Asymptotic Effect of Interleaver Structure on the Performance of Turbo Codes

Mohammad Hadi Baligh and Amir K. Khandani

Coding & Signal Transmission Laboratory

Department of Electrical & Computer Engineering

University of Waterloo

Waterloo, Ontario, Canada, N2L 3G1

Technical Report UW-E&CE#2004-03

March 19, 2004

Asymptotic Effect of Interleaver Structure on the Performance of Turbo Codes

Mohammad Hadi Baligh and Amir K. Khandani
Coding & Signal Transmission Laboratory(www.cst.uwaterloo.ca)
Dept. of Elec. and Comp. Eng., University of Waterloo
Waterloo, ON, Canada, N2L 3G1
Tel: 519-8848552, Fax: 519-8884338
e-mail: {hadi, khandani}@cst.uwaterloo.ca

Abstract

Battail in [1] shows that an appropriate criterion for the design of long block codes is the closeness of the normalized weight distribution to a Gaussian distribution. A subsequent work shows that iterated product of single parity check codes satisfy this criterion [2]. Motivated by these earlier works, in the current article, we study the effect of the interleaver on the performance of Turbo codes for large block lengths, $N \to \infty$. A parallel concatenated Turbo code that consists of two component codes is considered. We demonstrate that for $N \to \infty$, the normalized weight of the systematic $\widehat{w_1} = \frac{w_1}{\sqrt{N}}$, and the parity check sequences $\widehat{w_2} = \frac{w_2}{\sqrt{N}}$ and $\widehat{w_3} = \frac{w_3}{\sqrt{N}}$ become a set of jointly Gaussian distributions for the typical values of $\widehat{w_i}, i = 1, 2, 3$, where the typical values of $\widehat{w_i}$ are defined as $\lim_{N \to \infty} \frac{\widehat{w_i}}{\sqrt{N}} \neq 0, 1$ for $i = 1, 2, 3$. To optimize the Turbo code performance in the waterfall region which is dominated by high-weight codewords, it is desirable to reduce ρ_{ij}, $i, j = 1, 2, 3$ as much as possible, where ρ_{ij} is the correlation coefficient between $\widehat{w_i}$ and $\widehat{w_j}$. It is shown that: (i) $\rho_{ij} > 0$, $i, j = 1, 2, 3$, (ii) $\rho_{12}, \rho_{13} \to 0$ as $N \to \infty$, and (iii) $\rho_{23} \to 0$ as $N \to \infty$ for "almost" any random interleaver. This indicates that for $N \to \infty$, the optimization of the interleaver has a diminishing effect on the distribution of high-weight error events, and hence, on the error performance in the waterfall region. In [3], it is shown that only certain low-weight codeword structures remain asymptotically probable.

We prove that for large block lengths, the number of low-weight codewords of these structure are some Poisson random variables. These random variables can be used to find the asymptotic probability mass function of the minimum distance of the Turbo code among all the possible interleavers. We find the mean and the variance of the union bound that is applied to the error floor region and study the effect of expurgating low-weight codewords (using the method proposed in [4] on the performance of Turbo codes.

I. INTRODUCTION

The advent of Turbo codes [5] is one of the most important developments in coding theory in many years. These codes can achieve a near Shannon-limit error correcting performance with a relatively simple decoding method. Turbo codes consist of some Recursive Convolutional Codes (RCCs) which are connected in parallel or serial through pseudo-random interleavers. Since the RCCs and also the interleaver has the linearity property[1], the resulting code is linear[2]. Consequently, the group property and distance invariance property hold.

Figure I presents a block diagram of an encoder of a rate 1/3 Turbo code with a block length N that is composed of two RCCs, where $b_1(i),\ i = 1, 2, \ldots, N$ are the systematic bits, and $b_2(i)$, $b_3(i)$ are the parity check bits. The weight of the code in Figure I is equal to the sum of the weights of the b_1, b_2 and b_3 sequences that are denoted by w_1, w_2, and w_3, respectively, over a block.

To decode a Turbo coded stream, an iterative method is used. A Turbo-decoder consists of two concatenated decoders, each using the received systematic stream and the corresponding received parity stream. Each decoder provides a soft output of transmitted bits by using the received data and the information provided by the other decoder. Repeating this procedure improves the estimation of the bit probability values.

[1]The effect of interleaving is equivalent to multiplying the input sequence by a permutation matrix which corresponds to a linear operation.

[2]This is based on neglecting the effect of the possible non-linearity caused by the method used to terminate the trellis.

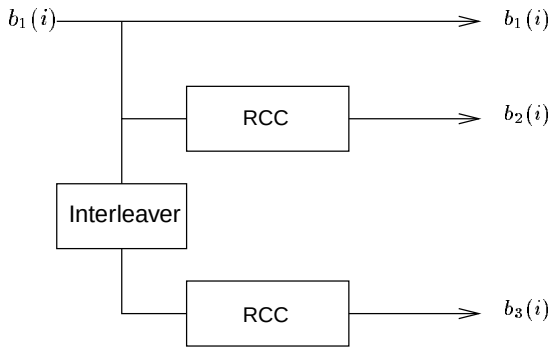

Fig. 1. Basic structure of the Turbo encoder.

One efficient algorithm for soft output decoding, based on the trellis diagram of the code known as the BCJR algorithm is presented in [6]. Another efficient soft decoding algorithm is derived from the coset decomposition principle in [7]. Also, there are some special methods for soft decoding such as sectionalized trellis diagrams [8] and the use of the codewords of the dual code [9].

A typical error performance of a Turbo code consists of two regions as illustrated in Figure 2. In the waterfall region, the error performance is determined by high-weight codewords, whereas in the error floor, the performance is determined by low-weight codewords.

Many researchers have concentrated on studying the weight distribution of Turbo codes and how these codes perform when maximum likelihood (ML) decoding is used. Although ML decoding is not feasible for Turbo codes, it provides insight into the performance of Turbo codes. Most of investigations focus on the average weight distribution of the Turbo codes among all the possible interleavers calculated in [10]. In [11], the asymptotic average weight distribution is calculated for large block lengths. In [12], according to the average weight spectrum, a simple approximation of the performance of parallel concatenated Turbo codes is obtained.

Using Gallager bounding techniques, [13]–[15] provide upper bounds on the performance of Turbo codes. The concept of average weight distribution has led to the

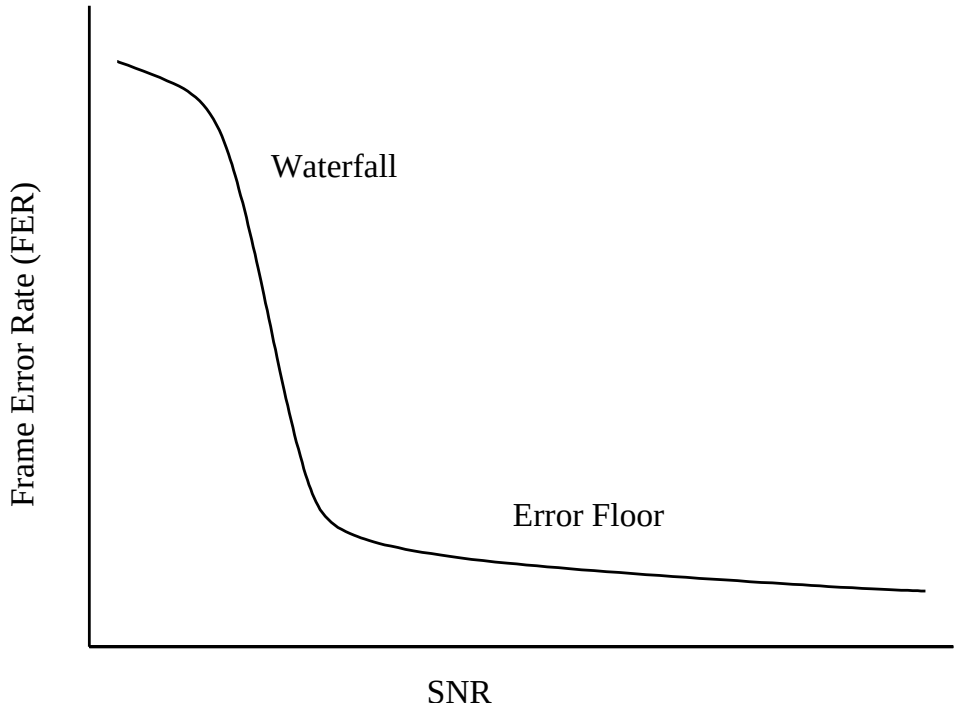

Fig. 2. Typical error performance of a Turbo code over an AWGN channel.

development of some results concerning the asymptotic performance of Turbo codes in [16].

In [17], [18], it is shown that Turbo codes belong to the class of weakly random-like codes; although their Frame Error Rate (FER) is poor, the Bit Error Rate (BER) remains low up to the neighborhood of the channel capacity. Reference [19] provides techniques to apply the channel coding theorem and the resulting error exponent, which was originally derived for random block-code ensembles, to ensembles of codes with fewer restrictive randomness requirements.

The structure and the number of low-weight codewords are studied in [3], [20], where it is reported that asymptotically probable low-weight codewords consist of one or more short error events and each event results from a weight two information sequence.

Some researchers have improved the performance of Turbo codes by optimizing the

interleaver structure. The effect of the chosen interleaver on the weight distribution for both low-weight and high-weight codewords is studied in [20]–[32]. These references provide some methods to design interleavers in order to decrease the number of low-weight codewords and/or to increase their weight to improve the performance of Turbo codes. These methods are more beneficial when the block length is relatively small. Reference [33] studies the design of nonsystematic Turbo codes to achieve higher minimum distances. The algorithm in [4] expurgates some low-weight codewords by injecting a zero in the lower-protected bit positions, and then punctures the resulting code to compensate for the loss in the effective code rate. In [34], the extrinsic information in the decoder is modified to exploit the source redundancy to enhance the system performance.

It is known that using a randomly chosen interleaver guarantees an excellent BER performance, but a certain number of low-weight codewords are generated, resulting in the appearance of an error floor and a small minimum distance. The effect of the interleaver structure on the minimum distance of the code is studied in [35]. References [36], [37] prove that the minimum Hamming distance of the Turbo codes cannot asymptotically grow at a rate higher than the third root of the codeword length. A systematic technique is introduced in [38] for obtaining sequences which are primary candidates for obtaining the minimum distance of parallel concatenated codes. The algorithm presented in [39] is improved by [40] and is applied to calculate the minimum distance of the Turbo codes. An interesting result is provided in [41] which denotes that for Low Density Parity Check (LDPC) code ensembles (which are closely related to Turbo codes) the capacity achieving codes do not have a large minimum distance. [1] shows that an appropriate criterion for the design of long block codes than the minimum Hamming distance is the closeness of the normalized weight distribution of the code to a Gaussian distribution. [2] substantiates this by showing that iterated-product codes have a weight distribution that is approximately Gaussian.

In [42], it is indicated that using more component codes improves the distance properties of the Turbo codes, resulting in a better performance when ML decoding is used. However, the sub-optimal iterative decoding does not perform well for multiple

component codes.

In this work, it is proved that the weights of the systematic and parity streams for their typical values, tend to a set of uncorrelated, and hence, independent, jointly Gaussian random variables for a randomly chosen interleaver and for any nontrivial recursive convolutional code. It is also demonstrated that any randomly chosen interleaver with the probability of one is the best interleaver in the waterfall region.

Low-weight codewords do not follow the Gaussian distribution and are more important in determining the performance of the code in the error floor (high SNR range). Unlike the waterfall, the optimization of the component codes and the interleaver affect the code performance in the error floor region. In [3], it is reported that as the block length increases, the low-weight codewords of a few special structures remain probable, and the expected number of low-weight codewords of each structure remains finite as the block length tends to infinity. In this paper, we show that the asymptotic probability mass function of the number of low-weight codewords of each structure is a Poisson random variable. The Poisson parameter of each structure is an increasing function of the systematic and parity weights. By means of these random variables, the probability mass function of the Turbo code minimum distance, and the mean and the variance of the union bound in the error floor region are calculated.

It is feasible to expurgate low-weight codewords, and thus, lower the error floor, because the number of low-weight codewords is small in comparison to the block length. Therefore, we discuss a method to expurgate the low-weight codewords following the method introduced in [4].

This article is organized as follows. In Section II, the effect of interleaver optimization to improve the waterfall region when the block length is large is examined. Section III concerns the interleaver optimization in the error floor region. In this part, we find the asymptotic statistical properties of the low-weight codewords and the asymptotic behavior of the error floor for large block Turbo codes.

II. Interleaver optimization for $N \to \infty$ in the waterfall region

It is assumed that the RCCs are generated by the transfer function $G(d) = N(d)/D(d)$. The impulse response of $G(d)$ is periodic with the period $P \leq 2^r - 1$, where r is the memory length of the code [43]. The main interest is in the group structure of the code-book, and also the periodicity property of the impulse response of $G(d)$. In this respect, we limit our attention to the structure of $D(d)$. This does not result in any loss of generality, because the group structure and also the periodicity property of the impulse response of $G(d)$ is not affected by the choice of $N(d)$. We consider a Turbo code with three output streams as reflected in Figure I. However, the discussions can be generalized to other configurations.

In general, the desire is that the period of the impulse response of $G(d)$ is as large as possible. If the period is equal to $2^r - 1$, the resulting impulse response is called a Maximum Length Sequence (MLS). For the rest of the paper, we assume that all the RCCs are MLS. The rules to determine all the possible configurations of $D(d)$ to obtain a maximum length sequence of period $2^r - 1$ (for the given r) are provided in [43]. It can be shown that any MLS-sequence satisfies the three postulates of randomness [43]. One consequence of this property is that in any period of an MLS-sequence, the number of ones is equal to 2^{r-1}, and the number of zeros is $2^{r-1} - 1$.

If the impulse response of $D(d)$ is considered to be a periodic sequence (started at infinity in the past), we obtain $P = 2^r - 1$ non-zero sequences which are time shifts of each other. Each sequence corresponds to a specific positioning of the impulse within the period. These sequences are referred to as different phases of the periodic signal. We assume that the different phases are labelled by integer numbers, say $1, \ldots, P$, where the label of a phase corresponds to the relative position of the corresponding impulse within the period. It can be shown that the set of phases of an MLS-sequence (plus the all-zero sequence) constitutes a group under binary addition [43]. The order of each element in this group is equal to two, indicating that the sum of each phase with itself results in the all-zero sequence (denoted as the zero phase).

Using the group property of phases, we conclude that the function of the numerator of $G(d)$ is to replace each phase with a linear combination of some other phases. This function is equivalent to a permutation (relabelling) of phases and does not play a role in the following discussions.

For the bit position k, $(k = 1, \ldots, N)$ within the i'th output stream, we refer to the set of systematic bit positions $j \leq k$ for which an impulse at position j results in a 1 at position k as $\mathcal{R}_i(k)$, $i = 1, 2, 3$. Obviously, $\mathcal{R}_1(k) = \{k\}$. If the bit position k is located in the L'th period, i.e., $L = \lceil k/P \rceil$, where $\lceil \, . \, \rceil$ denotes the ceiling function, then the number of positions belonging to $\mathcal{R}_i(k)$, $i = 2, 3$, within each of the periods $1, \ldots, L-1$ is equal to 2^{r-1} [43]. The number of positions within the L'th period (the period containing k itself) depends on the relative position of k within the L'th period and also on the numerator of $G(d)$. We are mainly interested in the large values of L (parity bits far from the boundaries) for which the effect of the elements within the L'th period itself is negligible. Thus, $|\mathcal{R}_2(k)| = |\mathcal{R}_3(k)| \simeq \lceil k/P \rceil 2^{r-1}$, where $|\, . \,|$ denotes the cardinality of the corresponding set.

The notation $b_i(k)$, $i = 1, 2, 3$, $k = 1, \ldots, N$, is used to refer to the k'th bit within the i'th output stream. Since each bit is zero or one with an equal probability, then $\overline{b_i(k)} = \overline{b_i^2(k)} = 1/2$.

To investigate the asymptotic weight distribution of Turbo codes, we show that $\widehat{w}_i = \frac{w_i}{\sqrt{N}}$, $i = 1, 2, 3$, referred to as the normalized weights, have a Gaussian distribution for their typical values when N is large. This is easily verified by noting that all the 2^N possible combinations within the three steams are equiprobable, and consequently, the positions within each of the three output streams are independent and identically distributed (iid) binary random variables (where zero and one are equally probable). Using the Central Limit Theorem, we conclude that $\widehat{w}_1$, $\widehat{w}_2$ and $\widehat{w}_3$, which are the normalized sum of N iid random variables, have a Gaussian distribution with mean $\sqrt{N}/2$ and variance $1/4$ for the large values of N.

In order to have a set of jointly Gaussian random variables, not only do the marginal weight distributions need to be Gaussian, but also the conditional distributions should

be Gaussian. When the systematic weight w_1 is known, the parity bits are no longer independent of each other, because only $\binom{N}{w_1}$ out of 2^N codewords represent a systematic weight of w_1, and hence, remain probable. Under these circumstances, the parity bits in each stream tend to be an m-*dependent* sequence and the Central Limit Theorem can still be applied. In the following, using the properties of m-dependent random variables, we show that the conditional weight distributions of $\widehat{w_2}$ and $\widehat{w_3}$ given $\widehat{w_1}$ are Gaussian for the typical values of $\widehat{w_1}$. As a result, noting that the marginal distributions are Gaussian, we can conclude that $\widehat{w_1}$, $\widehat{w_2}$ and $\widehat{w_3}$ are a set of jointly Gaussian random variables.

Definition: m-*dependent sequence*

A sequence $X_1, X_2, \ldots$ of random variables is called m-dependent if and only if $\{X_{a-r}, X_{a-r+1}, \ldots, X_a\}$ and $\{X_b, X_{b+1}, \ldots, X_{b+s}\}$ are independent sets of variables when $b - a > m$ [44]; that is, an m-dependent sequence is a sequence of dependent random variables for which the dependency lasts, at most, for m elements.

Theorem: *Central Limit Theorem for the sum of dependent random variables*

If $X_1, X_2, \ldots$ is a sequence of m-dependent, uniformly bounded random variables and $S_n = X_1 + X_2 + \cdots + X_N$, with the standard deviation V_N. Then, if $V_N/N^{1/3} \to \infty$ as $N \to \infty$, $\overline{G}_N(x) \to \Phi(x)$ for all x, as $N \to \infty$, where $\overline{G}_N$ is the cumulative distribution function (cdf) of $\{S_N - E(S_N)\}/V_N$ [44].

As indicated in the theorem, if the standard deviation of the sum of N consecutive elements of a stream of m-dependent random variables grows faster than the third root of their number, the Central Limit Theorem can still be applied. In order to apply this theorem on the conditional weight distributions, we prove the following proposition.

Proposition: Assuming the systematic weight is w_1, each parity stream is an m-dependent sequence, and the variance of its weight is given by

$$\sigma^2_{w_2|w_1} = \frac{N}{4}\left(1 + \frac{2(1-\frac{2w_1}{N})^{(P+1)/2}}{1-(1-\frac{2w_1}{N})^{(P+1)/2}}\right). \tag{1}$$

Proof: See Appendix. ■

With this proposition and the Central Limit Theorem for the dependent variables, the conditional parity weight distributions, when the systematic weight is given, asymp-

totically become Gaussian distributions. A similar approach is valid for the conditional weight distribution of $\widehat{w_3}$, given $\widehat{w_1}$ and $\widehat{w_2}$. As a result, $\widehat{w_1}$, $\widehat{w_2}$ and $\widehat{w_3}$ are a set of jointly Gaussian random variables, since their marginal and conditional distributions are Gaussian.

A set of jointly Gaussian random variables can be completely described by their mean vector and covariance matrix. The marginal mean and variance of $\widehat{w_1}$, $\widehat{w_2}$ and $\widehat{w_3}$ are $\sqrt{N}/2$ and $1/4$, respectively. The correlation coefficients between $\widehat{w}_i$ and $\widehat{w_j}$ denoted as ρ_{ij}, $i, j = 1, 2, 3$, can be written as

$$\rho_{ij} = \frac{\overline{\widehat{w}_i \widehat{w}_j} - \overline{\widehat{w}_i}\ \overline{\widehat{w}_j}}{\sigma_{\widehat{w}_i}\sigma_{\widehat{w}_j}} = 4\left[\overline{\widehat{w}_i \widehat{w}_j} - \frac{N}{4}\right], \tag{2}$$

and

$$\overline{\widehat{w}_i \widehat{w}_j} = \frac{1}{N}\sum_m \sum_n \overline{b_i(m) b_j(n)}, \tag{3}$$

where the expectation is taken over all the possible 2^N combinations of the input. The total weight of the output sequence is equal to $\widehat{w} = \widehat{w}_1 + \widehat{w}_2 + \widehat{w}_3$ which has a Gaussian distribution of the mean,

$$\mu_{\widehat{w}} = 3\frac{\sqrt{N}}{2}, \tag{4}$$

and the variance,

$$\sigma^2_{\widehat{w}} = \frac{3 + 2\rho_{12} + 2\rho_{13} + 2\rho_{23}}{4}. \tag{5}$$

Noting that sequences with a weight smaller than the mean value result in higher probabilities of error as compared to sequences with a weight larger than the mean, we conclude that the main objective in the code design is to sharpen the peak of the Probability Density Function (pdf) of the normalized Hamming weight $\widehat{w}$ which is equivalent to minimizing the pdf variance. This is equivalent to minimizing the ρ_{ij} coefficients. In the following, we first show that the $\rho_{ij} \geq 0$; therefore, the minimum possible value for the correlation coefficients is zero. When the block length increases, $\rho_{1j},\ j = 2, 3$ become zero for any nontrivial RCC. Also, ρ_{23} tends to zero with the probability of one for the randomly chosen interleavers. Consequently, the asymptotic weight distribution by using a randomly chosen interleaver is optimized with the probability of one.

Theorem: $\rho_{ij} \geq 0$ for $i, j = 1, 2, 3$.

Proof: Any of the pairs $b_i(m)$, $b_j(n)$ for $i, j = 1, 2, 3$ and $m, n = 1, \ldots, N$, can take four different values; namely, $\{00, 01, 10, 11\}$. The set of the input sequences that results in the value of 00 form a sub-group of all the possible 2^N input combinations. This is a direct consequence of the linearity and the group property of the code. Due to the group property of the set of corresponding coset leaders, two situations can occur. There is either only one coset with the coset leader 11, or there are three cosets with the coset leaders 01, 10 and 11. The important point is that in both of these cases, the 00 sub-group and its cosets contain the same number of input sequences. Therefore, for the probability of the pair $b_i(m)$, $b_j(n)$, the following two cases exist:

Case I: $b_i(m), b_j(n)$ take the values $00, 11$, each with the probability of $1/2$, resulting in $\overline{b_i(m)b_j(n)} = 1/2$, so that

$$\overline{b_i(m)b_j(n)} - \overline{b_i(m)}\ \overline{b_j(n)} = \frac{1}{4}. \tag{6}$$

Case II: $b_i(m), b_j(n)$ take the values $00, 01, 10, 11$, each with the probability of $1/4$, resulting in $\overline{b_i(m)b_j(n)} = 1/4$, so that

$$\overline{b_i(m)b_j(n)} - \overline{b_i(m)}\ \overline{b_j(n)} = 0. \tag{7}$$

The important point is that in both cases, we have

$$\overline{b_i(m)b_j(n)} - \overline{b_i(m)}\ \overline{b_j(n)} \geq 0. \tag{8}$$

This indicates that the correlation coefficients ρ_{ij}, $i, j = 1, 2, 3$ are always nonnegative. ■

Theorem: $\rho_{12}, \rho_{13} \to 0$ as $N \to \infty$.

Proof: For ρ_{12} and ρ_{13} (the interaction of the systematic stream with each of the parity checks), Case II in the previous two cases is valid, resulting in $\rho_{12}, \rho_{13} \to 0$ as $N \to \infty$. Note that $b_1(m)$ and $b_2(n)$ are independent of each other, if $b_1(m)$ is not mapped (through interleaving) to a bit position within $\mathcal{R}_2(n)$, or if $\mathcal{R}_2(n)$ contains at

least two elements. This is valid except for some trivial cases which have a vanishing effect on the overall result. ■

Theorem: $\rho_{23} \to 0$ for $N \to \infty$ with the probability of one (for almost any random interleaver).

Proof: If $\mathcal{R}_2(m)$ differs from $\mathcal{R}_3(n)$, even by one bit position, then $b_2(m)$ and $b_3(n)$ are independent of each other. This results in $\overline{b_2(m)b_3(n)} = \overline{b_2(m)}\ \overline{b_3(n)} = 1/4$. This is the case, unless $|m-n| < P/2$, and the elements of $\mathcal{R}_2(m)$ and $\mathcal{R}_3(n)$ contain the same input bits (before and after interleaving). Consequently, the corresponding interleaver has a restriction on the mapping of the many bit positions. Obviously, the fraction of such interleavers tends to zero as $N \to \infty$. Therefore, for almost any random interleaver, $\rho_{23} \to 0$ as $N \to \infty$. ■

As a result, the typical weight distribution of Turbo codes is not a function of the chosen RCC and interleaver (for nontrivial RCCs and interleavers), and hence, the interleaver optimization has a diminishing effect on the asymptotic performance of the Turbo code in its waterfall region.

The Gaussian weight distribution approximation is valid for the typical values of the Hamming weight. The number of low-weight codewords cannot be approximated by a continuous distribution, and as we will see in the next section, low-weight codewords appear only in certain structures and for each of these structures, their number is a Poisson random variable. However, as the SNR increases, the error performance is determined by codewords of lower weights. In the Appendix, in order to provide insight into the range of the SNR for which codewords of typical weights are dominant, we apply the union bound on the weight distribution to find the dominant weight in the error performance. Also, the cutoff rate which is based on applying the union bound on the weight distribution is calculated according to this assumption and compared to the random coding cutoff rate.

III. Interleaver optimization for $N \to \infty$ in the error floor region

A. Asymptotic behavior of low-weight codewords

The error floor is caused by low-weight codewords. The number of low-weight codewords and their weights are determined by the RCC and the interleaver structure. To have an approximation of the mean and the variance of the error floor among all the possible interleavers, the statistical properties of the low-weight codewords are required.

The probable low-weight codewords for large block lengths consist of some short error events[3] with the systematic weight of two in both the RCCs [3]. These short error events are caused by two nonzero systematic bits that are separated by an integer multiple of the RCC impulse response period. In other words, an asymptotically probable codeword has an even systematic weight of $w_1 = 2M$. Each RCC leaves the all-zero state M times. This is equal to at least M repetitions of the RCC impulse response in each encoder. This phenomenon produces $\frac{K(P+1)}{2}$ nonzero parity bits, where $K \geq 2M$ is the number of RCC impulse response repetitions in the parity check sequences. We call this structure of low-weight codewords as the structure of type (M, K), where $K \geq 2M$.

To calculate the mean and the variance of the error floor, it is necessary to compute the statistics of each low-weight structure. For structure type (M, K), there are, at most, $2M - K$ short error events with a duration of more than P. It is known that there are

$$\binom{K-1}{2M-1} \tag{9}$$

ways to choose $2M$ positive integer numbers with a summation of K. Equivalently, the structure of type (M, K) can be divided into $\binom{K-1}{2M-1}$ substructures, each having the same statistical properties as the codewords of type $(M, 2M)$. For the rest, we first, calculate the statistical properties of these codewords, and then generalize the result to the other structures.

There are

$$\binom{N}{M} \tag{10}$$

[3]A short error event means leaving the zero-state and returning back to it for the first time.

systematic inputs consisting of N pairs of nonzero bits separated by P. This can be easily verified by determining the place of the first element of each pair. The overlapping pairs are neglected, because $N \gg M$. Such a structure produces $M(P+1)/2$ bits in the first convolutional encoder. There are the same number of parity bits in the second convolutional encoder, if the interleaver maps that systematic stream to another stream of the same structure. There are $\binom{N}{2M}$ possible ways to interleave a stream of the weight $2M$. However, among them, only $\binom{N}{M}$ result in M pairs of nonzero bits separated by P; that is, a suspected low-weight stream changes to another one with the probability of

$$\frac{\binom{N}{M}}{\binom{N}{2M}}. \tag{11}$$

If (11) is multiplied by (10), the average number of these low-weight codewords is

$$\frac{\binom{N}{M}^2}{\binom{N}{2M}} \simeq \binom{2M}{M}. \tag{12}$$

In other words, we have a large number of binary random variables with a very small probability of success which, on the average, result in a nonzero finite number of low-weight codewords. These random variables are asymptotically independent because occupying a bit position in the interleaved stream by a certain bit does not asymptotically affect the probability for the other bits. As a result, the number of low-weight codewords is a Poisson random variable of parameter $\binom{2M}{M}$. If (12) is multiplied by (9), the Poisson parameter for the structure of type (M, K) is

$$\binom{2M}{M}\binom{K-1}{2M-1}. \tag{13}$$

As an example, there are approximately N codewords with the systematic weight of two, consisting of two nonzero bits separated by P. After interleaving, the distance between these two bits remains at P with the probability of $2/N$. This occurs because these two bits can occupy about $N^2/2$ different places after interleaving, and only N of the new places are separated by P. Then, the average number of low-weight codewords of this structure is two. On the other hand, there are four codewords with the systematic weight of 2 and parity weight of $3(P+1)/2$, averaged over all the possible interleavers,

because there are two possible structures for this situation: distance P before interleaving and $2P$ after interleaving, and vice versa, and the Poisson parameter for each of two substructures is two.

In a linear binary code-book, the binary addition of two or more low-weight codewords results in another low-weight codeword. The new codeword is decomposable when the original low-weight codewords do not have common nonzero bit positions. Decomposable codewords can be easily ignored, because the new codeword does not change the Voronoi region of the all-zero codeword and if each of the original low-weight codewords is expurgated, the decomposable codeword no longer exists. As a result, the Poisson parameters are more precise when only the indecomposable structures are counted. In the asymptotic weight distribution of Turbo codes, there are some low-weight codewords with a systematic weight greater than two which can be decomposed into smaller low-weight codewords. We assume that such a low-weight codeword has a systematic weight of $2M > 2$, and consists of some smaller low-weight codewords which can be partitioned to k_i codewords of the systematic weight $2i$ for $i = 1, \ldots, M-1$. Of course, k_i's are nonnegative integers that satisfy

$$\sum_{m=1}^{M-1} m k_m = M. \tag{14}$$

Again, only codewords of type $(M, 2M)$ are considered. The same approach that was previously applied is still valid for the codewords of type (M, K), when $K > 2M$. The distribution of such structures is a Poisson random variable, because it is the addition of a large number of identical, low-probable codewords. The total number of low-weight codewords of type $(M, 2M)$ is $\binom{2M}{M}$. It is assumed that the average number (Poisson parameter) of indecomposable codewords with the same type is λ_M. If the number of indecomposable low-weight codewords of type $(m, 2m)$ is X_m, then the number of decomposable codewords of type $(M, 2M)$ consisting of k_m, $m = 1, \ldots, M-1$ codewords of type $(m, 2m)$ is

$$\prod_{m=1}^{M-1} \binom{X_m}{k_m}. \tag{15}$$

X_m is a Poisson-distributed random variable with the parameter of λ_m. Therefore, the average number of decomposable codewords is

$$\binom{2M}{M} - \lambda_M = \sum_{\sum_{m=1}^{M-1} m k_m = M} E\left[\prod_{m=1}^{M-1} \binom{X_m}{k_m}\right] = \sum_{\sum_{m=1}^{M-1} m k_m = M} \prod_{m=1}^{M-1} E\left[\binom{X_m}{k_m}\right], \quad (16)$$

because different Poisson random variables are independent.

Lemma: The expected value of $\binom{X_m}{k_m}$, where X_m is a Poisson-distributed random variable of parameter λ_m is

$$E\left[\binom{X_m}{k_m}\right] = \frac{\lambda_m^{k_m}}{k_m!}. \quad (17)$$

Proof:

$$\begin{aligned} E\left[\binom{X_m}{k_m}\right] &= \sum_{k=k_m}^{\infty} \binom{k}{k_m} P\{X_m = k\} \\ &= \sum_{k=k_m}^{\infty} \frac{k!}{k_m!(k-k_m)!} e^{-\lambda_m} \frac{\lambda_m^k}{k!} \\ &= \frac{e^{-\lambda_m}\lambda_m^{k_m}}{k_m!} \sum_{k=k_m}^{\infty} \frac{\lambda_m^{k-k_m}}{(k-k_m)!} \\ &= \frac{e^{-\lambda_m}\lambda_m^{k_m}}{k_m!} \sum_{k=0}^{\infty} \frac{\lambda_m^k}{k!} \\ &= \frac{e^{-\lambda_m}\lambda_m^{k_m}}{k_m!} e^{\lambda_m} \\ &= \frac{\lambda_m^{k_m}}{k_m!}. \end{aligned} \quad (18)$$

■

Using this lemma, we can see that

$$\lambda_M = \binom{2M}{M} - \sum_{\sum_{m=1}^{M-1} m k_m = M} \prod_{m=1}^{M-1} \frac{\lambda_m^{k_m}}{k_m!} \quad , \quad \text{for } M > 1. \quad (19)$$

This recursive equation in conjunction with the fact that $\lambda_1 = 2$ yields the Poisson parameter for the indecomposable, low-weight codewords of the type $(M, 2M)$. With the same approach, the Poisson parameter of codewords of type (M, K) is the multiplication of λ_M by (9).

The new Poisson parameters lead us to find the asymptotic probability mass function (pmf) of the minimum distance over all the possible interleavers. Figure 3 represents the pmf of the minimum distance of a large-block Turbo code with $P = 7$. This pmf is calculated by using the fact that the smallest low-weight structure with a nonzero number determines the minimum distance.

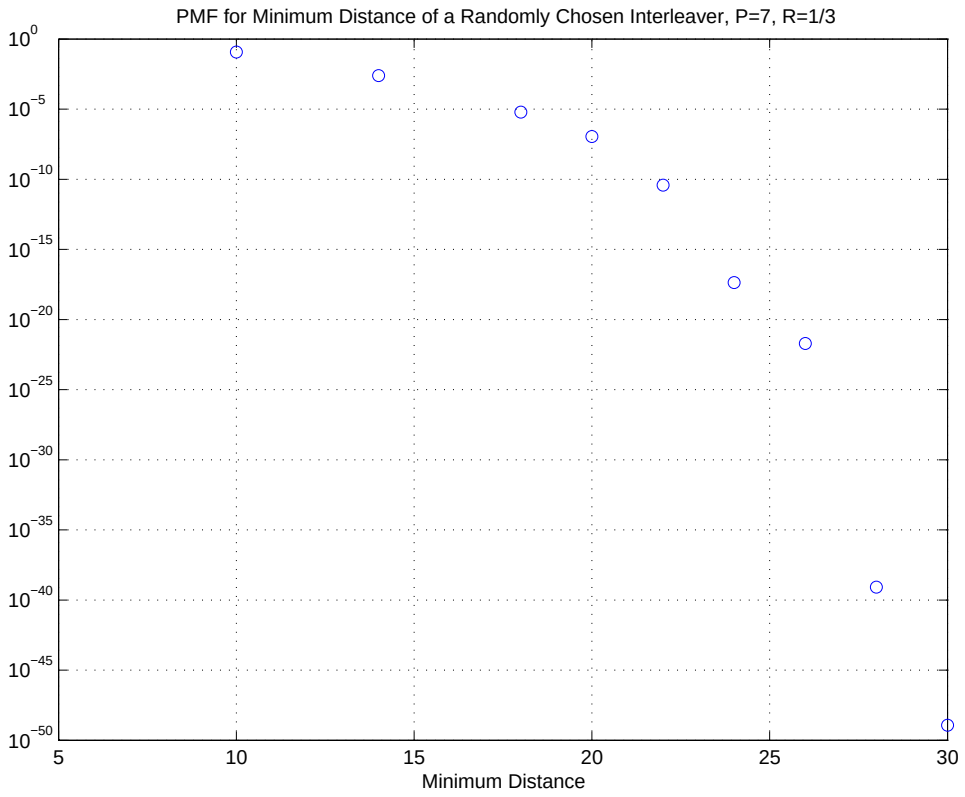

Fig. 3. Asymptotic probability mass function of the minimum distance of Turbo codes as $N \to \infty$ when RCC impulse response is 7.

B. Error floor for large block Turbo codes

In this section, the asymptotic behavior of the error floor is studied. Using the results of the previous section, we calculate the mean and the variance of the union bound on the error floor. Suppose that we sort the probable structures in the ascending order of their

weights. Obviously, the least weight belongs to codewords of type $(1, 2)$. The weight of such codewords is $2 + 2(P + 1)/2 = P + 3$. Suppose that the number of codewords of the i'th structure is Y_i which is determined by a Poisson distribution with the parameter λ_i. The error floor, by using the union bound, is bounded by $P_e \leq P_u = \sum_i Y_i p_i$, where $p_i = Q\left(\sqrt{\frac{2E_c w_i}{N_0}}\right)$ is the corresponding error for any codeword of the i'th structure, where E_c is the energy per channel use and is equal to E_b/R, where E_b is the energy per information bit and R is the code rate. The mean of this upper bound can be determined by $\sum_i \lambda_i p_i$. Using the same approach as that for small error events, we find that these random variables are asymptotically independent. Thus, the variance of P_u can be calculated as $\sigma^2_{P_u} = \sum_i \lambda_i p_i^2$.

In Figure 4, the mean and the standard deviation of the error floor for Turbo codes of the rate $\frac{1}{3}$ and memory lengths, 2, 3, and 4 (i.e., RCC impulse responses of 3, 7, and 15) are shown. As it is desirable, both the mean and the standard deviation decrease when the SNR increases. As the SNR increases, the ratio between them becomes $\sqrt{2}$. This occurs because, for this region of signal to noise ratio values, only the codewords of the lowest weight structure remain important. Figure 5 exhibits the effect of neglecting decomposable low-weight codewords on the mean of the error floor for a large-block Turbo code with $P = 3$. When P is relatively small, removing the decomposable low-weight codewords results in a tighter bound on the error floor.

C. Expurgating low-weight codewords

Low-weight codewords in Turbo codes occur when a low-weight information stream results in a few parity bits in both recursive convolutional encoders. As mentioned before, the average number of low-weight codewords in which more than two nonzero systematic bits cause a short error event is zero for large block lengths. The important point is that the average number of such low-weight codewords does not increase with the block length N [3]. The number of low-weight codewords is a nonnegative integer with a finite average, and consequently, the probability of having an infinite number of such

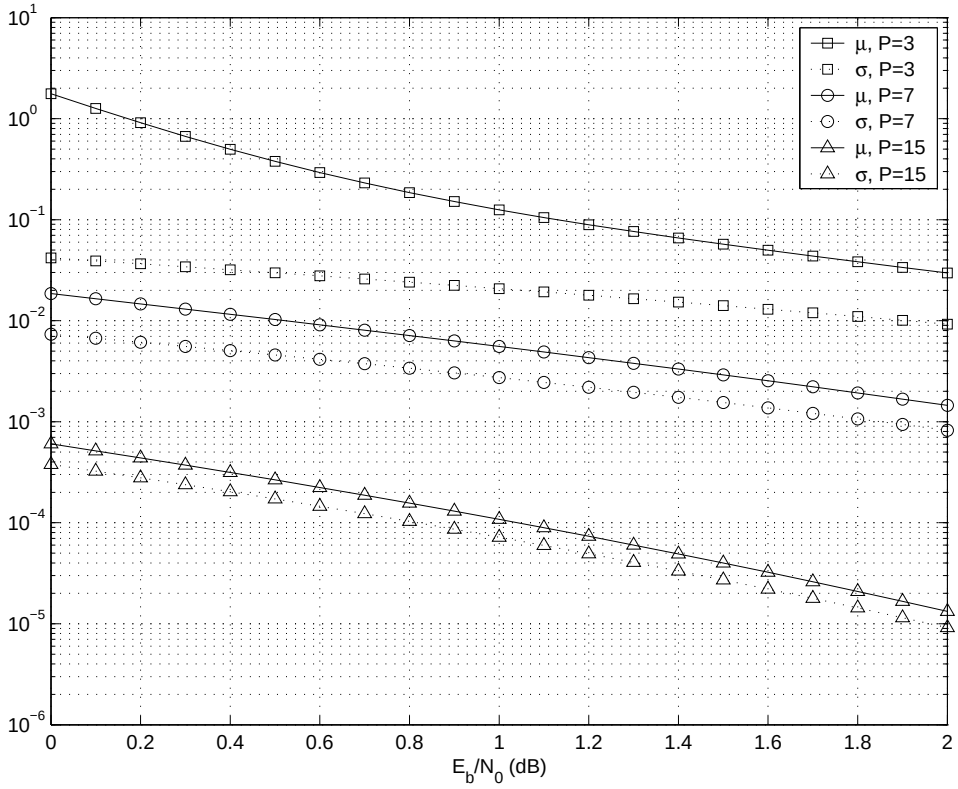

Fig. 4. The mean and standard deviation of the union bound on the error floor.

low-weight codewords approaches zero for large block lengths.

We can remove the effect of these low-weight codewords on the error floor region by expurgating them. Expurgating low-weight codewords decreases the dependency of the Turbo-code performance on the RCCs and the interleaver structure, since the remaining codewords tend to the Gaussian weight distribution.

To expurgate these codewords, one way is to set one information bit in each low-weight codeword to zero as presented in [4]. However, no further puncturing is required to maintain the code rate, because when the block length is sufficiently large, the number of these bits is small in comparison with the block length, and consequently, the code rate is not affected.

In Figure 6, the effect of expurgating low-weight codewords on the asymptotic mean

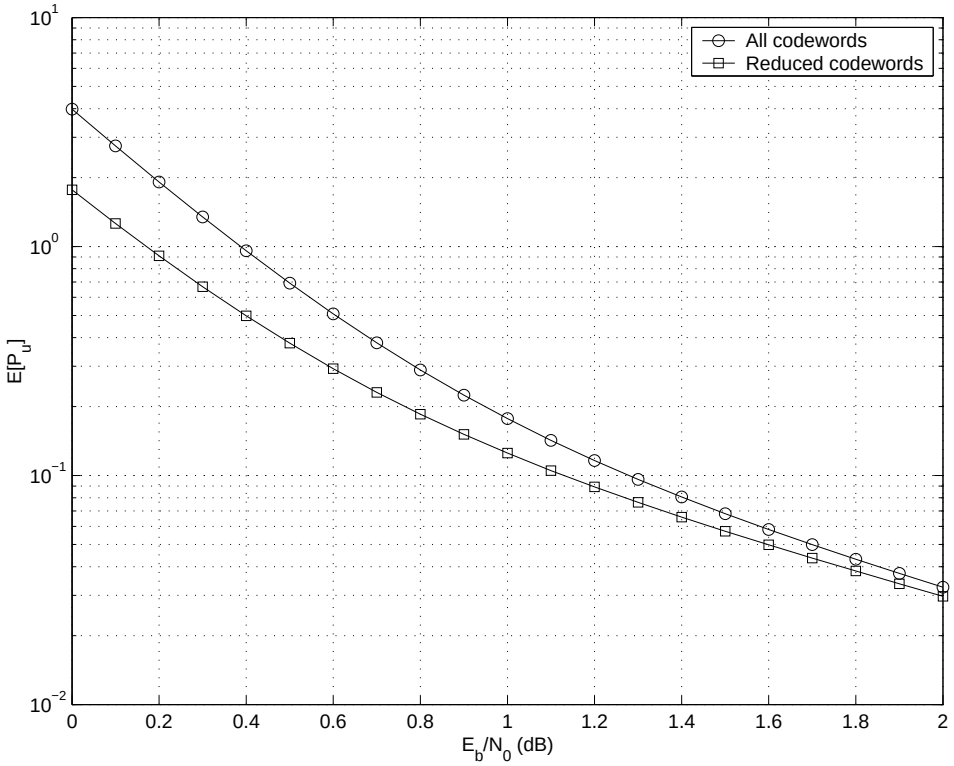

Fig. 5. The error floor for a large-block Turbo-code with $P = 3$.

of the error floor after expurgating codewords of the first low-weight structure (type (1,2), systematic weight 2 and parity weight $P+1$), and the second one (type (1,3), systematic weight 2 and parity weight $3(P+1)/2$) for a code of the rate $1/3$ and $P = 7$ is shown. On the average, there are two and four codewords of these two structures, respectively. The number of codewords in each of these two types does not exceed ten with probabilities 8×10^{-6} and 0.0028, respectively. Figure 7 presents the effect of the expurgation on a Turbo-code of the length 10000 and rate of 1/3 by using RCCs with three memory bits ($P = 7$). The interleaver is chosen randomly. Simulation results show that by using this randomly chosen interleaver, three low-weight codewords with the systematic weight of two and parity weight of less than or equal to 12 (having the first or the second structure) exist.

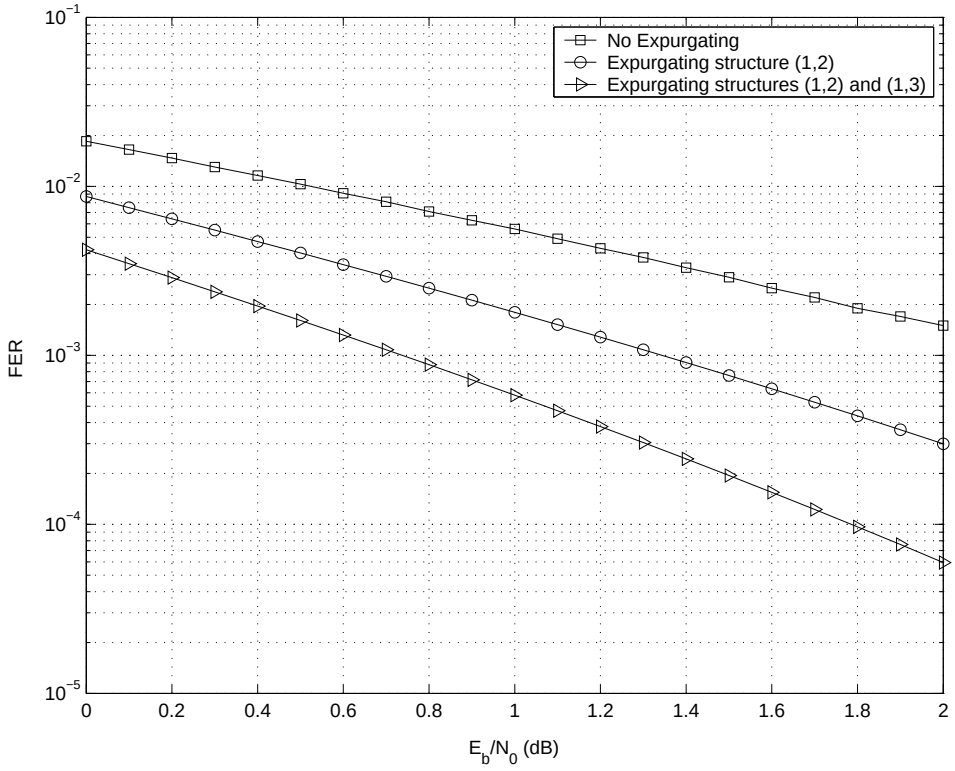

Fig. 6. Effect of expurgating two low-weight codeword structures on the asymptotic performance of a Turbo code of rate 1/3.

APPENDIX

A. Probability of error for large block Turbo codes

The Gaussian approximation of the Turbo code weight distribution is the same as the weight distribution of random codes. This assumption remains valid when high-weight codewords dominate the performance. One of the tools to characterize random coding is the cutoff rate. The weight of the dominant codewords in computing the cutoff rate provides insight into the validity of the Gaussian approximation. We compute the cutoff rate using the Gaussian distribution, and compare it to the random coding cutoff rate; namely, $R_0 = 1 - \log_2(1 + e^{-E_N/N_0})$, where E_N is the channel symbol energy, and N_0 is the one-sided Gaussian power spectrum of noise [45].

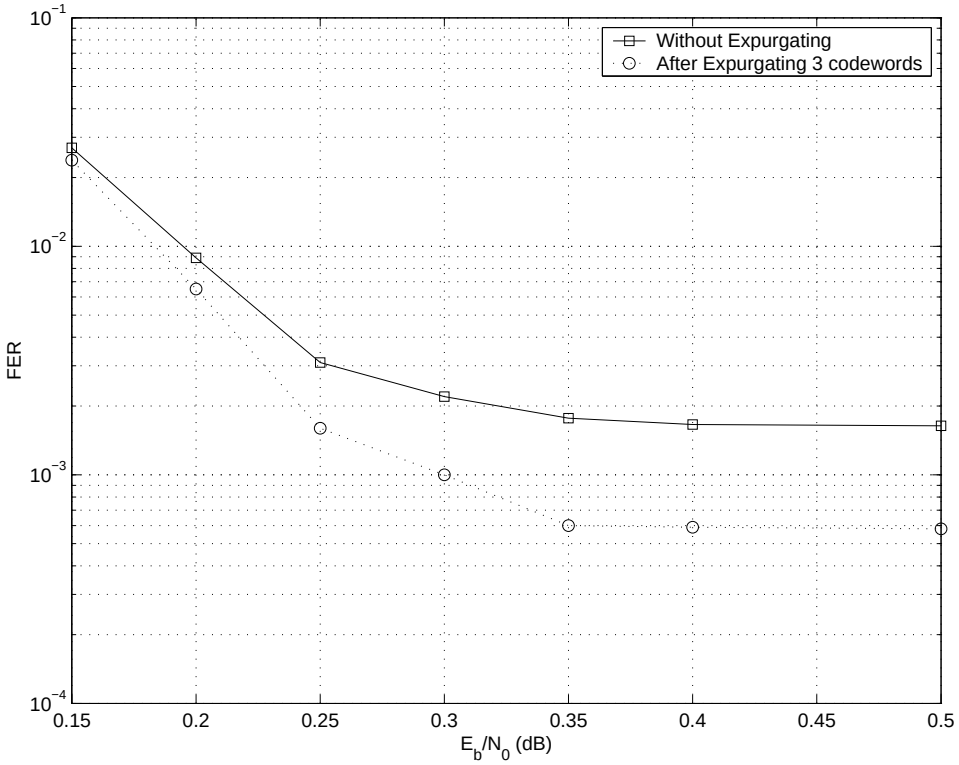

Fig. 7. Effect of expurgating three low-weight codewords on the performance of a Turbo code of rate 1/3 and block length of 10000.

For a Turbo code of the rate R and block length N, the normalized weight distribution function can be modelled as a Gaussian distribution with the mean $\frac{\sqrt{N}}{2R}$ and variance $\frac{1}{4R}$, where the code rate R is achieved by employing a larger number of parallel concatenated RCCs and/or puncturing which does not affect the Gaussian assumption. The number of codewords of the normalized weight between $\widehat{w}$ and $\widehat{w} + \Delta\widehat{w}$, under the Gaussian distribution, is

$$N_{\widehat{w}} \simeq \frac{2^N \Delta\widehat{w}}{\sqrt{\frac{\pi}{2R}}} \exp\left[-2R\left(\widehat{w} - \frac{\sqrt{N}}{2R}\right)^2\right]. \tag{20}$$

The term in the union bound that corresponds to the probability of an error event of the

normalized weight $\widehat{w}$ (using the BPSK modulation) is

$$p_{\widehat{w}} = Q\left(\sqrt{\frac{2\widehat{w}\sqrt{N}E_N}{N_0}}\right). \tag{21}$$

The dominant codewords in the error probability are around the peak of $N_{\widehat{w}}p_{\widehat{w}}$, which occurs at $\widehat{w_p} = \frac{\sqrt{N}}{2R}\left(1 - \frac{E_N}{2N_0}\right)$. The Gaussian assumption is valid when $\lim_{N\to\infty} \frac{R\widehat{w_p}}{\sqrt{N}} \neq 0, 1$. It is easy to see that $\frac{R\widehat{w_p}}{\sqrt{N}} < \frac{1}{2}$, and consequently, we only require that $\frac{R\widehat{w_p}}{\sqrt{N}} > 0$, resulting in $\frac{E_N}{N_0} < 2$ (equivalent to 3 dB). After the break point of $E_N/N_0 = 3$ dB is reached, the behavior of the Turbo code cannot be modelled anymore by using the Gaussian distribution.

In practice, Turbo codes are used in much lower ranges of signal to noise ratios than the break point. For example, the value $\frac{E_N}{N_0} = 3$ dB corresponds to the value of $\frac{E_b}{N_0} = 7.7$ dB (E_b stands for energy per information bit) for a code of the rate $1/3$, or to $\frac{E_b}{N_0} = 6$ dB for a code of the rate $1/2$. These values are substantially higher than those of the ranges of $\frac{E_b}{N_0}$ used in practical systems. In other words, the dominant codewords follow the Gaussian assumption for the SNRs of interest.

To find the cutoff rate under the Gaussian assumption, using the union bound, we have

$$P_e < \sum_{\widehat{w}=0}^{\frac{\sqrt{N}}{R}} N_{\widehat{w}}p_{\widehat{w}}. \tag{22}$$

By using the inequality $Q(x) < \frac{1}{2}\exp(-\frac{x^2}{2})$ and the Gaussian distribution assumption, (22) can be rewritten as

$$P_e < \frac{2^N}{\sqrt{\frac{2\pi}{R}}} A \int_0^{\frac{\sqrt{N}}{R}} \exp\left(-2R\left[\widehat{w} - \frac{\sqrt{N}}{2R}(1 - \frac{E_N}{2N_0})\right]^2\right) d\widehat{w}, \tag{23}$$

where

$$A = \exp\left(-\frac{N}{2R}\left[1 - \left(1 - \frac{E_N}{2N_0}\right)^2\right]\right), \tag{24}$$

and hence,

$$P_e < 2^{N-1}AB, \tag{25}$$

where

$$B = Q\left[\sqrt{\frac{N}{R}}\left(\frac{E_N}{2N_0} - 1\right)\right] - Q\left[\sqrt{\frac{N}{R}}\left(\frac{E_N}{2N_0} + 1\right)\right]. \tag{26}$$

For $\frac{E_N}{N_0} < 2$ and $N \to \infty$,

$$\lim_{N\to\infty} Q\left[\sqrt{\frac{N}{R}}\left(\frac{E_N}{2N_0} - 1\right)\right] = 1, \tag{27}$$

and,

$$\lim_{N\to\infty} Q\left[\sqrt{\frac{N}{R}}\left(\frac{E_N}{2N_0} + 1\right)\right] = 0. \tag{28}$$

Hence, B can be approximated as 1.

Let us define

$$R_T = \frac{1}{2\ln(2)}\left[\frac{E_N}{N_0} - \frac{1}{4}\left(\frac{E_N}{N_0}\right)^2\right]. \tag{29}$$

We can see that if $R < R_T$, then the probability of error converges to 0 as $N \to \infty$. Figure 8 reflects the difference between R_0 and R_T around the break point of $\frac{E_N}{N_0} = 3$ dB ($\frac{E_b}{N_0} = 7.7$ dB for a code of the rate $1/3$).

B. Proof of the Proposition

To prove the proposition, we need the following lemma.

Lemma: Suppose we partition a stream of N bits consisting of w number of ones and $N - w$ number of zeros into K groups. Each group consists of $N_k,\ k = 1, \ldots, K$ ($\sum_k N_k = N$) bits. We show the event in which the k'th group has an odd Hamming weight by O_k. For $N \to \infty$, if

$$\lim_{N\to\infty} N_k/N \neq 0,\ k = 1, \ldots, N, \tag{30}$$

then O_1, O_2, ..., O_{K-1} tend to be independent events of the probability of $1/2$ as N goes to infinity (for the typical values of w).

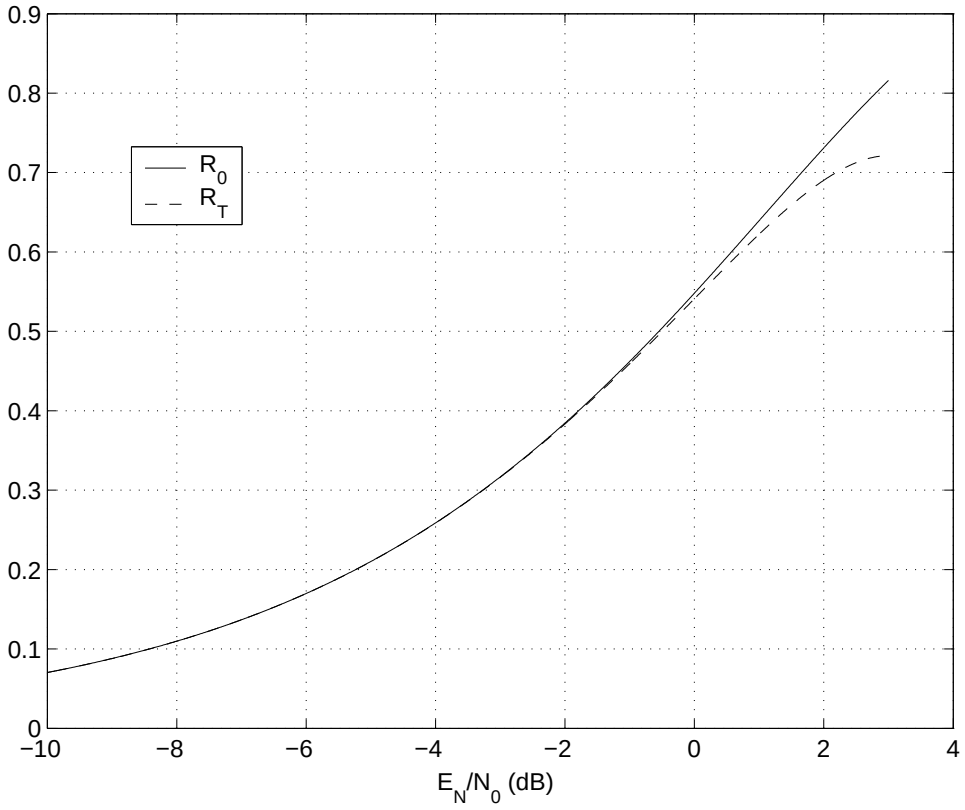

Fig. 8. Comparison between R_0 and R_T versus $\frac{E_N}{N_0}$.

Proof: The Hamming weight of the k'th group is shown by W_k. Then, the probability mass function of W_k can be written as

$$P_{W_k}(w_k) = \frac{\binom{N-N_k}{w-w_k}\binom{N_k}{w_k}}{\binom{N}{w}}, \;\; w_k = 0, 1, \ldots, N_k. \tag{31}$$

This probability mass function is an increasing function with respect to w_k for $0 < w_k < w_t$, where $w_t = \lfloor \frac{wN_k}{N} \rfloor$ is the typical value for the Hamming weight of the k'th subsequence, and is decreasing for $w_t < w_k < \min\{w, N_k\}$.

It is easy to see that for an integer random variable with a monotonic probability mass function, the difference between the sum of the probabilities for even and odd numbers is smaller than the boundary probabilities. The probability mass function that

is described by (31) can be separated into two monotonic functions. For $N \to \infty$, the boundary probabilities specified by (31) (i.e., the probabilities at $w = 0, N, w_t$) are 0, and so,

$$P\{W_k \text{ is odd}\} = P\{W_k \text{ is even}\} = \frac{1}{2}. \tag{32}$$

The same approach is valid for the k'th group ($k < K$), when the Hamming weight of the first $k-1$ groups are known, and hence, it is odd-weighted with the probability of $1/2$. Obviously, the Hamming weight of the K'th group, given the Hamming weights of the other groups, is known.

■

Proposition: Assuming the systematic weight is w_1, each parity stream is an m-dependent sequence, and the variance of its weight is given by

$$\sigma^2_{w_2|w_1} = \frac{N}{4}\left(1 + \frac{2(1-\frac{2w_1}{N})^{(P+1)/2}}{1-(1-\frac{2w_1}{N})^{(P+1)/2}}\right). \tag{33}$$

Proof: Consider two arbitrary parity bits (far from the boundaries) named pb_1 and pb_2 in one parity stream. We show that these two bits are independent of each other, when their distance is large. The proof can be extended to two sets of parity bits. According to the distance between pb_1 and pb_2, two situations can occur.

Case I: The distance between these parity bits is not an integer multiple of the RCC impulse response period P. We divide the information bits into four subsets, depending on whether they trigger these two parity bits or not. We call the four groups $C_k, k = 0, 1, 2, 3$. The members of the C_0 trigger none of the parity bits. Members of C_1 and C_2 trigger just the first parity bit and the second parity bit, respectively. Finally, C_3 consists of bits that trigger both parity bits. Similarly, we show the event of having an odd Hamming weight in the C_i by $O_i,\ i = 0, 1, 2, 3$. All the systematic bits after the second parity bit are in C_0. For any P information bits preceding the first parity bit, there is at least one bit in each of $C_i,\ i = 1, 2, 3$. Hence,

$$\frac{|C_i|}{N} \neq 0,\ \ i = 0, 1, 2, 3, \tag{34}$$

where $|.|$ denotes the cardinality of a set. As a result, C_i's satisfy the conditions in the preceding lemma. It is easy to see that

$$pb_1 = O_1 \oplus O_3, \quad pb_2 = O_2 \oplus O_3, \tag{35}$$

in which $\oplus$ is the binary addition (pb_1 is one if just one of O_1 and O_3 happens, and is zero, otherwise.) Since, O_1, O_2 and O_3 are equiprobable identical independent events, then pb_1 and pb_2 are equiprobable independent bits.

Case II: The distance between those two bits is an integer multiple of impulse response period P, say kP. In this case, C_1 is empty, but C_0 and C_3 still satisfy the condition in the lemma. C_2 has only $k(P+1)/2$ elements. However, as long as the distance between the two parity bits is large (when k is large which is true for almost any two typical bits), the conditions of the Lemma are satisfied, and O_2 and O_3 become equiprobable identical independent events. As a result pb_1 and pb_2 are independent.

To apply the Central Limit Theorem to the m-dependent sequence of the parity stream, we have to find the variance of the conditional parity weight. This variance is a function of the cross correlation between the near parity bits that are separated by an integer multiple of P (all the other pairs of the parity bits are uncorrelated). To compute this correlation, we note that when the distance between the parity bits is kP (k is a relatively small integer), the elements of C_2 can be considered to be iid bits, and each of them is one with the probability of $\frac{w_1}{N}$. Then,

$$cov[b_2(i), b_2(i+kP)] = \frac{1}{4}\left(1 - \frac{2w_1}{N}\right)^{k(P+1)/2}, \tag{36}$$

because the probability of having an odd parity within these $k(P+1)/2$ bits is

$$P\{O_2 = 1\} = \frac{1 - \left(1 - \frac{2w_1}{N}\right)^{k(P+1)/2}}{2}. \tag{37}$$

The covariances of the other pairs are zero. Since, the parity weight is $w_2 = \sum_{i=0}^{N} b_2(i)$, then

$$\sigma^2_{w_2|w_1} = \sum_{i=1}^{N} \sigma^2_{b_2(i)} + 2 \sum_{1 \le i < j \le N} cov[b_2(i), b_2(j)]. \tag{38}$$

As a result,

$$\begin{aligned}
\sigma^2_{w_2|w_1} &= \sum_{i=1}^{N} \frac{1}{4} + 2 \sum_{i=1}^{N} \sum_{k=1}^{\lfloor (N-i)/P \rfloor} cov[b_2(i), b_2(i+kP)] \\
&= \frac{N}{4} + 2 \sum_{i=1}^{N} \sum_{k=1}^{\lfloor (N-i)/P \rfloor} \frac{1}{4} \left(1 - \frac{2w_1}{N}\right)^{k(P+1)/2} \\
&\simeq \frac{N}{4} + 2 \sum_{i=1}^{N} \sum_{k=1}^{\infty} \frac{1}{4} \left(1 - \frac{2w_1}{N}\right)^{k(P+1)/2} \\
&= \frac{N}{4} \left(1 + \frac{2(1 - \frac{2w_1}{N})^{(P+1)/2}}{1 - (1 - \frac{2w_1}{N})^{(P+1)/2}}\right).
\end{aligned} \tag{39}$$

■

REFERENCES

[1] G. Battail, "Construction explicite do bons code longs," *Ann. Télécommun.*, pp. 392–404, 1989.

[2] E. Biglieri and V. Volski, "Approximately gaussian weight distribution of the iterated product of single-parity-check codes," *IEE Electronics Letters*, vol. 30, pp. 923–924, June 1994.

[3] L. C. Perez, J. Seghers, and D. J. Costello Jr., "A distance spectrum interpretation of turbo codes," *IEEE Trans. on Inform. Theory*, vol. 42, pp. 1698–1709, November 1996.

[4] F. Daneshgaran, M. Mondin, and P. Mulassano, "Turbo codes optimization via trace-bit injection and selective puncturing," in *IEEE International Conference on Communications*, April-May 2002, vol. 3, pp. 1706–1710.

[5] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes (1)," in *IEEE International Conference on Communications*, Geneva, Switzerland, May 1993, pp. 1064–1070.

[6] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. on Inform. Theory*, vol. 20, pp. 284–287, March 1974.

[7] L. Ping and K.L. Yeung, "Symbol-by-symbol decoding of the golay code and iterative decoding of concatenated Golay codes," *IEEE Trans. on Inform. Theory*, vol. 45, pp. 2558–2562, November 1999.

[8] Ye Liu, Shu Lin, and M.P.C. Fossorier, "MAP algorithms for decoding linear block codes based on sectionalized trellis diagrams," *IEEE Trans. on Communications*, vol. 48, pp. 577–586, April 2000.

[9] S. Riedel, "Symbol-by-symbol MAP decoding algorithm for high-rate convolutional codes that use reciprocal dual codes," in *IEEE Journal on Selected Areas in Communications*, February 1998, vol. 16, pp. 175–185.

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

[11] I. Sason, E. Teletar, and R. Urbanke, "On the asymptotic inputoutput weight distributions and thresholds of convolutional and turbo-like encoders," *IEEE Trans. on Inform. Theory*, vol. 48, pp. 3052–3061, December 2002.

[12] O. Y. Takeshita, M. P. C. Fossorier, and D. J. Costello, "A new technique for computing the weight spectrum of turbo codes," *IEEE Comm. Letters*, vol. 3, pp. 251–253, August 1999.

[13] T. M. Duman and M. Salehi, "New performance bounds for turbo codes," *IEEE Trans. on Communications*, vol. 46, pp. 717–723, June 1998.

[14] 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. on Inform. Theory*, vol. 46, pp. 24–47, January 2000.

[15] I. Sason and S. Shamai, "Variations on the Gallager bounds, connections, and applications," *IEEE Trans. on Inform. Theory*, vol. 48, pp. 3029–3051, December 2002.

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

[17] G. Battail, "A conceptual framework for understanding turbo codes," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 245–254, Feb. 1998.

[18] G. Battail, C. Berrou, and A. Glavieux, "Psuedo-random recursive convolutional coding for near-capacity performance," in *IEEE Globecom Conference*, Houston, USA, Nov. 1993, pp. 23–27.

[19] N. Shulman, "Random coding techniques for nonrandom codes," *IEEE Trans. on Inform. Theory*, vol. 45, pp. 2101–2104, Sep. 1999.

[20] P. Robertson, "Improving decoder and code structure of parallel concatenated recursive systematic (turbo) codes," in *IEEE Universal Personal Communications Conference*, San Diego, USA, September 1994, pp. 183–187.

[21] H.R. Sadjadpour, N.J.A. Sloane, M. Salehia, and G. Nebe, "Interleaver design for turbo codes," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 831–837, May 2001.

[22] O. Y. Takeshita and D. J. Costello, "New deterministic interleaver designs for turbo codes," *IEEE Trans. on Inform. Theory*, vol. 46, pp. 1988–2006, September 2000.

[23] A. K. Khandani, "Optimization of the interleaver structure for turbo-codes," in *Canadian Workshop on Information Theory*, Kingston, Canada, June 1999, pp. 25–28.

[24] A. K. Khandani, "Design of the turbo-code interleaver using hungarian method," *IEE Electronics Letters*, vol. 34, pp. 63–65, January 1998.

[25] M. Ferrari, F. Scalise, and S. Bellini, "Prunable s-random interleavers," in *IEEE International Conference on Communications (ICC)*, New York, USA, April 2002, vol. 3, pp. 1711–1715.

[26] D. Truhachev, M. Lentmaier, O. Wintzell, and K. Sh. Zigangirov, "On the minimum distance of turbo codes," in *IEEE International Symp. on Inform. Theory*, Lausanne, Switzerland, July 2002, p. 84.

[27] M. Eroz and A. R. Hammons, "On the design of prunable interleavers for turbo codes," in *Vehicular Technology Conference*. IEEE, May 1999, vol. 2, pp. 1669–1673.

[28] F. Daneshgaran and M. Mondin, "Optimized turbo codes for delay constrained applications," *IEEE Trans. on Inform. Theory*, vol. 48, pp. 293–305, Jan. 2002.

[29] F. Daneshgaran and M. Mondin, "Design of interleavers for turbo codes: iterative interleaver growth algorithms of polynomial complexity," *IEEE Trans. on Inform. Theory*, vol. 45, pp. 1845–1859, Sept 1999.

[30] W. Feng, J. Yuan, and B.S. Vucetic, "A code-matched interleaver design for turbo codes," *IEEE Trans. on Communications*, vol. 50, pp. 926–937, June 2002.

[31] J. Yuan, B. Vucetic, and W. Feng, "Combined turbo codes and interleaver design," *IEEE Transactions on Communications*, vol. 47, pp. 484–487, April 1999.

[32] J.A. Briffa and V. Buttigieg, "Interleaving and termination in unpunctured symmetric turbo codes," *IEE Proceedings on Communications*, vol. 149, pp. 6–12, Feb. 2002.

[33] F. Vatta, B. Scanavino, A. Banerjee, and D.J. Costello, "On the design of nonsystematic turbo codes," in *IEEE Int. Symp. on Inform. Theory*, Yokohama, Japan, June-July 2003, p. 320.

[34] G. Zhu and F. Alajaji, "Turbo codes for nonuniform memoryless sources over noisy channels," *IEEE Communications Letters*, vol. 6, pp. 64–66, Feb. 2002.

[35] K. Wu, H. Li, and Y. Wang, "Influence of interleaver on minimum turbo code distance," *IEEE Electronics Letters*, vol. 35, pp. 1456–1458, Aug. 1999.

[36] M. Breiling and J. B. Huber, "Upper bound on the minimum distance of turbo codes," *IEEE Trans. on Communications*, vol. 49, pp. 808–815, May 2001.

[37] M. Breiling and J. B. Huber, "Combinatorial analysis of the minimum distance of turbo codes," *IEEE Trans. on Inform. Theory*, vol. 47, pp. 2737–2750, Nov. 2001.

[38] F. Daneshgaran and M. Mondin, "Permutation fixed points with application to estimation of minimum distance of turbo codes," *IEEE Trans. on Inform. Theory*, vol. 46, pp. 2336–2349, Nov. 2000.

[39] R. Garello, P. Pierleoni, and S. Benedetto, "Computing the free distance of turbo codes and serially concatenated codes with interleavers: algorithms and applications," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 800–812, 2001.

[40] E. Rosnes and O. Ytrehus, "On algorithms for determination of turbo code weight distribution," in *IEEE Int. Symp. on Inform. Theory*, Lausanne, Switzerland, July 2002, p. 82.

[41] C. Di, R. Urbanke, and T. Richardson, "Weight distributions: How deviant can you be," in *IEEE Int. Symp. on Inform. Theory*, July 2001, p. 50.

[42] C. Tanriover, B. Honary, J. Xu, and S. Lin, "Improving turbo code error performance by multifold coding," *IEEE Comm. Letters*, vol. 6, pp. 193–195, May 2002.

[43] S. W. Golomb, *Shift Register Sequences*, San Francisco, Holden-Day, 1967.

[44] J. K. Patel and C. B. Read, *Handbook of the Normal Distribution*, Dekker Inc., second edition, 1996.

[45] John G. Proakis, *Digital Communication*, McGraw-Hill, fourth edition, 2001.