \lfloor (d^o - 1)/2 \rfloor \geq 1$. ■

V. Example

We give an example of a serial turbo scheme for which the ensemble (E2) has better average performances than (E1) and (E3) and we comment the reasons of the improvement.

We consider the following outer and inner encoders:

$$\phi^o = \left[1, \tfrac{1}{1+D+D^3}\right] \qquad \phi^i = \begin{bmatrix} \frac{1}{1+D} & 0 \\ 0 & \frac{1}{1+D} \end{bmatrix}$$

These encoders satisfy the assumptions of our analysis:

- both encoders are non-catastrophic;
- ϕ^o has free distance $d_f^o = 4$ and $d_s^o = 3$;
- ϕ^i is recursive and symbol-recursive.

We calculate the parameters α and d^* of Thm. 1 for these encoders. To do so, we need to notice that all the words $\boldsymbol{c}$ of the outer code such that $\mathrm{w_H}(\boldsymbol{c}) = d_f^o$ are obtained when input is $1 + D + D^3$ or its shifts and have $\mathrm{w}_m(\boldsymbol{c}) = (3, 1)$, $\mathrm{w_s}(\boldsymbol{c}) = (0, 2, 1)$ (listing the symbols in the order $(0,1), (1,0), (1,1)$). To calculate $n_i(\boldsymbol{h})$, we use Prop. 1. We find that α is:

(E1) $\alpha_1 = \lfloor (d_f^o + 1)/2 \rfloor = 2$.

(E2) $\alpha_2 = 3$. In fact $\alpha_2 \geq \lfloor (d_f^o + 1)/2 \rfloor = 2$, where equality could be reached only with $|\boldsymbol{h}| = 4$, $n_o(|\boldsymbol{h}|) = 1$, $n_i(\boldsymbol{h}) = 2$, 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 $n_i(\boldsymbol{h}) = 1$, which gives $f(\boldsymbol{h}) = 3$ and so $\alpha_2 = 3$.

(E3) $\alpha_3 = 2$. In fact $\alpha_3 \geq \lfloor (d_f^o + 1)/2 \rfloor = 2$, and equality is reached with $\boldsymbol{h} = (0, 2, 1)$, which has $|\boldsymbol{h}| = 3$, $n_o(|\boldsymbol{h}|) = 1$, $n_i(\boldsymbol{h}) = 1$ and so $f(\boldsymbol{h}) = 2$.

By an exhaustive listing of all small-weight codewords, we can also find $\mathcal{H}$, noting that $\boldsymbol{h} \in \mathcal{H}$ implies $h = 4$ for (E1), $|\boldsymbol{h}| \leq 8$ for (E2) and $|\boldsymbol{h}| \leq 6$ for (E3). We then find the exponent d^* for our ensembles: $d_1^* = 2$, $d_2^* = 3$, $d_3^* = 3$. At high SNR, the Union-Bhattacharyya bound is known to be tight with respect to actual performances. So, to show the improvement of average performances of (E2) with respect to (E1), we have fixed a SNR per information bit $E_b/N_0 = 10$ dB and we have computed, for increasing N, the average union bound for (E1) and (E2), by using (6) and Prop. 2. The resulting curves are plotted in Fig. 1 and confirm the analyzed behavior.

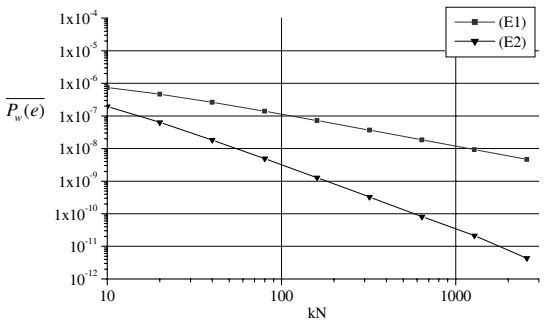

Fig. 1. Behavior of $\overline{P_w(e)}$ for ensembles (E1) and (E2) at E_b/N_0 = 10 dB

Now notice which are the properties of the component encoders that make (E2) behave better than the classical uniform interleaver (E1). The outer encoder is such that all its codewords with Hamming weight d_f^o have an odd weight in each of the two components. The inner encoder has a stronger recursiveness property than the usual one: we can say that it is recursive with respect to any component. In fact, not only $\mathrm{w_H}(\boldsymbol{u}) = 1 \Rightarrow \mathrm{w_H}(\phi^i(\boldsymbol{u})) = \infty$, but also

$$\exists j \in \{1, 2\} : \mathrm{w_H}(\boldsymbol{u}_j) = 1 \;\Rightarrow\; \mathrm{w_H}(\phi^i(u_1, u_2)) = \infty\,.$$

The inner encoder of our example is very simple, but the same property holds true also for many other encoders. Clearly, any pair of rate-1 non-catastrophic recursive encoders $1/p(D)$, $1/q(D)$ can give the same result, when they act separately on the channels as $\phi^i = \begin{bmatrix} 1/p(D) & 0 \\ 0 & 1/q(D) \end{bmatrix}$. Anyway, ϕ^i doesn't need to be a diagonal matrix, i.e. the encoder is not obliged to process the two inputs separately: the only property needed is recursiveness with respect to each input component, as defined above. For instance: $\phi^i = \begin{bmatrix} 1/(1+D) & 1 \\ 1+D & 1/(1+D) \end{bmatrix}$.

Even if they are very simple, the encoders of our example give the best possible α_2 for $d_f^o = 4$ and $m = 2$. Using the fact that $m = 2$ and Prop. 1 we have a tighter estimation of α_2 than simply $\alpha_2 \leq d_f^o$: for any $\boldsymbol{h}$ with $|\boldsymbol{h}| = d_f^o$, $n_o(\boldsymbol{h}) = 1$ and $n_i(\boldsymbol{h}) = \lfloor h_1/2 \rfloor + \lfloor h_2/2 \rfloor \geq d_f^o/2 - 1$, thus giving $\alpha_2 \leq f(\boldsymbol{h}) \leq d_f^o/2 + 1$. In our example, we have exactly $\alpha_2 = d_f^o/2 + 1 = 3$.

References

[1] S. Benedetto, G. Montorsi, "Design of Parallel Concatenated Convolutional Codes", *IEEE Trans. on Inf. Th.*, vol. 44, no. 5, pp. 591-600, 1996.

[2] S. Benedetto, D. Divsalar, G. Montorsi and F. Pollara, "Serial Concatenation of Interleaved Codes: Performance Analysis, Design, and Iterative Decoding", *IEEE Trans. on Inf. Th.*, vol. 44, no. 3, pp. 909-926, 1998.

[3] F. Fagnani, "Performance of parallel concatenated coding schemes", accepted for publication on *IEEE Trans. on Information Theory*, 2006.

[4] F. Fagnani, R. Garello, B. Scanavino and S. Zampieri, "Geometrically Uniform Parallel Concatenated Coded Modulation Schemes. Part I: Analysis. Part II: Design", 2005, available online: `http://calvino.polito.it/~fagnani/turbo/turbo.html`

[5] F. Fagnani and F. Garin, "Analysis of serial concatenation schemes for non-binary modulations", *Proceedings of ISIT 2005 (Adelaide, SA, Australia)*, pp. 745–749, Sept. 2005.

[6] H. Jin and R. J. McEliece, "Coding theorems for turbo code ensembles" *IEEE Trans. on Inf. Th.*, vol. 48, no. 6, pp. 1451–1461, 2002.