Space-Time Turbo Codes: Decorrelation Properties and Performance Analysis for Fading Channels

Sriram Vishwanath, Wei Yu, Rohit Negi and Andrea Goldsmith
Department of Electrical Engineering
Stanford University, Stanford, CA 94305, USA
Email : *sriram,andrea@wsl.stanford.edu, weiyu,negi@leland.stanford.edu*

Abstract— **This paper studies the decorrelation property of the constituent codes in Space-Time Turbo Codes (STTCs), where different constituent codes are transmitted over different antennas. We show that the time correlation between constituent codes falls of with respect to STTC block length N as $1/\sqrt{N}$ for large N, and hence goes to zero as $N \to \infty$. We then use this result to obtain analytical expressions for the frame error rate (FER) of STTCs in flat fading. The decorrelation property of STTCs helps explain the exceptional performance of STTCs on slowly varying flat fading channels.**

I. Introduction

Although receiver diversity has long been popular, interest in multiple transmit antennas for fading channels has recently exploded due to the huge capacity increases promised by Foschini and Gans in [1]. Specifically, [1] demonstrates that for a system with n transmit and m receive antennas with perfect receiver side information (RSI) and independent complex Gaussian propagation coefficients, the capacity increases linearly with the minimum of m and n.

A desire to tap into this potential capacity gain has led to the development of space-time codes, studied in great detail in [2] and [3]. The codes proposed in these papers are based on trellis codes or orthogonal block codes. Recently, space-time turbo codes have been investigated [4] [5] and have shown good results. In particular, example codes simulated in [5] are 2.5 dB away from outage capacity at a frame error rate of 0.01[1]. An explanation for this near-capacity performance has not yet emerged. The good performance of turbo codes on AWGN and single-antenna fading channels is often attributed to the good distance spectrum properties of the code. But for space-time turbo codes (STTCs), we hypothesize that the decorrelation of the constituent codes plays a crucial role in explaining their good performance. In this paper, we find that the time correlation between the constituent codes of a STTC decreases as the block size (or equivalently, the interleaver length) increases. Thus, if the constituent codes of a STTC are transmitted over different transmit antennas, these signals are nearly orthogonal for large block length codes. The importance of this decorrelation was determined in [6] for the case of no side information (SI) at either the receiver or the transmitter. In particular, [6] concludes that the (Shannon) capacity-achieving code for a multiple-antenna system must have transmitted signals that are mutually orthogonal with respect to time among the transmitter antennas. In addition, the design criteria presented in [2] and [3] indicate that, with RSI, a good code design is obtained by transmitting nearly orthogonal sequences over the antennas. Thus, with or without RSI, near-orthogonality of sequences from different antennas is an important design criterion for space-time codes. After we explicitly quantify the decorrelation of constituent codes, we apply this property to find approximate analytic expressions for the performance of STTCs on block-fading channels with RSI.

The rest of the paper is organized as follows. The system model is described in Section II. In Section III we analyze the correlation between a Bernoulli 1/2 {+1,-1} sequence and its randomly permuted version. In Section IV these results are used to characterize the correlation between the output sequences of the constituent encoders in the turbo code. The application of this correlation property to performance analysis of STTCs is discussed in Section V, followed by our conclusions in Section VI.

II. System Model

The structure of the space-time turbo encoder is shown in Figure 1(a). The turbo encoder consists of two identical recursive convolutional encoders (RCCs) connected in parallel, and hence it is also referred to as a parallel concatenated code. The RCCs are referred to as the constituent encoders. There is an interleaver[2] between the data stream $\{x_1, x_2, \ldots, x_N\}$ (denoted X) and the second encoder of length N. The interleaved sequence $\{y_1, y_2, \ldots, y_N\}$ is denoted Y. The data $\{x_1, x_2, \ldots, x_N\}$ are assumed to be Bernoulli 1/2 i.i.d. random variables. The RCCs are assumed to be rate one encoders, which means that they output N coded bits for every N input bits. We can relax this assumption to include a special class of STTCs that employ identical rate $1/n$ constituent encoders. This special class consists of those STTCs whose constituent encoders can be broken down into n *invertible* rate one sub-encoders working in parallel. By invertible we mean that each sub-encoder has a one-to-one map between the input data sequence and its output encoded sequence. The modulation scheme is assumed to be BPSK[3]. There are two transmit antennas in the system, and each constituent encoder output is modulated and sent through a different antenna. The channel between any transmit and receive antenna pair is modeled as an independent block-fading channel which is constant for one block length (N) and decorrelates between blocks. The fade coefficients are circularly symmetric com-

This work was partially supported by ONR Young Investigator award N00014-99-1-0578.

[1] The example code in question uses two identical 4 state constituent encoders and QPSK modulation, and is shown in Figure 2 of this paper.

[2] An interleaver in this context refers to a pseudo-random permuter.

[3] QPSK can also be used, since it consists of two orthogonal BPSK signals.

plex Gaussian random variables of unit variance. There is a suboptimal iterative turbo decoder at the receiver. The receiver may or may not have a channel estimator (RSI).

III. Decorrelation Property of Random Permutations

Before studying the decorrelation of constituent codes in turbo codes, we first analyze the intrinsic decorrelation property of random permutations.

Let $\{x_1, x_2, \cdots, x_N\}$ be a length-N i.i.d. Bernoulli sequence taking values on $\{+1, -1\}$ with equal probability, and let $\{y_1, y_2, \cdots, y_N\}$ be a permutation of $\{x_i\}$. The correlation between $\{x_i\}$ and $\{y_i\}$ is defined to be:

$$R_N(\pi, \{x_i\}) = \frac{1}{N}\sum_{i=1}^{N} x_i y_i, \tag{1}$$

where π denotes the permutation. The correlation depends on the sequence $\{x_i\}$ and the permutation π. The aim is to show that for large N, a random permutation produces $\{y_i\}$ which is uncorrelated with $\{x_i\}$. To this end, we characterize the asymptotic behavior of the ensemble average correlation over all permutations: $\lim_{N\to\infty} \mathbf{E}|R_N|$, where the expectation is taken over all i.i.d. sequences $\{x_i\}$ and all permutations $\pi \in \mathcal{P}$, where $\mathcal{P}$ denotes the set of all possible permutations, and every permutation in $\mathcal{P}$ is assumed to be equally likely.

First, let us consider what we expect the limiting behavior to be. A typical permutation randomly scrambles the input bits. So, we expect $\{y_i\}$ to have the same input distribution as $\{x_i\}$, and to be uncorrelated with $\{x_i\}$. If $\{x_i\}$ and $\{y_i\}$ were independent Bernoulli, their correlation should go to zero as N goes to infinity. Moreover, by a Central Limit Theorem argument, it is not hard to see that the correlation goes down approximately as $1/\sqrt{N}$. In our case, $\{x_i\}$ and $\{y_i\}$ are not independent. In fact, given a permutation, one determines the other. Nevertheless, if we average over all permutations, the same kind of asymptotic behavior holds.

We now state the main result in this section. The proof of this theorem can be found in the appendix.

Theorem: Let $\{x_i\}$, $\{y_i\}$ and R_N be defined as in (1). Then, $\forall\ \epsilon > 0$ and large N, $\mathbf{E}|R_N| = O(N^{-\frac{1}{2}+\epsilon})$. In particular, $\lim_{N\to\infty} \mathbf{E}|R_N| = 0$.

Proof: The idea is to average over the absolute value of all correlations for all permutations using a counting argument. A rigorous proof is in the appendix.

IV. Decorrelation Property of Turbo Codes

We now turn to the decorrelation property of constituent codes for turbo codes. As the intuitive argument in the previous section indicates, a typical permutation randomly permutes the input sequence to produce another i.i.d. Bernoulli 1/2 sequence. Thus, the turbo code in Figure 1(a) can be modified to look like the one in Figure 1(b), where two independent Bernoulli 1/2 sequences are encoded by two convolutional codes. Since each convolutional code is an information lossless encoder, a Bernoulli 1/2 input will necessarily lead to a Bernoulli 1/2 output. A simple argument to see this is that the entropy of the sequence cannot decrease after encoding. Since Bernoulli 1/2 sequences have the maximum entropy, the output has to be a Bernoulli 1/2 sequence. This is shown more rigorously in [7]. Thus the output of the two encoders are again Bernoulli 1/2 sequences. Thus one output can be regarded as the permuted version of the other, and by the previous section, the expectation of the magnitude of the correlation between them is bounded by $1/\sqrt{N}$.

For a more precise argument, we realize that presenting a counting argument as done in proving the decorrelation of random permutations is difficult. Hence we present a simplified argument that uses the modified turbo encoder presented in Figure 1(b). Note that this figure treats the two constituent encoders as independent convolutional encoders, with independent inputs X and Y. Denoting the rate one constituent encoders by RCC_1 and RCC_2 and their outputs by $C_1 = c_1^1, c_1^2, \ldots, c_1^N$ and $C_2 = c_2^1, c_2^2, \ldots, c_2^N$ respectively, we observe that finding the time correlation between C_1 and C_2 is equivalent to finding the time correlation between the outputs of two identical and independent convolutional encoders. To characterize this time correlation we look at the expected value of the magnitude of the correlation over all input patterns X and Y, i.e $\mathbf{E}|C_N|$ where $|C_N| = \frac{1}{N}\sum_{t=1}^{N} c_1^t c_2^t$. To simplify computation,we use Jensen's inequality to get $[\mathbf{E}|C_N|]^2 \le \mathbf{E}[C_N^2]$. Hence, instead of dealing with $\mathbf{E}|C_N|$, we shall work with $\mathbf{E}[C_N^2]$. We characterize $\mathbf{E}[C_N^2]$ as follows: Since the two constituent encoders are identical , they have the same distance spectrum a_z, where z denotes Hamming weight. In other words, a_z is the number of codewords of the *second* constituent encoder (RCC_2) at a Hamming distance of z from any one codeword c of the first constituent encoder (RCC_1). Since an inter-codeword distance of z implies a time correlation of $N - 2z$, we find $E[C_N^2]$ to be

$$\mathbf{E}[C_N^2] = \frac{1}{2^N}\sum_z a_z \frac{(N-2z)^2}{N^2} \tag{2}$$

The expression above is exactly the same as finding $\mathbf{E}(\frac{(N-2z)^2}{N^2})$ for any one of the constituent codes, where the expectation is over all codewords of that constituent code, N is the block size and z is the Hamming weight of the output codeword. We can expand $\frac{(N-2z)^2}{N^2}$ as $1 - 4\frac{z}{N} + 4\frac{z^2}{N^2}$ and then equation (2) can be rewritten as

$$\mathbf{E}[C_N^2] = 1 - \frac{4}{N}\mathbf{E}(z) + \frac{4}{N^2}\mathbf{E}(z^2). \tag{3}$$

To evaluate the expressions above, we utilize an argument in [7] which shows that, for an information lossless binary filter, the filter output is a Bernoulli 1/2 {0,1} sequence if and only if the filter input is a Bernoulli 1/2 {0,1} sequence. Since the input sequence under consideration is a Bernoulli 1/2 sequence, and the constituent encoder is an invertible (and hence information lossless) encoder, we find that

$$\mathbf{E}(z) = \frac{N}{2}, \tag{4}$$

$$\mathbf{E}(z^2) = \frac{N}{4} + \frac{N^2}{4}. \tag{5}$$

Substituting the above expressions in equation (3) yields $\mathbf{E}[C_N^2] = \frac{1}{N}$, which along with Jensen's inequality yields

$$\mathbf{E}|C_N| \leq \frac{1}{\sqrt{N}} \tag{6}$$

for STTCs with rate one constituent encoders. For rate $1/n$ constituent encoders, we use the idea of n parallel rate one encoders to get

$$\mathbf{E}|C_N| \leq \frac{n}{\sqrt{N}}. \tag{7}$$

This completes the proof of the decorrelation property. We performed a Monte-Carlo simulation of the decorrelation of constituent codes in an example code to verify this result. The example coder used is the same as that of Figure 3 of [5] and is shown in Figure 2. It uses S-random interleavers[4] (marked as π in Figure 2) as the code interleavers, and QPSK modulation. From the simulation results in Figure 3, we not only observe that $1/\sqrt{N}$ is an upper bound on $\mathbf{E}|C_N|$, but also that $\mathbf{E}|C_N|$ falls off as $CN^{-1/2}$, where C is a constant. In this particular case, the best fit was obtained using $C = 0.77$.

V. Performance of STTCs on Block-Fading Channels

An important application of the decorrelation property is in finding the error-rate performance of STTCs on block-fading channels with m receive antennas and perfect receiver side information. To this end, we proceed as follows: Let us denote the complex Gaussian channel gains between transmit antenna i and receive antenna j by $\alpha_{i,j}$. Then the probability that code sequences $c_1 = \{c_1^t\}$ and $c_2 = \{c_2^t\}; t = 1, 2, \ldots, N$ are transmitted and the receiver decides in favor of $e_1 = \{e_1^t\}$ and $e_2 = \{e_2^t\}; t = 1, 2, \ldots, N$ is given by [2]

$$P(e/c, \alpha_{i,j}, i = 1, 2; j = 1, \ldots, m) = \frac{1}{2} e^{(-d^2(c,e)\frac{E_s}{4N_0})} \tag{8}$$

where $e = e_1; e_2$, $c = c_1; c_2$, E_s/N_0 is the symbol energy to noise ratio and

$$d^2(c, e) = \sum_{j=1}^{m} \sum_{t=1}^{N} |\alpha_{1,j}(c_1^t - e_1^t) + \alpha_{2,j}(c_2^t - e_2^t)|^2.$$

The decorrelation characteristics of the constituent codes indicate that there exists at least one interleaver for which the expected correlation taken over all codewords is less than $1/\sqrt{N}$. To make the probability of error expressions tractable, we make the approximation

$$d^2(c, e) \approx \sum_{j=1}^{m} \sum_{t=1}^{N} |\alpha_{1,j}|^2 (c_1^t - e_1^t)^2 + |\alpha_{2,j}|^2 (c_2^t - e_2^t)^2. \tag{9}$$

A union bound on the frame error rate (FER) of STTCs for a particular value of channel SNR ($\alpha_{i,j}, i = 1, 2; j = 1, \ldots, m$) can be obtained by summing (8) over all possible decoded sequences e given a particular transmitted sequences c as[5]

$$P_f \leq \sum_e P(e/c, \alpha_{i,j}), \tag{10}$$

We can obtain the overall frame error rate on the fading channel $\overline{P_f}$ by averaging P_f over the distribution of $\alpha_{i,j}, i = 1, 2; j = 1, \ldots, m$. However, we realize that the union bound is very weak for low SNR values, i.e when any of the $\alpha_{i,j} \leq K$, where K is a suitably chosen threshold value. Thus, we use two different techniques for estimating P_f at low SNR (i.e for $0 \leq \alpha_{i,j} \leq K$). First, we set P_f equal to 0.5 (call it case 1), and second, we approximate it by that of uncoded transmission(case 2). Thus the average frame error rate $\overline{P_f}$ is given in the two cases by

$$\overline{P_f} = 1/2 * (1 - e^{-Km}) + \sum_{w,h,z} \frac{a_{w,h} a_{w,z}}{\binom{N}{w}} \left(\frac{e^{-K(1+hE_s/N_0)}}{1 + E_s/N_0 h} \frac{e^{-K(1+zE_s/N_0)}}{1 + E_s/N_0 z} \right)^m \tag{11}$$

for case 1 and

$$\overline{P_f} = 1/2 * \frac{1 - e^{-k_1 m(1+2E_s/N_0)}}{(1 + 2E_s/N_0)^m} + \sum_{w,h,z} \frac{a_{w,h} a_{w,z}}{\binom{N}{w}} \left(\frac{e^{-K(1+hE_s/N_0)}}{1 + E_s/N_0 h} \frac{e^{-K(1+zE_s/N_0)}}{1 + E_s/N_0 z} \right)^m \tag{12}$$

for case 2, where $a_{w,z}$ is the number of codewords of any one constituent encoder of input weight w and output weight z, and K is picked in each case such that the $\overline{P_f}$ is minimized. Note that in both (11) and (12), the first term dominates at low SNR and the second at high SNRs. Note also that we have used the *uniform interleaver argument* in the above expressions. Similar expressions can be obtained for the bit error rate. Although this result is for very large values of N, we can use this approximation for STTCs with practical values of N. As an example, we apply this result to the STTC given in Figure 2, which has a blocklength of $N = 260$ and one receive antenna (m=1). Simulation results are compared with the approximations in Figure 4. We find that the curve obtained in case 1 is an upper bound on the performance of STTCs, but is weak for large values of average received SNR, while case 2 provides a good approximation of the simulation results for midrange SNR values.

VI. Conclusions

We characterize the decorrelation of random permutations and of the constituent codes in STTCs. We find that the correlation decreases as $1/\sqrt{N}$, where N is the interleaver size. We then apply this to obtain computable approximations for the error rates of STTCs in a block-fading channel.

[4] S-Random interleavers are a particular class of good interleavers used in turbo codes.

[5] Please refer to [8] for details on the union bounding technique.

Appendix

Theorem: Let $\{x_i\}$, $\{y_i\}$ and R_N be defined as in (1). Then, $\forall\ \epsilon > 0$ and large N, $\mathbf{E}|R_N| = O(N^{-\frac{1}{2}+\epsilon})$. In particular, $\lim_{N\to\infty} \mathbf{E}|R_N| = 0$.

Lemma: Let $0 < x < 1$, then:

$$\left(\frac{n}{n-n^x}\right)^{n-n^x} \cdot \left(\frac{n}{n+n^x}\right)^{n+n^x} \le e^{-\frac{n^{2x}}{2(n+n^x)}} \tag{13}$$

Proof of Lemma: We will first establish the following fact. For $y \ge 1$,

$$\left(1-\frac{1}{y}\right)^y \le \frac{1}{e}\left(1-\frac{1}{2y}\right). \tag{14}$$

The right hand side of the inequality is the first two terms of the Taylor expansion of $f(z) = (1-z)^{1/z}$ at $z = 0$. The concavity of $f(z)$ ensures that the inequality holds. Then,

$$\begin{aligned}
&\left(\frac{n}{n-n^x}\right)^{n-n^x} \cdot \left(\frac{n}{n+n^x}\right)^{n+n^x} \\
\le\ & \left[e \cdot \frac{1}{e}\left(1-\frac{1}{2(n-n^x)/n^x}\right)\right]^{n^x} && (15)\\
\le\ & e^{-\frac{n^{2x}}{2(n+n^x)}}, && (16)
\end{aligned}$$

where (14), and $(1+1/y)^y \le e$ for $y > 0$ is used in (15). And $(1-1/y)^y \le 1/e$ for $y > 1$ is used (16). □

Proof of Theorem: The basic idea is to count over all permutations. First, let's restrict the set of $\{x_i\}$'s to sequences which contain equal number of +1's and −1's. Averaging over all such $\{x_i\}$'s and all permutations is equivalent to fixing one such $\{x_i\}$ and averaging over all permutations for this sequence. Let the length of $\{x_i\}$ be N. Since we are interested in the limiting behavior, we can let N be an integer multiple of 4. For notational convenience, let $n = N/4$.

The total number of permutations is $(4n)!$. First, count how many permutations produce zero correlation. In a sequence with $(2n)$ +1's and $(2n)$ −1's, to produce zero correlation, exactly (n) +1's need to be mapped to +1's, and the other (n) +1's need to be mapped to −1's. One way to count the number of all such permutations is to note that there are $\begin{pmatrix} 2n \\ n \end{pmatrix}$ ways to choose (n) +1's from $(2n)$ +1's, there are $(2n)(2n-1)\cdots(n+1)$ ways to map (n) +1's to +1's, and same number of ways to map +1's to −1's, and finally $(2n)!$ ways to map −1's. The total number of permutations that produce zero correlation is:

$$\binom{2n}{n}\frac{(2n)!}{n!}\cdot\frac{(2n)!}{n!}\cdot(2n)! = \frac{(2n)!^4}{n!^4}.$$

In a similar manner, it is easy to see that the number of permutations which produces correlation equal to k/n (or $-k/n$) is $(2n)!^4/(n-k)!^2(n+k)!^2$. So,

$$\mathbf{E}|R_{4n}| = \sum_{k=1}^{n} 2\cdot\frac{k}{n}\cdot\frac{(2n)!^4}{(n-k)!^2(n+k)!^2}\cdot\frac{1}{(4n)!}, \tag{17}$$

where the expectation is taken over all permutations and all inputs sequences of equal number of +1's and −1's. The factor of 2 is included because correlation can be positive or negative.

We proceed to bound the above sum. Fix n, let ϵ be the smallest positive number such that $n^{\frac{1}{2}+\epsilon}$ is an integer. (Note, $\lim_{n\to\infty}\epsilon = 0$.) Let $x = \frac{1}{2}+\epsilon$. Then

$$\begin{aligned}
&\sum_{k=1}^{n}\frac{k}{n}\cdot\frac{(2n)!^4}{(n-k)!^2(n+k)!^2}\cdot\frac{1}{(4n)!}, \\
\le\ & \sum_{k=1}^{n^x-1}\frac{k}{n}\frac{(2n)!^4}{n!^2n!^2(4n)!} + \sum_{k=n^x}^{n}\frac{k}{n}\frac{(2n)!^4}{(n-n^x)!^2(n-n^x)!^2(4n)!} && (18)\\
\le\ & \frac{n^x(n^x-1)}{\sqrt{2\pi}n^{3/2}} + \\
& \frac{[(n^2+n)-(n^{2x}-n^x)]}{\sqrt{2\pi}(n^2-n^{2x})}\left(\frac{n}{n-n^x}\right)^{2(n-n^x)}\left(\frac{n}{n+n^x}\right)^{2(n+n^x)} && (19)\\
=\ & O(n^{-\frac{1}{2}+2\epsilon}) + O(n)\cdot O(e^{-n^{\epsilon}}) && (20)\\
=\ & O(n^{-\frac{1}{2}+2\epsilon}),
\end{aligned}$$

where the Stirling's approximation $n! \approx \sqrt{2\pi n}(n/e)^n$ is used in (19), and the lemma is used in (20). Thus, we have shown that the theorem holds if $\{x_i\}$ is constrained to have equal number of +1's and −1's. In the remainder of the proof, we work to remove this constraint.

First, there will be a slight change of notation. Without loss of generality, let's assume that $\{x_i\}$ has more +1's than −1's. The distribution of correlations is skewed to positive, so we only need to bound the positive correlation. Consider all $\{x_i\}$'s with exactly $(2k)$ more +1's than −1's. Let $(n+k)$ be the number of +1's, and $(n-k)$ be the number of −1's in the sequence. Note that $N = 2n$. First, choose any $(2k)$ +1's out of $(n+k)$ +1's. We can group all permutations according to how these $(2k)$ +1's are mapped. The observation is that the set of permutations where *all* of these $(2k)$ +1's are mapped to +1's has the highest average correlation within the set. So, the ensemble correlation over all permutations is bounded by the ensemble correlation over the smaller set where $(2k)$ +1's are mapped exclusively to +1's. For sequences in this smaller set, excluding $(2k)$ +1's, the rest of the sequence has equal number of +1's and −1's, so its average correlation is $O((n-k)^{-\frac{1}{2}+\epsilon})$. Then, for the entire sequence,

$$\mathbf{E}|R_{2n}| \le \frac{2(n-k)}{2n}\cdot O\left(\frac{1}{(n-k)^{\frac{1}{2}-\epsilon}}\right) + \frac{2k}{2n}, \tag{21}$$

where the expectation is over all permutations for $\{x_i\}$'s with exactly $2k$ more +1's than −1's. Such a sequence has probability of occurrence equal to $\begin{pmatrix} 2n \\ n-k \end{pmatrix} 2^{-2n}$. So, taking expectation over all possible $\{x_i\}$'s,

$$\mathbf{E}|R_{2n}| \le \sum_{k=0}^{n}\binom{2n}{n-k}2^{-2n}\left[\frac{2(n-k)}{2n}O\left(\frac{1}{(n-k)^{\frac{1}{2}-\epsilon}}\right) + \frac{2k}{2n}\right].$$

The first term in the bracket on the right hand side is bounded by $O\left[\left(\frac{n-k}{n}\right)^{\frac{1}{2}+\epsilon} \cdot n^{-\frac{1}{2}+\epsilon}\right]$, so it is bounded by $O(n^{-\frac{1}{2}+\epsilon})$. To bound the second term asymptotically, we again use the Stirling's formula and the lemma. Fix n, let ϵ be the smallest positive number such that $n^{\frac{1}{2}+\epsilon}$ is an integer. Let $x = \frac{1}{2} + \epsilon$. Then

$$\begin{aligned} & \sum_{k=1}^{n} \frac{1}{2^{2n}} \frac{(2n)!}{(n-k)!(n+k)!} \frac{k}{n} \\ \leq \quad & \sum_{k=1}^{n^x - 1} \frac{1}{2^{2n}} \frac{(2n)!}{n!n!} \frac{k}{n} + \sum_{k=n^x}^{n} \frac{1}{2^{2n}} \frac{(2n)!}{(n-n^x)!(n+n^x)!} \frac{k}{n} \\ = \quad & O(n^{-\frac{1}{2}+2\epsilon}). \end{aligned}$$

Therefore, the ensemble average correlation over i.i.d. Bernoulli sequences of length N and over all permutations is $O(N^{-\frac{1}{2}+2\epsilon})$. In particular, it decreases to zero as $N \to \infty$. $\square$

References

[1] G.J. Foschini and M.J Gans , “On Limits of Wireless Communications in a Fading Environment when Using Multiple Antennas”, *Wireless Personal Communications*, pp 311-335, 1998.

[2] V. Tarokh, N.Seshadri and A.R. Calderbank, “Space-Time Codes for High Data Rate Wireless Communication : Performance Criterion and Code Construction”, *IEEE Trans. Inf. Theory*, pp 744-765, March 1998.

[3] Jiann-Ching Guey, Michael P.Fitz, Mark R.Bell and Wen-Yi Kuo, “Signal Design for Transmitter Diversity Wireless Communication Systems over Rayleigh Fading Channels”,*Proc. IEEE Vehicular Tech. Conf.*, pp 136-140, 1996

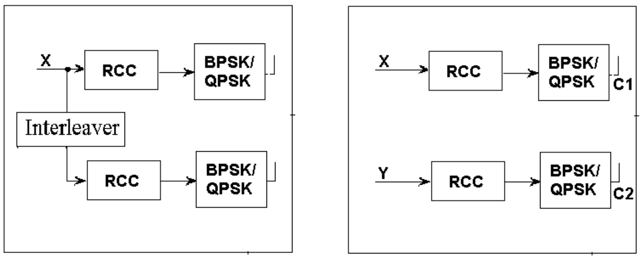
[4] Andrej Stefanov and Tolga M.Duman “Turbo Coded Modulation for Wireless Communications with Antenna Diversity”, *Proc. IEEE Vehicular Tech. Conf.*, pp. 1565-1569, 1999.

[5] Youjian Liu, Michael P.Fitz and Oscar Y.Takeshita “QPSK Space Time Turbo Codes”, *Proc. IEEE Int. Conf. Commun.*, pp.292-296, 2000.

[6] Thomas L. Marzetta and Bertrand M. Hochwald, “Capacity of a Mobile Multiple-Antenna Communication Link in Rayleigh Flat Fading”, *IEEE Trans. Inf. Theory*, pp. 139-157, January 1999.

[7] Philip Fire, “Boolean Operations on Binary Markov Chains”, Ph.D dissertation, Stanford University, April 1964.

[8] Sergio Benedetto and Guido Montorsi, “Design of Parallel Concatenated Convolutional Codes”, *IEEE Trans. Commun.*, pp.591-600, May 1996.



(a) Turbo Encoder (b) Modified Turbo Encoder

Fig. 1.

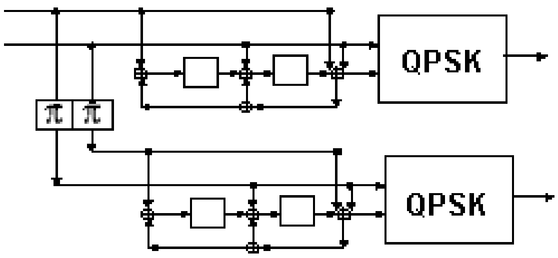


Fig. 2. Example turbo encoder with 4 state component code

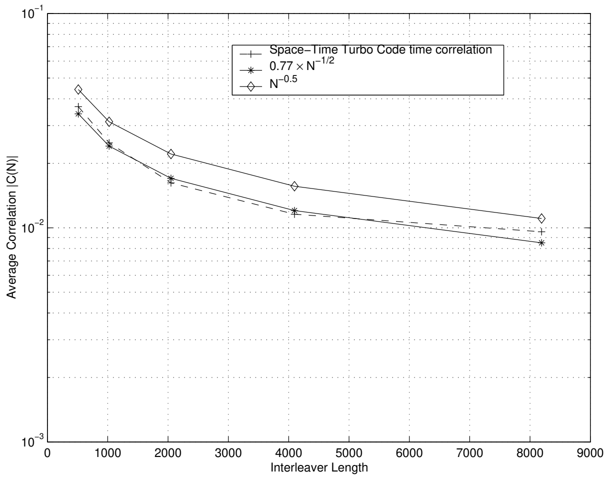


Fig. 3. Average correlation of an example turbo code compared with $1/\sqrt{N}$

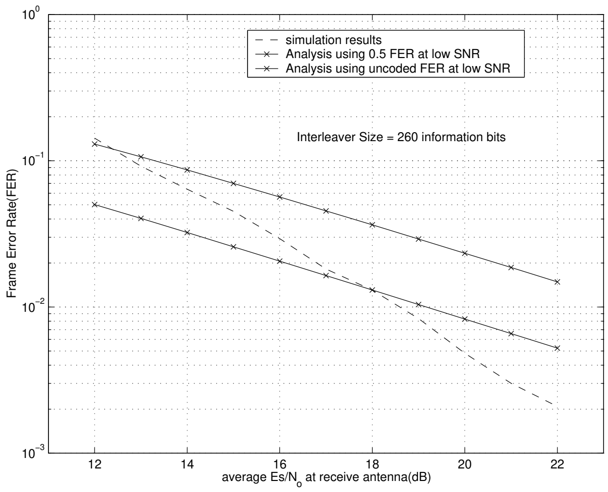


Fig. 4. Comparison between simulation results and analytic approximations for example encoder in Figure 4 with interleaver length 260