# Turbo Codes for Binary Markov Sources[1]

Guang-Chong Zhu and Fady Alajaji
Dept. of Mathematics and Statistics
Queen's University, Kingston
Ontario, Canada, K7L 3N6
E-mail: {zhugc, fady}@mast.queensu.ca

*Abstract* — **The reliable transmission via Turbo codes of binary stationary ergodic Markov sources over noisy channels is investigated. The first constituent Turbo decoder is designed to exploit the source redundancy according to a modified version of Berrou's original decoding method that employs the Gaussian assumption on the extrinsic information. Due to interleaving, the second constituent decoder is unable to adopt the same decoding method. However, its extrinsic information is modified via a weighted correction term. Substantial gains (from 1.29 up to 3.57 dB) over the original Turbo codes are demonstrated. Furthermore, the gaps to the Shannon limit are in the range of 0.94 to 1.45 dB. The proposed joint source-channel Turbo coding scheme is also shown to outperform two traditional tandem coding schemes while keeping a lower system complexity.**

## I. Introduction

Due to Shannon's well-known separation principle, source and channel coding are traditionally implemented independently, resulting in the so-called *tandem* scheme. Thus in almost all the theory and practice of error-control coding, the source is usually assumed to be uniform and memoryless. Obviously, this is seldom the case in natural sources (e.g., image and speech sources), which often exhibit substantial amounts of redundancy in the form of non-uniformity and/or memory. Source coding, if ideally designed, should entirely eliminate such redundancy and produce a uniform memoryless bit stream. However, most existing source coding schemes are sub-optimal, retaining a certain amount of *residual* redundancy in their output. Therefore, transmission of sources with a considerable amount of natural or residual redundancy is an important issue. It was indeed shown (e.g., [1]) that when the source redundancy is exploited in the channel coding design, the system performance can be significantly improved.

Turbo codes [2] have been regarded as one of the most exciting breakthroughs in channel coding, and their excellent performance has been demonstrated over additive white Gaussian noise (AWGN) channels and Rayleigh fading channels. However, most of the work on Turbo codes has mainly focused on uniform memoryless sources. To the best of our knowledge, only limited attention has been paid to the problem of exploiting the source redundancy in Turbo codes. In essence, this is a joint source-channel coding issue. The design of Turbo codes for the transmission of non-uniform memoryless sources has been recently studied in [12, 13], where close to Shannon limit performance was achieved. On the other hand, designing Turbo codes for sources with memory was considered in [5, 7, 8, 9]. In this work, we consider stationary ergodic binary first-order Markov sources. We investigate the design of (systematic) Turbo codes for transmitting such sources over BPSK-modulated AWGN and Rayleigh fading channels with known channel state information. Our goal is to try to achieve a performance that is as close to the Shannon limit as possible. The proposed framework can be extended to high-order Markov sources by using non-binary Turbo codes and non-binary modulation.

[1]This work was supported in part by NSERC of Canada and PREA of Ontario.

## II. Turbo Decoder Design

Consider a stationary ergodic binary first-order Markov source $\{U_k\}_{k=1}^{L}$ with transition matrix:

$$\Pi = [\pi_{ij}] = \begin{bmatrix} q_0 & 1-q_0 \\ 1-q_1 & q_1 \end{bmatrix},$$

where $\pi_{ij} \triangleq Pr\{U_k = j | U_{k-1} = i\}$, $i, j \in \{0,1\}$. Also, denote the source marginal distribution by $p_0 \triangleq 1 - p_1 \triangleq Pr\{U_k = 0\}$. By stationarity, it can be easily shown that

$$p_0 = 1 - p_1 = \frac{1-q_1}{2-q_0-q_1}. \quad (1)$$

In general, the above source is an *asymmetric* Markov source; its redundancy is in the form of both memory and non-uniformity. When $q_0 = 1 - q_1$, the source reduces to a memoryless source with marginal distribution $p_0 = q_0$ and $p_1 = q_1$; in this case the source redundancy is purely in the form of non-uniformity. When $q_0 = q_1 \neq 1/2$, the source becomes a *symmetric* Markov chain with a *uniform* marginal distribution, i.e., $p_0$=$p_1$=1/2; the source redundancy is therefore strictly in the form of memory. When the source is symmetric, we denote $q_0 = q_1 \triangleq q$.

### A. First Constituent Decoder

To exploit the source memory, the BCJR algorithm employed in the Turbo code decoder [2] has to be modified. Given that at time $k$ the encoder is in state $S_k$, then an input bit $U_k$ would bring the encoder state into $S_{k+1}$, with a parity bit $X_k^{1p}$ generated ($X_k^{2p}$ denotes the parity bit of the second constituent decoder). To keep the notation consistent, we denote the systematic bit by $X_k^s$, which is identical to $U_k$. The pair $(X_k^s, X_k^{1p})$ is denoted by $X_k$, and after transmission over the channel, it becomes $Y_k = (Y_k^s, Y_k^{1p})$, the noise corrupted version. For the sake of brevity, the sequence $\{Y_k\}_{k=1}^{L}$ is denoted as $Y_1^L$. Given that the $L$-tuple $Y_1^L = y_1^L$ is received at the channel output, we have

$$Pr\{U_k=i|Y_1^L=y_1^L\}=\sum_s Pr\{U_k=i,S_k=s|Y_1^L=y_1^L\}$$
$$=\sum_s \frac{Pr\{U_k=i,S_k=s,Y_1^k=y_1^k,Y_{k+1}^L=y_{k+1}^L\}}{p(y_1^k,y_{k+1}^L)}$$
$$=\sum_s \frac{Pr\{U_k=i,S_k=s,Y_1^k=y_1^k\}\cdot p(y_{k+1}^L|i,s,y_1^k)}{p(y_1^k)\cdot p(y_{k+1}^L|y_1^k)}.$$

Since given $U_k=i$ and $S_k=s$, the observation $Y_{k+1}^L=y_{k+1}^L$ does not depend on $Y_1^k=y_1^k$, hence if we define

$$\alpha_k(i,s) \triangleq Pr\{U_k=i,S_k=s|Y_1^k=y_1^k\}, \quad (2)$$
$$\beta_k(i,s) \triangleq \frac{p(y_{k+1}^L|i,s)}{p(y_{k+1}^L|y_1^k)}, \quad (3)$$

then the above conditional probability becomes

$$Pr\{U_k=i|y_1^L\}=\sum_s \alpha_k(i,s)\beta_k(i,s), \quad (4)$$

where $\alpha_k(i,s)$ and $\beta_k(i,s)$ can be computed via the following recursive relations:

$$\alpha_k(i,s) = \frac{\sum_{i',s'}\gamma(i,s,y_k|i',s')\alpha_{k-1}(i',s')}{\sum_{i,s}\sum_{i',s'}\gamma(i,s,y_k|i',s')\alpha_{k-1}(i',s')}, \quad (5)$$
$$\beta_k(i,s) = \frac{\sum_{i',s'}\gamma(i',s',y_{k+1}|i,s)\beta_{k+1}(i',s')}{\sum_{i,s}\sum_{i',s'}\gamma(i',s',y_{k+1}|i,s)\alpha_k(i,s)}, \quad (6)$$

where

$$\gamma(i,s,y_k|i',s')$$
$$=Pr\{U_k=i,S_k=s,Y_k=y_k|U_{k-1}=i',S_{k-1}=s'\}$$
$$=p(y_k^s|i)\cdot p(y_k^p|i,s)\cdot Pr\{S_k=s|U_k=i,S_{k-1}=s'\}\cdot Pr\{U_k=i|U_{k-1}=i'\}$$
$$\triangleq p(y_k^s|i)\cdot p(y_k^p|i,s)\cdot q(s|i,s')\cdot \pi(i|i'). \quad (7)$$

Solving (5) and (6) requires boundary conditions. Since the encoder starts and terminates at the all-zero state, and the source has marginal distribution as given in (1), we have

$$\alpha_0(0,0)=p_0,\quad \alpha_0(1,0)=p_1,$$
$$\alpha_0(i,s)=0,\quad i=0,1,\quad \forall\ s\neq 0,$$
$$\beta_L(0,0)=p_0,\quad \beta_L(1,0)=p_1,$$
$$\beta_L(i,s)=0,\quad i=0,1,\quad \forall\ s\neq 0.$$

Note that in the above boundary conditions, the source redundancy in the form of non-uniformity is exploited.

For iterative decoding, the log-likelihood ratio (LLR) needs to be decomposed. Combining (4)-(7), we can decompose the LLR of $U_k$ into two separate terms:

$$\Lambda_k^{(1)}(U_k) \triangleq \log\frac{Pr\{U_k=1|Y_1^L=y_1^L\}}{Pr\{U_k=0|Y_1^L=y_1^L\}} = L_{ch}(U_k)+L_{ex}^{(1)}(U_k), \quad (8)$$

where the channel transition term is

$$L_{ch}(U_k)\triangleq\log\frac{p(y_k^s|1)}{p(y_k^s|0)} \quad (9)$$

and the new extrinsic information term is

$$L_{ex}^{(1)}(U_k)\triangleq\log\frac{\sum_s\sum_{i',s'}r(y_k^p,1,s,i',s')}{\sum_s\sum_{i',s'}r(y_k^p,0,s,i',s')}, \quad (10)$$

where for $i=0,1$,

$$r(y_k^p,i,s,i',s') \triangleq p(y_k^p|i,s,s')q(s|i,s')\pi(i|i')\alpha_{k-1}(i',s')\beta_k(i,s).$$

In comparison with the decomposition of the LLR in the case of Turbo coding of uniform i.i.d. sources [2] where

$$\Lambda_k^{iid}(U_k)=L_{ch}(U_k)+L_{ex}^{iid}(U_k)+L_{ap}^{iid}(U_k),$$

we observe that in (8), $L_{ex}^{(1)}(U_k)$ is in essence the combination of $L_{ex}^{iid}(U_k)$ and $L_{ap}^{iid}(U_k)$, and that $L_{ap}^{iid}(U_k)$ is actually the extrinsic information generated from the second constituent decoder. An important principle in iterative decoding is that the estimation generated by a constituent decoder should not be fed back to itself; otherwise the noise corruption will be highly correlated [2]. On the other hand, the decomposition in (8) renders the first constituent decoder unable to "update" its *a priori* term by using the extrinsic information generated from the second constituent decoder in the same way as in the memoryless source case. Therefore, the extrinsic information term for the first constituent decoder must be further modified. This issue is not clearly addressed in [8, 9]. The method we adopt here is from Berrou's original version of the BCJR algorithm [2]. The input to the first decoder now has three components: $\mathbf{y}_k=(y_k^s,y_k^{1p},y_k^{ex(2)})$, where $y_k^{ex(2)}$ is the extrinsic information from the second decoder, $L_{ex}^{(2)}(U_k)$, after de-interleaving. Using the Gaussian assumption on $y_k^{ex(2)}$ as in [2, 6], in (7), $\gamma(i,s,y_k|i',s')$ has one more factor described by the following density

$$p(y_k^{ex(2)}|i)=\frac{1}{\sqrt{2\pi\sigma_i^2}}e^{-\frac{\left[y_k^{ex(2)}-(2i-1)M_i\right]^2}{2\sigma_i^2}},\quad i=0,1,$$

where $M_i$ and $\sigma_i^2$ are estimated on-line.

Finally, in the decomposition of $\Lambda^{(1)}(U_k)$, due to interleaving, $y_k^{ex(2)}$ can be regarded as weakly correlated with $Y_k^s=y_k^s$ and $Y_{1k}^p=y_{1k}^p$; therefore, the LLR soft-output generated by the first decoder is:

$$\Lambda_k^{(1)}=L_{ch}(U_k)+L_{ex}^{(1)}(U_k)+L_{ap}^{(1)}(U_k),$$

where the *a priori* term becomes

$$L_{ap}^{(1)}(U_k) \triangleq \log\frac{p(y_k^{ex(2)}|1)}{p(y_k^{ex(2)}|0)} = \log\frac{\sigma_0}{\sigma_1}+\frac{(y_k^{ex(2)}+M_0)^2}{2\sigma_0^2}-\frac{(y_k^{ex(2)}-M_1)^2}{2\sigma_1^2}.$$

When the source is symmetric, we have $M_1=M_0\triangleq M$, and $\sigma_1^2=\sigma_0^2\triangleq\sigma^2$. Therefore

$$L_{ap}^{(1)}(U_k)=\frac{2M}{\sigma^2}y_k^{ex(2)}.$$

At this point, we can use $L_{ex}^{(1)}(U_k)$ as the extrinsic information for the second constituent decoder.

### B. Second Constituent Decoder

For the second constituent decoder, due to interleaving, the Markovian property in the input sequence is destroyed; this renders the second constituent decoder unable to adopt the same modifications as in Section II.A.

Instead, we employ Robertson's classical Turbo decoding algorithm [11] with the following modification to the extrinsic information generated from the second decoder:

$$L_{ex}^{(2)}(U_k) = c_1 L_{ex}(U_k) + c_2 \log \left[ \frac{\sum_i \pi(1|i) Pr\{\hat{U}_{k-1} = i\}}{\sum_i \pi(0|i) Pr\{\hat{U}_{k-1} = i\}} \right],$$

where $L_{ex}(\cdot)$ is the extrinsic information defined in [11] and $c_1, c_2 \in [0, 1]$. The values of $c_1$ and $c_2$ are empirically chosen to yield the best possible improvement. In our simulations, $c_1 = 0.8$ and $c_2 = 0.2$ were the best choice. $Pr\{\hat{U}_{k-1} = i\}$ (with $i$=0, 1) is directly obtained from the extrinsic information described by

$$Pr\{\hat{U}_{k-1} = 1\} = 1 - Pr\{\hat{U}_{k-1} = 0\} = \frac{e^{L_{ex}(U_{k-1})}}{1 + e^{L_{ex}(U_{k-1})}},$$

for $k \geq 2$ with $Pr\{\hat{U}_1 = 0\} = p_0 = 1 - Pr\{\hat{U}_1 = 1\}$.

## III. Turbo Encoder Design and Optimization

When the Markov source distribution is biased, the input sequence to the encoder may contain long segments of 1s or 0s. As a result, similar to the case of heavily biased non-uniform i.i.d. sources [12], the encoder structure plays an important role in determining the system performance. Therefore, encoder structure optimization is also necessary in the design of Turbo codes for Markov sources. Furthermore, the two constituent decoders employ different decoding algorithms and thus exploit the source redundancy with different degrees. Therefore, the two constituent encoders do not have to be identical. Due to the increased number of possible encoder structures by using different constituent encoders, an exhaustive search is computationally expensive. Our search for (sub-)optimal encoder structures with four memory elements (16 states) is performed as follows.

1) Fix the second constituent encoder as, for example, (31, 23), find (by simulation) the best feed-forward and feedback polynomials of the first constituent encoder via the iterative steps described in [12].

2) Fix the best structure for the first constituent encoder as found in step 1), find the best feed-forward and feedback polynomials of the second constituent encoder.

The initial structure of the second constituent encoder in step 1) is selected according to our results in [12].

## IV. Shannon Limit

For an asymmetric Markov source, the rate-distortion function $R(D)$ has no closed-form expression, but relatively tight upper and lower bounds for $R(D)$ can be obtained [3]. Therefore, for a desired BER level and a given overall rate $r$ in source symbols/channel symbol, we may substitute $R(D)$ under the Hamming distortion measure with its upper/lower bound in the condition of Shannon's Information Transmission Theorem [10]

$$r \cdot R(D) < C(E_b/N_0), \tag{11}$$

to obtain an upper/lower bound on the corresponding Shannon limit, where $C$ is the channel capacity, $E_b$ is the average energy per source bit and $N_0/2$ is the variance of the additive Gaussian noise. For binary *symmetric* Markov sources with transition probability $q$, Gray proved that [4]

$$R(D) = H_b(q) - H_b(D), \quad \text{if } 0 \leq D \leq D_c, \tag{12}$$

where $D_c$ is the *critical distortion* defined by

$$D_c = \frac{1}{2}\left(1 - \sqrt{1 - \frac{(1-q)^2}{q^2}}\right), \quad \text{for} \quad q > 1/2.$$

## V. Numerical Results and Discussion

We present simulation results of our designed systematic Turbo codes for the transmission of binary symmetric Markov sources over BPSK-modulated AWGN and Rayleigh fading channels. All simulated Turbo codes have 16-state constituent encoders and use the same pseudo-random interleaver as in [2]. The sequence length is $L = 512 \times 512 = 262144$ 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 in the Turbo code decoder is 20. Simulations are performed for rate $R_c = 1/3$ with $q$=0.8 and 0.9; since the corresponding critical distortion values are $1.59 \times 10^{-2}$ and $3.10 \times 10^{-3}$, respectively, (all above the $10^{-5}$ BER level of our interest), the Shannon limit can be computed exactly using (11) and (12). The best encoder we found for $q$=0.9 has the first constituent encoder as (31, 23), and the second as (35, 23). For $q$=0.8, (35, 23) for both constituent encoders turned out to be the best choice.

Fig. 1 shows the performance of our rate-1/3 systematic Turbo codes for AWGN channels. Berrou's (37, 21) original code which does not exploit the source memory is also shown for the sake of comparison. At the $10^{-5}$ BER level, when $q$=0.8, our system offers a gain of 1.29 dB over Berrou's code; when $q = 0.9$, the gain becomes 3.03 dB. Furthermore, we observe that for $q = 0.8$ and 0.9, the gaps to the Shannon limit are 0.94 dB and 1.36 dB, respectively. Fig. 2 illustrates similar results for Rayleigh fading channels with known channel state information. When $q = 0.8$, the gain due to exploiting the source memory in our design is 1.55 dB, which brings the performance at a distance of 1.08 dB away from the Shannon limit. When $q = 0.9$, the gain increases to 3.57 dB; in this case, the gap to the Shannon limit is 1.45 dB. The Shannon limit (SL) values and the performance gaps of our system vis-a-vis SL are summarized in Table 1.

Finally, we compare our joint source-channel coding scheme with two tandem coding schemes. Similar to [13], the tandem schemes consist of a Huffman code (performing near optimal data compression) followed by a standard symmetric Turbo code. The one with constituent encoders (35, 23) offers a lower error-floor performance than that of its (37, 21) peer at the price of a slight performance loss in the water-fall region. The comparison is made at the same overall rate of $r = R_c/R_s = 1/2$ source symbols/channel symbol, where $R_c$ and $R_s$ are the source coding and channel coding rates, respectively. Our system has $R_c = 1/2$ and $R_s = 1$; the tandem scheme, however, has $R_c = 1/3$, therefore it needs to have $R_s = 2/3$. Using an $8^{th}$-order Huffman code, when

$q$=0.848315, we can achieve $R_s$=0.666667. The sequence length is $L$=12000, and at least 60000 blocks are simulated to produce a reliable average performance in the error-floor region. An S-random interleaver with $S$=10 is adopted as in [13] and the number of decoder iterations is 20. Results for AWGN channels are shown in Fig. 3. We observe that although initially the tandem schemes offer better water-fall performance, they are quickly outperformed by our system due to their high BER performance from medium to high SNRs. Our system, however, enjoys a lower complexity since no source encoding and decoding are performed, and offers a superior BER performance which is robust to channel errors.

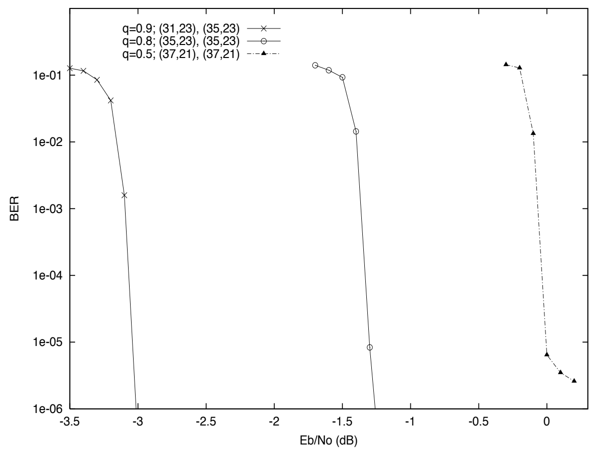


Figure 1: Turbo codes for binary symmetric Markov sources, $R_c$=1/3, $L$=262144, AWGN channel.

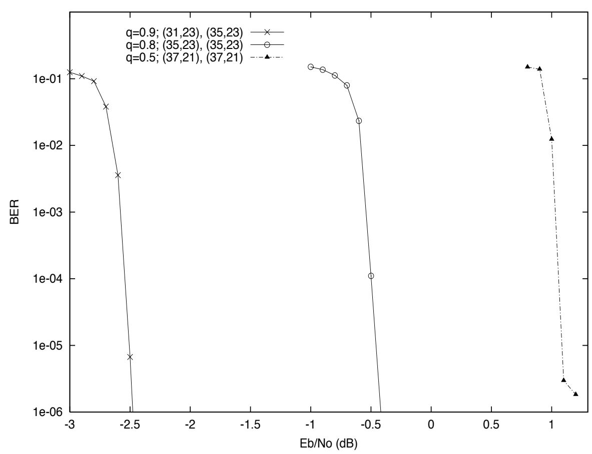


Figure 2: Turbo codes for binary symmetric Markov sources, $R_c$=1/3, $L$=262144, Rayleigh channel.

| Transition probability | AWGN | | Rayleigh | |
|---|---|---|---|---|
| | SL | gap | SL | gap |
| $q = 0.8$ | -2.24 | 0.94 | -1.56 | 1.08 |
| $q = 0.9$ | -4.40 | 1.36 | -3.96 | 1.45 |

Table 1: Shannon limit (SL) and performance gaps in $E_b/N_0$ (in dB) at BER=$10^{-5}$, $R_c$=1/3.

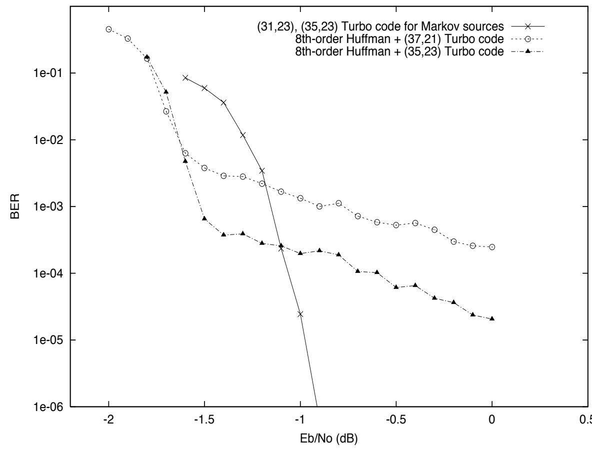


Figure 3: Proposed system ($R_s = 1, R_c = 1/2$) versus two tandem schemes ($R_s = 2/3$, $R_c = 1/3$), $q$=0.848315, $L$=12000, AWGN channel.

## References

[1] 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, pp. 325–336, Sept. 1996.

[2] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes(1)," *Proc. IEEE Int. Conf. on Commun.*, pp. 1064–1070, Geneva, Switzerland, May 1993.

[3] R. E. Blahut, "Computation of channel capacity and rate-distortion functions", *IEEE Trans. Inform. Theory*, vol. 18, pp. 460-473, July 1972.

[4] R. M. Gray, "Information rates for autoregressive processes," *IEEE Trans. Inform. Theory*, vol. 16, pp. 412-421, Jul. 1970.

[5] Z. Cai, K. R. Subramanian and L. Zhang, "Source-controlled channel decoding using nonbinary turbo codes," *Elect. Lett.*, vol. 37, pp. 39–40, 4*th* Jan. 2001.

[6] G. Colavolpe, G. Ferrari, and R. Raheli, "Extrinsic information in Turbo decoding: a unified view," *Proc. Globecom'99*, pp. 505–509, Rio de Janeiro, Brazil, Dec. 1999.

[7] T. Fazel and T. Fuja, "Joint source-channel decoding of block-encoded compressed speech," *Proc. Conf. Inform. Sciences and Systems*, pp. 1–6, Princeton, NJ, Mar. 2000.

[8] 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.

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

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

[11] P. Robertson, "Illuminating the structure of parallel concatenated recursive systematic (turbo) codes," *Proc. Globecom'94*, vol. 3, pp. 1298-1303, San Francisco, CA, Nov. 1994.

[12] 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.

[13] G.-C. Zhu, F. Alajaji, J. Bajcsy and P. Mitran, "Transmission of non-uniform memoryless sources over wireless fading channels via non-systematic Turbo codes," *Proc. IASTED Wireless Optical Commun. Conf.*, pp. 119–124, Banff, Alberta, Canada, Jul. 2002.