Asymptotic Effect of Interleaver Structure on the Performance of Turbo-codes

M. H. Baligh & A. K. Khandani
Coding & Signal Transmission Lab.
Dept. of Elec. and Comp. Eng.,
University of Waterloo,
Waterloo, ON, N2L 3G1
{hadi,khandani}@cst.uwaterloo.ca

Abstract — **This work studies the effect of the interleaver optimization on the performance of Turbo-codes for large block lengths, $N \to \infty$. For $N \to \infty$, the weight of the systematic and parity check sequences, denoted by w_1, w_2 and w_3, respectively, tend to a jointly Gaussian distribution for typical values of $w_i, i = 1, 2, 3$ (typical values of w_i are defined as $\lim_{N\to\infty} \frac{w_i}{N} \neq 0, 1$ for $i = 1, 2, 3$). To optimize the code performance, it is desirable that the corresponding correlation coefficients, denoted as ρ_{ij}, $i, j = 1, 2, 3$, to be as small as possible. It is however shown that: (i) $\rho_{ij} > 0$, $i, j = 1, 2, 3$, (ii) $\rho_{12}, \rho_{13} \to 0$ as $N \to \infty$, and (iii) $\rho_{23} \to 0$ as $N \to \infty$ for almost any random interleaver. This means that for $N \to \infty$, the optimization of the interleaver has a diminishing effect on the distribution of large weight error events. We discuss methods to expurgate the low weight code-words (lower the error floor) without affecting the code rate. The resulting expurgated code has an "average spectrum" [1] and consequently meets the best known random coding error exponent (i.e., achieves the channel capacity). We also present a condition on the channel Signal-to-Noise-Ratio (SNR) such that the dominant error events satisfy the Gaussian assumption and show that this condition is satisfied in cases of practical interest. This means that the asymptotic performance of the expurgated code is not affected by the choice of the interleaver for values of channel SNR of practical interest.**

I. Introduction

The advent of Turbo-codes [2] is perhaps the most important development in coding theory in many years. These codes can achieve near Shannon-limit error correcting performance with a relatively simple decoding method. The basic idea of Turbo-codes is to make use of some Recursive Convolutional Codes (RCC) which are connected in parallel through pseudo-random interleavers. Note that as the RCCs and also the interleaver have linearity property[1], the resulting code is linear[2], and consequently, group property and distance invariance property hold.

Figure 1 shows the block diagram of the encoder of a rate 1/3 Turbo-code composed of two RCCs, where $b_1(m)$ is the systematic bit, and $b_2(m)$, $b_3(m)$ are the parity check bits. The weight of the code in Fig. 1 is equal to the sum of the weights of the $b_1(m)$, $b_2(m)$ and $b_3(m)$ sequences over a block, which are denoted by w_1, w_2, and w_3, respectively.

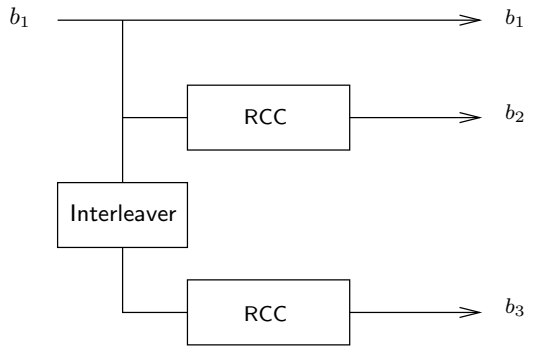

Figure 1: Basic structure of the Turbo encoder.

We present results for the performance of Turbo-codes for large block lengths, $N \to \infty$. For $N \to \infty$, the weight of the systematic and parity check sequences, denoted by w_1, w_2 and w_3, respectively, tend to a jointly Gaussian distribution for typical values of $w_i, i = 1, 2, 3$ (i.e., $\lim_{N\to\infty} \frac{w_i}{N} \neq 0, 1$ for $i = 1, 2, 3$). To optimize the code performance, it is desirable that the corresponding correlation coefficients, denoted as ρ_{ij}, $i, j = 1, 2, 3$, to be as small as possible. It is however shown that: (i) $\rho_{ij} > 0$ for $i, j = 1, 2, 3$, (ii) $\rho_{12}, \rho_{13} \to 0$ as $N \to \infty$, and (iii) $\rho_{23} \to 0$ for $N \to \infty$ with probability one (for almost any random interleaver). This means that for $N \to \infty$ and $\lim_{N\to\infty} \frac{w_i}{N} \neq 0, 1$ for $i = 1, 2, 3$, the weights w_1, w_2 and w_3 tend to become independent and consequently the optimization of the interleaver has a diminishing effect on the code performance.

For the low weight code-words, the assumption of Gaussian distribution is no longer valid. It is well known that some low weight code-words may exist which produce low weight parity strings in both the RCCs. Such low weight code-words contribute to an error floor on the code performance. The number of such low weight code-words and their weights depend on the code polynomial and the interleaver structure and hence the interleaver optimization plays an important role in this region. In [3], it is shown that the average number of these low weight code-words tends to zero when $N \to \infty$, except for some special code-words in which the systematic stream consists of certain pairs of 1's.

We discuss methods to expurgate the low weight code-words (lower the error floor) without affecting the code rate.

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

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

The key point is that the average number (averaged over all possible interleavers) of such error events does not increase with the block length [3], and consequently, the number of such error events remain finite with high probability (for most interleavers). The resulting expurgated code has an "average spectrum"[1] and consequently achieves the best known random coding error exponent (i.e., achieves the channel capacity). We also present a condition on the channel SNR such that the dominant error events satisfy the Gaussian assumption and show that this condition is satisfied in cases of practical interest. This means that the asymptotic performance of the expurgated code is not affected by the choice of the interleaver for values of code rate and channel SNR of practical interest.

II. Interleaver optimization for $N \to \infty$

We assume that the RCCs are generated by the transfer function $G(d) = N(d)/D(d)$. Using the results of [4], we know that the impulse response of $G(d)$ is periodic with period $p \leq 2^r - 1$ where r is the memory length of the code. We are mainly interested in the group structure 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 $N(d)$.

In general, we would like the period of the impulse response of $G(d)$ to be as large as possible. If the period is equal to $2^r - 1$, the resulting impulse response is called a maximum length sequence (MLS). Inthe rest we assume that all 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 given r) are given in [4]. It can be shown that any MLS-sequence satisfy the three randomness postulates [4]. One consequence of this property is that in any period of an MLS-sequence, the number of ones is equal to 2^{r-1} and the number of zeros is $2^{r-1} - 1$.

If we look at the impulse response of $D(d)$ as a periodic sequence (started infinite in the past), we obtain $K = 2^r - 1$ non-zero sequences which are time shifts of each other. Each sequence corresponds to a specific positioning of the impulse within the period. We refer to these sequences as different phases of the periodic signal. We assume that different phases are labeled by integer numbers, say $1, \ldots, K$, where the label of a phase corresponds to the relative position of the corresponding impulse within the period. It can be shown that the set of phases of an MLS-sequence (plus the all-zero sequence) constitute a group under binary addition [4]. The order of each element in this group is equal to two, meaning that the sum of each phase with itself results in the all-zero sequence (denoted as the zero phase).

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

For bit position $k = 1, \ldots, N$ within the jth output stream, we refer to the set of bit positions $i \leq k$ for which an impulse at position i results in a 1 at position k as $\mathcal{R}_j(k)$, $j = 1, 2, 3$. Obviously, $\mathcal{R}_1(k) = \{k\}$. If the bit position k is located in the Mth period, i.e., $M = \lceil k/p \rceil$, then the number of positions within each of the periods $1, \ldots, M-1$ which belong to $\mathcal{R}_j(k)$, $j = 2, 3$, is equal to 2^{r-1}. The number of positions within the Mth period (the period containing k itself) depends on the relative position of k within the Mth period and also on the numerator of $G(d)$. We are mainly interested in large values of M for which the effect of the elements within the Mth period itself is negligible. This means $|\mathcal{R}_2(k)|, |\mathcal{R}_3(k)| \simeq \lceil k/p \rceil 2^{r-1}$ where $|\,.\,|$ denotes the cardinality of the corresponding set.

We use the notation $b_i(m)$, $i = 1, 2, 3$, $m = 1, \ldots, N$, to refer to the mth bit within the ith output stream. We have the following statistical expectations: $\overline{b_i(m)} = \overline{b_i^2(m)} = 1/2$.

In the following, we first show that w_1, w_2 and w_3 have a Gaussian distribution for high input weight codewords. This is easily verified noting that all the 2^N possible combinations within the three steams are allowed, and consequently, the positions within each of the three output streams are iid binary random variables (where 0 and 1 are equally probable). Using the Central Limit Theorem, we conclude that w_1, w_2 and w_3 which are the sum of N iid random variables have a Gaussian distribution with mean $N/2$ and Variance $N/4$. On the other hand, when concerning with high weight code-words, the conditional weight distributions are also Gaussian and hence w_1, w_2 and w_3 are jointly Gaussian. To show this, let us consider either of the parity sequences when w_1 is known and is in the range of its typical values, i.e., $\lim_{N\to\infty} \frac{w_1}{N} \neq 0, 1$. Under these circumstances, it is easy to show that the sequence of bits in either of the two parities will be an m-dependent sequence [5], and consequently, the corresponding weight (as a random variable conditioned on the systematic weight) will have a Gaussian distribution.

In this case, to show that w_1, w_2 and w_3 are independent, we need to show that the corresponding correlation coefficients ρ_{ij}, $i, j = 1, 2, 3$, tend to zero as $N \to \infty$, where $\rho_{ij} = (4/N)(\overline{w_i w_j} - \overline{w_i}\ \overline{w_j}) = (4/N)[\overline{w_i w_j} - (N/2)^2]$. We have $\overline{w_i w_j} = \sum_m \sum_n \overline{b_i(m) b_j(n)}$ where the expectation is over all possible 2^N combinations of the input. The overall weight of the output sequence is equal to $w = w_1 + w_2 + w_3$ which has a Gaussian distribution of mean $3N/2$ and variance $N(3 + 2\rho_{12} + 2\rho_{13} + 2\rho_{23})/4$. Noting that sequences with a weight smaller than the mean value result in larger probability of error as compared to sequences with a weight larger than the mean, we conclude that the main objective in the code design is to sharpen the peak of the PDF of w which is equivalent to minimizing its variance. This is in turn equivalent to minimizing the ρ_{ij} coefficients. In the following, we first show that the $\rho_{ij} > 0$, so the minimum possible value for each of them is zero. Noting the properties of the Gaussian distribution, this is equivalent to having independence between the coded streams.

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

Any of the pairs $b_i(m), b_j(n)$, $i, j = 1, 2, 3$, $m, n = 1, \ldots, N$, can take four different values, namely $\{00, 01, 10, 11\}$. The set of the input sequences resulting in the value of 00 form a subgroup of the group of all possible 2^N input combinations. This is a direct consequence of the linearity, and consequently, the group property of the code. Noting the group property of the set of corresponding coset leaders, two cases may happen. There is either only one coset with the coset leader 11, or there are three cosets with the cosets leaders 01, 10 and 11. The important point is that in both of these cases, using the basic results of the group theory, we conclude that the 00 sub-group and its cosets contain the same number of input sequences. This means that for the probability of the pair $b_i(m), b_j(n)$,

we have the following two cases:

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

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

The important point is that in both cases, we have $\overline{b_i(m)b_j(n)} - \overline{b_i(m)}\ \overline{b_j(n)} > 0$. This means that the correlation coefficients ρ_{ij}, $i, j = 1, 2, 3$ are always positive.

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

For ρ_{12} and ρ_{13} (interaction of the systematic stream with each of the parity checks), Case II of the above two cases is valid resulting in $\rho_{12}, \rho_{13} \to 0$ for $N \to \infty$. Note that $b_1(m)$ and $b_2(n)$ will be independent of each other if $b_1(m)$ is not mapped (through interleaving) to a bit position within $\mathcal{R}_2(n)$, or otherwise, if $\mathcal{R}_2(n)$ contains at least two elements. This will be always valid unless for some trivial cases which will have a vanishing effect on the overall result.

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

If $\mathcal{R}_2(m)$, $\mathcal{R}_3(n)$ contain at least one bit which are different from each other then $b_2(m)$ and $b_3(n)$ will be 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 will be the case unless $m = n$ and the elements of $\mathcal{R}_2(m)$ and $\mathcal{R}_3(m)$ contain the same input bits (before and after interleaving). This means that the corresponding interleaver has restriction on the mapping of infinitely many bit positions. Obviously, the faction of such interleavers tend to zero as $N \to \infty$. This means that for almost any random interleaver, we have $\rho_{23} \to 0$ as $N \to \infty$.

III. Low weight codewords

Low weight code-words in Turbo-codes occur when a low weight input results in a small weight for the parity sequences. It means that every time each RCC leaves the zero state by a 1 in the systematic input stream it returns to zero-state after a small duration. In [3], it is shown that for $N \to \infty$, the average number of low weight code-words decreases to zero except for the code-words which are composed of several non-overlapping short error events[3] caused by two information bits separated by an integer multiple of the impulse response in each RCC. In other words, the average number of low weight code-words in which more than two ones cause a short error event is zero for large block lengths. The key point is that the average number of such low weight code-words does not increase with N [3]. The number of low weight code-words is a nonnegative integer with a finite average, and consequently, the probability of having an infinite number of such low weight code-words tends to zero for large block lengths.

We can remove the effect of these low weight code-words on the error floor region by expurgating them. Expurgating low weight code-words decreases the dependency of Turbo-code performance on the RCCs and interleaver structure and the remaining code-words satisfy the Gaussian assumption.

To expurgate these code-words one way is to set one information bit in each low weight code-word to zero. If the block length is sufficiently large, the number of these bits will be small in comparison with the block length, and consequently, the code rate will not be affected.

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

Another method to increase the minimum distance is to create large block length interleavers from shorter ones. For example, to create an interleaver of length $2N$ from a known interleaver of length N, we can concatenate two such interleavers and switch one information bit in each low weight code-word between the two interleavers. The number of bits to be exchanged remains finite with a high probability (for most interleavers). The resulting Turbo-code of block length $2N$ has a minimum distance twice the original Turbo-code of block length N. By repeating this procedure, we can increase the minimum distance and remove the error floor.

IV. Probability of error for large block Turbo-codes

In this section, we compute the condition on the channel SNR such that the Gaussian assumption is valid. We also compute the union bound on the Frame Error Probability (FER) using the Gaussian distribution and find the cutoff rate based on this assumption. Comparing the result with the true cutoff rate, namely $R_0 = 1 - \log_2(1 + e^{-E_N/N_0})$ where E_N is the channel symbol energy and N_0 is one-sided Gaussian power spectrum of noise [6], gives an indication of the region in which the Gaussian distribution approximation remains valid.

For an expurgated Turbo-code with code rate R and block length N the weight distribution function can be modeled as a Gaussian distribution with mean $\frac{N}{2R}$ and variance $\frac{N}{4R}$ where the code rate R is acheived by using a larger number of parallel concatenated RCCs and/or puncturing which does not affect the Gaussian assumption. The number of code-words of weight w is

$$N_w \simeq \frac{2^N}{\sqrt{\frac{\pi N}{2R}}} \exp\left[-\frac{(w - \frac{N}{2R})^2}{\frac{N}{2R}}\right] \tag{1}$$

The term in the union bound corresponding to the probability of an error event of weight w (using BPSK modulation) is

$$p_w = Q\left(\sqrt{\frac{2wE_N}{N_0}}\right) \tag{2}$$

Dominant code-words in the error probability occur around the peak of $N_w p_w$, which is $w_p = \frac{N}{2R}(1 - \frac{E_N}{2N_0})$. In order for the Gaussian assumption to be valid, we require that $\lim_{N\to\infty} \frac{Rw_p}{N} \neq 0, 1$. It is easy to see that $\frac{Rw_p}{N} < 1$, and consequently, we only require that $\frac{Rw_p}{N} > 0$, resulting in $\frac{E_N}{N_0} < 2$ (equivalent to 3 dB). After reaching this break point of $E_N/N_0 = 3$ dB, the behavior of Turbo-code cannot be modeled any more using Gaussian distribution.

In practice, Turbo-codes are used in much lower range of SNR. 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 rate $1/3$, or to $\frac{E_b}{N_0} = 6$ dB for a code of rate $1/2$. These values are substantially higher than the ranges of E_b/N_0 used in practical systems. In other words, for SNRs of interest the dominant code-words of expurgated Turbo-code follow the Gaussian assumption.

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

$$P_e < \int_0^{\frac{N}{R}} N_w p_w dw \tag{3}$$

By using the inequality $Q(x) < \frac{1}{2}\exp(-\frac{x^2}{2})$, we can reduce (3) to

$$P_e < \frac{2^{N-1}}{\sqrt{\frac{\pi N}{2R}}} A \int_0^{\frac{N}{R}} \exp\left(-\frac{\left[w - \frac{N}{2R}\left(1 - \frac{E_N}{2N_0}\right)\right]^2}{\frac{N}{2R}} \right) dw \quad (4)$$

where

$$A = \exp\left(-\frac{N}{2R}\left[1 - \left(1 - \frac{E_N}{2N_0} \right)^2 \right] \right) \quad (5)$$

and hence,

$$P_e < 2^{N-1} AB \quad (6)$$

where,

$$B = Q\left[\sqrt{\frac{N}{R}} \left(\frac{E_N}{2N_0} - 1 \right) \right] - Q\left[\sqrt{\frac{N}{R}} \left(\frac{E_N}{2N_0} + 1 \right) \right] \quad (7)$$

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

$$\lim_{N\to\infty} Q\left[\sqrt{\frac{N}{R}} \left(\frac{E_N}{2N_0} - 1 \right) \right] = 1 \quad (8)$$

and,

$$\lim_{N\to\infty} Q\left[\sqrt{\frac{N}{R}} \left(\frac{E_N}{2N_0} + 1 \right) \right] = 0 \quad (9)$$

and hence B can be approximated as 1. Let us define

$$R_T = \frac{1}{2\ln(2)} \left[\frac{E_N}{N_0} - \frac{1}{4}\left(\frac{E_N}{N_0} \right)^2 \right]. \quad (10)$$

We can see that if $R < R_T$, then the probability of error converges to 0 as $N \to \infty$. Figure 2 shows the difference between R_0 and R_T around the break point of $E_N/N_0 = 3$ dB.

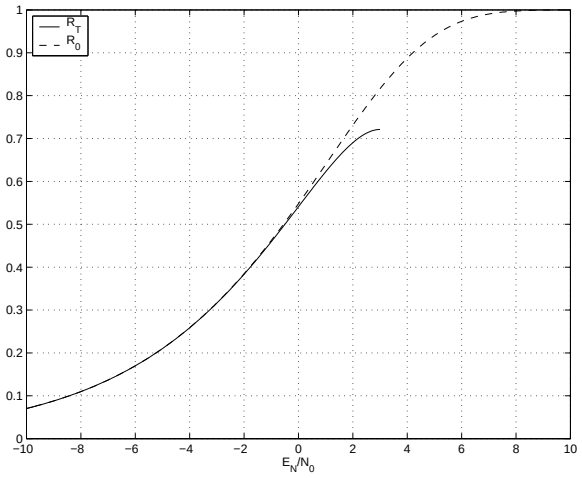

Figure 2: Comparison between R_0 and R_T versus $\frac{E_N}{N_0}$.

References

[1] 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.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-codes," *IEEE Int. Conf. on Comm. 1993 (ICC'93)* (Geneva, Switzerland), pp. 1064—1070, May 1993.

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

[4] S. W. Golomb, "Shift register sequences," San Francisco, Holden-Day, 1967.

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

[6] John G. Proakis, *Digital Communications*, Mc Graw-Hill, Fourth Edition, 2001.