Transmission of Non-Uniform Memoryless Sources via Non-Systematic Turbo Codes

Guang-Chong Zhu, Fady Alajaji, Jan Bajcsy and *Patrick Mitran*

Abstract

We investigate the joint source-channel coding problem of transmitting non-uniform memoryless sources over BPSK-modulated additive white Gaussian noise (AWGN) and Rayleigh fading channels via Turbo codes. In contrast to previous work, recursive non-systematic convolutional encoders are proposed as the constituent encoders for heavily biased sources. We prove that under certain conditions and when the length of the input source sequence tends to infinity, the encoder state distribution and the marginal output distribution of each constituent recursive convolutional encoder become asymptotically uniform regardless of the degree of source non-uniformity. We also give a conjecture (which is empirically validated) on the condition for the higher order distribution of the encoder output to be asymptotically uniform irrespective of the source distribution. Consequently, these conditions serve as design criteria for the choice of good encoder structures. As a result, the outputs of our selected non-systematic Turbo codes are suitably matched to the channel input since a uniformly distributed input maximizes the channel mutual information and hence achieves capacity. Simulation results show substantial gains by the non-systematic codes over previously designed systematic Turbo codes; furthermore, their performance is within 0.74 to 1.17 dB from the Shannon limit. Finally, we compare our joint source-channel coding system with two tandem schemes which employ a 4^{th}-order Huffman code (performing near optimal data compression) and a Turbo code that either gives excellent waterfall bit error rate (BER) performance or good error-floor performance. At the same overall transmission rate, our system offers robust and superior performance at low BERs ($\leq 10^{-4}$), while its complexity is lower.

Keywords: Non-systematic Turbo codes, non-uniform i.i.d. sources, joint source-channel coding, AWGN and Rayleigh fading channels, Shannon limit.

IEEE Transactions on Communications: submitted November 2002, revised September 2003 and November 2003. This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Premier's Research Excellence Award (PREA) of Ontario. Parts of this work were presented at the Conference on Information Sciences and Systems, Princeton, USA, March 2002, and the Second IASTED International Conference on Wireless and Optical Communications, Banff, Canada, July 2002.

G.-C. Zhu was with Department of Mathematics & Statistics, Queen's University, Kingston, ON K7L 3N6, Canada; he is now with the Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, ON M5S 3G4, Canada. F. Alajaji is with the Department of Mathematics & Statistics, Queen's University, Kingston, ON K7L 3N6, Canada (E-mail: fady@mast.queensu.ca). J. Bajcsy is with the Department of Electrical and Computer Engineering, McGill University, Montreal, Quebec H3A 2A7, Canada. P. Mitran is with the Division of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA.

1 Introduction

In most of the theory and practice of error-control coding, the input to the channel encoder is assumed to be uniform, independent and identically distributed (i.i.d.); i.e., the source generates a memoryless binary stream $\{U_k\}_{k=0}^{\infty}$, where $Pr\{U_k = 0\} = 1/2$. In reality, however, a substantial amount of redundancy resides in natural sources. For example, many uncompressed binary images (e.g., facsimile and medical images) may contain as much as 80% of redundancy in the form of non-uniformity (e.g., [1, 2]); this corresponds to a probability $p_0 \stackrel{\triangle}{=} Pr\{U_k = 0\} = 0.97$. In this case, a source encoder would be used. Such an encoder is *optimal* if it can eliminate all the source redundancy and generates uniform i.i.d. outputs. However, most existing source encoders are sub-optimal, particularly fixed-length encoders that are commonly used for transmission over noisy channels; thus, the source encoder output contains some *residual* redundancy. For example, the 4.8 kbits/s US FS 1016 CELP speech vocoder produces line spectral parameters that contain 41.5% of residual redundancy due to non-uniformity and memory [3]. Variable-length or entropy codes (e.g., Huffman codes) which can be asymptotically optimal (for sufficiently large blocks of source symbols) could be employed instead of fixed length codes or in conjunction with them; however, error-propagation problems in the presence of channel noise are inevitable and are sometimes catastrophic. Therefore, the reliable communication of sources with a considerable amount of redundancy (residual if compressed, or natural if not) is an important issue. This in essence is a joint source-channel coding problem. Blizard [4], Koshelev [5] and Hellman [6] are among the first few who proposed convolutional coding for the joint source-channel coding of sources with natural redundancy, where the source statistics are utilized at the receiver. Specifically, the convolutional encoding of such sources (Markov and non-uniform sources) over memoryless channels and their decoding via sequential decoders employing a decoding metric that is dependent on both source and channel distributions, was studied in [4, 5]. The computational complexity of such sequential decoders is analyzed, and it is shown that for a range of transmission rates, the expected number of computations per decoded bit is finite.

In [6], a lossless joint source-channel coding theorem for convolutional codes is established. It is proved that for a discrete memoryless source and channel pair, there exist convolutional codes of rate R that can be used to perform reliable joint source and channel coding (under maximum a posteriori decoding) as long as $R < C/H$, where H is the source entropy and C is the channel capacity. Recently, several studies (e.g., [1]–[3], [7]–[22], etc.) have also shown that appropriate use of the source redundancy can significantly improve the system performance.

Turbo codes have been regarded as one of the most exciting breakthroughs in channel coding. The original work by Berrou *et al.* demonstrated excellent performance of Turbo codes for uniform i.i.d. sources over AWGN channels [23]. The work was later extended to Rayleigh fading channels showing comparable performance [24]. Recently, McEliece pointed out that Turbo codes have also great potential on nonstandard channels (asymmetric, non-binary, multiuser, etc.) [25]. However, the above works focus on uniform i.i.d. sources. In [26, 27], the authors considered using Turbo codes for sources with memory. To the best of our knowledge, the issue of designing Turbo codes for non-uniform i.i.d. sources has not been fully studied, except for the recent work in [28, 29], in which standard systematic Turbo codes are considered, where each constituent encoder is a recursive systematic convolutional (RSC) encoder. Although the gains achieved by these codes are considerable with respect to the original Berrou code, their performance gaps vis-a-vis the Shannon limit – also known as the optimal performance theoretically achievable (OPTA) – are still relatively big for heavily biased sources (e.g., with p_0=0.8, 0.9). Analysis on the encoder output reveals that the drawback lies in the *systematic* structure, which results in a mismatch[1] between the biased distribution of the systematic bit stream and the uniform input distribution needed to achieve

[1] A similar mismatch in the context of the design of scalar quantizers for non-uniform memoryless sources over binary symmetric channels (BSC) was also observed in [15, 17] and addressed via the use of a rate-one convolutional encoder. This method is however unsuitable for our problem as it will result in error-propagation at the receiver due to our use of large data blocks in an attempt to achieve the Shannon limit.

channel capacity. As we will show in this paper, when some constraints are satisfied, recursive non-systematic convolutional (RNSC) encoders can generate asymptotically uniform outputs even for extremely biased sources. But it is known that the capacity of a binary input AWGN or Rayleigh channel is achieved by a uniform i.i.d. channel input. Furthermore, it was shown in [30] by Shamai and Verdú that the empirical distribution of any good code – a code with rate close to capacity and vanishing probability of error for sufficiently long blocklengths – should approach the capacity achieving input distribution[2]. Therefore, we propose using RNSC Turbo encoders. Simulation results demonstrate substantial gains over systematic Turbo codes. The OPTA gaps for heavily biased sources are hence significantly reduced.

This paper is organized as follows. In Section 2, we illustrate the need to examine non-systematic Turbo codes instead of systematic codes for the transmission of strongly non-uniform sources by evaluating the capacity loss incurred when such sources are directly (e.g., via a systematic bit stream) sent over BPSK-modulated AWGN or Rayleigh fading channels. In Section 3, we give two asymptotic properties (when the input sequence length tends to infinity) of recursive convolutional encoders whose input is a non-uniform i.i.d. source with arbitrary degree of non-uniformity. Design criteria for good non-systematic encoder structures based on these two properties are introduced in Section 4. The iterative decoding design for such non-systematic Turbo codes is also addressed. Simulation results and performance comparisons to the Shannon limit are presented in Section 5. In Section 6, we compare our joint source-channel coding system with two tandem coding schemes that employ a nearly optimal source code followed by Berrou's (37,21) Turbo code [23] (which gives excellent waterfall BER performance) or a systematic (35,23) Turbo code (which gives a slightly inferior waterfall BER performance but a lower error-floor performance). Finally, conclusions are given in Section 7.

[2]More precisely, it is shown in [30, Theorem 4] that for any discrete memoryless channel, and for any fixed integer $k > 0$, the kth-order empirical distribution of any regular good code sequence converges (in the Kullback Leibler divergence sense) to the k-product of the capacity-achieving distribution.

2 Capacity Loss Due to Channel Input Mismatch

In this section, we illustrate why systematic codes are not well matched to symmetric channels (due to their systematic bit stream) when the source is very biased (p_0=0.8, 0.9). This is achieved by examining the capacity loss incurred in BPSK-modulated AWGN channels and memoryless Rayleigh fading channels with known channel state information [24] under such biased sources as input. The channel capacity is the largest rate at which information can be transmitted (via a block code) and recovered with a vanishingly low probability of error. For discrete-input memoryless channels, it is well known that the capacity is given by the maximum of the mutual information between the channel input and output: $C = \max_{p(x)} I(X;Y)$, where the maximization is taken with respect to all input distributions $p(x)$. When the channel is symmetric, the capacity is achieved by a uniform channel input distribution. For AWGN channels, the capacity is

$$C_{AWGN} = 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y+1)^2}{2\sigma^2}} \log_2(1 + e^{\frac{2y}{\sigma^2}}) dy.$$

Similarly, for Rayleigh fading channels, the capacity is

$$C_{Ray} = 1 - \int_{0}^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y+a)^2}{2\sigma^2}} \cdot 2ae^{-a^2} \cdot \log_2(1 + e^{\frac{2ay}{\sigma^2}}) dy\ da.$$

Note that the average energy per channel symbol $E_s = 1$, and $\sigma^2 = N_0/2$ is the variance of the additive white Gaussian noise. Therefore, both C_{AWGN} and C_{Ray} are functions of the channel signal-to-noise ratio (CSNR) E_s/σ^2.

When a non-uniform i.i.d. source (with distribution $p_0 \neq 1/2$) is directly fed into either channels, the channel input is obviously non-uniform and the capacity cannot be fully exploited since the uniform distribution achieves capacity for such channels. Therefore, the actual achievable capacity for AWGN channels with such a biased input is

$$\begin{aligned} C_{AWGN}^{bias} &= I(X;Y)|_{P(X=0)=p_0} \\ &= -p_0 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y+1)^2}{2\sigma^2}} \log_2(p_0 + p_1 e^{\frac{2y}{\sigma^2}}) dy \\ &\quad -p_1 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-1)^2}{2\sigma^2}} \log_2(p_0 e^{\frac{-2y}{\sigma^2}} + p_1) dy, \end{aligned}$$

and for Rayleigh fading channels, it is

$$\begin{aligned} C_{Ray}^{bias} &= I(X;Y)|_{P(X=0)=p_0} \\ &= -p_0 \int_0^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y+a)^2}{2\sigma^2}} \cdot 2ae^{-a^2} \cdot \log_2 \left[p_0 + p_1 e^{\frac{2ay}{\sigma^2}} \right] dy\ da \\ &\quad -p_1 \int_0^{\infty} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-a)^2}{2\sigma^2}} \cdot 2ae^{-a^2} \cdot \log_2 \left[p_0 e^{\frac{-2ay}{\sigma^2}} + p_1 \right] dy\ da. \end{aligned}$$

As a result, to transmit the non-uniform source directly, the capacity cannot be fully realized due to the mismatch between the channel input distribution and the capacity-achieving distribution. This is numerically illustrated in Table 1 for various values of CSNR, where we note that in certain cases C^{bias}/C can be as low as 41.4%. This leads us to conclude that for the transmission of very biased non-uniform sources, we should eliminate the systematic bit stream in the encoder structure, and hence investigate the design of non-systematic codes in our attempt to avoid such mismatch incurred performance loss.

3 Asymptotic Properties of Recursive Encoders

In this section, we prove two asymptotic properties of recursive convolutional encoders when the input sequence is a non-uniform i.i.d. source $\{U_k\}$ with distribution p_0 and length approaching infinity. Throughout this section, we assume that $p_0 \in (0.0, 1.0)$ and that $p_0 \neq 1/2$. We first show that, regardless of the value of p_0, when certain conditions are satisfied, the parity output of the encoder is asymptotically uniform. Based upon empirical observations, we also give a conjecture on the condition when the encoder output has higher order asymptotically uniform distributions (the joint distribution of $m > 1$ consecutive parity bits). We next show that, if the tap coefficient of the last memory element is unity, all of the encoder states will be reached with asymptotically equal probability. The conditions for the above two asymptotic properties are both necessary and sufficient. These two properties offer pertinent design criteria which will be used in the next section for constructing good encoder structures of Turbo codes for heavily biased non-uniform i.i.d. sources.

We first study the asymptotic distribution of a single parity check output of a constituent recursive convolutional encoder (either systematic or non-systematic, see Figure 1). We begin by quoting the following result from [33, Theorem, p. 1692 with $Q = 2$] and [34, Lemma 3, p. 1855].

Lemma 1: Consider the sequence $\{Y_n z_n\}$, where $\{Y_n\}$ is a stationary sequence of independent symbols from GF(2) with probability $P(Y_n = 0) = p_0$, and $\{z_n\}$ is a deterministic sequence containing an infinite number of nonzero terms. If $W_k = \sum_{n=0}^{k} Y_n z_n$, where the summation is modulo-2, then

$$\lim_{k\to\infty} P(W_k = 0) = \frac{1}{2}.$$

Using the above lemma, we can now establish the following result.

Theorem 1: For a recursive convolutional encoder with feedback polynomial $F(D)$ and feed-forward polynomial $G(D)$, the necessary and sufficient condition that the encoder output is asymptotically uniform for a non-uniform i.i.d. source with distribution p_0 (regardless of the value of p_0) is that $G(D)$ is not divisible by $F(D)$ in $GF(2)$.

Proof: a) *Necessary Part*: Suppose the source generates $\{U_k\}_{k=0}^{\infty}$, and the encoder produces an output sequence of $\{X_k\}_{k=0}^{\infty}$. Denote the encoder input and output in polynomial forms as $U(D) = \sum_{k=0}^{\infty} U_k D^k$ and $X(D) = \sum_{k=0}^{\infty} X_k D^k$, respectively. Also, define $E(D) \triangleq G(D)/F(D)$. Then for a recursive convolutional encoder with feedback polynomial $F(D)$ and feed-forward polynomial $G(D)$, we have

$$X(D) = U(D) \cdot \frac{G(D)}{F(D)} \triangleq U(D) \cdot E(D).$$

Suppose that $G(D)$ is divisible by $F(D)$ in GF(2). Then $E(D)$ is a polynomial of D with finite degree. Therefore, any bit of the encoder output X_k is essentially a modulo-2 summation of a *finite* number of bits from $\{U_k\}_{k=0}^{\infty}$. However, a finite modulo-2 summation of non-uniform i.i.d. variables would still be non-uniform; therefore, the encoder output $\{X_k\}_{k=0}^{\infty}$ is non-uniform.

b) *Sufficient Part*: When $G(D)$ is not divisible by $F(D)$ in GF(2), $E(D)$ is an infinite polynomial of D.

$$E(D) = e_0 + e_1 D + e_2 D^2 + \cdots + e_k D^k + \cdots ,$$

where e_k is binary. Then

$$X_k = \sum_{l=0}^{k} U_l e_{k-l}, \quad k = 0, 1, 2, \cdots , \text{ (where summation is in mod-2).}$$

Letting $z_l \stackrel{\triangle}{=} e_{k-l}$, where $l = 0, 1, \cdots , k$. Since $E(D) = G(D)/F(D)$ is a rational polynomial fraction, when $G(D)$ is not divisible by $F(D)$ in GF(2), $\{e_l\}$ is a periodic series. Therefore, when $k \to \infty$, the number of 1s in $\{e_l\}_{l=0}^{k}$ (and hence in $\{z_l\}_{l=0}^{k}$) also goes to infinity. From Lemma 1 we obtain that $\lim_{k\to\infty} P(X_k = 0) = 1/2$. □

In [30], Shamai and Verdú proved that for any fixed positive integer k, the k^{th}-order empirical distribution of any good code (i.e., a code approaching capacity with asymptotically vanishing probability of error) converges to the input distributions that achieve channel capacity. The capacity of a binary-input AWGN or Rayleigh channel is achieved when its mutual information is maximized by an independent and identically distributed uniform input. Therefore, having higher order uniform distributions (in addition to the first order distribution) in encoder outputs is also desirable. A related result has been established by Leeper in [34], which states that for any binary source (with unknown statistics), recursive convolutional encoders can produce a parity output whose first and second order distributions can be arbitrarily close (within a given value δ) to uniform for an appropriately chosen large value of the source sequence under the following conditions: (i) the source is first passed through the equivalent of a BSC with an arbitrarily small but nonzero error probability ϵ; (ii) the encoder's memory size is greater than a certain value determined by δ and ϵ. For our system, which does not satisfy the above (rather stringent) conditions, we can state the following conjecture based on extensive simulations.

Conjecture: Suppose a non-uniform i.i.d. source with distribution p_0 is input to a recursive convolutional encoder with feedback polynomial $F(D)$ and feed-forward polynomial $G(D)$. If $G(D)/F(D)$ is in its minimal form, where M is the degree of $F(D)$, then the m^{th}-order distribution of the encoder output is asymptotically uniform for any p_0 and $\forall\ m = 1, 2, \cdots, M$.

A proof for the above conjecture with arbitrary $G(D)$ and $F(D)$ is not obvious. However, the conjecture has been verified by empirical estimations of all possible combinations of recursive convolutional encoders with memory 4.

The above results indicate promising potentials when RNSC encoders are used as the constituent Turbo code encoders for the transmission of non-uniform i.i.d. sources. However, in this large family of candidate encoders, some structures offer inferior performance. Consider a heavily biased non-uniform i.i.d. sequence which is used as the input of a recursive convolutional encoder; for some structures, with high probability the state transition is confined within a few encoder states while other states may rarely or never be reached. This inefficient use of the encoder memory would result in performance degradations. Therefore, we next establish the condition for the asymptotic uniformity on the distribution of the encoder states. This result will be a useful design criterion in eliminating poor encoder structures.

Lemma 2: For a recursive convolutional encoder as depicted in Fig. 1, at time $k+1$, each state s_{k+1} can be reached from only two distinct (state, input) pairs $(s_k^{(0)}, u_{k+1}^{(0)})$ and $(s_k^{(1)}, u_{k+1}^{(1)})$, where the two states at time k satisfy $s_k^{(0)} \neq s_k^{(1)}$. The necessary and sufficient condition for $u_{k+1}^{(0)} \neq u_{k+1}^{(1)}$ is $f_M = 1$.

Proof: The encoder state is determined by the content of each memory element of the shift register. Let the state at time k be $s_k = (r_{1k}, r_{2k}, \cdots, r_{Mk})$. By the shift register's structure, we have

$$r_{1,k+1} = u_{k+1} \oplus \sum_{j=1}^{M} f_j \cdot r_{jk}, \tag{1}$$

$$r_{j,k+1} = r_{j-1,k}, \quad j = 2, 3, \cdots, M, \tag{2}$$

where $\oplus$ and the summation are modulo-2. Therefore, s_{k+1} can only be reached from two states which only differ at r_{Mk}. Now rewrite (1) as follows,

$$u_{k+1} = r_{1,k+1} \oplus f_M \cdot r_{Mk} \oplus \sum_{j=1}^{M-1} f_j \cdot r_{jk}.$$

For both states $s_k^{(0)}$ and $s_k^{(1)}$, the summand produces the same result. When $f_M = 0$, the input u_{k+1} is independent of the encoder state being $s_k^{(0)}$ or $s_k^{(1)}$. When $f_M = 1$, different values of r_{Mk} require different values of u_{k+1}. □

Theorem 2: Consider a recursive convolutional encoder with feedback $\{f_1, f_2, \cdots, f_M\}$, and let a non-uniform i.i.d. source with distribution p_0 be its input. Then the encoder state distribution is asymptotically uniform (as the source sequence length tends to infinity) regardless of the value of p_0 iff $f_M = 1$.

Proof: The shift register is a fully connected (irreducible) time-invariant Markov chain when p_0; so the state distribution will asymptotically converge to the steady state distribution. Then it is equivalent to show that the uniform distribution is or is not the steady state distribution. From Lemma 2, there are only two (state, input) pairs $(s_k^{(0)}, u_{k+1}^{(0)})$ and $(s_k^{(1)}, u_{k+1}^{(1)})$ that may transit to state s_{k+1}. Without loss of generality, denote the possible encoder state as $0, 1, \cdots, 2^M - 1$. Assuming that at time k the state distribution is uniform, i.e., $Pr(S_k = s) = 2^{-M}$ for $s = 0, 1, \cdots, 2^M - 1$, then at time $k+1$, if $f_M = 1$, we have

$$\begin{aligned} Pr(S_{k+1} = s_{k+1}) &= Pr(S_k = s_k^{(0)}, U_{k+1} = u_{k+1}^{(0)}) + Pr(S_k = s_k^{(1)}, U_{k+1} = u_{k+1}^{(1)}) \\ &= Pr(U_{k+1} = u_{k+1}^{(0)})2^{-M} + Pr(U_{k+1} = u_{k+1}^{(1)})2^{-M} \ = \ 2^{-M}, \end{aligned}$$

where the second equality is due to the independence of S_k and U_{k+1}, and the last equality holds because $u_{k+1}^{(0)} \neq u_{k+1}^{(1)}$ by Lemma 2. Therefore, $f_M = 1$ yields that the uniform distribution is the steady distribution of the encoder states. On the other hand, if $f_M = 0$, by Lemma 2, $u_{k+1}^{(0)} = u_{k+1}^{(1)}$; then $Pr(S_{k+1} = s_{k+1}) = 2Pr(U_{k+1} = u_{k+1}) \cdot 2^{-M}$. Thus, when the source is biased, the uniform distribution is not the steady state distribution when $f_M = 0$.

In other words, when $f_M = 0$, biased sources result in a biased distribution of the encoder states, and the steady state distribution is not equal to the uniform distribution. Hence for any p_0, the encoder state distribution is asymptotically uniform iff $f_M = 1$. $\square$

4 Non-Systematic Turbo Codes

In most of the Turbo codes literature, the Turbo code encoders are *systematic*; however, non-systematic Turbo codes were recently proposed by Costello *et al.* [35], and followed by a series of works in [36, 37, 38]. The scenario in these works is still for uniform i.i.d. sources. The motivation for using non-systematic Turbo codes as an alternative to their systematic peers is due to their larger code space; therefore, there is a potential that better codes *might* be found. In [38], Massey *et al.* construct an asymmetric non-systematic Turbo code that outperforms Berrou's (37,21) code by about 0.2 dB at the 10^{-5} BER level with a block length size of 4096. For larger block length, we expect that Massey's code to still outperform Berrou's code, although it is not clear if the 0.2 dB gain would be maintained.

As illustrated in Section 2, when non-uniform i.i.d. sources get heavily biased, a possibly significant performance loss may result due to the systematic structure. Furthermore, in light of Shamai and Verdú's result [30], Theorem 1 and the conjecture, we note that non-systematic Turbo codes seem to provide a suitable solution for the source-channel coding of such sources.

4.1 Design of Good Encoder Structures

Figs. 2 and 3 show our proposed non-systematic Turbo encoders. In a) the first constituent encoder has two parity outputs X_k^{1h} and X_k^{1g} while the second has only one parity output X_k^{2g}; so the overall rate is 1/3. This structure has been extensively used in [35]-[38]. In b) both constituent encoders have two parity outputs and the overall rate is 1/4. Structure b) can achieve the same overall rate of 1/3 via puncturing. It is also clear that structure a) is a special case of structure b), obtained by completely puncturing X_k^{2h}; therefore, a generally designed decoder for structure b) can also be employed for structure a).

Ideally, good design criteria are analytically based on the minimization of the error probability. However, to the best of our knowledge, all available error bounds for Turbo codes are obtained by averaging over a code ensemble or by assuming uniform interleaving and maximum likelihood decoding, while Turbo codes in fact employ random interleaving with each constituent decoder adopting the BCJR algorithm [39], which is a maximum *a posteriori* decoding algorithm. Furthermore, the bounds are useful only in the error floor region at high signal-to-noise ratios (SNRs); they are not tight in the waterfall region [40, 41, 42]. Since our goal is to obtain the best waterfall performance, we revert to other methods to find good design criteria.

As in [28, 29], we focus on symmetric 16-state encoders. The feedback tap coefficients are $\{f_0, f_1, f_2, f_3, f_4\}$. We always have $f_0 = 1$ since it provides the tap for the encoder input. Denote the feedback polynomial as $F(D)=f_0+f_1D+f_2D^2+f_3D^3+f_4D^4$, and the two feed-forward polynomials as $G(D)$ and $H(D)$, respectively. Then for such an RNSC encoder, there are altogether $2^4 \times (2^5 - 1) \times (2^5 - 1) = 15,376$ possible combinations; an exhaustive search over the entire code space is clearly not feasible. However, the two asymptotic properties studied in Section 3 serve as good design criteria to eliminate poor encoder structures.

First, as proved by Theorem 2 in Section 3, we choose to have the last feedback tap coefficient $f_4 = 1$ to fully exploit the encoder memory. For the feed-forward polynomials, having $G(D) = 1$ or $G(D) = D^m$, where $m = 1, \cdots, 4$ yields the same performance, since the only difference is a shift of time delay. The same holds for having $G(D) = 1 + D^2$ or $G(D) = D^2 + D^4$. Therefore, without loss of generality, we can choose to fix the first tap coefficient of both feed-forward polynomials to 1. Furthermore, for obvious reasons we do not want the two feed-forward polynomials to be identical. Then the total number of possible encoders in our search space is reduced down to $2^3 \times 2^4 \times (2^4 - 1) = 1920$, which is still impractical for an exhaustive search. Second, according to Theorem 1, we should only consider the choices of $G(D)$ and $F(D)$ such that $G(D)$ is not divisible by $F(D)$. Furthermore, by our empirically verified conjecture, given $f_4 = 1$, the selection of relatively prime $F(D)$ and $G(D)$ can guarantee that the encoder output has an asymptotically uniform 4^{th} order distribution. Thus additional inferior candidate encoder structures can be eliminated. Finally, again to avoid an exhaustive search, we take advantage of the optimization results found in systematic Turbo codes [29]; i.e., we first find the best pair of $F(D)$ and $G(D)$, and then search for the best second feed-forward polynomial $H(D)$. The iterative search procedure for (sub-)optimal encoder structures, given a source distribution, is implemented as follows.

1) Using a systematic Turbo code, fix the feed-forward polynomial $G(D) = 1$ and search for the best feedback polynomial $F(D)$ with $f_0 = f_4 = 1$;

2) With the $F(D)$ found in step 1), search for the best $G(D)$ among all remaining possible choices with the condition that the greatest common divisor of $F(D)$ and $G(D)$ is 1;

3) With the $G(D)$ found in step 2), go back to step 1). If $F(D)$ coincides with the one obtained in step 1), go to step 4); otherwise, proceed to step 2);

4) For a non-systematic Turbo code, fix the pair $(F(D), G(D))$ found above and search for the best second feed-forward polynomial $H(D)$.

4.2 Decoder Design

When RNSC encoders are used as constituent encoders, unlike in systematic Turbo codes, the *a posteriori* log-likelihood ratio $\Lambda(U_k)$ in the Turbo decoder [23, 43] can only be decomposed into two terms:

$$\Lambda(U_k) = L_{ex}(U_k) + L_{ap}(U_k),$$

where the new extrinsic term involves two parity sequences. For AWGN channels, we have

$$L_{ex}(U_k) = \log \frac{\sum_s \sum_{s'} \gamma(y_k^h, y_k^g | 1, s, s') \cdot \alpha_{k-1}(s') \cdot \beta_k(s)}{\sum_s \sum_{s'} \gamma(y_k^h, y_k^g | 0, s, s') \cdot \alpha_{k-1}(s') \cdot \beta_k(s)},$$

where the summations for s and s' are both over all 2^M possible states, and for $i = 0, 1$,

$$\gamma(y_k^h, y_k^g | i, s, s') = p(y_k^h | U_k = i, S_k = s) \cdot p(y_k^g | U_k = i, S_k = s) \cdot Pr\{U_k = i | S_k = s, S_{k-1} = s'\},$$

and where S_k is the encoder state at time k, y_k^h and y_k^g are the noise corrupted version of x_k^h and x_k^g, which are the parity bits generated from the two feed-forward polynomials.

For Rayleigh fading channels, the extrinsic term needs to be modified to appropriately incorporate the channel statistics. The extrinsic term therefore becomes

$$L_{ex}(U_k) = \log \frac{\sum_{s,s'} \gamma(y_k^h, y_k^g | 1, s, s', a_k^h, a_k^g) \cdot \alpha_{k-1}(s') \cdot \beta_k(s)}{\sum_{s,s'} \gamma(y_k^h, y_k^g | 0, s, s', a_k^h, a_k^g) \cdot \alpha_{k-1}(s') \cdot \beta_k(s)},$$

where for $i = 0, 1$,

$$\begin{aligned} \gamma(y_k^h, y_k^g | i, s, s', a_k^h, a_k^g) &= p(y_k^h | U_k = i, S_k = s, a_k^h) \cdot p(y_k^g | U_k = i, S_k = s, a_k^g) \\ &\quad \cdot Pr\{U_k = i | S_k = s, S_{k-1} = s'\}, \end{aligned}$$

where a_k^h and a_k^g are the fading factors. When the source is non-uniform i.i.d., $\log((1-p_0)/p_0)$ is used as the initial a priori input $L_{ap}(U_k)$ to the first decoder at the first iteration; then at the following iterations, $L_{ex} + \log((1-p_0)/p_0)$ is used as the new extrinsic information for both constituent decoders at each iteration.

5 Numerical Results and Discussion

In this section, we present simulation results of our non-systematic Turbo codes for uniform i.i.d. sources over BPSK-modulated AWGN and Rayleigh fading channels with known channel state information[3]. In the following, all non-systematic Turbo codes adopt the decoder discussed in Section 4.2, while all systematic codes employ the BCJR algorithm with the modification proposed in [29].

[3]It is assumed throughout that the decoder has perfect knowledge of the source distribution p_0. Simulations on the effect of mismatch in p_0 show little performance loss if the value of p_0 used at the decoder is within a reasonable distance from the true p_0. For example, when the true distribution is $p_0 = 0.9$ and the decoder assumes it is 0.8, the performance degradation is no more than 0.1 dB. Furthermore, information on p_0 can be sent to the receiver as an overhead with negligible bandwidth loss. Finally, if no overhead information is sent, p_0 can possibly be estimated at the decoder (e.g., see [27] where the source statistics are estimated at the receiver in the context of Turbo decoding of Hidden Markov sources).

5.1 Performance Evaluations

The performance is measured in terms of the bit error rate (BER) versus E_b/N_0, where $E_b = E_s/R_c$ is the average energy per source bit. All simulated Turbo codes have 16-state constituent encoders and use the same pseudo-random interleaver introduced in [23]. The sequence length is $N = 512 \times 512 = 262144$ and at least 200 blocks are simulated; this would guarantee a reliable BER estimation at the 10^{-5} level with 524 errors. The number of iterations employed in the decoder is 20; note that additional iterations result in minor improvements. All results are presented for Turbo codes with structure b) encoders (see Fig. 3) as they provide a better performance than codes with structure a) encoders. Simulations are performed for rates $R_c = 1/3$ and $R_c = 1/2$ with p_0=0.8 and 0.9. From our simulations, for both rates 1/3 and 1/2, the best found RNSC encoder structure for p_0=0.8 and for both channels has each constituent encoder with feedback polynomial 35 and feed-forward polynomials 23 and 25; this is denoted by the octal triplet (35,23,25). For p_0=0.9 the best structure is (31,23,27). Several other encoders give very competitive performance, such as (35,23,31) for p_0=0.8, (31,23,35) and (31,23,37) for p_0=0.9. We hereafter characterize all non-systematic Turbo codes by triplets as described above (while their systematic peers are represented by the conventional octal pair).

Fig. 4 shows the performance over AWGN channels of our rate-1/3 non-systematic Turbo codes in comparison with their systematic peers investigated in [28, 29], as well as with Berrou's (37,21) code, which offers the best waterfall performance (among 16-state encoders) for uniform i.i.d. sources. At the 10^{-5} BER level, when p_0=0.8, our (35,23,25) non-systematic Turbo code offers a 0.45 dB gain over its (35,23) systematic peer; when p_0=0.9, our (31,23,27) code offers an improvement of 0.89 dB over the systematic (31,23) code[4]. In comparison with Berrou's (37,21) code performance, the gains achieved by exploiting the source redundancy

[4]It should also be indicated that our non-systematic joint source-channel Turbo codes maintain a similar level of performance gains over their systematic peers when the sequence length is shorter than 262144. For example, for a sequence length of $N = 128 \times 128 = 16384$, with p_0=0.9 and the same rate and channel

and encoder optimization are therefore 1.48 dB and 3.25 dB for p_0=0.8 and 0.9, respectively.

Fig. 5 shows similar results over AWGN channels for rate-1/2. In this case, the gains are generally more significant. In comparison with the best systematic Turbo codes, at the 10^{-5} BER level, for p_0=0.8 and 0.9, the gains achieved are 0.69 dB and 1.56 dB, respectively. Furthermore, the gains over Berrou's code due to combining the optimized encoder with the modified decoder that exploits the source redundancy are 1.57 dB (p_0=0.8) and 3.72 dB (p_0=0.9).

Simulations over Rayleigh channels are also provided in Figs. 6 and 7. For rate-1/3 codes, at a 10^{-5} BER, when p_0=0.8, our (35,23,25) non-systematic Turbo code provides a 0.40 dB gain over its (35,23) systematic peer; when p_0=0.9, the improvement is 1.01 dB with encoder structure (31,23,27). In comparison with Berrou's (37,21) code, the gains achieved by exploiting the source redundancy and encoder optimization are 1.76 dB and 3.87 dB for p_0=0.8 and 0.9, respectively. For rate-1/2, as shown in Fig. 7, the gains are more pronounced. In comparison with the best systematic Turbo codes, at a 10^{-5} BER, for p_0=0.8 and 0.9, the gains are 0.77 dB and 1.84 dB, respectively. Furthermore, the gains due to combining the optimized encoder with the modified decoder are 2.01 dB (for p_0=0.8) and 4.71 dB (for p_0=0.9).

To achieve a desired rate via puncturing, using different puncturing patterns may result in different performances. For example, when structure b) is used for an overall rate of 1/3, we may choose to puncture 1/4 of each parity sequence according to various patterns, or we may puncture half of two parity sequences, and leave the other two sequences intact. Simulations show the best puncturing pattern is to keep the parity sequence generated by feed-forward polynomial 23 intact and puncture half of the one generated by the other feed-forward polynomial. The performance of this puncturing pattern is about 0.2 dB better than

conditions as in Fig. 4, our (31,23,27) non-systematic code offers around 0.9 dB gain over its (31,23) systematic peer; a similar gain is obtained for $N = 32 \times 32 = 1024$. These gains are interesting to note for practical situations where delay may be limited.

other patterns; in particular it is 0.3 dB better than the performance offered by structure a). For an overall rate of 1/2, structure b) is also better than a), and the best puncturing pattern is to delete all even (odd) position bits of the sequences generated by feed-forward polynomial 23, and delete all odd (even) position bits of the sequences generated by the other feed-forward polynomial.

In previous work on non-systematic Turbo codes for uniform i.i.d. sources, it is verified via extensive simulations that most non-systematic codes show inferior performance to their systematic peers [36]-[38], except for the one in [38] employing a non-systematic "quick-look-in" constituent code, which is basically "close" to a systematic code. Also, it is pointed out in [37, 38] that at low values of E_b/N_0, the initial extrinsic estimates for the information bits provided by non-systematic Turbo codes are not as good as those of systematic Turbo codes due to the lack of received channel values; this motivates their choice of a close-to-systematic code in the non-systematic family. However, as shown by our results, systematic Turbo codes are not a suitable choice for heavily biased non-uniform sources; therefore, close-to-systematic codes may not be desired in this case. Another encoder structure called "big-numerator-little-denominator" [35], which provides gains over Berrou's code for uniform sources, may also be unsuitable for non-uniform sources since the denominator results in non-uniform higher order output distributions.

In the scenario of non-uniform i.i.d. sources, there are two factors playing against each other: the a priori knowledge of p_0 at the decoder, and the systematic structure. When the source is heavily biased, systematic codes considerably underperform due to the distribution mismatch between the source and the capacity achieving channel input. Furthermore, the knowledge of a biased p_0 can give good initial extrinsic estimations of the information bits at the decoder, thus eliminating the need for the systematic structure. This may explain why non-systematic Turbo codes, which do not suffer from any distribution mismatch (at least asymptotically), offer superior performance over their systematic counterparts. On the other hand, when the source distribution is close to uniform, the distribution mismatch due to a systematic structure becomes minor. Furthermore, very little useful knowledge about the information bits is provided from p_0 at the decoder, while a systematic structure, even when its bits are received corrupted at the receiver due to channel noise, provides more reliable extrinsic estimations in the initial iterative decoding stages for low E_b/N_0. In particular, as pointed out in [36]-[37], when the source is exactly uniform, there is no useful knowledge from p_0 at all, and thus the systematic structure plays a critically important role. To investigate the role of the systematic bits for less biased non-uniform i.i.d. sources, we also study the performance of so-called "extended" systematic Turbo codes. The results are briefly summarized in the following.

The encoder of an "extended" systematic Turbo code consists of one systematic output, two RNSC encoders, each of which produces two parity outputs. The overall rate is therefore 1/5. Higher rates (e.g., 1/3 and 1/2) are obtained by partial puncturing of the systematic part and partial puncturing of the four parity sequences. For p_0=0.6, 0.7 and 0.8, and for R_c=1/2 and 1/3, we performed simulations using the following puncturing patterns: delete $j/8$ of the systematic sequence, where $j = 0, 1, \cdots, 8$, and delete evenly the four parity sequences to maintain the desired overall rates. Our simulations demonstrate that when p_0=0.7, starting from $j = 0$, the performance is improved as j increases up to $j = 4$, which yields approximately a 0.1 dB gain in E_b/N_0 at the 10^{-5} BER level over the system with pattern $j = 0$. Then the performance degrades as j increases from 4 to 8. Therefore, when p_0=0.7, puncturing half of the systematic sequence yields the best performance. When p_0=0.8, we observe a monotonic performance improvement as j increases from 0 up to 8, which indicates that purely non-systematic Turbo codes should be preferred. When p_0=0.6, we observe exactly the opposite behavior: preserving all systematic gives the best performance. These observations indicate, as remarked in [35]-[38], that a systematic encoding structure is especially important when not enough a priori knowledge about the source is available at the decoder.

5.2 Shannon Limit

Shannon's *Lossy Joint Source-Channel Coding Theorem* states that, for a given memoryless source and channel pair[5] and for sufficiently large source block lengths, the source can be transmitted via a source-channel code over the channel at a transmission rate of R_c source symbols/channel symbols and reproduced at the receiver end within an end-to-end distortion given by D if the following condition is satisfied [32]:

$$R_c \cdot R(\mathrm{D}) < C, \tag{3}$$

where C is the channel capacity and $R(\mathrm{D})$ is the source rate-distortion function. For a discrete binary non-uniform i.i.d. source with distribution p_0, we have that D=P_e (the BER) under the Hamming distortion measure; then $R(\mathrm{D})$ becomes

$$R(P_e) = \begin{cases} h_b(p_0) - h_b(P_e), & 0 \le P_e \le \min\{p_0, p_1\} \\ 0, & P_e > \min\{p_0, p_1\} \end{cases}$$

where $p_1 = 1 - p_0$, and $h_b(x) = -x \log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function.

As seen in Section 2, the capacity of an AWGN or Rayleigh fading channel is a function of CSNR, or equivalently, E_b/N_0; therefore, the optimum value of E_b/N_0 to guarantee a BER of P_e – called the Shannon limit or OPTA – can be solved using (3) assuming equality. The Shannon limit cannot be explicitly solved for our BPSK-modulated channels due to the lack of a closed form expression; so it is computed via numerical integration.

For the simulations of the above subsection, the OPTA values at the 10^{-5} BER level are computed. The OPTA gaps, which are the distances between our system performance and the corresponding OPTA values, are provided in Table 2. We observe that the OPTA gaps

[5] The above theorem also holds for wider classes of sources and channels with memory (e.g., stationary ergodic sources and channels with additive stationary ergodic noise) [31].

are significantly reduced by non-systematic Turbo codes. For example, for AWGN channels and $R_c = 1/2$, when p_0=0.8, systematic Turbo codes provide a performance which is 1.56 dB away from OPTA; on the other hand, for our non-systematic Turbo codes the OPTA gap is 0.87 dB. When p_0=0.9, the OPTA gap is reduced from 2.61 dB down to 1.05 dB.

6 Comparison with Tandem Schemes

Traditionally, source and channel coding are designed separately, resulting in a so-called *tandem* coding scheme. That is, the source is compressed first, and then channel coded. This is justified by Shannon's separation principle [44], which states that there is no loss of optimality in such separation as long as unlimited coding delay and complexity are available. However, in practice, joint source-channel coding often outperforms traditional tandem coding when delay and system complexity are constrained.

We next compare the performance of our joint source-channel system with that of two tandem schemes for the same overall transmission rate. Each tandem scheme consists of a 4^{th}-order Huffman code followed by a rate R_c=1/3 Turbo code. The overall rate for both tandem and joint source-channel coding systems is $r = R_c/R_s$=1/2 source symbol/channel symbol (in the joint coding scheme, $R_c = 1/2$ and $R_s = 1$ since no source coding is performed). Therefore, the Huffman code needs to be of rate R_s=2/3 code bits/source symbol. Since the average rate of the Huffman code depends on the source distribution, we need to find the value of p_0 which renders the Huffman code rate (not the entropy) R_s=2/3 code bits/source symbol with satisfactory accuracy. By using the bisection method, we obtain that when p_0=0.83079, a 4^{th}-order Huffman code has rate R_s=0.666668 code bits/source symbol. Therefore, simulations are performed for this value of p_0.

Berrou's pseudo-random interleaver [23] requires the sequence length to be an even power of 2; this inflexibility is an obstacle in the design of the tandem scheme, since the Huffman code is a variable-length code. The S-random interleaver [45], however, can take an input sequence of arbitrary length and it yields good BER performance. We thus herein adopt the S-random interleaver in the Turbo code. We also do not terminate the first constituent Turbo encoder, because otherwise errors in the tail bits of the Turbo-decoded sequence would introduce irrecoverable errors in the Huffman-decoded sequence. For fair comparison, our system also uses the S-random interleaver and its first constituent encoder is not terminated.

The tandem scheme is implemented as follows: 1) the source generates a non-uniform i.i.d. sequence with length N, and p_0=0.83079; 2) the Huffman encoder produces a compressed sequence with variable length, whose mean is approximately $(2/3)N$; 3) an S-random interleaver is generated for this given length; 4) the sequence is Turbo-encoded using the S-random interleaver generated in 3); 5) the sequence is BPSK modulated and transmitted over a Rayleigh fading channel; 6) the sequence is Turbo-decoded and then Huffman-decoded.

Another issue is the choice of the source sequence length N. Due to error propagation in the Huffman decoder, a few errors in the Turbo-decoded sequence could result in a big percentage of errors in the final Huffman-decoded sequence. Furthermore, what matters is not only the number of errors in the Turbo-decoded sequence, but also the positions of the erroneous bits. This is due to the fact that an error occurring in the first few bits of the Turbo-decoded sequence has a much longer propagating effect than an error occurring in the tail bits. Therefore, a sufficiently large number of blocks is necessary to obtain a good average performance. After several tests, we chose to use N=12000 and 60000 blocks.

For a given N, the larger the "spread" S of the S-random interleaver is, the better is the performance. However, in practice, generating an S-random interleaver with a large S requires a substantial amount of computing time, and sometimes such an interleaver may not be successfully generated. Thus, in order to reduce the computation time and to guarantee the successful generation of S-random interleavers of arbitrary size, S is set to equal 10.

Fig. 8 shows the performance of our system versus that of two tandem schemes over Rayleigh channels. Twenty iterations are used in the Turbo decoder. In the first tandem scheme, the Turbo code with R_c=1/3 is Berrou's (37,21) code, which offers an excellent waterfall performance for uniform sources among all 16-state codes. However, due to a relatively high error-floor provided by Berrou's code, this tandem scheme suffers from a high-BER performance caused by error propagation in the Huffman decoder. Thus, we also evaluate a second tandem scheme using the (35,23) Turbo code, which has a significantly lower error-floor at the expense of a slight waterfall performance loss. Although at very high BERs, both tandem schemes are better than our system, their error-floors occur at high BERs (10^{-3} for the (37,21) code, and 10^{-4} for the (35,23) code). Therefore, at low BER levels, our system offers superior performance than both tandem schemes. Interestingly, most traditional joint source-channel coding schemes outperform tandem schemes at high BER levels (e.g., [2, 3, 7]), while in the context of Turbo codes, the opposite result is observed. Alternative source-channel coding systems using jointly designed Huffman and Turbo codes have been recently proposed in [46, 47]. It would be interesting to make performance comparisons with these schemes. However, our system has lower complexity since the source encoding and decoding parts are omitted.

7 Conclusion

In this work, the joint source-channel coding issue of transmitting non-uniform memoryless sources via Turbo codes over AWGN and Rayleigh channels is investigated. Necessary and sufficient conditions are proved for recursive convolutional encoders having asymptotically uniform state and marginal output distributions regardless of the degree of source non-uniformity. Therefore, recursive non-systematic Turbo source-channel codes are proposed, and the outputs of our selected codes are suitably matched to the channel input as they nearly maximize the channel mutual information. Simulation results demonstrate substantial coding gains (up to 1.84 dB) over systematic Turbo codes designed in [28, 29], and the OPTA gaps are significantly reduced. Finally, our system is compared with two tandem schemes, which employ a near-optimal Huffman code followed by a standard Turbo code. Our system offers substantially better performance at low BERs and enjoys a lower complexity.

Acknowledgment

The authors would like to thank the anonymous reviewers for their valuable suggestions, which were very helpful in improving the presentation of this paper.

References

[1] P. Burlina, F. Alajaji and R. Chellappa, "Transmission of two-tone images over noisy communication channels with memory," Technical Report, CAR-TR-814, Center for Automation Research, Univ. of Maryland, College Park, 1996.

[2] J. Kroll and N. Phamdo, "Source-channel optimized trellis codes for bitonal image transmission over AWGN channels," *IEEE Trans. Image Processing,* vol. 8, pp. 899–912, Jul. 1999.

[3] F. Alajaji, N. Phamdo and T. Fuja, "Channel codes that exploit the residual redundancy in CELP-encoded speech," *IEEE Trans. Speech Audio Processing*, vol. 4, pp. 325–336, Sept. 1996.

[4] R. B. Blizard, "Convolutional coding for data compression," Martin Marietta Corp., Denver Div., Rep. R-69-17, 1969.

[5] V. N. Koshelev, "Direct sequential encoding and decoding for discrete sources," *IEEE Trans. Inform. Theory,* vol. 19, pp. 340–343, May 1973.

[6] M. E. Hellman, "Convolutional source encoding," *IEEE Trans. Inform. Theory,* vol. 21, no. 6, pp. 651–656, Nov. 1975.

[7] F. Alajaji, N. Phamdo, N. Farvardin and T. Fuja, "Detection of binary Markov sources over channels with additive Markov noise," *IEEE Trans. Inform. Theory,* pp. 230–239, Jan. 1996.

[8] T. Fingscheidt, T. Hindelang, R. V. Cox and N. Seshadri, "Joint source-channel (de)coding for mobile communications," *IEEE Trans. Commun.*, vol. 50, No. 2, pp. 200–212, Feb. 2002.

[9] J. Hagenauer, "Source controlled channel decoding," *IEEE Trans. Commun.*, vol. 43, pp. 2449–2457, Sept. 1995.

[10] F. Lahouti and A. Khandani, "Sequence MMSE source decoding over noisy channels using the residual redundancies," *Proc. Allerton Conf. Commun., Cont., Comput.*, Monticello, Oct. 2001.

[11] M. Park and D. Miller, "Improved image decoding over noisy channels using minimum mean-squared estimation and a Markov mesh," *IEEE Trans. Image Proc.,* vol. 8, no. 6, pp. 863–867, Jun. 1999.

[12] M. Park and D. Miller, "Joint source-channel decoding for variable-length encoded data by exact and approximate MAP sequence estimation," *IEEE Trans. Commun.,* vol. 48, no. 1, pp. 1–6, Jan. 2000.

[13] M. Adrat, P. Vary and J. Spittka, "Iterative source-channel decoding using extrinsic information of softbit-source decoding," *Proc. IEEE ICASSP 2001*, pp. 2653–2656, Salt Lake City, Utah, May 2001.

[14] N. Görtz, "On the iterative approximation of optimal joint source-channel decoding," *IEEE J. Selec. Areas Commun.*, vol. 19, no. 9, pp. 1662–1670, Sept. 2001.

[15] N. Phamdo, *Quantization over Discrete Noisy Channels under Complexity Constraints*, Ph.D. Dissertation, University of Maryland, 1993.

[16] N. Phamdo and N. Farvardin, "Optimal detection of discrete Markov sources over discrete memoryless channels—applications to combined source-channel coding," *IEEE Trans. Inform. Theory,* vol. 40, no. 1, pp. 186–193, Jan. 1994.

[17] N. Phamdo and N. Farvardin, "Scalar quantization of memoryless sources over memoryless channels using rate-one convolutional codes," *Proc. Int. Symp. Inform. Theory*, p. 235, Trondheim, Norway, Jun. 1994.

[18] K. Sayood and J. C. Borkenhagen, "Use of residual redundancy in the design of joint source-channel codes," *IEEE Trans. Commun.*, vol. 39, pp. 838–846, Jun. 1991.

[19] M. Skoglund, "Soft decoding for vector quantization over noisy channels with memory," *IEEE Trans. Inform. Theory,* vol. 45, no. 4, pp. 1293–1307, May 1999.

[20] M. Skoglund, "Bit-estimate based decoding for vector quantization over noisy channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1309–1317, Aug. 2000.

[21] R. E. Van Dyck and D. Miller, "Transport of wireless video using separate, concatenated, and joint source-channel coding", *Proc. of the IEEE*, vol. 87, pp. 1734-1750, Oct. 1999.

[22] W. Xu, J. Hagenauer and J. Hollmann, "Joint source-channel decoding using the residual redundancy in compressed images," *Proc. Int. Conf. Commun.*, pp. 142–148, Dallas, June 1996.

[23] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *IEEE Trans. Commun.*, vol. 44, pp. 1261–1271, Oct. 1996.

[24] E. K. Hall and S. G. Wilson, "Design and analysis of Turbo codes on Rayleigh fading channels," *IEEE Journal Selected Areas Commun.*, vol. 16, pp. 160–174, Feb. 1998.

[25] R. J. McEliece, "Are Turbo-like codes effective on nonstandard channels?" *IEEE Inform. Theory Soc. Newsletter*, vol. 51, no. , pp. 1–8, Dec. 2001.

[26] J. Garcia-Frias and J. D. Villasenor, "Combining hidden Markov source models and parallel concatenated codes," *IEEE Commun. Lett.,* vol. 1, pp. 111-113, Jul. 1997.

[27] J. Garcia-Frias and J. D. Villasenor, "Joint turbo decoding and estimation of hidden Markov sources," *IEEE J. Selec. Areas Commun.*, vol. 19, pp. 1671–1679, Sept. 2001.

[28] G.-C. Zhu and F. Alajaji, "Design of Turbo codes for non-equiprobable memoryless sources," *Proc. 39th Allerton Conf. Commun., Cont., Computing,* Monticello, IL, Oct. 2001.

[29] G.-C. Zhu and F. Alajaji, "Turbo codes for non-uniform memoryless sources over noisy channels," *IEEE Commun. Lett.*, vol. 6, pp. 64–66, Feb. 2002.

[30] S. Shamai and S. Verdú, "The empirical distribution of good codes," *IEEE Trans. Inform. Theory*, vol. 43, pp. 836–846, May 1997.

[31] T. Berger, *Rate Distortion Theory: A Mathematical Basis for Data Compression*, Prentice Hall, 1971.

[32] R. J. McEliece, *The Theory of Information and Coding,* 2nd Edition, Addison-Wesley, 2002.

[33] I. J. Fair, V. K. Bhargava and Q. Wang, "On the power spectral density of self-synchronizing scrambled sequences," *IEEE Trans. Inform. Theory*, vol. 44, no. 4, pp. 1687-1693, July 1998.

[34] D. G. Leeper, "A universal digital data scrambler," *Bell Syst. Tech. J.*, vol. 52, no. 10, pp. 1851–1865, Dec. 1973.

[35] D. J. Costello, *Jr.*, P. C. Massey, O. M. Collins and O. Y. Takeshita, "Some reflections on the mythology of turbo-codes," in *Proc. 3rd ITG Conf. Source Channel Coding*, pp. 157–160, Munich, Germany, Jan. 2000.

[36] O. M. Collins, O. Y. Takeshita and D. J. Costello, *Jr.*, "Iterative decoding of non-systematic Turbo codes," *Proc. of Int. Symp. Inform. Theory*, Sorrento, Italy, pp. 172, June 2000.

[37] D. J. Costello, *Jr.*, H. A. Cabral and O. Y. Takeshita, "Some thoughts on the equivalence of systematic and nonsystematic convolutional encoders," *Codes, Graphs and Systems,* (edited by R. E. Blahut and R. Koetter), Kluwer Academic, 2002.

[38] P. C. Massey and D. J. Costello, *Jr.*, "Turbo codes with recursive nonsystematic quick-look-in constituent codes," *Proc. Int. Symp. Inform. Theory*, p. 141, Washington, DC, Jun. 2001.

[39] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory,* vol. 20, pp. 248-287, Mar. 1974.

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

[41] T. M. Duman and M. Salehi, "New performance bounds for turbo codes," *IEEE Trans. Commun.,* vol. 46, No. 6, pp. 717–723. June 1998.

[42] S. Shamai and I. Sason, "Variations on Gallager's bounding techniques: performance bounds for Turbo codes in Gaussian and fading channels," *Proc. 2nd Int. Symp. on Turbo Codes and Related Topics*, Brest, France, pp. 27–34, Sept. 2000.

[43] J. Hagenauer, E. Offer and L. Papke, "Iterative decoding of binary block and convolutional codes," *IEEE Trans. Inform. Theory,* vol. 42, pp. 429–445, Mar. 1996.

[44] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, 1948.

[45] D. Divsalar and F. Pollara, "Multiple Turbo codes for deep-space communications," TDA Progress Report 42-121, Jet Propulsion Laboratory, Pasadena, CA, pp. 66–77, May 15, 1995.

[46] L. Guivarch, J.-C. Carlach and P. Siohan, "Joint source-channel soft decoding of Huffman codes with Turbo-codes," *Proc. DCC 2000*, pp. 83–92, Snowbird, Utah, Mar. 2000.

[47] J. Hagenauer and R. Bauer, "The Turbo principle in joint source channel decoding of variable length codes," *Proc. IEEE Inform. Theory Work.*, pp. 128–130, Australia, Sept. 2001.

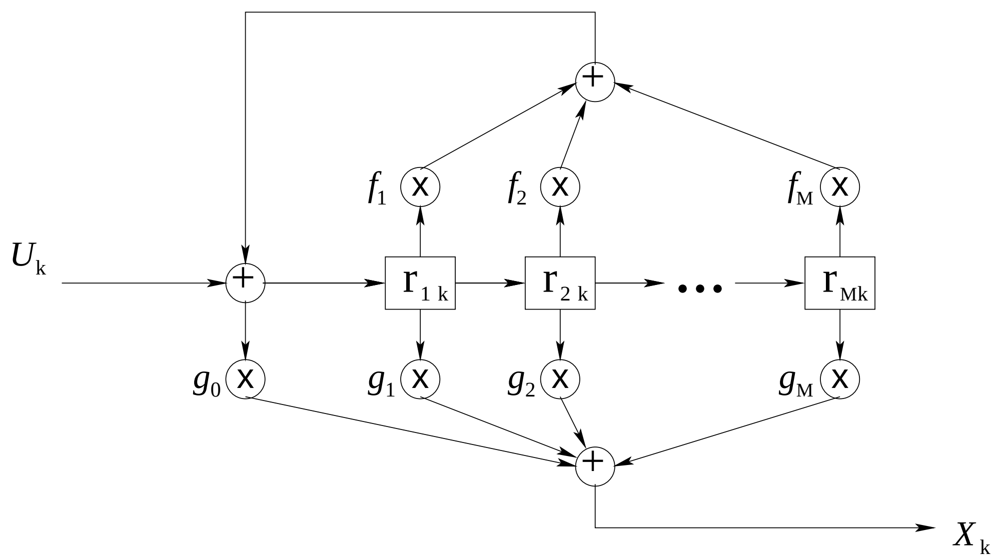

Figure 1: General structure of a recursive convolutional encoder.

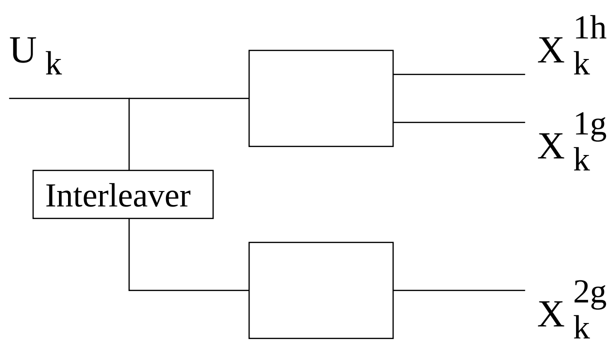

Figure 2: Non-Systematic Turbo encoder structure a).

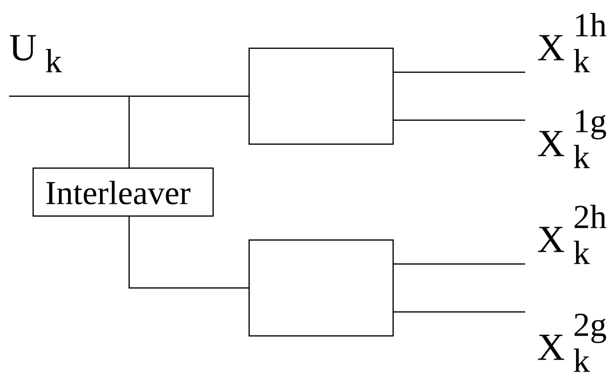

Figure 3: Non-Systematic Turbo encoder structure b).

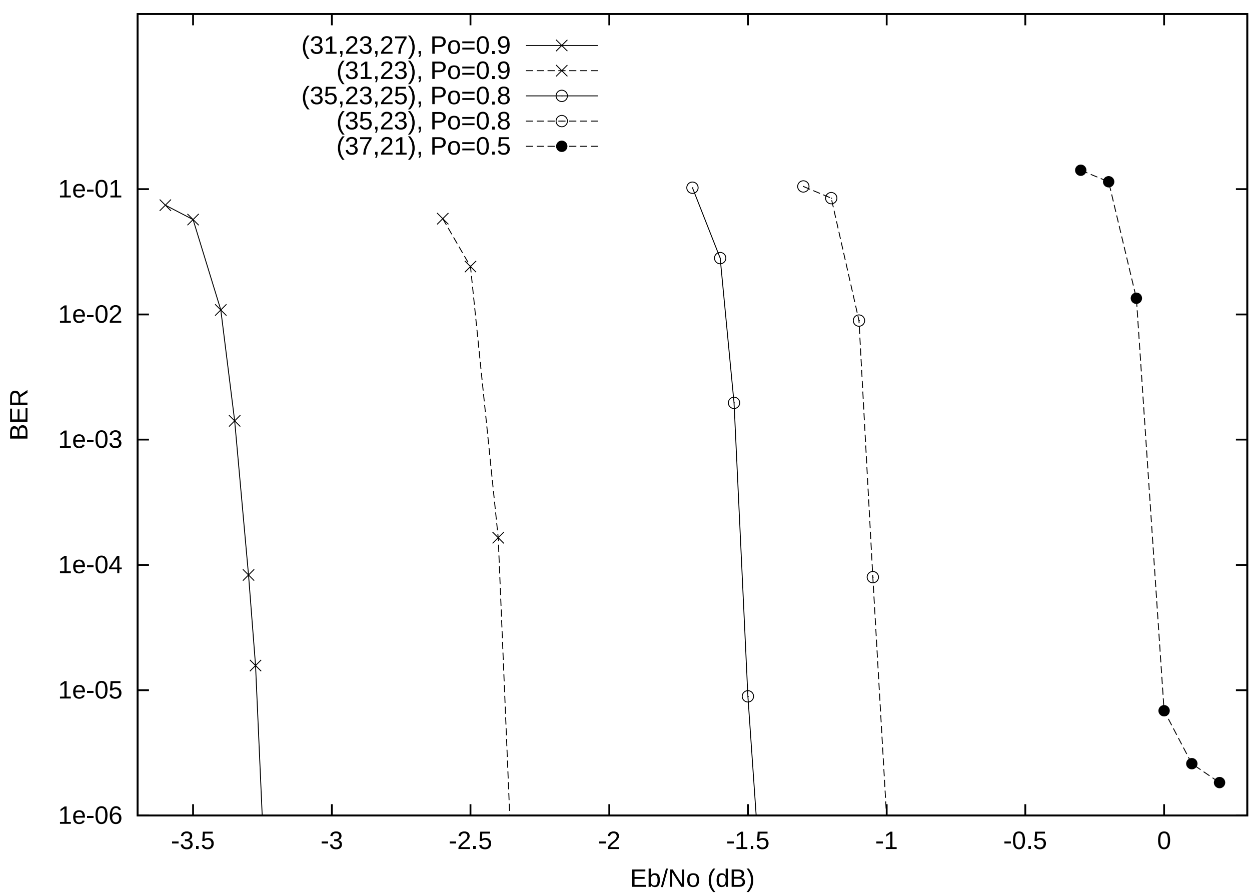

Figure 4: Turbo codes for non-uniform i.i.d. sources, R_c=1/3, N=262144, AWGN channel.

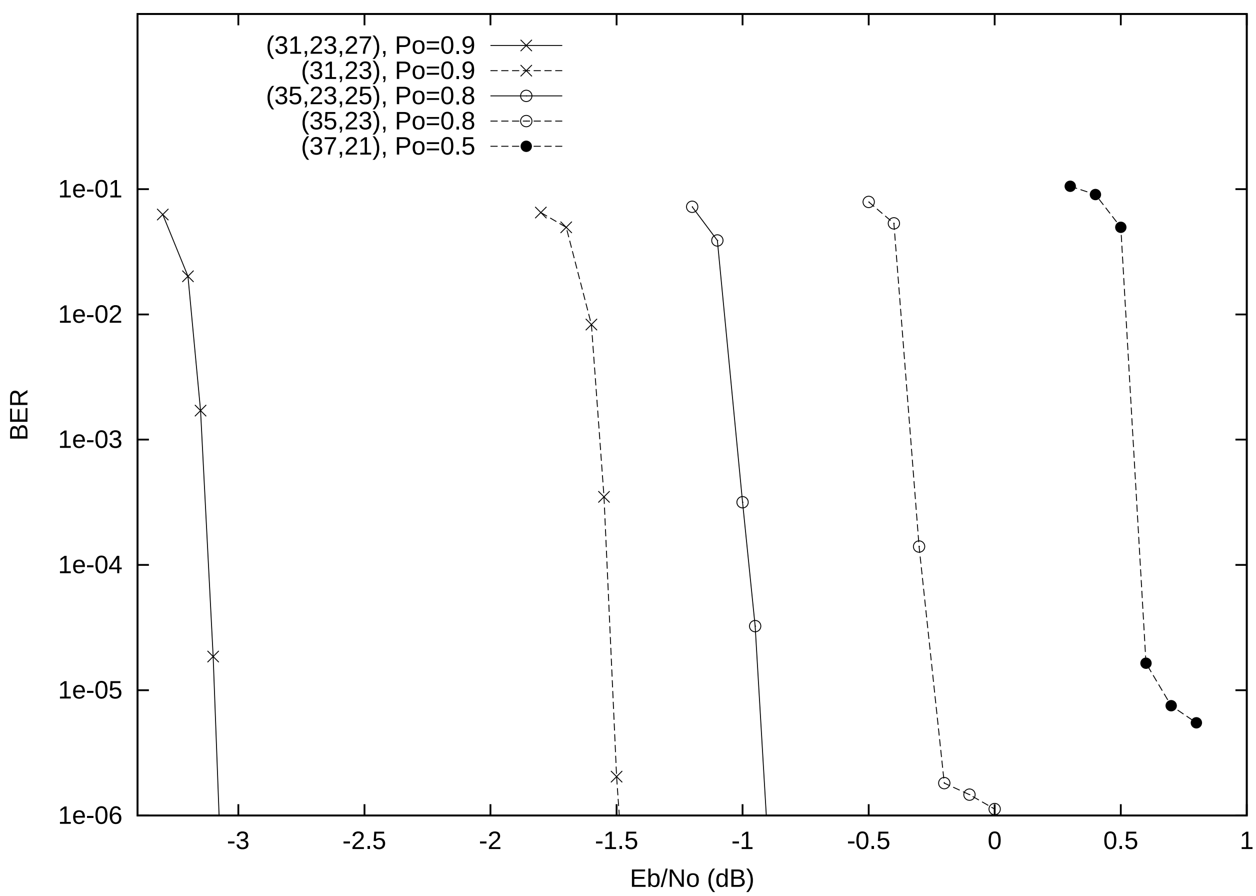

Figure 5: Turbo codes for non-uniform i.i.d. sources, R_c=1/2, N=262144, AWGN channel.

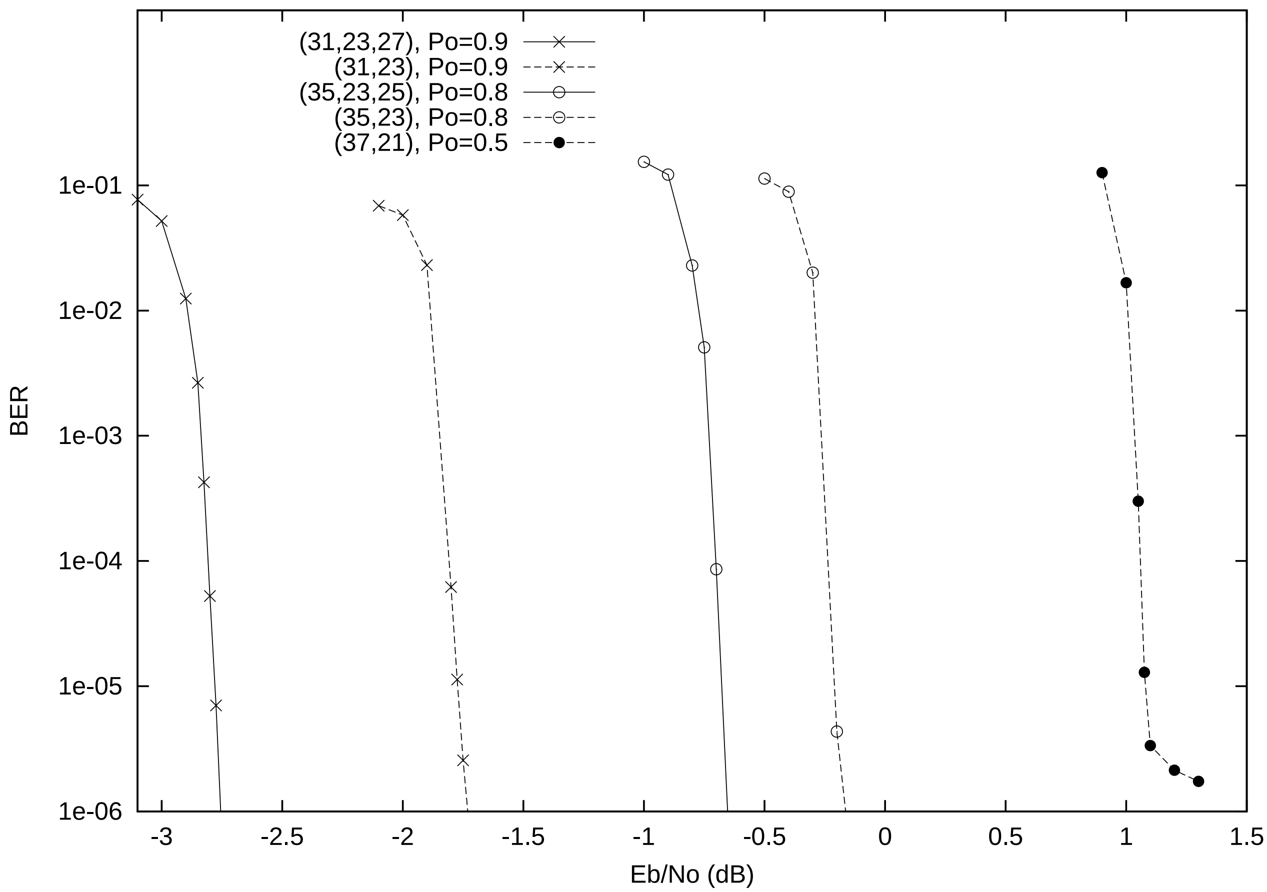

Figure 6: Turbo codes for non-uniform memoryless sources, R_c=1/3, N=262144, Rayleigh fading channel.

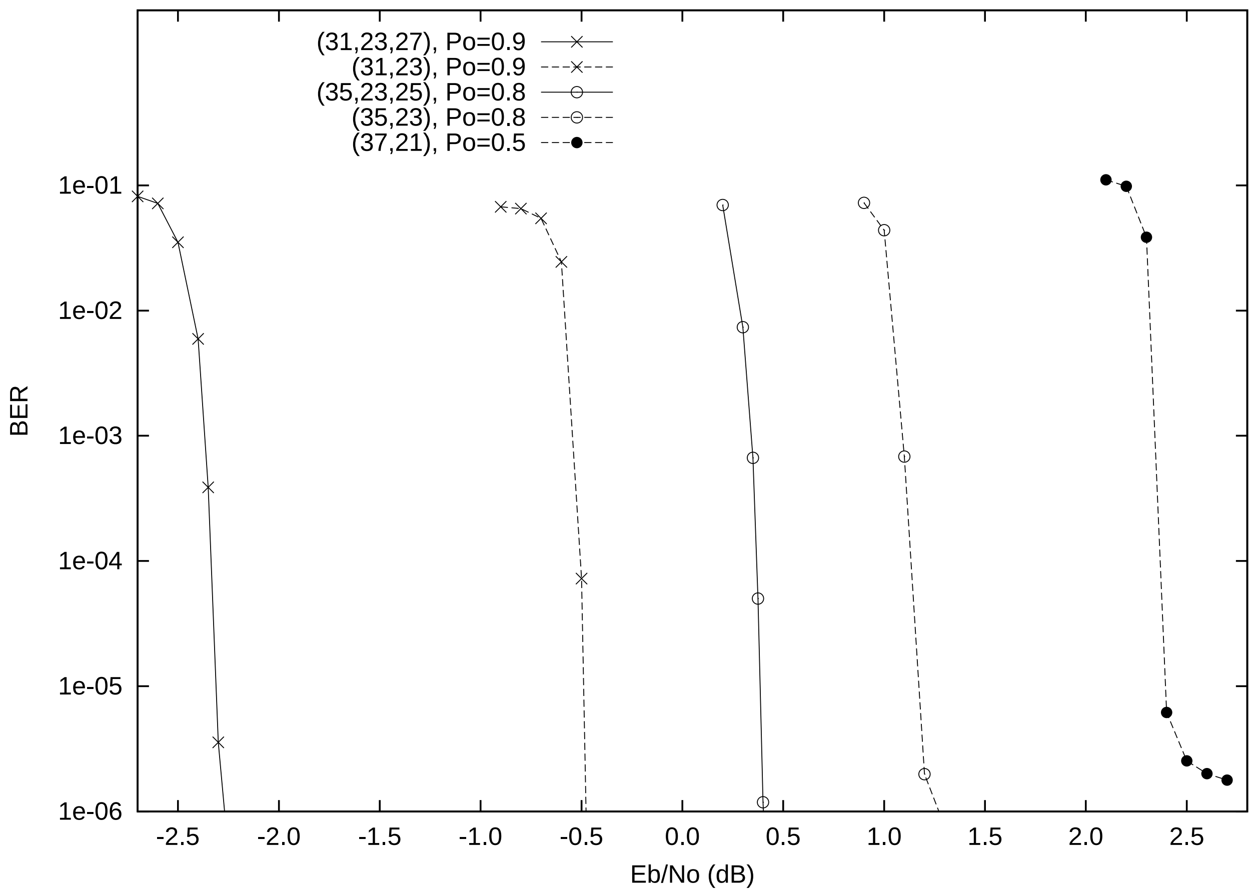

Figure 7: Turbo codes for non-uniform memoryless sources, R_c=1/2, N=262144, Rayleigh fading channel.

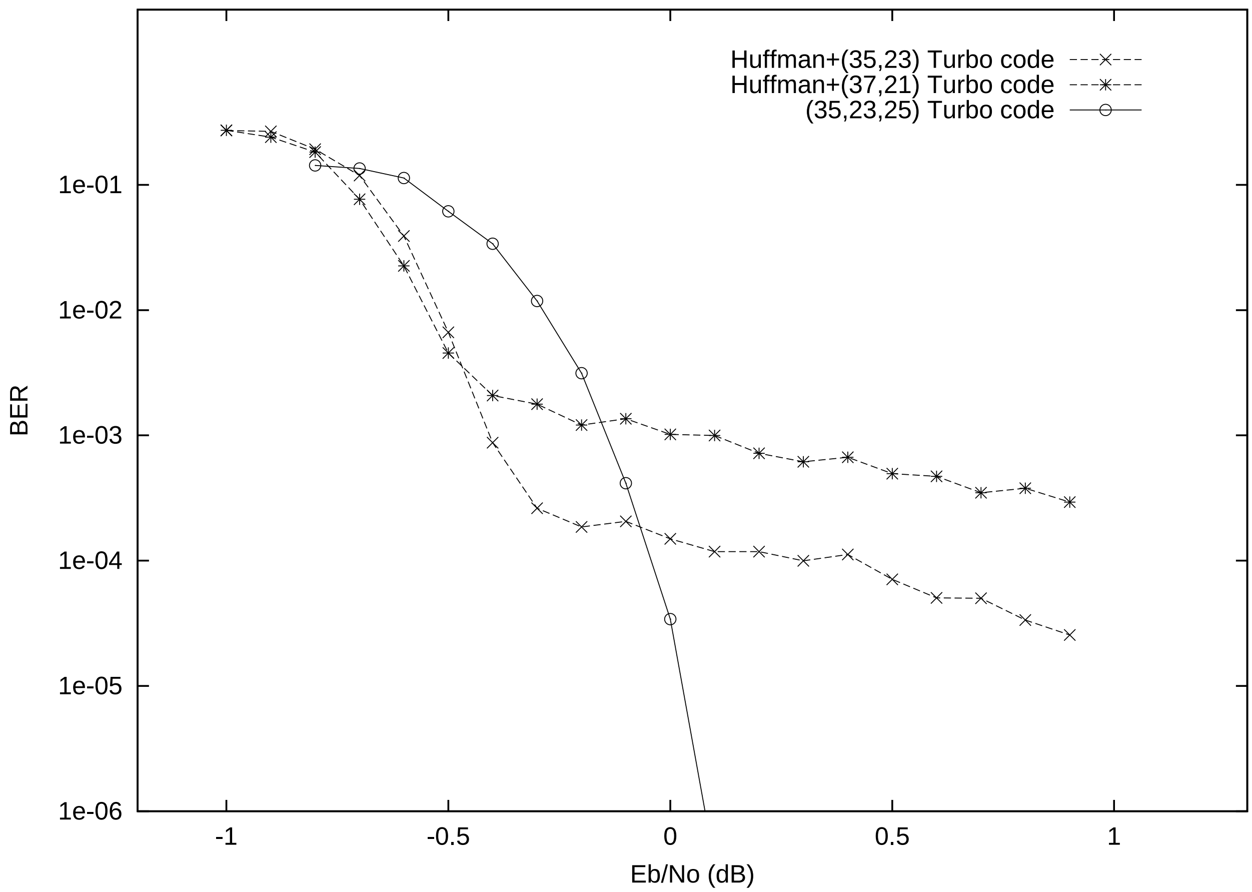

Figure 8: Performance comparison of our system ($R_s = 1, R_c = 1/2$) with that of tandem schemes ($R_s = 2/3$, $R_c = 1/3$), p_0=0.83079, N=12000, Rayleigh fading channel.

Channel	CSNR	C	$p_0 = 0.8$		$p_0 = 0.9$	
			C^{bias}	$\frac{C^{bias}}{C}$	C^{bias}	$\frac{C^{bias}}{C}$
AWGN	-4.0	0.415	0.284	68.5%	0.172	41.4%
	-2.0	0.564	0.392	69.6%	0.242	42.9%
	0.0	0.721	0.509	70.6%	0.320	44.4%
	2.0	0.860	0.614	71.4%	0.392	45.6%
	4.0	0.951	0.684	71.9%	0.442	46.4%
Rayleigh	-4.0	0.348	0.240	69.0%	0.147	42.2%
	-2.0	0.454	0.317	69.8%	0.197	43.3%
	0.0	0.566	0.398	70.4%	0.250	44.2%
	2.0	0.671	0.476	71.0%	0.302	45.0%
	4.0	0.763	0.544	71.3%	0.348	45.6%

Table 1: Illustration of channel input mismatch for AWGN and Rayleigh fading channels: C vs. C^{bias} for p_0=0.8, 0.9.

Channel	R_c	p_0	OPTA	STC OPTA gaps [29]	NSTC OPTA gaps
AWGN channel	1/3	0.8	-2.24	1.19	0.74
		0.9	-4.40	2.02	1.13
	1/2	0.8	-1.81	1.56	0.87
		0.9	-4.14	2.61	1.05
Rayleigh channel	1/3	0.8	-1.56	1.28	0.88
		0.9	-3.96	2.18	1.17
	1/2	0.8	-0.73	1.88	1.11
		0.9	-3.47	2.99	1.15

Table 2: OPTA gaps for systematic Turbo codes (STC) and non-systematic Turbo codes (NSTC) in E_b/N_0 (dB) at BER=10^{-5}.