The performance of serial turbo codes does not concentrate*

Federica Garin†, Giacomo Como‡, Fabio Fagnani§.

October 27, 2009

Abstract

Minimum distances and maximum likelihood error probabilities of serial turbo concatenations with random interleaver are analyzed. It is shown that, with high probability, the minimum distance of serial turbo codes grows as a positive power of the block-length, while their error probability decreases sub-exponentially fast in the block-length, on sufficiently good memoryless channels. Such a typical code behavior contrasts the performance of the average serial turbo code, whose error probability is dominated by an asymptotically negligible fraction of bad interleavers, and decays only as a negative power of the block-length. The analysis proposed in this paper relies on precise estimations of the minimum distance distribution, whose scaling law is shown to depend both on the free distance of the outer constituent encoder, and on the effective free distance of the inner encoder. Hence, despite the lack of concentration of the maximum likelihood error probability around its expected value, the main design parameters suggested by the average-code analysis turn out to characterize also the performance of the typical serial turbo code.

Keywords: Turbo codes, serially concatenated codes, minimum distance, error probability, typical code analysis.

1 Introduction

Serially concatenated convolutional codes with random interleaver, briefly serial turbo codes, were introduced in [4], 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 [5] in order to explain the performance of Berrou et al.'s parallel turbo codes [6]. In a nutshell, the idea consists in fixing the outer and the inner constituent encoders, and in estimating the maximum likelihood (ML) error probability averaged over all possible interleavers. The main result in [4] is an upper bound to the average error probability which goes to zero as a negative power of the interleaver length. The exponent of such power law decay, 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. The effect of the inner constituent encoder was analyzed by considering the limit performance in the high signal-to-noise ratio. The fundamental design parameter characterizing the performance in this regime is the *effective free distance* of the inner encoder, i.e. the smallest weight of codewords corresponding

*A preliminary version of this work has appeared as [8]

†DEI, Università di Padova, garin@dei.unipd.it, corresponding author

‡LIDS, MIT, giacomo@mit.edu

§Dip. di Matematica, Politecnico di Torino, fabio.fagnani@polito.it

to inputs of weight two. These ideas have been rigorously formalized first in [13] and then, in a more general setting, in [12], 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*. Numerical simulations of serial turbo codes confirm the hierarchies of the two aforementioned design parameters, suggesting that a typical serial turbo code should perform closely to the average code, i.e. that the performance of serial turbo ensembles concentrates around its average.

In the present paper, we shall disprove such a conjecture. Indeed, we shall show that, for almost all choices of the interleaver, serial turbo codes have ML error probability subexponentially decreasing to zero in the block-length. This proves that, due to the presence of an asymptotically vanishing fraction of bad codes, the average-code analysis provides too conservative a prediction of the behaviour of the *typical serial turbo code*. In fact an analogous phenomenon is known to occur for LDPC code ensembles [10], as well as for random (linear) code ensembles at low rates [2]. However, despite the lack of concentration of the serial turbo code ensemble's performance, we shall show that the free distance of the outer encoder and the effective free distance of the inner encoder turn out to be the fundamental parameters characterizing the scaling law of the performance of the typical serial turbo code.

The analysis presented in this paper relies on precise estimations of the probability distribution of the minimum distance, inspired both by the the tail estimations of [14] and the deterministic upper bounding techniques devised in [3]. Our main results show that, with high probability, the minimum distance of serial turbo codes scales as $d_e^i N^{1-2/d_f^o}$, where d_e^i is the effective free distance of the inner constituent encoder, N is the blocklength, and d_f^o is the free distance of the outer constituent encoder. While the dependence of the typical minimum distance on d_f^o is the same highlighted in [14, 3], the dependence on d_e^i is a novel contribution. Such a scaling law for the minimum distance of the typical turbo code will be proven through both a detailed study of the left tail of the minimum distance's probability distribution (Theorem 1), and of a deterministic upper bound (Theorem 2). The former may be thought of as a generalization of the results of [14], with the main improvement consisting in the characterization of d_e^i as a linear scaling parameter for the minimum distance. The latter generalizes some of the results of [3], and improves asymptotically on the best known deterministic bound for minimum distance of serial turbo codes, presented in [16]. Finally, by means of code-expurgation techniques, these results will allow us to show that the ML error probability of the typical turbo code decreases sub-exponentially fast in the block-length (Theorem 3).

The remainder of the paper is organized as follows. In Section 2 we introduce in a formal way the serially concatenated codes. Section 3 gathers some fundamental estimations on the weight enumerators of convolutional codes which will be used throughout the paper. Section 4 contains all the main results on minimum distances of serial codes. Finally, in Section 5 we prove our main results on the typical behavior of minimum distance and ML error probability and a number of related results. The most technical proofs are deferred to Appendix A, while Appendix B contains some extensions.

2 Problem setting

In this section we establish some notation on convolutional encoders, and introduce the serial turbo code ensemble.

2.1 Convolutional codes

We start by briefly recalling some fundamental facts on convolutional codes which will be used throughout this paper. We shall denote by $\mathbb{Z}$ the set of integers and by $\mathbb{Z}_2 = \{0, 1\}$ the

binary field. Given a $\mathbb{Z}_2$-vector space V, the space of Laurent series with coefficients in V will be denoted by $V((D))$. We shall consider the following subspaces of $V((D))$: the subspace of formal power series $V[[D]]$, the subspace of polynomials $V[D]$, the subspace of Laurent polynomials $V[D, D^{-1}]$, the subspace of rational functions $V(D)$. If $v \in V((D))$, $v(t)$ denotes the coefficient in v of D^t, so that we can write $v = \sum_t v(t)D^t$. We shall often identify v with the sequence $(v(t))_{t\in\mathbb{Z}}$. Given $v \in V((D))$, we define the *support* of v as $\mathrm{supp}(v) := \{t \in \mathbb{Z} \mid v(t) \neq 0\}$. If $v = \sum_t v(t)D^t$, we define $v^- = \sum_{t<0} v(t)D^t$. If $v \in \mathbb{Z}_2^r((D))$, we define $\mathrm{w_H}(v) := \sum_t \mathrm{w_H}(v(t))$, with $\mathrm{w_H}(x)$ denoting Hamming weight of a string $x \in \mathbb{Z}_2^r$.

A convolutional encoder is any $\phi \in \mathbb{Z}_2^{r\times s}(D) \cap \mathbb{Z}_2^{r\times s}[[D]]$. It naturally induces a map (denoted with the same symbol) $\phi : \mathbb{Z}_2^s((D)) \to \mathbb{Z}_2^r((D))$ by simple matrix right multiplication (vectors are intended as column vectors throughout the paper). If $\phi = \sum_{t=0}^{M} \phi(t)D^t \in \mathbb{Z}_2^{r\times s}[D]$ for some finite M, it is called polynomial. A convolutional encoder ϕ is said to be *recursive* if each column j contains at least one entry ϕ_{ij} which is not polynomial; equivalently, ϕ is recursive if, for all input u with Hamming weight $\mathrm{w_H}(u) = 1$, the $\mathrm{w_H}(\phi u) = +\infty$. The encoder is said to be *non-catastrophic* if $\phi u \in \mathbb{Z}_2^s[D^{-1}, D]$ implies $u \in \mathbb{Z}_2^r[D^{-1}, D]$. The *free distance* and the *effective free distance* of ϕ are defined as

$$d_f := \min\{\mathrm{w_H}(\phi u) \mid u \neq 0\}, \qquad d_e := \min\{\mathrm{w_H}(\phi u) \mid \mathrm{w_H}(u) = 2\},$$

respectively.

Rationality of convolutional encoders is equivalent to the existence of a linear state-space realization with a finite number of states: given $\phi \in \mathbb{Z}_2^{r\times s}((D)) \cap \mathbb{Z}_2^{r\times s}[[D]]$, there exist a state space $X = \mathbb{Z}_2^\mu$ and matrices $F \in Z_2^{\mu\times\mu}$, $G \in Z_2^{\mu\times r}$, $H \in \mathbb{Z}_2^{s\times\mu}$, $W \in \mathbb{Z}_2^{s\times r}$, such that $y(D) = \phi(D)u(D)$ if and only if there exists a state sequence $x(D) \in \mathbb{Z}_2^\mu(D)$ such that, for all t, $x_{t+1} = Fx_t + Gu_t$ and $y_t = Hx_t + Wu_t$. For a given realization (F, G, H, W), if $u \in \mathbb{Z}_2^r[[D]]$, upon fixing $x(0) = 0$, $x(D)$ is uniquely determined; we will say that such $x(D)$ is the state sequence associated with $u(D)$. The state realization can be pictorially represented as a trellis: for each $t \in \mathbb{N}$, draw 2^μ states, corresponding to the state space X; then draw an edge from state x at time t to state x' at time $t + 1$, with input label a and output label b if and only if $x' = Fx + Ga$ and $b = Hx + Wa$. The minimal realization (i.e., having the smallest μ) is unique up to a change of basis, and has observability and controllability properties which are essential for defining the terminated encoders (see below) and for proving Lemma 1. In this paper we will always assume that we are using the minimal trellis.

The *block-termination* of a convolutional encoder ϕ after N trellis steps is defined as follows. Fix $N \in \mathbb{N}$, consider an input u supported inside $[0, N - 1]$, and let x be the associated state sequence. The output ϕu, however, may well be not supported in the same interval. Indeed, it can happen that $x(N) \neq 0$. However, there exists an integer $\nu \geq 0$ (called *constraint length* and not depending on the particular u or on N), and an input $\tilde{u}$ coinciding with u on $[0, N-1]$ and supported inside $[0, N + \nu - 1]$ such that $x_{N+\nu} = 0$. The output is then also supported in $[0, N + \nu - 1]$. In the following, we shall assume that, for every input u, the terminating extension $\tilde{u}$ has been chosen in such a way to be a linear function of u (it is easy to see that this can always be done). The block termination of ϕ after N trellis steps is a $\mathbb{Z}_2$-linear map $\phi_N : \mathbb{Z}_2^{rN} \to \mathbb{Z}_2^{s(N+\nu)}$ defined by

$$\phi_N(u(0), u(1), \ldots, u(N-1)) = (y(0), y(1), \ldots, y(N-1))$$

if

$$\begin{aligned}&\phi\big(u(0) + u(1)D + \cdots + u(N-1)D^{N-1} + \cdots + \tilde{u}(N+\nu-1)D^{N+\nu-1}\big)\\ &= y(0) + y(1)D + \cdots + y(N+\nu-1)D^{N+\nu-1}.\end{aligned}$$

Notice that, whenever convenient, the space $\mathbb{Z}_2^{rN}$ is identified with the subspace of $\mathbb{Z}_2^r[D]$ consisting of the polynomials of degree up to $N - 1$.

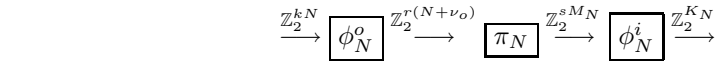

Figure 1: The serially concatenated encoding scheme

2.2 Serially concatenated convolutional codes with random interleaver

We start from two convolutional encoders $\phi^o \in \mathbb{Z}_2^{r\times k}(D) \cap \mathbb{Z}_2^{r\times k}[[D]]$ and $\phi^i \in \mathbb{Z}_2^{l\times s}(D) \cap \mathbb{Z}_2^{l\times s}[[D]]$. Let ν_o and ν_i be their corresponding constraint lengths and let N be such that s divides $r(N+\nu_o)$. Let M_N be such that $sM_N = r(N+\nu_o)$, and let $K_N := l(M_N+\nu_i) = l(\frac{r}{s}(N+\nu_o)+\nu_i)$. Consider the block terminations of ϕ^o and ϕ^i after N and M_N trellis steps, respectively:

$$\phi_N^o : \mathbb{Z}_2^{kN} \to \mathbb{Z}_2^{r(N+\nu_o)}, \qquad \phi_N^i : \mathbb{Z}_2^{sM_N} \to \mathbb{Z}_2^{K_N}.$$

Finally let π_N be a permutation of length sM_N and denote by the same symbol $\pi_N : \mathbb{Z}_2^{sM_N} \to \mathbb{Z}_2^{sM_N}$ the corresponding linear isomorphism. The serially concatenated encoder considered in this paper is the composition

$$\phi_N^i \circ \pi_N \circ \phi_N^o : \mathbb{Z}_2^{kN} \to \mathbb{Z}_2^{K_N}$$

depicted in Fig.1. We shall refer to ϕ^o as the *outer encoder*, to ϕ^i as the *inner encoder*, and to π_N as *the interleaver*.

In order to avoid extremely cumbersome notation, we shall expose our results in full detail under some simplifying assumptions. From now on, unless differently specified, we shall assume that:

- ϕ^o is non-catastrophic, and its free distance d_f^o is even and satisfies $d_f^o > 2$;
- ϕ^i is non-catastrophic and recursive, has scalar input ($s = 1$) and is proper rational, i.e. $\phi^i = \frac{1}{q(D)}[p_1(D), \ldots, p_l(D)]^T$ with $\deg(p_i) < \deg(q_i)$ for all i.

The only essential assumptions, however, are non-catastrophicity and recursiveness of ϕ^i. We shall discuss in Appendix B how to treat the case of odd d_f^o, while addressing the interested reader to [11] for further generalizations.

In the rest of this paper, we shall investigate the performance of the above-described serially concatenated coding schemes, assuming that the interleaver Π_N is a random variable uniformly distributed on the group of permutations of M_N symbols. This is the classical 'uniform interleaver' ensemble of [5, 4]. Since the interleaver Π_N is random, the minimum distance $\mathrm{d}_N^{\min} := \min\{\mathrm{w_H}(\phi_N^i \circ \pi_N \circ \phi_N^o(u)) : u \neq 0)\}$ is a random variable. Similarly, assuming transmission over binary-input output-symmetric memoryless channel with ML decoding, the word error probability of the serial turbo code is a random variable, to be denoted by $P(e|\Pi_N)$. While the focus of most of the literature [4, 12] has been of the error probability of the *average serial turbo code*, $\mathbb{E}[P(e|\Pi_N)]$, in this paper we shall be concerned with the minimum distance and error probability of the *typical serial turbo code*, namely with the high-probability behaviour of $\mathrm{d}_N^{\min}$ and $P(e|\Pi_N)$, as N goes to infinity.

We end this section by establishing the following notational convention, to be used throughout the paper. When dealing with quantities depending on many parameters, such as $w, d, N, n \ldots$, we shall implicitly assume that all the parameters are depending on N, but we shall avoid cumbersome notation $w_N, d_N \ldots$. Hence, a statement such as '$f(w,d,N) = o(N^a)$ for $N \to \infty$, $d = o(N)$ and $w \leq d$' means that if $d = d_N$, $w = w_N$ satisfying $w_N \leq d_N$ and $d_N/N \to 0$ when $N \to \infty$, then $\lim_{N\to\infty} f(w_N, d_N, N)/N = 0$. When we say '$w$ is constant' we mean it does not depend on N.

3 Weight enumerators of the constituent encoders

In this section, we recall some well-known definitions and properties of convolutional encoders, and we state the bounds on the weight enumerators of outer and inner encoder, which will be used in the following sections. The proofs of such bounds, many of which relying on arguments developed in [14], are given in Appendix A.1.

Consider a convolutional encoder $\phi \in \mathbb{Z}_2^{r\times s}(D) \cap \mathbb{Z}_2^{r\times s}[[D]]$. A sequence $u \in \mathbb{Z}_2^s((D))$ (and also its image ϕu) is called an *error event* if there exist $t_1 < t_2$ such that $\mathrm{supp}(u) \subseteq [t_1, t_2]$ and the corresponding state sequence x is such that $\mathrm{supp}(x) = [t_1 + 1, t_2]$. Notice that this implies that necessarily $u(t_1) \neq 0$ and $\mathrm{supp}(\phi u) \subseteq [t_1, t_2]$. The *length* of such error event is $t_2 - t_1 + 1$, while the discrete interval $[t_1, t_2]$ is called its *active window*. Every finitely supported input sequence u such that ϕu has also finite support, can be obtained as the summation of a finite number of error events with non overlapping active windows. The following useful result was proved in [9, Lemma 20].

Lemma 1 *Given a non-catastrophic convolutional encoder, there exists a constant η such that any of its error events with output Hamming weight w has length not greater than ηw.*

Let ν be the constraint length of ϕ and consider the block termination of length N, $\phi_N : \mathbb{Z}_2^{rN} \to \mathbb{Z}_2^{s(N+\nu)}$. An error event for ϕ_N is any input $(u(0), \dots u(N-1))$ such that

$$u(0) + u(1)D + \cdots + u(N-1)D^{N-1} + \cdots + \tilde{u}(N+\nu-1)D^{N+\nu-1}$$

(where $\tilde{u}$ is the usual linear terminating extension of u) is an error event for ϕ. Such an error event is said to be *regular* if its active window $[t_1, t_2]$ lies inside $[0, N-1]$ (the termination $\tilde{u}$ is 0). Otherwise, the error event is called *terminating*. It is clear that any input for ϕ_N can be written as the sum of a finite number of regular error events plus, possibly, a terminating one, all having disjoint active windows.

Consider $\phi^o \in \mathbb{Z}_2^{r\times k}(D)$ and $\phi^i \in \mathbb{Z}_2^{l\times 1}(D)$. We shall denote by η_o and η_i the constants defined in Lemma 1 for ϕ^o and ϕ^i respectively.

For the outer encoder, we define the enumerating coefficient $A_d^{o,N}$ to be the number of inputs of ϕ_o^N with output weight d. For it, we need only the following simple upper bound, which holds true for all non-catastrophic terminated convolutional encoders. It is proved in Sect. A.1.1.

Lemma 2 *The following inequalities hold true:*

(a) If $\lfloor d/d_f^o \rfloor < N/2$, then $A_d^{o,N} \leq 2^{(k\eta_o+\eta_o+1)d+1}\binom{N}{\lfloor d/d_f^o \rfloor}$

(b) $A_{d_f^o}^{o,N} \leq m_f^o N$, where m_f^o is the number of different error events for ϕ^o starting at 0 and producing output weight d_f^o.

As for the inner encoder, we shall need a weight enumerator which considers both input and output weight. Define $A_{w,\leq d}^{i,N}$ to be the number of inputs of ϕ_N^i with input weight w and output weight not greater than d. Another weight enumerator which will play a key role is $R_{w,\leq d,n}^{i,N}$, defined as the number of inputs of ϕ_N^i with input weight w and output weight not greater than d, consisting of exactly n regular error events.

By recursiveness of ϕ^i, $\mathrm{w_H}(\phi^i u)$ is infinite for all weight-one inputs u. On the contrary, it is well known that there exists a positive integer δ such such that $\mathrm{w_H}(\phi^i(1 + D^\delta))$ is finite (and hence d_e^i is finite). Let δ_i the smallest such value. Then, $\mathrm{w_H}(\phi^i(1 + D^\delta))$ is finite if and only if $\delta = a\delta_i$ for some positive integer a. Also, having assumed that ϕ^i is proper rational, one has that $\mathrm{w_H}(1 + D^{a\delta_i}) = a\mathrm{w_H}(1 + D^{\delta_i}) = a\, d_e^i$. These considerations show, in particular,

that any error event for ϕ^i has input weight 2 or larger. When considering ϕ^i_N, however, one has to be slightly more careful: regular error events have indeed weight at least 2, while this is not necessarily true for a terminating event u which could have weight 1, the remaining weight being in the extended part $\tilde{u}$ and not counted in the weight of u. The bounds we shall give rely on such input-weight limitation of error events. Notice in particular that , for every even w, the input sequences contributing to $R^{i,N}_{w,\leq d,w/2}$ will exclusively be composed of error events of input weight equal to 2.

For the weight enumerator coefficients of ϕ^i_N, we have the following estimations, proved in Sect. A.1.3, and Sect. A.2.1, respectively.

Lemma 3 *The following estimations hold:*

(a) If w is even,

$$R^{i,N}_{w,\leq d,w/2} \leq \frac{(2e)^w}{w^w} M_N^{w/2} \left\lfloor \frac{d}{d^i_e} \right\rfloor^{w/2}$$

(b) For $d \leq M_N/2\eta_i$,

$$A^{i,N}_{w,\leq d} \leq \begin{cases} R^{i,N}_{w,\leq d,w/2} + \frac{d}{N}\frac{C^w}{w^w} N^{\lfloor w/2 \rfloor} d^{\lceil w/2 \rceil} & \text{if } w \text{ is even} \\ \frac{C^w}{w^w} N^{\lfloor w/2 \rfloor} d^{\lceil w/2 \rceil} & \text{if } w \text{ is odd} \end{cases}$$

where C is a constant only depending on the constituent encoders.

Lemma 4 *If w is even and $\frac{d^i_e w}{2} \leq d \leq \frac{d^i_e M_N}{2\delta}$*

$$R^{i,N}_{w,\leq d,w/2} \geq \frac{2^{w/2}}{w^w} M_N^{w/2} \left\lfloor \frac{d}{d^i_e} \right\rfloor^{w/2} .$$

4 Minimum distance of the typical serial turbo code

In this section, we state and prove our main results on the minimum distance of the typical serial turbo code. Our results will indicate that, with high probability, $\mathrm{d}^{\min}_N$ scales as $d^i_e N^\beta$, where

$$\beta := 1 - \frac{2}{d^o_f} \in (0,1) .$$

First, we shall provide precise estimates of the left tail of the distribution of $\mathrm{d}^{\min}_N$. These estimates, stated in Theorem 1, extend some of the results of [14]. Then, we shall prove a deterministic upper bound on $\mathrm{d}^{\min}_N$. Such a bound, stated in Theorem 2, generalizes of the results of [3]. The most novel contribution of both Theorems 1 and 2 with respect to the existing literature consists in highlighting the role of the effective free distance of the inner encoder, d^i_e, as a linear scaling parameter for $\mathrm{d}^{\min}_N$.

We start by observing that a standard application of the union bound gives the useful estimation (see [14, Lemma 6]):

$$\mathbb{P}(\mathrm{d}^{\min}_N \leq d) \leq \sum_{w=d^o_f}^{\eta_i d} \binom{M_N}{w}^{-1} A^{o,N}_w A^{i,N}_{w,\leq d} , \qquad \forall d \leq K_N \tag{1}$$

The limitation $w \leq \eta_i$ is due to the remark that any terminating or regular error event of ϕ^i_N with output weight w has input weight bounded from above by $s\eta_i$ (and here we are considering $s = 1$).

Now, using the estimations on the weight enumerators established in the previous section, we obtain the following result on minimum distances, which generalizes [14, Thm.2.a].

Proposition 1 *For $N \to \infty$, if $d = o(N^\beta)$, then*

$$\mathbb{P}(\mathrm{d}_N^{\min} \le d) \le m_f^o \left(\frac{2e}{\sqrt{r}}\right)^{d_f^o} N^{1-\frac{d_f^o}{2}} \left\lfloor \frac{d}{d_e^i} \right\rfloor^{\frac{d_f^o}{2}} + o\left(N^{1-\frac{d_f^o}{2}} d^{\frac{d_f^o}{2}}\right)$$

Proof: It follows from (1), Lemmas 2 and 3, and (6) that

$$\mathbb{P}(\mathrm{d}_N^{\min} \le d) \quad \le \quad m_f^o \left(\tfrac{2e}{\sqrt{r}}\right)^{d_f^o} N^{1-d_f^o/2} \left\lfloor d/d_e^i \right\rfloor^{d_f^o/2} + S_e + S_o\,,$$

where

$$S_e := \frac{d}{N} \sum_{\substack{w \ge d_f^o \\ w \text{ even}}} C^w N^{\lfloor w/d_f^o \rfloor - \lceil w/2 \rceil} d^{\lceil w/2 \rceil}\,, \quad S_o := \sum_{\substack{w \ge d_f^o \\ w \text{ odd}}} C^w N^{\lfloor w/d_f^o \rfloor - \lceil w/2 \rceil} d^{\lceil w/2 \rceil}\,.$$

The terms S_e, S_o may be estimated as follows

$$S_e \le \left(\frac{d}{N}\right)^{1/2} \sum_{w \ge d_f^o + 1} \left[C N^{1/d_f^o} \left(\frac{d}{N}\right)^{1/2} \right]^w\,, \qquad S_o \le \sum_{w \ge d_f^o} \left[C N^{1/d_f^o} \left(\frac{d}{N}\right)^{1/2} \right]^w$$

Notice now that, since $d = o(N^\beta)$, then $C N^{1/d_f^o} \left(\frac{d}{N}\right)^{1/2} \to 0$. Both sums are thus convergent and are dominated by their first term. It follows that both $S_e = o\left(N^{1-d_f^o/2} d^{d_f^o/2}\right)$, and $S_o = o\left(N^{1-d_f^o/2} d^{d_f^o/2}\right)$. ■

It is possible to obtain also a lower bound for the left tail of the minimum distance distribution, showing that asymptotically the upper bound in Proposition 1 is tight. This lower bound, stated below as Proposition 2 is a novel result. Its proof combines techniques similar to those of [14, Thm. 2b] with the inclusion-exclusion principle [1, p. 124].

First of all, we fix an error event u^* for ϕ^o having active window $[0, T-1]$ for some T, and with an output $c^* = \phi^o u^*$ such that $\mathrm{w_H}(c^*) = d_f^o$. Note that $2 \le T \le d_f^o \eta_o$. Consider $N > T$. Define $c_j^* = D^j c^*$; clearly, if $|i-j| \ge T$, then c_i^* and c_j^* have non-overlapping supports. Define the set of indexes $J := \{d_f^o \eta_o\, i\,,\ i \in \mathbb{Z}^+\} \cap \{0, 1, \ldots, N-1-d_f^o\eta_o\}$, so that $i, j \in J$ with $i \ne j$ clearly ensures $|i-j| \ge d_f^o \eta_o \ge T$. For $j \in J$ and $d \in \mathbb{N}$, define the events

$$E_j^*(d) := \left\{ \mathrm{w_H}(\phi_N^i(\Pi_N(c_j^*))) \le d \right\} \cap \left\{ \phi_N^i(\Pi_N(c_j^*)) \text{ has } d_f^o/2 \text{ regular events} \right\}$$

Clearly, for any j, $E_j^*(d)$ implies $\mathrm{d}_N^{\min} \le d$, so that $\mathbb{P}(\mathrm{d}_N^{\min} \le d) \ge \mathbb{P}(\cup_{j \in J} E_j^*(d))$. The following lemma, proved in Sect. A.2.1, provides an expression for $\mathbb{P}\big(E_j^*(d)\big)$ and shows that, asymptotically, the events $E_j(d)$ are 'almost' pairwise independent.

Lemma 5

(a) for all $j \in J$, $\mathbb{P}(E_j^*(d)) = \binom{M_N}{d_f^o}^{-1} R^{i,N}_{d_f^o, \le d, d_f^o/2}$.

(b) for all $i, j \in J$ with $i \ne j$,

$$\mathbb{P}(E_i^*(d) \cap E_j^*(d)) \le \binom{M_N}{d_f^o} \binom{M_N - d_f^o}{d_f^o}^{-1} \mathbb{P}(E_i^*(d)) \mathbb{P}(E_j^*(d))$$

We shall get our lower bound by estimating the probability of the union $\cup_{j\in J} E_j^*(d)$ with the inclusion-exclusion principle.

Proposition 2 *For $N \to \infty$, if $d = o(N^\beta)$ and $d \geq \frac{1}{2} d_f^o d_e^i$, then*

$$\mathbb{P}(\mathrm{d}_N^{\min} \leq d) \geq \frac{2^{d_f^o/2}}{r^{d_f^o/2} e^{d_f^o} d_f^o \eta_o} N^{1-\frac{d_f^o}{2}} \left\lfloor \frac{d}{d_e^i} \right\rfloor^{\frac{d_f^o}{2}} + o\left(N^{1-\frac{d_f^o}{2}} d^{\frac{d_f^o}{2}}\right)$$

Proof: Using the inclusion-exclusion principle we obtain

$$\mathbb{P}(\mathrm{d}_N^{\min} \leq d) \geq \mathbb{P}\left(\bigcup_{j\in J} E_j^*(d)\right) \geq \sum_{j\in J} \mathbb{P}(E_i^*(d)) - \sum_{\substack{i,j\in J \\ i<j}} \mathbb{P}(E_i^*(d) \cap E_j^*(d))$$

We lower bound the first summation using Lemmas 5(a), and 4, and (6). Also remember that $|J| = \lfloor N/(d_f^o \eta_o) \rfloor$. We get:

$$\sum_{j\in J} \mathbb{P}(E_i^*(d)) \geq |J| \frac{R^{i,N}_{d_f^o, \leq d, d_f^o/2}}{\binom{M_N}{d_f^o}} \geq \left\lfloor \frac{N}{d_f^o \eta_o} \right\rfloor \frac{2^{d_f^o/2}}{e^{d_f^o}} M_N^{-d_f^o/2} \left\lfloor \frac{d}{d_e^i} \right\rfloor^{d_f^o/2} \tag{2}$$

We now upper bound the second summation using Lemmas 5 (b), and 3, and estimation (6):

$$\sum_{\substack{i,j\in J \\ i<j}} \mathbb{P}(E_i^*(d) \cap E_j^*(d)) \leq \binom{|J|}{2} \frac{\binom{M_N}{d_f^o}}{\binom{M_N - d_f^o}{d_f^o}} \left[\frac{R^{i,N}_{d_f^o, \leq d, d_f^o/2}}{\binom{M_N}{d_f^o}}\right]^2 \leq \Gamma\,,$$

where

$$\Gamma := \frac{1}{2} \left(\frac{N}{d_f^o \eta_o}\right)^2 \frac{\binom{M_N}{d_f^o}}{\binom{M_N - d_f^o}{d_f^o}} (2e)^{2d_f^o} \left(M_N^{-d_f^o/2} \left\lfloor \frac{d}{d_e^i} \right\rfloor^{d_f^o/2}\right)^2 .$$

Notice that $1 \leq \binom{M_N}{d_f^o}\binom{M_N - d_f^o}{d_f^o}^{-1} \leq \left(1 + \frac{d_f^o}{M_N - 2d_f^o + 1}\right)^{d_f^o}$. Hence, $\lim_{N\to\infty} \binom{M_N}{d_f^o}\binom{M_N - d_f^o}{d_f^o}^{-1} = 1$. This implies that

$$\Gamma = O\left(N^{1-\frac{d_f^o}{2}} d^{\frac{d_f^o}{2}}\right)^2 = o\left(N^{1-\frac{d_f^o}{2}} d^{\frac{d_f^o}{2}}\right).$$

Together with (2), the foregoing yields the result. ■

We may combine Propositions 1 and 2, in the following:

Theorem 1 *For all sufficiently large N, for all $\varepsilon \in (0, \beta)$,*

$$C_1 N^{-\varepsilon d_f^o/2} \leq \mathbb{P}\left(\mathrm{d}_N^{\min} \leq d_e^i N^{\beta-\varepsilon}\right) \leq C_2 N^{-\varepsilon d_f^o/2}\,,$$

for two positive constants C_1 and C_2, depending on the outer encoder only.

Theorem 1 provides fundamental insight into the effect of the constituent convolutional encoders on the minimum distance of the typical serial turbo code. On the one hand, it shows that, with probability approaching one as N goes to infinity, the minimum distance grows as a positive power of N. The exponent of such a power law growth, β, depends only on the free distance of the outer encoder, d_f^o, in an increasing way. This is in line with the results of [14]. On the other hand, it shows that the minimum distance of the typical turbo scales linearly in

the effective free distance of the inner encoder, d_e^i. While the effect of d_e^i on the average error probability of serial turbo codes has been studied in [4, 12], up to our knowledge no results have previously appeared in the literature relating d_e^i to the minimum distance. Such a scaling effect of d_e^i on $\mathrm{d}_N^{\min}$ is particularly relevant for moderate values of N. It is confirmed by the following deterministic upper bound on $\mathrm{d}_N^{\min}$.

Theorem 2 *For all* $N \geq \max\{2d_f^o\eta_o, \frac{1}{2}d_f^o\delta_i\}$

$$\mathrm{d}_N^{\min} \leq K d_e^i N^\beta \log N\,,$$

where $K := d_f^o\delta_i^2(8d_f^o\eta_o)^{2/d_f^o}$.

Theorem 2, whose proof is deferred to Sect. A.2.2, may be thought of as a generalization of [3, Thm.2]. There, only the case when the outer encoder is a repetition code was considered, while we extend it to general serial turbo codes. Indeed, [3, Thm. 3] covers a more general class of encoders with growing memory, which includes serial turbo codes, but, when specialized to the constant-memory case, gives a weaker bound than Theorem 2. The result we obtain is also asymptotically tighter than the current best known bound for serial turbo codes, presented in [16]. Moreover, our modification of [3, Thm.2] unveils the fundamental role played by d_e^i.

5 Error probability of the typical serial turbo code

In this section, we discuss implications of the previous results to the analysis of the error probability of the typical serial turbo code. For the sake of concreteness –even if the results can be easily generalized to binary-input output-symmetric memoryless channels– we shall assume the channel to be the binary additive white Gaussian noise channel: when $\omega \in \{0,1\}$ is transmitted, the output of the channel is $(-L)^\omega + \Omega$, where $L \in \mathbb{R}$ and Ω is a Gaussian variable $\Omega \sim \mathcal{N}(0,\sigma^2)$. The signal-to-noise ratio is $\mathrm{SNR} := L^2/(2\sigma^2)$.

As already mentioned, the focus of most of the previous literature on the analysis and design of serial turbo codes has been on the error probability of the average code, for which it is known [4, 12] that

$$C_1 N^{-\lfloor (d_f^o-1)/2\rfloor} \leq \mathbb{E}(P(e|\Pi_N)) \leq C_2 N^{-\lfloor (d_f^o-1)/2\rfloor}\,,$$

for some constants C_1, C_2 whose dependence on d_e^i in the high SNR regime can be made explicit.

However, the error probability of the average code turns out to be way larger than that of the typical code. Indeed, the former is dominated by an asymptotically negligible fraction of poorly performing codes. In the sequel, we shall use expurgation techniques in order to show that the decay rate of the typical serial turbo code is of order $\exp(-N^\beta)$.

We define, for every $N \in \mathbb{N}$ and $\varepsilon > 0$, the event $E_N^\varepsilon := \{\mathrm{d}_N^{\min} > N^{\beta-\varepsilon}\}$. It follows from Theorem 2 that

$$\mathbb{P}(E_N^\varepsilon) \geq 1 - C_1 N^{-\varepsilon d_f^o/2}\,. \tag{3}$$

Simply using the union-Batthacharyya bound, we get the following estimation of the average error probability of the serial turbo codes conditioned on the event E_N^ε.

Proposition 3 *If the SNR is sufficiently high, for all* $\varepsilon \in (0,\beta)$*, and sufficiently large* N*,*

$$\mathbb{E}[P(e|\Pi_N)\,|E_N^\varepsilon] \leq C_2 \exp(-2N^{\beta-\varepsilon})$$

for some positive constant C_2 *depending on* ε *but not on* N*.*

Proof: We use the union-Bhattacharyya bound for serial turbo codes established in [4]. Upon denoting by χ the indicator function of the event E_N^ε, one has

$$\mathbb{E}[P(e|\Pi_N)|E_N^\varepsilon] = \mathbb{P}(E_N^\varepsilon)^{-1}\mathbb{E}[P(e|\Pi_N)\,\chi] \leq \mathbb{P}(E_N^\varepsilon)^{-1} \sum_{h=N^{\beta-\varepsilon}}^{K_N} \sum_{w=d_f^o}^{\eta_i h} \frac{A_w^{o,N} A_{w,h}^{i,N}}{\binom{M_N}{w}} \gamma^h \,.$$

where $\gamma = \exp(-\mathrm{SNR})$. By Theorem 1, $\mathbb{P}(E_N^\varepsilon) \stackrel{N\to\infty}{\longrightarrow} 1$. So, for some $c > 0$, $\mathbb{P}(E_N^\varepsilon) \geq c$. We estimate $A_{w,h}^{i,N} \leq A_{w,\leq h}^{i,N}$ by Lemma 3 and $A_w^{o,N}$ by Lemma 2, so we can find a positive C such that:

$$\mathbb{E}[P(e|\Pi_N)|E_N^\varepsilon] \leq c^{-1} \sum_{h=N^{\beta-\varepsilon}}^{K_N} \sum_{w=d_f^o}^{\eta_i h} C^w \Big(\frac{h}{w}\Big)^{\frac{w}{2}} \Big(\frac{w}{N}\Big)^{\frac{w}{2}-\frac{w}{d_f^o}} \gamma^h.$$

Then, we remark that the function $g(z) := (a/z)^z$ has maximum value $g(a/e) = e^{a/e}$ and hence $(h/w)^{w/2} \leq e^{h/(2e)}$. Moreover, $w \leq \tilde{c}N$ for some $\tilde{c} \geq 1$, so $(w/N)^{\frac{w}{2}-\frac{w}{d_f^o}} \leq \tilde{c}^{(\frac{1}{2}-\frac{1}{d_f^o})w}$. Hence, as $w \leq \eta_i h$, we can find a constant $\tilde{C} \geq 1$ such that:

$$\mathbb{E}[P(e|\Pi_N)|E_N^\varepsilon] \leq \sum_{N^{\beta-\varepsilon}}^{K_N} (\tilde{C}\gamma)^h \,.$$

For $\gamma \leq e^{-2}/\tilde{C}$ the series is convergent and the claim follows. ∎

From Proposition 3 and the deterministic bound on $\mathrm{d}_N^{\min}$, Theorem 2, we can obtain the following result, characterizing the asymptotic decay rate of the error probability of the typical serial turbo code.

Theorem 3 *For sufficiently high SNR, for any $\varepsilon > 0$, and sufficiently large N,*

$$\mathbb{P}\Big(\exp(-N^{\beta+\varepsilon}) \leq P(e|\Pi_N) \leq \exp(-N^{\beta-\varepsilon})\Big) \geq 1 - CN^{-\varepsilon d_f^o/2}\,,$$

for some positive constant C.

Proof: By Proposition 3 and Markov's inequality, one gets that

$$\begin{aligned} \mathbb{P}\left(P(e|\Pi_N) \geq \exp(-N^{\beta-\varepsilon})\middle| E_N^\varepsilon\right) &\leq \mathbb{P}\left(P(e|\Pi_N) \geq \frac{\mathbb{E}\left[P(e|\Pi_N)|\,E_N^\varepsilon\right]}{\exp(-N^{\beta-\varepsilon})}\middle| E_N^\varepsilon\right) \\ &\leq C_2 \exp(-N^{\beta-\varepsilon})\,. \end{aligned}$$

Thus, by (3), one gets

$$\begin{aligned} \mathbb{P}\left(P(e|\Pi_N) \geq \exp(-N^{\beta-\varepsilon})\right) &\leq 1 - \mathbb{P}(E_N^\varepsilon) + \mathbb{P}\left(P(e|\Pi_N) \geq \exp(-N^{\beta-\varepsilon})\middle|\,\overline{E_N^\varepsilon}\right)\mathbb{P}(E_N^\varepsilon) \\ &\leq C_1 N^{-\varepsilon d_f^o/2} + C_2 \exp(-N^{\beta-\varepsilon}) \\ &\leq CN^{-\varepsilon d_f^o/2}\,, \end{aligned} \tag{4}$$

where the last inequality holds with $C := 2\max\{C_1, C_2\}$, for sufficiently large N.

On the other hand, using the inequality $P(e|\Pi_N) \geq p^{\mathrm{d}_N^{\min}}$, where $p = 1/2\,\mathrm{erfc}(\sqrt{\mathrm{SNR}})$ is the equivocation probability of the channel, and Theorem 2, one gets that deterministically,

$$P(e|\Pi_N) \geq \exp(-N^{\beta+o(1)})\,. \tag{5}$$

Then, the claim is an immediate consequence of (4) and (5). ∎

We conclude this section by observing that both Theorems 1 and 3 imply weak probabilistic convergence results, since the left tails of $\mathrm{d}_N^{\min}$ and $P(e|\Pi_N)$ decrease slowly in N. Indeed, one may prove [8] that, while converging in distribution to β, with probability one both the growth rate $X_N := \frac{\log \mathrm{d}_N^{\min}}{\log N}$ and the decay rate $Y_N := \frac{\log(-\log P(e|\Pi_N))}{\log N}$) densely cover the interval $[\alpha, \beta]$, where $\alpha = 1 - 2/\lceil d_f^o/2 \rceil$.

6 Conclusion

In this paper we have studied the behaviour of the minimum distance and ML error probability of serial turbo concatenations with random interleaver. We have shown that the minimum distance of the typical serial turbo code grows as a positive power of the block-length, whose exponent is an increasing function of the free distance of the outer encoder, and scales linearly with the effective free distance of the inner constituent encoder. Such a scaling law has been proven by means of a detailed study of the left tail of the minimum distance's probability distribution, and of a deterministic upper bound. As a consequence, we have characterized the subexponential decay rate of the ML error probability of the typical turbo code. In spite of the lack of concentration around the performance of the average code, our results confirm the centrality of two main design parameters for serial turbo codes suggested by the average-code analysis.

A Proofs

In the present appendix, we provide the proofs of some of the statements of Sect.s 3 and 4. Throughout, we shall make repeated use of the following well known combinatorial estimations

$$\frac{n^m}{m^m} \leq \binom{n}{m} \leq \frac{(en)^m}{m^m} \tag{6}$$

$$\binom{n-m}{m} \leq e^{n+m} \tag{7}$$

$$t^t(w-t)^{w-t} \geq (w/2)^w \quad \text{for all } t \in [0,w] \tag{8}$$

$$\frac{1}{(t-1)^{(t-1)}} \leq \frac{e\,t}{t^t} \tag{9}$$

A.1 Proofs of the results presented in Sect. 3

Our proof techniques are based on ideas from [14]. We retrace here the proofs in all detail, both since [14] has not appeared yet, and in order to be able to underline the role of d_e^i.

A.1.1 Proof of Lemma 2

Our arguments parallel those of [14, Lemma 3]. We start by introducing some notation:

- $R_d^{o,N}$ and $T_d^{o,N}$ denote, respectively, the number of inputs to ϕ_N^o having output weight d and consisting exclusively of regular error events, or containing a terminal error event. We thus have $A_d^{o,N} = R_d^{o,N} + T_d^{o,N}$.
- $R_{(d_1,\ldots,d_n)}^{o,N}$ is the number of inputs to ϕ_N^o consisting of n regular error events whose output weights are $d_1,\ldots,d_n$, respectively. Similarly, $T_{(d_1,\ldots,d_n)}^{o,N}$ is the number of inputs to ϕ_N^o consisting of $n-1$ regular error events having output weights, in order, $d_1,\ldots,d_{n-1}$ and a final terminating one of weight d_n.

Suppose that $d_1+\cdots+d_n = d$. It holds

$$R_{(d_1,\ldots,d_n)}^{o,N} \leq 2^{kd\eta_o}\binom{N}{n}$$

In fact, we are considering n error events, with lengths at most $d_1\eta_o$, ..., $d_n\eta_o$ respectively, so that the sum of their lengths is bounded by $d\eta_o$. The number of inputs in the active windows of these error events are thus at most $2^{kd\eta_o}$. The only remaining freedom is in the choice of the starting points of the error event, and the number of possibilities is clearly bounded by $\binom{N}{n}$.

Hence, one has

$$\begin{aligned} R_d^{o,N} &= \sum_{n=1}^{\lfloor d/d_f^o \rfloor} \sum_{\substack{d_1,\ldots,d_n: \\ \sum_i d_i = d, d_i \geq 1}} R_{(d_1,\ldots,d_n)}^{o,N} \\ &\leq \sum_{n=1}^{d} \binom{d}{n} 2^{kd\eta_o} \binom{N}{\lfloor d/d_f^o \rfloor} \\ &\leq 2^{(k\eta_o+1)d} \binom{N}{\lfloor d/d_f^o \rfloor}, \end{aligned} \tag{10}$$

where we are using the fact that $\lfloor d/d_f^o \rfloor \leq N/2$. Similarly,

$$T_{(d_1,\ldots,d_n)}^{o,N} \leq 2^{kd\eta_o} \binom{N}{n-1} d\eta_o$$

because the n-th event, being terminating and having length at most $d\eta_o$, starts in a position between $N - d\eta_o$ and $N-1$ on the trellis. Therefore,

$$\begin{aligned} T_d^{o,N} &= \sum_{n=1}^{\lceil d/d_f^o \rceil} \sum_{\substack{d_1,\ldots,d_n: \\ \sum_i d_i = d, d_i \geq 1}} T_{(d_1,\ldots,d_n)}^{o,N} \\ &\leq 2^d 2^{kd\eta_o} \binom{N}{\lceil d/d_f^o \rceil - 1} d\eta_o \\ &\leq 2^{(k\eta_o+\eta_o+1)d} \binom{N}{\lfloor d/d_f^o \rfloor} \end{aligned} \tag{11}$$

Summing up (10) and (11) we get (a). The tighter estimation (b) when $d = d_f^o$ is easily obtained from the observation that inputs with output weight d_f^o necessarily consist of just one error event starting in the interval $[0, N-1]$. ■

A.1.2 Proof of Lemma 3

Our arguments parallel those of [14, Lemma 1]. Similarly to what we have done before, we need to introduce several auxiliary weight enumerators for ϕ^i:

- let $R_{w,\leq d}^{i,N}$ (respectively, $T_{w,\leq d}^{i,N}$) denote the number of inputs for ϕ_N^i having input weight w, output weight not larger than d, and containing n regular error events (resp. $n-1$ regular error events plus a terminating one);
- let $R_{w,\leq d,n}^{i,N}$ (respectively, $T_{w,\leq d,n}^{i,N}$) denote the number of inputs for ϕ_N^i having input weight w, output weight not larger than d, and consisting of n regular events (resp. $n-1$ regular error events plus a terminating one);
- Fix two vectors of integers $\boldsymbol{w} = (w_1, \ldots, w_n)$ and $\boldsymbol{b} = (b_1, \ldots, b_n)$ with $w_i > 0$ and $b_i \in [0, N-1]$. Let $R_{\boldsymbol{w},\boldsymbol{b},\leq d,n}^{i,N}$ (respectively, $T_{\boldsymbol{w},\boldsymbol{b},\leq d,n}^{i,N}$) denote the number of weight-w inputs to ϕ_N^i such that: the output has weight not larger than d, and contains n regular error events (resp. $n-1$ regular error events plus a terminating one); for all $1 \leq j \leq n$ the j-the error event starts in position b_j and has input weight w_j.

(a): For any input word with $w/2$ error events and input weight w, recursiveness of ϕ^i forces input weight 2 for each error event. So the input words contributing to $R^{i,N}_{w,\leq d,w/2}$ can be written as

$$u(D) = \sum_{t=1}^{w/2} D^{b_t}(1 + D^{\delta a_t})$$

with $b_t > \delta a_{t-1}$ (so that the error events have disjoint active windows). We also have the restriction $\mathrm{w_H}\left(\phi^i(D)u(D)\right) \leq d$, but we can obtain an upper bound on the number of such words by imposing a weaker condition.

Notice that

$$\mathrm{w_H}\left(\phi^i(D)\sum_{t=1}^{w/2} D^{b_t}(1 + D^{\delta a_t})\right) = \sum_{t=1}^{w/2} \mathrm{w_H}\left(\phi^i(D)(1 + D^{\delta a_t})\right) \geq d^i_e \sum_{t=1}^{w/2} a_t$$

The restriction $\mathrm{w_H}\left(\phi^i(D)u(D)\right) \leq d$ thus implies $d^i_e \sum_{t=1}^{w/2} a_t \leq d$ and there are $\binom{\lfloor d/d^i_e \rfloor}{w/2}$ choices for $a_1, \ldots, a_{w/2}$ satisfying this relation. Finally, there are at most $\binom{M_N}{\frac{w}{2}}$ choices for the starting positions $b_1, \ldots, b_{w/2}$ of the error events. Summing up, and using (6) and (8), we obtain

$$R^{i,N}_{w,\leq d,w/2} \leq \binom{\lfloor d/d^i_e \rfloor}{\frac{w}{2}}\binom{M_N}{\frac{w}{2}} \leq \frac{1}{w^w}(2M_N)^{w/2}\lfloor d/d^i_e \rfloor^{w/2} e^w$$

This yields (a).

(b): We start by considering the case when w is even. We first show that

$$R^{i,N}_{\boldsymbol{w},\boldsymbol{b},\leq d,n} \leq \binom{d\eta_i}{w-n} \tag{12}$$

Notice indeed that $R^{i,N}_{\boldsymbol{w},\boldsymbol{b},\leq d,n}$ is smaller than the number of binary words of length $d\eta_i$ with exactly $w - n$ ones, because it is possible to exhibit an injective map between the words we want to count and such words. Given an input word (of length M_N) producing n error events having input weights $w_1, \ldots, w_n$, fixed starting points $b_1, \ldots, b_n$, and total output weight $\leq d$, map it into a word of length $d\eta_i$ in the following way: remove all the zeros outside the active windows of the error events, and furthermore remove the bit corresponding to the starting point of each error event (which is surely a one). The word obtained in such a way has surely length $< d\eta_i$, then add dummy zeros at the end to get a word of length $d\eta_i$; the number of ones is $w - n$. This map is injective since the starting points of the error events are fixed and known. This proves (12).

Now, we consider the decomposition

$$R^{i,N}_{w,\leq d,n} = \sum_{\substack{\boldsymbol{w}=(w_1,\ldots,w_n):\\ w_j\geq 2,\, \sum w_j = w}} \sum_{\substack{\boldsymbol{b}=(b_1,\ldots,b_n):\\ 0\leq b_1\leq\cdots\leq b_n\leq M_N-1}} R^{i,N}_{\boldsymbol{w},\boldsymbol{b},\leq d,n}$$

(again the constraint $w_j \geq 2$ comes from the recursiveness of ϕ^i), and we estimate using (12)

$$\begin{aligned}
\sum_{n=1}^{w/2-1} R^{i,N}_{w,\leq d,n} &\leq \sum_{n=1}^{w/2-1} \binom{w-n-1}{n-1}\binom{M_N}{n}\binom{d\eta_i}{w-n} \\
&\leq \sum_{n=1}^{w/2-1} e^{w+n-1}\frac{(eM_N)^n}{n^n}\frac{(ed\eta_i)^{w-n}}{(w-n)^{w-n}} && \text{by (6) and (7)} \\
&\leq \frac{e^{5w/2}}{(w/2)^w}\sum_{n=1}^{w/2-1} M_N^n(\eta_i d)^{w-n} && \text{by (8)} \\
&\leq \frac{e^{5w/2}\eta_i^{w/2}}{(w/2)^w}\frac{d^{\frac{w}{2}}M_N^{\frac{w}{2}}}{\frac{M_N}{d\eta_i}-1}
\end{aligned}$$

Finally, we have to consider weight enumerators of type T.

$$T^{i,N}_{w,\leq d} = \sum_{n=1}^{w/2} T^{i,N}_{w,\leq d,n} = \sum_{n=1}^{w/2} \sum_{\substack{\boldsymbol{w}=(w_1,\dots,w_n):\\ \sum w_j = w \\ w_j\geq 2\ \forall j<n,\ w_n\geq 1}} \sum_{\substack{\boldsymbol{b}=(b_1,\dots,b_n):\\ 0\leq b_1\leq\cdots\leq b_n\leq M_N-1 \\ b_n\geq M_N-d\eta_i}} T^{i,N}_{\boldsymbol{w},\boldsymbol{b},\leq d,n}$$

Everything is similar to the regular case, except for the additional condition $b_n \geq M_N - d\eta_i$. This comes from the remark that the terminating event has clearly output weight smaller than d, hence of length smaller than $d\eta_i$. Being it a terminating event, it cannot start before $M_N - d\eta_i$. Moreover, the recursiveness imposes $w_j \geq 2$ for the regular events, while for the terminating event only $w_n \geq 1$ is required.

With the same proof as for the estimation (12) of $R^{i,N}_{\boldsymbol{w},\boldsymbol{b},\leq d,n}$, we have also $T^{i,N}_{\boldsymbol{w},\boldsymbol{b},\leq d,n} \leq \binom{d\eta_i}{w-n}$, so that

$$\begin{aligned}
T^{i,N}_{w,\leq d} &\leq \sum_{n=1}^{w/2} \sum_{\substack{\boldsymbol{w}=(w_1,\dots,w_n):\\ \sum w_j = w \\ w_j\geq 2\ \forall j<n,\ w_n\geq 1}} \sum_{\substack{\boldsymbol{b}=(b_1,\dots,b_n):\\ 0\leq b_1\leq\cdots\leq b_n\leq M_N-1 \\ b_n\geq M_N-d\eta_i}} \binom{d\eta_i}{w-n} \\
&\leq \sum_{n=1}^{w/2}\binom{w-n}{n-1}\binom{M_N}{n-1}d\eta_i\binom{d\eta_i}{w-n} \\
&\leq e^{5w/2-3}\,d\eta_i\sum_{n=1}^{w/2}\frac{M_N^{n-1}(d\eta_i)^{w-n}}{(n-1)^{(n-1)}(w-n)^{(w-n)}} && \text{by (6) and (7)} \\
&\leq e^{5w/2-2}\frac{w}{2}\frac{d\eta_i}{M_N}\sum_{n=1}^{w/2}\frac{M_N^n(d\eta_i)^{w-n}}{n^n(w-n)^{(w-n)}} && \text{by (9)} \\
&\leq \frac{e^{5w/2-2}}{(w/2)^w}\frac{w}{2}\frac{d\eta_i}{M_N}\sum_{n=0}^{w/2} M_N^n(d\eta_i)^{w-n} && \text{by (8)} \\
&\leq \frac{e^{5w/2-2}}{(w/2)^w}\frac{w}{2}\frac{M_N^{w/2}(d\eta_i)^{w/2}}{\frac{M_N}{d\eta_i}-1}
\end{aligned}$$

Now everything follows from the fact that

$$A^{i,N}_{w,\le d} = R^{i,N}_{w,\le d,w/2} + \sum_{n=1}^{w/2-1} R^{i,N}_{w,\le d,n} + T^{i,N}_{w,\le d}$$

The case of odd w needs slightly more attention. We start with the analysis of $R^{i,N}_{w,\le d,\lfloor w/2\rfloor}$. Input words contributing to this term are made of $w/2-1$ events with input weight 2 and one event with input weight 3:

$$u(D) = \sum_{t=1}^{\lfloor w/2\rfloor -1} D^{b_t}(1+D^{\delta a_t}) + D^b(1+D^a+D^{a'})$$

All the error events have disjoint support, which implies, the weaker condition that $b_1 < \cdots < b_{\lfloor w/2\rfloor -1}$ and $b \neq b_1, \ldots, b_{\lfloor w/2\rfloor -1}$. The overall output weight is $\le d$, and this implies, as in (b), the weaker condition $d^i_e \sum_{t=1}^{\lfloor w/2\rfloor -1} a_t \le d$ and $a < a' < \eta_i d$. There are $\binom{\eta_i d}{2}$ choices for such a, a', $\binom{\lfloor d/d^i_e\rfloor}{\lfloor w/2\rfloor -1}$ choices for $a_1, \ldots, a_{\lfloor w/2\rfloor -1}$, no more than $\lfloor w/2\rfloor \binom{M_N}{\lfloor w/2\rfloor}$ choices for $b_1, \ldots, b_{\lfloor w/2\rfloor -1}, b$, where the factor $\lfloor w/2\rfloor$ comes from the choice of the position where to put the error event of weight 3 in between the other events. Summarizing:

$$\begin{aligned}
R^{i,N}_{w,\le d,\lfloor w/2\rfloor} &\le \left\lfloor \frac{w}{2} \right\rfloor \binom{M_N}{\lfloor w/2\rfloor} \binom{\eta_i d}{2} \binom{\lfloor d/d^i_e\rfloor}{\lfloor w/2\rfloor -1} \\
&\le \frac{\mu_i^2}{4e^2} \frac{we^w\, M_N^{\lfloor w/2\rfloor}\, d^2 \left\lfloor \frac{d}{d^i_e} \right\rfloor^{\lfloor \frac{w}{2}\rfloor -1}}{\lfloor \frac{w}{2}\rfloor^{\lfloor \frac{w}{2}\rfloor} \left(\lfloor \frac{w}{2}\rfloor -1\right)^{\lfloor \frac{w}{2}\rfloor -1}} && \text{by (6)} \\
&\le \frac{\mu_i^2}{16} \frac{w^3 (2e)^w}{w^w} M_N^{\lfloor w/2\rfloor}\, d^2 \left\lfloor \frac{d}{d^i_e} \right\rfloor^{\lfloor \frac{w}{2}\rfloor -1} && \text{by (8) and (9)}
\end{aligned}$$

The remaining regular terms are estimated exactly as in the case when w is even:

$$\sum_{n=1}^{\lfloor w/2\rfloor -1} R^{i,N}_{w,\le d,n} \le \frac{e^{5w/2} \eta_i^{\lceil \frac{w}{2}\rceil}}{(w/2)^w} \frac{d^{\lceil \frac{w}{2}\rceil} M_N^{\lfloor \frac{w}{2}\rfloor}}{\frac{M_N}{d\eta_i} - 1}$$

We now pass to studying the terms $T^{i,N}_{w,\le d}$. Differently from the even case, we shall consider the main term $T^{i,N}_{w,\le d,\lceil w/2\rceil}$ separately. Inputs contributing to $T^{i,N}_{w,\le d,\lceil w/2\rceil}$ consist of $\lfloor w/2\rfloor$ regular error events, each with input weight 2, and one terminating event with input weight 1, with overall output weight $\le d$. We represent such inputs as

$$u(D) = \sum_{t=1}^{\lfloor w/2\rfloor} D^{b_t}(1+D^{\delta a_t}) + D^{M_N - l}$$

and we observe that the following conditions hold: $0 \le b_1 < \cdots < b_{\lfloor w/2\rfloor} < M_N$, $l \le \eta_i d$, $d^i_e \sum_t a_t \le d$. We thus get:

$$T^{i,N}_{w,\le d,\lceil w/2\rceil} \le \binom{M_N}{\lfloor w/2\rfloor} d\eta_i \binom{\lfloor d/d^i_e\rfloor}{\lfloor w/2\rfloor} \le \frac{\eta_i}{2} \frac{w(2e)^w}{w^w} M_N^{\lfloor w/2\rfloor}\, d \left\lfloor \frac{d}{d^i_e} \right\rfloor^{\lfloor w/2\rfloor}. \tag{13}$$

The remaining terms are estimated as in the even case,

$$\sum_{n=1}^{\lfloor w/2\rfloor} T^{i,N}_{w,\le d,n} \le \frac{e^{5w/2-2}}{(w/2)^w}\frac{w}{2}\ \frac{M_N^{\lfloor w/2\rfloor}(d\eta_i)^{\lceil w/2\rceil}}{\frac{M_N}{d\eta_i}-1}$$

This completes the proof of Lemma 3. ∎

A.1.3 Proof of Lemma 4

We shall use ideas similar to those of [14, Lemma 2]. We consider a subclass of inputs contributing to the term $R^{i,N}_{w,\le d,w/2}$, exactly those which can be written as

$$\sum_{t=1}^{w/2} D^{i_t+h_{t-1}\delta} + D^{i_t+h_t\delta}$$

with $0 \le i_1 < i_2 < \cdots < i_{w/2} < M_N - \delta\lfloor d/d_e^i\rfloor$, and $0 = h_0 < h_1 < h_2 < \cdots < h_{w/2} \le \lfloor d/d_e^i\rfloor$. It is evident that they have input weight w and consist of $w/2$ disjoint error events. The only property which remains to be verified is whether they produce output weight not exceeding d. In fact, the t-th error event has input $D^{i_t+h_{t-1}}(1+D^{\delta(h_t-h_{t-1})})$, so that the output has weight $\mathrm{w_H}\big(\phi^i((1+D^{\delta(h_t-h_{t-1})}))\big) \le d_e^i\,(h_t-h_{t-1})$. The total output weight is thus bounded above by $d_e^i\sum_{t=1}^{w/2}(h_t-h_{t-1}) = d_e^i h_{w/2} \le d$.

Observe that, for every choice of the two $w/2$-uples $(i_1, i_2, \ldots, i_{w/2})$ and $(h_1, h_2, \ldots, h_{w/2})$, one obtains distinct inputs. It follows that

$$R^{i,N}_{w,\le d,w/2} \ge \binom{M_N - \delta\lfloor d/d_e^i\rfloor}{w/2}\binom{\lfloor d/d_e^i\rfloor}{w/2}$$

The final estimation follows applying (6) (notice that, because of the assumption made, $w/2 \le M_N - \delta\lfloor d/d_e^i\rfloor$ and $w/2 \le \lfloor d/d_e^i\rfloor$) and the inequality $M_N - \delta\lfloor d/d_e^i\rfloor \ge \frac{M_N}{2}$. ∎

A.2 Proofs of the results presented in Section 4

A.2.1 Proof of Lemma 5

This proof follows part of the proof of [14, Thm. 2.b].

The first statement is immediate, let's prove the second one. Let $c_i^* = \sum_{m=1}^{d_f^o} D^{t_m} \in \mathbb{Z}_2[D]$. Given a multi-index $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_{d_f^o}) \in [M_N]^{d_f^o}$, where $[M_N] := \{0, \ldots, M_N - 1\}$, define the event $E_{\boldsymbol{\tau}} := \{\Pi_N(D^{t_m}) = D^{\tau_m}\,\forall m = 1, \ldots, d_f^o\}$. Clearly

$$\mathbb{P}\big(E_i^*(d)\cap E_j^*(d)\big) = \sum_{\boldsymbol{\tau}\in[M_N]^{d_f^o}} \mathbb{P}\big(E_i^*(d)\cap E_{\boldsymbol{\tau}}\big)\mathbb{P}\big(E_j^*(d)\big|E_i^*(d)\cap E_{\boldsymbol{\tau}}\big)$$

Then, notice that

$$\mathbb{P}\big(E_j^*(d)\big|E_i^*(d)\cap E_{\boldsymbol{\tau}}\big) = \mathbb{P}\big(E_j^*(d)\big|E_{\boldsymbol{\tau}}\big) \le \frac{R^{i,N}_{d_f^o,\le d,d_f^o/2}}{\binom{M_N-d_f^o}{d_f^o}} = \mathbb{P}\big(E_j^*(d)\big)\frac{\binom{M_N}{d_f^o}}{\binom{M_N-d_f^o}{d_f^o}}\,.$$

Therefore,

$$\mathbb{P}\big(E_i^*(d)\cap E_j^*(d)\big) \le \sum_{\boldsymbol{\tau}\in[M_N]^{d_f^o}} \mathbb{P}\big(E_i^*(d)\cap E_{\boldsymbol{\tau}}\big)\mathbb{P}\big(E_j^*(d)\big)\frac{\binom{M_N}{d_f^o}}{\binom{M_N-d_f^o}{d_f^o}}\,.$$

Now, observe that $\sum_{\boldsymbol{\tau}\in[M_N]^{d_f^o}} \mathbb{P}(E_i^*(d)\cap E_{\boldsymbol{\tau}}) = \mathbb{P}(E_i^*(d))$. From this, the claim immediately follows. ■

A.2.2 Proof of Theorem 2

The key idea, first introduced in [3], consists in turning the problem of finding codewords of small weight into the problem of finding a generalized cycle on an hypergraph. We describe here the construction of the suitable hypergraph, adapting the construction from [3] to our setting, and then we state the Lemma on hypergraphs given in [3] which completes the proof.

Let c^*, J, c_j^* be defined as in Section 4. The aim is to show that, for any interleaver, it is possible to find a suitable subset of the c_j^*, with cardinality growing at most as $c\log N$, such that the corresponding output has weight smaller than $KN^\beta \log N$.

Define a map $\boldsymbol{\sigma} : J \to \mathbb{Z}_\delta^{d_f^o}$ by associating to an index $j \in J$ a vector $(\sigma_1(j),\dots,\sigma_{d_f^o}(j))$ in the following way: if $c_j^* = \sum_{m=1}^{d_f^o} D^{t_m}$ (with t_m increasing sequence) and $\pi(D^{t^m}) = D^{\tau_m}$, then $\sigma_m(j) = \tau_m \mod \delta$. By the pigeonhole principle, clearly there exists $U \subseteq J$ with $|U| \geq \left\lceil \frac{|J|}{\delta^{d_f^o}} \right\rceil$ such that $\boldsymbol{\sigma}(i) = \boldsymbol{\sigma}(j)$ for all $i, j \in U$.

From now on, we shall consider only c_j^* with $j \in U$. The idea is that, as all the ones in these words are permuted to positions at a distance multiple of δ, when applying ϕ^i any pair of them gives an output weight which is proportional to the distance within the input ones. So, the aim is to find a subset of indexes $S \subseteq U$ such that the corresponding c_j's form pairs of ones in such a way that we number of pairs grows at most logarithmically in N, and that the distance within ones of the same pair grows at most as N^β.

Now look at $[M_N] = \{0,\dots,M_N-1\}$ and divide it in b intervals $I_1,\dots,I_b$, each of length $\lfloor M_N/b \rfloor$ (except for a possibly longer one at the end); b is a parameter depending on N that will be properly chosen later.

Define an hypergraph $H = (V, E)$ in the following way. Take a d_f^o-partite vertex set V being the union of d_f^o disjoint copies of $W = \{I_1,\dots,I_b\}$. The set of hyperedges E has cardinality $|U|$ and is d_f^o-regular in the sense that $E \subseteq W^{d_f^o}$, i.e. every hyperedge contains exactly one vertex from each of the d_f^o copies of W. Any hyperedge in E corresponds to an index $j \in U$, and is defined as $e = (I_{h_1},\dots,I_{h_{d_f^o}}) \in W^{d_f^o}$ where, denoting $c_j^* = \sum_{m=1}^{d_f^o} D^{t_m}$ as before, h_m is such that $\pi(D^{t_m}) \in I_{h_m}$.

Define the degree of a vertex in the hypergraph as the number of hyperedges that contain that vertex. The following lemma holds true:

Lemma 6 ([3], Lemma 3) *Given a k-partite, k-regular hypergraph (V, E) with b vertices in each part, if $4b^{\lceil k/2\rceil} \leq |E|$, then there exists a non-empty subset $S \subset E$, with $|S| \leq k\log b$, such that in the induced subhypergraph (V, S) every vertex has even degree (possibly zero).* ■

We shall show here that this lemma implies Theorem 2. In the above construction of the hypergraph H, choose b such that $4b^{d_f^o/2} \leq \left\lceil \frac{1}{\delta^{d_f^o}} \left\lfloor \frac{N}{d_f^o \eta_o} \right\rfloor \right\rceil$; this ensures $4b^{d_f^o/2} \leq |E| = |U|$, so that we can apply Lemma 6 and find the subset S. There is a bijection from S to a subset $\tilde{S} \subset U$: any $s \in S$ corresponds to some c_j^*, $j \in \tilde{S}$. Observe that $c := \sum_{j\in\tilde{S}} c_j^*$ is clearly a feasible output of the inner encoder. Then, $\phi_{i,N}(\pi(c))$ is a possible output of the serial scheme.

By construction $\pi(c)$ is composed of $|S|d_f^o/2$ pairs of 1's. Each pair lives in the same interval I_j and has distance a multiple of δ. Hence,

$$\mathrm{w_H}\big(\phi_N^i(\pi(c))\big) \leq \frac{|S|d_f^o}{2} d_e^i \frac{N}{b}$$

Finally use the bound on $|S|$ which is the key contribution of Lemma 6: $|S| \leq d_f^o \log b$. By the way b was chosen, this completes the proof of the claim. ■

B Generalizing to the case of odd d_f^o

In this section, we discuss how one of the simplifying assumptions we made, i.e. that d_f^o be even, can be removed. The reader is referred to [11] for further generalizations, in particular the cases when the inner encoder has non-scalar input ($s > 1$), or is not proper-rational, and the case when $d_f^o = 2$.

Throughout this section, we shall consider the case when $d_f^o \geq 3$ is odd. Most of the results extend to this case, and in particular we will prove that the growth rate for $\mathrm{d}_N^{\min}$ is N^β (Thm. 4).

First, notice that Lemmas 2 and 3 hold true without any modification. This allows one to prove the following result, consisting in the analogous of Proposition 1 for odd d_f^o.

Proposition 4 *For $N \to \infty$, if $d = o(N^\beta)$, then*

$$\mathbb{P}(\mathrm{d}_N^{\min} \leq d) = O\left(N^{1-\lceil d_f^o/2 \rceil} d^{\lceil d_f^o/2 \rceil}\right) + O\left(N^{2-d_f^o} d^{d_f^o}\right)$$

Proof: From (1), by estimating the enumerating coefficients of the constituent encoders with Lemmas 2 and 3, one gets:

$$\mathbb{P}(\mathrm{d}_N^{\min} \leq d) \leq \sum_{w=d_f^o}^{\eta_i d} C^w N^{\lfloor w/d_f^o \rfloor - \lceil w/2 \rceil} d^{\lceil w/2 \rceil} \tag{14}$$

for some $C > 0$ depending on ϕ^o and ϕ^i, but neither on N nor d. Now, we shall separately consider different terms. Write $w = a d_f^o + b$, with integers $a \geq 1$, $0 \leq b < d_f^o$. Then, observe that

$$N^{\lfloor w/d_f^o \rfloor - \lceil w/2 \rceil} d^{\lceil w/2 \rceil} = \begin{cases} \left(\frac{d}{N}\right)^{b/2} \left(N^{1-d_f^o/2} d^{d_f^o/2}\right)^a & \text{if } a+b \text{ is even} \\ \left(\frac{d}{N}\right)^{\frac{b+1}{2}} \left(N^{1-d_f^o/2} d^{d_f^o/2}\right)^a & \text{if } a+b \text{ is odd} \end{cases}$$

As $N \to \infty$, If $d = o(N^\beta)$, then $N^{1-d_f^o/2} d^{d_f^o/2} \to 0$. Hence, $\sum_a (N^{1-d_f^o/2} d^{d_f^o/2})^a$ converges for all sufficiently large N. We may split the summation in (14) in the following four terms (with the notation $[d_f^o] = \{0, 1, \ldots, d_f^o - 1\}$):

- $\displaystyle\sum_{\substack{b \in [d_f^o] \\ b \text{ even}}} \left(\frac{d}{N}\right)^{b/2} \sum_{\substack{a \in \mathbb{Z}^+ \\ a \text{ even}}} \left(N^{1-d_f^o/2} d^{d_f^o/2}\right)^a \leq c_1 N^{2-d_f^o} d^{d_f^o}$
- $\displaystyle\sum_{\substack{b \in [d_f^o] \\ b \text{ odd}}} \left(\frac{d}{N}\right)^{b/2} \sum_{\substack{a \in \mathbb{Z}^+ \\ a \text{ odd}}} \left(N^{1-d_f^o/2} d^{d_f^o/2}\right)^a \leq c_2 \left(\frac{d}{N}\right)^{1/2} N^{1-d_f^o/2} d^{d_f^o/2}$
- $\displaystyle\sum_{\substack{b \in [d_f^o] \\ b \text{ even}}} \left(\frac{d}{N}\right)^{\frac{b+1}{2}} \sum_{\substack{a \in \mathbb{Z}^+ \\ a \text{ odd}}} \left(N^{1-d_f^o/2} d^{d_f^o/2}\right)^a \leq c_3 \frac{d}{N} N^{2-d_f^o} d^{d_f^o}$
- $\displaystyle\sum_{\substack{b \in [d_f^o] \\ b \text{ odd}}} \left(\frac{d}{N}\right)^{\frac{b+1}{2}} \sum_{\substack{a \in \mathbb{Z}^+ \\ a \text{ even}}} \left(N^{1-d_f^o/2} d^{d_f^o/2}\right)^a \leq c_4 \left(\frac{d}{N}\right)^{1/2} N^{1-d_f^o/2} d^{d_f^o/2}$

for some constants $c_1, c_2, c_3, c_4 > 0$. Finally, the claim follows upon observing that $\frac{d}{N} N^{2-d_f^o} d^{d_f^o} = o\left(N^{2-d_f^o} d^{d_f^o}\right)$. ■

More precise results highlighting the dependence on d_e^i as in Proposition 1 could be obtained, as well a lower bound on the left tail of $\mathrm{d}_N^{\min}$ analogous to Proposition 2. However, we shall not pursue this direction here. Also, one may try to extend the deterministic upper bound on $\mathrm{d}_N^{\min}$. It turns out that Theorem 2 holds true in the case when d_f^o is odd, with the growth parameter β replaced by the larger parameter

$$\tilde{\beta} := 1 - \frac{1}{\lceil d_f^o/2 \rceil} = 1 - \frac{2}{d_f^o + 1}\,.$$

However, it is still possible to prove that N^β is the actual growth rate of $\mathrm{d}_N^{\min}$, using a second-order method, as shown below.

Theorem 4 *Assume that d_f^o is odd. If $d/N^\beta \to \infty$, then $\mathbb{P}(\mathrm{d}_N^{\min} \le d) \to 1$.*

Proof: We follow the same arguments used to prove the lower bound for the left tail of the minimum distance distribution for even d_f^o. We fix an error event u^* for ϕ^o having active window $[0, T-1]$ for some T, and with an output $c^* = \phi^o u^*$ such that $\mathrm{w_H}(c^*) = d_f^o$. We consider $I = \{0, 1, \ldots, N - 1 - d_f^o \eta_o\}$ and error events $c_i^* = D^i c^*$ for $i \in I$. Assume that $N > T$.

For $i, j \in I$, we define:

$$E_{ij}^*(d) := \Big\{ \Pi(c_i^*) = \sum_{t=1}^{d_f^o} D^{b_t} \text{ and } \Pi(c_j^*) = \sum_{t=1}^{d_f^o} D^{b_t + l_t \delta} \\ \text{for some } 0 \le b_1 < \cdots < b_{d_f^o} \le M_N,\, l_t \ge 1,\, \sum_{t=1}^{d_f^o} l_t \le \left\lfloor \frac{d}{d_e^i} \right\rfloor \Big\}$$

Now define the random variable $Z := \sum_{i,j \in I,\, i \ne j} \mathbb{1}_{E_{ij}^*(d)}$. Clearly

$$\mathbb{P}(\mathrm{d}_N^{\min} \le d) \ge \mathbb{P}\Big(\bigcup_{i,j \in I,\, i \ne j} E_{ij}^*(d) \Big) = 1 - \mathbb{P}(Z = 0)$$

A standard argument, consequence of Chebyshev's inequality [1, Thm. 4.3.1], gives

$$\mathbb{P}(Z = 0) \le \frac{\mathbb{E}(Z^2)}{[\mathbb{E}(Z)]^2} - 1\,,$$

so that

$$\mathbb{P}(\mathrm{d}_N^{\min} \le d) \ge 2 - \frac{\mathbb{E}(Z^2)}{[\mathbb{E}(Z)]^2} = 2 - \frac{\sum_{\substack{i,j,k,l \in I \\ i \ne j, k \ne l}} \mathbb{P}\big(E_{ij}^*(d) \cap E_{kl}^*(d)\big)}{\Big[\sum_{\substack{i,j \in I \\ i \ne j}} \mathbb{P}\big(E_{ij}^*(d)\big)\Big]^2} \tag{15}$$

We now estimate the right hand term in a number of steps:

- a look at the proof of Lemma 4 (with $w = 2d_f^o$), gives:

$$\mathbb{P}\big(E_{ij}^*(d)\big) \ge \frac{1}{\binom{M_N}{2d_f^o}} \frac{2^{d_f^o}}{(d_f^o)^{2d_f^o}} M_N^{d_f^o} \left\lfloor \frac{d}{d_e^i} \right\rfloor^{d_f^o}$$

- with a similar proof to Lemma 5 (i.e. using the same conditioning trick) you find that, if i, j, k, l are all distinct:

$$\mathbb{P}\big(E^*_{ij}(d) \cap E^*_{kl}(d)\big) \leq \frac{\binom{M_N}{2d^o_f}}{\binom{M_N - 2d^o_f}{2d^o_f}} \mathbb{P}\big(E^*_{ij}(d)\big)\mathbb{P}\big(E^*_{kl}(d)\big)$$

- simple counting gives that, if i, j, k are all distinct,

$$\mathbb{P}\big(E^*_{ij}(d) \cap E^*_{ik}(d)\big) \leq \frac{1}{\binom{M_N}{3d^o_f}} \binom{M_N}{d^o_f} \binom{\lfloor d/d^i_e \rfloor}{d^o_f}^2$$

  and the same bound holds for $\mathbb{P}\big(E^*_{ij}(d) \cap E^*_{kj}(d)\big)$

so that we can split the summation in Eq. (15) in the following terms:

- $\dfrac{\sum_{\substack{i,j,k,l \in I \\ i,j,k,l \text{ distinct}}} \mathbb{P}\big(E^*_{ij}(d) \cap E^*_{kl}(d)\big)}{\left[\sum_{\substack{i,j \in I \\ i \neq j}} \mathbb{P}\big(E^*_{ij}(d)\big)\right]^2} \overset{N \to \infty}{\longrightarrow} 1$
- $\dfrac{\sum_{\substack{i,j,k \in I \\ i,j,k \text{ distinct}}} \left[\mathbb{P}\big(E^*_{ij}(d) \cap E^*_{ik}(d)\big) + \mathbb{P}\big(E^*_{ij}(d) \cap E^*_{kj}(d)\big)\right]}{\left[\sum_{\substack{i,j \in I \\ i \neq j}} \mathbb{P}\big(E^*_{ij}(d)\big)\right]^2} \leq c_1 \dfrac{1}{N}$
  for some constant $c_1 > 0$;
- $\dfrac{\sum_{\substack{i,j \in I \\ i \neq j}} \mathbb{P}\big(E^*_{ij}(d)\big)}{\left[\sum_{\substack{i,j \in I \\ i \neq j}} \mathbb{P}\big(E^*_{ij}(d)\big)\right]^2} \leq c_2 \dfrac{1}{N^{2-d^o_f} d^{d^o_f}} = c_2 \left(\dfrac{N^\beta}{d}\right)^{d^o_f}$
  for some constant $c_2 > 0$.

This ends the proof. ■

Acknowledgments

The authors thank Rüdiger Urbanke for an interesting discussion on the topics of this paper.

References

[1] N. Alon, and J. Spencer, "The probabilistic method", 3rd Ed., J. Wiley & Sons, Hoboken, NJ, USA, 2008.

[2] A. Barg, and G. D. Forney, Jr., "Random codes: minimum distances and error exponents", *IEEE Trans. Inf. Theory*, vol. 48, pp. 2568–2573, 2002.

[3] L. Bazzi, M. Mahdian, and D.A. Spielman, "The Minimum Distance of Turbo-like Codes", *IEEE Trans. Inf. Theory*, vol. 55, pp. 6–15, 2009.

[4] S. Benedetto, D. Divsalar, G. Montorsi and F. Pollara, "Serial concatenation of interleaved codes: Performance analysis, design and iterative decoding", *IEEE Trans. Inf. Theory*, vol. 44, pp. 909–926, 1998.

[5] S. Benedetto and G. Montorsi, "Design of parallel concatenated convolutional codes", *IEEE Trans. Communicat.*, vol. 44, pp. 591–600, 1996.

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

[7] M. Breiling, "A logarithmic upper bound on the minimum distance of turbo codes", *IEEE Trans. Inf. Theory*, vol. 50, pp. 1692–1710, 2004.

[8] G. Como, F. Fagnani, F. Garin, "ML Performances of Serial Turbo Codes do not Concentrate", *Proc. of the 4th International Symposium on Turbo Codes and Related Topics (Munich, Germany)*, April 2006.

[9] F. Fagnani, "Performance of parallel concatenated coding schemes", *IEEE Trans. Inf. Theory*, vol. 54, pp. 1521–1535, 2008.

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

[11] F. Garin, *Generalized serial turbo coding ensembles: analysis and design*, Ph.D. Thesis, Politecnico di Torino, Torino, Italy, March 2008.

[12] F. Garin and F. Fagnani, "Analysis of serial turbo codes over Abelian groups for symmetric channels", *Siam J. Discr. Math.*, vol. 22, pp. 1488–1526, 2008.

[13] H. Jin and R. J. McEliece, "Coding theorems for turbo code ensembles", *IEEE Trans. Inf. Theory*, vol. 48, pp. 1451–1461, 2002.

[14] N. Kahale and R. Urbanke, "On the minimum distance of parallel and serially concatenated codes", submitted, 1997.

[15] D. J. C. MacKay, "Good Error Correcting Codes Based On Very Sparse Matrices", *IEEE Trans. Inf. Theory*, vol. 45, pp. 399–431, 1999.

[16] A. Perotti and S. Benedetto, "An Upper Bound on the Minimum Distance of Serially Concatenated Convolutional Codes", *IEEE Trans. Inf. Theory*, vol. 52, pp. 5501-5509, 2006.