Non-Systematic Turbo Codes for Non-Uniform I.I.D. Sources over AWGN Channels[1]

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

Dept. of Mathematics & Statistics
Queen's University
Kingston, Ontario,
Canada, K7L 3N6
zhugc@mast.queensu.ca
fady@mast.queensu.ca

Dept. of Electrical Engineering
McGill University
Montreal, Québec,
Canada, H3A 2A7
jbajcsy@tsp.ece.mcgill.ca
pmitran@tsp.ece.mcgill.ca

Abstract — **In this work, we investigate the joint source-channel coding issue of transmitting non-uniform independent and identically distributed (i.i.d.) sources over additive white Gaussian noise (AWGN) channels via Turbo codes. The source redundancy in the form of non-uniformity is exploited in the Turbo decoder via a modified extrinsic information. In contrast to previous work, non-systematic recursive convolutional encoders are proposed as the constituent encoders, which produce almost uniform outputs for heavily biased sources. 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 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.**

I. Introduction and Motivation

In almost all the theory and practice of error-control coding, the source that is encoded for transmission over the channel is assumed to be uniform i.i.d.; i.e., the source is assumed to generate memoryless 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 documents and medical images) may contain as much as 80% of redundancy in the form of non-uniformity (e.g., [8, 14]); this corresponds to a probability

$$p_0 \stackrel{\triangle}{=} Pr\{D_k = 0\} = 0.97.$$

[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.

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 and generates uniform i.i.d. outputs. However, most existing source encoders are only sub-optimal (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 contains 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., [1]-[4], [10, 13, 16, 20, 21], etc.) have shown that appropriate use of the source redundancy can significantly improve the system performance.

Turbo codes [6, 7] have been regarded as one of the most exciting breakthroughs in channel coding, and excellent performance has been demonstrated for uniform i.i.d. sources over AWGN channels. In [11] the authors considered using Turbo codes for sources with memory. However, 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 [22, 23], where the source redundancy in the form of non-uniformity was exploited in the Turbo decoder via a modified extrinsic information term, and the encoder structure was optimized in accordance with the source distribution. As a result, significant coding gains were achieved in these works over the standard Berrou Turbo code, and the performance results compared fairly well to the *Shannon limit*, also known as the *optimal performance theoretically achievable (OPTA).*

In [22, 23], the Turbo encoders used are recursive *systematic* convolutional (RSC) encoders. Although the gains achieved in these works are considerably significant, the OPTA gaps for heavily biased sources are still relatively big. Also, when p_0 increases, the gaps become wider. Analysis on the encoder output reveals that the drawback lies in the *systematic* structure. Due to the feedback, recursive non-systematic con-

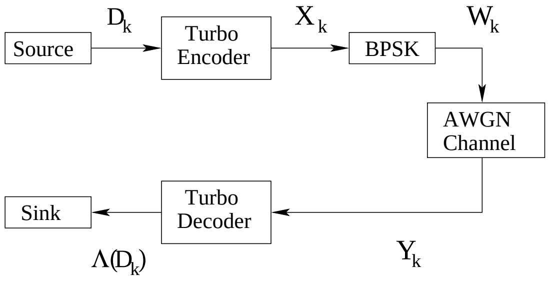

Figure 1: Block diagram of the system model.

volutional (RNSC) encoders can generate almost uniformly distributed output even for very biased sources. From information theory (e.g., [9]), we know that the capacity of a binary input AWGN channel is maximized by a uniform i.i.d. channel input; therefore, we propose using RNSC encoders as the constituent Turbo encoders. Simulation results demonstrate substantial gains over systematic Turbo codes. The OPTA gaps for heavily biased sources are hence significantly reduced.

II. 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 $p_0 = Pr\{D_k = 0\} \neq 1/2$. 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 code. The data sequence is Turbo encoded and binary phase-shift keying (BPSK) modulated; then it is transmitted through an AWGN channel whose matched filter output is described by

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

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 $N_0/2$. We assume that W_k and N_k are independent of each other. At the receiver end, the sequence is fed into the Turbo decoder, which iteratively computes the log-likelihood ratio (LLR) $\Lambda(D_k)$ of each bit D_k. The Turbo encoder used here is recursive non-systematic, and the Turbo decoder is modified accordingly for the encoder structure as well as to exploit the source redundancy in the form of non-uniformity.

III. 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 [5], 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.* [7], extraordinary performance has been demonstrated by using Turbo codes for uniform i.i.d. sources over AWGN channels.

Designing Turbo codes for non-uniform i.i.d. sources has been recently studied in [22, 23], in which the Turbo decoder is modified to take advantage of the source redundancy in the form of non-uniformity, and the Turbo encoder is optimized for a given source probability distribution. For AWGN channels, when rate=1/3 and p_0=0.9, the optimization of the encoder yields a 1.18 dB gain over the Berrou (37,21) code, while exploiting the source redundancy gives an impressive 2.46 dB gain. Despite these significant coding gains, 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 the rate is 1/2, OPTA gaps of 1.56 dB and 2.61 dB are achieved for p_0=0.8 and 0.9, respectively.

Note that in [22, 23], the encoders are *systematic*, which is commonly used in almost all the Turbo codes literature. When the source is heavily biased, this systematic structure becomes 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 source sequence) contains much more 0's than 1's. 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 can be almost uniformly distributed even for a very heavily biased source input.

It has been shown in [17] 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. Since the capacity of a binary input AWGN channel is achieved when its mutual information is maximized by a uniformly distributed input, 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

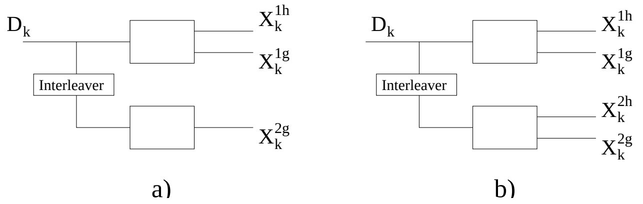

Figure 2: Non-Systematic Turbo encoder structures.

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 [22, 23], 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 [22, 23], and search for the other best feed-forward polynomial. Searches are performed separately (with puncturing) for rate-1/2 and rate-1/3 encoders.

IV. Decoder Modifications

When RSC encoders are used as constituent encoders, the log-likelihood ratio (LLR) in the BCJR algorithm [5] employed by the Turbo decoder can be decomposed into three terms [7]:

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

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,

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

where for $i = 0, 1$,

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

and where E_k is the encoder state at time k, 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. $\alpha_k(e)$ and $\beta_k(e)$ are defined and can be recursively computed as in [7]. Also, as in [22, 23], since 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[1]; 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.

V. Simulation Results and Discussions

In this section, we present simulation results of our non-systematic Turbo codes for uniform i.i.d. sources over BPSK-modulated AWGN 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 [7]. 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) encoders as they have a better performance than the codes with structure a) encoders. Simulations are performed for rates $R_c = 1/3$ and $R_c = 1/2$ with 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) and (31,23,37) for p_0=0.9 also give a good performance that is very close to the one 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 [22, 23], as well as with Berrou's (37,21) code, which offers the best water-fall 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, the improvement is 0.89 dB with the encoder structure (31,23,27). In comparison with Berrou's (37,21) code performance, the gains achieved by exploiting the source redundancy and encoder optimization are therefore 1.48 dB and 3.25 dB for p_0=0.8 and 0.9, respectively.

Fig. 4 shows similar results for rate-1/2. We observe that 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 0.69 dB and 1.56 dB, respectively. Furthermore, the gains due to

[1] This simple modification of appropriately using the source information in the Turbo decoder for non-uniform sources is also briefly mentioned in [12] and [11]; but it is not explicitly studied or assessed.

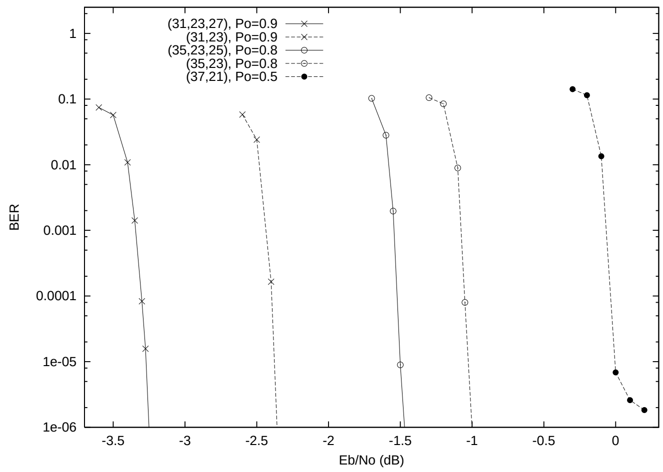

Figure 3: Turbo codes for non-uniform i.i.d. sources, R_c=1/3, N=262144, AWGN channel. The reference dotted curves are from [22, 23].

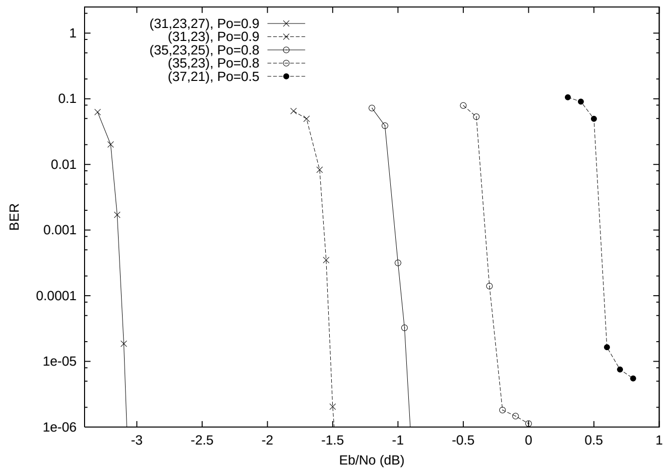

Figure 4: Turbo codes for non-uniform i.i.d. sources, R_c=1/2, N=262144, AWGN channel. The reference dotted curves are from [22, 23].

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).

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 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 0.2 dB better than the 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 from feed-forward 23, and delete all odd (even) position bits of the sequences generated from the other feed-forward polynomial.

VI. Shannon Limit

In his landmark papers [18], [19], 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 source rate-distortion function. For a discrete binary non-equiprobable memoryless source with distribution p_0, we have that $D = P_e \triangleq \text{BER}$ under the Hamming distortion measure [9]; then $R(D)$ becomes

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

where $h_b(\cdot)$ is the binary entropy function:

$$h_b(x) = -x \log_2 x - (1-x)\log_2(1-x).$$

For AWGN 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 BER of P_e can be solved using (1) assuming equality. This optimum value of E_b/N_0 is called the Shannon limit, or OPTA. The Shannon limit cannot be explicitly solved for a BPSK-modulated input due to the lack of a closed form expression; so it is computed via numerical integration.

For the above simulations, the OPTA values at the 10^{-5} BER level are computed and provided in Table 1. The OPTA gaps, which are the distances between our system performances and the corresponding OPTA values, are provided in Table 2. We can clearly remark that the OPTA gaps are significantly reduced by non-systematic Turbo codes: less than 0.9 dB for $p_0 = 0.8$ and within 1 dB for $p_0 = 0.9$.

Source distribution	$R_c = 1/2$	$R_c = 1/3$
$p_0 = 0.8$	-1.81	-2.24
$p_0 = 0.9$	-4.14	-4.40

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

Source distribution	Systematic [22, 23]		Non-Systematic	
	$R_c = 1/2$	$R_c = 1/3$	$R_c = 1/2$	$R_c = 1/3$
$p_0 = 0.8$	1.56	1.19	0.87	0.74
$p_0 = 0.9$	2.61	2.02	1.05	1.13

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

VII. Conclusions

In this work, the joint source-channel coding issue of designing Turbo codes for transmitting non-uniform i.i.d. sources over AWGN channels is investigated. Recursive non-systematic Turbo codes are proposed for the considered sources. The non-systematic Turbo encoder output, which is almost uniformly distributed for even heavily biased sources, is suitably matched to the channel input as it nearly maximizes the channel mutual information. Simulation results show substantial coding gains (up to 1.56 dB) achieved in comparison with systematic Turbo codes designed in [22, 23], and the OPTA gaps are significantly reduced.

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, A. Glavieux and P. Thitimajshima, "Near Shannon limit error correcting coding and decoding: Turbo codes," *Proc. IEEE Int. Conf. Commun.*, Geneva, Switzerland, pp. 1064-1070, 1993.

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

[8] 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.

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

[10] T. Fingsheidt, T. Hindelang, R. V. Cox and N. Sheshadri, "Joint source-channel (de)coding for mobile communications," *IEEE Trans. Commun.*, Vol. 50, No. 2, pp. 200-212, Feb. 2002.

[11] 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.

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

[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] S. Shamai and S. Verdú, "The empirical distribution of good codes," *IEEE Trans. Inform. Theory*, Vol. 43, No. 3, pp. 836–846, May 1997.

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

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

[20] 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.

[21] 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.

[22] 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.

[23] G.-C. Zhu and F. Alajaji, "Turbo codes for non-uniform memoryless sources over noisy channels," *IEEE Communications Letters,* Vol. 6, No. 2, pp. 64–66, Feb. 2002.