Transmission of Non-Uniform Memoryless Sources over Wireless Fading Channels via Non-Systematic Turbo Codes[1]

Guang-Chong Zhu, Fady Alajaji
Dept of Math & Stats
Queen's University
Kingston, Ontario,
Canada, K7L 3N6
email: {zhugc, fady}@mast.queensu.ca

Jan Bajcsy and Patrick Mitran
Dept of ECE
McGill University
Montreal, Quebec,
Canada, H3A 2A7
email: {jbajcsy, pmitran}@tsp.ece.mcgill.ca

ABSTRACT

In this work, we investigate the joint source-channel coding problem of transmitting non-uniform memoryless sources over wireless Rayleigh fading channels via Turbo codes. The source redundancy in the form of non-uniformity is exploited in the Turbo decoder by incorporating the source statistics in the modified extrinsic information. In contrast to previous work, non-systematic recursive convolutional encoders are proposed as the constituent encoders, as they produce almost uniform outputs irrespective of the degree of non-uniformity in the source. As a result, unlike the outputs of systematic encoders, they 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 achieved over previously designed systematic Turbo codes, and the gaps to the optimal Shannon limit are therefore significantly reduced. In comparison with a tandem scheme, our system offers robust and superior performance at low BER levels (below 10^{-4}), while the complexity is lower.

KEY WORDS

Non-systematic Turbo codes, non-uniform memoryless sources, joint source-channel coding, Rayleigh fading channels, Shannon limit.

1 Introduction and Motivation

Ideally, in channel coding, the source is assumed to be uniform memoryless, i.e., the source generates independent and identically distributed (i.i.d.) bit streams $\{D_k\}_{k=1}^{\infty}$, where $Pr\{D_k = 0\} = Pr\{D_k = 1\} = 1/2$. In reality, however, substantial amount of redundancy is often observed 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 the *a priori* probability $p_0 \stackrel{\triangle}{=} Pr\{D_k = 0\} = 0.97$. In this case, a source encoder would then be used. A source encoder is said to be *optimal* if it can eliminate all the source redundancy, hence generating uniform memoryless outputs. However, most existing source encoders are only suboptimal (particularly fixed-length encoders that are commonly used for transmission over noisy channels); therefore, the source encoder output contains a certain amount of *residual* redundancy. For example, the 4.8 kbits/s US Federal Standard 1016 CELP speech vocoder produces line spectral parameters that contain 41.5% of residual redundancy due to non-uniformity and memory [3]. Therefore, the reliable communication of sources with a considerable amount of residual or natural redundancy is an important issue. Several studies (e.g., [3, 4, 5, 6, 7, 8], etc.) have shown that appropriate use of the source redundancy can significantly improve the system performance.

Turbo codes [9] are considered among the most exciting breakthroughs in channel coding due to their excellent performance which was initially demonstrated for uniform memoryless sources sent over additive white Gaussian noise (AWGN) channels. Later, the work was extended by Hall and Wilson for Rayleigh fading channels [10]. In [11] the authors considered using Turbo codes for sources with memory over AWGN channels. However, to the best of our knowledge, the issue of designing Turbo codes for non-uniform memoryless sources has not been fully studied, except for the recent work in [12, 13] where the source redundancy in the form of non-uniformity is exploited in the Turbo decoder via a modified extrinsic information term. Furthermore, the encoder structure is optimized in accordance with the source distribution. Significant coding gains are achieved by combining this optimized encoder structure with the appropriately modified decoder. Also, the performance results are compared to the *Shannon limit*, also known as the *optimum performance theoretically achievable (OPTA)*.

In wireless communications, the Rayleigh fading channel is commonly used to describe the statistical time varying nature of the received envelope of a flat fading signal, or the envelope of an individual multipath component. In this work, we investigate the design of joint source-channel Turbo codes that are suitable for transmitting non-

[1]This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Premier's Research Excellence Award (PREA) of Ontario.

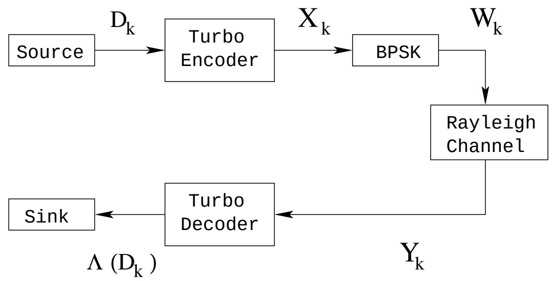

Figure 1. Block diagram of the system model.

uniform memoryless sources over Rayleigh fading channels. Instead of compressing the non-uniform source and channel coding it via standard source and channel codes, the two operations are combined into one via an appropriately designed joint source-channel Turbo codes. Our objective is to strive to come as close to the Shannon limit as possible. In [12, 13], the Turbo encoders used are recursive *systematic* convolutional (RSC) encoders. Although the gains achieved in these works are significant, the OPTA gaps (the difference between the system performance and the optimal Shannon limit performance) for heavily biased sources are still relatively big. Furthermore, when p_0 increases, the gaps became wider. Analysis on the encoder output reveals that the drawback lies in the *systematic* structure. Due to the feedback, recursive non-systematic convolutional (RNSC) encoders can generate almost uniformly distributed output even for very biased sources. From information theory [14, 15], we know that the capacity of a binary input Rayleigh fading channel is maximized by a uniform memoryless channel input; therefore, we propose using RNSC encoders as the constituent Turbo encoders. Simulation results demonstrate substantial gains over systematic Turbo codes and the OPTA gaps are significantly reduced. A similar study for non-systematic Turbo codes over AWGN channels was recently given in [16]. A performance comparison of our joint source-channel system with the traditional tandem system, where the source and channel codes are separately designed, is also presented.

2 System Model

The block diagram of the system we are considering is depicted in Fig. 1. The source generates a non-uniform memoryless bitstream $\{D_k\}_{k=1}^{\infty}$, where $p_0 = Pr\{D_k = 0\} \neq 1/2$. The data sequence is Turbo encoded and then binary phase-shift keying (BPSK) modulated; after being transmitted through the Rayleigh fading channel, the sequence is fed into the Turbo decoder, which iteratively computes the log-likelihood ratio (LLR) $\Lambda(D_k)$ of each bit. The channel model considered is fully interleaved Rayleigh fading channel, which is described by

$$Y_k = A_k W_k + N_k, \qquad k = 1, 2, 3, \cdots,$$

where $W_k \in \{-1, +1\}$ is the BPSK signal of unit energy and $\{N_k\}$ is an i.i.d. Gaussian noise sequence with zero mean and variance $N_0/2$. The amplitude fading process $\{A_k\}$, also known as the channel state information, is assumed to be i.i.d. (via full channel interleaving) and Rayleigh distributed. We assume that $\{A_k\}$ is known at the decoder, and that A_k, W_k, and N_k are independent of each other.

3 Non-Systematic Turbo Codes

Turbo codes use two (or more) simple convolutional encoders in parallel concatenation linked by an interleaver; in the decoder, constituent decoders are placed in serial concatenation with an interleaver in between, and a deinterleaver is used in the feedback loop from the second constituent decoder to the first. Each constituent decoder employs the BCJR algorithm [17], and the decoding process is realized in an iterative fashion by exchanging the extrinsic information between the two constituent decoders. In the original work by Berrou *et al.* [9], extraordinary performance was demonstrated by using Turbo codes for uniform memoryless sources over AWGN channels.

Design of Turbo codes for non-uniform memoryless sources has been recently studied in [12, 13]. It is shown that with a modified Turbo decoder that exploits the source redundancy in the form of non-uniformity, Berrou's (37,21) code exhibits an obvious performance degradation when the source is non-uniform, while a source-dependent encoder optimization can significantly improve the performance. For example, for Rayleigh fading channels, when the rate R_c=1/3 and p_0=0.9, the optimization of the encoder yields a 1.08 dB gain over the Berrou (37,21) code; in comparison with a Turbo code that does not exploit the source redundancy, a source-optimized Turbo code gives an impressive 3.01 dB gain.

Despite the significant coding gains achieved in the above works, the performance can be further improved since the gaps to the Shannon limit are still relatively wide for heavily biased sources. Furthermore, when p_0 increases, the OPTA gaps become wider. For example, when R_c=1/2, for p_0=0.8 and 0.9, the OPTA gaps are 1.88 dB and 2.99 dB, respectively.

Note that in [12, 13], the encoders are *systematic*, which is commonly used in almost all the Turbo codes literature. When the source is heavily biased, this systematic structure would become a drawback. For example, when p_0=0.9, as part of the Turbo encoder outputs, the systematic sequence (which is identical to the original information sequence) contains many more 0's than 1's, which renders the codebook with a considerably small minimum Hamming distance.

If the encoder is non-recursive, when the source is heavily biased, the parity output would also be heavily biased. However, this is not the case when the encoder is recursive. Due to the feedback structure, the parity output

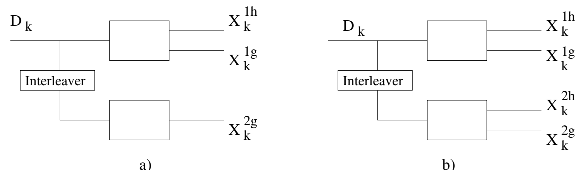

Figure 2. Non-systematic Turbo encoder structures.

can be almost uniformly distributed even for a very heavily biased source input.

We know that the capacity of a binary input Rayleigh fading channel is achieved when its mutual information is maximized by a uniformly distributed input; Furthermore, it has been shown in [18] that the 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. Thus, in our search for good codes, we should only consider codes whose empirical distributions are close to the capacity-achieving distributions. This implies that, if a non-systematic encoder is adopted in conjunction with a recursive structure, the above drawback can be resolved since the resulting joint source-channel Turbo code is more suitably matched to the channel, and therefore an improved performance is expected.

Fig. 2 shows our proposed non-systematic Turbo encoders. In a) the first constituent encoder has two parity outputs while the second has only one parity output; so the overall rate is 1/3. 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 by puncturing. Structure a) is virtually a special case of structure b) obtained by completely puncturing X_k^{2h}; therefore, a generally designed decoder for structure b) can also be used for structure a).

In [12, 13], the RSC encoders are optimized for a given source distribution by choosing the best feedback and feed-forward polynomials iteratively. For RNSC encoders, an exhaustive search for the best structure is computationally impractical. In our simulations, we fix the best feedback and feed-forward polynomials found in [12, 13], and search for the other best feed-forward polynomial.

4 Decoder Modifications

When RSC encoders are used as constituent encoders, the LLR in the BCJR algorithm [17] employed by the Turbo decoder can be decomposed into three terms [9]:

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

where $L_{ch}(D_k)$, $L_{ex}(D_k)$ and $L_{ap}(D_k)$ are the channel transition term, the extrinsic term and the *a priori* term, respectively.

When RNSC encoders are used as constituent encoders, $\Lambda(D_k)$ can only be decomposed into two terms:

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

where the new extrinsic term involves two parity sequences. Also, for Rayleigh fading channels, the extrinsic term needs to be modified to appropriately incorporate the channel statistics. The extrinsic term therefore becomes

$$L_{ex}(D_k) = \log \frac{\sum_{e,e'} \gamma(\mathbf{y}_k|1,e,e',\mathbf{a}_k) \cdot \alpha_{k-1}(e') \cdot \beta_k(e)}{\sum_{e,e'} \gamma(\mathbf{y}_k|0,e,e',\mathbf{a}_k) \cdot \alpha_{k-1}(e') \cdot \beta_k(e)},$$

where for $i = 0, 1$,

$$\begin{aligned}\gamma(\mathbf{y}_k|i,e,e',\mathbf{a}_k) \quad = \quad & p(y_k^h|D_k = i, E_k = e, a_k^h) \\ & \cdot p(y_k^g|D_k = i, E_k = e, a_k^g) \\ & \cdot Pr\{D_k = i|E_k = e, E_{k-1} = e'\},\end{aligned}$$

and where E_k is the encoder state at time k, $\mathbf{a}_k = (a_k^h, a_k^g)$ is the fading factor, $\mathbf{y}_k = (y_k^h, y_k^g)$ is the noise corrupted version of $\mathbf{x}_k = (x_k^h, x_k^g)$, which is the pair of parity bits generated from the two feed-forward polynomials. $\alpha_k(e)$ and $\beta_k(e)$ are defined and can be recursively computed as in [9]. Also, as in [12, 13], when the source is non-uniform i.i.d., $\log((1-p_0)/p_0)$ is used as the initial *a priori* input to the first decoder at the first iteration, then it can be verified via the BCJR algorithm's derivation that this term will appear in the output $\Lambda(D_k)$ as an extra term. In our design, we then use $L_{ex} + \log((1-p_0)/p_0)$ as the new extrinsic information for both constituent decoders at each iteration.

5 Simulation Results and Discussion

In this section, we present simulation results of our non-systematic Turbo codes for uniform memoryless sources over BPSK-modulated Rayleigh fading channels. The performance is measured in terms of bit error rate (BER) versus E_b/N_0, where E_b is the average energy per source bit and $N_0/2$ is the variance of the Gaussian additive noise process. All simulated Turbo codes have 16-state constituent encoders and use the same pseudo-random interleaver introduced in [9]. The sequence length is $N = 512 \times 512 = 262144$ and 200 blocks are used; this would guarantee a reliable BER estimation at the 10^{-5} level with 524 errors. The number of iterations used in the decoder is 20. All presented results are for Turbo codes with structure b) in Fig. 2 as they have a better performance than the codes with structure a) encoders. Simulations are performed for rates $R_c = 1/3$ and $R_c = 1/2$, and for p_0=0.8 and 0.9. From our simulations, for both rates 1/3 and 1/2, the best RNSC encoder structure found for p_0=0.8 has each constituent encoder with the feedback polynomial 35 and feed-forward polynomials 23 and 25, denoted by (35,23,25); for p_0=0.9 the best structure is (31,23,27). Several other encoders give very competitive performance; for example, (35,23,21) and (35,23,31) for p_0=0.8, (31,23,35)

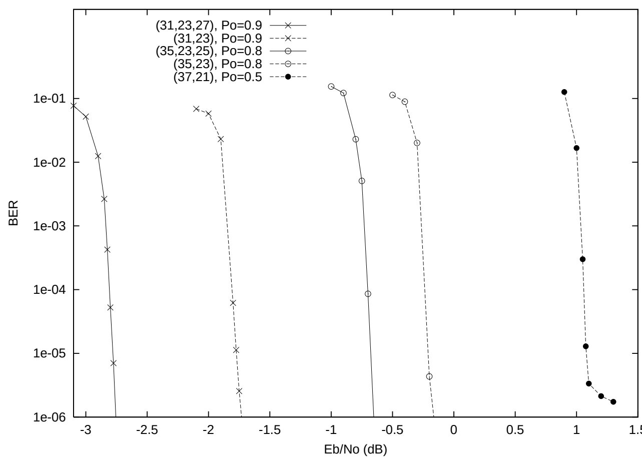

Figure 3. Turbo codes for non-uniform memoryless sources, R_c=1/3, N=262144, Rayleigh fading channel.

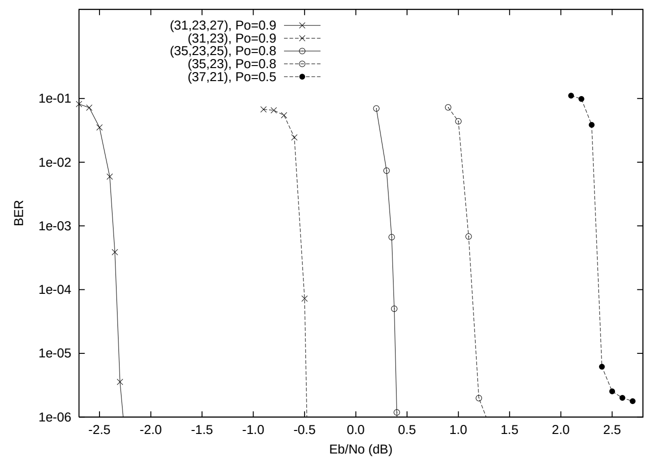

Figure 4. Turbo codes for non-uniform memoryless sources, R_c=1/2, N=262144, Rayleigh fading channel.

Source distribution	$R_c = 1/2$	$R_c = 1/3$
$p_0 = 0.8$	-0.73	-1.56
$p_0 = 0.9$	-3.47	-3.96

Table 1. OPTA values in E_b/N_0 at BER=10^{-5} level (in dB), Rayleigh fading channel.

Source	Systematic [12, 13]		Non-Systematic	
distribution	$R_c = 1/2$	$R_c = 1/3$	$R_c = 1/2$	$R_c = 1/3$
$p_0 = 0.8$	1.88	1.28	1.11	0.88
$p_0 = 0.9$	2.99	2.18	1.15	1.17

Table 2. OPTA gaps in E_b/N_0 at BER=10^{-5} level (in dB), Rayleigh fading channel.

and (31,23,37) for p_0=0.9 also give good performance very close to those offered by the above encoders.

Fig. 3 shows the performances of our rate-1/3 non-systematic Turbo codes in comparison with their systematic peers investigated in [12, 13], as well as with Berrou's (37,21) code, which offers the best water-fall performance (among 16-state encoders) for uniform memoryless sources. At the 10^{-5} BER level, when p_0=0.8, our (35,23,25) non-systematic Turbo code offers a 0.40 dB gain over its (35,23) systematic peer, which brings the performance only 0.88 dB away from OPTA; when p_0=0.9, the improvement is 1.01 dB with the encoder structure (31,23,27), which narrows the OPTA gap from 2.18 dB down to 1.17 dB. In comparison with Berrou's (37,21) code performance, the gains achieved by exploiting the source redundancy and encoder optimization are therefore 1.76 dB and 3.87 dB for p_0=0.8 and 0.9, respectively.

Fig. 4 shows similar results for rate-1/2 and the gains are generally more significant. In comparison with the best systematic Turbo code performances, at the 10^{-5} BER level, for p_0=0.8 and 0.9, the gains achieved are 0.77 dB and 1.84 dB, respectively; the OPTA gaps are therefore reduced to 1.11 dB and 1.15 dB. Furthermore, the gains due to combining the optimized encoder with the modified decoder that exploits the source redundancy are 2.01 dB (p_0=0.8) and 4.71 dB (p_0=0.9). The OPTA values and the OPTA gaps, which are computed as outlined in [12, 13], are provided in Table 1 and 2.

To achieve a desired rate by puncturing, using different puncturing patterns may result in a difference in the performance. 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 that the best puncturing pattern is to keep the parity sequence generated from feed-forward 23 intact and puncture half of the one generated from the other feed-forward polynomial. The performance of this puncturing pattern is about 0.1 to 0.2 dB better than other attempted 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 from feed-forward 23, and delete all odd (even) position bits of the sequences generated from the other feed-forward polynomial.

6 Comparison with Tandem Scheme

Traditionally, source and channel coding are usually implemented separately, resulting in a so-called *tandem* scheme.

That is, the source is compressed first, and then channel-coded. In this section, we compare the performance of our joint source-channel system with that of a tandem scheme for the same overall transmission rate. The 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 systems is r=1/2 source symbol/channel symbol; 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 probability 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 can find that when p_0=0.83079, a 4^{th}-order Huffman encoder would produce an average length/source bit of 0.666668 code bits/source symbol. Therefore, in this section, simulations are performed for this value of p_0.

Berrou's pseudo-random interleaver [9] requires that the sequence length has to be an even power of 2; this inflexibility becomes an obstacle in the design of the tandem scheme, since the Huffman code is a variable-length code. The S-random interleaver [19], however, can take an input sequence with arbitrary length and yields good BER performance; therefore, in this section we adopt the S-random interleaver in the Turbo encoder and decoder. Another small modification to the Turbo code is that the first constituent encoder is not terminated; 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 adopts the S-random interleaver, and also with the first constituent encoder 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. On the other hand, what matters is not only the number of errors in the Turbo-decoded sequence, but also the positions of the erroneous bits. Considering how Huffman decoding is performed, we can see that an error among the first few bits in an input sequence is much more disastrous than an error in the tail bits. Therefore, a sufficiently large number of blocks is necessary to obtain a good average performance. Considering computation limitations, after several trials, we choose to use N=12000, and 60000 blocks are used. The number of iterations in the Turbo decoder is also 20.

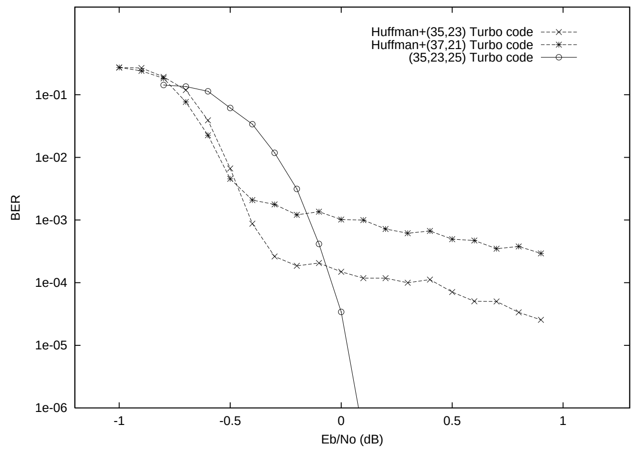

Figure 5. Performance comparison of our system ($R_s = 1, R_c = 1/2$) with that of tandem schemes ($R_s = 2/3, R_c = 1/3$), p_0=0.83079, N=12000, Rayleigh fading channel.

For a given sequence length, the larger we choose the "spread" S of the S-random interleaver, the better performance we can achieve. However, in practice, generating an S-random interleaver with a large spread requires a substantial amount of computation time, and sometimes such an S-random interleaver may not be generated successfully. Therefore, in order to reduce the computation time and also to guarantee the successful generation of S-random interleavers of arbitrary size, the spread S is chosen to be only 10. Furthermore, a newly designed S-random interleaver generator [20] is adopted, which is significantly faster than the original one in [19].

Figure 5 shows the performance comparison of our system with two tandem schemes. The comparison is made in terms of $E_b/N_0 = E_s/(rN_0)$, where $r = R_c/R_s$, and E_s is the average energy per channel symbol. In the first tandem scheme, the rate R_c=1/3 Turbo code we simulated is Berrou's (37,21) code, which offers the best 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 provide simulation results of 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 BER levels, both tandem schemes offer better performance than that of our joint source-channel coding system, their error-floor occur at high BER levels (10^{-3} for the (37,21) code, and 10^{-4} for the (35,23) code). Therefore, at low

BER levels, our joint source-channel coding system offers superior performance than both tandem schemes; it also has lower complexity than the tandem scheme.

7 Conclusions

In this work, the joint source-channel coding issue of transmitting non-uniform memoryless sources via Turbo codes over Rayleigh fading channels is investigated. Recursive non-systematic Turbo codes are proposed for the considered sources, since their output is almost uniformly distributed for even heavily biased sources. It is thus suitably matched to the channel input as it nearly maximizes the channel mutual information. Simulation results show substantial coding gains (up to 1.84 dB) achieved in comparison with systematic Turbo codes designed in [12, 13], and the OPTA gaps are significantly reduced. In comparison with the tandem scheme, our system has lower complexity, offers substantially better performance at low BER levels, and is more robust to channel errors.

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 and Audio Processing*, Vol. 4, pp. 325–336, Sept. 1996.

[4] F. Alajaji, S. A. Al-Semari and P. Burlina, "Visual communication via trellis coding and transmission energy allocation," *IEEE Trans. Commun.*, Vol. 47, pp. 1722–1728, Nov. 1999.

[5] S. A. Al-Semari, F. Alajaji and T. Fuja, "Sequence MAP decoding of trellis codes for Gaussian and Rayleigh channels," *IEEE Trans. Veh. Technol.*, Vol. 48, pp. 1130–1140, Jul. 1999.

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

[7] 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.

[8] 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, TX, Jun. 1996.

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

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

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

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

[13] 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.

[14] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley series in Telecommunications, 1991.

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

[16] G.-C. Zhu, F. Alajaji, J. Bajcsy and P. Mitran, "Non-systematic Turbo codes for non-uniform i.i.d. sources over AWGN channels," *Proc. 2002 Conf. on Inform. Sciences and Systems*, Princeton, NJ, Mar. 2002.

[17] 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. IT-20, pp. 248-287, Mar. 1974.

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

[19] 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.

[20] P. Mitran, "Design and applications of Turbo source codes," M.Eng. thesis (preprint), McGill University, July 2002.