

# Analysis of the Asymptotic Performance of Turbo Codes

**M. Hadi Baligh, Amir K. Khandani & I. Bahceci**

## Abstract

Battail [1989] shows that an appropriate criterion for the design of long block codes is the closeness of the normalized weight distribution to Gaussian. A subsequent work by Biglieri and Volski [1994] shows that iterated product of single parity check codes satisfies this criterion. Motivated by these works, in the current article, we study the performance of turbo codes for large block lengths,  $N \rightarrow \infty$ . We show that for a parallel concatenated code that consists of two component codes, for  $N \rightarrow \infty$ , the normalized weight of the systematic sequence  $\hat{w}_1 = \frac{w_1}{\sqrt{N}}$ , and the parity check sequences  $\hat{w}_2 = \frac{w_2}{\sqrt{N}}$  and  $\hat{w}_3 = \frac{w_3}{\sqrt{N}}$  become jointly Gaussian for the typical values of  $\hat{w}_i, i = 1, 2, 3$ , where the typical values of weight are defined as  $\lim_{N \rightarrow \infty} \frac{w_i}{N} \neq 0, 1$  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 the correlation coefficients between  $\hat{w}_i$  and  $\hat{w}_j$ ,  $\rho_{ij}, i \neq j = 1, 2, 3$ . We show that: (i)  $\rho_{ij} > 0, i, j = 1, 2, 3$ , (ii)  $\rho_{12}, \rho_{13} \rightarrow 0$  as  $N \rightarrow \infty$ , and (iii)  $\rho_{23} \rightarrow 0$  as  $N \rightarrow \infty$  for “almost” any random interleaver. This indicates that for  $N \rightarrow \infty$ , the optimization of the interleaver has a diminishing effect on the distribution of high-weight error events, and consequently, on the error performance in the waterfall region. We show that for the typical weights, the Gaussian weight distribution approaches the average spectrum defined by Poltyrev [1994]. We also apply the tangential sphere bound (TSB) on Gaussian distribution and show that the

This work has been presented in part in the Canadian Workshop on Information Theory (CWIT'99) [1] and in part in the Conference on Information Sciences and Systems (CISS'02) [2] and is reported in part in [3]. Submitted to IEEE Transactions on Information Theory on March 2006.

Coding & Signal Transmission Lab, Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L

code performs very close to the capacity for lower code rates. We also study the statistical properties of the low-weight codeword structures. We prove that for large block lengths, the number of low-weight codewords of these structures are some Poisson random variables. These random variables can be used to evaluate the asymptotic probability mass function of the minimum distance of the turbo code among all the possible interleavers. We show that the number of indecomposable low-weight codewords of different types tend to a set of independent Poisson random variables. We find the mean and the variance of the union bound in the error floor region over the interleaver ensemble and study the effect of expurgating low-weight codewords on the performance. We show that the weight distribution in the transition region between Poisson and Gaussian follows a negative binomial distribution. We also calculate the interleaver gain for multi-component turbo codes based on these Poisson random variables, and we show that the asymptotic error performance for multi-component codes in different weight regions converges to zero either exponentially (in the Gaussian region) or polynomially (in the Poisson and negative binomial regions) with respect to the block length, with the code-rate and energy values close to the channel capacity.

**Keywords:** Turbo codes, asymptotic performance, weight distribution, Gaussian distribution, waterfall region, error floor, TSB.

## I. INTRODUCTION

The advent of turbo codes [4] 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 two or more recursive convolutional codes (RCCs) which are concatenated in parallel or serially via pseudo-random interleavers. Since the RCCs and also the interleaver have the linearity property<sup>1</sup>, the resulting code is linear. Consequently, the group property and the distance invariance property hold.

Figure 1 presents a rate 1/3 turbo code with two RCCs. The coded bits are obtained by multiplexing the systematic bits  $b_1(i)$ ,  $i = 1, 2, \dots, N$ , and parity check bits  $b_2(i)$ ,  $b_3(i)$ . The weight of the code in

<sup>1</sup>The effect of interleaving is equivalent to multiplying the input sequence by a permutation matrix which corresponds to a linear operation.

Figure 1 is equal to the sum of the weights of sequences  $\{b_1\}$ ,  $\{b_2\}$  and  $\{b_3\}$ , over a block, denoted by  $w_1$ ,  $w_2$ , and  $w_3$ , respectively.

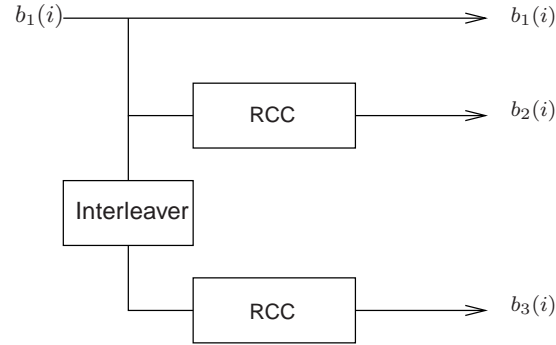


Fig. 1. Basic structure of the turbo encoder.

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 region, the performance is determined by low-weight codewords.

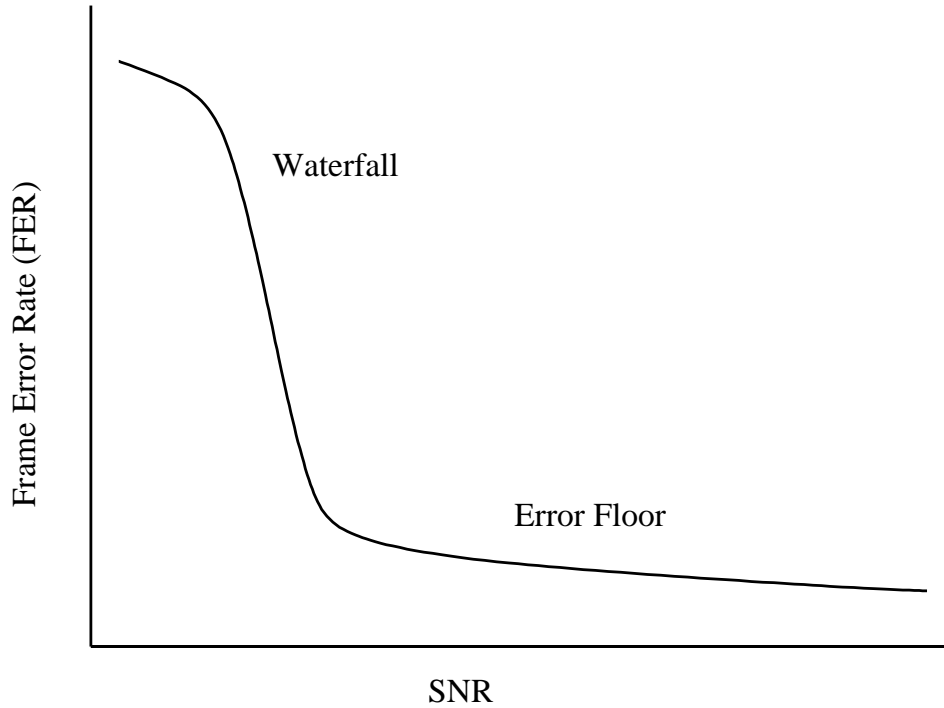


Fig. 2. Typical error performance of a turbo code over an AWGN channel.

Weight distribution plays an important role in assessing the performance of the code using maximum likelihood (ML) decoding. While ML decoding is not computationally feasible for turbo codes, it provides insight into the potential performance of these codes. Because of the existence of the interleaver, the analysis based on the actual weight distribution becomes very complicated. Benedetto and Montorsi introduce “uniform interleaving” technique and evaluate the “average weight distribution” of the code, which is defined as the average weight distribution among all codes generated with various possible interleavers [5].

In [6], the asymptotic average weight distribution is calculated for large block lengths. In [7], according to the average weight spectrum, a simple approximation of the performance of parallel concatenated turbo codes is obtained. Based on the concept of average weight distribution, Jin and McEliece in [8] prove that for a uniformly interleaved turbo code and for some region of signal to noise ratio, the asymptotic performance of serial and parallel concatenated turbo codes approaches zero, at least as fast as  $N^{-\beta}$ , where  $\beta$  is called the “interleaver gain” [5]. They also show that with uniform interleaving an interleaver gain of  $J - 2 - \epsilon$  is obtained for a parallel concatenated code consisting of  $J$  component codes.

In [9], [10], it is shown that turbo codes belong to the class of weakly random-like codes; although their frame error rate (FER) performance is poor, the bit error rate (BER) remains low up to the code rates within the vicinity of the channel capacity. Shulman provides techniques to apply the channel coding theorem and the error exponent, which was originally derived for random block-code ensembles, to the ensembles of codes with fewer restrictive randomness requirements [11]. Based on the weight distribution of turbo codes and by using Gallager bounding technique [12] and tangential sphere bound (TSB) [13], some upper bounds on the performance of turbo codes are derived in [14]–[16]. In particular, it is shown that TSB is the tightest bound derived by Gallager bounding technique and the TSB error exponent for a code with “average spectrum” is very close to the random coding exponent, especially for low-rate codes [13].

It is known that using a pseudo-random interleaver in turbo codes guarantees an excellent BER performance, but a certain number of low-weight codewords are generated, resulting in a small minimum distance and the appearance of an error floor. The structure and the number of such low-weight codewords are studied in [17] and [18], where it is reported that asymptotically probable low-weight codewords consist of one or more short error events<sup>2</sup> and each such event is due to an information sequence of weight two. The effect of the interleaver structure on the minimum distance of the code is studied in [19]. Breiling proves that the minimum Hamming distance of turbo codes cannot asymptotically grow at a rate higher than  $O(\log N)$  [20]. In [21], the variance of the error floor over all possible interleavers for a turbo code with a finite block length is evaluated and it is shown that the variation (the standard deviation divided by the mean) of the error floor increases with signal to noise ratio and decreases with the block length.

Battail shows that an appropriate criterion for the design of large block codes is the closeness of the normalized weight distribution to Gaussian rather than having a large minimum distance [22]. Biglieri and Volski substantiate this by showing that iterated-product codes have a weight distribution that is approximately Gaussian [23]. This line of work is followed by [24] which shows that for codes with rates approaching one, the weight distribution is asymptotically Gaussian as the block length increases. For codes with lower code rates, if the minimum distance of the dual code tends to infinity, the cumulative weight distribution asymptotically tends to that of a Gaussian. This provides a sufficient condition on the systematic parity-check matrix of the code in order to have a Gaussian distribution. Note that this condition is rather restrictive and it cannot be applied to the turbo code structure shown in Figure 1.

In this paper, we address the weight distribution of turbo codes and show that the weight distribution can be classified into three different regions. The number of the high-weight codewords follow a Gaussian distribution, the number of the low-weight codewords form a set of Poisson random variables and the transition region from low-weight to high-weight codewords has a negative binomial distribution. We

<sup>2</sup>A short error event is defined by leaving the all zero state and returning to it after a few information bits.

prove that for high-weight codewords, the weights of the systematic and parity streams tend to a set of uncorrelated, and hence, independent, Gaussian random variables for a randomly chosen interleaver and for any nontrivial recursive convolutional code. We show that with probability one, in the waterfall region, a randomly chosen interleaver performs as well as the best interleaver. The performance of a code with average spectrum is very close to that of a capacity-achieving random code with binary phase shift keying (BPSK) signaling over an AWGN channel [13]. We apply the tangential sphere bound (TSB) on the frame error rate of turbo codes assuming a Gaussian weight distribution and find the signal-to-noise ratio (SNR) region for which the error exponent is positive and hence, the error probability converges to zero as the block length increases. We show that the corresponding achievable rate is very close to the capacity for code rates of interest. Note that since this study is based on weight spectrum, the performance analysis in this paper is valid for maximum likelihood (ML) decoding of the code.

We also investigate the effect of the interleaver optimization on the error floor region. It is known that the low-weight codewords do not follow the Gaussian distribution and they are more important in determining the performance of the code in the error floor region (at high SNR). Therefore, unlike in the waterfall region, the optimization of the component codes and the interleaver affect the performance in the error floor region. In [17], 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 such 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. We also show that indecomposable low-weight codewords constitute a set of independent Poisson random variables. After this work was completed [3], we became aware of [25] which studies the asymptotic behavior of minimal (indecomposable) low-weight codewords, based on detour generating function of convolutional codes. In this work, however, we study the statistical properties based on asymptotically possible low-weight codewords and then, we derive the mean (and the variance) of the number of decomposable and indecomposable low-weight codewords. We show

that with Maximal Length Sequence (MLS) component codes, the Poisson parameters of these random variables are only functions of the number of memory elements in each component code. 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 evaluated and it is shown that the error probability converges to zero for multi-component codes as  $N \rightarrow \infty$ . Furthermore, we show that in the transition region between Poisson (for finite weight codewords) and Gaussian (for typical codewords), the weight spectrum has a negative binomial distribution. The error probability corresponding to codewords in this region tends to zero as the block length increases. The three different weight regions and the corresponding weight distributions are shown in Figure 3, where  $w \sim o(N)$  denotes weights where  $\lim_{N \rightarrow \infty} w = \infty$  and  $\lim_{N \rightarrow \infty} \frac{w}{N} = 0$ , and  $w \sim O(N)$  denotes weights linearly increasing with  $N$ .

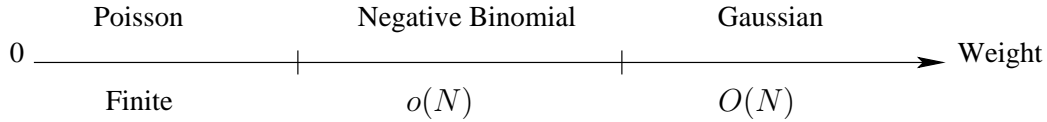


Fig. 3. The weight distribution for different regions of weight.

In [26], it is indicated that using  $J > 2$  component codes improves the distance properties of turbo codes. In this paper, we show that for a turbo code with  $J$  component codes and randomly chosen interleavers, the interleaver gain is  $J - 2$  which is the same as for the uniformly interleaved code reported in [8]. As a result, for  $J > 2$ , the error floor asymptotically tends to zero, with probability one.

Our results show that the overall performance of multi-component turbo codes is very close to the capacity for BPSK signalling over an AWGN channel, because: (i) the error probability due to high-weight codewords exponentially tends to zero for SNR values and rates close to the capacity, and (ii) the low-weight codewords result in an error floor which decreases polynomially as the block length increases. Finally, observing that the number of low-weight codewords is small, we discuss a method to expurgate the low-weight codewords following the method introduced in [27], [28], and show that the

interleaver gain can be increased for multi-component turbo codes by expurgating low-weight codewords.

This article is organized as follows. In Section II, we study the asymptotic weight distribution of turbo codes for their typical weights and examine the effect of the interleaver optimization to improve the waterfall region when the block length is large. Section III is concerned with asymptotic behavior of turbo code and its weight distribution in the error floor region. We also find the asymptotic statistical properties of the low-weight codewords and the asymptotic behavior of the error floor for large block turbo codes. Section IV concludes the article.

## II. ASYMPTOTIC PERFORMANCE OF TURBO CODES IN THE WATERFALL REGION<sup>3</sup>

### A. Parallel concatenated turbo code structure

Consider a turbo code with two RCCs and let the transfer function  $G(d) = N(d)/D(d)$ . The impulse response of  $G(d)$  is periodic with period  $P \leq 2^r - 1$ , where  $r$  is the memory length of the constituent code [29]. 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  $N(d)$ . Although we consider a parallel concatenated code with three output streams as shown in Figure 1, the discussions can be generalized to other configurations.

In general, it is desirable to have the period of the impulse response of  $G(d)$  as large as possible. If the period is equal to  $2^r - 1$ , the 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 a given  $r$ ) are provided in [29]. It can be shown that any MLS satisfies the three postulates of randomness [29]. One consequence of this property is that in any period of an MLS, the number of ones is  $2^{r-1}$ , and the number of zeros is  $2^{r-1} - 1$ .



<sup>3</sup>Results in subsections A and B have been presented in part in the Canadian Workshop on Information Theory (CWIT'99)



If the impulse response of  $D(d)$  is considered to be a periodic sequence (starting from minus infinity), 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 labeled by integer numbers, say  $1, \dots, P$ , 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 a maximum length sequence (plus the all-zero sequence) constitutes a group under binary addition [29]. 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 by the zero phase). Using the group property of phases, we conclude that the role 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 any role in the following discussions.


For bit position  $k$ ,  $k = 1, \dots, N$ , within the  $i^{\text{th}}$  output stream, we denote by  $\mathcal{R}_i(k)$ ,  $i = 1, 2, 3$ , the set of systematic bit positions  $\{j | j \leq k\}$  for which an impulse at position  $j$  results in a nonzero parity bit in position  $b_i(k)$ . Obviously,  $\mathcal{R}_1(k) = \{k\}$ . If the bit position  $k$  is located in the  $L^{\text{th}}$  period, i.e.,  $L = \lceil k/P \rceil$ , where  $\lceil \cdot \rceil$  denotes the ceiling function, then the number of positions belonging to  $\mathcal{R}_i(k)$ ,  $i = 2, 3$ , within each of the periods  $1, \dots, L - 1$  is equal to  $2^{r-1}$  [29]. The number of positions within the  $L^{\text{th}}$  period (the period containing  $k$  itself) depends on the relative position of  $k$  within the  $L^{\text{th}}$  period and also on the numerator of  $G(d)$ . We are mainly interested in large values of  $L$  (i.e., for parity bits far from the boundaries resulting in  $L \gg 1$ ) for which the effect of the elements within the  $L^{\text{th}}$  period itself is negligible. Thus,  $|\mathcal{R}_2(k)| = |\mathcal{R}_3(k)| \simeq \lceil k/P \rceil 2^{r-1}$ , where  $|\cdot|$  denotes the cardinality of the corresponding set.

The notation  $b_i(k)$ ,  $k = 1, \dots, N$ , refers to the  $k^{\text{th}}$  bit within the  $i^{\text{th}}$  output stream for  $i = 1, 2, 3$ . Assuming equally likely transmitted bits, i.e.  $P\{b_1(k) = 1\} = P\{b_1(k) = 0\} = \frac{1}{2}$ , we have  $\overline{b_1(k)} = \overline{b_1^2(k)} = 1/2$ . On the other hand, with equiprobable input sequences, all  $2^N$  possible combinations within the three streams are equiprobable, and consequently, the bit values within each of the three output

streams are independently and identically distributed (iid) Bernoulli random variables with parameter  $1/2$ . As a result,  $\overline{b_i(k)} = \overline{b_i^2(k)} = 1/2$ ,  $i = 2, 3$ .

### B. Analysis of weight distribution and asymptotical performance

To investigate the asymptotic weight distribution of turbo codes, we first show that the normalized weights,  $\hat{w}_i = \frac{w_i}{\sqrt{N}}$ ,  $i = 1, 2, 3$ , have a Gaussian distribution for their typical<sup>4</sup> values when  $N$  is large, since the bit values within each of the three output streams are iid Bernoulli random variables with parameter  $1/2$ . Using the Central Limit Theorem, we conclude that  $\hat{w}_1$ ,  $\hat{w}_2$  and  $\hat{w}_3$ , which are sum of  $N$  iid random variables normalized by  $\sqrt{N}$ , have a Gaussian distribution with mean  $\sqrt{N}/2$  and variance  $1/4$  for large values of  $N$ . This approximation is most accurate for normalized weight near  $\sqrt{N}/2$ .

 in order to have a set of jointly Gaussian weight distributions, both the marginal and the conditional distributions of the weights shall be Gaussian. When the systematic weight  $w_1$  is known, the parity bits are no longer independent from 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  $\hat{w}_2$  and  $\hat{w}_3$  given  $\hat{w}_1$  are Gaussian for the typical values of  $\hat{w}_1$ . Since the marginal distributions are Gaussian,  $\hat{w}_1$ ,  $\hat{w}_2$  and  $\hat{w}_3$  constitute a set of jointly Gaussian random variables.

**Definition:**  $m$ -dependent sequence [30]: A sequence  $X_1, X_2, \dots$  of random variables is called  $m$ -dependent if and only if  $\{X_{a-r}, X_{a-r+1}, \dots, X_a\}$  and  $\{X_b, X_{b+1}, \dots, X_{b+s}\}$  are independent sets of variables when  $b - a > m$ ; that is, an  $m$ -dependent sequence is a sequence of dependent random variables for which the dependency lasts, at most, for  $m$  elements.

*Theorem 1: Central Limit Theorem for the sum of dependent random variables* [30]: If  $X_1, X_2, \dots$  is a sequence of  $m$ -dependent, uniformly bounded random variables and  $S_n = X_1 + X_2 + \dots + X_N$ , with the standard deviation  $V_N$ . Then, if  $V_N/N^{1/3} \rightarrow \infty$  as  $N \rightarrow \infty$ ,  $\overline{G}_N(x) \rightarrow \Phi(x)$  for all

<sup>4</sup>Typical weights are weights which are far from the boundaries, i.e.  $\lim_{N \rightarrow \infty} w_i/N \neq 0, 1$ .

$x$ , as  $N \rightarrow \infty$ , where  $\overline{G}_N$  is the cumulative distribution function (CDF) of  $\{S_N - E(S_N)\}/V_N$  and  $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$ .

As indicated by 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  $N$ , 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 1:* Given that the systematic weight is  $w_1$ , each parity stream is an  $m$ -dependent sequence, and the variance of its weight is given by

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

*Proof:* See the appendix B. ■

With this proposition and Theorem 1, the conditional parity weight distributions given the normalized systematic weight  $\hat{w}_1$ , asymptotically become Gaussian. A similar approach is valid for the conditional weight distribution of  $\hat{w}_3$ , given  $\hat{w}_1$  and  $\hat{w}_2$ . As a result,  $\hat{w}_1$ ,  $\hat{w}_2$  and  $\hat{w}_3$  are 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 mean and the variance of  $\hat{w}_i$  are  $\sqrt{N}/2$  and  $1/4$ , respectively, for  $i = 1, 2, 3$ . The correlation coefficients between  $\hat{w}_i$  and  $\hat{w}_j$  denoted by  $\rho_{ij}$ ,  $i, j = 1, 2, 3$ , can be written as

$$\rho_{ij} = \frac{\overline{\hat{w}_i \hat{w}_j} - \overline{\hat{w}_i} \overline{\hat{w}_j}}{\sigma_{\hat{w}_i} \sigma_{\hat{w}_j}} = 4 \left[ \overline{\hat{w}_i \hat{w}_j} - \frac{N}{4} \right], \quad (2)$$

and

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

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

$$\mu_{\hat{w}} = 3 \frac{\sqrt{N}}{2}, \quad (4)$$

and variance,

$$\sigma_{\hat{w}}^2 = \frac{3 + 2\rho_{12} + 2\rho_{13} + 2\rho_{23}}{4}. \quad (5)$$

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

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

*Proof:* Any of the pairs  $b_i(m)$ ,  $b_j(n)$  for  $i, j = 1, 2, 3$  and  $m, n = 1, \dots, N$ , can take four different values,  $\{00, 01, 10, 11\}$ . The set of the input sequences that result 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 which means the two bits are identical, or there are three cosets with the coset leaders 01, 10 and 11, which means with equiprobable input sequences, these two bits are independent. 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 probability  $1/2$ , resulting in  $\overline{b_i(m)b_j(n)} = 1/2$ , so that

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

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

1/4, so that

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

In both cases, we have

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

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

*Theorem 3:*  $\rho_{12}, \rho_{13} \rightarrow 0$  as  $N \rightarrow \infty$ .

*Proof:* As indicated in Theorem 2, for linear binary codes, if the input sequences is a set of independent Bernoulli random variables with probability 1/2, every pair of bits (systematic or parity) are either identical or independent. Positive cross-correlation between Hamming weights of different streams occurs only for identical pairs. A parity bit can be equal to a systematic bit if it is triggered only by that systematic bit. Within  $N$  bits in each parity stream, the number of bits which are triggered by only one systematic bit is upper bounded by  $P$ . Therefore,  $\rho_{12}$  and  $\rho_{13}$ , are upper bounded by  $P/N$  and tend to 0 as  $N \rightarrow \infty$ . ■

*Theorem 4:*  $\rho_{23} \rightarrow 0$  for  $N \rightarrow \infty$  with probability one (for almost any random interleaver).

*Proof:* With the same approach, we are looking for parity bits in the two parity streams which are always identical. This happens for parity bits which are triggered by exactly the same set of systematic bits. 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 for  $|m - n| < P$ , 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. Given  $m$  and  $m - P < n < m + P$ , the probability that a randomly chosen interleaver maps  $\mathcal{R}_2(m)$  to  $\mathcal{R}_3(n)$  is

$$\frac{1}{\binom{N}{|\mathcal{R}_2(m)|}}, \quad (9)$$

if  $|\mathcal{R}_2(m)| = |\mathcal{R}_3(n)|$  and is 0, otherwise. The number of  $(m, n)$  pairs satisfying  $|\mathcal{R}_2(m)| = |\mathcal{R}_3(n)|$  is upper bounded by  $2NP$ . For  $m > P$ , (9) goes to 0 faster than  $N^{-(P+1)/2}$ , while the number of



pairs grows only linearly with  $N$  and hence, such identical pairs are negligible compared to the total number of pairs of parity bits<sup>5</sup>. Therefore, for any randomly chosen interleaver,  $\rho_{23} \rightarrow 0$  as  $N \rightarrow \infty$  with probability 1. ■

As a result, the typical weight distribution of turbo codes is not a function of the chosen RCC and interleaver (for nontrivial MLS RCCs and interleavers), and hence, the interleaver optimization has a diminishing effect on the asymptotic ML performance of the turbo code in its waterfall region. Note that although the RCC optimization does not improve the asymptotic weight distribution of the code and hence the performance in the waterfall region with ML decoding, it may affect the performance of the iterative decoding algorithm because different constituent codes result in different extrinsic information transfer (EXIT) charts [31]–[33].

### C. Tangential sphere bound for Gaussian spectrum

Poltyrev shows that for codes with average spectrum, tangential sphere bounding provides an error exponent which is very close to that of capacity-achieving random coding [13]. Average spectrum is defined as [13]

$$\overline{A}_w = \begin{cases} 2^{n(h(\omega) - h(\alpha)) + o(n)} & w = \omega n \geq \alpha n \\ 0 & w = \omega n < \alpha n \end{cases}, \quad (10)$$

where  $n = N/R$  is the code length and  $0 < \alpha < \frac{1}{2}$  is the root to the following equation

$$R = 1 - h(\alpha) = 1 + \alpha \log_2(\alpha) + (1 - \alpha) \log_2(1 - \alpha). \quad (11)$$


The Gaussian distribution, described by

$$A_\omega = \frac{2^{nR}}{\sqrt{\frac{\pi}{2}}} \exp \left[ -2n \left( \omega - \frac{1}{2} \right)^2 \right], \quad (12)$$

is slightly different from the average spectrum given by (10). The Gaussian weight distribution predicts a nonzero number of codewords in the region  $0 < w < \alpha n$ . Here, we apply the TSB on the error



<sup>5</sup>This is true neglecting a few parity bits at the beginning of the two streams which have negligible impact on the overall performance.



probability for the Gaussian distribution to  predict the performance of the code in the waterfall region where the dominant codewords are in the typical region. In the next section, we show that the weight distribution for low weights does not follow the Gaussian distribution and then, we evaluate the effect of the low weight codewords on the overall performance of the code.

The Gallager region in TSB is a cone whose apex is located at the origin and its axis denoted by the  $Z$  axis connects the origin to the all-zero codeword. We normalize the space by dividing each axes by  $\sqrt{n}$ . Note that with BPSK signalling, the all-zero codeword is located at  $\left(\sqrt{\frac{E_N}{n}}, \sqrt{\frac{E_N}{n}}, \dots, \sqrt{\frac{E_N}{n}}\right)$ , where  $E_N$  is the energy per channel use. This cone is produced by rotating the line  $r = z \tan \theta$  about the  $Z$  axis, where  $r$  is the distance to the  $Z$  axis in the polar coordinates. Note that all codewords are on the surface of an  $n$ -dimensional sphere with radius  $\sqrt{E_N}$ . The Euclidean distance between two codewords in the normalized space is  $2\sqrt{\frac{w_d E_N}{n}}$ , where  $w_d$  is the Hamming distance between the two codewords.

We use TSB to determine the error exponent for the Gaussian weight spectrum. To that end we need the following lemma<sup>6</sup>.

*Lemma 1:* For large even integer  $n$ , if  $Y$  is a chi-squared random variable with mean  $n$  and  $n$  degrees of freedom, then for  $y = \beta n > n$ ,

$$P\{Y > y\} < \frac{n e^{-y/2} (y/2)^{n/2-1}}{(n/2)!} = O\left(\exp\left\{-\frac{n}{2}(\beta - 1 - \log \beta)\right\}\right). \quad (13)$$

 *Proof:*  See Appendix A. ■

*Theorem 5:* The probability of error for a code of rate  $R$  of length  $n$  and  $N = nR$  information bits whose weight distribution is given by (12) approaches zero as  $N \rightarrow \infty$ , if

$$E_2 = \min_{0 < \omega < \frac{\sqrt{N_0/2}}{\sqrt{E_N} + \sqrt{N_0/2}}} \left\{ 2 \left( \omega - \frac{1}{2} \right)^2 - \frac{1}{2} \log \left( 1 - \frac{2\omega}{1 - \omega} \frac{E_N}{N_0} \right) \right\} - R \log 2 > 0, \quad (14)$$

where  $E_N$  is the energy per channel use and  $N_0$  is the one-sided noise spectrum.

<sup>6</sup>A similar result without proof can be found in [34].

*Proof:* Consider a thin disk of radius  $c = \sqrt{E_N} \tan \theta$  and height  $\epsilon \rightarrow 0$  around the all-zero codeword as shown in Figure 4. Note that the surface of the disk is an  $n - 1$  dimensional sphere perpendicular to the  $Z$  axis. This disk is the portion of the cone which is confined by  $\sqrt{E_N} - \frac{\epsilon}{2} < Z < \sqrt{E_N} + \frac{\epsilon}{2}$ .

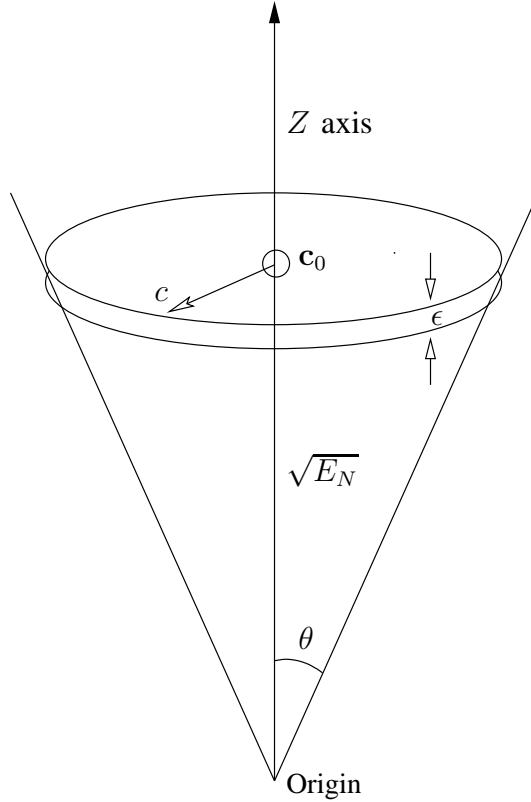


Fig. 4. Gallager region used for TSB.

Each dimension is normalized by  $\sqrt{n}$ . Therefore, the noise component on each dimension is asymptotically zero. Hence, the probability that the received vector falls inside the disk given that the all-zero codeword is transmitted is equivalent to the probability that it falls inside the entire cone. In other words, the cone and the disk around the all-zero codeword are the same in  $n - 1$  dimensions and differ in only one dimension and the noise component along that dimension is zero with probability one.

If the thin disk is used as the Gallager region, assuming that the all-zero codeword is transmitted, we obtain the Gallager upper bound as

$$P_e \leq P\{\mathbf{r} \notin \mathfrak{R}\} + \sum_{i \neq 0} P\{\mathbf{r} \in \mathfrak{R}, |\mathbf{r} - \mathbf{c}_i| < |\mathbf{r} - \mathbf{c}_0|\}, \quad (15)$$



where  $\mathfrak{R}$  is the Gallager region (i.e., the thin disk),  $\mathbf{c}_0$  is the all-zero codeword,  $\mathbf{r} = \mathbf{c}_0 + \mathbf{n}$  is the received vector ( $\mathbf{n}$  is the noise vector) and the summation is over all nonzero codewords  $\mathbf{c}_i$  whose median planes with the all-zero codeword intersect with the Gallager region. In the following, we find an upper bound for each summand in the error probability bound in (15) and its associated error exponent. The probability of error converges to zero as  $N \rightarrow \infty$  if all error exponents are positive.

The probability that the received vector is outside the Gallager region, given that the all-zero codeword is transmitted, is upper bounded by

$$\begin{aligned} P\{\mathbf{r} \notin \mathfrak{R}\} &\leq P\{|n_1| > \sqrt{n\epsilon}/2\} + P\left\{\frac{1}{n} \sum_{i=2}^n n_i^2 > c^2\right\} \\ &< P\{|n_1| > \sqrt{n\epsilon}/2\} + P\left\{\frac{1}{n} \sum_{i=1}^n n_i^2 > c^2\right\}, \end{aligned} \quad (16)$$

where  $n_1 = n_Z$  is the noise component along the  $Z$  axis,  $\sum_{i=1}^n n_i^2$  is the total noise energy in all  $n$  dimensions and  $c$  is the radius of the thin disk. The noise component along the  $Z$  axis is a zero mean Gaussian random variable with variance  $N_0/2$ , where  $N_0$  is the one-sided noise power spectrum.

Therefore,

$$P\{|n_1| > \sqrt{n\epsilon}/2\} = 2Q\left(\sqrt{\frac{n\epsilon^2}{2N_0}}\right) < \exp\left(-\frac{n\epsilon^2}{4N_0}\right), \quad \text{for } \epsilon > 0. \quad (17)$$

This indicates a positive exponent of  $E_0 = \epsilon^2/4N_0$  for  $\epsilon > 0$ .

On the other hand, the total noise energy is a *chi-squared* random variable with  $n$  degrees of freedom and mean  $nN_0/2$ . Using Lemma 1, the error exponent for the second summand in (16) can be obtained as

$$E_1 = \frac{1}{2} \left( \frac{c^2}{N_0/2} - 1 - \log \frac{c^2}{N_0/2} \right), \quad \text{for } c > \sqrt{N_0/2}. \quad (18)$$

Next, we find an upper bound on the second summand in (15), which is equal to

$$\lim_{n \rightarrow \infty} \sum_{i \neq 0} P\{\mathbf{r} \in \mathfrak{R}, |\mathbf{r} - \mathbf{c}_i| < |\mathbf{r} - \mathbf{c}_0|\} = \int_0^{\omega_{\max}} A_\omega P\{\mathbf{r} \in \mathfrak{R}, |\mathbf{r} - \mathbf{c}_\omega| < |\mathbf{r} - \mathbf{c}_0|\} d\omega, \quad (19)$$

where  $\mathbf{c}_\omega$  is a codeword of weight  $w = \omega n$  and  $\omega_{\max} = \frac{c}{\sqrt{E_N} + c} = \frac{\tan \theta}{1 + \tan \theta}$ , because only the median planes between the all-zero codeword and codewords of weight  $0 < w < n\omega_{\max} = n \frac{c}{\sqrt{E_N} + c}$

intersect with the Gallager region as shown in Figures 5 and 6. Note that  $P\{\mathbf{r} \in \mathfrak{R}, |\mathbf{r} - \mathbf{c}_\omega| < |\mathbf{r} - \mathbf{c}_0|\}$  represents the probability that the received vector falls in the shaded area in Figure 5 given that the all-zero codeword is transmitted. Since, the disc is very thin (i.e.,  $\epsilon \simeq 0$ ), the distance from the all-zero codeword to the the intersection of the Gallager region and the median plane is  $\sqrt{\frac{\omega E_N}{1-\omega}}$ .

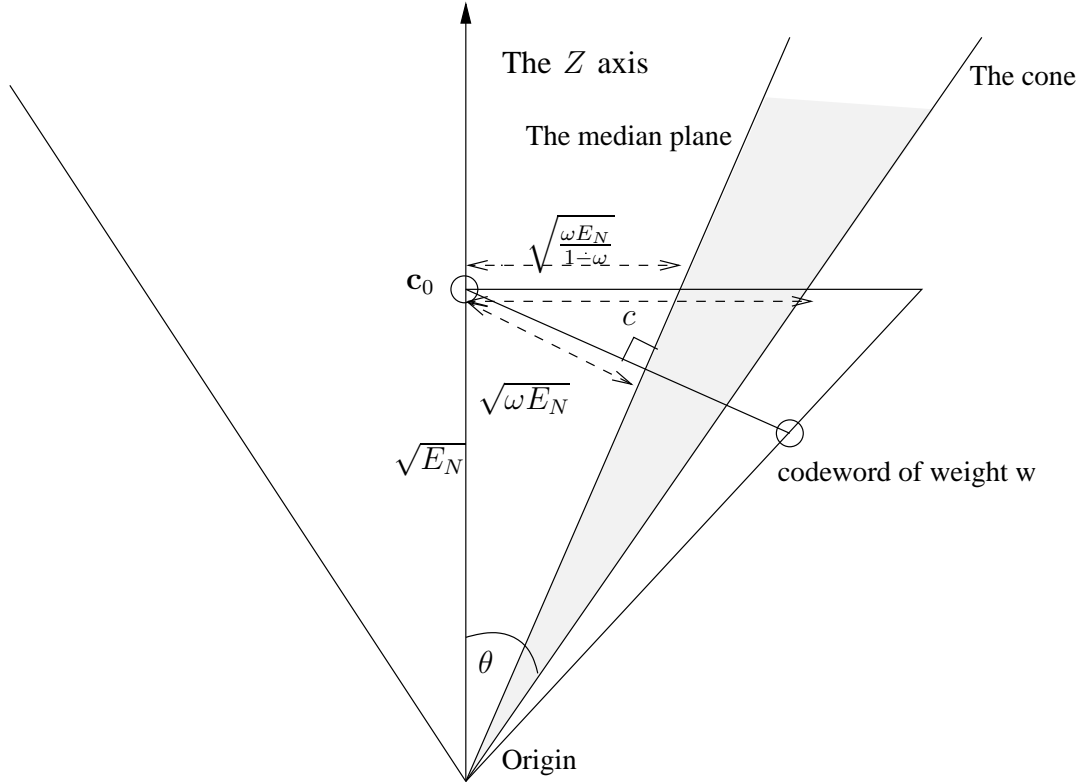


Fig. 5. The median plane between the all-zero codeword and a codeword of weight  $w = \omega n$  (side view).

If  $c$  is chosen to be  $\sqrt{N_0/2} + \epsilon$ ,  $\epsilon \rightarrow 0$ , the error exponent defined by (18) is positive for  $\epsilon > 0$ . On the other hand, if we omit the height of the thin disk and the noise component along with the  $Z$  axis,  $n_Z$ , the disk transforms to the  $n - 1$  dimensional noise sphere. For large  $n$ , the  $n - 1$  dimensional normalized white Gaussian noise is uniformly distributed within a sphere with radius  $\sqrt{N_0/2}$  [35]. The volume of the noise sphere (noted by  $\mathcal{S}_1$  in Figure 6) is  $\frac{\pi^{(n-1)/2}}{\Gamma(\frac{n+1}{2})}(N_0/2)^{(n-1)/2}$ . Note that for  $\sqrt{N_0/2} + \epsilon$ ,  $\epsilon \rightarrow 0$ , the cone approaches the  $n - 1$  dimensional noise sphere and becomes tangent to it. The intersection of the median plane between the all-zero codeword and a codeword of weight  $w = \omega n$  with the  $n - 1$  dimensional noise sphere confines a cap (the shaded area,  $\mathcal{S}_2$ , in Figure 6 noted

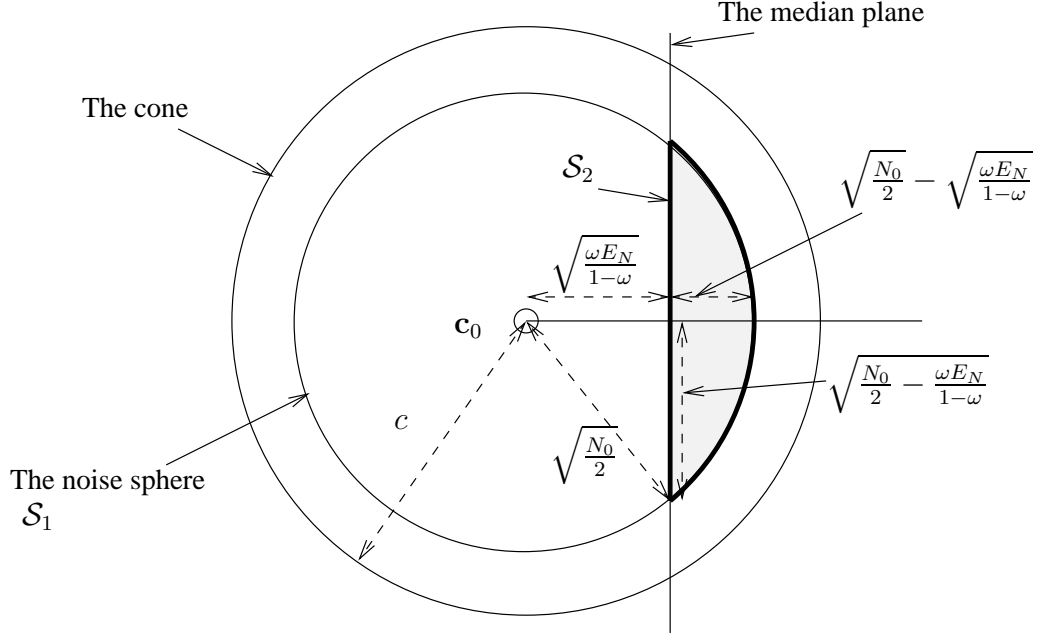


Fig. 6. The intersection of the median plane and the Gallager region (top view).

by  $\mathcal{S}_2$ ) with radius  $\sqrt{\frac{N_0}{2} - \frac{\omega}{1-\omega} E_N}$  and height  $\sqrt{\frac{N_0}{2} - \frac{\omega}{1-\omega} E_N}$ . If the received vector given that the all-zero codeword is transmitted falls inside this cap, an error occurs. The volume of this cap is upper bounded by the volume of a  $n-1$  dimensional sphere with radius  $\sqrt{\frac{N_0}{2} - \frac{\omega}{1-\omega} E_N}$ , which is  $\frac{\pi^{(n-1)/2}}{\Gamma(\frac{n+1}{2})} \left( \frac{N_0}{2} - \frac{\omega}{1-\omega} E_N \right)^{(n-1)/2}$ . Note that this upper bound is tight since the sphere cap and the sphere of radius  $\sqrt{\frac{N_0}{2} - \frac{\omega}{1-\omega} E_N}$  differ in only one dimension. Comparing the volume of the sphere cap and the noise sphere, for a codeword  $\mathbf{c}_i$  of weight  $w = \omega n$ ,

$$P\{\mathbf{r} \in \mathfrak{R}, |\mathbf{r} - \mathbf{c}_i| < |\mathbf{r} - \mathbf{c}_0|\} = \frac{\text{vol}\{\mathcal{S}_2\}}{\text{vol}\{\mathcal{S}_1\}} < \frac{\frac{\pi^{(n-1)/2}}{\Gamma(\frac{n+1}{2})} \left( \frac{N_0}{2} - \frac{\omega}{1-\omega} E_N \right)^{\frac{n-1}{2}}}{\frac{\pi^{(n-1)/2}}{\Gamma(\frac{n+1}{2})} (N_0/2)^{\frac{n-1}{2}}} = \left( 1 - \frac{2\omega}{1-\omega} \frac{E_N}{N_0} \right)^{\frac{n-1}{2}}, \quad (20)$$

whose exponent is

$$-\frac{1}{2} \log\left(1 - \frac{2\omega}{1-\omega} \frac{E_N}{N_0}\right). \quad (21)$$

Using the weight distribution described by (12) for codewords of weight  $0 < w < n \frac{\sqrt{N_0/2}}{\sqrt{E_N} + \sqrt{N_0/2}}$ , we

finally arrive at the exponent for the second summand in (15)

$$E_2 = \min_{0 < \omega < \frac{\sqrt{N_0/2}}{\sqrt{E_N} + \sqrt{N_0/2}}} \left\{ 2 \left( \omega - \frac{1}{2} \right)^2 - \frac{1}{2} \log \left( 1 - \frac{2\omega}{1 - \omega} \frac{E_N}{N_0} \right) \right\} - R \log 2. \quad (22)$$

Sine the error exponent in (18) is positive for  $c = \sqrt{N_0/2} + \epsilon$ , then the overall error exponent is positive if the error exponent in (22) is positive. ■

Figure 7 shows the minimum signal to noise ratio for which the TSB error exponent for the Gaussian weight spectrum given in (22) is positive and compares it to that of the average spectrum. It also shows the capacity of BPSK signalling over an AWGN channel. The achievable rates predicted by the TSB for the Gaussian spectrum and for the average spectrum are very close to the BPSK capacity for code rates less than 1/2. Note that Figure 7 only shows a lower bound on the achievable rates. For higher SNR values, where the TSB achievable rates deviate from the capacity of BPSK signalling, a tighter upper or lower bound is required to determine the actual achievable rates and to answer whether turbo codes are capacity-achieving codes.

Figure 8 shows the dominant weight  $\omega_c n$  on the error exponent in (22) for different SNR values:

$$\omega_c = \arg \min_{0 < \omega < \frac{\sqrt{N_0/2}}{\sqrt{E_N} + \sqrt{N_0/2}}} \left\{ 2 \left( \omega - \frac{1}{2} \right)^2 - \frac{1}{2} \log \left( 1 - \frac{2\omega}{1 - \omega} \frac{E_N}{N_0} \right) \right\}. \quad (23)$$

As the SNR increases, the error exponent is dominated by the error probability due to the codewords of lower weights. For  $E_N/N_0 > 2$  (=3 dB), the dominant weight in (22) becomes zero as shown in Figure 8 and the Gaussian approximation is no longer valid. In the Appendix C, we use the union bound to evaluate the cutoff rate for a Gaussian weight spectrum and provide more insight to the region where the Gaussian assumption is acceptable.

The derivations in this section remain valid for parallel concatenated turbo codes with  $J > 2$  component codes as all the systematic and parity weights are Gaussian and each parity stream is an  $m$ -dependent sequence conditioned on the weight of the systematic and the other parity streams. If one punctures one or more of the parity streams to increase the code rate, the Central Limit Theorem is still applicable, if the puncturing leaves an infinite number of parity bits when  $N \rightarrow \infty$ .

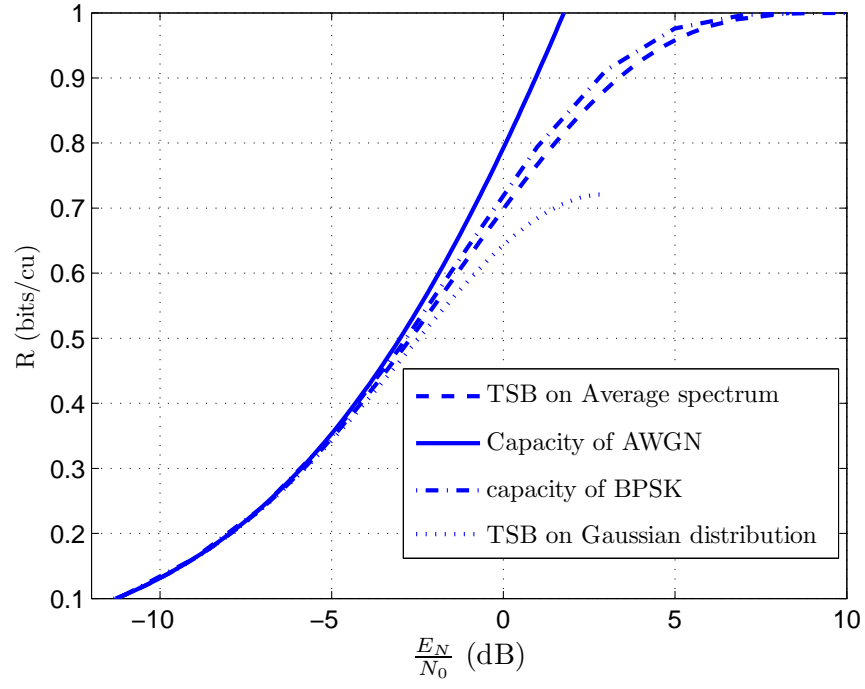


Fig. 7. TSB bound vs capacity.

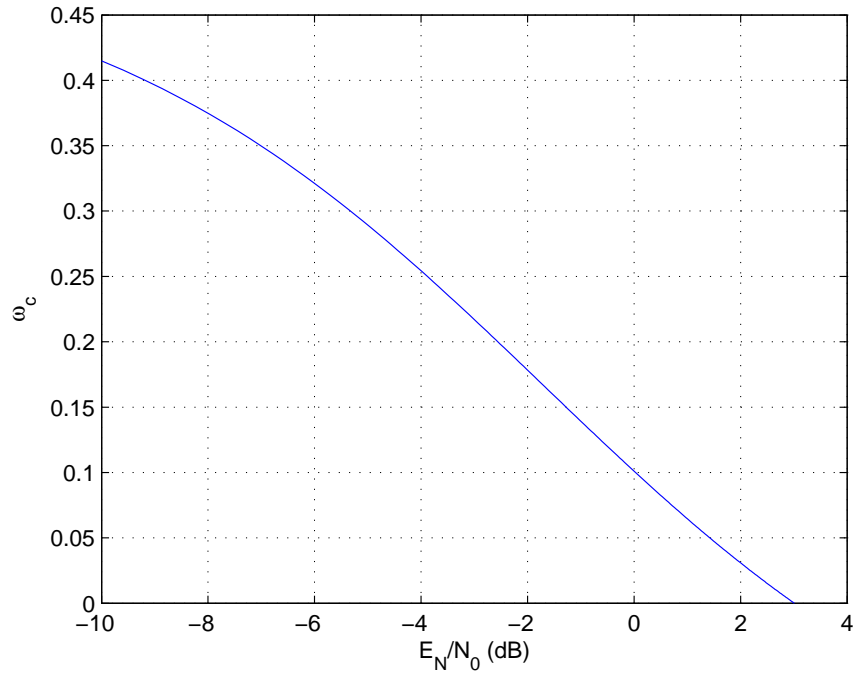





Fig. 8. Dominant weight in the error exponent evaluation.

The Gaussian weight distribution approximation is valid for the typical values of the Hamming weight. As the SNR increases, the ML error performance is dominated by the codewords of lower weights. The number of low-weight codewords cannot be approximated by a continuous distribution. As we will see in Section III, low-weight codewords appear only in certain structures. We next study the statistical properties of low weight codewords and their effect on the overall performance.

### III. ASYMPTOTIC PERFORMANCE OF TURBO CODES IN THE ERROR FLOOR REGION

#### A. Asymptotic statistics of low-weight codewords

Consider the turbo code shown in Figure 1 with two component codes. Lemma 1 in [17] proves that in an asymptotically large turbo code, the probable low-weight codewords consist of some short single error events<sup>7</sup> with the systematic weight of two in both RCCs. It shows that the average number of low weight codewords of other structures tends to 0 as the block length increases. Each of these short error events is 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  $w_1 = 2M$ ,  $M \in \mathbb{N}$ . Each RCC leaves the all-zero state  $M$  times and returns to it after an integer multiple of  $P$  transitions. 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 total number of RCC impulse response repetitions in the parity check sequences. Such a structure is denoted by type  $(M, K)$  where  $K \geq 2M$ . For a code consisting of  $J$  constituent codewords, the low-weight codeword of type  $(M, K)$  consists of  $M$  short error events in each parity stream. The systematic stream and all its  $J - 1$  interleaved versions in a code with  $J$  component codes contain  $M$  pairs of ones, where each pair  separated by an integer multiple of  $P$  as shown in Figure 9.

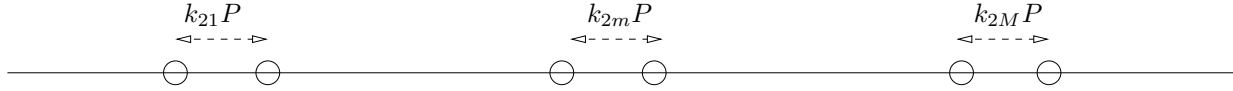
To calculate the mean and the variance of the error floor, it is necessary to compute the statistical properties of each low-weight structure.  One can show that there are  $\binom{K-1}{2M-1}$  ways to choose  $2M$

<sup>7</sup>A single error event means leaving the zero-state and returning back to it for the first time  after a few trellis transitions.

Before Interleaving

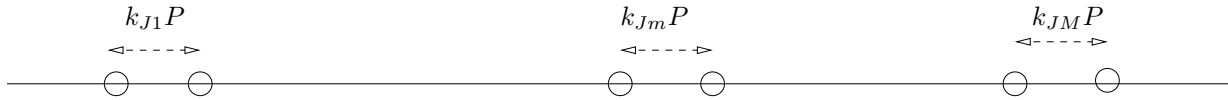


After first Interleaver



⋮

After Interleaver  $J - 1$



$$\sum_{j=1}^J \sum_{m=1}^M k_{jm} = K$$

Systematic stream consisting of  $M$  pairs of ones (ones are shown by circles, zeros elsewhere)

Fig. 9. The structure of low-weight codewords of type  $(M, K)$ .

positive integers whose sum is  $K$ . Equivalently, the structure of type  $(M, K)$  can be divided into  $\binom{K-1}{2M-1}$  substructures. The number of codewords of each substructure has the same statistical properties as the number of codewords of type  $(M, 2M)$ , when  $N \rightarrow \infty$ .

**Theorem 6:** The number of codewords of type  $(M, K)$  is a Poisson random variable with parameter  $\lambda_{M,K}^{(2)} = \binom{2M}{M} \binom{K-1}{2M-1}$ , where superscript (2) is used to denote the presence of two component codes.

**Proof:** We first calculate the statistical properties of the number of codewords of type  $(M, 2M)$  and then generalize the result to the other structures. There are asymptotically<sup>8</sup>  $\binom{N}{M}$  systematic input

<sup>8</sup>The exact number of such combinations is  $\binom{N - M(P+1)}{M}$  as the first systematic bits of the  $M$  pairs should be separated by  $P + 1$  bit positions, and for large  $N$ ,  $\lim_{N \rightarrow \infty} \binom{N - M(P+1)}{M} = \binom{N}{M}$ .

combinations consisting of  $M$  pairs of ones, each pair with  $P - 1$  zeroes in between. This can be verified by determining the place of the first element of each pair. The overlapping pairs are neglected, because  $N \gg M$ . Such a structure generates  $M(P + 1)/2$  parity 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  $\binom{N}{2M}$  ways to interleave a stream of weight  $2M$ . However, among them, only asymptotically  $\binom{N}{M}$  result in  $M$  pairs of ones, each pair with  $P - 1$  zeroes in between; that is, the Bernoulli event that a low parity-weight generating stream changes to another one occurs with probability  $\frac{\binom{N}{M}}{\binom{N}{2M}}$ . The number of these Bernoulli events is  $\binom{N}{M}$ . These Bernoulli events are asymptotically independent because occupying a bit position in the interleaved stream by a certain information bit does not asymptotically affect the probability for the other bits. As a result, the number of low-weight codewords of type  $(M, 2M)$  is asymptotically a Poisson random variable with parameter

$$\lambda_{M,2M}^{(2)} = \frac{\binom{N}{M}}{\binom{N}{2M}} \binom{N}{M}. \quad (24)$$

For  $N \rightarrow \infty$ , (24) converges to

$$\lambda_{M,2M}^{(2)} = \binom{2M}{M}. \quad (25)$$

Finally, the Poisson parameter for the structure of type  $(M, K)$  is computed by multiplying  $\lambda_{M,2M}^{(2)}$  by  $\binom{K-1}{2M-1}$ , which is

$$\lambda_{M,K}^{(2)} = \binom{2M}{M} \binom{K-1}{2M-1}. \quad (26)$$

■

### B. Asymptotic statistics of indecomposable low-weight codewords

In a linear binary codebook<sup>9</sup>, 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 ignored,

<sup>9</sup>A codebook is defined as the set of all possible codewords.



because: (i) the decomposable codewords do not contribute to the walls of the Voronoi region of the all-zero codeword, and (ii) if each of the original low-weight codewords is expurgated, the associated decomposable codewords no longer exist.

The Poisson parameters calculated by (26) include both decomposable and indecomposable low-weight codewords. These Poisson random variables are not independent. For example, the number of codewords of type  $(M_1 + M_2, K_1 + K_2)$  depends on the number of codewords of types  $(M_1, K_1)$  and  $(M_2, K_2)$ .

*Theorem 7:* The number of indecomposable codewords of type  $(M, K)$  is a Poisson random variable with parameter  $\hat{\lambda}_{M,K}^{(2)} = \frac{2^{2M}}{2M} \binom{K-1}{2M-1}$ .

*Proof:* Again, we begin with codewords of type  $(M, 2M)$ . A codeword of type  $(M, 2M)$  consists of  $2M$  systematic bits. There are asymptotically  $\binom{N}{2M}$  ways to choose  $2M$  bits out of the  $N$  systematic bits. Consider these bits as the  $2M$  nodes of a graph. These bits form  $M$  pairs before and  $M$  pairs after interleaving. We denote each pair before interleaving by a red edge and each pair after interleaving by a blue edge. A graph with two edges for each node consists of one or more loops. Each loop represents an indecomposable codeword. We have one and only one indecomposable codeword of type  $(M, 2M)$ , if and only if there is only one loop in the graph. There are  $(2M-1)!$  ways to form only one loop with  $2M$  nodes in such a way that each node has one blue edge and one red edge. For each edge in the graph, the probability that the corresponding systematic bits are separated by  $P$  trellis positions is  $2/N$ . Since the relative position of bits in different pairs are asymptotically independent, all pairs are separated by  $P$  before and after interleaving with probability  $\left(\frac{2}{N}\right)^{2M}$ . The number of low-weight codewords of type  $(M, 2M)$  is the summation of many Bernoulli events with a low probability which is a Poisson random variable. Noting the above statements, the parameter of this random variable is

$$\hat{\lambda}_{M,2M}^{(2)} = \binom{N}{2M} (2M-1)! \left(\frac{2}{N}\right)^{2M}. \quad (27)$$

For large  $N$ , (27) can be written as

$$\hat{\lambda}_{M,2M}^{(2)} = \frac{2^{2M}}{2M}. \quad (28)$$

The Poisson parameter for the number of indecomposable low-weight codewords of type  $(M, K)$  is the multiplication of  $\hat{\lambda}_{M,2M}^{(2)}$  by  $\binom{K-1}{2M-1}$ :

$$\hat{\lambda}_{M,K}^{(2)} = \frac{2^{2M}}{2M} \binom{K-1}{2M-1}. \quad (29)$$

■

The Poisson random variables describing the number of low-weight codewords of different structures are asymptotically independent. To show this independency, we use the following theorem.

*Theorem 8:* For  $m \in \mathbb{N}$  ( $\mathbb{N}$  is the set of positive natural numbers), let  $\mu_1, \dots, \mu_m$  be non-negative numbers. For each  $N \in \mathbb{N}$ , let  $(X_1(N), \dots, X_m(N))$  be a vector of non-negative integer-valued random variables defined on the same space. If for all  $(r_1, \dots, r_m) \in \mathbb{N}^m$

$$\lim_{N \rightarrow \infty} E \{(X_1(N))_{r_1} \cdots (X_m(N))_{r_m}\} = \prod_{i=1}^m \mu_i^{r_i}, \quad (30)$$

where  $(x)_r = x(x-1) \cdots (x-r+1)$ , then the random vector  $(X_1(n), \dots, X_m(n))$  converges in distribution to a vector of independent Poisson random variables with mean  $(\mu_1, \dots, \mu_m)$  [36]

*Theorem 9:* Poisson random variables described by (29) are independent.

*Proof:* Now, we apply Theorem 8 to the Poisson random variables described in Theorem 7. Suppose that  $X_1, \dots, X_m$  represent  $m$  random variables describing the number of codewords of  $m$  different structures of type  $((M_1, K_1), \dots, (M_m, K_m))$ . For the  $i^{\text{th}}$  structure,  $X_i = \sum v_i$ , where  $v_i$  is the Bernoulli event of a particular finite subgraph in Theorem 7 making a low-weight codeword. For such random variables, (30) can be written as

$$\lim_{N \rightarrow \infty} E \{(X_1(N))_{r_1} \cdots (X_m(N))_{r_m}\} = \lim_{N \rightarrow \infty} E \left\{ \prod_{i=1}^m \sum_{\{i_1, \dots, i_r\}} v_{i_1} \cdots v_{i_r} \right\}. \quad (31)$$

For  $N \rightarrow \infty$ , different finite subgraphs form low-weight codewords independent from each other. This is because as the block length increases, the number of bit positions occupied by other pairs is negligible with respect to the block length. As a result,

$$\lim_{N \rightarrow \infty} E \{(X_1(N))_{r_1} \cdots (X_m(N))_{r_m}\} = \lim_{N \rightarrow \infty} \prod_{i=1}^m \sum_{\{i_1, \dots, i_r\}} E \{v_{i_1}\} \cdots E \{v_{i_r}\}. \quad (32)$$

Note that  $v_{ij}$ ,  $j = 1, \dots, r$  are iid random variables and as previously shown,

$$\sum_{i_j} E \{v_{i_j}\} = \mu_i = \hat{\lambda}_{M_i, K_i}^{(2)}, \quad (33)$$

where  $\mu_i = \hat{\lambda}_{M_i, K_i}^{(2)}$  is the corresponding Poisson parameter in (29) for the structure of type  $(M_i, K_i)$ .

Therefore,

$$\lim_{N \rightarrow \infty} E \{(X_1(N))_{r_1} \cdots (X_m(N))_{r_m}\} = \lim_{N \rightarrow \infty} \prod_{i=1}^m \mu_i^m, \quad (34)$$

an hence,  $X_1, \dots, X_m$  form a set of independent Poisson random variables. ■

Using Poisson parameters in (29), we can evaluate the asymptotic probability mass function of the minimum distance over all possible interleavers. Note that the set ~~of the structure~~ with the lowest weight and a nonzero cardinality determines the minimum distance. In other words, if these structures are sorted in the ascending order of their weights (i.e.,  $w_i \leq w_{i+1}$ ,  $i = 1, 2, \dots$ ) and  $Y_i$  is the number of low-weight codewords of the  $i^{\text{th}}$  structure, then the minimum distance of the code is  $w_i$  if

$$Y_j = 0, \quad j = 1, 2, \dots, i-1 \quad \text{and} \quad Y_i \neq 0. \quad (35)$$

The probability of this event can be obtained as

$$P\{w_{\min} = w_i\} = P\{Y_1 = 0, Y_2 = 0, \dots, Y_{i-1} = 0, Y_i \neq 0\}. \quad (36)$$

Since the number of indecomposable codewords of different types are independent Poisson random variables,

$$P\{w_{\min} = w_i\} = P\{Y_i \neq 0\} \prod_{j=1}^{i-1} P\{Y_j = 0\} = \exp \left\{ - \sum_{j=1}^{i-1} \lambda_j \right\} (1 - \exp \{-\lambda_i\}), \quad (37)$$

where  $\lambda_i$  denotes the Poisson parameter of random variable  $Y_i$ . Table I and Figure 10 show the pmf of the minimum distance of a large block turbo code with the 2, 3 or 4 memory elements in each RCC among all possible interleavers.

TABLE I  
ASYMPTOTIC PMF OF THE TURBO CODE MINIMUM DISTANCE AS  $N \rightarrow \infty$

$P = 3$		$P = 7$		$P = 15$	
$w_{\min}$	$Pr$	$w_{\min}$	$Pr$	$w_{\min}$	$Pr$
6	8.64e-001	10	8.64e-001	18	8.64e-001
8	1.32e-001	14	1.32e-001	26	1.32e-001
10	2.46e-003	18	2.47e-003	34	2.47e-003
12	6.14e-006	20	6.03e-006	36	6.03e-006
14	3.77e-011	22	1.12e-007	42	1.12e-007
16	1.92e-022	24	3.77e-011	44	3.77e-011
18	5.03e-045	26	4.24e-018	50	4.24e-018
20	1.76e-090	28	1.92e-022	52	1.92e-022
22	5.02e-186	30	8.19e-040	54	8.19e-040

### C. Error floor for large block size turbo codes

Using the results of the previous section, we can 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 minimum 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  $i^{\text{th}}$  structure is  $Y_i$  which is determined by a Poisson distribution with parameter  $\lambda_i$ . With the union bound, the error floor can be bounded as

$$P_e \leq P_u = \sum_i Y_i p_i, \quad (38)$$

where  $p_i = Q\left(\sqrt{\frac{2E_N w_i}{N_0}}\right)$  is the corresponding error for any codeword of the  $i^{\text{th}}$  structure, where  $E_N$  is the energy per channel use. The mean of this upper bound can be determined by

$$E[P_u] = \sum_i \lambda_i p_i. \quad (39)$$

As mentioned earlier, the upper bound becomes tighter if only indecomposable codewords are considered. On the other hand, since the Poisson random variables corresponding to the number of inde-

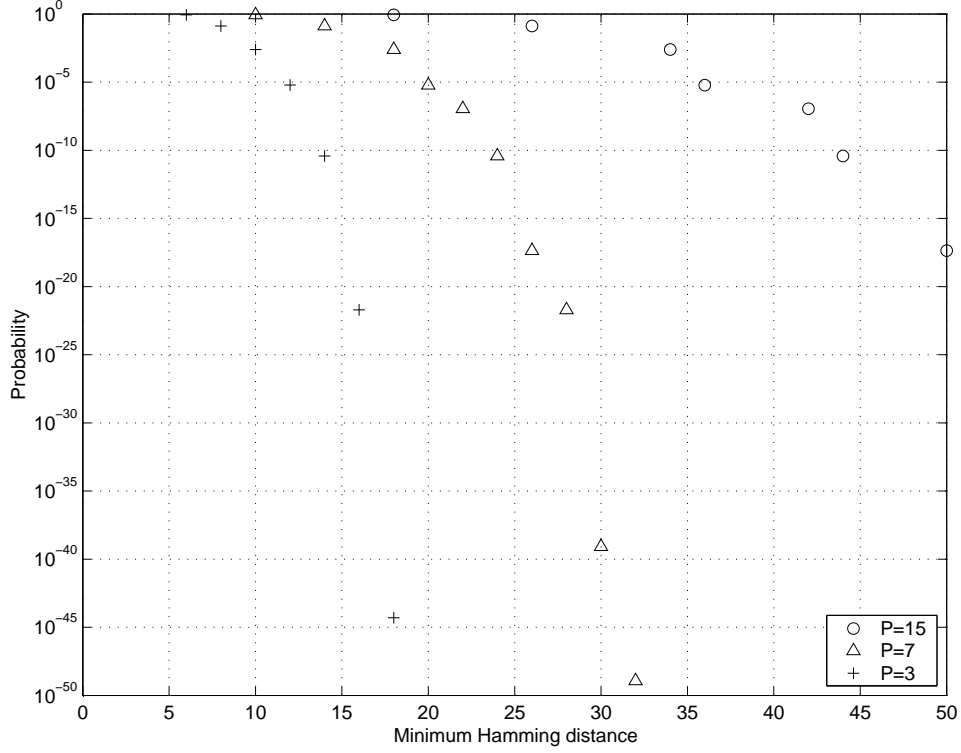


Fig. 10. Asymptotic pmf of the turbo code minimum distance as  $N \rightarrow \infty$ .

composable low-weight codewords are asymptotically independent, the variance of  $P_u$  can be evaluated by

$$\sigma_{P_u}^2 = \sum_i \lambda_i p_i^2. \quad (40)$$

In section III-D, we will study the convergence rate of (39) and (40) with respect to the block length.

Figure 11 shows the mean and the standard deviation of the union bound on the error floor with ML decoding, by using the Poisson parameters in (26). As expected, both the mean and the standard deviation decrease when the SNR increases. As the SNR increases, the ratio between them converges to  $\sqrt{2}$ . This is because for this region of signal to noise ratio values, only the codewords of the lowest weight structure, i.e., type (1,2), remain effective and the Poisson parameter for this type is two.

As mentioned above, we can obtain a tighter upper bound when we restrict the evaluation of the error performance to the indecomposable codewords only. In Figure 12, the union bound on the average error floor using the Poisson distribution of the indecomposable low-weight codewords in (29) for a code with

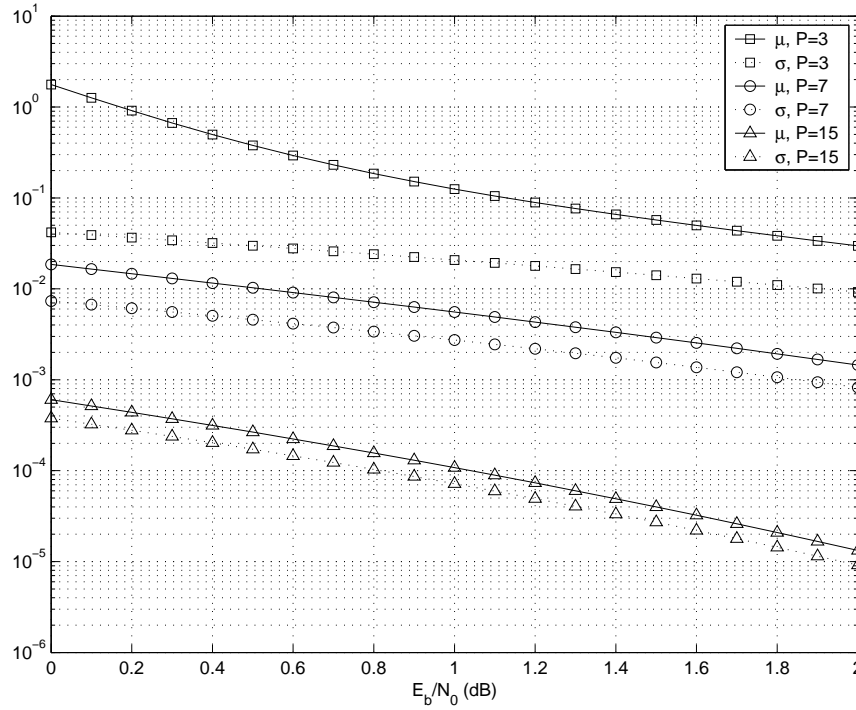


Fig. 11. The mean and standard deviation of the union bound on the error floor for  $P = 3, 7$  and  $15$ .

$P = 3$  is compared to the union bound evaluated by using Poisson parameters in (26) corresponding to all low-weight codewords.

The mean and the standard deviation of the performance of a turbo code of length 10000 and rate 1/3 consisting of two component codes each having four memory elements is evaluated by simulation by using 30 different pseudo-random interleavers and are shown in figure 13. They are also compared to the corresponding mean and standard deviation of the error floor in (39) and (40) for  $P = 15$ . For low SNR values, the high-weight codewords dominate the performance and the iterative decoder might not converge to the original transmitted codeword or a low-weight error. At higher SNR values, the low-weight codewords become dominant and the mean and the standard deviation of performance of the code follow (39) and (40).

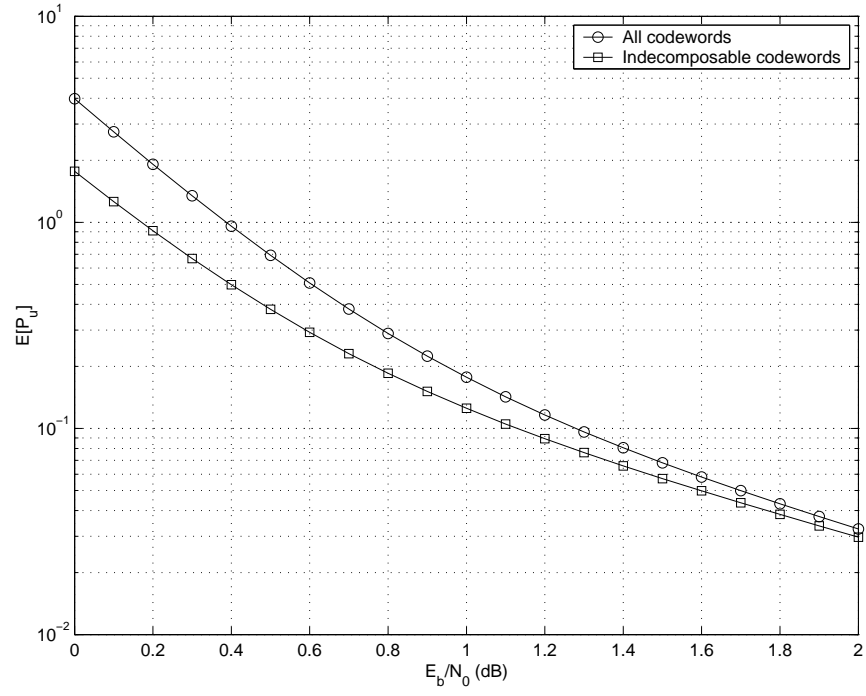


Fig. 12. The error floor for a large-block turbo-code with  $P = 3$ .

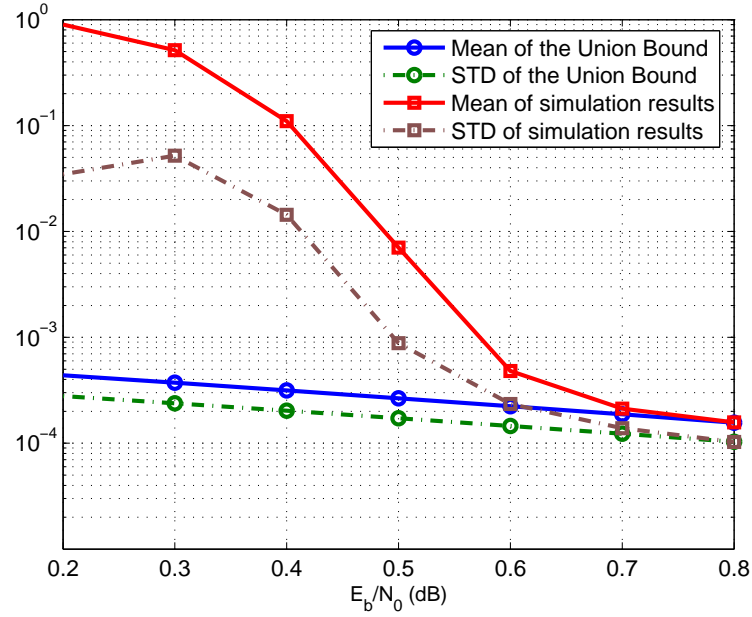


Fig. 13. The mean and standard deviation of FER for a code with 30 different randomly chosen interleavers ( $P = 15$ ).

#### D. Turbo codes with multiple constituent codes

In [26], it is shown that turbo codes with more component codes have a better performance when ML decoding is used. In [8], it is shown that for a randomly interleaved turbo code, the error floor decreases as  $O(N^{-J+2+\epsilon})$  when  $J$  component codes are used. In this section, we investigate the asymptotic behavior of multi-component turbo codes based on the Poisson distribution of low-weight codewords.

We focus on a parallel concatenated code consisting of  $J$  component codes. These codes are concatenated via  $J - 1$  randomly chosen interleavers. The rate of this code is  $\frac{1}{J+1}$ . Higher code rates ~~of less than one~~ can be achieved by puncturing and lower rates can be **obtained** by using component codes of rate less than one.

Again, we only need to concentrate on the codewords consisting of short single error events due to the systematic sequence of weight-two. This is because by using the same methods as in [17], **we can show** that the average number of codewords consisting of  $w_1$  systematic bits and  $A_j \leq \lfloor \frac{w_1}{2} \rfloor$ ,  $j = 1, \dots, J$  short error events in the  $j^{\text{th}}$  encoder ( $j = 1, \dots, J$ ) is given by  $O(N^{-w_1(J-1)+\sum_j A_j})$ . Within low weight codewords of even systematic weight  $w_1$ , those codewords with

$$A_j = M = \left\lfloor \frac{w_1}{2} \right\rfloor, \quad j = 1, \dots, J \quad (41)$$

have higher probability and hence are dominant in the performance. The parity weight of such codewords is  $\frac{K(P+1)}{2}$  where  $K \geq JM$ . For low-weight patterns that do not satisfy (41), **the average number of such codewords converges to 0 by the order of  $O(N^{-w_1(J-1)+\sum_j A_j})$ . For  $w_1 = 2M$ , this order is  $O(N^{-(MJ-2M+1)})$  or faster and for  $w_1 = 2M + 1$ , it converges to 0 as  $O(N^{-(MJ-2M+J-1)})$  or faster. As a result, for a given  $M$ , the average number of codewords that do not satisfy (41) is at least  $O(N^{-(MJ-2M+1)})$ .**

In the following, we find the order of the error floor based on these low-weight codewords. This error floor is mainly determined by codewords that satisfy (41). We also find the effect of codewords that do not satisfy (41) on the error performance and show that this term can be ignored in the overall performance.



We first study the statistical properties of codewords of weight  $2M + \frac{JM(P+1)}{2}$  (i.e.,  $K = JM$ ). The number of low-weight patterns of systematic weight  $2M$  is asymptotically<sup>10</sup>  $\binom{N}{M}$ . After each interleaver, only asymptotically  $\binom{N}{M}$  out of  $\binom{N}{2M}$  possible outcomes are still low-weight patterns. As a result, the average number of low-weight codewords with the systematic weight of  $2M$  and parity weight of  $\frac{JM(P+1)}{2}$  is

$$\lambda_{M,JM}^{(J)} = \frac{\binom{N}{M}^J}{\binom{N}{2M}^{J-1}} \simeq \frac{(2M!)^{J-1}}{(M!)^J} N^{-M(J-2)}. \quad (42)$$

For  $J = 2$ , e.g., a turbo code with two component codes,  $\lambda_{M,JM}^{(J)}$  is finite and non-zero. For  $J > 2$ , this average goes to zero as  $N$  increases. For structures with parity weight of  $\frac{K(P+1)}{2}$  where  $K \geq JM$ , we have

$$\lambda_{M,K}^{(J)} = \lambda_{M,JM}^{(J)} \binom{K-1}{JM-1}. \quad (43)$$

Using the union bound, the mean of the error floor can be bounded as

$$E\{P_e\} \leq \sum_M \sum_{K \geq JM} \lambda_{M,K}^{(J)} Q \left( \sqrt{\left(2M + \frac{K(P+1)}{2}\right) \frac{2E_N}{N_0}} \right) + \sum_M O(N^{-(MJ-2M+1)}). \quad (44)$$

~~The last term corresponds to codewords not satisfying (41).~~ The term  $\sum_M O(N^{-(MJ-2M+1)})$  corresponds to the codewords that do not satisfy (41). Therefore,

$$\begin{aligned} E\{P_e\} &\leq \sum_M \frac{(2M!)^{J-1}}{(M!)^J} N^{-M(J-2)} \sum_{K \geq JM} \binom{K-1}{JM-1} Q \left( \sqrt{\left(2M + \frac{K(P+1)}{2}\right) \frac{2E_N}{N_0}} \right) \\ &\quad + \sum_M O(N^{-(MJ-2M+1)}). \end{aligned} \quad (45)$$

For  $N \rightarrow \infty$ , both summations in (45) are dominated by the term corresponding to  $M = 1$ :

$$\begin{aligned} E\{P_e\} &\leq 2^{J-1} N^{-J+2} \sum_{K \geq J} \binom{K-1}{J-1} Q \left( \sqrt{\left(2 + \frac{K(P+1)}{2}\right) \frac{2E_N}{N_0}} \right) + O(N^{-(J-1)}) \\ &= O(N^{-J+2}), \end{aligned} \quad (46)$$

<sup>10</sup>The exact number of such combinations is  $\binom{N-M(P+1)}{M}$  as the first systematic bits of the  $M$  pairs should be separated by  $P+1$  bit positions. If  $\lim_{N \rightarrow \infty} \frac{M}{N} = 0$ , then  $\lim_{N \rightarrow \infty} \frac{\binom{N-M(P+1)}{M}}{\binom{N}{M}} = 1$ .

which indicates an interleaver gain of at least  $J - 2$ . A lower bound on the error probability can be found by considering only codewords of type  $(1, J)$ . With a uniform interleaver, at least one low weight codeword of type  $(1, J)$  exists with the asymptotic probability of

$$P_{(1,J)} = 1 - \left[ 1 - \left( \frac{N}{\binom{N}{2}} \right)^{J-1} \right]^N \simeq 1 - \left[ 1 - N \left( \frac{2}{N} \right)^{J-1} \right] = 2^{J-1} N^{-J+2}. \quad (47)$$

This is because there are asymptotically<sup>11</sup>  $N$  systematic pairs of distance  $P$ . The probability that a certain pair does not result in a low weight codeword is  $\left[ 1 - \left( \frac{N}{\binom{N}{2}} \right)^{J-1} \right]$  and these events are asymptotically independent.

$$E\{P_e\} > P_{(1,J)} Q \left( \sqrt{\left( 2 + \frac{J(P+1)}{2} \right) \frac{2E_N}{N_0}} \right) = O(N^{-J+2}). \quad (48)$$

This shows that the interleaver gain is lower and upper bounded by  $J - 2$  and hence, it is  $J - 2$ . For  $J = 2$ , The average error floor will be bounded below away from zero and the error term corresponding to low-weight codewords that do not satisfy (41) is  $O(N^{-1})$ . In other words, the mean and the variance of the union bound on the error floor converges to values in (39) and (40) on the order of  $O(N^{-1})$ . From (46), we also conclude that the error floor for turbo codes with uniform interleavers having more than two component codes decays inversely with the block length. This behavior is different from what we have seen in the waterfall region. Note that the performance of turbo code in the waterfall region is determined by high-weight codewords and the error probability in this region decays exponentially with the block length of the code. The BER for this code can be bounded by

$$E\{P_b\} \leq \sum_M \frac{(2M!)^{J-1}}{(M!)^J} N^{-M(J-2)-1} \sum_{K \geq JM} 2M \binom{K-1}{JM-1} Q \left( \sqrt{\left( 2M + \frac{K(P+1)}{2} \right) \frac{2E_N}{N_0}} \right). \quad (49)$$

Here, the effect of codewords that do not satisfy (41) is on the order of  $O(N^{-J})$  and hence negligible compared to the error probability. Equation (49) indicates an interleaver gain of  $J - 1$ . As a result, although the FER performance of a turbo code consisting of two component codes is bounded below by  $O(1)$ , its BER asymptotically tends to zero in the error floor region with almost any random interleaver.

<sup>11</sup>The exact number of such pairs is  $N - P$ .

If one punctures one or more of the parity streams to increase the code rate, the number of low-weight codewords remain unchanged but the weight of each codeword decreases. This increases the error floor in (45) and (49), but does not change the decay rate of the error floor for bit and frame error probabilities, which are  $O(N^{-J+2})$  and  $O(N^{-J+1})$ , respectively.

### E. Transition region

As discussed earlier, the weight distribution follows Gaussian and Poisson distributions for high (i.e.,  $w \sim O(N)$ ) and low weight (i.e., finite weight) codewords, respectively. There are codewords which cannot be categorized as part of Gaussian or Poisson distribution regions and fall between the two weight regions. This transition region is specified by codewords of weight  $w \sim o(N)$ . In this section, we study this transition region to cover the the gap between the Poisson and Gaussian regions. We show that this weight region has little effect on the overall performance of the code.

Because of RCC properties, error events<sup>12</sup> start with a nonzero systematic bit and end with another nonzero systematic bit. To account for the the codewords belonging to the transition region, we only need to consider the codewords with systematic weight of order  $o(N)$ . Note that, the effect of codewords with overall weight  $w \sim O(N)$  on the ML decoding performance is negligible<sup>13</sup>

We note that each error event starts and ends with a nonzero systematic bit, and consists of  $o(N)$  zero and a finite number<sup>14</sup> of nonzero systematic bits. Given that the systematic weight is  $w_1 \sim o(N)$ , the distribution of the systematic bits, which is given by Bernoulli  $(w_1/N)$ , is not altered by this fact.

We divide the systematic stream into  $w_1 - 1$  contiguous segments, such that each segment starts with a nonzero systematic bit and ends with the next nonzero systematic bit. In other words, the first segment starts with the first nonzero systematic bit and ends with the second one in the stream. The second segment starts with the second one and ends the third one in the stream, and so on. A nonzero

<sup>12</sup>An error event is defined by leaving the all-zero state and returning back to it for the first time

<sup>13</sup>The number of such codewords grows polynomially with the block length but the pairwise error probability between these code words and all-zero codeword decays exponentially.

<sup>14</sup>The average number of nonzero bits before the RCC comes back to the all-zero state is  $P$ .

systematic bit drives the RCC into a nonzero state, if it was in the all-zero state, and to one of  $P - 1$  non-zero states or the all-zero state, if it was in a nonzero state. As a result, each RCC in one of  $P$  nonzero states with probability  $\frac{P}{P+1}$ . Since  $w_1$  is very large, the law of the large number applies and in each RCC encoder, the encoder remains in the zero state for about  $\left\lfloor \frac{w_1}{P+1} \right\rfloor$  segments. Thus, with  $J$  component codes, the total number of nonzero segments<sup>15</sup> is about  $\left\lfloor \frac{P}{P+1} J w_1 \right\rfloor$ . Assuming that each systematic bit is one with probability  $\frac{w_1}{N}$ , the number of systematic bits in each segment is a geometric random variable with parameter  $\frac{w_1}{N}$ . As a result, the total number of systematic bits of the nonzero segments in all  $J$  RCC encoders is a negative binomial random variable with parameters  $\left( \left\lfloor \frac{P}{P+1} J w_1 \right\rfloor, \frac{w_1}{N} \right)$ . Note that the overall parity weight is approximately  $\left\lfloor \frac{P+1}{2P} x \right\rfloor$ , where  $x$  is the total number of bits (zero or nonzero) in the RCC output stream corresponding to nonzero segments. Since there are  $\binom{N}{w_1}$  different systematic inputs of weight  $w_1$ , the overall number of codewords of systematic weight  $w_1 \sim o(N)$  and parity weight  $w_p$  is

$$A_{w_1, w_p}^{(J)} \simeq \binom{N}{w_1} \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad (50)$$

where  $p = \frac{w_1}{N}$ ,  $x = \left\lfloor \frac{2P}{P+1} w_p \right\rfloor$  and  $r = \left\lfloor \frac{P}{P+1} J w_1 \right\rfloor$  is the total number of segments. By the help of the Stirling's approximation, the weight enumerating function in (50) can be approximated by

$$A_{w_1, w_p}^{(J)} \simeq \frac{N^{w_1}}{w_1!} \frac{\left( \frac{2P}{P+1} w_p \right)^{\frac{P}{P+1} J w_1}}{\left\lfloor \frac{P}{P+1} J w_1 \right\rfloor!} \left( \frac{w_1}{N} \right)^{\frac{P}{P+1} J w_1}. \quad (51)$$

Using the union bound, the upper bound on the error probability can be expressed as

$$P_{e_{NB}} < \sum_{w_1} \sum_{w_p} A_{w_1, w_p}^{(J)} Q \left( \sqrt{\frac{2E_N}{N_0}} (w_1 + w_p) \right), \quad (52)$$

where  $P_{e_{NB}}$  is the error probability considering codewords in the negative binomial region. For  $w_1, w_p \sim o(N)$ , the right hand-side of (52) converges to zero as  $N \rightarrow \infty$ . The error probability in (52) is given for a code with a uniform interleaver. For a code with a randomly chosen interleaver, with probability

one, the weight enumerating function has the same order as the weight distribution with the average

<sup>15</sup>Segments in which an RCC circulates over  $P$  nonzero states.



interleaver described in (50) and hence, the effect of the transition region on the overall performance is negligible.


#### F. Expurgating low-weight codewords

For a parallel concatenated turbo code with two constituent codes, the average number of low-weight codewords in which more than two nonzero systematic bits cause a short single error event is asymptotically zero for large block lengths [17]. The important point is that the average number of such low-weight codewords does not increase with the block length  $N$ .

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 proposed in [27], [28]. 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 change in the code rate is negligible.

One other way to avoid those low-weight structures is to build a good interleaver. Many different algorithms exist  build such interleavers. One  famous method is to employ S-random interleavers [37] which avoid low-weight codewords with systematic weight of two.

Using Table I, one can evaluate the average number of  ~~large block random interleaver generations~~ to find a good interleaver resulting in a code which is free from certain low-weight structures. For example, it takes on average about 400 trials to find a good interleaver without codewords of types (1,2) and (1,3). This number grows very fast with the number of such patterns. For example, to find a large block interleaver that guarantees the minimum distance of 30 in a code with  $P = 7$ , one should try on average, more than  $10^{39}$  different interleavers.

In Figure 14, the effect of expurgating low-weight codewords of a turbo code with two component

codes on the asymptotic mean of the error floor is shown. The figure depicts the mean of the error floor when codewords of types (1,2) and (1,3) are expurgated. Figure 15 presents the simulation results to show the effect of expurgation on a turbo code of length 10000, which shows a 2-fold reduction in the error floor after expurgating codewords of types (1,2) and (1,3).

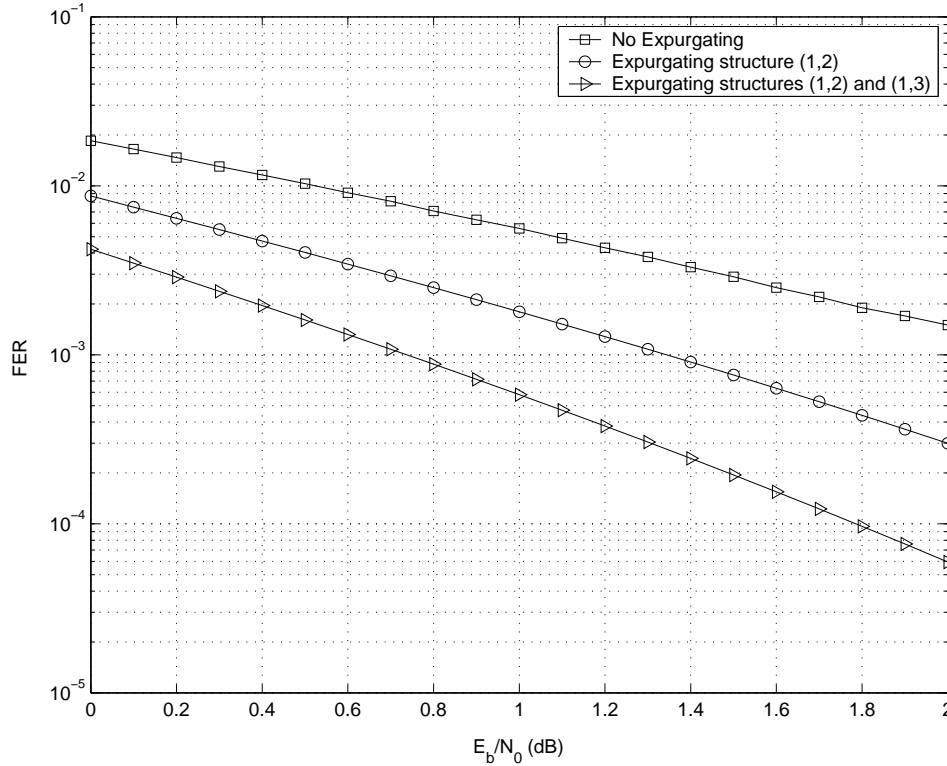


Fig. 14. Effect of expurgating codewords of type (1,2) and (1,3) on the asymptotic performance of a code of rate 1/3 and  $P = 7$ .

In multi-component turbo codes, the Poisson parameters decrease with  $N$ . However, some low-weight codewords may still exist for large (but finite) block size turbo codes. Since the number of low-weight codewords decreases as the block length increases, it is possible to expurgate all codewords of weight less than a certain threshold. This threshold can be increased unbounded as the block length increases. This is because the total number of low-weight codewords with the systematic weight  $2M$  and  $K \leq K_{\max}$  is

$$\sum_{K=JM}^{K_{\max}} A_{M,K}^{(J)} = \frac{(2M!)^{J-1}}{(M!)^J} N^{-M(J-2)} \sum_{K=JM}^{K_{\max}} \binom{K-1}{JM-1} = O(N^{-M(J-2)} K_{\max}^{MJ}). \quad (53)$$

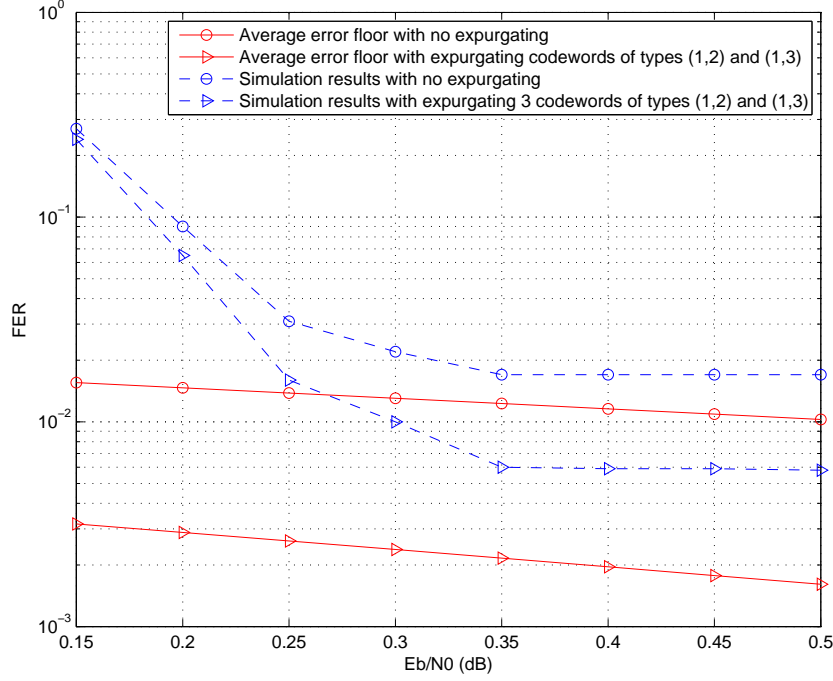


Fig. 15. Effect of expurgating codewords of type (1,2) and (1,3) on the performance of a turbo code of rate 1/3 with  $N=10000$  and  $P = 7$ .

For  $K_{\max} = o(N^{(J-2)/J})$ , the number of low-weight codewords in (53) is negligible compared to the block length and hence, expurgating those low-weight codewords does not change the code rate for  $N \rightarrow \infty$ . If all the low-weight codewords of systematic weight  $1, 2, \dots, 2M - 1$  are expurgated, then the interleaver gain increases to  $M(J - 2)$  without affecting the code rate.

#### IV. CONCLUSION

We studied the asymptotic performance of turbo codes. Our analysis is based on code weight distribution. We show that for a turbo code with a large block length, the weight spectrum has three different regions: (i) the low-weight codeword region where the number of codewords has a Poisson distribution, (ii) the high-weight region where the weight distribution is Gaussian, and (iii) the transition region where the weight distribution is negative binomial. The performance of turbo codes in the waterfall region is dominated by high-weight codewords. For almost any random interleaver and any nontrivial

recursive constituent code, the normalized weight distribution of turbo codes is asymptotically Gaussian and the code spectrum is very close to random coding weight distribution. Interleaver optimization has little effect on the asymptotic performance of the code in the waterfall region. This Gaussian distribution approaches the average spectrum defined in [13]. We evaluated the TSB bound for the Gaussian weight spectrum and determined the code rate and the SNR region where the TSB error exponent is positive and we showed that for SNR values of interest, the achievable rate is close to the capacity of BPSK signalling over AWGN channel. For higher SNR values, these achievable rates deviate from the capacity and a tighter upper or lower bound is required to determine the actual achievable rates.

In the error floor region (large SNR values), the performance of the code is dominated by low-weight codewords and for a code with two RCCs, the number of these codes remains finite as the block length increases. For large block lengths and a randomly chosen interleaver, only low-weight codewords of certain structures are probable and the average number of low-weight codewords of other structures converges to 0 as the block length increases. The number of indecomposable codewords of each structure is characterized by a set of independently distributed Poisson random variables. The frame error rate for codes with two component codes is bounded away from zero with the order of  $O(1)$  for a large block length but expurgating some low-weight codewords lowers the error floor but still bounded below by  $O(1)$ . Multi-component codes on the other hand, have a positive interleaver gain and the error floor disappears as the block length increases. The overall asymptotic error probability for these codes converges to zero either exponentially (in the Gaussian region) or polynomially (for Poisson and negative binomial regions).

## APPENDIX

### A. Proof of Lemma 1

*Proof:* For the chi-squared random variable  $Y$  with mean  $n$  and  $n$  degrees of freedom,

$$P\{Y > y\} = \frac{\Gamma\left(\frac{n}{2}, \frac{y}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}, \quad (54)$$



where  $\Gamma(\alpha, x)$  is the *incomplete gamma function* defined by

$$\Gamma(\alpha, x) = \int_x^\infty t^{\alpha-1} e^{-t} dt, \quad (55)$$

and the *gamma function*,  $\Gamma(\alpha)$ , is

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt. \quad (56)$$

For integer  $\alpha = m$ ,

$$\Gamma(m) = (m-1)!. \quad (57)$$

and

$$\Gamma(m, x) = (m-1)! e^{-x} \sum_{k=0}^{m-1} \frac{x^k}{k!} = (m-1)! e^{-x} e_m(x), \quad (58)$$

where  $e_m(x) = \sum_{k=0}^{m-1} \frac{x^k}{k!}$  is the *exponential sum function*. For  $x > m$ ,  $e_m(x)$  can be upper bounded by

$$e_m(x) = \sum_{k=0}^{m-1} \frac{x^k}{k!} < m \frac{x^{k_c}}{k_c!}, \quad (59)$$

where  $k_c$  is

$$k_c = \arg \max_{0 \leq k < m} \frac{x^k}{k!} = m-1. \quad (60)$$

and hence,

$$e_m(x) < m \frac{x^{m-1}}{(m-1)!}. \quad (61)$$

Then, by replacing  $x = y/2$  and  $m = n/2$ , (54) is upper bounded by

$$P\{Y > y\} < \frac{n}{2} \frac{e^{-y/2} (y/2)^{n/2-1}}{(n/2)!}. \quad (62)$$

For odd  $n$ , we add an independent chi-squared random variable  $Y_1$  with mean one and one degree of freedom to  $Y$  to form the random variable  $Y' = Y + Y_1$  which will be a chi-squared random variable with mean  $n+1$  and  $n+1$  degrees of freedom. Since  $Y_1 \geq 0$ , for  $y > n+1$ ,

$$P\{Y > y\} < P\{Y' = Y + Y_1 > y\} < \frac{n+1}{2} \frac{e^{-y/2} (y/2)^{(n+1)/2-1}}{((n+1)/2)!}. \quad (63)$$

For large even  $n$  and  $y/n = \beta > 1$  and by using the Stirling's approximation,

$$P\{Y > \beta n\} < \frac{1}{2} \frac{e^{-\beta n/2} (\beta n/2)^{n/2-1}}{\sqrt{\pi n} (n/2e)^{n/2}} = \frac{n}{2} \frac{e^{-\beta n/2} (\beta e)^{n/2}}{\sqrt{\pi n} \beta/2} \quad \text{for } \beta > 1. \quad (64)$$

This indicates an exponent of  $\frac{1}{2}(\beta - 1 - \log \beta)$  in the probability defined by (54). ■

### B. Proof of Proposition 1

To prove the proposition, we need the following lemma.

*Lemma 2:* Suppose we partition a stream of  $N$  bits consisting of  $w$  ones and  $N - w$  zeros into  $K$  groups. Each group consists of  $N_k$ ,  $k = 1, \dots, K$ ,  $\sum_k N_k = N$  bits. We denote by  $\mathcal{O}_k$ , the event in which the  $k^{\text{th}}$  group has an odd Hamming weight. For  $N \rightarrow \infty$ , if

$$\lim_{N \rightarrow \infty} N_k/N \neq 0, \quad k = 1, \dots, K, \quad (65)$$

then  $\mathcal{O}_1, \mathcal{O}_2, \dots, \mathcal{O}_{K-1}$  tend to be independent events of the probability  $1/2$  as  $N$  goes to infinity (for the typical values of  $w$ ).

*Proof:* The Hamming weight of the  $k^{\text{th}}$  group is shown by  $W_k$ . Then, the probability mass function of  $W_k$  can be written as

$$P_{W_k}(w_k) = \frac{\binom{N-N_k}{w-w_k} \binom{N_k}{w_k}}{\binom{N}{w}}, \quad w_k = 0, 1, \dots, N_k. \quad (66)$$

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  $k^{\text{th}}$  subsequence, and is decreasing for  $w_t < w_k < \min\{w, N_k\}$ .

An integer random variable with a monotonic probability mass function is almost equally likely to be an even or an odd number. In fact, the difference between the two probabilities is less than the boundary probabilities. For example, suppose that  $X$  is a random variable with a monotonically increasing probability mass function defined for  $2a < x < 2b$ ,  $x, a, b \in \mathbb{Z}$ . Then,

$$\begin{aligned} P\{X \text{ is even}\} &= \sum_{x=a}^b P\{X = 2x\} = \sum_{x=a}^{b-1} P\{X = 2x\} + P\{X = 2b\} \\ &\leq \sum_{x=a}^{b-1} P\{X = 2x + 1\} + P\{X = 2b\} = P\{X \text{ is odd}\} + P\{X = 2b\}. \end{aligned} \quad (67)$$

The probability mass function that is described by (66) can be separated into two monotonic functions. For  $N \rightarrow \infty$ , the boundary probabilities specified by (66) (i.e., the probabilities at  $w = 0, N, w_t$ ) are 0, and so,

$$P\{W_k \text{ is odd}\} = P\{W_k \text{ is even}\} = \frac{1}{2}. \quad (68)$$

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

We are now ready to prove Proposition 1. Assuming the systematic weight is  $w_1$ , we show that each parity stream is an  $m$ -dependent sequence, and the variance of its weight is given by (1).

*Proof:* Consider two arbitrary parity bits (far from the boundaries) named  $pb_1$  and  $pb_2$  in a given parity stream. We show that these two bits are independent of each other, when the distance between them is large. The proof can be easily 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 denote these four groups by  $C_k, k = 0, 1, 2, 3$ . The members of the  $C_0$  trigger none of the parity bits. Members of  $C_1$  and  $C_2$  trigger only the first parity bit and the second parity bit, respectively. Finally,  $C_3$  consists of bits that trigger both parity bits. Similarly, we denote by  $O_i$ , the event that  $C_i, i = 0, 1, 2, 3$  has an odd weight. Systematic bits located after both parity bit position do not affect them and hence, they belong to set  $C_0$ . 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,

$$\frac{|C_i|}{N} \neq 0, \quad i = 0, 1, 2, 3, \quad (69)$$

where  $|\cdot|$  denotes the cardinality of a set. As a result,  $C_i$ 's satisfy the conditions in Lemma 2. It is easy to see that

$$pb_1 = O_1 \oplus O_3, \quad pb_2 = O_2 \oplus O_3, \quad (70)$$

in which  $\oplus$  is the binary addition ( $pb_1$  is one if only 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,  $pb_1$  and  $pb_2$  are equiprobable independent bits.

*Case II:* The distance between the two parity 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 2.  $C_2$  has only  $k(P+1)/2$  elements since in each period  $P$ , only  $(P+1)/2$  bits trigger a certain parity bit. 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 independent and identically distributed 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 parity bit pairs 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 probability  $\frac{w_1}{N}$ . Then,

$$\text{cov}[b_2(i), b_2(i + kP)] = \frac{1}{4} \left(1 - \frac{2w_1}{N}\right)^{k(P+1)/2}, \quad (71)$$

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

$$P\{O_2 = 1\} = \frac{1 - \left(1 - \frac{2w_1}{N}\right)^{k(P+1)/2}}{2}. \quad (72)$$

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

$$\sigma_{w_2|w_1}^2 = \sum_{i=1}^N \sigma_{b_2(i)}^2 + 2 \sum_{1 \leq i < j \leq N} \text{cov}[b_2(i), b_2(j)]. \quad (73)$$

As a result,

$$\begin{aligned}
\sigma_{w_2|w_1}^2 &= \sum_{i=1}^N \frac{1}{4} + 2 \sum_{i=1}^N \sum_{k=1}^{\lfloor (N-i)/P \rfloor} \text{cov}[b_2(i), b_2(i+kP)] \\
&= \frac{N}{4} + 2 \sum_{i=1}^N \sum_{k=1}^{\lfloor (N-i)/P \rfloor} \frac{1}{4} \left(1 - \frac{2w_1}{N}\right)^{k(P+1)/2} \\
&\simeq \frac{N}{4} + 2 \sum_{i=1}^N \sum_{k=1}^{\infty} \frac{1}{4} \left(1 - \frac{2w_1}{N}\right)^{k(P+1)/2} \\
&= \frac{N}{4} \left(1 + \frac{2(1 - \frac{2w_1}{N})^{(P+1)/2}}{1 - (1 - \frac{2w_1}{N})^{(P+1)/2}}\right).
\end{aligned} \tag{74}$$

■

### C. Probability of error for large block turbo codes

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 determine 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 under this assumption and it is compared to the random coding cutoff rate.<sup>16</sup>

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 given by  $R_0 = 1 - \log_2(1 + e^{-E_N/N_0})$ , where  $E_N$  is the channel symbol energy, and  $N_0$  is the one-sided noise power spectrum [38].

For a turbo code of rate  $R$  and block length  $N$ , the normalized weight distribution function can be modeled as a Gaussian distribution with the mean  $\frac{\sqrt{N}}{2R}$  and variance  $\frac{1}{4R}$ , where the code rate  $R$  can be obtained 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  $\hat{w}$

<sup>16</sup>Results in this subsection have been presented in part in the Conference on Information Sciences and Systems (CISS'02).

and  $\hat{w} + \Delta\hat{w}$ , under the Gaussian distribution, is

$$N_{\hat{w}} \simeq \frac{2^N \Delta\hat{w}}{\sqrt{\frac{\pi}{2R}}} \exp \left[ -2R \left( \hat{w} - \frac{\sqrt{N}}{2R} \right)^2 \right]. \quad (75)$$

The term in the union bound that corresponds to the probability of an error event of the normalized weight  $\hat{w}$  (using the BPSK modulation) is

$$p_{\hat{w}} = Q \left( \sqrt{\frac{2\hat{w}\sqrt{N}E_N}{N_0}} \right). \quad (76)$$

The dominant codewords in the error probability are around the peak of  $N_{\hat{w}}p_{\hat{w}}$ , which occurs at

$$\hat{w}_p = \frac{\sqrt{N}}{2R} \left( 1 - \frac{E_N}{2N_0} \right). \quad (77)$$

The Gaussian assumption is valid when  $\lim_{N \rightarrow \infty} \frac{R\hat{w}_p}{\sqrt{N}} \neq 0, 1$ . **It is easy to see** that  $\frac{R\hat{w}_p}{\sqrt{N}} < \frac{1}{2}$ , and consequently, we only require that  $\frac{R\hat{w}_p}{\sqrt{N}} > 0$ , resulting in  $\frac{E_N}{N_0} < 2$  (equivalent to 3 dB). After the break point of  $E_N/N_0 = 3$  dB, the behavior of the turbo code performance cannot be modeled anymore by using the Gaussian weight distribution.

In practice, turbo codes are used in much lower ranges of signal to noise ratio values than the above 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 the ranges of  $\frac{E_b}{N_0}$  used in practical turbo coded 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

$$P_e < \sum_{\hat{w}=0}^{\frac{\sqrt{N}}{R}} N_{\hat{w}} p_{\hat{w}}. \quad (78)$$

By using the inequality  $Q(x) < \frac{1}{2} \exp(-\frac{x^2}{2})$  and the Gaussian distribution assumption, (78) can be rewritten as

$$P_e < \frac{2^N}{\sqrt{\frac{2\pi}{R}}} A \int_0^{\frac{\sqrt{N}}{R}} \exp \left( -2R \left[ \hat{w} - \frac{\sqrt{N}}{2R} \left( 1 - \frac{E_N}{2N_0} \right) \right]^2 \right) d\hat{w}, \quad (79)$$

where

$$A = \exp \left( -\frac{N}{2R} \left[ 1 - \left( 1 - \frac{E_N}{2N_0} \right)^2 \right] \right), \quad (80)$$

and hence,

$$P_e < 2^{N-1} AB, \quad (81)$$

where

$$B = Q \left[ \sqrt{\frac{N}{R}} \left( \frac{E_N}{2N_0} - 1 \right) \right] - Q \left[ \sqrt{\frac{N}{R}} \left( \frac{E_N}{2N_0} + 1 \right) \right]. \quad (82)$$

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

$$\lim_{N \rightarrow \infty} Q \left[ \sqrt{\frac{N}{R}} \left( \frac{E_N}{2N_0} - 1 \right) \right] = 1, \quad (83)$$

and,

$$\lim_{N \rightarrow \infty} Q \left[ \sqrt{\frac{N}{R}} \left( \frac{E_N}{2N_0} + 1 \right) \right] = 0. \quad (84)$$

Hence,  $B$  can be approximated as 1.

Let us define

$$R_T = \frac{1}{2 \ln(2)} \left[ \frac{E_N}{N_0} - \frac{1}{4} \left( \frac{E_N}{N_0} \right)^2 \right]. \quad (85)$$

We can see that if  $R < R_T$ , then the error probability approaches to 0 as  $N \rightarrow \infty$ . Figure 16 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).

## Acknowledgements

Authors wish to thank Professor Rüdiger L. Urbanke for providing insightful ideas and suggestions and for his helpful comments to improve the quality of this paper.

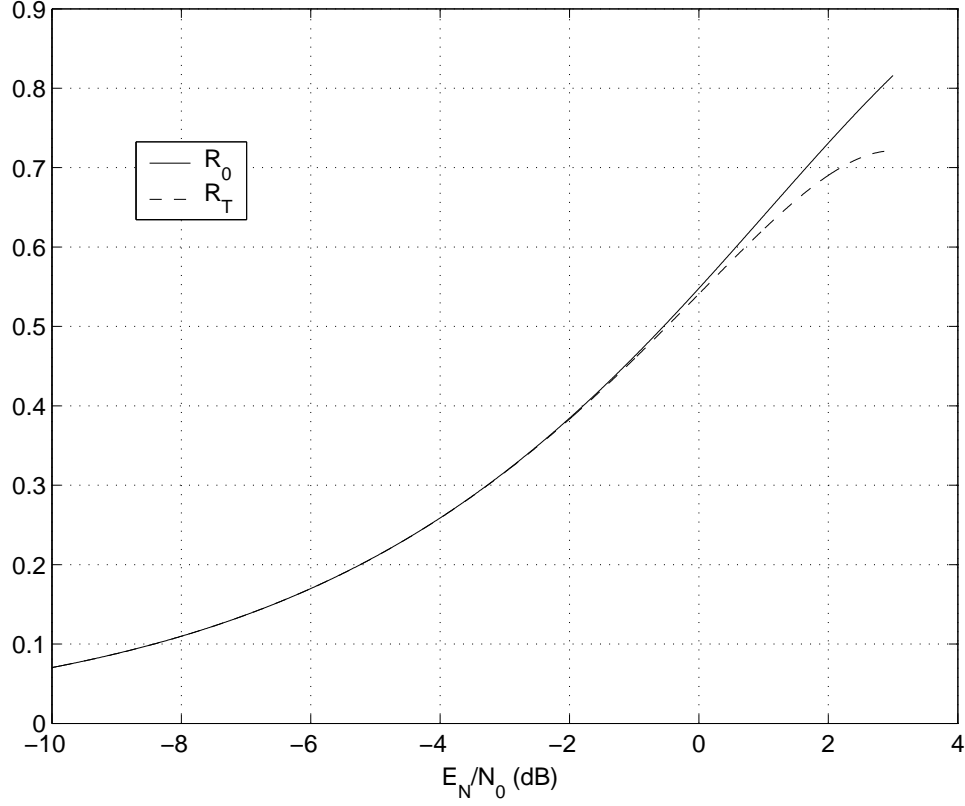


Fig. 16. Comparison between  $R_0$  and  $R_T$  versus  $\frac{E_N}{N_0}$ .

## REFERENCES

- [1] A. K. Khandani, "Optimization of the interleaver structure for turbo-codes," in *Canadian Workshop on Information Theory*, Kingston, Canada, June 1999, pp. 25–28.
- [2] M. H. Baligh and A. K. Khandani, "Asymptotic effect of interleaver structure on the performance of turbo-codes," in *Conference on Information Sciences and Systems*, Princeton, NJ, USA, March 2002, pp. 766–769.
- [3] M. H. Baligh and A. K. Khandani, "Asymptotic effect of interleaver structure on the performance of turbo codes," Tech. Rep. 2004-03, University of Waterloo, March 2004.
- [4] 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.
- [5] 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.
- [6] 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.



- [7] 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.
- [8] H. Jin and R. J. McEliece, "Coding theorems for turbo code ensembles," *IEEE Trans. on Inform. Theory*, vol. 48, pp. 1451–1461, June 2002.
- [9] G. Battail, "A conceptual framework for understanding turbo codes," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 245–254, Feb. 1998.
- [10] G. Battail, C. Berrou, and A. Glavieux, "Pseudo-random recursive convolutional coding for near-capacity performance," in *IEEE Globecom Conference*, Houston, USA, Nov. 1993, pp. 23–27.
- [11] N. Shulman, "Random coding techniques for nonrandom codes," *IEEE Trans. on Inform. Theory*, vol. 45, pp. 2101–2104, Sep. 1999.
- [12] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. on Inform. Theory*, vol. 11, pp. 3–18, Jan. 1965.
- [13] G. Poltyrev, "Bounds on the decoding error probability of binary linear codes via their spectra," *IEEE Trans. on Inform. Theory*, vol. 40, pp. 1284–1292, July 1994.
- [14] T. M. Duman and M. Salehi, "New performance bounds for turbo codes," *IEEE Trans. on Communications*, vol. 46, pp. 717–723, June 1998.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] M. Breiling, "A logarithmic upper bound on the minimum distance of turbo codes," *IEEE Trans. on Inform. Theory*, vol. 50, pp. 1692–1710, Aug. 2004.
- [21] A. H. S. Mohammadi and W. Zhuang, "Variance of the turbo code performance bound over the interleavers," *IEEE Trans. on Inform. Theory*, vol. 48, pp. 2078–2086, July 2002.
- [22] G. Battail, "Construction explicite de bons codes longs," *Ann. Télécommun.*, vol. 44, pp. 392–404, July-Aug. 1989.
- [23] E. Biglieri and V. Volski, "Approximately Gaussian weight distribution of the iterated product of single-parity-check codes," *IEEE Electronics Letters*, vol. 30, pp. 923–924, June 1994.

- [24] D. Yue and E. Yang, "Asymptotically Gaussian weight distribution and performance of multicomponent turbo block codes and product codes," *IEEE Transactions on Communications*, vol. 52, pp. 728–736, May 2004.
- [25] T. Richardson and R. Urbanke, "On the distribution of low-weight codewords for turbo codes," in *42<sup>nd</sup> Annual Allerton Conference on Communication, Control, and Computing*, Allerton, IL, USA, Sep. 2004.
- [26] 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.
- [27] 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.
- [28] M. Öberg and P. H. Siegel, "Application of distance spectrum analysis to turbo code performance improvement," in *35<sup>th</sup> Annual Allerton Conference on Communication, Control, and Computing*, Allerton, IL, USA, Sep. 1997, pp. 701–710.
- [29] S. W. Golomb, *Shift Register Sequences*, San Francisco, Holden-Day, 1967.
- [30] J. K. Patel and C. B. Read, *Handbook of the Normal Distribution*, Dekker Inc., second edition, 1996.
- [31] S. ten Brink, "Convergence of iterative decoding," *IEEE Electronic Letters*, vol. 35, pp. 806–808, May 1999.
- [32] H. El-Gamal and A. R. Hammons Jr, "Analyzing the turbo decoder using the Gaussian approximation," *IEEE Trans. on Inform. Theory*, vol. 47, pp. 671–686, February 2001.
- [33] D. Divsalar, S. Dolinar, and F. Pollara, "Iterative turbo decoder analysis based on density evolution," *IEEE Journal on Selected Areas in Communications*, vol. 19, pp. 891–907, May 2001.
- [34] G. Poltyrev, "On coding without restrictions for the AWGN channel," *IEEE Trans. on Inform. Theory*, vol. 40, pp. 409–417, March 1994.
- [35] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, John Wiley, New York, 1965.
- [36] T. Richardson and R. Urbanke, *Modern Control Theory*, To be published.
- [37] S. Dolinar and D. Divsalar, "Weight distributions for turbo codes using random and nonrandom permutations," *JPL TDA Progress Report*, vol. 42-122, pp. 56–65, Aug. 1995.
- [38] John G. Proakis, *Digital Communication*, McGraw-Hill, fourth edition, 2001.