Design of Turbo Codes for Non-Equiprobable Memoryless Sources*

Guang-Chong Zhu and Fady Alajaji

Mathematics and Engineering

Department of Mathematics and Statistics

Queen's University, Kingston, Ontario, Canada, K7L 3N6

Tel: (613) 533-2423, Fax: (613) 533-2964

E-mail: fady@shannon.mast.queensu.ca

zhugc@shannon.mast.queensu.ca

Abstract

This work addresses the problem of designing Turbo codes for non-uniform binary memoryless or independent and identically distributed (i.i.d.) sources over noisy channels. The extrinsic information in the decoder is modified to exploit the source redundancy in the form of non-uniformity; furthermore, the constituent encoder structure is optimized for the considered non-uniform i.i.d. source to further enhance the system performance. Some constituent encoders are found to substantially outperform Berrou's $(37, 21)$ encoder. Indeed, it is shown that the bit error rate (BER) performance of the newly designed Turbo codes is greatly improved as significant coding gains are obtained. Comparisons to the optimal Shannon limit are also provided.

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

1 Introduction

In channel coding, the source is usually assumed to be a binary memoryless Bernoulli (1/2) source, i.e., it generates uniform i.i.d. bit streams $\{d_k\}_{k=1}^{\infty}$, where $Pr\{d_k = 0\} = Pr\{d_k = 1\} = 1/2$. In reality, however, natural sources (e.g., image and speech sources) often exhibit substantial amounts of redundancy in the form of memory and/or non-uniformity [2]; in this case, a source encoder would be used. An ideal source encoder would be able to eliminate all the source redundancy and hence produce a uniform i.i.d. sequence of bits, which would then be used as the input of the channel encoder. However, most existing source encoders are suboptimal. As a result, the input to the channel encoder contains a certain amount of *residual* redundancy. For example, the 4.8 kbits/s US Federal Standard 1016 CELP speech vocoder produces an output that contains 41.5% of residual redundancy due to non-uniformity and memory [3]. For uncompressed sources,

* This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

it has been observed that many binary images (e.g., facsimile and medical images) may contain as much as 80% of redundancy in the form of non-uniformity (e.g., [7], [14]), this corresponds to the *a priori* probability $Pr\{d_k = 0\} = 0.97$. Therefore, transmission of sources with a considerable amount of residual or natural redundancy is an important practical issue. Several studies (e.g., [1], [4], [11], [16] and [19], etc.) have shown that appropriate use of the source redundancy can significantly improve the system performance.

Turbo codes [6] have demonstrated excellent performance for uniform i.i.d. sources over additive white Gaussian noise (AWGN) channels; to the best of our knowledge, the issue of using Turbo codes for non-uniform i.i.d. sources has not been systematically studied. In essence, this is a joint source-channel coding problem. In this work, we investigate the issue of designing Turbo codes for non-uniform i.i.d. sources sent over noisy channels. Our objective is to strive to come as close to the Shannon limit as possible. The extrinsic information in the Turbo decoder is modified to take advantage of the source redundancy (this simple method was briefly mentioned in [13] and [9]; however, its performance was not explicitly studied, particularly *vis-a-vis* the Shannon limit); as a result, the BER performance of the Turbo coded system is significantly enhanced. We also observe that while the original Berrou Turbo code which uses the $(37, 21)$ convolutional code in each constituent encoder offers extraordinary performance (excellent "waterfall" region at very low SNR's) for uniform i.i.d. sources, it provides a relatively poor performance with a considerably high error floor when the source is non-uniform. An analysis of the encoder's structure reveals to us that it is important to find more suitable constituent encoders for non-uniform i.i.d. sources. Through a systematic search we performed by simulations, we show that some constituent encoders substantially outperform Berrou's $(37, 21)$ encoders for a given non-uniform i.i.d. source. Significant coding gains are further achieved by combining this optimized encoder structure with the appropriately modified decoder that exploits the source non-uniformity.

2 Turbo Codes for Non-uniform I.I.D. Sources

A non-uniform i.i.d. source is described by the non-equiprobable probability distribution of the bit stream. The source emits a sequence of bits $\{d_k\}_{k=1}^{\infty}$ with probability $Pr\{d_k = 0\} = p_0$, $k = 1, 2, \cdots$.

Turbo codes [6] 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 [5], and the decoding process is realized in an iterative fashion by exchanging the extrinsic information between the two constituent decoders. Extraordinary BER performance has been demonstrated by using Turbo codes for uniform i.i.d. sources (with p_0=1/2) over AWGN channels [6] and Rayleigh fading channels [10].

In the following, we consider the problem of designing Turbo codes for the transmission of non-uniform i.i.d. binary sources over AWGN and Rayleigh fading channels. We propose some modifications for the Turbo decoder in order to take advantage of the source redundancy in the form of non-uniformity. We also optimize the Turbo encoder structure with respect to the considered non-uniform i.i.d. source.

2.1 Modifications of the Decoder Extrinsic Information

In the BCJR algorithm used by the Turbo decoder, we observe that the log-likelihood ratio (LLR) produced by the Turbo decoder can be decomposed into three terms:

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

where $L_{ch}(d_k)$, $L_{ex}(d_k)$, and $L_{ap}(d_k)$ are the channel transition term, the extrinsic information term, and the *a priori* term, respectively, [6], [12]. The extrinsic information produced by one constituent decoder is used as the *a priori* estimation for the other constituent decoder. At the first iteration, for the first constituent decoder, if the source is uniform i.i.d., which is the case investigated in [6], we have

$$L_{ap}(d_k) = \log \frac{P(d_k = 1)}{P(d_k = 0)} = 0,$$

since $P(d_k = 1) = P(d_k = 0) = 1/2$.

When the source is non-uniform i.i.d., $\log((1 - p_0)/p_0) \neq 0$ is used as the initial *a priori* input for the first decoder in the first iteration. As a result, in the output $\Lambda(d_k)$ produced by the first decoder, $L_{ap}(d_k) = \log((1 - p_0)/p_0)$. By passing this term together with the extrinsic information from the first decoder to the second decoder, a BER performance gain is observed[1]. It can be shown via the BCJR algorithm's derivation that $\log((1 - p_0)/p_0)$ 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 decoders at each iteration. With this simple procedure, the performance is greatly improved.

2.2 Optimizing the Encoder Structure

In the original design of Turbo codes, Berrou *et al.* used a 16-state $(37, 21)$ recursive systematic convolutional (RSC) code in both constituent encoders. From our simulations, we found that a Turbo code using Berrou's encoder performs poorly for non-uniform i.i.d. sources over a wide range of E_b/N_0 values, where E_b is the average energy per information bit, and $N_0/2$ is the additive noise variance. Analysis of the tap coefficients reveals that for non-uniform i.i.d. sources, especially when their probability distributions are heavily biased, many possible states of this encoder may rarely or never be reached. For example, when $p_0 = 0.9$, suppose the $(37, 21)$ encoder starts at the all zero state 0000, where each digit represents the content of each shift register, then the encoder would remain in this state until a "1" arrives, which would cause a transition to state 1000. Since $p_0 = 0.9$, with high probability, a state transition would occur among very limited number of states, such as 1000, 1100, 0110, 0011, and 0001, etc. Furthermore, the parity sequence generated by this encoder has low weight. As a result, these drawbacks would cause performance degradations in the decoder. Therefore, for non-uniform i.i.d. sources, it is more important to find appropriate encoder structures that can overcome the above mentioned problems and hence offer better BER performances.

[1] This simple modification of appropriately using the source information in the Turbo decoder for non-uniform sources was also briefly mentioned in [13] (see Remark 4.d on page 433) and [9] but not explicitly studied and evaluated. One of our goals in this paper is to explicitly assess the gains achieved by this method and examine how close we can come to the Shannon limit.

With this motivation, we performed a systematic search for better constituent encoders for a given non-uniform probability distribution of the source. In our simulations, we only focused on 16-state encoders. Denote the coefficients of the feedback and feed-forward polynomials of a 16-state RSC encoder in binary form as $\{f_0, f_1, f_2, f_3, f_4\}$ and $\{g_0, g_1, g_2, g_3, g_4\}$, respectively, where $f_i, g_j = 0$ or 1, $i, j = 0, 1, \cdots, 4$. Altogether there are $2^4 \times (2^5 - 1) = 496$ possible combinations; therefore, an exhaustive search is impractical. Instead, for a given non-uniform i.i.d. source, we choose to determine the sub-optimal encoder among those with $f_0 = f_4 = g_0 = g_4 = 1$. The total number of such encoders is $2^3 \times 2^3 = 64$. Again, to avoid an exhaustive search, the sub-optimal encoders are obtained through the following iterative steps:

1) Fix the feed-forward polynomial, (e.g., $\{g_0, g_1, g_2, g_3, g_4\} = \{1, 1, 1, 1, 1\}$), find (by simulation) the best feedback polynomial among the remaining possible choices;

2) Fix the feedback polynomial as the one found in step 1), find the best feed-forward polynomial among all remaining possible choices;

3) Fix the feed-forward polynomial as found in step 2), go back to step 1), if the feedback polynomial coincides with the one obtained in step 1), stop (otherwise, proceed to the next step).

Via this procedure, for a given non-uniform probability distribution, several encoders were found to outperform Berrou's $(37, 21)$ encoder significantly; they also have considerably lower error floors. Among them, the $(35, 23)$ encoder gives the best performance for $p_0 = 0.7$ and 0.8; when $p_0 = 0.9$, the best encoder is $(31, 23)$. However, when the source is uniform i.i.d, Berrou's $(37, 21)$ encoder gives a better performance in the water-fall region than the above encoders, though its error floor is higher. For uniform sources, improving the error floor region of Turbo codes is usually achieved at the expense of the waterfall region; however, for the non-uniform case, this tendency seems to decrease as p_0 increases.

3 Simulation Results and Discussions

In this section, we present simulation results of Turbo codes for non-uniform i.i.d. sources over BPSK-modulated AWGN and Rayleigh fading channels. All simulated Turbo codes use the same pseudo-random interleaver introduced in [6]. 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. Simulations for our selected codes and the Berrou code are performed for rates $R_c = 1/3$ and $R_c = 1/2$, for p_0=0.5, 0.7, 0.8 and 0.9. To show the performance gains due to using the modified extrinsic information, simulations are also performed by using the unchanged extrinsic information; i.e., the decoder has no knowledge of the probability distribution and assumes the source is uniform[2].

[2] In this case, regardless of the source generated at the encoder ($p_0 = 0.9$, 0.8, 0.7 or 0.5), the performance of the system with unchanged extrinsic information at the decoder varies very slightly with p_0; so for this system, we can assume that $p_0 = 0.5$ at both the encoder and the decoder.

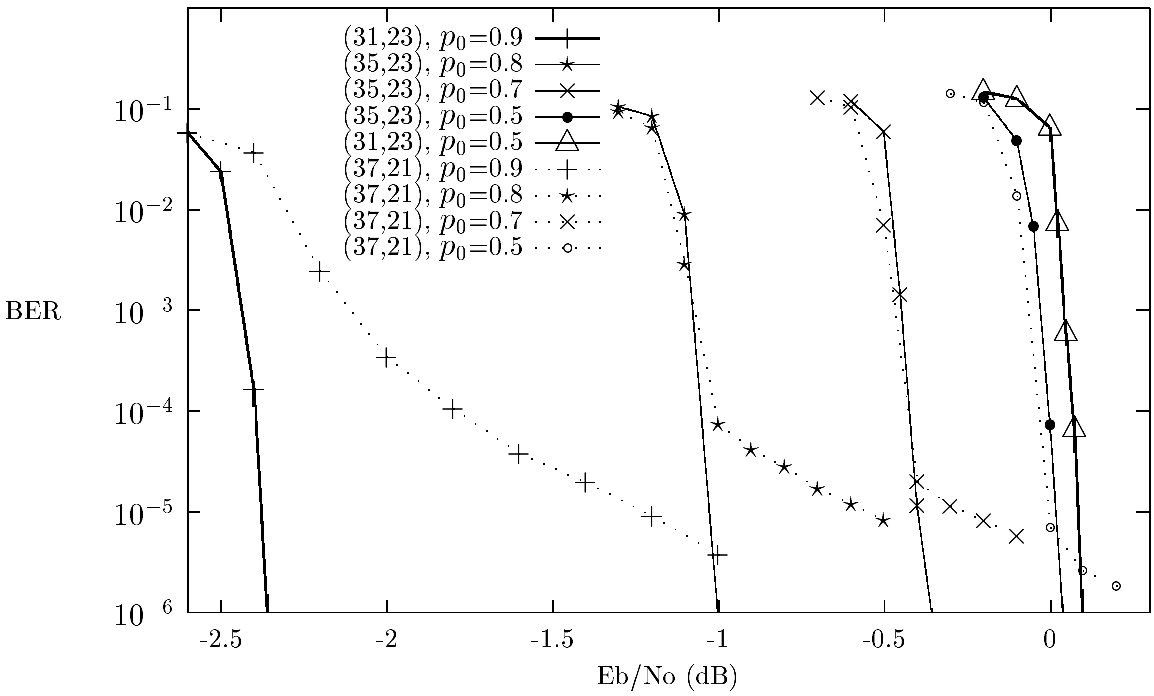

Figure 1: Turbo codes for non-uniform i.i.d. sources, R_c=1/3, N=262144, AWGN channel.

3.1 Performance Evaluations

Figure 1 shows the performance comparison of Berrou's rate-1/3 $(37, 21)$ Turbo code and our selected Turbo codes for transmitting uniform and non-uniform i.i.d. sources (with p_0=0.5, 0.7, 0.8 and 0.9) over AWGN channels by using the modified extrinsic information in the decoder. The $(35, 23)$ code offers the best performance from p_0=0.7 up to 0.8; when $p_0 = 0.9$, the $(31, 23)$ seems to be the best choice. At the 10^{-5} BER level, the gains over the Berrou $(37, 21)$ code due to optimization of the encoder for p_0=0.7, 0.8 and 0.9 are 0.09, 0.51 and 1.18 dB, respectively. Furthermore, the gains due to using the modified extrinsic information for p_0=0.7 and 0.8 (with the $(35, 23)$ code) are 0.43 dB and 1.08 dB, respectively; for p_0=0.9 (with the $(31, 23)$ code) it is 2.46 dB. Similar results can be observed in Figure 2, in which the rate is 1/2. For example, when p_0=0.9, our selected $(31, 23)$ code gives a gain of 1.08 dB over the Berrou $(37, 21)$ code at the 10^{-5} BER level, while the gain achieved due to exploiting the source redundancy in the form of non-uniformity is 2.20 dB.

In Figure 3, we provide the performance comparison for the Rayleigh fading channel with known channel state information. When p_0=0.7, the gain due to encoder optimization is visible only below the 10^{-5} BER level, because the performance offered by Berrou's code has lower error floor than that in the AWGN channel case. When p_0=0.8, our $(35, 23)$ Turbo code offers a 0.33 dB gain at the 10^{-5} BER level over its $(37, 21)$ peer; when p_0=0.9, the gain is further increased up to 1.08 dB by using the $(31, 23)$ encoder. Furthermore, the gains due to exploiting the source redundancy in the form of non-uniformity become bigger than those obtained in the AWGN channel case: when p_0=0.7 and 0.8, the gains at a BER of 10^{-5} produced by the $(35, 23)$ Turbo code are 0.54 dB and 1.36 dB, respectively; when p_0=0.9, the $(31, 23)$ Turbo code realizes a significant

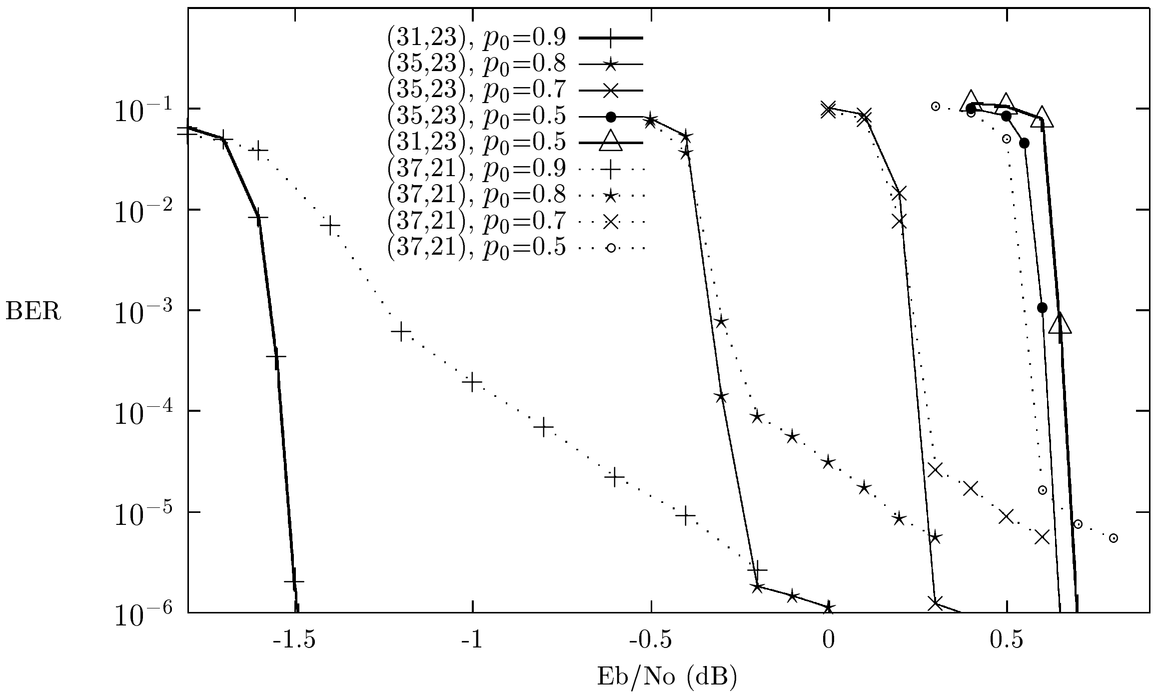

Figure 2: Turbo codes for non-uniform i.i.d. sources, R_c=1/2, N=262144, AWGN channel.

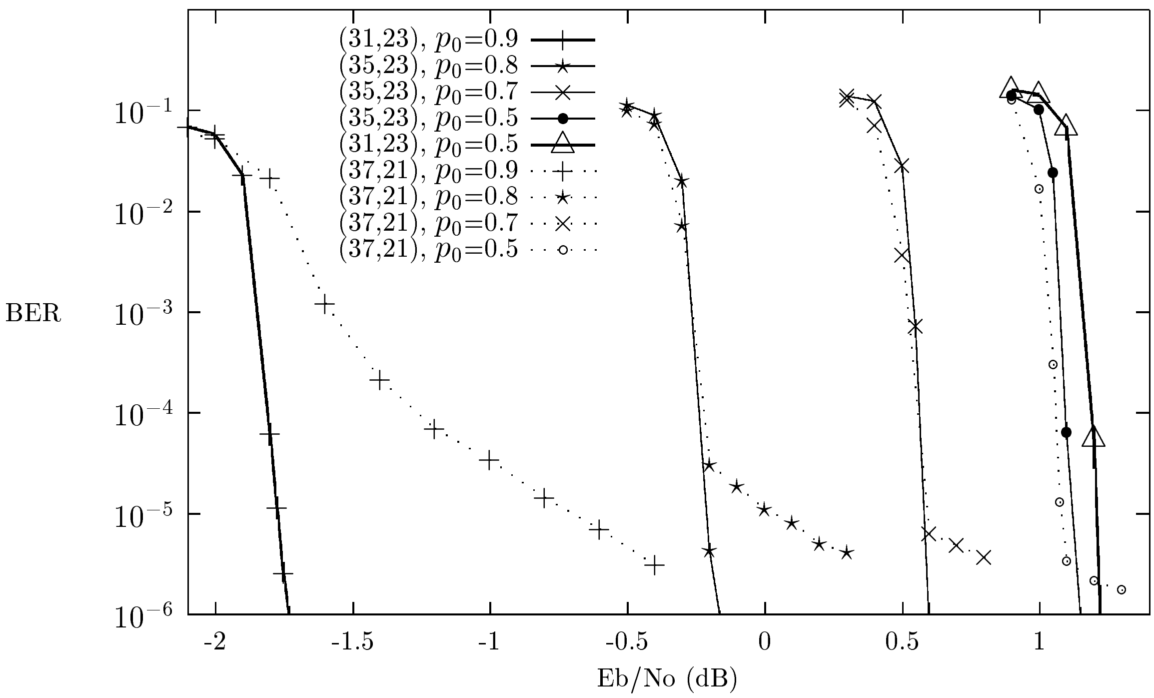

Figure 3: Turbo codes for non-uniform i.i.d. sources, R_c=1/3, N=262144, Rayleigh fading channel.

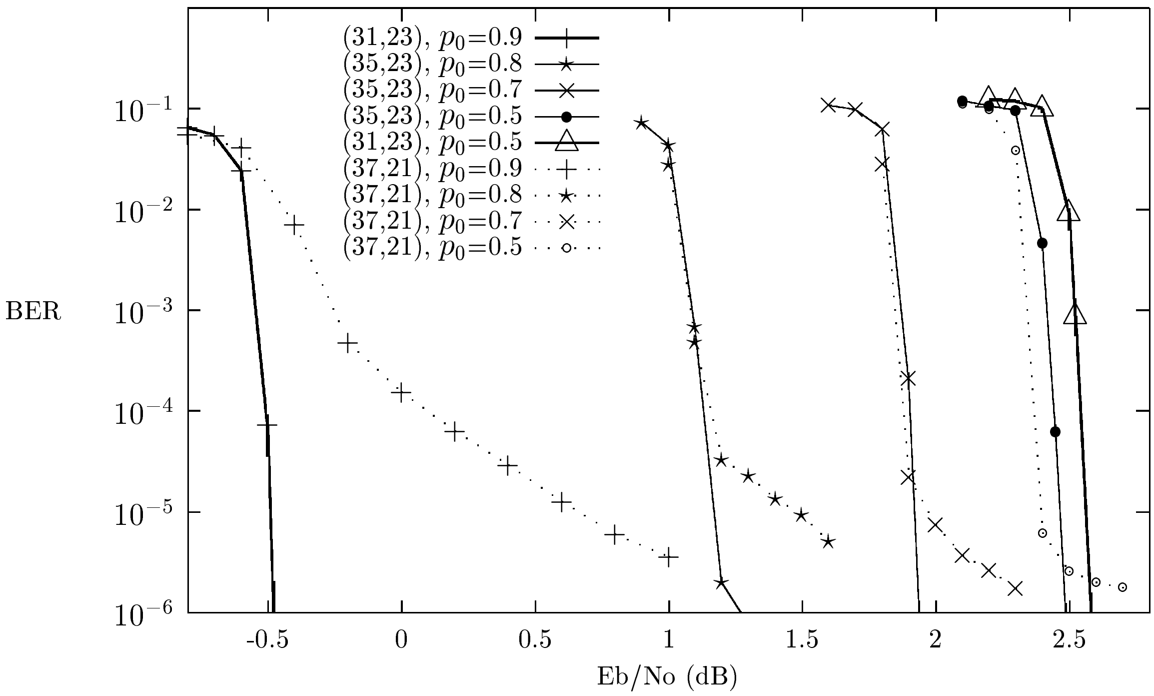

Figure 4: Turbo codes for non-uniform i.i.d. sources, R_c=1/2, N=262144, Rayleigh fading channel.

3.01 dB gain. Similar performances for rate-1/2 are shown in Figure 4. For example, at the same BER level, and for $p_0 = 0.9$, by using our selected $(31, 23)$ code, the gain achieved by encoder optimization is 1.03 dB, while the gain achieved due to using the modified extrinsic information is also 3.01 dB.

3.2 Shannon Limit

In his landmark papers [17], [18], Shannon established the *Lossy Information Transmission Theorem*, also known as the *Joint Source-Channel Coding Theorem with Fidelity Criterion* [15]. From this theorem, we know that for a given memoryless source and a given memoryless channel with capacity C, 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 [15]:

$$R_c \cdot R(D) < C, \tag{1}$$

where $R(D)$ is the rate-distortion function. For a discrete binary non-uniform i.i.d. source with distribution p_0, we have that $D = P_e$ (BER) under the Hamming distortion measure [8]; then $R(D)$ becomes

$$R(P_e) = \begin{cases} h_b(p_0) - h_b(P_e), & 0 \le P_e \le \min\{p_0, 1 - p_0\} \\ 0, & P_e > \min\{p_0, 1 - p_0\} \end{cases}$$

where $h_b(\cdot)$ is the binary entropy function: $h_b(x) = -x \log_2 x - (1 - x) \log_2(1 - x)$.

For AWGN and Rayleigh fading channels, the channel capacity is a function of E_b/N_0; i.e., $C = C(E_b/N_0)$. Therefore, the optimum value of E_b/N_0 to guarantee a bit error rate of P_e can be solved using (1) assuming equality. This optimum value of E_b/N_0 is called the *Shannon limit*, or the *optimal performance theoretically achievable (OPTA)*. The Shannon limit cannot be explicitly solved for a BPSK-modulated input due to the lack of a closed form expression [10]; so it is computed via numerical integration.

For the above simulations, we computed the OPTA at the 10^{-5} BER level for rates 1/2 and 1/3, and for $p_0 = 0.7$, 0.8 and 0.9. Table 1 provides the gaps in E_b/N_0 between the performance of the designed systems and the corresponding OPTA values. We observe that the performances of our selected Turbo codes are significantly closer to the OPTA limit than those offered by their $(37, 21)$ peer. When p_0 increases, the gaps become wider; meanwhile, the gains achieved by using our selected encoders over the $(37, 21)$ encoder become more significant. Designing more sophisticated Turbo-based joint source-channel codes that further fill the OPTA gap for heavily biased sources (e.g., with $p_0 = 0.9$) is a challenging and interesting future work. Finally, we observe that the OPTA gaps become smaller when the rate is lower.

Turbo Coding System	AWGN		Rayleigh	
	$R_c = 1/2$	$R_c = 1/3$	$R_c = 1/2$	$R_c = 1/3$
(37,21), $p_0 = 0.7$	1.07	0.89	1.17	0.91
(35,23), $p_0 = 0.7$	0.87	0.80	1.16	0.91
(37,21), $p_0 = 0.8$	2.02	1.70	2.18	1.61
(35,23), $p_0 = 0.8$	1.56	1.19	1.88	1.28
(37,21), $p_0 = 0.9$	3.69	3.20	4.12	3.26
(31,23), $p_0 = 0.9$	2.61	2.02	2.99	2.18

Table 1: OPTA gaps in E_b/N_0 at BER=10^{-5} level (in dB).

4 Conclusion

In this work, we investigate the joint source-channel coding issue of designing Turbo codes for non-uniform i.i.d. sources over noisy channels. Both AWGN and Rayleigh fading channels are considered. The source redundancy in the form of non-uniformity is exploited by the decoder via a modified extrinsic information, and the constituent encoders are optimized to further enhance the performance. Simulation results demonstrate significant coding gains due to exploiting the source redundancy in the decoder as well as using the optimized encoder structure. When p_0 increases, the gap between the performance and OPTA becomes wider. In general, better performance can be obtained when the rate decreases. Our results are sub-optimal, further gains might be achieved by a more complete encoder optimization and the use of asymmetric Turbo codes.

References

[1] F. Alajaji, S. A. Al-Semari and P. Burlina, “Visual communication via trellis coding and transmission energy allocation,” *IEEE Trans. Commun.*, Vol. 47, No. 11, pp. 1722–1728, Nov. 1999.

[2] F. Alajaji, N. Phamdo, N. Farvardin and T. Fuja, “Detection of binary Markov sources over channels with additive Markov noise,” *IEEE Trans. Inform. Theory,* Vol. 42, pp. 230–239, Jan. 1996.

[3] F. Alajaji, N. Phamdo and T. Fuja, “Channel codes that exploits the residual redundancy in CELP-encoded speech,” *IEEE Trans. Speech and Audio Processing*, Vol. 4, No. 5, pp. 325–336, Sept. 1996.

[4] 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.

[5] 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.

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

[7] 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.

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

[9] 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.

[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, No. 2, pp. 160–174, Feb. 1998.

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

[12] J. Hagenauer, P. Robertson, and L. Papke, “Iterative (turbo) decoding of systematic convolutional codes with MAP and SOVA algorithms,” *Proc. ITG Conf. on “Source and Channel Coding,”* Frankfurt, Germany, pp. 1-9, Oct. 1994.

[13] J. Hagenauer, E. Offer and L. Papke, “Iterative decoding of binary block and convolutional codes,” *IEEE Trans. Inform. Theory,* Vol. 42, No. 2, pp. 429–445, Mar. 1996.

[14] 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.

[15] R. J. McEliece, *The Theory of Information and Coding,* Addison-Wesley, 1977.

[16] 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.

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

[18] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, pt. 4, pp. 142–163, Mar. 1959.

[19] 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.