On the Minimum Distance of Parallel and Serially Concatenated Codes

Nabil Kahale
Room C-258
Tel/Fax: (973) 360-8447/8178
e-mail: kahale@research.att.com
AT&T Research, 180 Park Ave, Florham Park, NJ 07932

Rüdiger Urbanke
Room 2C-254
Tel/Fax: (908) 582-3657/3340
e-mail: ruediger@lucent.com
Bell Labs, 600 Mountain Avenue, Murray Hill, NJ 07974

Abstract

We show that with high probability the minimum distance of parallel concatenated codes with k parallel branches and recursive component codes grows like $n^{\frac{k-2}{k}}$ in the interleaving length n. In particular, this growth rate is independent of the choice of recursive component codes. As a specific case, the minimum distance of standard turbo codes with only two branches does not grow like any power of n. For serially concatenated codes with two recursive component codes the minimum distance grows with high probability like $n^{\frac{d_o^*-2}{d_o^*}}$, where d_o^* is the free distance of the *outer* code. The result is still valid if the outer encoder is non-recursive.

Index Terms – turbo-codes, minimum distance.

I. Introduction

A. Motivation

Ever since the introduction of turbo codes [1] there has been a strong interest in the communications community to explain their sensational behavior and to develop guidelines for their design. The two principle research questions are (i) to investigate the properties of the *codes* and (ii) to investigate the properties of the associated recursive *decoding* algorithm.

In this paper we present new results concerning the codes. As stated more precisely later, we show that the minimum distance of parallel concatenated codes with k parallel branches and recursive component codes grows like $n^{\frac{k-2}{k}}$ in the interleaving length n. For serially concatenated codes the minimum distance grows like $n^{\frac{d_o^*-2}{d_o^*}}$. Hereby, d_o^* is the free distance of the outer code. It is well known that for low signal-to-noise ratios the good performance of turbo-codes is not based on their large minimum distance but rather on their small multiplicity of near neighbors. For larger signal-to-noise ratios, on the other hand, the relatively modest minimum distance of turbo codes (especially of parallel concatenated codes)

is clearly visible as a floor in the error probability. Hence, at least in this region, we expect to gain some valuable information by investigating the growth rate of the minimum distance as a function of the interleaving length n.

In [2, 3, 4, 5, 6, 7, 8] the average weight distribution of various ensembles of parallel or serially concatenated codes was investigated. In [2, 3, 4, 5, 8] the ensembles were defined by a fixed choice of component codes and a uniform choice over all interleavers, whereas in [6, 7] the component codes where also chosen randomly from the set of time-varying convolutional codes. Based on these ensemble averages of the weight distribution and the particular channel model one can then give upper bounds on their error probability under a maximum likelihood decoding. For sufficiently large signal-to-noise ratios these bounds are tight and one can quantify the "interleaving gain", which is defined to be the decrease in the error probability with increasing interleaving length n. A slightly simpler task is to assume that the signal-to-noise ratio is sufficiently large so that the minimum distance term dominates the error probability. One can then investigate how the corresponding coefficient behaves as a function of the interleaver length n. This determines the behavior of the ensemble error probability for (not too large) changes of n. This approach was taken in [2, 3, 4, 8]. We note that one has to be careful in interpreting these results since in the range where this approximation is valid the ensemble probability of error is dominated by a few very bad codes and "typical codes" might behave quite differently.

Another way of quantifying the "interleaving gain" is to determine the growth rate of the minimum distance which is done in this paper, see also [9]. We note that our results concern individual codes rather than ensemble behavior.

From the point of view of the code designer our result for parallel codes is disappointing. After all, the growth rate of the minimum distance is *independent* of the choice of component codes. Also, for the particular case of the original turbo-codes with $k = 2$ we see that the minimum distance does not grow like any power of n. Note that the growth rate of the minimum distance is an increasing function of k. Although we did not include the effects of puncturing in our derivations, it is tempting to conjecture that for constant rates, parallel concatenated codes with larger k (and also an appropriately larger amount of puncturing) are better than those with smaller k.

For serially concatenated codes we see that we can achieve growth rates close to linear if we pick an outer code with large free distance. These results are consistent with the design rules in [4] where it was suggested to pick an outer code with a large distance d_o^*.

In the remaining of this section we introduce the necessary notation and review a few basic facts about the recursive component codes. The main result is stated in Section II. The proofs are then given in Section III.

B. Systematic Recursive Convolutional Encoders

In this paper we focus on turbo codes whose component codes are binary convolutional codes of rate 1/2. Of particular interest for turbo codes are systematic recursive convolutional encoders. Hence, in the present section we will review those properties of convolutional

codes and encoders which are relevant to our discussion of turbo codes.

A binary systematic recursive convolutional encoder with memory m and rate $1/2$ is defined in terms of a rational function $G(D) = p(D)/q(D)$ of degree m where $\gcd(p, q) = 1$. We assume that the feedback is non-trivial, i.e., the degree of q is at least 1. Fig. 1 depicts the particular example $G(D) = (1 + D^4)/(1 + D + D^2 + D^3 + D^4)$. For a given rational function

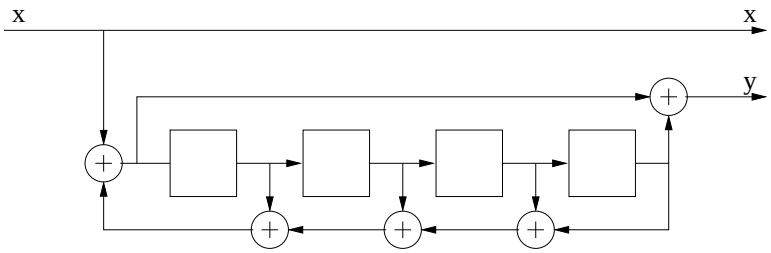

Fig. 1. A binary systematic recursive covolutional encoder with memory m and rate $1/2$ defined by $G(D) = (1 + D^4)/(1 + D + D^2 + D^3 + D^4)$.

G and length n, $n \in \mathsf{N}$, we associate a code $C = C(G, n)$ as follows: Let the input x be $x = (x_1, \cdots, x_n, \underbrace{0, \cdots, 0}_{m\times})$, where the first n components are elements of GF(2) and the last m components are zero. Associated to each input x is an output $y = (y_1, \cdots, y_{n+m})$, which is the result of passing x through a linear filter. Hereby, for the first n steps the filter is equal to $G(D)$, whereas for the last m steps the filter is defined as $G'(D) = p(D)$, i.e., we remove the feedback. This termination technique was suggested in [10]. Let this encoding map be denoted by $y = \gamma(x)$, where for simplicity we suppress the dependency of γ on G and n in our notation. Then $C(G, n) := \{(x, \gamma(x)) : (x_1, \cdots, x_n, \underbrace{0, \cdots, 0}_{m\times}), x_i \in \mathrm{GF}(2)\}$.

To every input x there is an associated state sequence $s = (s_0, \cdots, s_{n+m})$, where s_i is the content of the shift register after x_i has entered the encoder. By definition $s_0 = 0$, i.e., the encoder starts in the zero state. Also, since the last m steps of the encoding operation clear the shift register we have $s_{n+m} = 0$.

Example 1: For the encoder depicted in Fig. 1 and $n = 10$, Table I lists the state sequence s as well as the output y associated to the input $x = (1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0)$. $\diamond$

Every non-zero codeword $(x, \gamma(x)) \in C$ consists of possibly several *detours*. Detours are defined in terms of the state sequence s as continuous sections starting and ending in the zero state and taking on non-zero state values in between. More formally, the beginning b_i and ending e_i of the ith detour can be defined as follows. Let $e_0 = 0$. Then b_i and e_i, if defined, are given by

$$\begin{aligned} b_i &= \min\{j \in \{e_{i-1} + 1, \cdots, n\} : s_j \neq 0\}, \\ e_i &= \min\{j \in \{b_i + 1, \cdots, n + m\} : s_j = 0\}. \end{aligned}$$

The length of a detour l_i is defined as $l_i := e_i - b_i + 1$ and the number of detours is denoted by t.

k	x	s	y	k	x	s	y
0		0000					
1	1	1000	1	8	1	1000	1
2	0	1100	1	9	1	0100	0
3	0	0110	0	10	0	1010	1
4	0	0011	0	11	0	0101	0
5	0	0001	1	12	0	0010	1
6	1	0000	1	13	0	0001	0
7	0	0000	0	14	0	0000	1

TABLE I

THE STATE SEQUENCE s AS WELL AS THE OUTPUT y ASSOCIATED TO THE INPUT $x = (1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0)$ FOR THE CODE C DEFINED BY $G(D) = (1 + D^4)/(1 + D + D^2 + D^3 + D^4)$ AND $n = 10$.

Example 2: In Example 1 the number of detours is 2, $b_1 = 1$, $e_1 = 6$, $b_2 = 8$, and $e_2 = 14$. $\diamond$

Every encoder has an associated directed graph called a *state diagram* [11, p. 442]. Its vertices are the states and vertex v is connected to vertex u if the state corresponding to vertex u can be reached from the state corresponding to vertex v in one step by an input of either 0 or 1. We label the edges with the sum of the input and output weight corresponding to this transition. Fig. 2 shows the state diagram for the encoder depicted in Fig. 1. There

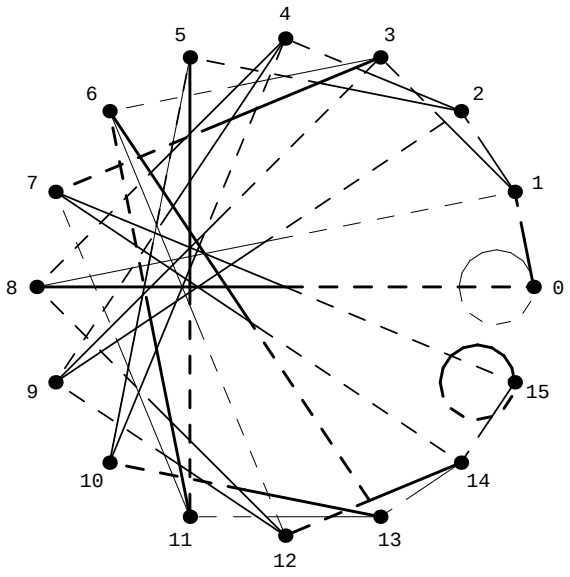

Fig. 2. The state diagram for the encoder of Fig. 1. Thin, medium, and thick lines correspond to weights 0, 1, and 2, respectively. Solid lines indicate outgoing and dashed lines indicate incoming edges.

are only finitely many states, namely 2^m. Further, due to the condition $\gcd(p, q) = 1$ there are no cycles of zero weight in the state diagram except for the self-loop at the zero state. Call a path trough the state diagram *admissible* if it does not contain an initial segment corresponding to cycles at the zero state. Then there exists a constant η, $\eta = \eta(G)$, such

that any admissible path of length at least η has non-zero weight. (Note that we do not have to worry about cycles corresponding to the zero state at the end of an admissible path since the transition from a non-zero state to the zero state always requires a non-zero input.)

Example 3: For our continuing example an inspection of the state diagram in Fig. 2 reveals that $\eta = 4$. This corresponds to 1 plus the length of the longest path consisting only of thin lines (lines with associated weight zero). $\diamond$

We can use this fact to bound the length of a detour given its weight. Because we chose to disconnect the feedback for the last m steps we will have to distinguish between two kinds of detours. Note that for any detour $b_i \leq n$. We say that a detour is *regular* if $e_i \leq n$, whereas we call a detour *terminating* if $e_i > n$. If a regular detour has weight d then it can have length at most $d\eta$. To see this assume to the contrary that the length of some regular detour of weight d is strictly larger than $d\eta$. Partition the detour into sections of length η, padding the last section with zeros if necessary. Note that each such section corresponds to an admissible path and, hence, has non-zero weight. This leads to a contradiction since there are at least $d+1$ such sections. A moment's thought shows that the same statement is true also for terminating detours if we replace η by $\eta + m$. We will assume in the sequel that this has been done so that the length of any detour (regular or terminating) of weight d is bounded by ηd. In general, if a codeword consists of several detours and has weight d then the total length of all its detours is bounded by $d\eta$.

Assume that the input x to a convolutional encoder as the one given in Fig. 1 is a single 1 followed by an infinite number of 0's. The output y will then be of infinite weight. If on the other hand x has weight 2 then the output y will have *finite* weight if the spacing of the 1's is chosen correctly [4]. This follows directly from the well-known fact that for every polynomial $q(D)$ with binary coefficients there exists a natural number δ such that $q(D)$ divides $1 + D^\delta$ [12]. Let δ be the smallest such natural number. Let ω be the weight of the detour that corresponds to an input with 2 1's separated by δ. By linearity, an input with 2 1's separated by $j\delta$, $j \in \mathsf{N}$, has then weight at most $j\omega$.

C. Parallel and Serially Concatenated Codes

A parallel concatenated code with k branches is defined as follows. Fix a rational function G of degree m and a length n. Further let $\pi = (\pi_1, \cdots, \pi_k)$, where $\pi_i : \{1, \cdots, n+m\} \to \{1, \cdots, n+m\}$, $1 \leq i \leq k$, is a permutation on $n+m$ letters which fixes the last m letters. The associated code $P = P(G, n, \pi)$ is defined as $P(G, n, \pi) := \{(x, \gamma(\pi_1(x)), \cdots, \gamma(\pi_k(x))) : (x_1, \cdots, x_n, \underbrace{0, \cdots, 0}_{m\times}), x_i \in \mathrm{GF}(2)\}$. Without loss of generality we can always assume that π_1 is the identity map. Fig. 3 shows an example with $G(D) = (1+D^4)/(1+D+D^2+D^3+D^4)$, k=2, and $\pi_1 = \mathrm{id}$. We note that we could define parallel concatenated codes in a more general way by allowing different component codes for each branch. Nevertheless, as we will see in the sequel the growth rate of the minimum distance is *independent* of the choice of component codes and so our choice entails no essential loss of generality.

For fixed G, n, and k, let $\mathcal{P} = \mathcal{P}(G, n, k)$ be the ensemble of codes generated by varying π over all possible choices. In the sequel we let $P = P(G, n, \pi)$ be a random code chosen

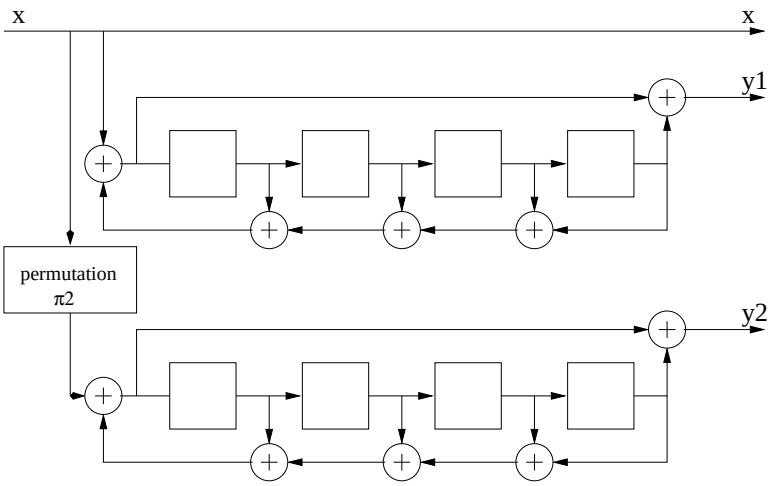

Fig. 3. A parallel concatenated code with $G(D) = (1 + D^4)/(1 + D + D^2 + D^3 + D^4)$, k=2, and $\pi_1 = \mathrm{id}$.

uniformly from $\mathcal{P}(G, n, k)$.

Serially concatenated codes are defined in an analogous way. Fix a rational function G_o of degree m_o, a rational function G_i of degree m_i, and a length n. Further, let π, $\pi : \{1, \cdots, 2(n + m_o) + m_i\} \to \{1, \cdots, 2(n + m_o) + m_i\}$, be a permutation on $2(n + m_o) + m_i$ letters which fixes the last m_i letters. The associated code $S = S(G_o, G_i, n, \pi)$ is defined as $S(G_o, G_i, n) := \{(x \cdot \gamma_o(x) \cdot \underbrace{0 \cdots 0}_{m_i \times}, \gamma_i(\pi(x \cdot \gamma_o(x) \cdot \underbrace{0 \cdots 0}_{m_i \times}))) : (x_1, \cdots, x_n, \underbrace{0, \cdots, 0}_{m_o \times}), x_i \in \mathrm{GF}(2)\}$, where $\cdot$ denotes concatenation of sequences. Note that we append m_i zeros to the output of the first encoder so that the input to the second encoder has the proper form.

Fig. 4 shows an example with $G_o(D) = G_i(D) = (1 + D^4)/(1 + D + D^2 + D^3 + D^4)$. We note

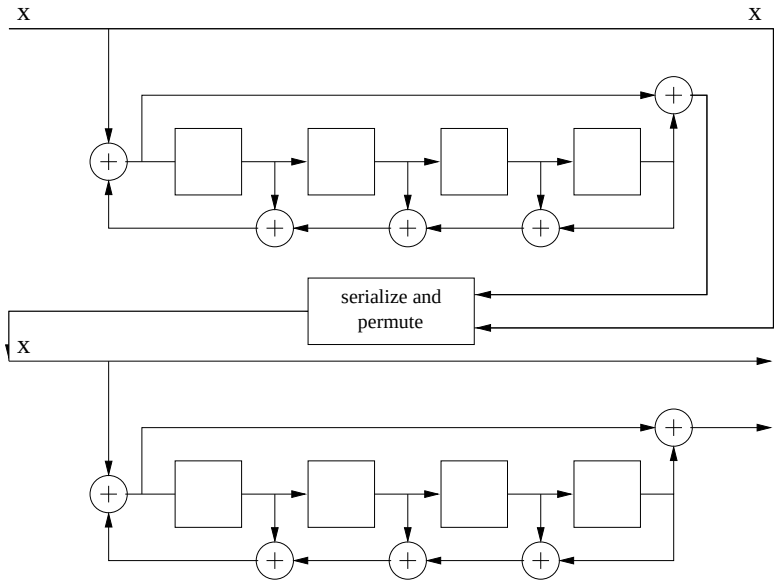

Fig. 4. A serially concatenated code with $G_o(D) = G_i(D) = (1 + D^4)/(1 + D + D^2 + D^3 + D^4)$.

that our main result holds valid if we replace the outer recursive convolutional encoder by an equivalent non-recursive encoder, i.e., the proof depends only on the parameters of the outer *code* but is independent of the particular outer *encoder*. This is not true for the inner encoder which needs to be recursive for our main result to hold. Contrary to the case of parallel concatenated codes, the minimum distance of serially concated codes *does* depend on the choice of the component codes. For large signal-to-noise ratios we should pick a serially concatenated code with as large as possible distance of the outer component code.

For fixed G_o, G_i, and n let $\mathcal{S} = \mathcal{S}(G_o, G_i, n)$ be the ensemble of codes generated by varying π over all possible choices. In the sequel we let $S = S(G_o, G_i, n, \pi)$ be a random code chosen uniformly from $\mathcal{S}(G_o, G_i, n)$.

II. Main Result and Discussion

The main result of this paper is the determination of the growth rate of the minimum distance of a parallel or serially concatenated code.

Given a code C, let $C_{\leq d}$ denote the set of *non-zero* codewords of C with weight at most d. For parallel concatenated codes we have

Theorem 1: Let $\mathcal{P}(G, n, k)$ be the ensemble of parallel concatenated codes with encoder defined by the rational function G of degree m and k parallel branches. Let P be a random code chosen uniformly from $\mathcal{P}(G, n, k)$. Then for every $\epsilon > 0$,

$$(a)\ \Pr\{|P_{\leq n^{\frac{k-2}{k}-\epsilon}}| = 0\} \overset{n\to\infty}{\to} 1 \ \text{ and } \ (b)\ \Pr\{|P_{\leq n^{\frac{k-2}{k}+\epsilon}}| = 0\} \overset{n\to\infty}{\to} 0.$$

In words, the minimum distance of a parallel concatenated code with k branches grows like $n^{\frac{k-2}{2}}$ in the interleaving length n. Note in particular, that Theorem 1 asserts that the minimum distance of a code with only two branches does not grow like any power of n, whereas the minimum distance of a code with 3 branches grows like $n^{1/3}$. It is quite surprising that the growth rate is independent of the choice of the component codes.

As mentioned in the introduction, this suggests the following conjecture: Parallel codes with a larger number of branches (properly punctured to maintain a constant rate) are better than codes with a smaller number of branches. This conjecture seems also reasonable since codes with a larger number of branches contain more "randomness." Unfortunately, it seems hard to extend our results to include puncturing. Also, this behavior might not be apparent from simulations since the performance of the suboptimum decoding algorithm is also a function of the number of branches.

For serially concatenated codes our result shows a different behavior. The growth rate *does* depend on the choice of the component codes, namely it depends on the *free distance* of the *outer* code.

Theorem 2: Let $\mathcal{S}(G_o, G_i, n)$ be the ensemble of serially concatenated codes with encoders defined by the rational functions G_o and G_i respectively. Let S be a random code chosen uniformly from $\mathcal{S}(G_o, G_i, n)$. Then for every $\epsilon > 0$,

$$(a)\ \Pr\{|S_{\leq n^{\frac{d_o^*-2}{d_o^*}-\epsilon}}| = 0\} \overset{n\to\infty}{\to} 1 \ \text{ and } \ (b)\ \Pr\{|S_{\leq n^{\frac{d_o^*-2}{d_o^*}+\epsilon}}| = 0\} \overset{n\to\infty}{\to} 0.$$

where d_o^* is the free distance of the outer code.

In words, the minimum distance of a serially concatenated code grows like $n^{\frac{d_o^*-2}{d_o^*}}$ in the interleaving length n. We see that we can achieve growth rates close to linear if we pick an outer code with a large minimum distance. Note that our results are consistent with the design rules in [4] where it was also suggested to pick an outer code with a large d_o^*.

III. Proofs

A. The Quantities $|C^w_{\leq d}|$ and $|C_d|$

Before we can start the proofs of Theorem 1 and 2 we need to have a closer look at the convolutional component codes.

Given a systematic code C, let $C^w_{\leq d}$ denote the set of codewords of C with weight at most d which stem from inputs of weight w. How large can $C^w_{\leq d}$ be? This is answered in

Lemma 1: For a given rational function G and a given natural number n let $C = C(G, n)$. For $w \geq 1$,

$$|C^w_{\leq d}| \leq \Theta(1)^w \left(\frac{nd}{w^2}\right)^{w/2}.$$

where $\Theta(1)$ is a constant independent of n.

Proof. First note that a codeword in $C^1_{\leq d}$ consists only of one terminating detour. But there are at most $d\eta$ starting values for this terminating detour since by the remarks in Section I-B the length of a detour of weight d is bounded above by $d\eta$. Hence $|C^1_{\leq d}| \leq d\eta$ and the lemma clearly holds for $w = 1$. For $w \geq 2$, we first show that

$$|C^w_{\leq d}| \leq \Theta(1)^w \sum_{t=1}^{\lfloor \frac{w}{2} \rfloor} \binom{n}{t}\binom{d\eta}{w-t} \tag{1}$$

We split $C^w_{\leq d}$ into $R^w_{\leq d}$ and $T^w_{\leq d}$ where R denotes codewords with only regular detours whereas T denotes codewords including a terminating detour. Note that each non-zero codeword consists of at least one detour and that an element of $R^w_{\leq d}$ consists of at most $\lfloor \frac{w}{2} \rfloor$ detours since each regular detour requires an input weight of at least 2 whereas an element of $T^w_{\leq d}$ can consist of as many as $\lceil \frac{w}{2} \rceil$ detours. Let $C^{(w_1,\cdots,w_t)}_{\leq d}$ be the set of codewords with weight at most d consisting of t detours, the ith detour having input weight equal to w_i. Clearly,

$$\begin{aligned}
|R^w_{\leq d}| &= \sum_{t=1}^{\lfloor \frac{w}{2} \rfloor} \sum_{(w_1,\cdots,w_t):\sum w_i = w; w_i \geq 2, 1 \leq i \leq t} |R^{(w_1,\cdots,w_t)}_{\leq d}|, \\
|T^w_{\leq d}| &= \sum_{t=1}^{\lceil \frac{w}{2} \rceil} \sum_{(w_1,\cdots,w_t):\sum w_i = w; w_i \geq 2, 1 \leq i < t; w_t \geq 1} |T^{(w_1,\cdots,w_t)}_{\leq d}|.
\end{aligned}$$

We claim that $|R^{(w_1,\cdots,w_t)}_{\leq d}| \leq \binom{n}{t}\binom{d\eta}{w-t}$ and that $|T^{(w_1,\cdots,w_t)}_{\leq d}| \leq \binom{n}{t-1}\binom{d\eta}{w-t}d\eta$. To show this let $C^{((w_1,b_1),\cdots,(w_t,b_t))}_{\leq d}$ be the subset of $C^{(w_1,\cdots,w_t)}_{\leq d}$ where the ith detour starts at position b_i.

We now exhibit an injective map from $C_{\leq d}^{((w_1,b_1),\cdots,(w_t,b_t))}$ into the set of binary $(d\eta)$-tuples of weight $w-t$. Let $(x,y) \in C_{\leq d}^{((w_1,b_1),\cdots,(w_t,b_t))}$. The map is composed of two simpler maps $(x,y) \to x \to x'$. Clearly the first map $(x,y) \to x$ is injective. The second map $x \to x'$ consists roughly speaking of deleting all components of x which are not part of a detour. More precisely,

$$x' := (x_{b_1+1}, \cdots, x_{e_1}, x_{b_2+1}, \cdots, x_{e_2}, \cdots, x_{b_t+1}, \cdots, x_{e_t}).$$

Note that by definition $x_{b_i} = 1$, $1 \leq i \leq t$, and that $x_{e_i} = 1$, $1 \leq i \leq t-1$. In words, the first input of a regular or terminating detour is non-zero and the last input of a regular detour is non-zero. Because of this and since w_i,$1 \leq i \leq t$, is known it is possible to dissect x' again into its corresponding detours. We simply start from the left of x' and look for the (w_1-1)st non-zero element which marks the end of the first detour and we continue in this manner. Since further b_i,$1 \leq i \leq t$, is known it is possible to reconstruct x from x'. This shows that $x \to x'$ and, therefore, $(x,y) \to x \to x'$ is injective.

Example 4: As in Example 1 let

$$(x,y) := ((1,0,0,0,0,1,0,1,1,0,0,0,0,0),(1,1,0,0,1,1,0,1,0,1,0,1,0,1)).$$

Then $x = (1,0,0,0,0,1,0,1,1,0,0,0,0,0)$ and $x' = (0,0,0,0,1,1,0,0,0,0,0)$. To reconstruct x from x' and $((w_1,b_1),(w_2,b_2)) = ((2,1),(2,8))$ we first dissect x' into its detours. Since $w_1 = 2$ we know that the first detour ends where the first 1 appears in x'. The remaining part then belongs to the second detour. Hence, adding the leading 1's we know that the first detour is $(1,0,0,0,0,1)$ and that the second one is $(1,1,0,0,0,0,0)$. Finally, these detours start at positions 1 and 8, respectively. $\diamond$

As remarked in section I-B, the total length of all t detours is bounded above by $d\eta$. It follows that $|R_{\leq d}^{((w_1,b_1),\cdots,(w_t,b_t))}|, |T_{\leq d}^{((w_1,b_1),\cdots,(w_t,b_t))}| \leq |C_{\leq d}^{((w_1,b_1),\cdots,(w_t,b_t))}| \leq \binom{d\eta}{w-t}$. Since there are at most $\binom{n}{t}$ starting values for t regular detours it follows that $|R_{\leq d}^{(w_1,\cdots,w_t)}| \leq \binom{n}{t}\binom{d\eta}{w-t}$ and since there at most $\binom{n}{t-1}d\eta$ starting values for $(t-1)$ regular and one terminating detour it follows that $|T_{\leq d}^{(w_1,\cdots,w_t)}| \leq \binom{n}{t-1}\binom{d\eta}{w-t}d\eta$. Hence, we can bound $|R_{\leq d}^{w}|$ by

$$\begin{aligned}
|R_{\leq d}^{w}| &= \sum_{t=1}^{\lfloor \frac{w}{2} \rfloor} \sum_{(w_1,\cdots,w_t):\sum w_i = w; w_i \geq 2, 1 \leq i \leq t} |R_{\leq d}^{(w_1,\cdots,w_t)}| \\
&\leq \sum_{t=1}^{\lfloor \frac{w}{2} \rfloor} \sum_{(w_1,\cdots,w_t):\sum w_i = w; w_i \geq 2, 1 \leq i \leq t} \binom{n}{t}\binom{d\eta}{w-t} \\
&\overset{(a)}{\leq} \sum_{t=1}^{\lfloor \frac{w}{2} \rfloor} \binom{w-t-1}{t-1}\binom{n}{t}\binom{d\eta}{w-t} \\
&= \Theta(1)^w \sum_{t=1}^{\lfloor \frac{w}{2} \rfloor} \binom{n}{t}\binom{d\eta}{w-t}.
\end{aligned}$$

To see (a) note that there are exactly $\binom{w-1}{t-1}$ ordered partitions of a positive integer w into t positive parts, $w = w_1 + \cdots + w_t$, [13, p. 31]. Requiring that all components are at least 2 is equivalent to asking for the number of ordered partitions of $w - t$ into t positive parts.

In the same manner we can bound $|T^w_{\leq d}|$.

$$\begin{aligned}
|T^w_{\leq d}| &= \sum_{t=1}^{\lceil \frac{w}{2} \rceil} \sum_{(w_1,\cdots,w_t):\sum w_i = w; w_i \geq 2, 1 \leq i < t; w_t \geq 1} |T^{(w_1,\cdots,w_t)}_{\leq d}| \\
&\leq \sum_{t=1}^{\lceil \frac{w}{2} \rceil} \sum_{(w_1,\cdots,w_t):\sum w_i = w; w_i \geq 2, 1 \leq i < t; w_t \geq 1} \binom{n}{t-1}\binom{d\eta}{w-t} d\eta \\
&\leq \sum_{t=1}^{\lceil \frac{w}{2} \rceil} \binom{w-t}{t-1}\binom{n}{t-1}\binom{d\eta}{w-t} d\eta \\
&\leq \Theta(1)^w \sum_{t=1}^{\lfloor \frac{w}{2} \rfloor} \binom{n}{t}\binom{d\eta}{w-t}.
\end{aligned}$$

Eq. 1 now follows since $|C^w_{\leq d}| = |R^w_{\leq d}| + |T^w_{\leq d}|$. We now bound the right-hand side of Eq. 1 using the inequality $\binom{n}{s} \leq (ne/s)^s$ [14, p. 206]. For $1 \leq t \leq w/2$,

$$\begin{aligned}
\binom{n}{t}\binom{d\eta}{w-t} &\leq \left(\frac{ne}{t}\right)^t \left(\frac{d\eta e}{w-t}\right)^{w-t} \\
&\leq \frac{(ne)^t (d\eta e)^{w-t}}{(w/2)^w} \\
&\leq \frac{(ne)^{w/2} (d\eta e)^{w/2}}{(w/2)^w},
\end{aligned}$$

where the third inequality holds if $n \geq d\eta$. The second inequality follows from the fact that the function $t^t(w-t)^{w-t}$ attains its minimum at $t = w/2$ when t varies continuously in the interval $[0, w]$. Thus, if $n \geq d\eta$,

$$\sum_{t=1}^{\lfloor \frac{w}{2} \rfloor} \binom{n}{t}\binom{d\eta}{w-t} \leq w(2e\eta^{1/2})^w \left(\frac{nd}{w^2}\right)^{w/2},$$

and the lemma follows in this case. On the other hand, the lemma clearly holds if $n \leq d\eta$ since $|C^w_{\leq d}| \leq \binom{n}{w} \leq (ne/w)^w$. ■

We will also need the following lower bound on $|C^w_{\leq d}|$.

Lemma 2: If w is even and $2 \leq d \leq 2n + m$, then $|C^w_{\leq d}| \geq \Theta(1)^w \left(\frac{nd}{w^2}\right)^{w/2}$.

Proof. Consider the following codewords of input weight w which are the sum (over GF(2)) of $w/2$ regular detours, each of input weight 2. The tth detour has its two non-zero inputs spaced $j_t\delta$ apart and, hence, has weight at most $j_t\omega$. Assume now that $n \geq 2\delta d$. We can map any pair of sequences $1 \leq i_1 < i_2 < \cdots < i_{w/2} \leq n - \delta\lfloor d/\omega \rfloor$ and $1 \leq h_1 < h_2 < \cdots <$

$h_{w/2} \leq \lfloor d/\omega \rfloor$ to such a codeword by setting $b_t = i_t + \delta h_{t-1}$ and $e_t = i_t + \delta h_t$, for $1 \leq t \leq w/2$, where $h_0 = 0$. Such a map can be shown to be injective by induction on t. By construction, each such codeword has indeed input weight w. Further, since $\sum_{1 \leq t \leq w/2} j_t = h_{w/2} \leq \lfloor d/\omega \rfloor$, the codeword has weight at most d. The number of (i_t) and (h_t) sequences satisfying the above conditions is $\binom{n-\delta\lfloor d/\omega \rfloor}{w/2}$ and $\binom{\lfloor d/\omega \rfloor}{w/2}$, respectively, and so

$$|C^w_{\leq d}| \geq \binom{n - \delta\lfloor d/\omega \rfloor}{w/2}\binom{\lfloor d/\omega \rfloor}{w/2} \geq \Theta(1)^w \frac{(n-\delta d)^{w/2} d^{w/2}}{w^w}. \tag{2}$$

Thus the lemma holds if $n \geq 2\delta d$. On the other hand, if $n \leq 2\delta d$ then, by replacing d with $\lfloor n/(2\delta) \rfloor$ and applying Eq. 2, we see that $|C^w_{\leq d}| \geq \Theta(1)^w (n/w)^w$. Hence the lemma also holds if $n \geq 2\delta d$. ■

Next we determine an upper bound for $|C_d|$, the number of codewords of weight d. Although much tighter bounds can be given we will only need the following rough estimate.

Lemma 3:

$$|C_d| \leq \Theta(1)^d \binom{n}{\lfloor d/d^* \rfloor},$$

where d^* is the free distance of the code.

Proof. We proceed in a similar fashion as before. Again we split C_d into codewords consisting exclusively of regular detours and those which contain one terminating detour. We denote the corresponding sets by R_d and T_d, respectively. An element of R_d consists of at most $\lfloor \frac{d}{d^*} \rfloor$ detours since each regular detour has weight at least d^*. An element of T_d on the other hand can have as many as $\lceil \frac{d}{d^*} \rceil$ detours. Hence,

$$\begin{aligned} |R_d| &= \sum_{t=1}^{\lfloor \frac{d}{d^*} \rfloor} \sum_{(d_1,\cdots,d_t):\sum d_i = d} |R_{(d_1,\cdots,d_t)}|, \\ |T_d| &= \sum_{t=1}^{\lceil \frac{d}{d^*} \rceil} \sum_{(d_1,\cdots,d_t):\sum d_i = d} |T_{(d_1,\cdots,d_t)}|. \end{aligned}$$

Let $C_{((d_1,b_1),\cdots,(d_t,b_t))}$ be the set of codewords consisting of t detours where the ith detour has weight d_i and starts at position b_i. Using a similar argument as in Lemma 1, one can show that there exists an injective map from $C_{((d_1,b_1),\cdots,(d_t,b_t))}$ into the set of binary $(d\eta)$-tuples. This shows that $|R_{((d_1,b_1),\cdots,(d_t,b_t))}|$, $|T_{((d_1,b_1),\cdots,(d_t,b_t))}| \leq |C_{((d_1,b_1),\cdots,(d_t,b_t))}| \leq \Theta(1)^d$. Since there are at most $\Theta(1)^d \binom{n}{\lfloor d/d^* \rfloor}$ starting values for at most $\lfloor d/d^* \rfloor$ regular detours and at most $\Theta(1)^d \binom{n}{\lfloor d/d^* \rfloor} d\eta$ starting values for at most $\lfloor d/d^* \rfloor$ regular detours and one terminating detour it follows that $|R_{(d_1,\cdots,d_t)}|$, $|T_{(d_1,\cdots,d_t)}| \leq \Theta(1)^d \binom{n}{\lfloor d/d^* \rfloor}$. The claim now follows since in each case the range of the summation is upper bounded by $\Theta(1)^d$. ■

B. Lower Bound on the Minimum Distance for Parallel Concatenated Codes

In this section we will prove Theorem 1.a. In words, we will show that the minimum distance of a parallel concatenated code with k branches grows at least like $n^{\frac{k-2}{k}}$ in the interleaving

length n. The proof is based on the following straightforward

Lemma 4: Let $\mathcal{P}(G,n,k)$ be the ensemble of parallel concatenated codes with encoder defined by the rational function G of degree m and k parallel branches. Let P be a random code chosen uniformly from $\mathcal{P}(G,n,k)$. Then for any $d \in \mathbb{N}$

$$\Pr\{|P_{\leq d}| > 0\} \leq \sum_{1 \leq w \leq d} \frac{|C^w_{\leq d}|^k}{\binom{n}{w}^{k-1}}.$$

Proof.

$$\begin{aligned}
\Pr\{|P_{\leq d}| > 0\} &\overset{(a)}{\leq} \sum_{1 \leq w \leq d} \Pr\{|P^w_{\leq d}| > 0\} \\
&\overset{(b)}{\leq} \sum_{1 \leq w \leq d} \Pr\{\exists x \in \mathrm{GF}(2)^n \times \{0\}^m : (\pi_i(x), \gamma(\pi_i(x))) \in C^w_{\leq d}\ \forall i, 1 \leq i \leq k\} \\
&\overset{(c)}{\leq} \sum_{1 \leq w \leq d} \sum_{x \in \mathrm{GF}(2)^n \times \{0\}^m} \Pr\{(\pi_i(x), \gamma(\pi_i(x))) \in C^w_{\leq d}\ \forall i, 1 \leq i \leq k\} \\
&\overset{(d)}{=} \sum_{1 \leq w \leq d} \sum_{x \in \mathrm{GF}(2)^n \times \{0\}^m : \mathrm{weight}(x) = w} \left(\frac{|C^w_{\leq d}|}{\binom{n}{w}} \right)^k \\
&= \sum_{1 \leq w \leq d} \frac{|C^w_{\leq d}|^k}{\binom{n}{w}^{k-1}}.
\end{aligned}$$

Hereby, (a) follows from an application of the union bound. To see (b) note that if $c \in P^w_{\leq d}$, $c = (x, \gamma(\pi_1(x)), \cdots, \gamma(\pi_k(x)))$ then necessarily $(\pi_i(x), \gamma(\pi_i(x))) \in C^w_{\leq d}\ \forall i, 1 \leq i \leq k$. Step (c) follows from another application of the union bound. For step (d) first note that $\Pr\{(\pi_i(x), \gamma(\pi_i(x))) \in C^w_{\leq d}\ \forall i, 1 \leq i \leq k\}$ is clearly zero if the weight of x is unequal to w. Hence, the sum over x can be restricted to elements of weight w. Further, for each branch the choice of the permutation π_i is uniform over all permutations and, hence, the probability that $(\pi_i(x), \gamma(\pi_i(x))) \in C^w_{\leq d}$ is equal to $\frac{|C^w_{\leq d}|}{\binom{n}{w}}$. The claim now follows since these permutations are chosen independently. ■

Proof of Theorem 1.a. We have

$$\begin{aligned}
&\Pr\{|P_{\leq d}| > 0\} \\
&\overset{(a)}{\leq} \sum_{1 \leq w \leq d} \frac{|C^w_{\leq d}|^k}{\binom{n}{w}^{k-1}} \\
&\overset{(b)}{\leq} \sum_{1 \leq w \leq d} \frac{\Theta(1)^w \left(\frac{nd}{w^2}\right)^{wk/2}}{\left(\frac{n}{w}\right)^{w(k-1)}} \\
&\leq \sum_{1 \leq w \leq d} \Theta(1)^w \left(n^{\frac{2-k}{2}} d^{\frac{k}{2}} \right)^w,
\end{aligned}$$

where (a) is Lemma 4 and (b) follows from Lemma 1 and the inequality $\binom{n}{w} \geq (n/w)^w$. It follows that for any $\epsilon > 0$, $\Pr\{|P_{\leq n^{\frac{k-2}{k}-\epsilon}}| > 0\} \stackrel{n\to\infty}{\to} 0$ and, hence, $\Pr\{|P_{\leq n^{\frac{k-2}{k}-\epsilon}}| = 0\} \stackrel{n\to\infty}{\to} 1$. ■

C. Upper Bound on the Minimum Distance for Parallel Concatenated Codes

In the previous section we showed that the minimum distance of a parallel concatenated code with k branches grows at least like $n^{\frac{k-2}{k}}$ in the interleaving length n. We will now prove the converse, namely that it can not grow faster than that.

On the one hand this analysis is made more difficult since for any two inputs x and x' the maps $\pi_i(x)$ and $\pi_i(x')$ are dependent. On the other hand, we are fortunate since it turns out that it suffices to focus on inputs of weight 2.

Let P be a code chosen uniformly from the ensemble $\mathcal{P}(G, n, k)$. Let $\pi = (\pi_1, \cdots, \pi_k)$ be the associated permutation. Without loss of generality we can always assume that π_1 is the identity. We will employ the so called second moment method to upper bound the probability that the code does not contain codewords of weight d.

Lemma 5: ([15, p. 19]) Let X be a random variable with finite mean $E[X]$ and finite second moment $E[X^2]$. Then $\Pr\{X = 0\} \leq \frac{E[(X-E[X])^2]}{E[X]^2}$.

Proof. $\Pr\{X = 0\} \leq \Pr\{|X - E[X]| \geq E[X]\} \leq \frac{E[(X-E[X])^2]}{E[X]^2}$, where the last step is the well-known Chebyshev inequality [16, p. 48]. ■

Proof of Theorem 1.b: Recall that ω is the weight of a detour corresponding to an input of two $1's$ spaced δ apart. Define $\mathcal{I} := \{1, \cdots, n - \delta\lceil\frac{d}{k\omega}\rceil\}$ and $\mathcal{J} := \{1, 2, \cdots, \lceil\frac{d}{k\omega}\rceil\}$. To simplify notation and without loss of essential generality, we will assume that $\frac{d}{k\omega}$ is an integer. Let E_{ij} be the event that $\{|\pi_l(i) - \pi_l(i + j\delta)|\}_{l=2}^{k} \subseteq \delta\mathcal{J}$, where $i \in \mathcal{I}$ and $j \in \mathcal{J}$. We claim that if E_{ij} occurs then the associated code has minimum distance no larger than d. To see this, look at the codeword corresponding to an input x of weight 2, where the first non-zero symbol is at position i and the second one at position $i + \delta j$. By definition, the permutations map these two non-zero inputs to positions which are a multiple of δ apart for all branches (recall that π_1 is the identity). Further, the total separation is at most $d\frac{\delta}{\omega}$ since the separation of the two 1's per branch is at most $\delta\frac{d}{k\omega}$. It follows from the remarks in Section I-B that the total weight is at most d. Define $X := \sum_{i\in\mathcal{I};j\in\mathcal{J}} \mathbf{1}_{E_{ij}}$. Then $\Pr\{|P_{\leq d}| = 0\} \leq \Pr\{X = 0\}$. In order to apply Lemma 5 we need to bound $E[X]$ and $E[X^2]$. We start by lower bounding $E[X]$. We have

$$\begin{aligned} E[X] &= E[\sum_{i\in\mathcal{I};j\in\mathcal{J}} \mathbf{1}_{E_{ij}}] \\ &= \sum_{i\in\mathcal{I};j\in\mathcal{J}} E[\mathbf{1}_{E_{ij}}] \\ &= \frac{d(n - \delta\frac{d}{k\omega})}{k\omega} E[\mathbf{1}_{E_{11}}]. \end{aligned}$$

There are $(n(n-1))^{k-1}$ ways in which the two non-zero bits can be mapped to the remaining

$k-1$ branches and out of those at least $\left(\frac{2d}{k\omega}(n-2\delta\frac{d}{k\omega})\right)^{k-1}$ have the property that $\{|\pi_l(1)-\pi_l(\delta)|\}_{l=2}^{k}\in\delta\mathcal{J}$. It follows that $E[X]\geq\frac{d(n-\delta\frac{d}{k\omega})}{k\omega}\left(\frac{2d(n-2\delta\frac{d}{k\omega})}{n(n-1)k\omega}\right)^{k-1}=\Theta(1)\frac{d^k}{n^{k-2}}$.

Next we upper bound $E[X^2]$. We have

$$\begin{aligned}
E[X^2] &= \sum_{i,l\in\mathcal{I};j,h\in\mathcal{J}} E[\mathbf{1}_{E_{ij}}\mathbf{1}_{E_{lh}}] \\
&= \sum_{|\{i,i+j\delta\}\cap\{l,l+h\delta\}|=2} E[\mathbf{1}_{E_{ij}}\mathbf{1}_{E_{lh}}] + \sum_{|\{i,i+j\delta\}\cap\{l,l+h\delta\}|=1} E[\mathbf{1}_{E_{ij}}\mathbf{1}_{E_{lh}}] + \\
&\quad \sum_{|\{i,i+j\delta\}\cap\{l,l+h\delta\}|=0} E[\mathbf{1}_{E_{ij}}\mathbf{1}_{E_{lh}}] \\
&\leq E[X]+\Theta(1)\frac{d^{2k}}{n^{2k-3}}+\left(\frac{n(n-1)}{(n-2)(n-3)}\right)^{k-1}E[X]^2.
\end{aligned}$$

The last step requires some explanation. Clearly,

$$\sum_{|\{i,i+j\delta\}\cap\{l,l+h\delta\}|=2} E[\mathbf{1}_{E_{ij}}\mathbf{1}_{E_{lh}}] = \sum_{i\in\mathcal{I};j\in\mathcal{J}} E[\mathbf{1}_{E_{ij}}] = E[X].$$

Next look at the case $|\{i,i+j\delta\}\cap\{l,l+h\delta\}|=0$. We first exhibit a simple upper bound on $\Pr\{E_{ij},E_{lh}\}$. We have

$$\begin{aligned}
&\Pr\{E_{ij},E_{lh}\} \\
\overset{(a)}{=} &\sum_{((i_2',j_2'),\cdots,(i_k',j_k'))} \Pr\{E_{ij},(i,i+j\delta)\overset{\pi}{\to}((i_2',j_2'),\cdots,(i_k',j_k')),E_{lh}\} \\
= &\sum_{((i_2',j_2'),\cdots,(i_k',j_k'))} \Pr\{E_{ij},(i,i+j\delta)\overset{\pi}{\to}((i_2',j_2'),\cdots,(i_k',j_k'))\}\cdot \\
&\Pr\{E_{lh}|(i,i+j\delta)\overset{\pi}{\to}((i_2',j_2'),\cdots,(i_k',j_k'))\} \\
\overset{(b)}{\leq} &\sum_{((i_2',j_2'),\cdots,(i_k',j_k'))} \Pr\{E_{ij},(i,i+j\delta)\overset{\pi}{\to}((i_2',j_2'),\cdots,(i_k',j_k'))\}\Pr\{E_{lh}\}\left(\frac{n(n-1)}{(n-2)(n-3)}\right)^{k-1} \\
= &\left(\frac{n(n-1)}{(n-2)(n-3)}\right)^{k-1}\Pr\{E_{ij}\}\Pr\{E_{lh}\}.
\end{aligned}$$

In step (a) we have used the law of total probability. Further, by

$$(i,i+j\delta)\overset{\pi}{\to}((i_2',j_2'),\cdots,(i_k',j_k'))$$

we mean that the permutation π_2 for the second branch maps the positions i and $i+j\delta$ into the positions i_2' and j_2', respectively, and so on. To see step (b) note the following. For each branch the permutation on the remaining $n-2$ components is still uniform. There are $\binom{n-2}{2}$ (unordered) ways of mapping the positions l and h into these remaining $n-2$ positions instead of $\binom{n}{2}$ which applies for the unconditioned case. How many of those will result in the event E_{lh}? Certainly not more than for the unconditioned case. Finally, the permutations for

each branch are chosen independently. From $\Pr\{E_{ij}, E_{lh}\} \leq \left(\frac{n(n-1)}{(n-2)(n-3)}\right)^{k-1} \Pr\{E_{ij}\}\Pr\{E_{lh}\}$ we conclude that $E[\mathbf{1}_{E_{ij}}\mathbf{1}_{E_{lh}}] \leq \left(\frac{n(n-1)}{(n-2)(n-3)}\right)^{k-1} E[\mathbf{1}_{E_{ij}}]E[\mathbf{1}_{E_{lh}}]$. The claim now follows by expanding the summation over the whole range. The case $|\{i, i+j\delta\} \cap \{l, l+h\delta\}| = 1$ can be dealt with in a similar manner and we will be brief. There are at most $\frac{4d^2}{\omega^2 k^2}(n - \delta\frac{d}{k\omega}) = \Theta(1)d^2 n$ such terms. For each such term condition on the fact that the common component is mapped to a particular position. Note that the permutation is still uniform on the remaining components. The claim then follows since at most $\left(\frac{4d^2}{\omega^2 k^2}\right)^{k-1} = \Theta(1)d^{2(k-1)}$ permutations out of all possible $((n-1)(n-2))^{k-1} = \Theta(1)n^{2(k-1)}$ result in $\mathbf{1}_{E_{ij}}\mathbf{1}_{E_{hl}} = 1$.

Hence,

$$\begin{aligned}
\Pr\{|P_{\leq d}| = 0\} &\leq \Pr\{X = 0\} \\
&\leq \frac{E[X^2] - E[X]^2}{E[X]^2} \\
&\leq \frac{1}{E[X]} + \frac{\Theta(1)\frac{d^{2k}}{n^{2k-3}}}{E[X]^2} + \left(\frac{n(n-1)}{(n-2)(n-3)}\right)^{k-1} - 1 \\
&= \Theta(1)\frac{n^{k-2}}{d^k} + \Theta(1)\frac{1}{n} + \left(\frac{n(n-1)}{(n-2)(n-3)}\right)^{k-1} - 1.
\end{aligned}$$

The second and the third term converge to zero when n approaches infinity. Finally, the first term approaches zero if d grows at least like $n^{\frac{k-2}{k}+\epsilon}$ for some $\epsilon > 0$. This proves that $\Pr\{|P_{\leq n^{\frac{k-2}{k}+\epsilon}}| = 0\} \stackrel{n\to\infty}{\to} 0$. ■

D. Lower Bound on the Minimum Distance for Serially Concatenated Codes

In this section we will prove Theorem 2.a. In words, we will show that the minimum distance of a serially concatenated code grows at least like $n^{\frac{d_o^*-2}{d_o^*}}$ in the interleaving length n where d_o^* is the free distance of the outer code. The proof is based on the following straightforward

Lemma 6: Let $\mathcal{S}(G_o, G_i, n)$ be the ensemble of serially concatenated codes with outer and inner encoder defined by the rational functions G_o and G_i, respectively. Let S be a random code chosen uniformly from $\mathcal{S}(G_o, G_i, n)$. Then for any $d \in \mathbb{N}$

$$\Pr\{|S_{\leq d}| > 0\} \leq \sum_{w=d_o^*}^{d} \frac{|C_w(G_o, n)||C^w_{\leq d}(G_i, 2(n+m_o))|}{\binom{2(n+m_o)}{w}}.$$

Proof. We have

$$\begin{aligned}
& \Pr\{|S_{\leq d}| > 0\} \\
= \; & \Pr\{\exists x \in \mathrm{GF}(2)^{2(n+m_o)} : x \in C(G_o, n) \text{ and } (\pi(x \cdot \{0\}^{m_i}), \gamma_i(\pi(x \cdot \{0\}^{m_i})) \in C_{\leq d}(G_i, 2(n+m_o))\} \\
\overset{(a)}{\leq} \; & \sum_{w=d_o^*}^{d} \sum_{\substack{x \in \mathrm{GF}(2)^{2(n+m_o)}, \\ \text{weight}(x)=w}} \Pr\{x \in C_w(G_o, n) \text{ and } (\pi(x \cdot \{0\}^{m_i}), \gamma_i(\pi(x \cdot \{0\}^{m_i})) \in C^w_{\leq d}(G_i, 2(n+m_o))\} \\
\overset{(b)}{=} \; & \sum_{w=d_o^*}^{d} \sum_{\substack{x \in \mathrm{GF}(2)^{2(n+m_o)}, \\ \text{weight}(x)=w}} \Pr\{x \in C_w(G_o, n)\} \Pr\{(\pi(x \cdot \{0\}^{m_i}), \gamma_i(\pi(x \cdot \{0\}^{m_i})) \in C^w_{\leq d}(G_i, 2(n+m_o))\} \\
= \; & \sum_{w=d_o^*}^{d} |C_w(G_o, n)| \frac{|C^w_{\leq d}(G_i, 2(n+m_o))|}{\binom{2(n+m_o)}{w}}.
\end{aligned}$$

Step (a) follows from an application of the union bound and in step (b) we have used the fact that the interleaver makes the two events independent. ■

Proof of Theorem 2.a:

$$\begin{aligned}
\Pr\{|S_{\leq d}| > 0\} \; & \overset{(a)}{\leq} \; \sum_{w=d_o^*}^{d} \frac{|C_w(G_o, n)||C^w_{\leq d}(G_i, 2(n+m_o))|}{\binom{2(n+m_o)}{w}} \\
& \overset{(b)}{\leq} \; \sum_{w=d_o^*}^{d} \Theta(1)^w \frac{\binom{n}{\lfloor w/d_o^* \rfloor} \left(\frac{nd}{w^2}\right)^{w/2}}{\binom{2(n+m_o)}{w}} \\
& \overset{(c)}{\leq} \; \sum_{w=d_o^*}^{d} \left(\Theta(1) d^{1/2} n^{-(1/2 - 1/d_o^*)}\right)^w.
\end{aligned}$$

Step (a) is Lemma 6 and (b) follows from Lemma 1 and Lemma 3. In (c) we have used the estimates $\binom{n}{\lfloor w/d_o^* \rfloor} \leq \Theta(1)^w \left(\frac{n}{w}\right)^{w/d_o^*}$ and $\binom{2(n+m_o)}{w} \geq \left(\frac{n}{w}\right)^w$. Choosing $d = n^{\frac{d_o^*-2}{d_o^*} - \epsilon}$ for some $\epsilon > 0$ it follows that $\Pr\{|S_{\leq n^{\frac{d_o^*-2}{d_o^*}-\epsilon}}| > 0\} \leq \Theta(1) \sum_{w=d_o^*}^{d} n^{-w\epsilon/2} \overset{n\to\infty}{\to} 0$ and, hence, $\Pr\{|S_{\leq n^{\frac{d_o^*-2}{d_o^*}-\epsilon}}| = 0\} \overset{n\to\infty}{\to} 1$. ■

E. Upper Bound on the Minimum Distance for Serially Concatenated Codes

We want to prove that the minimum distance of a serially concatenated code can not grow faster than $n^{\frac{d_o^*-2}{d_o^*}}$. Although Theorem 2.b is true in general, for simplicity of notation we will only consider the case of of even d_o^*. The case of odd d_o^* can be proved by similar methods by looking at words of weight $2d_o^*$.

As for the case of parallel concatenated codes we will use the second moment method. In particular we will use

Lemma 7: Let E_i, $i \in \mathcal{I}$, be events in a given probability space such that $\Pr\{E_i, E_j\} \leq (1+\epsilon)\Pr\{E_i\}\Pr\{E_j\}$ for all $j \in \mathcal{I}_i \subseteq \mathcal{I}$. If $X = \sum_{i\in\mathcal{I}} \mathbf{1}_{E_i}$ then $\Pr\{X=0\} \leq \frac{\max_{i\in\mathcal{I}}|\mathcal{I}_i^c|}{E[X]} + \epsilon$.

Proof. We need to prove that $E[X^2] \leq \max_{i\in\mathcal{I}}|\mathcal{I}_i^c|E[X] + (1+\epsilon)E[X]^2$. The claim then follows from Lemma 5. But

$$\begin{aligned}
E[X^2] &= \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}} E[\mathbf{1}_{E_i}\mathbf{1}_{E_j}] \\
&= \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}_i^c} E[\mathbf{1}_{E_i}\mathbf{1}_{E_j}] + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}_i} E[\mathbf{1}_{E_i}\mathbf{1}_{E_j}] \\
&\leq \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}_i^c} E[\mathbf{1}_{E_i}] + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}_i} E[\mathbf{1}_{E_i}]E[\mathbf{1}_{E_j}](1+\epsilon) \\
&\leq \sum_{i\in\mathcal{I}} |\mathcal{I}_i^c| E[\mathbf{1}_{E_i}] + (1+\epsilon)\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}_i} E[\mathbf{1}_{E_i}]E[\mathbf{1}_{E_j}] \\
&\leq \max_{i\in\mathcal{I}}|\mathcal{I}_i^c| \sum_{i\in\mathcal{I}} E[\mathbf{1}_{E_i}] + (1+\epsilon)\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}} E[\mathbf{1}_{E_i}]E[\mathbf{1}_{E_j}] \\
&= \max_{i\in\mathcal{I}}|\mathcal{I}_i^c| E[X] + (1+\epsilon)E[X]^2.
\end{aligned}$$

■

Proof of Theorem 2.b: We will now define the appropriate events E_i for our case. Let $c^* \in C(G_o, n)$ be a codeword of weight d_o^* starting at position 1. Define C^* to be the set of codewords containing c^* and all its shifts. More precisely, let c^* have representation $c^* = (x^*, \gamma_o(x^*))$, $x^* \in \mathrm{GF}(2)^{n+m_o}$, and let $\mathsf{S}^i x^*$ denote x^* with its first n components cyclically shifted i position to the right. Note that x^* corresponds to a detour of length at most $d_o^*\eta_o$ and, hence, for any $i \in \{0, \cdots, n - d_o^*\eta_o - 1\}$, $(\mathsf{S}^i x^*, \gamma_o(\mathsf{S}^i x^*))$ is also an element of $C_{d_o^*}(G_o, n)$. Define $c_i^* := (\mathsf{S}^i x^*, \gamma_o(\mathsf{S}^i x^*))$ and let $C^* := \{c_i^*\}_{i=0}^{n-d_o^*\eta_o-1}$.

Let E_i, $1 \leq i \leq |C^*|$, be the event that $\pi(c_i^* \cdot \underbrace{0\cdots 0}_{m_i\times}) \in C_{\leq d}^{d_o^*}(G_i, 2(n+m_o))$. Since the permutation π is chosen uniformly it follows that $\Pr\{E_i\}$ is independent of i and is simply given by $|C_{\leq d}^{d_o^*}(G_i, 2(n+m_o))|/\binom{2(n+m_o)}{d_o^*}$. For every $i \in \mathcal{I}$ let $\mathcal{I}_i$ be defined as $\mathcal{I}_i := \{j \in \mathcal{I} : |i-j| > d_o^*\eta_o\}$. Note that for a pair $(i,j) \in \mathcal{I}\times\mathcal{I}_i$ the two corresponding code words c_i^* and c_j^* do not overlap. We claim that for $(i,j) \in \mathcal{I}\times\mathcal{I}_i$,

$$\Pr\{E_i, E_j\} \leq \Pr\{E_i\}\Pr\{E_j\}\frac{\binom{2(n+m_o)}{d_o^*}}{\binom{2(n+m_o)-d_o^*}{d_o^*}} \leq \Pr\{E_i\}\Pr\{E_j\}(1 + \frac{d_o^*}{2(n+m_o-d_o^*)+1})^{d_o^*}.$$

The proof is very similar to the one used for parallel concatenated codes for the non-overlapping case and, hence, we will omit the details. Let $X = \sum_i \mathbf{1}_{E_i}$ and note that a code S has a non-zero code word of weight less or equal to d if its corresponding realization

of X is non-zero. Hence, applying Lemma 7

$$\begin{aligned}\Pr\{|S_{\leq d}| = 0\} &\leq \Pr\{X = 0\} \\ &\leq \frac{\max_{i\in\mathcal{I}} |\mathcal{I}_i^c|}{E[X]} + \epsilon \\ &\leq \frac{2d_o^*\eta_o}{E[X]} + (1 + \frac{d_o^*}{2(n+m_o-d_o^*)+1})^{d_o^*} - 1.\end{aligned}$$

The second term on the right hand side clearly tends to zero as n tends to infinity. Hence, it remains to bound the first term on the right. But

$$E[X] = |C^*|E[\mathbf{1}_{E_1}] = |C^*|\frac{|C^{d_o^*}_{\leq d}(G_i, 2(n+m_o))|}{\binom{2(n+m_o)}{d_o^*}} \geq \Theta(1)n\left(\frac{nd}{{d_o^*}^2}\right)^{d_o^*/2} n^{-d_o^*} = \Theta(1)n^{1-d_o^*/2}d^{d_o^*/2},$$

where the second before last step follows from Lemma 2. This shows that $\frac{2d_o^*\eta_o}{E[X]} \stackrel{n\to\infty}{\to} 0$ if $d = n^{\frac{d_o^*-2}{d_o^*}+\epsilon}$ for any $\epsilon > 0$. ■

IV. Acknowledgments

We would like to thank Joseph Boutros for initial stimulating discussions about this topic.

References

[1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding," in *Proceedings of ICC'93*, (Geneve, Switzerland), pp. 1064–1070, May 1993.

[2] S. Benedetto and G. Montorsi, "Unveiling turbo codes: some results on parallel concatenated coding schemes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 409–428, Mar. 1996.

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

[4] S. Benedetto and G. Montorsi, "Serial concatenation of interleaved codes: Performance analysis, design, and iterative decoding," *?*, vol. ?, p. ?, ? ?

[5] D. Divsalar, S. Dolinar, F. Pollara, and R. McEliece, "Transfer function bounds on the performance of turbo codes." TDA Progress Report 42-122, Communications Systems and Research Section, California Institute of Technology, 95.

[6] E. Telatar and R. Urbanke, "On the ensemble performance of turbo codes," in *1997 IEEE International Symposium on Information Theory*, (Ulm, Germany), p. 105, 1997.

[7] I. Sason and S. S. (Shitz), "On union bounds for random turbo codes." Department of Electrical Engineering, Technion, 97.

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

[9] R. D. Divsalar, S. Dolinar and F. Pollara, "Weight distributions for turbo codes using random and nonrandom permutations." TDA Progress Report 42-122, Communications Systems and Research Section, California Institute of Technology, 95.

[10] D. Divsalar and F. Pollara, "Turbo codes for deep-space communications." TDA Progress Report 42-120, Communications Systems and Research Section, California Institute of Technology, 95.

[11] R. Blahut, *Theory and Practice of Error Control Codes*. New York: Addison-Wesley, 1984.

[12] R. Lidl and H. Niederreiter, *Introduction to Finite Fields and Their Applications*. New York: Cambridge University Press, 1994.

[13] M. Hall, *Combinatorial Analysis*. New York: John Wiley & Sons, Inc., 1986.

[14] L. Ahlfors, *Complex Analysis*. New York: McGraw-Hill, Inc., 1979.

[15] J. Spencer, *Ten Lectures on the Probabilistic Method*. Philadelphia: CBMS 52, Siam, 1987.

[16] K. Chung, *A Course in Probability Theory*. New York: Academic Press, Inc., 1974.