The performance of serial turbo codes does not concentrate

Garin, Federica; Como, Giacomo; Fagnani, Fabio

Published in:
IEEE Transactions on Information Theory

DOI:
10.1109/TIT.2012.2184673

2012

The performance of serial turbo codes does not concentrate

Federica Garin, Giacomo Como, and Fabio Fagnani

***Abstract*—Minimum distances and maximum likelihood error probabilities of serial turbo codes with uniform interleaver are analyzed. It is shown that, for a fraction of interleavers approaching one as the block-length grows large, the minimum distance of serial turbo codes grows as a positive power of their block-length, while their error probability decreases exponentially fast in some positive power of their 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 poorly performing interleavers, and decays only as a negative power of the block-length. The analysis proposed in this paper relies on precise bounds of the minimum distance of the typical serial turbo code, whose scaling law is shown to depend both on the free distance of its outer constituent encoder, which determines the exponent of its sublinear growth in the block-length, and on the effective free distance of its inner constituent encoder. The latter is defined as the smallest weight of codewords obtained when the input word of the inner encoder has weight two, and appears as a linear scaling factor for the minimum distance of the typical serial turbo code. 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. By showing for the first time that the typical serial turbo code's minimum distance scales linearly in the effective free distance of the inner constituent encoder, the presented results generalize, and improve upon, the probabilistic bounds of Kahale and Urbanke, as well as the deterministic upper bound of Bazzi, Mahdian, and Spielman, where only the dependence on the outer encoder's free distance was proved.**

***Index Terms*—Turbo codes, serially concatenated codes, minimum distance, error probability, typical code analysis.**

I. Introduction

Serially concatenated convolutional codes with random interleaver, briefly serial turbo codes, were introduced in [5], 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 [6] in order to explain the performance of Berrou et al.'s parallel turbo codes [8]. In a nutshell, the idea consists in fixing the outer and the inner constituent encoders, and in studying the maximum likelihood (ML) error probability averaged over all possible interleavers. The main result in [5] is an upper bound to the average error probability which decays to zero as a negative power of the interleaver length. The exponent of such power law decay, usually referred to as 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 (SNR) regime. The fundamental design parameter characterizing the performance in this regime is the *effective free distance* of the inner encoder, defined as the smallest weight of codewords obtained when the input word of the inner encoder has weight two. These ideas have been rigorously formalized first in [24] and then, in a more general setting, in [22], where also a lower bound is proved differing from the upper bound only by a multiplicative constant, thus showing that the bound is tight for the *average serial turbo code*.

In fact, the average code analysis has been the main tool used in the literature to study the performance of turbo and turbo-like codes in the 'waterfall' SNR region, see e.g. [14], [10], [34], [1], [27], [23] for a (non-exhaustive) list of examples of papers on the average error probability of serial turbo-like ensembles, including recent work. The effectiveness of the design based on the average performance might lead one to believe that there is a concentration phenomenon, i.e., almost all codes perform closely to the average one. In this paper, we shall prove that this is not the case, as the typical serial turbo code performs much better than the average one. Nevertheless, as explained in the sequel, the typical serial turbo code analysis shows the relevance of the same design parameters highlighted by the average code analysis, namely, the free distance of the outer encoder and the effective free distance of the inner encoder.

A notable exception to the aforementioned literature based on the average turbo code analysis is provided by the early manuscript [26], whose focus is on the probability distribution of the minimum distance of parallel and serial turbo code ensembles, rather than on the ML error probability of the average turbo code. A related line of research has focused on deterministic bounds on the minimum distance, initiated by Breiling [9] for parallel turbo codes, and developed in the serial case in [4], [32]. A side research effort has also concerned algorithms for numerical computation of minimum distance, see in particular [20].

It is shown in [26] that, with high probability, the minimum distance of serial turbo codes grows like $N^{1-2/d_{\mathrm{f}}^{\mathrm{o}}}$, where N is the block-length, and $d_{\mathrm{f}}^{\mathrm{o}}$ is the free distance of the outer

An earlier version of this work has been presented at the 4th International Symposium on Turbo Codes and Related Topics held in Munich, Germany, on April 3–7, 2006.

F. Garin is with NeCS, INRIA Rhône-Alpes, Grenoble, France. E-mail: federica.garin@inria.fr

G. Como is with the Department of Automatic Control, Lund University, Lund, Sweden. E-mail: giacomo.como@control.lth.se

F. Fagnani is with Dipartimento di Matematica, Politecnico di Torino, Torino, Italy. E-mail: fabio.fagnani@polito.it

constituent encoder, and the scaling is up to some unspecified constants which depend both on the inner and on the outer encoders, but not on the block-length. This result implies that, for almost all choices of the interleaver, serial turbo codes have ML error probability decreasing to zero exponentially in a positive power of the block-length, thus showing that, due to the presence of an asymptotically vanishing fraction of bad codes, the average-code analysis provides too conservative a prediction of the behavior of the *typical serial turbo code*.

In fact, an analogous phenomenon has long been known to occur for other code ensembles, and this has motivated a considerable research effort in the analysis of the distance spectra of such ensembles. Early results for random and linear code ensembles at low rates, as well as low-density parity-check (LDPC) code ensembles appear already in Gallager's thesis [19, Ch. 3], while more recent rigorous results are reported, e.g., in [3] and [28, Ch. 6] for binary random and linear code ensembles, [29] and [28, Ch. 11] for binary LDPC code ensembles, [7], [12], and [13] for code ensembles over groups for non-binary input channels. For a related stream of literature based on the application of non-rigorous but powerful techniques of statistical physics to the analysis of LDPC codes, see, e.g., [30], [18], [31], [35] and [28, Ch. 21]. It is worth mentioning that, in contrast to the ML error probability, other parameters of these code ensembles, such as the weight-enumerating coefficients, may concentrate in some cases, see, e.g., [3] for random and linear code ensembles and [33] for regular LDPC code ensembles.

However, despite the lack of concentration of the serial turbo code ensemble's performance, the results in [26] show that the scaling law of the typical serial turbo code's minimum distance is characterized by the outer encoder's free distance, $d_{\mathrm{f}}^{\mathrm{o}}$, which is the same main design parameter suggested by the average code analysis [5], [24], [22]. On the other hand, no design parameter of the inner encoder emerges from the analysis proposed by [26], [4].

The main contribution of the present paper consists in showing that the scaling law of the performance of the typical serial turbo code does depend also on the inner constituent encoder's effective free distance, to be denoted by $d_{\mathrm{e}}^{\mathrm{i}}$. We shall prove (see Theorem 1) that, with high probability, the minimum distance of serial turbo codes scales like

$$d_{\mathrm{e}}^{\mathrm{i}} N^{1-2/d_{\mathrm{f}}^{\mathrm{o}}} ,$$

up to some constants which depend on the outer encoder only. This result generalizes and improves upon the aforementioned probabilistic bounds of [26, Thm. 2]. We shall also prove (see Theorem 2) a deterministic upper bound on the minimum distance of serial turbo codes, which shows an analogous dependance on the inner and outer encoder's parameters. This result generalizes and improves upon some of the bounds of [4], with the main improvement consisting in highlighting the dependance of the bound on the inner encoder's parameters. Also, it improves asymptotically on the best known deterministic bound for minimum distance of serial turbo codes, presented in [32]. Finally, by means of code-expurgation techniques, these results will allow us to show (see Theorem 3) that the ML error probability of the typical turbo code decreases exponentially fast in a positive power of the block-length.

The analysis performed in this paper involves, on the one hand, precise bounds on the cumulative distribution function (CDF) of the serial turbo code's minimum distance, whose proofs heavily rely on the combinatorial ideas developed in [26]. On the other hand, our proof of the deterministic upper bound makes use of some of the techniques devised in [4]. For all the probabilistic bounds, we shall present completely self-contained proofs. Our choice is in the interest of readability, both since the manuscript [26] has not been published yet, and because our results do not follow from the statements in [26] but rather involve some suitable modification of the arguments therein. Moreover, we shall consider a family of constituent encoders which is more general than the one defined in [26], where only systematic recursive convolutional encoders of rate $1/2$ were used.

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 bounds on the weight-enumerating coefficients 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 I while Appendix II contains some extensions.

Before proceeding, we establish 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 'as N grows large, if $d = o(N)$ and $w \leq d$, then $f(w, d, N) = o(N^a)$' means that if $d = d_N$, $w = w_N$ satisfy $w_N \leq d_N$ and d_N/N vanishes, as N grows large, then $f(w_N, d_N, N)/N^a$ converges to 0. When we say 'w is constant' we mean it does not depend on N. We shall also write $f(N) = \omega(g(N))$ to mean $g(N) = o(f(N))$.

II. Problem setting

In this section we establish some notation on convolutional encoders, and introduce the serial turbo code ensemble. Since we do not want to put a priori limitations on the rate of constituent encoders and/or their structure (e.g., systematic encoders), we shall consider below general convolutional encoders.

A. Convolutional encoders

In this section, we recall a few definitions and properties of convolutional encoders that are essential for this paper. We refer the reader to [16] and [25] for classical results on convolutional encoders, and to [17], [15], [22] for more details on those properties which are useful in the study of turbo-like concatenations.

Denote by $\mathbb{Z}_+$ the set of non-negative integers, and consider a map

$$\phi : (\mathbb{Z}_2^k)^{\mathbb{Z}_+} \to (\mathbb{Z}_2^r)^{\mathbb{Z}_+} ,$$

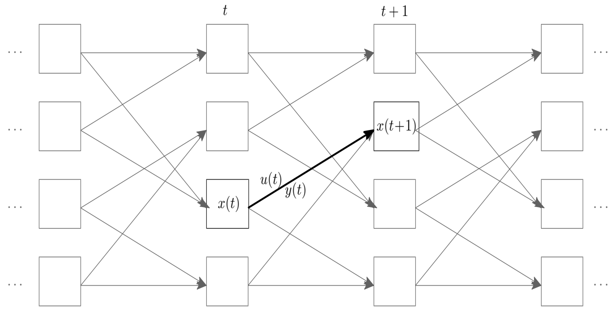

Fig. 1. Section of the trellis associated to a convolutional encoder. At time $t \geq 0$, the state is $x(t) \in \mathbb{Z}_2^\mu$. Then, in response to an input $u(t) \in \mathbb{Z}_2^k$, an output $y(t) = Hx(t) + Wu(t) \in \mathbb{Z}_2^r$ is produced, and the state is updated as $x(t+1) = Fx(t) + Gu(t) \in \mathbb{Z}_2^\mu$.

i.e., ϕ maps an input word which is an infinite sequence of vectors[1] having k bits each into an output word which is an infinite sequence of vectors having r bits each. We say that the map ϕ is a *convolutional encoder* if it admits a linear finite state-space realization. This means that the relationship between the input and the output words (codewords) can be described by a linear dynamical system with finite memory. More precisely, there exist a state space $X = \mathbb{Z}_2^\mu$ and matrices F, G, H, W of suitable dimensions and with binary entries, such that $y = \phi(u)$ if and only if there exists a (unique) state sequence $x \in (\mathbb{Z}_2^\mu)^{\mathbb{Z}_+}$ such that $x(0) = 0$ and, for all t,

$$x(t+1) = Fx(t) + Gu(t)\,, \qquad y(t) = Hx(t) + Wu(t)\,. \quad (1)$$

We shall say that x is the state sequence associated with u.

The state realization is usually pictorially represented as a labeled graph, called trellis. To construct the trellis, for each $t \in \mathbb{Z}_+$, draw 2^μ points, corresponding to elements of the state space X; then draw an edge from state x at time t to state x' at time $t+1$, with input label $a \in \mathbb{Z}_2^k$ and output label $b \in \mathbb{Z}_2^r$ if and only if $x' = Fx + Ga$ and $b = Hx + Wa$ (see Figure 1).

The minimal realization (i.e., the one having the smallest μ) of a given convolutional code is unique (up to a change of basis for the state space), and has the observability and controllability properties which are essential for defining the terminated encoders (see below) and for proving Lemma 1. In this paper we shall always assume that we are using a minimal realization, in a fixed choice of coordinates for the state space, and we shall refer to it as the trellis of the encoder.

A convolutional encoder ϕ is said to be *recursive* if, for every input word u with Hamming weight[2] $\mathrm{w_H}(u) = 1$, the corresponding codeword $\phi(u)$ has infinite Hamming weight. The encoder is said to be *non-catastrophic* if every codeword $\phi(u)$ having finite Hamming weight comes from an input word u which also has finite Hamming weight. The *free distance* and the *effective free distance* of ϕ are defined, respectively, as

$$d_{\mathrm{f}} := \min\{\mathrm{w_H}(\phi(u)) \ : \ u \neq 0\}\,, \quad (2)$$

$$d_{\mathrm{e}} := \min\{\mathrm{w_H}(\phi(u)) \ : \ \mathrm{w_H}(u) = 2\}\,. \quad (3)$$

[1]Throughout this paper, vectors are column vectors.

[2]Throughout this paper, Hamming weight is to be intended bit-wise, i.e., the number of ones in the word, and not the number of non-zero vectors.

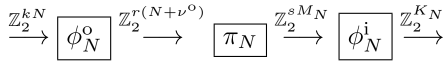

Fig. 2. A serially concatenated encoding scheme.

Given $u \in (\mathbb{Z}_2^k)^{\mathbb{Z}_+}$, we define the *support*[3] of u as $\mathrm{supp}(u) := \{t \in \mathbb{Z} \ : \ u(t) \neq 0\}$. The *block-termination* of a convolutional encoder ϕ after N trellis steps is defined as follows. Fix $N \in \mathbb{Z}_+$, consider an input word u with $u(t) = 0$ for all $t \geq N$, and let x be the associated state sequence. Notice that the state sequence x and the output word $y = \phi(u)$ may not be supported in the same interval. Indeed, it can happen that $x(N) \neq 0$ and $y(N) \neq 0$. However, thanks to the controllability of the minimal realization (see, e.g., [36] or [17]) there exists an integer $\nu \in [0, \mu]$ (called *constraint length* and not depending on the particular u nor on N), and an input word $\tilde{u}$ coinciding with u on $[0, N-1]$ and supported inside $[0, N+\nu-1]$ such that the associated state sequence $\tilde{x}$ has $\tilde{x}_{N+\nu} = 0$ and thus also the corresponding output word is supported in $[0, N+\nu-1]$. Moreover, the pole placement theorem (see, e.g., [36]) ensures that it is always possible to choose the terminating inputs $\tilde{u}(N), \ldots, \tilde{u}(N+\nu-1)$ to be a linear state-feedback, i.e., to have the form $\tilde{u}(t) = -Kx(t)$ for all $t = N, \ldots, N+\nu-1$, for a suitable $K \in \mathbb{Z}_2^{k\times\mu}$ which depends only on the encoder ϕ, not on u nor on N. In this paper, we shall assume that, given a convolutional encoder ϕ, a matrix K has been chosen allowing one to construct the terminating inputs. Then, the block termination of ϕ after N trellis steps is defined as the map

$$\phi_N : \mathbb{Z}_2^{kN} \to \mathbb{Z}_2^{r(N+\nu)}$$

which associates to an input word

$$(u^T(0), u^T(1), \ldots, u^T(N-1))^T$$

the output word

$$(y^T(0), y^T(1), \ldots, y^T(N-1), y^T(N), \ldots, y^T(N+\nu-1))^T$$

such that

$$\phi\big(u(0), u(1), \ldots, u(N-1), \tilde{u}(N), \ldots, \tilde{u}(N+\nu-1), 0, \ldots\big)$$
$$= \big(y(0), y(1), \ldots, y(N-1), y(N), \ldots, y(N+\nu-1), 0, \ldots\big)\,,$$

where $\tilde{u}(N), \ldots, \tilde{u}(N+\nu-1)$ is the above-described terminating input obtained as a linear state-feedback. Such a choice of the terminating input immediately implies that ϕ_N is a $\mathbb{Z}_2$-linear block encoder.

B. Serially concatenated convolutional encoders with random interleaver

We start from two convolutional encoders

$$\phi^{\mathrm{o}} : (\mathbb{Z}_2^k)^{\mathbb{Z}_+} \to (\mathbb{Z}_2^r)^{\mathbb{Z}_+}\,, \qquad \phi^{\mathrm{i}} : (\mathbb{Z}_2^s)^{\mathbb{Z}_+} \to (\mathbb{Z}_2^l)^{\mathbb{Z}_+}\,.$$

Let ν^{o} and ν^{i} be their corresponding constraint lengths and let N be a positive integer such that s divides $r(N+\nu^{\mathrm{o}})$. Let M_N be such that

$$sM_N = r(N+\nu^{\mathrm{o}})\,,$$

[3]Notice that the size of the support is the number of non-zero vectors in the sequence u. Hence, $|\mathrm{supp}(u)| = \mathrm{w_H}(u)$ when $k = 1$, while the equality need not hold true in general for $k > 1$.

TABLE I
THE RELEVANT PARAMETERS OF THE CONSTITUENT ENCODERS OF THE SERIAL SCHEME IN FIGURE 2

$d_{\mathrm{f}}^{\mathrm{o}}$	free distance of ϕ^{o}, see (2)
$d_{\mathrm{e}}^{\mathrm{i}}$	effective free distance of ϕ^{i}, see (3)
δ^{i}	defined in Sect. III
η^{o}, η^{i}	defined in Lemma 1
ν^{o}, ν^{i}	constraint lengths of ϕ^{o} and ϕ^{i}, see Sect. II-A
μ^{o}	memory (size of minimal state space) of ϕ^{o}, see Sect. II-A

and let

$$K_N := l(M_N + \nu^{\mathrm{i}}) = l(\tfrac{r}{s}(N + \nu^{\mathrm{o}}) + \nu^{\mathrm{i}})\,.$$

Consider the block terminations of ϕ^{o} and ϕ^{i} after N and M_N trellis steps, respectively

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

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

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

depicted in Figure 2. We shall refer to ϕ^{o} as the *outer encoder*, to ϕ^{i} as the *inner encoder*, and to π_N as the *interleaver*. Table I summarizes the parameters of ϕ^{o} and ϕ^{i} that will be used along this paper.

Throughout this paper we shall make the following assumptions on the constituent encoders:

Assumption 1. *The outer encoder $\phi^{\mathrm{o}} : (\mathbb{Z}_2^k)^{\mathbb{Z}_+} \to (\mathbb{Z}_2^r)^{\mathbb{Z}_+}$ is non-catastrophic, and its free distance $d_{\mathrm{f}}^{\mathrm{o}}$ is even and satisfies $d_{\mathrm{f}}^{\mathrm{o}} > 2$.*

Assumption 2. *The inner encoder $\phi^{\mathrm{i}} : (\mathbb{Z}_2^s)^{\mathbb{Z}_+} \to (\mathbb{Z}_2^l)^{\mathbb{Z}_+}$ is non-catastrophic and recursive, has scalar input (i.e., $s = 1$) and is proper rational (i.e., the matrix F of its minimal state space representation (1) is invertible).*

Among such assumptions, the ones which are truly needed in order to obtain the claimed asymptotic behavior of minimum distance and error probability are the following: non-catastrophicity of both encoders, $d_{\mathrm{f}}^{\mathrm{o}} > 2$ and recursiveness of ϕ^{i}. The other assumptions have been introduced for simplicity: they allow one to avoid cumbersome notation and definitions, to have simpler proofs, and to easily underline underline the role of $d_{\mathrm{e}}^{\mathrm{i}}$ (the effective free distance) as the main design parameter for the inner encoder. In Appendix II we shall briefly comment on which results can be obtained in the most general case, with a particular focus on the case of odd $d_{\mathrm{f}}^{\mathrm{o}}$, while we refer the interested reader to the first author's Ph.D. thesis [21] for further detail.

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 element uniformly distributed on the group of permutations of M_N symbols. This is the classical 'uniform interleaver' ensemble of [6], [5]. Since the interleaver Π_N is random, the minimum distance

$$d_N^{\min} := \min\{\mathrm{w_H}(\phi_N^{\mathrm{i}} \circ \pi_N \circ \phi_N^{\mathrm{o}}(u)) : u \neq 0)\}$$

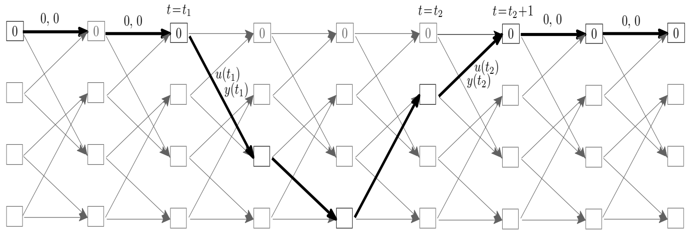

Fig. 3. An error event with active window $[t_1, t_2]$.

is a random variable itself. Similarly, assuming transmission over a 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 (see, e.g.,[5], [22]) has been on 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 behavior of $d_N^{\min}$ and the distribution of $P(e|\Pi_N)$, as N goes to infinity.

III. WEIGHT-ENUMERATING COEFFICIENTS OF THE CONSTITUENT ENCODERS

This sections deals with the input-output weight-enumerating coefficients of the constituent encoders. We define the error events and the weight-enumerating coefficients, we recall some properties of convolutional encoders related with the weight of codewords, and we state the bounds on the weight-enumerating coefficients of outer and inner encoder, which will be used in the following sections. The proofs of such bounds, many of which rely on variations of the arguments developed in [26], are deferred to Appendix I-A.

Consider a convolutional encoder $\phi \in (\mathbb{Z}_2^k)^{\mathbb{Z}_+} \to (\mathbb{Z}_2^r)^{\mathbb{Z}_+}$. We say that an input word $u \in (\mathbb{Z}_2^k)^{\mathbb{Z}_+}$ is an *error event* if there exist $t_1 < t_2$ such that u has support $\mathrm{supp}(u) \subseteq [t_1, t_2]$ and such that the corresponding state sequence x has support equal to the discrete interval $\mathrm{supp}(x) = [t_1 + 1, t_2]$. Notice that this implies that $u(t_1) \neq 0$ and that the corresponding codeword $y = \phi(u)$ has support $\mathrm{supp}(y) \subseteq [t_1, t_2]$. The *length* of the error event is defined as $t_2 - t_1 + 1$ and the discrete interval $[t_1, t_2]$ is called the *active window*. See Figure 3 for a pictorial representation.

Every finitely supported input sequence u such that $\phi(u)$ 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 [15, 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, $\phi_N : \mathbb{Z}_2^{kN} \to \mathbb{Z}_2^{r(N+\nu)}$. An error event for ϕ_N is any input word $(u^T(0), \ldots, u^T(N-1))$ such that

$$\big(u(0), u(1), \cdots, u(N-1), \tilde{u}(N), \cdots, \tilde{u}(N+\nu-1), 0, \dots\big)$$

is an error event for ϕ (where $\tilde{u}$ is the usual linear terminating extension of u). 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 word 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 $\phi^{\rm o} : (\mathbb{Z}_2^k)^{\mathbb{Z}_+} \to (\mathbb{Z}_2^r)^{\mathbb{Z}_+}$ and $\phi^{\rm i} : \mathbb{Z}_2^{\mathbb{Z}_+} \to (\mathbb{Z}_2^l)^{\mathbb{Z}_+}$ to be the outer and inner encoder of the turbo encoder described in the previous section (notice that we are considering $s = 1$). We shall denote by $\eta^{\rm o}$ and $\eta^{\rm i}$ the constants defined in Lemma 1 for $\phi^{\rm o}$ and $\phi^{\rm i}$ respectively.

For the outer encoder, we define the weight-enumerating coefficient $A_d^{{\rm o},N}$ to be the number of input words of $\phi_N^{\rm o}$ whose corresponding codewords have weight d. For it, we need only the following simple upper bound, which holds true for all non-catastrophic terminated convolutional encoders, and is mainly a restatement of [26, Lemma 3]. Its proof is provided in Appendix I-A1.

Lemma 2. *If $\phi^{\rm o}$ is non-catastrophic, then the following inequalities hold true.*

(a) *If $\lfloor d/d_{\rm f}^{\rm o} \rfloor < N/2$, then*

$$A_d^{{\rm o},N} \leq 2^{(k\eta^{\rm o}+\eta^{\rm o}+1)d+1} \binom{N}{\lfloor d/d_{\rm f}^{\rm o} \rfloor};$$

(b) *If $m_{\rm f}^{\rm o}$ denotes the number of different error events for $\phi^{\rm o}$ starting at $t_1 = 0$ and producing output weight $d_{\rm f}^{\rm o}$, then*

$$A_{d_{\rm f}^{\rm o}}^{{\rm o},N} \leq m_{\rm f}^{\rm o} N\,.$$

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

Because of the assumption of recursiveness, the inner encoder's output $\phi^{\rm i}(u)$ has infinite Hamming weight whenever the input word u has weight 1. In contrast, it is well known that there exists an input word of Hamming weight 2 which produces a codeword with finite weight (see e.g. [22, Proposition 3.6] for a proof). Having assumed that $\phi^{\rm i}$ has scalar input ($s = 1$), the codewords corresponding to weight-2 input words have the following useful property. Let $\delta^{\rm i}$ be the smallest possible relative distance between the positions of the non-zero entries of a weight-2 input word u such that $\phi^{\rm i}(u)$ has finite Hamming weight. Let $\bar{u}$ be the weight-2 input word with a one in position 0 and a one in position $\delta^{\rm i}$, and let $\bar{y} := \phi^{\rm i}(u)$ be the corresponding output word. Then, it is easy to see that, if u is a weight-2 input word, then $\phi^{\rm i}(u)$ has finite weight if and only if the positions of the two non-zero entries of u are at a distance multiple of $\delta^{\rm i}$, say $a\delta^{\rm i}$ for $a \geq 1$. Moreover, under the assumption that $\phi^{\rm i}$ is proper rational, such an output word is made of a consecutive disjoint copies of $\bar{y}$ and thus it has Hamming weight $a{\rm w_H}(\bar{y}) \geq {\rm w_H}(\bar{y})$. In particular, this means that ${\rm w_H}(\bar{y}) = d_{\rm e}^{\rm i}$. The case when the inner encoder has non-scalar input or is not proper rational is discussed in Appendix II.

Recursiveness of $\phi^{\rm i}$ ensures that any error event for $\phi^{\rm i}$ has input weight 2 or larger. When considering $\phi_N^{\rm i}$, 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 the input-weight limitation of error events imposed by recursiveness. Notice in particular that, for every even w, the input words contributing to $R_{w,\leq d,w/2}^{{\rm i},N}$ will exclusively be composed of regular error events each having input weight equal to 2.

For the weight-enumerating coefficients of $\phi_N^{\rm i}$, we have the two bounds stated below. The following lemma is proved in Appendix I-A2. While its part (b) follows from minor changes to the arguments in [26, Lemma 1], its part (a) is a key novel contribution, since it explicitly captures the dependence of the leading term on the inner encoder's effective free distance $d_{\rm e}^{\rm i}$. In fact, part (a) of the following lemma will turn out to be a fundamental ingredient in the next section, when showing the linear scaling of $d_N^{\min}$ in $d_{\rm e}^{\rm i}$. In contrast, the bound of [26, Lemma 1] depends on a term, therein denoted by $\Theta(1)$, which can be traced back to equal $4e\sqrt{\eta^{\rm i}}$, and cannot be chosen inversely proportional to $\sqrt{d_{\rm e}^{\rm i}}$: therefore, [26, Lemma 1] does not allow one to prove the linear scaling of $d_N^{\min}$ on $d_{\rm e}^{\rm i}$.

Lemma 3. *Let Assumption 2 be satisfied. Then, the following inequalities hold true.*

(a) *If w is even, then*

$$R_{w,\leq d,w/2}^{{\rm i},N} \leq \frac{(2e)^w}{w^w} M_N^{w/2} \left\lfloor \frac{d}{d_{\rm e}^{\rm i}} \right\rfloor^{w/2}.$$

(b) *If $d \leq M_N/(2\eta^{\rm i})$, then*

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

where C is a constant only depending on the inner convolutional encoder.

The following result is essentially a restatement of [26, Lemma 2], with the dependence on $d_{\rm e}^{\rm i}$ made explicit, and is proved in Appendix I-A3.

Lemma 4. *Let Assumption 2 be satisfied. If w is even and*

$$\frac{d_{\rm e}^{\rm i} w}{2} \leq d \leq \frac{d_{\rm e}^{\rm i} M_N}{2\delta^{\rm i}}\,,$$

then

$$R_{w,\leq d,w/2}^{{\rm i},N} \geq \frac{2^{w/2}}{w^w} M_N^{w/2} \left\lfloor \frac{d}{d_{\rm e}^{\rm i}} \right\rfloor^{w/2}.$$

IV. 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, if $d_{\mathrm{f}}^{\mathrm{o}}$ is even, then the minimum distance $d_N^{\min}$ scales as $d_e^{\mathrm{i}} N^{\beta}$ with high probability, where

$$\beta := 1 - \frac{2}{d_{\mathrm{f}}^{\mathrm{o}}} \in (0,1)\,.$$

First, we shall provide precise upper and lower bounds of the CDF of $d_N^{\min}$. These bounds, stated in Theorem 1, improve upon some of those in [26]. Then, we shall prove a deterministic upper bound on $d_N^{\min}$. Such a bound, stated in Theorem 2, generalizes and improves upon some of the results of [4]. As explained in the Introduction, 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_e^{i}, as a linear scaling parameter for $d_N^{\min}$.

We start by observing that a standard application of the union bound gives the useful bound (see [26, Lemma 6])

$$\mathbb{P}(d_N^{\min} \le d) \le \sum_{w=d_{\mathrm{f}}^{\mathrm{o}}}^{\eta^{\mathrm{i}} d} \binom{M_N}{w}^{-1} A_w^{\mathrm{o},N} A_{w,\le d}^{\mathrm{i},N}\,, \qquad \forall d \le K_N\,. \tag{4}$$

The limitation $w \le \eta^{\mathrm{i}} d$ is due to the remark that any terminating or regular error event of ϕ_N^{i} with output weight d has input weight w bounded from above by $s\eta^{\mathrm{i}} d$ (and here we are considering $s=1$).

Now, using the bounds on the weight-enumerating coefficients established in the previous section, we obtain the following result on minimum distances, which is a refinement of [26, Thm. 2.a].

Proposition 1. *Let Assumptions 1 and 2 be satisfied. Assume that $d = o(N^{\beta})$, as N grows large. Then, there exists $N_0 \ge 0$ such that*

$$\mathbb{P}(d_N^{\min} \le d) \le C \left(N^{-\beta} \frac{d}{d_{\mathrm{e}}^{\mathrm{i}}} \right)^{\frac{d_{\mathrm{f}}^{\mathrm{o}}}{2}}\,,$$

for all $N \ge N_0$, where $C := 2 m_{\mathrm{f}}^{\mathrm{o}} \left(2e/\sqrt{r}\right)^{d_{\mathrm{f}}^{\mathrm{o}}}$.

Proof: Define $\xi_N := \left(N^{-\beta} d / d_{\mathrm{e}}^{\mathrm{i}}\right)^{1/2}$, and observe that the assumption $d = o(N^{\beta})$ implies that

$$\xi_N = o(1)\,, \qquad \frac{d}{N} = o\left(\xi_N\right)\,, \tag{5}$$

as N grows large. Now consider (4), and split the summation therein in three parts:

$$\mathbb{P}(d_N^{\min} \le d) \le S_{d_{\mathrm{f}}^{\mathrm{o}}} + S_{\text{odd}} + S_{\text{even}}\,, \tag{6}$$

where

$$S_{d_{\mathrm{f}}^{\mathrm{o}}} := \binom{M_N}{d_{\mathrm{f}}^{\mathrm{o}}}^{-1} A_{d_{\mathrm{f}}^{\mathrm{o}}}^{\mathrm{o},N} A_{d_{\mathrm{f}}^{\mathrm{o}},\le d}^{\mathrm{i},N}\,,$$

$$S_{\text{odd}} := \sum_{\substack{d_{\mathrm{f}}^{\mathrm{o}} < w \le \eta^{\mathrm{i}} d \\ w \text{ odd}}} \binom{M_N}{w}^{-1} A_w^{\mathrm{o},N} A_{w,\le d}^{\mathrm{i},N}\,,$$

and S_{even} is defined similarly to S_{odd}, considering terms with even $w > d_{\mathrm{f}}^{\mathrm{o}}$. Then, in order to obtain bounds on the weight-enumerating coefficients, we use the upper bounds from Lemmas 2 and 3, as well as the simple bound

$$\binom{M_N}{w} \ge \frac{M_N^w}{w^w}\,.$$

We obtain that, for some suitable positive constants K_1, K_2, K_3, K_4 (depending on the constituent convolutional encoders only)

$$S_{d_{\mathrm{f}}^{\mathrm{o}}} \le \xi_N^{d_{\mathrm{f}}^{\mathrm{o}}} \left(\frac{C}{2} + K_1 \frac{d}{N} \right)\,; \tag{7}$$

$$\begin{aligned} S_{\text{odd}} &\le \sum_{\substack{d_{\mathrm{f}}^{\mathrm{o}} < w \le \eta^{\mathrm{i}} d \\ w \text{ odd}}} K_2^w N^{\lfloor w/d_{\mathrm{f}}^{\mathrm{o}} \rfloor - \lceil w/2 \rceil} d^{\lceil w/2 \rceil} \\ &= \left(\frac{d}{N}\right)^{1/2} \sum_{\substack{d_{\mathrm{f}}^{\mathrm{o}} < w \le \eta^{\mathrm{i}} d \\ w \text{ odd}}} K_2^w N^{\lfloor w/d_{\mathrm{f}}^{\mathrm{o}} \rfloor - w/2} d^{w/2} \\ &\le \left(\frac{d}{N}\right)^{1/2} \sum_{w=d_{\mathrm{f}}^{\mathrm{o}}+1}^{\infty} (\tilde{K}_2 \xi_N)^w\,, \end{aligned} \tag{8}$$

where $\tilde{K}_2 := K_2 \sqrt{d_{\mathrm{e}}^{\mathrm{i}}}$;

$$\begin{aligned} S_{\text{even}} &\le \sum_{\substack{d_{\mathrm{f}}^{\mathrm{o}} < w \le \eta^{\mathrm{i}} d \\ w \text{ even}}} K_3^w N^{\lfloor w/d_{\mathrm{f}}^{\mathrm{o}} \rfloor - \frac{w}{2}} d^{\frac{w}{2}} + \frac{d}{N} K_4^w N^{\lfloor w/d_{\mathrm{f}}^{\mathrm{o}} \rfloor - \frac{w}{2}} d^{\frac{w}{2}} \\ &\le \left(1 + \frac{d}{N}\right) \sum_{w=d_{\mathrm{f}}^{\mathrm{o}}+2}^{\infty} (K_5 \xi_N)^w\,, \end{aligned} \tag{9}$$

where $K_5 := \sqrt{d_{\mathrm{e}}^{\mathrm{i}}} \max\{K_3, K_4\}$. It follows from (5) that

$$\tilde{K}_2 \xi_N \le \frac{1}{2}\,, \qquad K_5 \xi_N \le \frac{1}{2}\,, \tag{10}$$

$$K_1 \frac{d}{N} \le \frac{1}{6} C\,, \qquad 2 \tilde{K}_2^{d_{\mathrm{f}}^{\mathrm{o}}+1} \xi_N \left(\frac{d}{N}\right)^{\frac{1}{2}} \le \frac{1}{6} C \tag{11}$$

$$\left(1 + \frac{d}{N}\right) 2 K_5^{d_{\mathrm{f}}^{\mathrm{o}}+2} \xi_N^2 \le \frac{1}{6} C\,, \tag{12}$$

for sufficiently large N. From (7) and (11), it follows that

$$S_{d_{\mathrm{f}}^{\mathrm{o}}} \le \xi_N^{d_{\mathrm{f}}^{\mathrm{o}}} \left(\frac{1}{2} C + K_1 \frac{d}{N} \right) \le \xi_N^{d_{\mathrm{f}}^{\mathrm{o}}} \left(\frac{1}{2} C + \frac{1}{6} C \right)\,. \tag{13}$$

Equation (10) implies that the series in right-hand sides of both (8) and (9) are convergent, and dominated by twice their first term. From this remark, together with (11) and (12), it follows that

$$S_{\text{odd}} \le \left(\frac{d}{N}\right)^{\frac{1}{2}} 2 \left(\tilde{K}_2 \xi_N\right)^{d_{\mathrm{f}}^{\mathrm{o}}+1} \le \frac{1}{6} C \xi_N^{d_{\mathrm{f}}^{\mathrm{o}}}\,, \tag{14}$$

$$S_{\text{even}} \le \left(1 + \frac{d}{N}\right) 2 \left(K_5 \xi_N\right)^{d_{\mathrm{f}}^{\mathrm{o}}+2} \le \frac{1}{6} C \xi_N^{d_{\mathrm{f}}^{\mathrm{o}}}\,. \tag{15}$$

The claim follows by combining (6), (13), (14), and (15). ■

It is possible to obtain also a lower bound for the CDF of the minimum distance, showing that, asymptotically in the block-length, 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 [26, Thm. 2.b] with the inclusion-exclusion principle [2, p. 124].

First of all, we fix an error event u^* for the outer convolutional encoder ϕ^{o}, having active window $[0, T-1]$ for some T, and with an output $c^* = \phi^{\mathrm{o}}(u^*)$ such that $\mathrm{w_H}(c^*) = d_{\mathrm{f}}^{\mathrm{o}}$. Note that $2 \leq T \leq d_{\mathrm{f}}^{\mathrm{o}}\eta^{\mathrm{o}}$. Consider $N > T$. For a nonnegative integer j, define c_j^* as the codeword obtained by shifting c^* for j trellis steps, so that the active window is $[j, T+j-1]$; clearly, if $|j_2 - j_1| \geq T$, then $c_{j_1}^*$ and $c_{j_2}^*$ have non-overlapping supports.

Now consider the terminated encoder ϕ_N^{o}, and, with a slight abuse of notation, let c_j^* denote its codewords corresponding to the above-constructed codewords of ϕ^{o}. Define the set of indices $J := \{d_{\mathrm{f}}^{\mathrm{o}}\eta^{\mathrm{o}}\, j\,,\ j \in \mathbb{Z}_+\} \cap \{0, 1, \ldots, N-1-d_{\mathrm{f}}^{\mathrm{o}}\eta^{\mathrm{o}}\}$, so that if j_1 and j_2 both belong to J, and $j_1 \neq j_2$, then clearly $|j_2 - j_1| \geq d_{\mathrm{f}}^{\mathrm{o}}\eta^{\mathrm{o}} \geq T$. For $j \in J$ and $d \in \mathbb{Z}_+$, define the event

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

Clearly, for any j, $E_j^*(d)$ implies $d_N^{\min} \leq d$, so that

$$\mathbb{P}(d_N^{\min} \leq d) \geq \mathbb{P}(\cup_{j\in J} E_j^*(d))\,.$$

The following lemma provides an expression for $\mathbb{P}\big(E_j^*(d)\big)$ and shows that, asymptotically, the events $E_j^*(d)$ are 'almost' pairwise independent. Its proof, deferred to Appendix I-B1 closely parallels the arguments of part of the proof of [26, Thm. 2.a]. The main difference with respect to [26, Thm. 2.a] is in the definition of the event $E_j^*(d)$, which in our case has the additional restriction that $\phi_N^{\mathrm{i}}(\Pi_N(c_j^*))$ has $d_{\mathrm{f}}^{\mathrm{o}}/2$ regular events. Our definition does not significantly modify the proof of this result, but turns out to be a key point in order to show the role of $d_{\mathrm{e}}^{\mathrm{i}}$ in Proposition 2.

Lemma 5. *Let Assumptions 1 and 2 be satisfied. Then, for all $j_1 \neq j_2 \in J$,*

$$\mathbb{P}(E_j^*(d)) = \binom{M_N}{d_{\mathrm{f}}^{\mathrm{o}}}^{-1} R^{\mathrm{i},N}_{d_{\mathrm{f}}^{\mathrm{o}}, \leq d, d_{\mathrm{f}}^{\mathrm{o}}/2}\,, \tag{16}$$

$$\mathbb{P}(E_{j_1}^*(d) \cap E_{j_2}^*(d)) \leq \frac{\binom{M_N}{d_{\mathrm{f}}^{\mathrm{o}}}}{\binom{M_N - d_{\mathrm{f}}^{\mathrm{o}}}{d_{\mathrm{f}}^{\mathrm{o}}}} \mathbb{P}(E_{j_1}^*(d))\mathbb{P}(E_{j_2}^*(d))\,.$$

We shall obtain our lower bound by considering the probability of the union event $\bigcup_j E_j^*(d)$ and using the inclusion-exclusion principle.

Proposition 2. *Let Assumptions 1 and 2 be satisfied. Assume that $d \geq \frac{1}{2}d_{\mathrm{f}}^{\mathrm{o}}d_{\mathrm{e}}^{\mathrm{i}}$, and $d = o(N^\beta)$, as N grows large. Then, there exists $N_0 \geq 0$ such that, for all $N \geq N_0$,*

$$\mathbb{P}(d_N^{\min} \leq d) \geq K\left(N^{-\beta}\frac{d}{d_{\mathrm{e}}^{\mathrm{i}}}\right)^{d_{\mathrm{f}}^{\mathrm{o}}/2}\,,$$

where $K := \frac{1}{4}(1 - 2/d_{\mathrm{f}}^{\mathrm{o}})^{d_{\mathrm{f}}^{\mathrm{o}}/2} / \left(r^{d_{\mathrm{f}}^{\mathrm{o}}/2} e^{d_{\mathrm{f}}^{\mathrm{o}}} d_{\mathrm{f}}^{\mathrm{o}}\eta^{\mathrm{o}}\right)$.

Proof: Let us define $\xi_N := \left(N^{-\beta} d/d_{\mathrm{e}}^{\mathrm{i}}\right)^{1/2}$, and

$$\Gamma_1 := \sum_{j\in J}\mathbb{P}(E_j^*(d))\,,\ \Gamma_2 := \frac{1}{2}\sum_{\substack{j_1,j_2\in J\\ j_1\neq j_2}}\mathbb{P}(E_{j_1}^*(d)\cap E_{j_2}^*(d))\,.$$

Then, using the inclusion-exclusion principle we obtain

$$\mathbb{P}(d_N^{\min} \leq d) \geq \mathbb{P}\left(\bigcup_{j\in J} E_j^*(d)\right) \geq \Gamma_1 - \Gamma_2\,. \tag{17}$$

We give a lower bound for the first summation using Lemma 5, Lemma 4, and (26). Also, recall that $|J| = \lfloor N/(d_{\mathrm{f}}^{\mathrm{o}}\eta^{\mathrm{o}})\rfloor$. We get

$$\begin{aligned}\Gamma_1 &= |J|\, R^{\mathrm{i},N}_{d_{\mathrm{f}}^{\mathrm{o}}, \leq d, d_{\mathrm{f}}^{\mathrm{o}}/2}\binom{M_N}{d_{\mathrm{f}}^{\mathrm{o}}}^{-1}\\ &\geq \left\lfloor\frac{N}{d_{\mathrm{f}}^{\mathrm{o}}\eta^{\mathrm{o}}}\right\rfloor \frac{2^{d_{\mathrm{f}}^{\mathrm{o}}/2}}{e^{d_{\mathrm{f}}^{\mathrm{o}}}} M_N^{-d_{\mathrm{f}}^{\mathrm{o}}/2}\left\lfloor\frac{d}{d_{\mathrm{e}}^{\mathrm{i}}}\right\rfloor^{d_{\mathrm{f}}^{\mathrm{o}}/2}\\ &\geq 2K\xi_N^{d_{\mathrm{f}}^{\mathrm{o}}}\,,\end{aligned} \tag{18}$$

with the last inequality following from the fact that

$$\left\lfloor\frac{d}{d_{\mathrm{e}}^{\mathrm{i}}}\right\rfloor \geq \frac{d}{d_{\mathrm{e}}^{\mathrm{i}}}\left(1 - \frac{d_{\mathrm{e}}^{\mathrm{i}}}{d}\right) \geq \frac{d}{d_{\mathrm{e}}^{\mathrm{i}}}\left(1 - \frac{2}{d_{\mathrm{f}}^{\mathrm{o}}}\right)\,,$$

thanks to the assumption $d \geq \frac{1}{2}d_{\mathrm{f}}^{\mathrm{o}}d_{\mathrm{e}}^{\mathrm{i}}$, and from the inequalities

$$M_N \leq 2rN\,, \qquad \left\lfloor\frac{N}{d_{\mathrm{f}}^{\mathrm{o}}\eta^{\mathrm{o}}}\right\rfloor \geq \frac{N}{2d_{\mathrm{f}}^{\mathrm{o}}\eta^{\mathrm{o}}}\,,$$

which hold true for sufficiently large N.

Now, we find an upper bound for the second summation in (17) using Lemma 5, Lemma 3, and (26), as follows

$$\Gamma_2 \leq \binom{|J|}{2}\frac{\binom{M_N}{d_{\mathrm{f}}^{\mathrm{o}}}}{\binom{M_N - d_{\mathrm{f}}^{\mathrm{o}}}{d_{\mathrm{f}}^{\mathrm{o}}}}\left(R^{\mathrm{i},N}_{d_{\mathrm{f}}^{\mathrm{o}}, \leq d, d_{\mathrm{f}}^{\mathrm{o}}/2}\binom{M_N}{d_{\mathrm{f}}^{\mathrm{o}}}^{-1}\right)^2 \leq \overline{\Gamma}_2\,,$$

where

$$\overline{\Gamma}_2 := \frac{1}{2}\left(\frac{N}{d_{\mathrm{f}}^{\mathrm{o}}\eta^{\mathrm{o}}}\right)^2\frac{\binom{M_N}{d_{\mathrm{f}}^{\mathrm{o}}}}{\binom{M_N - d_{\mathrm{f}}^{\mathrm{o}}}{d_{\mathrm{f}}^{\mathrm{o}}}}(2e)^{2d_{\mathrm{f}}^{\mathrm{o}}} M_N^{-d_{\mathrm{f}}^{\mathrm{o}}}\left\lfloor\frac{d}{d_{\mathrm{e}}^{\mathrm{i}}}\right\rfloor^{d_{\mathrm{f}}^{\mathrm{o}}}\,.$$

Notice that

$$M_N = rN(1 + o(1))\,, \quad \binom{M_N}{d_{\mathrm{f}}^{\mathrm{o}}}\binom{M_N - d_{\mathrm{f}}^{\mathrm{o}}}{d_{\mathrm{f}}^{\mathrm{o}}}^{-1} = 1 + o(1)\,,$$

as N grows large, so that

$$\overline{\Gamma}_2 \leq \frac{(4e^2)^{d_{\mathrm{f}}^{\mathrm{o}}}}{2r^{d_{\mathrm{f}}^{\mathrm{o}}}(d_{\mathrm{f}}^{\mathrm{o}}\eta^{\mathrm{o}})^2}(1 + o(1))\,\xi_N^{2d_{\mathrm{f}}^{\mathrm{o}}}\,.$$

Since $d = o(N^\beta)$ by assumption, one has that $\xi_N = o(1)$, so that

$$\Gamma_2 \leq \overline{\Gamma}_2 \leq K\xi_N^{d_{\mathrm{f}}^{\mathrm{o}}}\,,$$

for sufficiently large N. Together with (17) and (18), the foregoing implies the claim. ■

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

Theorem 1. *Let Assumptions 1 and 2 be satisfied. Then, for every positive sequence $\{\varepsilon_N\}$ such that $\lim_{N\to\infty}\varepsilon_N = 0$, there exists a finite $N_0 \geq 0$ such that*

$$C_0^{\mathrm{o}}\varepsilon_N^{d_{\mathrm{f}}^{\mathrm{o}}/2} \leq \mathbb{P}\left(d_N^{\min} \leq d_{\mathrm{e}}^{\mathrm{i}}N^\beta\varepsilon_N\right) \leq C_1^{\mathrm{o}}\varepsilon_N^{d_{\mathrm{f}}^{\mathrm{o}}/2}\,,$$

for all $N \geq N_0$, where C_0^{o} and C_1^{o} are positive constants depending on the outer encoder only.

Theorem 1 provides some 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 the minimum distance of the typical serial turbo code grows as a positive power of the block-length. In fact, it implies that the probability that the minimum distance $d_N^{\min}$ grows any slower than N^β vanishes as N grows large. The exponent of this power law growth, β, depends only on the free distance of the outer encoder, $d_{\rm f}^{\rm o}$, in an increasing way. This is in line with the results of [26]. On the other hand, it shows that the minimum distance of the typical turbo code scales linearly in the effective free distance of the inner encoder, $d_{\rm e}^{\rm i}$. While the effect of $d_{\rm e}^{\rm i}$ on the average error probability of serial turbo codes has been studied in [5], [22], up to our knowledge no results have previously appeared in the literature relating $d_{\rm e}^{\rm i}$ to the minimum distance. Such a scaling effect of $d_{\rm e}^{\rm i}$ on $d_N^{\min}$ is particularly relevant for moderate block-lengths.

The result stated below provides a deterministic upper bound on the minimum distance $d_N^{\min}$, showing an analogous dependence on the parameters $d_{\rm f}^{\rm o}$ and $d_{\rm e}^{\rm i}$.

Theorem 2. *Let Assumptions 1 and 2 be satisfied. Then, for all*

$$N \geq 2^{2/d_{\rm f}^{\rm o}}\, 8 d_{\rm f}^{\rm o} \eta^{\rm o} (\delta^{\rm i})^{d_{\rm f}^{\rm o}} ,$$

and for every realization π_N of the interleaver Π_N, the minimum distance satisfies

$$d_N^{\min} \leq 6 r d_{\rm f}^{\rm o} (8 d_{\rm f}^{\rm o} \eta^{\rm o})^{2/d_{\rm f}^{\rm o}} (\delta^{\rm i})^2 d_{\rm e}^{\rm i}\ N^\beta \log N . \qquad (19)$$

It is worth comparing the upper bound (19) with the high probability scaling $N^\beta d_{\rm e}^{\rm i}$ implied by Theorem 1. On the one hand, the dependence on N of the right-hand side of (19) involves an additional factor $\log N$. On the other hand, the right-hand side of (19) shows a linear dependence on $d_{\rm e}^{\rm i}$, though multiplied by a factor $(\delta^{\rm i})^2$, which depends itself on the inner encoder, and is therefore related to $d_{\rm e}^{\rm i}$ itself. It is important to highlight the fact that, in contrast to Theorem 1, Theorem 2 holds for every choice of the interleaver, and not only with high probability with respect to its random choice. In fact, it may be conjectured that such greater strength of the statement could be the main reason for the additional factors in the upper bound (19).

Theorem 2, whose proof is deferred to Appendix I-B2, may be thought of as a generalization of [4, 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. Moreover, our modification of [4, Thm. 2] unveils the fundamental role played by the inner encoder's parameters $d_{\rm e}^{\rm i}$ and $\delta^{\rm i}$.

Indeed, [4] considers serial turbo codes as well, in an even more general setting with growing memory, but the result they obtain ([4, Thm. 3]), when specialized to the constant-memory case, gives a bound which is asymptotically weaker than Theorem 2. In fact, [4, Thm. 3] gives

$$d_N^{\min} \leq C N^{1-(r(\mu^{\rm o}+2))^{-1}}$$

for some positive constant C, and where $\mu^{\rm o}$ is the dimension of the state space of the outer encoder. It is easy to show that $d_{\rm f}^{\rm o} \leq r(\mu^{\rm o}+1)$ and thus that

$$\beta < 1 - (r(\mu^{\rm o}+2))^{-1} .$$

In fact, we can always construct a non-zero outer codeword of weight at most $r(\mu^{\rm o}+1)$, as follows. Take a non-zero input at time zero, and then drive the state back to zero by applying the termination procedure: the corresponding codeword is supported in $[0, \nu^{\rm o}] \subseteq [0, \mu^{\rm o}]$ and thus has weight at most $r(\mu^{\rm o}+1)$.

The result we obtain in Theorem 2 is also asymptotically tighter than the currently best known bound for serial turbo codes, presented in [32], which, as N grows large, grows as fast as $C\, N^{1-1/d_{\rm f}^{\rm o}}$.

V. 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-input additive white Gaussian noise channel: when $\omega \in \{0,1\}$ is transmitted, the output of the channel is $(-1)^\omega L + \Omega$, where $L \in (0, +\infty)$ and Ω is an independent Gaussian random variable $\Omega \sim \mathcal{N}(0, \sigma^2)$. The signal-to-noise ratio is

$$\rho := L^2/(2\sigma^2) .$$

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 [5], [22] that

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

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

However, the error probability of the average serial turbo code turns out to be much larger than that of the typical serial turbo code. Indeed, the former is dominated by an asymptotically negligible fraction of poorly performing codes. In the sequel, we shall use so-called expurgation techniques in order to show that the error probability of the typical serial turbo code decays faster than $\exp(-N^{\beta-\varepsilon})$, for all $\varepsilon > 0$.

We define, for every $N \geq 1$ and $\varepsilon > 0$, the event $E_N^\varepsilon := \{d_N^{\min} > N^{\beta-\varepsilon}\}$. It follows from Theorem 1 that

$$\mathbb{P}(E_N^\varepsilon) \geq 1 - C_1 N^{-\varepsilon d_{\rm f}^{\rm o}/2} . \qquad (20)$$

The following proposition gives an upper bound on the average word error probability of the serial turbo ensemble, conditioned on the event E_N^ε.

Proposition 3. *Let Assumptions 1 and 2 be satisfied. Then, there exists some finite $\rho_0 \geq 0$ such that, if the signal-to-noise ratio ρ satisfies $\rho \geq \rho_0$, then, for all $\varepsilon \in (0, \beta)$ there exist some finite constants $N_0 \geq 0$ and $C > 0$ such that*

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

for all $N \geq N_0$.

Proof: The main tool for this proof is the classical union-Bhattacharyya bound, introduced for the average error probability in serial ensembles in [5]. Here we use a modified version of it, where we consider the ensemble expurgated from the codes with low minimum distance

$$\mathbb{E}[P(e|\Pi_N)|E_N^\varepsilon] \leq \frac{1}{\mathbb{P}(E_N^\varepsilon)} \sum_{h=N^{\beta-\varepsilon}}^{K_N} \sum_{w=d_{\rm f}^{\rm o}}^{\eta^{\rm i}h} \frac{A_w^{{\rm o},N} A_{w,h}^{{\rm i},N}}{\binom{M_N}{w}} \gamma^h \,, \tag{21}$$

where $\gamma = \exp(-\rho)$.

To prove this bound, first notice that

$$\mathbb{E}[P(e|\Pi_N)|E_N^\varepsilon] = \frac{\mathbb{E}[\chi P(e|\Pi_N)]}{\mathbb{P}(E_N^\varepsilon)} \,,$$

where χ denotes the indicator function of the event E_N^ε. The union-Bhattacharyya bound (see e.g. [5] or [24]) gives

$$P(e|\Pi_N) \leq \sum_{h=1}^{K_N} A_h^{\text{serial},\Pi_N} \gamma^h \,,$$

where by $A_h^{\text{serial},\Pi_N}$ we denote the number of codewords with weight h of the serial code obtained from the given ensemble when the interleaver Π_N is sampled. Then (21) is obtained as follows

$$\begin{aligned}\mathbb{E}[P(e|\Pi_N)\,\chi] &\leq \mathbb{E}[\sum_{h=1}^{K_N} A_h^{\text{serial},\Pi_N}\,\chi\gamma^h] \\ &\leq \sum_{h=N^{\beta-\varepsilon}}^{K_N} \mathbb{E}[A_h^{\text{serial},\Pi_N}]\gamma^h \\ &= \sum_{h=N^{\beta-\varepsilon}}^{K_N} \sum_{w=d_{\rm f}^{\rm o}}^{\eta^{\rm i}h} A_w^{{\rm o},N} A_{w,h}^{{\rm i},N} \binom{M_N}{w}^{-1} \gamma^h \,,\end{aligned}$$

where the last equality is obtained by applying the expression [24, Eq. (7.1)]. The limitations $d_{\rm f}^{\rm o} \leq w \leq \eta^{\rm i}h$ come from the fact that, by definition of $d_{\rm f}^{\rm o}$ and by Lemma 1, if these inequalities are not satisfied then $A_w^{{\rm o},N} A_{w,h}^{{\rm i},N} = 0$.

By Theorem 1, $\mathbb{P}(E_N^\varepsilon)$ approaches 1, as N grows large. So, for some $c > 0$, $\mathbb{P}(E_N^\varepsilon) \geq c$, for large enough N. Now we need bounds for the weight-enumerating coefficients of the constituent encoders.

We start by considering the terms with $h \leq N/(2\eta^{\rm i})$. For the outer encoder, having $w \leq \eta^{\rm i} d \leq N/2$, we can apply Lemma 2 to find a bound for $A_w^{{\rm o},N}$. For the inner encoder we use the simple bound $A_{w,h}^{{\rm i},N} \leq A_{w,\leq h}^{{\rm i},N}$ and then, thanks to the inequality $d \leq N/(2\eta^{\rm i}) \leq K_N/(2\eta^{\rm i})$, we can apply Lemma 3. Hence, we can find a positive C_1 such that

$$\sum_{h=N^{\beta-\varepsilon}}^{N/(2\eta^{\rm i})} \sum_{w=d_{\rm f}^{\rm o}}^{\eta^{\rm i}h} \frac{A_w^{{\rm o},N} A_{w,h}^{{\rm i},N}}{\binom{M_N}{w}} \gamma^h \leq \sum_{h=N^{\beta-\varepsilon}}^{N/(2\eta^{\rm i})} \sum_{w=d_{\rm f}^{\rm o}}^{\eta^{\rm i}h} C_1^w \left(\frac{h}{w}\right)^{\frac{w}{2}} \left(\frac{w}{N}\right)^{\frac{w}{2}-\frac{w}{d_{\rm f}^{\rm o}}} \gamma^h.$$

Then, observe that the function $g(z) := (a/z)^z$ has maximum value $g(a/e) = e^{a/e}$, so that

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

Moreover, $w \leq \tilde{c}N$ for some $\tilde{c} \geq 1$, so

$$(w/N)^{\frac{w}{2}-\frac{w}{d_{\rm f}^{\rm o}}} \leq \tilde{c}^{(\frac{1}{2}-\frac{1}{d_{\rm f}^{\rm o}})w} \,.$$

Hence, as $w \leq \eta^{\rm i}h$, we can find a constant $C_2 \geq 1$ such that

$$\sum_{h=N^{\beta-\varepsilon}}^{N/(2\eta^{\rm i})} \sum_{w=d_{\rm f}^{\rm o}}^{\eta^{\rm i}h} \frac{A_w^{{\rm o},N} A_{w,h}^{{\rm i},N}}{\binom{M_N}{w}} \gamma^h \leq \sum_{h=N^{\beta-\varepsilon}}^{N/(2\eta^{\rm i})} (C_2\gamma)^h \,.$$

For the remaining terms, having $N/(2\eta^{\rm i}) < h \leq K_N$, we use the following trivial upper bounds on the weight-enumerating coefficients

$$A_w^{{\rm o},N} \leq \binom{M_N}{w} \quad \text{and} \quad A_{w,h}^{{\rm i},N} \leq \binom{K_N}{h} \,,$$

from which we have

$$\sum_{h=N/(2\eta^{\rm i})}^{K_N} \sum_{w=d_{\rm f}^{\rm o}}^{\eta^{\rm i}h} \frac{A_w^{{\rm o},N} A_{w,h}^{{\rm i},N}}{\binom{M_N}{w}} \gamma^h \leq \sum_{h=N/(2\eta^{\rm i})}^{K_N} \eta^{\rm i}h \binom{K_N}{h} \,.$$

Now notice that, under the assumption $N/(2\eta^{\rm i}) < h \leq K_N$, one has

$$\binom{K_N}{h} \leq \left(\frac{eK_N}{h}\right)^h \leq C_3^h$$

for some positive constant C_3 which depends only on $r, l, \nu^{\rm o}, \nu^{\rm i}, \eta^{\rm i}$. Finally, putting all terms together, we have proved that there exists some constant $C_4 \geq 1$ such that

$$\mathbb{E}[P(e|\Pi_N)|E_N^\varepsilon] \leq \sum_{h=N^{\beta-\varepsilon}}^{K_N} (C_4\gamma)^h \leq \sum_{h=N^{\beta-\varepsilon}}^{\infty} (C_4\gamma)^h \,.$$

Assuming that $\gamma < 1/C_4$, the series is convergent, and equal to $(C_4\gamma)^{N^{\beta-\varepsilon}}/(1-C_4\gamma)$. As we don't aim at optimizing constants, we can further assume that $\gamma \leq 1/(C_4 e^2)$, so that the claim easily follows with $C = C_4/(1-e^{-2})$. ■

It is worth pointing out that the constant C in Proposition 3 is independent from the signal to noise ratio ρ, provided that this is large enough.

From Proposition 3 and 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. *Let Assumptions 1 and 2 be satisfied. Then, there exists some finite $\rho_0 \geq 0$ such that, if the signal-to-noise ratio ρ satisfies $\rho \geq \rho_0$, then for all $\varepsilon \in (0, \beta)$ there exist some finite $N_0 \geq 0$ and $C > 0$ such that*

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

for all $N \geq N_0$.

Proof: By applying Markov's inequality to the random variable $P(e|\Pi_N)$ conditioned on the event E_N^ε, one gets

$$\mathbb{P}\Big(P(e|\Pi_N) \geq a\,\mathbb{E}\left[P(e|\Pi_N)\Big|E_N^\varepsilon\right]\Big|\,E_N^\varepsilon\Big) \leq \frac{1}{a} \,, \quad \forall a > 0 \,. \tag{22}$$

Now, consider the event

$$F_N^\varepsilon := \{P(e|\Pi_N) \geq \exp(-N^{\beta-\varepsilon})\} \,.$$

From Proposition 3 and inequality (22) with $a = C_2^{-1}\exp(N^{\beta-\varepsilon})$, one gets that

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

Let us denote the complement of the event E_N^ε by $\overline{E_N^\varepsilon}$. Then, it follows from (20) that

$$\begin{aligned}\mathbb{P}\left(F_N^\varepsilon\right) &= \mathbb{P}\left(F_N^\varepsilon \cap \overline{E_N^\varepsilon}\right) + \mathbb{P}\left(F_N^\varepsilon \cap E_N^\varepsilon\right)\\ &\leq 1-\mathbb{P}(E_N^\varepsilon) + \mathbb{P}\left(F_N^\varepsilon | E_N^\varepsilon\right)\mathbb{P}(E_N^\varepsilon)\\ &\leq C_1 N^{-\varepsilon d_{\mathrm{f}}^{\mathrm{o}}/2} + C_2\exp(-N^{\beta-\varepsilon})\\ &\leq CN^{-\varepsilon d_{\mathrm{f}}^{\mathrm{o}}/2}\,,\end{aligned} \tag{23}$$

where the last inequality holds with $C := C_1 + C_2$, for sufficiently large N.

On the other hand, using the inequality

$$P(e|\Pi_N) \geq p^{d_N^{\min}}\,, \tag{24}$$

where $p = \mathrm{erfc}(\sqrt{\rho})/2$ is the bit error probability of uncoded transmission (see e.g. [15] for a proof), and using Theorem 2, one gets that

$$P(e|\Pi_N) \geq \exp(-N^{\beta+o(1)})\,, \tag{25}$$

for every realization of the random interleaver Π_N. Then, the claim is an immediate consequence of (23) and (25). ■

We conclude this section by observing that both Theorems 1 and 3 only imply weak probabilistic convergence results, since the CDFs of $d_N^{\min}$ and $P(e|\Pi_N)$ decrease slowly in N. Indeed, one may prove [11] that, while converging in distribution to β, both the growth rate of the minimum distance, i.e.,

$$X_N := (\log N)^{-1}\log d_N^{\min}\,,$$

and the decay rate of the error probability, i.e.,

$$Y_N := (\log N)^{-1}\log(-\log(P(e|\Pi_N)))\,,$$

densely cover the interval $[\alpha,\beta]$ with probability one, where $\alpha = 1 - 2/\lceil d_{\mathrm{f}}^{\mathrm{o}}/2\rceil$.

VI. Conclusion

In this paper we have studied the behavior of the minimum distance and ML error probability of serial turbo codes with uniform 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 minimum distance's CDF, and of a deterministic upper bound. As a consequence, we have characterized the decay rate of the ML error probability of the typical turbo code, which turns out to be exponential in some positive power of the block-length.

This contrasts the asymptotic behavior of the ML error probability of the average serial turbo code, which is known to decay only as a negative power of the block-length. In spite of such lack of concentration of the typical code performance around the average code performance, our results confirm the centrality of the two main design parameters for serial turbo codes suggested by the average-code analysis, namely the free distance of the outer encoder, and the effective free distance of the inner encoder.

In the results that we have presented, we have considered the assumptions that the constituent convolutional encoders are non-catastrophic, that the outer encoder's free distance is even and greater than 2, and that the inner encoder is recursive, proper rational and with scalar input. As discussed in Appendix II, only some of these assumptions are indeed essential in order to obtain the claimed asymptotic scaling of the typical minimum distance and ML error probability (non-catastrophicity of both encoders, outer encoder's free distance greater than 2, inner encoder's recursiveness), while the other assumptions were introduced for simplicity.

Appendix I
Proofs

In the present appendix, we provide the proofs of some of the statements of Sections III and IV. Throughout, we shall make repeated use of the following well-known combinatorial bounds. For positive integers $m \leq n$, one has

$$\frac{n^m}{m^m} \leq \binom{n}{m} \leq \frac{(en)^m}{m^m}\,, \tag{26}$$

$$\binom{n}{m} \leq 2^n < e^n\,. \tag{27}$$

For reals $w \geq t \geq 0$, one has

$$t^t(w-t)^{w-t} \geq (w/2)^w \quad \text{for all } t\in[0,w]\,, \tag{28}$$

while, for $t > 1$,

$$\frac{1}{(t-1)^{(t-1)}} \leq \frac{e\,t}{t^t}\,. \tag{29}$$

Throughout this section, whenever we find it useful, we will write input and output words of the terminated encoders (finite strings of bits) as polynomials in the indeterminate D with binary coefficients, where the powers of D will simply be place-holders, indicating the position where the bits occur. This is a very common notation for convolutional encoders, where the powers of D denote the number of trellis steps and the coefficients are vectors of a suitable number of bits, but here we will rather use it for the terminated encoders, and powers of D will count the number of bits, not of vector labels (this distinction is important for the outer codewords in the proof of Theorem 2, while for the input words of the inner encoder the assumption $s = 1$ implies a one-to-one correspondence between bits and trellis steps).

A. Proofs of the results presented in Section III

Our proof techniques are based on ideas from [26]. We retrace here the proofs in all detail, both since [26] has not appeared yet, and in order to be able to underline the role of $d_{\mathrm{e}}^{\mathrm{i}}$.

1) Proof of Lemma 2: This is essentially a restatement of [26, Lemma 3]. We start by introducing some notation:

- Let $R_d^{\mathrm{o},N}$ and $T_d^{\mathrm{o},N}$ denote, respectively, the number of input words to ϕ_N^{o} having output weight d and consisting exclusively of regular error events, or containing a terminating error event. We thus have

$$A_d^{\mathrm{o},N} = R_d^{\mathrm{o},N} + T_d^{\mathrm{o},N}\,.$$

- Let $R^{\mathrm{o},N}_{(d_1,\dots,d_n)}$ be the number of input words to ϕ^{o}_N consisting of n regular error events whose output weights are $d_1,\dots,d_n$, respectively. Similarly, let $T^{\mathrm{o},N}_{(d_1,\dots,d_n)}$ be the number of input words to ϕ^{o}_N consisting of $n-1$ regular error events having output weights, in order, $d_1,\dots,d_{n-1}$, and a final terminating one of weight d_n.

Assume that $d_1+\cdots+d_n=d$. Then, one has that

$$R^{\mathrm{o},N}_{(d_1,\dots,d_n)} \le 2^{kd\eta^{\mathrm{o}}}\binom{N}{n}.$$

Indeed, we are considering n error events, with lengths at most $d_1\eta^{\mathrm{o}}$, ..., $d_n\eta^{\mathrm{o}}$ respectively, so that the sum of their lengths is bounded by $d\eta^{\mathrm{o}}$. Thus, the number of distinct choices for the bits in the input word inside the active windows of such error events are at most $2^{kd\eta^{\mathrm{o}}}$. The only remaining freedom is in the choice of the starting points of the error events, and the number of possibilities is clearly bounded by $\binom{N}{n}$.

Hence, one has

$$\begin{aligned} R^{\mathrm{o},N}_{d} &= \sum_{n=1}^{\lfloor d/d^{\mathrm{o}}_{\mathrm{f}}\rfloor}\sum_{\substack{d_1,\dots,d_n:\\ \sum_i d_i=d, d_i\ge 1}} R^{\mathrm{o},N}_{(d_1,\dots,d_n)} \\ &\le \sum_{n=1}^{d}\binom{d}{n}2^{kd\eta^{\mathrm{o}}}\binom{N}{\lfloor d/d^{\mathrm{o}}_{\mathrm{f}}\rfloor} \\ &\le 2^{(1+k\eta^{\mathrm{o}})d}\binom{N}{\lfloor d/d^{\mathrm{o}}_{\mathrm{f}}\rfloor}, \end{aligned} \tag{30}$$

where we are using assumption that $\lfloor d/d^{\mathrm{o}}_{\mathrm{f}}\rfloor \le N/2$. Similarly,

$$T^{\mathrm{o},N}_{(d_1,\dots,d_n)} \le 2^{kd\eta^{\mathrm{o}}}\binom{N}{n-1}d\eta^{\mathrm{o}}$$

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

$$\begin{aligned} T^{\mathrm{o},N}_{d} &= \sum_{n=1}^{\lceil d/d^{\mathrm{o}}_{\mathrm{f}}\rceil}\sum_{\substack{d_1,\dots,d_n:\\ \sum_i d_i=d, d_i\ge 1}} T^{\mathrm{o},N}_{(d_1,\dots,d_n)} \\ &\le \sum_{n=1}^{d}\binom{d}{n}2^{kd\eta^{\mathrm{o}}}\binom{N}{\lceil d/d^{\mathrm{o}}_{\mathrm{f}}\rceil-1}d\eta^{\mathrm{o}} \\ &\le 2^{(1+k\eta^{\mathrm{o}}+\eta^{\mathrm{o}})d}\binom{N}{\lfloor d/d^{\mathrm{o}}_{\mathrm{f}}\rfloor}. \end{aligned} \tag{31}$$

Summing up (30) and (31) we get statement (a) of Lemma 2. The tighter bound of statement (b) of Lemma 2 is easily obtained from the observation that input words with output weight $d^{\mathrm{o}}_{\mathrm{f}}$ necessarily consist of just one error event starting in the interval $[0,N-1]$. ■

2) Proof of Lemma 3: Our arguments parallel those of [26, Lemma 1]. The main novelty consists in proving separate bounds for the leading term (statement (a)), and the other ones (statement (b)). While the proof of part (b) is essentially the same as the one of [26, Lemma 1], with different handling of some of the constants involved, the proof of part (a) is novel, and fundamental in showing the correct scaling in $d^{\mathrm{i}}_{\mathrm{e}}$.

Similarly to what we have done before, we need to introduce several auxiliary weight-enumerating coefficients for ϕ^{i}:

- let $R^{\mathrm{i},N}_{w,\le d}$ (respectively, $T^{\mathrm{i},N}_{w,\le d}$) denote the number of input words for ϕ^{i}_N having input weight w and output weight not larger than d, and consisting exclusively of regular error events (resp., containing a terminating error event);
- let $R^{\mathrm{i},N}_{w,\le d,n}$ (respectively, $T^{\mathrm{i},N}_{w,\le d,n}$) denote the number of input words for ϕ^{i}_N 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 $\boldsymbol{w}=(w_1,\dots,w_n)$ and $\boldsymbol{b}=(b_1,\dots,b_n)$ with $w_i>0$ and $b_i\in[0,N-1]$. Let $R^{\mathrm{i},N}_{\boldsymbol{w},\boldsymbol{b},\le d,n}$ (respectively, $T^{\mathrm{i},N}_{\boldsymbol{w},\boldsymbol{b},\le d,n}$) denote the number of weight-w input words to ϕ^{i}_N 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 $1\le j\le n$ the j-the error event starts in position b_j and has input weight w_j.

In order to prove statement (a), we notice that, 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 $R^{\mathrm{i},N}_{w,\le d,w/2}$ can be written as

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

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

Notice that

$$\begin{aligned} d^{\mathrm{i}}_{\mathrm{e}}\sum_{t=1}^{w/2}a_t &\le \textstyle\sum_{t=1}^{w/2}\mathrm{w_H}\left(\phi^{\mathrm{i}}\left(1+D^{\delta^{\mathrm{i}}a_t}\right)\right) \\ &= \mathrm{w_H}\left(\phi^{\mathrm{i}}\left(\textstyle\sum_{t=1}^{w/2}D^{b_t}(1+D^{\delta^{\mathrm{i}}a_t})\right)\right). \end{aligned}$$

The restriction $\mathrm{w_H}\left(\phi^{\mathrm{i}}(u)\right)\le d$ thus implies

$$d^{\mathrm{i}}_{\mathrm{e}}\sum_{1\le t\le w/2}a_t\le d, \tag{32}$$

and there are $\binom{\lfloor d/d^{\mathrm{i}}_{\mathrm{e}}\rfloor}{w/2}$ choices for positive integers $a_1,\dots,a_{w/2}$ satisfying (32). Finally, there are at most $\binom{M_N}{\frac{w}{2}}$ choices for the starting positions $b_1,\dots,b_{w/2}$ of the error events. Summing up, and using (26), we obtain

$$R^{\mathrm{i},N}_{w,\le d,w/2}\le\binom{\lfloor d/d^{\mathrm{i}}_{\mathrm{e}}\rfloor}{\frac{w}{2}}\binom{M_N}{\frac{w}{2}}\le\left(\frac{2e}{w}\right)^{w}M_N^{w/2}\left\lfloor\frac{d}{d^{\mathrm{i}}_{\mathrm{e}}}\right\rfloor^{w/2}.$$

This yields statement (a) of Lemma 3.

In order to prove statement (b) of Lemma 3, we start by considering the case when w is even. We first show that

$$R^{\mathrm{i},N}_{\boldsymbol{w},\boldsymbol{b},\le d,n}\le\binom{d\eta^{\mathrm{i}}}{w-n}. \tag{33}$$

Notice indeed that $R^{\mathrm{i},N}_{\boldsymbol{w},\boldsymbol{b},\le d,n}$ is smaller than the number of binary words of length $d\eta^{\mathrm{i}}$ 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 $w_1,\dots,w_n$, fixed starting points $b_1,\dots,b_n$, and total output weight $\le d$, map it into a word of length $d\eta^{\mathrm{i}}$ 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 $< d\eta^{\mathrm{i}}$, then add dummy zeros at the end to get a word of length $d\eta^{\mathrm{i}}$; 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 (33).

Now, consider the decomposition

$$R^{\mathrm{i},N}_{w,\le d,n} = \sum_{\substack{\boldsymbol{w}=(w_1,\dots,w_n):\\ w_j\ge 2,\,\sum w_j=w}} \sum_{\substack{\boldsymbol{b}=(b_1,\dots,b_n):\\ 0\le b_1<\dots<b_n<M_N}} R^{\mathrm{i},N}_{\boldsymbol{w},\boldsymbol{b},\le d,n}\,,$$

where, once again, the constraint $w_j \ge 2$ comes from the recursiveness of ϕ^{i}. Using (33), we obtain the bound

$$\begin{aligned}\sum_{n=1}^{w/2-1} R^{\mathrm{i},N}_{w,\le d,n} &\le \sum_{n=1}^{w/2-1}\binom{w-n-1}{n-1}\binom{M_N}{n}\binom{d\eta^{\mathrm{i}}}{w-n}\\ &\le \sum_{n=1}^{w/2-1} e^{w-n-1}\frac{(eM_N)^n}{n^n}\frac{(ed\eta^{\mathrm{i}})^{w-n}}{(w-n)^{w-n}}\\ &\le \frac{e^{2w}}{(w/2)^w}\sum_{n=1}^{w/2-1} M_N^n(\eta^{\mathrm{i}}d)^{w-n}\\ &\le \frac{e^{2w}(\eta^{\mathrm{i}})^{w/2}}{(w/2)^w}\frac{d^{\frac{w}{2}}M_N^{\frac{w}{2}}}{\frac{M_N}{d\eta^{\mathrm{i}}}-1}\,,\end{aligned}$$

where the second inequality follows from (26) and (27), and the third one from (28).

Finally, we have to consider weight-enumerating coefficients of type T. For them, we have

$$\begin{aligned}T^{\mathrm{i},N}_{w,\le d} &= \sum_{1\le n\le \frac{w}{2}} T^{\mathrm{i},N}_{w,\le d,n}\\ &= \sum_{1\le n\le \frac{w}{2}} \sum_{\substack{\boldsymbol{w}=(w_1,\dots,w_n):\\ \sum w_j=w\\ w_j\ge 2\,\forall j<n,\, w_n\ge 1}} \sum_{\substack{\boldsymbol{b}=(b_1,\dots,b_n):\\ 0\le b_1<\dots<b_n<M_N\\ b_n\ge M_N-d\eta^{\mathrm{i}}}} T^{\mathrm{i},N}_{\boldsymbol{w},\boldsymbol{b},\le d,n}\,.\end{aligned}$$

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

With the same proof as for the bound (33) on $R^{\mathrm{i},N}_{\boldsymbol{w},\boldsymbol{b},\le d,n}$, we have also

$$T^{\mathrm{i},N}_{\boldsymbol{w},\boldsymbol{b},\le d,n} \le \binom{d\eta^{\mathrm{i}}}{w-n}\,,$$

so that

$$\begin{aligned}T^{\mathrm{i},N}_{w,\le d} &\le \sum_{n=1}^{w/2} \sum_{\substack{\boldsymbol{w}=(w_1,\dots,w_n):\\ \sum w_j=w\\ w_j\ge 2\,\forall j<n,\, w_n\ge 1}} \sum_{\substack{\boldsymbol{b}=(b_1,\dots,b_n):\\ 0\le b_1<\dots<b_n<M_N\\ b_n\ge M_N-d\eta^{\mathrm{i}}}} \binom{d\eta^{\mathrm{i}}}{w-n}\\ &\le \sum_{n=1}^{w/2}\binom{w-n}{n-1}\binom{M_N}{n-1}d\eta^{\mathrm{i}}\binom{d\eta^{\mathrm{i}}}{w-n}\\ &\le e^{2w-2}\,d\eta^{\mathrm{i}}\sum_{n=1}^{w/2}\frac{M_N^{n-1}(d\eta^{\mathrm{i}})^{w-n}}{(n-1)^{(n-1)}(w-n)^{(w-n)}}\\ &\le e^{2w-1}\frac{w}{2}\frac{d\eta^{\mathrm{i}}}{M_N}\sum_{n=1}^{w/2}\frac{M_N^n(d\eta^{\mathrm{i}})^{w-n}}{n^n(w-n)^{(w-n)}}\\ &\le \frac{e^{2w}}{(w/2)^w}\frac{w}{2}\frac{d\eta^{\mathrm{i}}}{M_N}\sum_{n=0}^{w/2}M_N^n(d\eta^{\mathrm{i}})^{w-n}\\ &\le \frac{e^{2w}}{(w/2)^w}\frac{w}{2}\frac{M_N^{w/2}(d\eta^{\mathrm{i}})^{w/2}}{\frac{M_N}{d\eta^{\mathrm{i}}}-1}\,,\end{aligned}$$

where the third inequality above follows from (26) and (27), the forth one from (29), and the fifth one from (28). Now, statement (b) of Lemma 3 follows from the fact that

$$A^{\mathrm{i},N}_{w,\le d} = R^{\mathrm{i},N}_{w,\le d,w/2} + \sum_{n=1}^{w/2-1} R^{\mathrm{i},N}_{w,\le d,n} + T^{\mathrm{i},N}_{w,\le d}\,. \tag{34}$$

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

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

All the error events have disjoint support, which implies the weaker condition that $b_1 < \dots < b_{\lfloor w/2\rfloor-1}$ and $b \ne b_1,\dots,b_{\lfloor w/2\rfloor-1}$. The overall output weight is $\le d$, and this implies the weaker condition $d^{\mathrm{i}}_{\mathrm{e}}\sum_{t=1}^{\lfloor w/2\rfloor-1} a_t \le d$ and $a < a' < \eta^{\mathrm{i}}d$. There are:

- $\binom{\eta^{\mathrm{i}}d}{2}$ choices for such a, a';
- $\binom{\lfloor d/d^{\mathrm{i}}_{\mathrm{e}}\rfloor}{\lfloor w/2\rfloor-1}$ choices for $a_1,\dots,a_{\lfloor w/2\rfloor-1}$;
- no more than $\lfloor w/2\rfloor\binom{M_N}{\lfloor w/2\rfloor}$ choices for $b_1,\dots,b_{\lfloor w/2\rfloor-1}, b$, 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,

$$\begin{aligned}R^{\mathrm{i},N}_{w,\le d,\lfloor w/2\rfloor} &\le \left\lfloor\tfrac{w}{2}\right\rfloor\tbinom{M_N}{\lfloor w/2\rfloor}\tbinom{\eta^{\mathrm{i}}d}{2}\tbinom{\lfloor d/d^{\mathrm{i}}_{\mathrm{e}}\rfloor}{\lfloor w/2\rfloor-1}\\ &\le \frac{(\eta^{\mathrm{i}})^2}{4e^2}\frac{we^w M_N^{\lfloor w/2\rfloor} d^2\left\lfloor\frac{d}{d^{\mathrm{i}}_{\mathrm{e}}}\right\rfloor^{\lfloor\frac{w}{2}\rfloor-1}}{\left\lfloor\frac{w}{2}\right\rfloor^{\lfloor\frac{w}{2}\rfloor}\left(\left\lfloor\frac{w}{2}\right\rfloor-1\right)^{\lfloor\frac{w}{2}\rfloor-1}}\\ &\le \frac{(\eta^{\mathrm{i}})^2}{16}\frac{w^3(2e)^w}{w^w}M_N^{\lfloor w/2\rfloor}d^2\left\lfloor\frac{d}{d^{\mathrm{i}}_{\mathrm{e}}}\right\rfloor^{\lfloor\frac{w}{2}\rfloor-1}\,,\end{aligned} \tag{35}$$

where the second inequality follows from (26), and the last inequality follows from (28) and (29).

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

$$\sum_{n=1}^{\lfloor w/2\rfloor-1} R^{\mathrm{i},N}_{w,\le d,n} \le \frac{e^{2w}(\eta^{\mathrm{i}})^{\lceil \frac{w}{2}\rceil}}{(w/2)^w}\frac{d^{\lceil \frac{w}{2}\rceil}M_N^{\lfloor \frac{w}{2}\rfloor}}{\frac{M_N}{d\eta^{\mathrm{i}}}-1}\,. \tag{36}$$

We now pass to studying the terms $T^{\mathrm{i},N}_{w,\le d}$. Differently from the even case, we shall consider the main term $T^{\mathrm{i},N}_{w,\le d,\lceil w/2\rceil}$ separately. Input words contributing to $T^{\mathrm{i},N}_{w,\le d,\lceil w/2\rceil}$ 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 input words as

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

and we observe that the following conditions hold

$$0\le b_1<\dots<b_{\lfloor w/2\rfloor}<M_N\,,$$

$$l\le \eta^{\mathrm{i}}d\,,\qquad d^{\mathrm{i}}_{\mathrm{e}}\sum_t a_t\le d\,.$$

Thus, we get

$$\begin{aligned} T^{\mathrm{i},N}_{w,\le d,\lceil w/2\rceil} &\le \binom{M_N}{\lfloor w/2\rfloor}d\eta^{\mathrm{i}}\binom{\lfloor d/d^{\mathrm{i}}_{\mathrm{e}}\rfloor}{\lfloor w/2\rfloor}\\ &\le \frac{\eta^{\mathrm{i}}}{2}\frac{w(2e)^w}{w^w}M_N^{\lfloor w/2\rfloor}\,d\left\lfloor\frac{d}{d^{\mathrm{i}}_{\mathrm{e}}}\right\rfloor^{\lfloor w/2\rfloor}\,. \end{aligned} \tag{37}$$

The remaining terms are bounded as in the even case,

$$\sum_{n=1}^{\lfloor w/2\rfloor} T^{\mathrm{i},N}_{w,\le d,n}\le \frac{e^{2w}}{(w/2)^w}\frac{w}{2}\,\frac{M_N^{\lfloor w/2\rfloor}(d\eta^{\mathrm{i}})^{\lceil w/2\rceil}}{\frac{M_N}{d\eta^{\mathrm{i}}}-1}\,. \tag{38}$$

By bounding the addends of the right-hand side of (34) as in (35), (36), (37), and (38), one finds that the leading terms are in fact the ones on the right-hand side of (35) and of (37), and statement (b) follows. This completes the proof of Lemma 3. ■

3) Proof of Lemma 4: We shall use ideas similar to those of [26, Lemma 2]. We consider a subclass of input words contributing to the term $R^{\mathrm{i},N}_{w,\le d,w/2}$, exactly those which can be written as

$$\sum_{1\le t\le w/2}\left(D^{i_t+h_{t-1}\delta^{\mathrm{i}}}+D^{i_t+h_t\delta^{\mathrm{i}}}\right)$$

with

$$0\le i_1<i_2<\dots<i_{w/2}<M_N-\delta^{\mathrm{i}}\lfloor d/d^{\mathrm{i}}_{\mathrm{e}}\rfloor\,,$$

$$0=h_0<h_1<h_2<\dots<h_{w/2}\le\lfloor d/d^{\mathrm{i}}_{\mathrm{e}}\rfloor\,.$$

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 word

$$D^{i_t+\delta^{\mathrm{i}}h_{t-1}}(1+D^{\delta^{\mathrm{i}}(h_t-h_{t-1})})\,,$$

so that the output has weight

$$\mathrm{w_H}\big(\phi^{\mathrm{i}}(1+D^{\delta^{\mathrm{i}}(h_t-h_{t-1})})\big)\le d^{\mathrm{i}}_{\mathrm{e}}\,(h_t-h_{t-1})\,.$$

Thus, the total output weight can be bounded from above as

$$d^{\mathrm{i}}_{\mathrm{e}}\sum_{t=1}^{w/2}(h_t-h_{t-1})=d^{\mathrm{i}}_{\mathrm{e}}h_{w/2}\le d\,.$$

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

$$R^{\mathrm{i},N}_{w,\le d,w/2}\ge\binom{M_N-\delta^{\mathrm{i}}\lfloor d/d^{\mathrm{i}}_{\mathrm{e}}\rfloor}{w/2}\binom{\lfloor d/d^{\mathrm{i}}_{\mathrm{e}}\rfloor}{w/2}\,. \tag{39}$$

Recall that, by assumption, $\frac{w}{2}\le\frac{d}{d^{\mathrm{i}}_{\mathrm{e}}}\le\frac{M_N}{2\delta^{\mathrm{i}}}$ and w is even. Hence,

$$\frac{w}{2}\le\left\lfloor\frac{d}{d^{\mathrm{i}}_{\mathrm{e}}}\right\rfloor,\quad M_N-\delta^{\mathrm{i}}\left\lfloor\frac{d}{d^{\mathrm{i}}_{\mathrm{e}}}\right\rfloor\ge\frac{M_N}{2}\,,\quad \frac{w}{2}\le M_N-\delta^{\mathrm{i}}\left\lfloor\frac{d}{d^{\mathrm{i}}_{\mathrm{e}}}\right\rfloor.$$

The final bound follows by applying (39) and (26). ■

B. Proofs of the results presented in Section IV

Throughout this subsection, we will use the words c^*, c^*_j and the set of indices J defined in Section IV.

1) Proof of Lemma 5: This proof closely follows part of the proof of [26, Thm. 2.b].

The first statement is immediate, let us prove the second one. Let

$$c^*_i=\sum_{m=1}^{d^{\mathrm{o}}_{\mathrm{f}}}D^{t_m}\,.$$

Given a multi-index

$$\boldsymbol{\tau}=(\tau_1,\dots,\tau_{d^{\mathrm{o}}_{\mathrm{f}}})\in[M_N]^{d^{\mathrm{o}}_{\mathrm{f}}}\,,$$

where $[M_N]:=\{0,\dots,M_N-1\}$, define the event

$$E_{\boldsymbol{\tau}}:=\{\Pi_N(D^{t_m})=D^{\tau_m}\ \forall m=1,\dots,d^{\mathrm{o}}_{\mathrm{f}}\}\,.$$

Clearly,

$$\mathbb{P}\big(E^*_{j_1}(d)\cap E^*_{j_2}(d)\big)=\sum_{\boldsymbol{\tau}}\mathbb{P}\big(E^*_{j_1}(d)\cap E_{\boldsymbol{\tau}}\big)\mathbb{P}\big(E^*_{j_2}(d)\big|E^*_{j_1}(d)\cap E_{\boldsymbol{\tau}}\big),$$

where the summation index $\boldsymbol{\tau}$ runs over all $[M_N]^{d^{\mathrm{o}}_{\mathrm{f}}}$.

Then, notice that

$$\mathbb{P}\big(E^*_{j_2}(d)\big|E^*_{j_1}(d)\cap E_{\boldsymbol{\tau}}\big)=\mathbb{P}\big(E^*_{j_2}(d)\big|E_{\boldsymbol{\tau}}\big)\,. \tag{40}$$

Also notice that

$$\mathbb{P}\big(E^*_{j_2}(d)\big|E_{\boldsymbol{\tau}}\big)\le R^{\mathrm{i},N}_{d^{\mathrm{o}}_{\mathrm{f}},\le d,d^{\mathrm{o}}_{\mathrm{f}}/2}\binom{M_N-d^{\mathrm{o}}_{\mathrm{f}}}{d^{\mathrm{o}}_{\mathrm{f}}}^{-1}\,. \tag{41}$$

Indeed, after having fixed the positions $\boldsymbol{\tau}$ where Π_N maps the $d^{\mathrm{o}}_{\mathrm{f}}$ ones of $c^*_{j_1}$, we need to find how many choices for the positions of the ones of $c^*_{j_2}$ will produce an output weight less than or equal to d, out of the $\binom{M_N-d^{\mathrm{o}}_{\mathrm{f}}}{d^{\mathrm{o}}_{\mathrm{f}}}$ ways to choose $d^{\mathrm{o}}_{\mathrm{f}}$ positions among $M_N-d^{\mathrm{o}}_{\mathrm{f}}$. The number of such favorable choices is bounded by the number of favorable choices that we would have if we could choose among all M_N positions, including the unavailable positions already assigned to $c^*_{j_1}$, i.e., $R^{\mathrm{i},N}_{d^{\mathrm{o}}_{\mathrm{f}},\leq d,d^{\mathrm{o}}_{\mathrm{f}}/2}$, which proves (41).

Eqs. (40) and (41), together with (16), give

$$\mathbb{P}\big(E^*_{j_2}(d)\big|E^*_{j_1}(d)\cap E_{\boldsymbol{\tau}}\big) \leq \mathbb{P}\big(E^*_{j_2}(d)\big)\binom{M_N}{d^{\mathrm{o}}_{\mathrm{f}}}\binom{M_N-d^{\mathrm{o}}_{\mathrm{f}}}{d^{\mathrm{o}}_{\mathrm{f}}}^{-1}.$$

Therefore,

$$\begin{aligned}&\mathbb{P}\big(E^*_{j_1}(d)\cap E^*_{j_2}(d)\big)\\ &\quad\leq \sum_{\boldsymbol{\tau}} \mathbb{P}\big(E^*_{j_1}(d)\cap E_{\boldsymbol{\tau}}\big)\mathbb{P}\big(E^*_{j_2}(d)\big)\binom{M_N-d^{\mathrm{o}}_{\mathrm{f}}}{d^{\mathrm{o}}_{\mathrm{f}}}^{-1}\binom{M_N}{d^{\mathrm{o}}_{\mathrm{f}}},\end{aligned}$$

where the summation index $\boldsymbol{\tau}$ runs over the set $[M_N]^{d^{\mathrm{o}}_{\mathrm{f}}}$. Finally, observe that

$$\sum_{\boldsymbol{\tau}\in[M_N]^{d^{\mathrm{o}}_{\mathrm{f}}}} \mathbb{P}(E^*_{j_1}(d)\cap E_{\boldsymbol{\tau}}) = \mathbb{P}(E^*_{j_1}(d))\,.$$

From this, the claim immediately follows. ■

2) Proof of Theorem 2: The key idea, introduced in [4], 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 [4] to our setting, and then we state the lemma on hypergraphs given in [4], which completes the proof. The aim is to show that, for any interleaver, it is possible to find a suitable subset of the codewords c^*_j, say $\{c^*_j : j\in\tilde{S}\}$, with cardinality growing at most as logarithmically with N, and such that the outer codeword $c := \sum_{j\in\tilde{S}} c^*_j$ produces a codeword $y=\phi^{\mathrm{i}}_N\circ\pi_N(c)$ of the serial code having weight $\mathrm{w_H}(y)$ non-zero and smaller than $KN^\beta\log N$, for some constant K.

Let $\mathbb{Z}_{\delta^{\mathrm{i}}}$ be the ring of integers modulo δ^{i}. Define a map $\boldsymbol{\sigma} : J\to\mathbb{Z}^{d^{\mathrm{o}}_{\mathrm{f}}}_{\delta^{\mathrm{i}}}$ by associating with an index $j\in J$ a vector $(\sigma_1(j),\ldots,\sigma_{d^{\mathrm{o}}_{\mathrm{f}}}(j))$ in the following way: if

$$c^*_j = \sum_{m=1}^{d^{\mathrm{o}}_{\mathrm{f}}} D^{t_m}\,,\qquad \pi_N(D^{t_m}) = D^{\tau_m}\,,$$

with $\{t_m\}_m$ an increasing sequence, then $\sigma_m(j) = \tau_m \mod \delta^{\mathrm{i}}$. By the pigeonhole principle, clearly there exists $U\subseteq J$ with $|U|\geq |J|/(\delta^{\mathrm{i}})^{d^{\mathrm{o}}_{\mathrm{f}}}$ such that $\boldsymbol{\sigma}(j_1)=\boldsymbol{\sigma}(j_2)$ for all $j_1,j_2\in U$.

This means that, for every $m=1,\ldots,d^{\mathrm{o}}_{\mathrm{f}}$, all the m-th ones in words c^*_j, with $j\in U$, are permuted by π_N to positions whose relative distance is a multiple of δ^{i}. Thus, applying ϕ^{i} to any pair of such ones gives an output weight which is proportional to the distance between the two ones. The goal is to find a non-empty subset of indices $\tilde{S}\subseteq U$, such that its cardinality $|\tilde{S}|$ is even and grows at most logarithmically with N, and such that for all $m=1,\ldots,d^{\mathrm{o}}_{\mathrm{f}}$, the ones being the m-th one of words c^*_j with $j\in\tilde{S}$ form pairs in such a way that after the permutation the distance within ones of the same pair grows at most as N^β. This will allow to construct an outer codeword $c=\sum_{j\in\tilde{S}} c^*_j$ which gives a codeword $y=\phi^{\mathrm{i}}_N\circ\pi_N(c)$ of the serial scheme, whose weight grows as most as a constant times $N^\beta\log N$.

In order to find the set $\tilde{S}$, consider the set $[M_N] = \{0,\ldots,M_N-1\}$ and divide it in b intervals $I_1,\ldots,I_b$, each of length at most $\lceil M_N/b\rceil$; b is a parameter depending on N that will be properly chosen later in this proof.

Define a hypergraph $H=(V,E)$ in the following way. Take a $d^{\mathrm{o}}_{\mathrm{f}}$-partite vertex set V being the union of $d^{\mathrm{o}}_{\mathrm{f}}$ disjoint copies of $W=\{I_1,\ldots,I_b\}$. The set of hyperedges E has cardinality $|U|$ and is $d^{\mathrm{o}}_{\mathrm{f}}$-regular in the sense that $E\subseteq W^{d^{\mathrm{o}}_{\mathrm{f}}}$, i.e., every hyperedge contains exactly one vertex from each of the $d^{\mathrm{o}}_{\mathrm{f}}$ copies of W. Any hyperedge in E corresponds to an index $j\in U$, and is defined as $e=(I_{h_1},\ldots,I_{h_{d^{\mathrm{o}}_{\mathrm{f}}}})\in W^{d^{\mathrm{o}}_{\mathrm{f}}}$ where, denoting as above

$$c^*_j = \sum_{m=1}^{d^{\mathrm{o}}_{\mathrm{f}}} D^{t_m}$$

with $\{t_m\}_m$ and increasing sequence, the index h_m is such that $\pi_N(D^{t_m})\in I_{h_m}$.

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 ([4], 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, we choose

$$b=\left\lfloor\left(\frac{|J|}{4(\delta^{\mathrm{i}})^{d^{\mathrm{o}}_{\mathrm{f}}}}\right)^{2/d^{\mathrm{o}}_{\mathrm{f}}}\right\rfloor = \left\lfloor\left(\frac{1}{4(\delta^{\mathrm{i}})^{d^{\mathrm{o}}_{\mathrm{f}}}}\left\lfloor\frac{N}{d^{\mathrm{o}}_{\mathrm{f}}\eta^{\mathrm{o}}}\right\rfloor\right)^{2/d^{\mathrm{o}}_{\mathrm{f}}}\right\rfloor.$$

This ensures that b is an integer satisfying

$$4b^{d^{\mathrm{o}}_{\mathrm{f}}/2}\leq\frac{|J|}{(\delta^{\mathrm{i}})^{d^{\mathrm{o}}_{\mathrm{f}}}}\leq|U|=|E|\,,$$

so that we can apply Lemma 6 and find the subset S.

By construction of the hypergraph, there is a bijection between hyperedges and indices in $U\subset J$; let $\tilde{S}\subset U$ be the indices corresponding to the hyperedges in S, so that any hyperedge $s\in S$ corresponds to some word c^*_j, $j\in\tilde{S}$. Let $c:=\sum_{j\in\tilde{S}} c^*_j\in\mathbb{Z}_2^{M_N}$, and observe that c is clearly a non-zero codeword of the outer code. Hence, $y:=\phi^{\mathrm{i}}_N(\pi_N(c))$ is a non-zero codeword of the serial turbo code.

By construction, $\pi_N(c)$ is composed of $|S|d^{\mathrm{o}}_{\mathrm{f}}/2$ pairs of ones. Each pair has both ones lying in a same interval I_j and at a distance multiple of δ^{i}. Hence,

$$\mathrm{w_H}\big(\phi^{\mathrm{i}}_N(\pi_N(c))\big)\leq\frac{|S|d^{\mathrm{o}}_{\mathrm{f}}}{2}\,d^{\mathrm{i}}_{\mathrm{e}}\left\lceil\frac{M_N}{b}\right\rceil\,.$$

Finally use the bound on $|S|$ which is the key contribution of Lemma 6: $|S|\leq d^{\mathrm{o}}_{\mathrm{f}}\log b$.

Our choice of b gives

$$\log(b)\leq\log(N^{2/d^{\mathrm{o}}_{\mathrm{f}}})=\frac{2}{d^{\mathrm{o}}_{\mathrm{f}}}\log(N)$$

and

$$\lceil M_N/b\rceil\leq 6r(8d^{\mathrm{o}}_{\mathrm{f}}\eta^{\mathrm{o}})^{2/d^{\mathrm{o}}_{\mathrm{f}}}(\delta^{\mathrm{i}})^2N^{1-2/d^{\mathrm{o}}_{\mathrm{f}}}\,,$$

which concludes the proof. ■

Appendix II
Generalizations

Parts of Assumptions 1 and 2 were stated for the sake of simplicity, and are in fact not essential for the validity of the results presented. In this appendix, we shortly discuss how such assumptions can be weakened, pointing out the role they played in the proofs and stating the results that can be obtained in greater generality, while we refer the interested reader to [21] for more details and proofs.

The following formulation is the one truly needed in order to obtain the claimed asymptotic behavior of the minimum distance and the error probability:

Assumption 3. *The outer encoder* $\phi^{\rm o} : (\mathbb{Z}_2^k)^{\mathbb{Z}_+} \to (\mathbb{Z}_2^r)^{\mathbb{Z}_+}$ *is non-catastrophic, and its free distance* $d_{\rm f}^{\rm o}$ *satisfies* $d_{\rm f}^{\rm o} \geq 3$.

Assumption 4. *The inner encoder* $\phi^{\rm i} : (\mathbb{Z}_2^s)^{\mathbb{Z}_+} \to (\mathbb{Z}_2^l)^{\mathbb{Z}_+}$ *is non-catastrophic and recursive.*

Non-catastrophicity of both constituent encoders and recursiveness of the inner encoder are needed in order to ensure the properties of the weight-enumerating coefficients (Lemmas 2 and 3), and to give the limitations on the input weights (due to Lemma 1 and to the absence of input-weight-1 inner codewords) in the summations in the proofs of Propositions 1 and 3.

The assumption $d_{\rm f}^{\rm o} \geq 3$ is needed in order to ensure that $\beta > 0$, and is essential in order to have minimum distance growing with high probability as some positive power of N. Indeed, when $d_{\rm f}^{\rm o} = 2$ (and thus $\beta = 0$), Theorem 2 still holds true, and states that, for any choice of the interleavers sequence, the minimum distance grows at most logarithmically with N. Moreover, a slight modification of the proof of Proposition 2 (see [21, Sect. 4.5.1]) allows one to prove that, when $d_{\rm f}^{\rm o} = 2$,

$$\mathbb{P}(d_N^{\min} \leq d_{\rm e}^{\rm i}) \geq c$$

for some positive constant c, which implies that

$$\mathbb{P}\left(P(e|\Pi_N) \geq p^{d_{\rm e}^{\rm i}}\right) \geq c\,,$$

where $p = \operatorname{erfc}(\sqrt{\rho})/2$ is the bit error probability of uncoded transmission.

The assumptions that the inner encoder $\phi^{\rm i}$ has scalar input ($s = 1$) and is proper rational (F is invertible) have been considered in order to simplify the analysis of the codewords of $\phi_N^{\rm i}$ made of error events with input weight 2 (proofs of Lemma 3 and Theorem 2), and to have clean expressions of the constants depending on $d_{\rm e}^{\rm i}$. Indeed, under such assumptions, an input word with weight two produces a finite-weight output word if an only if the two ones are separated by $a\delta^{\rm i} - 1$ zeros, and the output weight is $ad_{\rm e}^{\rm i}$, because the word is made of a shifted copies of the same error event, with non-overlapping support. When $\phi^{\rm i}$ is not proper rational, the above-mentioned error events have overlapping support, so that the weight is smaller than $a\delta^{\rm i}$: this allows one to prove bounds on the one side, while for the other side it is necessary to introduce another parameter of the inner encoder, for which the opposite inequality holds true. When $\phi^{\rm i}$ has non-scalar input ($s > 1$), we have to look separately at pairs of ones being in different components of the entry vector, so that we need to define s parameters $\delta^{\rm i}(j)$ and corresponding weights $d_{\rm e}^{\rm i}(j)$, one for each component $j = 1, \dots, s$; moreover, we need to take into account also possible pairs of ones where the second one is not in the same component as the first one (which turn out to have an asymptotically negligible role). For more details, see [21], Sections 4.5.2 and 4.5.3.

Removing the assumptions that $\phi^{\rm i}$ has scalar input ($s = 1$) and is proper rational (F is invertible) does not change any of the asymptotic results when N grows large: except for the value of the constants and their dependence on $d_{\rm e}^{\rm i}$, all the statements of this paper remain true under Assumptions 1 and 4.

Removing the assumption that $d_{\rm f}^{\rm o}$ is even requires some more effort, because of the key role that was played by words where an outer codeword with weight $d_{\rm f}^{\rm o}$ (or multiples of it) was producing inner codewords composed of error events each with input weight two. In the remainder of this section, we consider the case of odd $d_{\rm f}^{\rm o}$, and for simplicity we focus again on the simpler case where the inner encoder satisfies Assumption 2, while we replace Assumption 1 with the following:

Assumption 5. *The outer encoder* $\phi^{\rm o} : (\mathbb{Z}_2^k)^{\mathbb{Z}_+} \to (\mathbb{Z}_2^r)^{\mathbb{Z}_+}$ *is non-catastrophic, and its free distance* $d_{\rm f}^{\rm o}$ *is odd and satisfies* $d_{\rm f}^{\rm o} \geq 3$.

We will state and prove the main results (the asymptotic typical behavior of $d_N^{\min}$ and $P(e|\Pi_N)$, while we will refer the reader to [21] for details on some results we will only quickly mention.

Notice that, under Assumptions 5 and 2, Lemmas 2 and 3 hold true without any modification. However, Proposition 1 needs to be modified, because the dominant term in the summations is not the same, due to the ceilings and floors of the fractions in the exponents. The following Proposition holds true, where for simplicity we do not look at the explicit dependence of the constants on $d_{\rm e}^{\rm i}$ and on other parameters of the inner encoder such as the output weight of terminated error events with input weight 1 or of regular error events with input weight 3.

Proposition 4. *Let Assumptions 5 and 2 be satisfied. Assume that* $d = o(N^{\beta})$ *as* N *grows large. Then, there exists* $N_0 \geq 0$ *and* $C_1, C_2 > 0$, *depending on the constituent convolutional encoders only, such that, for all* $N \geq N_0$,

$$\mathbb{P}(d_N^{\min} \leq d) \leq C_1 \left(\frac{d}{N}\right)^{1/2} \left(N^{-\beta} d\right)^{d_{\rm f}^{\rm o}/2} + C_2 \left(N^{-\beta} d\right)^{d_{\rm f}^{\rm o}}\,.$$

Before giving the proof, we underline the fact that, differently from Proposition 1, we have two terms in this upper bound, and either one can be the dominant one, depending on how fast d grows with N: defining

$$\kappa = 1 - \frac{2}{d_{\rm f}^{\rm o} - 1}$$

(notice that $\kappa < \beta$), if $d = o(N^{\kappa})$ the dominant term is the first one, while otherwise it is the second one.

Proof: From (4), we use Lemmas 2 and 3 to find bounds on the weight-enumerating coefficients of the constituent encoders, and we get

$$\mathbb{P}(d_N^{\min} \le d) \le \sum_{w=d_\mathrm{f}^\mathrm{o}}^{\eta^\mathrm{i} d} C^w N^{\lfloor w/d_\mathrm{f}^\mathrm{o}\rfloor - \lceil w/2 \rceil} d^{\lceil w/2\rceil} \tag{42}$$

for some $C > 0$ depending on the constituent convolutional encoders only. For even d_f^o, the asymptotically dominant term in the summation was the one with $w = d_\mathrm{f}^\mathrm{o}$. Here, for odd d_f^o, we have different dominant terms: the ones with $w = d_\mathrm{f}^\mathrm{o}$ and with $w = d_\mathrm{f}^\mathrm{o} + 1$ dominate if $d = o(N^\kappa)$, and otherwise the dominant term is the one with $w = 2d_\mathrm{f}^\mathrm{o}$. To prove this, we consider separately the terms with odd and even w in (42). For the odd terms, using $\lfloor w/d_\mathrm{f}^\mathrm{o}\rfloor \le w/d_\mathrm{f}^\mathrm{o}$ and the fact that $\lceil w/2\rceil = (w+1)/2$ for odd w, we get

$$\sum_{\substack{d_\mathrm{f}^\mathrm{o} \le w \le \eta^\mathrm{i} d \\ w \mathrm{odd}}} C^w N^{\lfloor w/d_\mathrm{f}^\mathrm{o}\rfloor - \lceil w/2 \rceil} d^{\lceil w/2\rceil} \le \left(\frac{d}{N}\right)^{\frac{1}{2}} \sum_{w \ge d_\mathrm{f}^\mathrm{o}} \left(C N^{-\frac{\beta}{2}} d^{\frac{1}{2}}\right)^w . \tag{43}$$

For even w, we need to split once more the summation in two parts. A first summation will contain the terms with w multiple of d_f^o, for which $\lfloor w/d_\mathrm{f}^\mathrm{o}\rfloor = w/d_\mathrm{f}^\mathrm{o}$; notice that such terms have $w \ge 2d_\mathrm{f}^\mathrm{o}$. All the other terms will have

$$\lfloor w/d_\mathrm{f}^\mathrm{o}\rfloor \le \frac{w}{d_\mathrm{f}^\mathrm{o}} - \frac{1}{d_\mathrm{f}^\mathrm{o}} , \qquad w \ge d_\mathrm{f}^\mathrm{o} + 1 .$$

Hence,

$$\sum_{\substack{d_\mathrm{f}^\mathrm{o} < w \le \eta^\mathrm{i} d \\ w \text{ even}}} C^w N^{\lfloor w/d_\mathrm{f}^\mathrm{o}\rfloor - \lceil w/2 \rceil} d^{\lceil w/2\rceil} \le \sum_{w \ge 2d_\mathrm{f}^\mathrm{o}} \left(C N^{-\frac{\beta}{2}} d^{\frac{1}{2}}\right)^w + N^{-1/d_\mathrm{f}^\mathrm{o}} \sum_{w \ge d_\mathrm{f}^\mathrm{o}+1} \left(C N^{-\frac{\beta}{2}} d^{\frac{1}{2}}\right)^w . \tag{44}$$

Similarly to the proof of Proposition 1, we can use the assumption $d = o(N^\beta)$ to conclude that, for sufficiently large N, the series in (43) and (44) are convergent and each one is bounded by twice its first term. ■

Similarly to what was done for the even case with Proposition 2, a lower bound can be found, which ensures that the upper bound given in Proposition 4 is tight for $d = o(N^\kappa)$; this is useful in order to find $\alpha = 1 - 2/\lceil d_\mathrm{f}^\mathrm{o}/2\rceil$ such that the growth rate $X_N := (\log N)^{-1} \log d_N^{\min}$ and the decay rate $Y_N := (\log N)^{-1} \log(-\log(P(e|\Pi_N)))$ densely cover the interval $[\alpha, \beta]$ with probability one, but we will not discuss such issue here.

For even d_f^o, Proposition 1 (or equivalently the upper bound in Theorem 1) was completed by Theorem 2: the two results together imply that the growth rate $X_N := (\log N)^{-1} \log d_N^{\min}$ converges in probability to β. For odd d_f^o, it is indeed possible to prove a deterministic upper bound, analogous to Theorem 2, by a slight modification of the construction of the bipartite graph from the hypergraph in the proof of Theorem 2 (see the proof of [4, Thm. 2] for repeat-accumulate codes, or see [21]). Unfortunately, such a bound is of the form

$$d_N^{\min} \le C N^{\tilde\beta} \log N$$

where

$$\tilde\beta := 1 - \frac{1}{\lceil d_\mathrm{f}^\mathrm{o}/2\rceil} = 1 - \frac{2}{d_\mathrm{f}^\mathrm{o} + 1} > \beta .$$

However, as suggested in [26], it is still possible to prove that N^β is the actual growth rate of $d_N^{\min}$, using a second-order method, as shown below.

Theorem 4. *Let Assumptions 5 and 2 be satisfied. If $d = \omega(N^\beta)$ as N grows large, then there exist positive constants C_1, C_2, and N_0, such that*

$$\mathbb{P}(d_N^{\min} \le d) \ge 1 - \frac{C_1}{N} - C_2 \frac{N^\beta}{d} ,$$

for all $N \ge N_0$.

Proof: Let the outer codewords c^*, c_j^* and the set of indices J be the same as in Section IV and in Appendix I-B. We define events quite similar to the E_j^*'s involved in the proof of Proposition 2, but here we consider pairs of codewords c_j^*'s. More precisely, for $j_1, j_2 \in J$, we define

$$E_{j_1,j_2}^*(d) := \bigcup_{(\boldsymbol{b},\boldsymbol{e}) \in \mathcal{B}} E_{j_1,j_2}(\boldsymbol{b},\boldsymbol{e}) ,$$

where

$$E_{j_1,j_2}(\boldsymbol{b},\boldsymbol{e}) := \left\{ \Pi_N(c_{j_1}^*) = \sum_{t=1}^{d_\mathrm{f}^\mathrm{o}} D^{b_t} ,\ \Pi_N(c_{j_2}^*) = \sum_{t=1}^{d_\mathrm{f}^\mathrm{o}} D^{e_t} \right\} ,$$

$\boldsymbol{b} = (b_1, \ldots, b_{d_\mathrm{f}^\mathrm{o}})$, $\boldsymbol{e} = (e_1, \ldots, e_{d_\mathrm{f}^\mathrm{o}})$, and

$$\mathcal{B} := \Big\{ (\boldsymbol{b},\boldsymbol{e}) \text{ s.t. } 0 \le b_1 < e_1 < \cdots < b_{d_\mathrm{f}^\mathrm{o}} < e_{d_\mathrm{f}^\mathrm{o}} \le M_N , \quad e_t = b_t + l_t \delta^\mathrm{i} \forall t , \textstyle\sum_{t=1}^{d_\mathrm{f}^\mathrm{o}} l_t \le \lfloor d/d_\mathrm{e}^\mathrm{i} \rfloor \Big\} .$$

Now, let χ_{j_1,j_2} be the indicator of the event $E_{j_1,j_2}^*(d)$, and define the random variable

$$Z := \sum_{j_1,j_2 \in J,\, j_1 \ne j_2} \chi_{j_1,j_2} .$$

Clearly

$$\mathbb{P}(d_N^{\min} \le d) \ge \mathbb{P}\Big(\bigcup_{j_1,j_2 \in J,\, j_1 \ne j_2} E_{j_1,j_2}^*(d) \Big) = 1 - \mathbb{P}(Z = 0).$$

A standard argument, which follows from Chebyshev's inequality and is known as 'second-order method' [2, Thm. 4.3.1], gives

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

so that

$$\mathbb{P}(d_N^{\min} \le d) \ge 2 - \frac{\mathbb{E}(Z^2)}{[\mathbb{E}(Z)]^2} = 2 - \frac{\sum_{\substack{\boldsymbol{j} \in J^4 \\ j_1 \ne j_2, j_3 \ne j_4}} \Lambda_{\boldsymbol{j}}}{\Xi^2} , \tag{45}$$

where, for $\boldsymbol{j} = (j_1, j_2, j_3, j_4) \in J^4$,

$$\Lambda_{\boldsymbol{j}} := \mathbb{P}\big(E_{j_1,j_2}^*(d) \cap E_{j_3,j_4}^*(d)\big)$$

and

$$\Xi := \sum_{j,j' \in J, j \ne j'} \mathbb{P}\big(E_{j,j'}^*(d)\big) .$$

The following steps allow one to find bounds for Ξ and $\Lambda_{\boldsymbol{j}}$. First, notice that $\mathbb{P}\big(E^*_{j,j'}(d)\big)$ is the same for all pairs $j \neq j'$, so that $\Xi = |J|(|J|-1)\mathbb{P}\big(E^*_{j,j'}(d)\big)$. Then, notice that the union in the definition of $E^*_{j_1,j_2}(d)$ is a disjoint union, so that

$$\mathbb{P}\big(E^*_{j,j'}(d)\big) = \sum_{(\boldsymbol{b},\boldsymbol{e})\in\mathcal{B}} \mathbb{P}\big(E_{j,j'}(\boldsymbol{b},\boldsymbol{e})\big)\,.$$

Moreover,

$$\mathbb{P}\big(E_{j,j'}(\boldsymbol{b},\boldsymbol{e})\big) = \frac{(d_{\mathrm{f}}^{\mathrm{o}}!)^2(M_N - 2d_{\mathrm{f}}^{\mathrm{o}})!}{M_N!}$$

and the set $\mathcal{B}$ can be conveniently described in the following equivalent way (which was already used in the proof of Lemma 4):

$$\begin{aligned}\mathcal{B} := \big\{(\boldsymbol{b},\boldsymbol{e}) \text{ s.t. } \forall t,\ b_t = i_t + h_{t-1}\delta^{\mathrm{i}} \text{ and } e_t = i_t + h_t\delta^{\mathrm{i}},\\ 0 \le i_1 < i_2 < \cdots < i_{w/2} < M_N - \delta^{\mathrm{i}}\lfloor d/d_{\mathrm{e}}^{\mathrm{i}}\rfloor,\\ 0 = h_0 < h_1 < h_2 < \cdots < h_{w/2} \le \lfloor d/d_{\mathrm{e}}^{\mathrm{i}}\rfloor\big\}.\end{aligned}$$

from which it is clear that

$$|\mathcal{B}| = \binom{M_N - \delta^{\mathrm{i}}\lfloor d/d_{\mathrm{e}}^{\mathrm{i}}\rfloor}{d_{\mathrm{f}}^{\mathrm{o}}}\binom{\lfloor d/d_{\mathrm{e}}^{\mathrm{i}}\rfloor}{d_{\mathrm{f}}^{\mathrm{o}}}\,.$$

Thus we have the following explicit formula:

$$\mathbb{P}\big(E^*_{j,j'}(d)\big) = \binom{M_N - \delta^{\mathrm{i}}\lfloor d/d_{\mathrm{e}}^{\mathrm{i}}\rfloor}{d_{\mathrm{f}}^{\mathrm{o}}}\binom{\lfloor d/d_{\mathrm{e}}^{\mathrm{i}}\rfloor}{d_{\mathrm{f}}^{\mathrm{o}}}\frac{(d_{\mathrm{f}}^{\mathrm{o}}!)^2(M_N - 2d_{\mathrm{f}}^{\mathrm{o}})!}{M_N!}\,. \tag{46}$$

Then we consider $\Lambda_{\boldsymbol{j}}$. We use a similar proof as for Lemma 5, i.e., we condition on the events $E_{j_1,j_2}(\boldsymbol{b},\boldsymbol{e})$.

If j_1, j_2, j_3, j_4 are all distinct, then

$$\begin{aligned}\Lambda_{\boldsymbol{j}} &= \sum_{(\boldsymbol{b},\boldsymbol{e})\in\mathcal{B}} \mathbb{P}\big(E^*_{j_3,j_4}(d)\big|E_{j_1,j_2}(\boldsymbol{b},\boldsymbol{e})\big)\mathbb{P}\big(E_{j_1,j_2}(\boldsymbol{b},\boldsymbol{e})\big)\\ &\le \sum_{(\boldsymbol{b},\boldsymbol{e})\in\mathcal{B}} |\mathcal{B}|\frac{(d_{\mathrm{f}}^{\mathrm{o}}!)^2(M_N-4d_{\mathrm{f}}^{\mathrm{o}})!}{(M_N-2d_{\mathrm{f}}^{\mathrm{o}})!}\mathbb{P}\big(E_{j_1,j_2}(\boldsymbol{b},\boldsymbol{e})\big)\\ &= \mathbb{P}\big(E^*_{j_1,j_2}(d)\big)\mathbb{P}\big(E^*_{j_3,j_4}(d)\big)\frac{(M_N-4d_{\mathrm{f}}^{\mathrm{o}})!(M_N)!}{[(M_N-2d_{\mathrm{f}}^{\mathrm{o}})!]^2}\end{aligned} \tag{47}$$

so that $\Lambda_{\boldsymbol{j}} \le \mathbb{P}\big(E^*_{j_1,j_2}(d)\big)^2\big(1+O(1/N)\big)$ as N grows large.

When one of the indices is repeated, say $j_1 = j_3$, we have that

$$\begin{aligned}\Lambda_{\boldsymbol{j}} &= \sum_{(\boldsymbol{b},\boldsymbol{e})\in\mathcal{B}} \mathbb{P}\big(E^*_{j_1,j_4}(d)\big|E_{j_1,j_2}(\boldsymbol{b},\boldsymbol{e})\big)\mathbb{P}\big(E_{j_1,j_2}(\boldsymbol{b},\boldsymbol{e})\big)\\ &\le \sum_{(\boldsymbol{b},\boldsymbol{e})\in\mathcal{B}} \binom{\lfloor d/d_{\mathrm{e}}^{\mathrm{i}}\rfloor}{d_{\mathrm{f}}^{\mathrm{o}}}\frac{d_{\mathrm{f}}^{\mathrm{o}}!(M_N-3d_{\mathrm{f}}^{\mathrm{o}})!}{(M_N-2d_{\mathrm{f}}^{\mathrm{o}})!}\mathbb{P}\big(E_{j_1,j_2}(\boldsymbol{b},\boldsymbol{e})\big)\\ &= \mathbb{P}\big(E^*_{j_1,j_2}(d)\big)\binom{\lfloor d/d_{\mathrm{e}}^{\mathrm{i}}\rfloor}{d_{\mathrm{f}}^{\mathrm{o}}}\frac{d_{\mathrm{f}}^{\mathrm{o}}!(M_N-3d_{\mathrm{f}}^{\mathrm{o}})!}{(M_N-2d_{\mathrm{f}}^{\mathrm{o}})!}\end{aligned} \tag{48}$$

and the same bound holds true when $j_2 = j_4$.

Finally, it's clear that $\Lambda_{\boldsymbol{j}} = \mathbb{P}\big(E^*_{j_1,j_2}(d)\big)$ for all $\boldsymbol{j} \in J^4$ such that $j_3 = j_1$ and $j_4 = j_2$.

The above bounds allow one to prove that the right-hand side of (45) tends to one. In fact, we can split the summation into the following terms

$$\mathbb{P}(d_N^{\min} \le d) \ge 2 - S_4 - S_3 - S_2\,,$$

where

$$S_2 = \sum_{j_1=j_3\neq j_2=j_4} \frac{\Lambda_{\boldsymbol{j}}}{\Xi^2}\,, \qquad S_4 = \sum_{\substack{j_1,j_2,j_3,j_4\\ \text{distinct}}} \frac{\Lambda_{\boldsymbol{j}}}{\Xi^2}\,,$$

$$S_3 = \sum_{\substack{j_2\neq j_1=j_3\\ j_3\neq j_4\neq j_2}} \frac{\Lambda_{\boldsymbol{j}}}{\Xi^2} + \sum_{\substack{j_1\neq j_2=j_4\\ j_1\neq j_3\neq j_4}} \frac{\Lambda_{\boldsymbol{j}}}{\Xi^2}\,.$$

Remember that $|J|$ and M_N grow linearly with N, and that d/N^{β} grows unbounded by assumption. On the other hand, without loss of generality one may assume that d/N vanishes, since the deterministic upper bound guarantees that $d_N^{\min} \le CN^{\bar{\beta}}\log N$ for any choice of the interleavers sequence. Then, using (46), (47), (48), and the bound (26) for the binomial coefficients, it is easy to conclude that, as N grows large,

$$S_4 \le 1 + \frac{C_1}{N}, \quad S_3 \le \frac{C_2}{N}, \quad S_2 \le \frac{C_3}{N} + C_4(N^{\beta}d^{-1})^{d_{\mathrm{f}}^{\mathrm{o}}}$$

for some positive constants C_1, C_2, C_3, C_4. ■

Similarly to Section V, we will now show how the above results on the minimum distance imply results on the word error probability. We will use here the same notation

$$E_N^{\varepsilon} := \{d_N^{\min} > N^{\beta-\varepsilon}\}\,,\ F_N^{\varepsilon} := \{P(e|\Pi_N) \ge \exp(-N^{\beta-\varepsilon})\}\,.$$

A first result is that Proposition 3 holds true also when Assumption 5 replaces Assumption 1: the only modification in the proof is that now $\mathbb{P}(E_N^{\varepsilon})$ converges to 1 thanks to Proposition 4 instead of Theorem 1.

The following theorem is the analogous of Theorem 3 for odd $d_{\mathrm{f}}^{\mathrm{o}}$.

Theorem 5. *Let Assumptions 5 and 2 be satisfied. Then, there exists some finite $\rho_0 \ge 0$ such that, if the signal-to-noise ratio ρ satisfies $\rho \ge \rho_0$, then for all $\varepsilon \in (0, \beta-\kappa)$ there exist some finite $N_0 \ge 0$ and $C > 0$ such that, for all $N \ge N_0$,*

$$\mathbb{P}\Big(\exp(-N^{\beta+\varepsilon}) \le P(e|\Pi_N) \le \exp(-N^{\beta-\varepsilon})\Big) \ge 1 - CN^{-\varepsilon d_{\mathrm{f}}^{\mathrm{o}}}.$$

Proof: Similarly to the proof of Theorem 3, the upper bound follows from Proposition 3 and from Proposition 4 (which is the analogous for odd $d_{\mathrm{f}}^{\mathrm{o}}$ of Proposition 1)

$$\mathbb{P}\big(F_N^{\varepsilon}\big) \le 1 - \mathbb{P}\big(E_N^{\varepsilon}\big) + \mathbb{P}\big(E_N^{\varepsilon}\big|F_N^{\varepsilon}\big) \le \frac{C_1}{N^{\varepsilon d_{\mathrm{f}}^{\mathrm{o}}}} + C_2\exp(-N^{\beta-\varepsilon})\,.$$

The lower bound is obtained again using (24), but here the role of Theorem 2 is replaced by Theorem 4

$$\mathbb{P}\big(P(e|\Pi_N) \ge p^{N^{\beta+\varepsilon}}\big) \ge \mathbb{P}\big(d_N^{\min} \ge N^{\beta+\varepsilon}\big) \ge 1 - \frac{c_1}{N} - \frac{c_2}{N^{\varepsilon d_{\mathrm{f}}^{\mathrm{o}}}}\,.$$

Finally, notice that, for $\varepsilon \in (0, \beta-\kappa)$, $1/N = o\big(1/N^{\varepsilon d_{\mathrm{f}}^{\mathrm{o}}}\big)$ as N grows large. ■

ACKNOWLEDGMENTS

The authors thank Prof. Rüdiger Urbanke of EPFL for an interesting discussion on the topics of this paper. They are grateful to the Associate Editor and to an anonymous Referee for many detailed comments on an earlier version of this work, which significantly helped in improving its form.

References

[1] A. Abbasfar, D. Divsalar, and Y. Kung, "Accumulate-Repeat-Accumulate codes," *IEEE Trans. Comm.*, vol. 55, no. 4, pp. 692–702, April 2007.

[2] N. Alon and J. Spencer, "The Probabilistic Method," 3rd Ed., J. Wiley & Sons, Hoboken, NJ, USA, 2008.

[3] A. Barg and G. D. Forney, Jr., "Random codes: Minimum distances and error exponents," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2568–2573, September 2002.

[4] L. Bazzi, M. Mahdian, and D. A. Spielman, "The minimum distance of turbo-like codes," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, pp. 6–15, January 2009.

[5] 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, no. 3, pp. 909–926, May 1998.

[6] S. Benedetto and G. Montorsi, "Design of parallel concatenated convolutional codes," *IEEE Trans. Communicat.*, vol. 44, no. 5, pp. 591–600, May 1996.

[7] A. Bennatan and D. Burshtein, "On the application of LDPC codes to arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 417–438, March 2004.

[8] 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.

[9] M. Breiling, "A logarithmic upper bound on the minimum distance of turbo codes," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1692–1710, August 2004.

[10] C. Brutel and J. Boutros, "Serial concatenation of interleaved convolutional codes and M-ary continuous phase modulations," *Annals of Telecommunications*, vol. 54, no. 3-4, pp. 235–242, 1999.

[11] G. Como, F. Fagnani, and 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 3–7, 2006.

[12] G. Como and F. Fagnani, "Average type spectra and minimum distances of low-density parity-check codes over Abelian groups," *SIAM J. Discr. Math.*, vol. 23, no. 1, pp. 19–53, October 2008.

[13] G. Como, "Group codes outperform binary-coset codes on non-binary symmetric memoryless channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4321–4334, September 2010.

[14] D. Divsalar and F. Pollara, "Serial and hybrid concatenated codes with applications" *Proc. of the 1st International Symposium on Turbo Codes and Related Topics (Brest, France)*, pp. 80–87, 1997.

[15] F. Fagnani, "Performance of parallel concatenated coding schemes," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1521–1535, April 2008.

[16] G. D. Forney, Jr., "Convolutional codes I: Algebraic structure," *IEEE Trans. Inf. Theory*, vol. 16, no. 6, pp. 720–738, Nov. 1970.

[17] C. Fragouli and R. D. Wesel, "Convolutional codes and matrix control theory," *Proc. of the 7th International Conference on Advances in Communications and Control (Athens, Greece)*, June 28-July 2, 1999.

[18] S. Franz, M. Leone, A. Montanari, and F. Ricci-Tersenghi, "Dynamic phase transition for decoding algorithms," *Phys. Rev. E*, vol. 22, 046120, 2002.

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

[20] R. Garello, P. Pierleoni, and S. Benedetto, "Computing the free distance of turbo codes and serially concatenated convolutional codes: Algorithms and applications," *IEEE J. Sel. Areas Comm.*, vol. 19, pp. 800–812, May 2001.

[21] F. Garin, *Generalized serial turbo coding ensembles: Analysis and design*, Ph.D. Thesis, Department of Mathematics, Politecnico di Torino, Torino, Italy, March 2008.

[22] F. Garin and F. Fagnani, "Analysis of serial turbo codes over Abelian groups for symmetric channels," *SIAM J. Discr. Math.*, vol. 22, no. 4, pp. 1488–1526, October 2008.

[23] A. Graell i Amat, G. Montorsi, and F. Vatta, "Design and performance analysis of a new class of rate compatible serially concatenated convolutional codes," *IEEE Trans. Comm.*, vol. 57, no. 8, pp. 2280–2289, August 2009.

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

[25] R. Johannesson and K. S. Zigangirov, *Fundamentals of Convolutional Coding*, Wiley-IEEE Press, 1999.

[26] N. Kahale and R. Urbanke, "On the minimum distance of parallel and serially concatenated codes", unpublished manuscript, 1997. Available online: `http://lthcwww.epfl.ch/~ruediger/papers/weight.ps`

[27] K. Li, G. Yue, X. Wang, and L. Ping, "Low-rate Repeat-Zigzag-Hadamard codes," *IEEE Trans. Inf. Theory*, vol. 54, no. 2, pp. 531–543, February 2008.

[28] M. Mézard and A. Montanari, *Information, Physics, and Computation*, Oxford University Press, 2009.

[29] G. Miller and D. Burshtein, "Bounds on the maximum likelihood decoding error probability of low-density parity-check codes," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2696–2710, November 2001.

[30] A. Montanari, "The glassy phase of Gallager codes," *Eur. Phys. J. B*, vol. 23, no. 1, pp. 121–136, January 2001.

[31] T. Mora and O. Rivoire, "Error exponents of low-density parity-check codes on the binary erasure channel," *Proc. of the 2006 IEEE Information Theory Workshop (Punta del Este, Uruguay)*, pp. 81–85, March 13–17, 2006.

[32] A. Perotti and S. Benedetto, "An upper bound on the minimum distance of serially concatenated convolutional codes," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5501–5509, December 2006.

[33] V. Rathi, "On the asymptotic weight and stopping set distribution of regular LDPC ensembles", *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4212–4218, September 2006.

[34] I. Sason and S. Shamai, "Improved upper bounds on the ML decoding error probability of parallel and serial concatenated turbo codes via their ensemble distance spectrum," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, pp. 24–47, January 2000.

[35] N.S Skantzos, J. van Mourik, D. Saad, and Y. Kabashima, "Average and reliability error exponents in low-density parity-check codes," *J. Phys. A*, vol. 36, pp. 11131–11142, 2003.

[36] W. M. Wonham, "On pole assignment in multi-input, controllable linear systems," *IEEE Trans. Aut. Control*, vol. 12, pp. 660–665, 1967.