ML Performances of Serial Turbo Codes do not Concentrate!

Giacomo Como, Federica Garin, Fabio Fagnani

Dipartimento di Matematica, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy
E-mail: giacomo.como@polito.it, federica.garin@polito.it, fabio.fagnani@polito.it

Abstract

In this paper we investigate the typical behaviour of minimum distance and ML word error probability of a serial turbo concatenation with random interleaver, when the interleaver length N goes to infinity. Our main result shows that the word error probability $P(e)$ goes to zero subexponentially in N with probability one. While it is known that $\log \mathbb{E}[P(e)]/\log N$ converges to a constant, we prove that with probability one the sequence $\log(-\log(P(e)))/\log N$ approaches an interval $[\alpha, \beta] \subset (0,1)$, thus showing that the expected error rate is dominated by an asymptotically negligible fraction of bad interleavers. Our analysis is based on precise estimations of the minimum distance distribution.

1 Introduction

Serial turbo codes (serially concatenated convolutional codes with interleaver) were introduced in [3], together with an analytical explanation of the simulation results. The authors based their analysis on the so called 'uniform interleaver', a conceptual tool first introduced in [2] in order to explain the performances of Berrou et al.'s turbo codes [4]. Essentially the idea consists of fixing outer and inner encoder and estimating the ML error probability averaged over all possible interleavers. The main result in [3] consists in an upper bound to the average error probability which goes to zero as a negative power of the interleaver length N. The exponent of N, called the interleaver gain, was shown to depend only on the free distance of the outer encoder, which turns out to be the main design parameter of serial turbo codes. These ideas were rigorously formalized first in [9] and then, in a more general setting, in [7], where also a lower bound is proved differing from the upper bound only by a multiplicative constant, thus showing that the estimation is tight for the average serial turbo code. Since this average based analysis seemed to agree with simulation results in the sense that hierarchies of the design parameters were respected, it could be expected that a typical serial turbo code has an analogous behaviour, i.e. there is concentration phenomenon. From this, the quest for more precise probabilistic estimations.

In this paper we investigate this problem, showing that in fact there is no concentration of the ML error probability around its average value, since the ratio $P(e)/\mathbb{E}[P(e)]$ converges to zero with probability one. More precisely we shall prove that a typical sequence of serial turbo codes has error probability subexponentially decreasing to zero in N. The speed of this convergence turns out to depend (in a non deterministic way) on the free distance of the outer encoder, which is confirmed as the main design parameter for these coding schemes.

Our analysis is based on a precise estimation of the probability distribution of minimum distances, inspired both by the the tail estimations of [10] and the deterministic upper bounding techniques devised in [1].

Our result has to be considered as analogous of the well known behaviour of ML-decoded LDPC codes (see [8], [11]): for the (c,d)-regular LDPC ensemble 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. At the same time our results should be considered in contrast with the concentration results of [12] and [13], proved in the different context of iterative BP-decoding.

In Sec. 2, we introduce the setting and state our main result. Sec. 3 contains estimations of the probabilistic distribution of minimum distances. In Sec. 4 we prove strong probabilistic results first on the asymptotic distribution of the minimum distance sequence and then for that of the ML word error probabilities.

2 Problem setting and main result

Throughout this paper we will deal with the following coding scheme

$$\xrightarrow{kn \text{ bits}} \boxed{\phi_n^o} \xrightarrow{N_n \text{ bits}} \boxed{\pi_n} \xrightarrow{N_n \text{ bits}} \boxed{\phi_n^i} \xrightarrow{M_n \text{ bits}} \boxed{\text{Channel}}$$

where:

- the outer encoder ϕ_n^o is the termination after n trellis steps of a convolutional encoder $\phi^o \in \mathbb{Z}_2^{k\times m}(D)$ with controllability index ν^o;
- the inner encoder ϕ_n^i is the termination after $n+\nu^o$ trellis steps of a convolutional encoder $\phi^i \in \mathbb{Z}_2^{m\times r}(D)$ with controllability index ν^i;

This work was supported by the European Community Network of Excellence on Wireless COMmunications (NEWCOM).

- the interleaver π_n is a permutation of $N_n := m(n+\nu^o)$ bits;
- $M_n := r(n+\nu^o+\nu^i)$ is the blocklength;
- the channel is memoryless, binary-input output-symmetric, with Batthacharyya noise parameter γ (see [9], e.g. for the BIAWGNC $\gamma = e^{-E_s/N_0}$).

We will denote by $P(e|\pi_n)$ the word error probability of the above coding scheme, under maximum likelihood (ML) decoding, and by $\mathrm{d}_n^{\min}$ its minimum Hamming distance. All the asymptotic results about $P(e|\pi_n)$ will be stated *for a sufficiently good channel*, meaning that there exists $\overline{\gamma} > 0$ such that the result holds true provided that $\gamma < \overline{\gamma}$.

About the component encoders, we will assume that:
- ϕ^o is non-catastrophic, with free distance $d_f^o \geq 5$;
- ϕ^i is non-catastrophic and recursive.

These assumptions are essential to our results. The most used concatenation scheme, with two rate 1/2 systematic recursive encoders, is a particular case of our more general setting (systematic codes are surely non-catastrophic). Also Repeat-Accumulate and Repeat-Convolute codes fill in our setting.

For a fixed pair of component encoders ϕ^o and ϕ^i, for every n in $\mathbb{N}$ it is possible to introduce a probabilistic structure in the above serial turbo scheme, by considering as interleaver a random variable (r.v.) Π_n uniformly distributed over the set S_{N_n} of all permutations of N_n bits. We denote by $P(e|\Pi_n)$ the r.v. describing the ML word error probability of such a random coding scheme.

Consider a sequence $(\Pi_n)_{n\in\mathbb{N}}$ of *independent* random interleavers each uniformly distributed over S_{N_n}: from $(\Pi_n)_{n\in\mathbb{N}}$ we naturally obtain a sequence of random coding schemes. We call this probabilistic space the *serial turbo ensemble*; denote by $\mathbb{P}$ and $\mathbb{E}$ probability and expected value with respect to this space.

The well known results about the serial turbo ensemble consist of estimations of the average word error probabilities ([3], [9], [7]): there exist two positive constants C' and C'' such that

$$C' n^{-\lfloor (d_f^o-1)/2 \rfloor} \leq \mathbb{E}[P(e|\Pi_n)] \leq C'' n^{-\lfloor (d_f^o-1)/2 \rfloor}\,.$$

In this paper we will show that the typical asymptotic behaviour of the random sequence $(P(e|\Pi_n))_{n\in\mathbb{N}}$ is quite different from its means and is actually subexponentially decreasing to zero. Indeed, we will prove that, with probability one, for all $\varepsilon > 0$:

$$\lim_{n\to\infty} \frac{P(e|\Pi_n)}{\exp(-n^{\alpha-\varepsilon})} = 0\,;\ \lim_{n\to\infty} \frac{P(e|\Pi_n)}{\exp(-n^{\beta+\varepsilon})} = \infty\,,$$

where

$$\alpha := 1 - \frac{2}{\lceil d_f^o/2 \rceil}\,, \qquad \beta := 1 - \frac{1}{\lceil d_f^o/2 \rceil}\,. \quad (1)$$

Notice that both α and β are increasing functions of d_f^o and as $d_f^o \geq 5$, we have $0 < \alpha < \beta < 1$. So the typical behaviour does not concentrate and is much better then the average one. However, the key design parameter d_f^o is still the same, enlightened by the previous average-based analysis.

3 Estimation of minimum distance distribution

3.1 Properties of component encoders

In this paragraph we fix some notation and we recall a well-known property of convolutional encoders. We then give some estimations of the weight enumerating coefficients of our terminated convolutional encoders. In the next paragraphs we will apply these properties to the component encoders of our serial scheme. As a notation, superscripts 'o' or 'i' will refer to the outer and the inner encoder respectively.

Following [2], we will call an *error event* 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 $t_1 < t \leq t_2$ (our error events are [10]'s detours). We will call the interval $[t_1, t_2]$ the support of the error event.

Note that non-catastrophic encoders are surely injective and so there is a one-to-one correspondence between input words and codewords.

Property 1: Given a non-catastrophic convolutional encoder ϕ, there exists a positive μ such that any codeword of weight d comes from an input word of weight $w \leq \mu d$ (for systematic encoders, trivially $\mu = 1$).

Given the coding scheme described in Section 2, we define outer and inner weight enumerating coefficients. All the weights we will consider are Hamming weights (we denote by $\mathrm{w_H}(x)$ the weight of a word x). We define $A_{w,h}^{n,o}$ and $A_{w,h}^{n,i}$ to be the number of codewords of ϕ_n^o and ϕ_n^i respectively, having input weight w and output weight h. We will also write $A_h^{n,o} := \sum_w A_{w,h}^{n,o}$.

We give here some estimations of these weight enumerating coefficients. The following Lemmas are taken from [10]; we have slightly rearranged their proofs, extending their results to our more general setting and obtaining tighter bounds in Lemma 1 when w is odd.

Lemma 1 (Lemmas 1 and 2 in [10]): There exist some positive constants C_1, C_2, η, ω such that:

$$\sum_{h=1}^{d} A_{w,h}^{n,i} \leq \frac{C_1^w}{w^w} n^{\lfloor w/2 \rfloor} d^{\lceil w/2 \rceil}$$

and, if $n \geq \eta w$ and $\omega w \leq d \leq M_n$,

$$\sum_{h=1}^{d} A_{w,h}^{n,i} \geq \frac{C_2^w}{w^w} n^{\lfloor w/2 \rfloor} d^{\lceil w/2 \rceil}\,. \qquad \square$$

Lemma 2 (Lemma 3 in [10]): There exists a constant $C > 0$ such that

$$A_d^{n,o} \leq C^d \binom{n}{\lfloor d/d_f^o \rfloor}\,. \qquad \square$$

3.2 Upper bound for the left tail

We recall the upper bound for $\mathbb{P}(\mathrm{d}_n^{\min} \leq d)$ given in [10] (Thm. 2.a), here improved for odd d_f^o and generalized to our setting.

Lemma 3 (Lemma 6 in [10]):

$$\mathbb{P}(\mathrm{d}_n^{\min} \le d) \le \sum_{w=d_f^o}^{\mu^i d} \frac{1}{\binom{N_n}{w}} A_w^{n,o} \left(\sum_{h=1}^{d} A_{w,h}^{n,i} \right) \qquad \square$$

Theorem 1: There exists a constant $C > 0$ such that,

$$\mathbb{P}(\mathrm{d}_n^{\min} \le d) \le \sum_{w=d_f^o}^{\mu^i d} C^w n^{w/d_f^o - \lceil w/2 \rceil} d^{\lceil w/2 \rceil}$$

Proof sketch: The proof of this theorem follows the proof of Theorem 2.a in [10], (whose exact statement is the second part of the following Corollary 1) and is obtained applying Lemma 3 and then estimating the weight enumerating coefficients by Lemmas 1 and 2. ■

Corollary 1: If $(d_n)_{n\in\mathbb{N}}$ is a sequence such that $\frac{d_n}{n^\beta} \overset{n\to\infty}{\longrightarrow} 0$, there exists a constant $C > 0$ such that

$$\mathbb{P}(\mathrm{d}_n^{\min} \le d_n) \le C\, n \left(\frac{d_n}{n} \right)^{\lceil d_f^o/2 \rceil}$$

and hence $\mathbb{P}(\mathrm{d}_n^{\min} \le d_n) \to 0$ when $n \to \infty$. $\square$

3.3 Lower bound for the left tail

We recall some technical results given in [10] as a part of the proof of their Thm. 2.b (whose proofs can be generalized to our setting). We then use these results to establish a new lower bound for $\mathbb{P}(\mathrm{d}_n^{\min} \le d)$, given in the following Theorem 2.

First of all, we define some particular outer codewords we will use in the proof. Let c^* be a word of the outer code which has $\mathrm{w_H}(c^*) = d_f^o$ and is an error event with support $[0, a-1]$ for some constant a. We consider $n > a$. We define c_j^* as the shift to the right of c^* for j trellis steps; clearly, if $|j - l| \ge a$, then c_j^* and c_l^* have non-overlapping supports. For $j \in \{0, 1, \ldots, n-1-a\}$ and $d \in \mathbb{N}$, we define the events $E_j^*(d) := \left\{ \mathrm{w_H}(\phi_n^i(\Pi_n(c_j^*))) \le d \right\}$.

Lemma 4 ([10], part of proof of Thm. 2.b):

- if j and l are such that $|j - l| \ge a$

$$\mathbb{P}(E_j^*(d) \cap E_l^*(d)) \le \frac{\binom{N_n}{d_f^o}}{\binom{N_n - d_f^o}{d_f^o}} \mathbb{P}(E_j^*(d)) \mathbb{P}(E_l^*(d))$$

- for all j, $\mathbb{P}(E_j^*(d)) = \dfrac{\sum_{h=1}^{d} A_{d_f^o,h}^{n,i}}{\binom{N_n}{d_f^o}}$. $\square$

Note that $\frac{\binom{N_n}{d_f^o}}{\binom{N_n - d_f^o}{d_f^o}} \le \left(1 + \frac{d_f^o}{N_n - 2d_f^o + 1}\right)^{d_f^o} \overset{n\to\infty}{\longrightarrow} 1$ and hence it is surely bounded by some constant c.

Theorem 2: There exist some positive constants $C_1, C_2, \bar{n}, \omega$ such that, if $n > \bar{n}$ and $\omega d_f^o \le d \le M_n$

$$\mathbb{P}(\mathrm{d}_n^{\min} \le d) \ge C_1\, n \left(\frac{d}{n} \right)^{\lceil d_f^o/2 \rceil} - C_2 \left[n \left(\frac{d}{n} \right)^{\lceil d_f^o/2 \rceil} \right]^2$$

Proof:
Let $\bar{n} = \max\{\eta d_f^o, a\}$ with a as above, η as in Lemma 1, and consider $n > \bar{n}$. Let ω be as in Lemma 1.

$$\mathbb{P}(\mathrm{d}_n^{\min} \le d) \ge \mathbb{P}\left(\bigcup_{j=0}^{N_n} E_j^*(d) \right) \ge \mathbb{P}\left(\bigcup_{j\in J} E_j^*(d) \right), \text{ where}$$

$J := \{a\, r\,,\ r \in \mathbb{Z}^+\} \cap \{0, 1, \ldots, n-1-a\}$.

Note that $j, l \in J$, $j \ne l$ implies that $|j - l| \ge a$ and hence, by Lemma 4,

$$\mathbb{P}(E_j^*(d) \cap E_l^*(d)) \le c\, [\mathbb{P}(E_j^*(d))]^2 .$$

We use the union-intersection bound:

$$\begin{aligned} \mathbb{P}\left(\bigcup_{j\in J} E_j^*(d) \right) &\ge \sum_{j\in J} \mathbb{P}(E_i^*(d)) - \sum_{j\in J} \sum_{l\in J\setminus\{j\}} \mathbb{P}(E_j^*(d) \cap E_l^*(d)) \\ &\ge |J|\, \mathbb{P}(E_0^*(d)) - c\, [|J| \mathbb{P}(E_0^*(d))]^2 . \end{aligned}$$

By Lemmas 4 and 1, we can find two positive c_1 and c_2 (depending on d_f^o) such that, for all $n > \bar{n}$ and $\omega d_f^o \le d \le M_n$:

$$c_1 \left(\frac{d}{n} \right)^{\lceil d_f^o/2 \rceil} \le \mathbb{P}(E_j^*(d)) \le c_2 \left(\frac{d}{n} \right)^{\lceil d_f^o/2 \rceil}$$

To conclude, note that $c_3 n \le |J| \le c_4 n$ for some positive constants c_3 and c_4. ■

Corollary 2: If $(d_n)_{n\in\mathbb{N}}$ is a sequence such that $\frac{d_n}{n^\beta} \overset{n\to\infty}{\longrightarrow} 0$, there exists a constant $C > 0$ such that

$$\mathbb{P}(\mathrm{d}_n^{\min} \le d_n) \ge Cn \left(\frac{d_n}{n} \right)^{\lceil d_f^o/2 \rceil} \qquad \square$$

3.4 Deterministic upper bound

We have generalized the deterministic upper bound for the minimum distance obtained by Bazzi et al. for Repeat–Convolute codes ([1], Thm. 2) to our serial concatenation scheme. Actually Bazzi et al. also study serial turbo codes in an even more general setting ([1], Thm. 4), but we need a different estimation, where d_f^o plays the same role as the repetition parameter k in [1], Thm. 2.

Theorem 3: There exists a constant $K > 0$ such that

$$\mathrm{d}_n^{\min} \le K n^\beta \log n \qquad \square$$

The details of the proof will be given in a forthcoming paper; the outline follows the proof of Thm. 2 in [1].

4 Probabilistic conclusions

In this section we derive probabilistic results for the sequence of minimum distances based on the estimations of the previous section. Roughly speaking, we show that minimum distances almost grow as n to some positive exponent which is less then one and converges in a weak way to β, while in a strong way the sequence densely covers the whole interval $[\alpha, \beta]$, α and β being defined in (1). Finally we show how these results can be transferred to ML word error probabilities. We show

that typically $P(e|\Pi_n)$ is subexponentially decreasing to zero, again with a speed densely covering the interval $[\alpha, \beta]$ with probability one and weakly converging to β.

Remember that our probabilistic space is the serial turbo ensemble generated by a sequence of independent r.v.s $(\Pi_n)_{n\in\mathbb{N}}$, with each Π_n uniformly distributed over S_{N_n}. The main probabilistic tool we will use in our derivation is the Borel-Cantelli lemma ([5] Thm. 1.4.2) which states that, for every sequence of events $(A_n)_{n\in\mathbb{N}}$

(i) if $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$, then $\mathbb{P}(\{A_n \text{ i.o.}\}) = 0$;

(ii) if the A_n's are independent and $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$, then $\mathbb{P}(\{A_n \text{ i.o.}\}) = 1$;

where the event $\{A_n \text{ i.o.}\}$ ('A_n occurs infinitely often') is defined as

$$\{A_n \text{ i.o.}\} := \bigcap_{n\in\mathbb{N}} \Big(\bigcup_{l\geq n} A_l \Big) .$$

We define, for every $n \in \mathbb{N}$ and $x \in [0, 1]$,

$$E_n^x := \{ \mathrm{d}_n^{\min} \leq M_n^x \} ,$$
$$\theta(x) := 1 + \lceil d_f^o/2 \rceil (x - 1) .$$

Observe that $\theta(x)$ is an increasing function of x, and that $\theta(\alpha) = -1$, $\theta(\beta) = 0$. From Corollaries 1 and 2 it follows that, for $0 \leq x < \beta$, two positive constants C' and C'' exist such that

$$C' n^{\theta(x)} \leq \mathbb{P}(E_n^x) \leq C'' n^{\theta(x)} . \tag{2}$$

4.1 Minimum distances

Usually, asymptotics of the minimum distance of ensembles of codes are studied by defining the relative minimum distance $\delta_n = \mathrm{d}_n^{\min}/M_n$. In our case Theorem 3 directly implies that deterministically $\delta_n \overset{n\to\infty}{\longrightarrow} 0$ for any sequence of serial turbo codes. For this reason we propose the following non linear rescaling

$$X_n := \frac{\log(\mathrm{d}_n^{\min})}{\log(M_n)} .$$

With this rescaling, $(X_n)_n$ is a sequence of independent random variables taking values in $[0, 1]$, since $1 \leq \mathrm{d}_n^{\min} \leq M_n$. The meaning of X_n is to capture the exponent of the sublinear asymptotic behaviour of $\mathrm{d}_n^{\min}$. Notice that

$$E_n^x = \{X_n \leq x\} .$$

Our main results about $(X_n)_{n\in\mathbb{N}}$ are the two following theorems.

Theorem 4: With probability one:

(a) $(X_n)_{n\in\mathbb{N}}$ densely covers $[\alpha, \beta]$;

(b) $\liminf_n X_n = \alpha$;

(c) $\limsup_n X_n = \beta$.

Proof:

(a) We define for any $t, n \in \mathbb{N}$, and $s = 1, \ldots, 2^t$,

$$B_t^{s,n} := \Big\{ X_n \in \big[\alpha + \tfrac{s-1}{2^t}(\beta - \alpha), \alpha + \tfrac{s}{2^t}(\beta - \alpha)\big] \Big\} ,$$

$$B_t^s := \{B_t^{s,n} \text{ i.o.}\} , \quad B_t = \textstyle\bigcap_{s=1}^{2^t} B_t^s .$$

From (2), we have that

$$\begin{aligned}\mathbb{P}(B_t^{s,n}) &\geq C' n^{\theta\left(\alpha + \frac{s}{2^t}(\beta-\alpha)\right)} - C'' n^{\theta\left(\alpha + \frac{s-1}{2^t}(\beta-\alpha)\right)} \\ &= C' n^{\theta\left(\alpha + \frac{s}{2^t}(\beta-\alpha)\right)} \left(1 - \tfrac{C''}{C'} n^{-\frac{\beta-\alpha}{2^t}}\right) ,\end{aligned}$$

so that, since $\theta\left(\alpha + \frac{s}{2^t}(\beta - \alpha)\right) \geq -1$,

$$\sum_{n\in\mathbb{N}} \mathbb{P}\left(B_t^{s,n}\right) = \infty .$$

Thus, part (ii) of the Borel-Cantelli lemma lets us conclude that $\mathbb{P}(B_t^s) = 1$ for any $s = 1, \ldots, 2^t$, and so

$$\mathbb{P}(B_t) = \mathbb{P}\left(\textstyle\bigcap_{s=1}^{2^t} B_t^s\right) = 1 , \quad \forall t \in \mathbb{N} .$$

But then

$$\begin{aligned}\mathbb{P}\left(\{(X_n)_n \text{ densely covers } [\alpha, \beta]\}\right) &= \mathbb{P}\left(\textstyle\bigcap_{t\in\mathbb{N}} B_t\right) \\ &= \lim_{t\to\infty} \mathbb{P}(B_t) = 1 .\end{aligned}$$

(b) By (2) we have that, for every $\varepsilon > 0$

$$\sum_{n\in\mathbb{N}} \mathbb{P}(E_n^{\alpha-\varepsilon}) \leq \sum_{n\in\mathbb{N}} C n^{\theta(\alpha-\varepsilon)} < \infty ,$$

so that part (i) of the Borel-Cantelli lemma implies

$$\mathbb{P}\big(\{E_n^{\alpha-\varepsilon} \text{ i.o.}\} \big) = 0 .$$

Denoting by A^c the complement of an event A, we have $\{E_n^{\alpha-\varepsilon} \text{ i.o.}\}^c \subseteq \left\{ \liminf_{n\in\mathbb{N}} X_n \geq \alpha - \varepsilon \right\}$, so that

$$\begin{aligned}\mathbb{P}\left(\liminf_{n\in\mathbb{N}} X_n \geq \alpha \right) &= \mathbb{P}\left(\bigcap_{k\in\mathbb{N}} \left\{ \liminf_n X_n \geq \alpha - \tfrac{1}{k} \right\} \right) \\ &= \lim_{k\to\infty} \mathbb{P}\left(\left\{ \liminf_n X_n \geq \alpha - \tfrac{1}{k} \right\} \right) \\ &\geq \lim_{k\to\infty} \mathbb{P}\left(\left\{ E_n^{\alpha - 1/k} \text{ i.o.} \right\}^c \right) \\ &= 1 .\end{aligned}$$

Since by point (a) we have $\mathbb{P}\left(\liminf_n X_n \leq \alpha\right) = 1$, point (b) follows.

(c) Theorem 3 directly implies that $\limsup_n X_n \leq \beta$. Since point (a) implies that $\mathbb{P}\left(\limsup_n X_n \geq \beta\right) = 1$, point (c) follows. ■

Although Theorem 4 tells us that with probability one a random sequence of codes from the serial turbo ensemble has minimum distance exhibiting a chaotic behaviour, a weak form of convergence for the sequence of r.v.s $(X_n)_n$ can still be observed. Formally, we have to consider the sequence of probability measures instead of the probability space of sequences. We will denote by $X_n \overset{\mathbb{P}}{\to} X$ the convergence in probability (see [5] for definitions and properties). The following result is a restating of [10]'s Theorem 2 in our setting (and with an improvement when d_f^o is odd).

Theorem 5: $X_n \overset{\mathbb{P}}{\to} \beta$.

Proof: For every $\varepsilon > 0$, Corollary 1 and Theorem 3 guarantee that

$$\mathbb{P}(|X_n - \beta| < \varepsilon) \geq 1 - C\, n^{-\lceil d_f^o/2 \rceil \varepsilon} \overset{n\to\infty}{\longrightarrow} 1 . \quad ■$$

4.2 ML Error probabilities

In order to transfer our results about minimum distances to ML word error probabilities we use a classical tool of coding theory known as expurgation (see [8]). We estimate the averaged error probability conditioned on the complement events $(E_n^x)^c$ for some proper $x \in [0,\beta)$. By combining these estimations with (2) we derive strong probabilistic results about the asymptotic behaviour of $P(e|\Pi_n)$.

We define the following r.v.

$$Y_n := \frac{\log(-\log P(e|\Pi_n))}{\log n}\,;$$

the idea is that Y_n should capture the speed of the subexponential asymptotic decrease of $P(e|\Pi_n)$.

Proposition 1: If the channel is sufficiently good, for all $x \in [0,\beta)$,

$$\mathbb{E}[P(e|\Pi_n)\,|(E_n^x)^c] \le \exp(-K_x n^x)$$

for some positive constant K_x.

Proof: We use the Union-Bhattacharyya bound, remembering that $(E_n^x)^c = \{\mathrm{d}_n^{\min} > M_n^x\}$ and then, denoting by $\mathbb{1}_E$ the characteristic function of some event E:

$$\mathbb{E}[P(e|\Pi_n)|(E_n^x)^c] = \frac{1}{\mathbb{P}((E_n^x)^c)}\mathbb{E}[P(e|\Pi_n)\cdot \mathbb{1}_{(E_n^x)^c}]$$
$$\le \frac{1}{\mathbb{P}((E_n^x)^c)} \sum_{h=M_n^x}^{M_n} \sum_{w=d_f^o}^{\mu^i h} \sum_{l=1}^{\mu^o w} \frac{A_{l,w}^{n,o} A_{w,h}^{n,i}}{\binom{N_n}{w}} \gamma^h\,.$$

By Coroll. 1, $\mathbb{P}((E_n^x)^c) \overset{n\to\infty}{\longrightarrow} 1$. So, for some $c \ge 1$,

$$\frac{1}{\mathbb{P}((E_n^x)^c)} \le c\,.$$

We estimate $A_{w,h}^{n,i} \le \sum_{j=1}^{h} A_{w,j}^{n,i}$ by Lemma 1 and $\sum_{l=1}^{\mu^o w} A_{l,w}^{n,o}$ by Lemma 2, so we can find a positive C such that:

$$\mathbb{E}[P(e|\Pi_n)|(E_n^x)^c] \le c \sum_{h=M_n^x}^{M_n} \sum_{w=d_f^o}^{\mu^i h} C^w \left(\frac{h}{w}\right)^{\frac{w}{2}} \left(\frac{w}{n}\right)^{\frac{w}{2}-\frac{w}{d_f^o}} \gamma^h.$$

Then we remark that the function $g(s) := (a/s)^s$ has maximum value $g(a/e) = e^{a/e}$ and hence

$$(h/w)^{w/2} \le e^{h/(2e)}\,.$$

Moreover, $w \le N_n = m(n+\nu^o) \le \tilde{c}n$ for some $\tilde{c} \ge 1$, so $(w/n)^{\frac{w}{2}-\frac{w}{d_f^o}} \le \tilde{c}^{(\frac{1}{2}-\frac{1}{d_f^o})w}$. Hence, as $w \le \mu^i h$, we can find a constant $\bar{C} \ge 1$ such that:

$$\mathbb{E}[P(e|\Pi_n)|(E_n^x)^c] \le \sum_{h=M_n^x}^{M_n} (\bar{C}\gamma)^h \le \bar{c}(\bar{C}\gamma)^{M_n^x}$$

where the last inequality holds true, for some $\bar{c} > 0$, if $\gamma < 1/\bar{C}$. Notice that $\bar{C}\gamma < 1$ also implies that $\bar{c}(\bar{C}\gamma)^{M_n^x} \le \exp(-K_x n^x)$ for some positive K_x. ■

Lemma 5: There exists a constant K such that, deterministically, $P(e|\Pi_n) \ge \exp(-Kn^\beta \log n)$.

Proof: We use the inequality $P(e|\Pi_n) \ge p^{\mathrm{d}_n^{\min}}$, where p is the equivocation probability of the channel (see [6]; e.g. $p = 1/2\,\mathrm{erfc}(\sqrt{E_s/N_0})$ for the BIAWGNC). This, together with Theorem 3, gives the result. ■

Lemma 6: For any $x \in [0,\beta)$, there exist two positive constants K and C, depending on x but not on n, such that

$$\mathbb{P}\Big(P(e|\Pi_n) \ge \exp(-Kn^x)\Big) \ge Cn^{\theta(x)}\,.$$

Proof: Since $P(e|\Pi_n) \ge p^{\mathrm{d}_n^{\min}}$, by (2) we get

$$\begin{aligned}\mathbb{P}\left(P(e|\Pi_n) \ge p^{M_n^x}\right) &\ge \mathbb{P}\left(\mathrm{d}_n^{\min} \le M_n^x\right)\\ &= \mathbb{P}\left(E_n^x\right)\\ &\ge Cn^{\theta(x)}\,.\end{aligned}$$ ■

Lemma 7: For a sufficiently good channel, for any $x \in [0,\beta)$, there exist two positive constants K and K', depending on x but not on n, such that

$$\mathbb{P}\Big(P(e|\Pi_n) \ge \exp(-Kn^x)\Big) \le K' n^{\theta(x)}\,.$$

Proof: By Proposition 1 we have, for some $K_x > 0$

$$\mathbb{E}\left[P(e|\Pi_n)\,|(E_n^x)^c\right] \le \exp(-K_x n^x)\,,$$

so that, by Markov inequality, we get

$$\begin{aligned}&\mathbb{P}\left(P(e|\Pi_n) \ge \exp(-\tfrac{K_x}{2}n^x)\big|\,(E_n^x)^c\right)\\ &\le \mathbb{P}\left(P(e|\Pi_n) \ge \frac{\mathbb{E}\left[P(e|\Pi_n)|\,(E_n^x)^c\right]}{\exp(-\frac{K_x}{2}n^x)}\,\middle|\,(E_n^x)^c\right)\\ &\le \exp(-\tfrac{K_x}{2}n^x)\,.\end{aligned}$$

Thus, by (2) we get

$$\begin{aligned}&\mathbb{P}\left(P(e|\Pi_n) \ge \exp(-\tfrac{K_x}{2}n^x)\right)\\ &= \mathbb{P}\left(P(e|\Pi_n) \ge \exp(-\tfrac{K_x}{2}n^x)\big|\,E_n^x\right)\mathbb{P}(E_n^x) +\\ &\quad + \mathbb{P}\left(P(e|\Pi_n) \ge \exp(-\tfrac{K_x}{2}n^x)\big|\,(E_n^x)^c\right)\mathbb{P}((E_n^x)^c)\\ &\le \mathbb{P}(E_n^x) + \mathbb{P}\left(P(e|\Pi_n) \ge \exp(-\tfrac{K_x}{2}n^x)\big|\,(E_n^x)^c\right)\\ &\le Cn^{\theta(x)} + \exp(-\tfrac{K_x}{2}n^x)\end{aligned}$$

and the claim immediately follows with $K = K_x/2$, and for some $K' \ge C$. ■

Theorem 6: For a sufficiently good channel, with probability one it holds true:

(a) $(Y_n)_{n\in\mathbb{N}}$ densely covers $[\alpha,\beta]$;
(b) $\liminf_n Y_n = \alpha$;
(c) $\limsup_n Y_n = \beta$.

Proof:

(a) The proof is rather technical and will be given in a forthcoming paper. The main ideas are similar to those of the proof of Thm. 4 (a).

(b) For every $\varepsilon > 0$, by Lemma 7 we get

$$\sum_{n\in\mathbb{N}} \mathbb{P}\Big(P(e|\Pi_n) \ge \exp(-Kn^{\alpha-\varepsilon})\Big) \le \sum_{n\in\mathbb{N}} K' n^{\theta(\alpha-\varepsilon)} < \infty$$

Then point (i) of the Borel-Cantelli lemma implies

$$\mathbb{P}\left(\left\{P(e|\Pi_n) \ge \exp(-Kn^{\alpha-\varepsilon})\right\}\ \text{i.o.}\right) = 0$$

so that

$$\mathbb{P}\left(\liminf_n Y_n \geq \alpha - \varepsilon\right) \geq \mathbb{P}\left(\left\{\left\{P(e|\Pi_n) \geq \exp(-Kn^{\alpha-\varepsilon})\right\} \text{ i.o.}\right\}^c\right) = 1\,,$$

and

$$\begin{aligned}\mathbb{P}\left(\liminf_n Y_n \geq \alpha\right) &= \mathbb{P}\left(\bigcap_{k\in\mathbb{N}} \left\{\liminf_n Y_n \geq \alpha - 1/k\right\}\right) \\ &= \lim_{k\to\infty} \mathbb{P}\left(\liminf_n Y_n \geq \alpha - 1/k\right) = 1\,. \qquad (3)\end{aligned}$$

Moreover, by Lemma 6

$$\sum_{n\in\mathbb{N}} \mathbb{P}\left(P(e|\Pi_n) \geq \exp(-Kn^{\alpha})\right) \geq \sum_{n\in\mathbb{N}} Cn^{\theta(\alpha)} = \infty$$

and thus, by point (ii) of the Borel-Cantelli lemma:

$$\mathbb{P}\left(\liminf_n Y_n \leq \alpha\right) \geq \mathbb{P}\Big(\left\{P(e|\Pi_n) \geq \exp(-Kn^{\alpha})\right\} \text{i.o.}\Big) = 1$$

(c) Lemma 5 implies that, deterministically

$$\limsup_n Y_n \leq \beta\,.$$

Moreover, for every $\varepsilon > 0$, by Lemma 7 we have

$$\mathbb{P}\Big(P(e|\Pi_n) \geq \exp(-Kn^{\beta-\varepsilon})\Big) \leq Cn^{\theta(\beta-\varepsilon)} \overset{n\to\infty}{\longrightarrow} 0\,.$$

Thus a subsequence $(\Pi_{n_k})_{k\in\mathbb{N}}$ exists[1] such that

$$\sum_{k\in\mathbb{N}} \mathbb{P}\Big(P(e|\Pi_{n_k}) \geq \exp(-Kn_k^{\beta-\varepsilon})\Big) < \infty\,,$$

so that part (i) of the Borel-Cantelli lemma implies

$$\begin{aligned}&\mathbb{P}\left(\limsup_n Y_n \geq \beta - \varepsilon\right) \\ &\geq \mathbb{P}\Big(\left\{P(e|\Pi_n) \geq \exp(-Kn^{\beta-\varepsilon})\right\} \text{i.o.}\Big) \\ &\geq \mathbb{P}\Big(\left\{P(e|\Pi_{n_k}) \geq \exp(-Kn_k{}^{\beta-\varepsilon})\right\} \text{i.o.}\Big) = 1\,.\end{aligned}$$

By essentially the same derivation as in (3), we obtain

$$\mathbb{P}\left(\limsup_n Y_n \geq \beta\right) = 1\,. \qquad \blacksquare$$

Theorem 7: For a sufficiently good channel

$$Y_n \overset{\mathbb{P}}{\to} \beta\,.$$

Proof: This follows from Lemmas 5 and 7. $\blacksquare$

4.3 Other ensembles

From the same fixed component encoders ϕ^o and ϕ^i, it is possible to construct different ensembles, introducing other probabilistic structures for the interleaver sequence. For instance, instead of a sequence of independent interleavers $(\Pi_n)_{n\in\mathbb{N}}$ with Π_n uniformly distributed over S_{N_n} as in our serial turbo ensemble, we can consider a sequence of interleavers $(\Pi'_n)_{n\in\mathbb{N}}$ such that each Π'_n is still uniformly distributed over S_{N_n}, but possibly dependent on $\{\Pi'_i, i = 1, \ldots, n-1\}$.

A close look at our proofs shows that independence among the Π_n's is required only when using point (ii) of the Borel-Cantelli lemma. Hence, for the new ensemble based on $(\Pi'_n)_{n\in\mathbb{N}}$ we can state that, with probability one:

- $\liminf_n X'_n \geq \alpha$; $\limsup_n X'_n = \beta$,
- $\liminf_n Y'_n \geq \alpha$; $\limsup_n Y'_n = \beta$,

while $X'_n \overset{\mathbb{P}}{\to} \beta$ and $Y'_n \overset{\mathbb{P}}{\to} \beta$.

This means that introducing some dependence among the uniform interleavers cannot make performances worse while it could possibly improve them. It would be interesting to develop an analysis for these hierarchical structures.

5 Conclusions

We have analyzed the asymptotic behaviour of minimum distances and ML error probabilities of the serial turbo ensemble. We have proved that a typical sequence of codes from this ensemble has minimum distance sublinearly growing in the interleaver length and ML error probability subexponentially decreasing to zero. Both these asymptotic behaviours are characterized by a random parameter densely covering the interval $[\alpha, \beta]$, where α and β are increasing functions of the free distance of the outer encoder. This shows that there is no concentration of error probability around its average, which decreases only as a negative power of the interleaver length ([3],[9],[7]).

References

[1] L. Bazzi, M. Mahdian, D. Spielman. *The minimum distance of turbo-like codes* (2003). [Online.] Available: http://www.mathnet.or.kr/papers/MIT/Spielman/s_12.pdf

[2] S. Benedetto, G. Montorsi. *Design of Parallel Concatenated Convolutional Codes*. IEEE Trans. on Inf. Th. Vol. 44, No. 5, pp. 591–600, May 1996.

[3] S. Benedetto, D. Divsalar, G. Montorsi, 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, May 1998.

[4] C. Berrou, A. Glavieux, P. Thitimajshima. *Near Shannon Limit Error-Correctiong Coding and Decoding: Turbo Codes*. Proc. of ICC'93 (Genève, Switzerland), pp. 1064–1070, May 1993.

[5] V. Borkar. *Probability Theory*. NewYork: Springer-Verlag, 1995.

[6] F. Fagnani. *Performance of Parallel Concatenated Coding Schemes* (2004). [Online.] Available: http://calvino.polito.it/∼fagnani/turbo/turbo.html

[7] F. Fagnani, F. Garin. *Analysis of Serial Concatenation Schemes for Non-binary Modulations*. Proc. of ISIT 2005 (Adelaide, SA, Australia), pp. 745–749, Sept. 2005.

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

[9] H. Jin, R.J. McEliece. *Coding Theorems for Turbo Code Ensembles*. IEEE Trans. on Inf. Th. Vol. 48, No. 6, pp. 1451–1461, June 2002.

[10] N. Kahale, R. Urbanke. *On the Minimum Distance of Parallel and Serially Concatenated Codes* (1997). [Online.] Available: http://lthcwww.epfl.ch/papers/KaU.ps

[11] D.J.C. MacKay. *Good Error Correcting Codes Based On Very Sparse Matrices*. IEEE Trans. on Inf. Th. Vol. 45, No. 2, pp. 399–431, Mar. 1999.

[12] T. Richardson, R. Urbanke. *Concentrate!* Proc. of 37th Allerton Conf. (Monticello, Illinois, USA), Oct. 1999.

[13] T. Richardson, R. Urbanke. *The Capacity of Low-Density Parity-Check Codes under Message-Passing Decoding*. IEEE Trans. on Inf. Th. Vol. 47, No. 2, pp. 599–618, Feb. 2001.

[1] Given any real sequence $(a_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} a_n = 0$, you can find an increasing sequence of naturals $n_1 < n_2 < \ldots$ such that $\sum_{k\in\mathbb{N}} |a_{n_k}| < +\infty$.