Enhanced channel decoding via EM source-channel estimation

Joseph J. Boutros, Amira Alloum, and Gil I. Shamir
ENST, Paris, France; FT R&D Division, Issy-Les-Moulineaux, France; University of Utah, Salt Lake City, USA

***Abstract*— We investigate the joint source-channel estimation and decoding problem. We consider a non-uniform binary source transmitted over a binary-input output-symmetric channel, namely the BEC, BSC and AWGN channels. The source sequence is encoded via systematic and non-systematic low-density parity-check codes. The proposed joint source-channel iterative estimation technique relies on the Expectation Maximization (EM) algorithm that will be associated to the message passing LDPC decoding for both systematic and non systematic codes. Simulation results confirm the strong improvement in performance over the case in which source a priori information is not considered. Furthermore, within this proposed joint source-channel iterative estimation via Expectation Maximization no loss in error-rate performance is observed with respect to the perfect knowledge case.**

I. Introduction

Since the invention of low-density parity-check codes [5] (LDPC), recent developments of codes defined on graphs [7] considered the design and optimization of error-correcting codes for uniformly distributed sources. The unique exception appears in Mackay-Neal (MN) codes in [8] that describes the scrambling of low-density generator-matrix codewords and the use of source *a priori* probability in MN decoding. More recently, following the study of parallel turbo codes for non-uniform binary sources [11], a special class of non-systematic LDPC codes has been proposed for non-uniform i.i.d binary sources [10][1].
The full utilization of systematic and non-systematic LDPC codes for channel coding of non-uniform sources requires: 1- The knowledge of the source probability distribution, and 2- The knowledge of channel parameters, at the decoder side. In this paper, we show that both source state information (SSI) and channel state information (CSI) can be jointly estimated by the iterative decoder via Expectation-Maximization (EM) [3][9] with no loss in error rate performance, and at the expense of a negligible complexity.
The paper is organized as follows: we define the system model and the associated notations in Section II. In Section III, we briefly recall the statement of EM algorithm as a recursive low complexity approach to ML estimation. We then show in sections IV and V how EM can be integrated into the decoding of systematic and non systematic LDPC codes for both binary symmetric channel with erasures and additive white gaussian noise channel.

II. System Model and Notations

Let us consider a non-uniform binary independent identically distributed source. The source generates a binary sequence of length k denoted by $\mathbf{s} \triangleq (s_1, s_2, ..., s_k)$ where $s_i \in \{0,1\}$. We assume that the source sequence is directly fed to a channel encoder without any data compression. Our study is restricted to channel encoding via systematic and non-systematic binary LDPC codes of rate $R_c = k/n$. An LDPC codeword has length $n \geq k$ and is denoted by $\mathbf{x} \triangleq (x_1, ..., x_n)$. The codeword $\mathbf{x}$ is transmitted over a noisy channel defined by its CSI, the received noisy vector being $\mathbf{y}$. The LDPC decoder should retrieve $\mathbf{s}$ from $\mathbf{y}$ with prior knowledge of source and channel statistics.

A. Source State Information

The binary i.i.d. source is characterized by the parameter $\mu = P(s_i = 1)$, where $0 < \mu \leq \frac{1}{2}$. The source entropy is given by $H_s = H_2(\mu)$, $0 < H_s \leq 1$, where $H_2(x)$ is the binary entropy function. Our model is not restrictive, instead of making a direct estimation of a stationary finite-memory source as in [6], it is possible to convert a finite-memory source into a piecewise i.i.d. non-uniform source [4]. The EM estimation described in the sequel can easily tackle with the piecewise variation of the source.

B. Channel State Information

Two kinds of binary input symmetric output channels are considered: the binary symmetric channel with erasures (BECBSC) and the complex binary input additive white gaussian noise channel (BIAWGN). Characteristics and main parameters of both channels are summarized below.
The BECBSC channel has a binary input and a ternary output. The cross-over probability of its BSC part is λ and the erasure probability of its BEC part is ε. Channel observations at the decoder input are derived from normalized likelihoods

$$obs(x_i) = \frac{p(y_i|x_i)}{\sum_j p(y_j|x_j)} = \begin{cases} \lambda' & \text{if } y_i = x_i \\ 1-\lambda' & \text{if } y_i = \overline{x_i} \\ \frac{1}{2} & \text{if } y_i = \frac{1}{2} \end{cases}$$

where $\overline{x_i}$ denotes the binary complement of x_i, and $\lambda' = \dfrac{\lambda}{1-\varepsilon}$, where $0 \leq \lambda < \frac{1}{2}$ and $0 \leq \varepsilon < 1$. Consider the special case of BEC ($\lambda = 0$). Using classical information theoretical tools [2], it is easy to show that the maximum achievable erasure probability for systematic codes is

$$\varepsilon_{max} = \frac{1 - R_c}{1 - Rc(1 - H_s)} \tag{1}$$

However, for non-systematic codes, we have

$$\varepsilon_{max} = 1 - R_c H_s \tag{2}$$

For the special case of BSC ($\varepsilon = 0$), the maximum achievable cross-over probability for systematic codes is given by

$$H_s R_c = [1 - H_2(\lambda_{max})] - R_c[1 - H_2(q)] \tag{3}$$

where $q = \mu(1-\lambda_{max}) + (1-\mu)\lambda_{max}$. For non-systematic codes, the maximum achievable cross-over becomes

$$H_s R_c = 1 - H_2(\lambda_{max}) \tag{4}$$

The BIAWGN channel is defined as

$$y_i = Ae^{\Im\phi} x_i + \eta_i \quad \text{with} \quad \Im = \sqrt{-1} \tag{5}$$

where the three CSI parameters are amplitude A, phase ambiguity ϕ, and gaussian noise variance σ^2. Similar expressions for the capacity of BIAWGN with non-uniform sources can be found in [1]. Capacity limits versus H_s are illustrated in Figures 1 and 2 for $R_c = 1/2$.

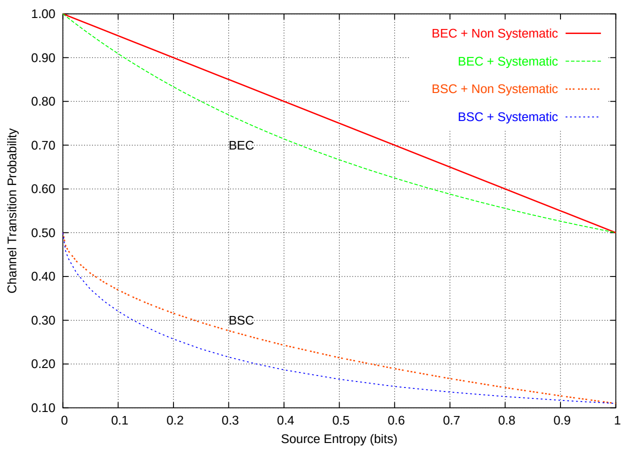

Fig. 1. Capacity limit versus source entropy, BEC and BSC.

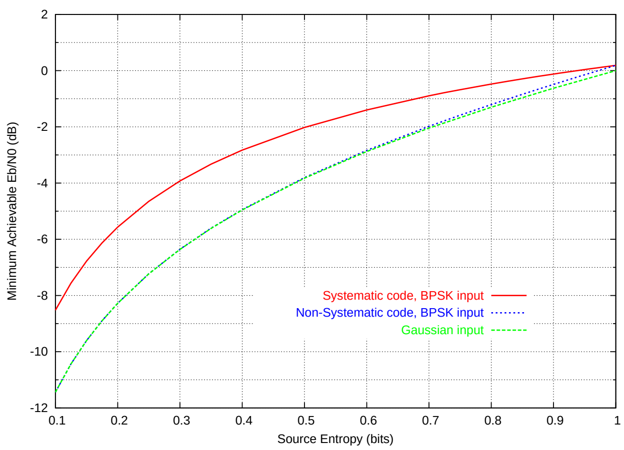

Fig. 2. Capacity limit versus source entropy, BIAWGN.

III. Brief Statement of the EM Algorithm

The general model is depicted on Fig. 3. The observation **y** is referred to as *incomplete data* and κ = (**x, y**) as *complete data*. When κ = (**x, y**) is available, maximum-likelihood estimation can be performed to determine SSI and CSI at the decoder side. In a coded communication system, the incomplete data **y** is the only available observation to the decoder. ML estimation of a parameter θ which includes both SSI and CSI is therefore obtained by maximizing the log-likelihood: $\widehat{\theta} = \arg\max_\theta \log p(\mathbf{y}|\theta)$. Unfortunately, analytical or numerical ML estimation by maximizing $\log p(\mathbf{y}|\theta)$ is not possible because an explicit expression for the log-likelihood does not exist, or when maximization over θ is an extremely difficult task. The EM algorithm is then employed to compute an update θ^{i+1}, knowing that the final estimate depends on the initial value θ^0. As the symbol vector **x** is missing, the log-likelihood function is replaced by the mathematical expectation over **x**, given the observed data and the current parameter value. The algorithm proceeds in two steps [3][9]

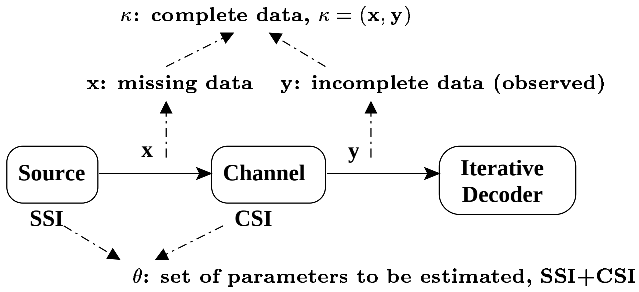

Fig. 3. General model of EM source-channel estimation.

- E-step: Compute *the Auxiliary function* $\mathcal{Q}$:

$$\begin{aligned} \mathcal{Q} &= \mathbf{E}[\log\, p(\mathbf{x},\mathbf{y}|\theta)|\mathbf{y},\theta^i] \\ &= \sum_{\mathbf{x}} log[P(\mathbf{y}|\mathbf{x},\theta)\, P(\mathbf{x}|\theta)]\, APP_i(\mathbf{x}) \end{aligned} \tag{6}$$

- M-step:

$$\theta^{i+1} = \arg\max_{\theta} \mathcal{Q}(\theta|\theta^i) \tag{7}$$

The auxiliary function is a sum over the codewords of a product containing channel observations (CSI), the source distribution (SSI), and *a posteriori* information provided by the decoder from the previous iteration. Afterwards, the decoder exploits updated source and channel statistics provided by the estimation module to enhance its decisions. The SSI part in the auxiliary function is

$$P(\mathbf{x}|\theta) \equiv P(\mathbf{s}|\theta) = \mu^{\omega_H(\mathbf{s})}\,(1-\mu)^{k-\omega_H(\mathbf{s})} \tag{8}$$

where $\omega_H(\mathbf{s})$ denotes the Hamming weight of **s**.

IV. EM Application to BECBSC

Let us consider the binary symmetric channel with erasures as introduced in section II. Let **e** denote the channel error vector and **E** the channel erasure vector. The channel observation is written as

$$P(\mathbf{y}|\mathbf{x},\theta) = (\lambda)^{\omega_H(e)}\,(1-\lambda-\varepsilon)^{n-\omega_H(e)-\omega_H(E)}\,\varepsilon^{\omega_H(E)} \tag{9}$$

A. Expectation step on BECBSC

We write the logarithm of the joint distribution function of parameters λ, ε and μ, by combining (8) and (9). The conditional quantity $P(\mathbf{x}|\mathbf{y},\theta^i)$ is equal to the codeword *a* *posteriori* probability at iteration i. After some algebraic manipulations, the auxiliary function becomes

$$\begin{aligned}\mathcal{Q}(\theta|\theta^i) &= \log[\frac{\lambda}{1-\lambda-\varepsilon}]\mathcal{E}_x[\omega_H(e)] + \log[\frac{\varepsilon}{1-\lambda-\varepsilon}]\mathcal{E}_x[\omega_H(E)] \\ &+ \log[\frac{\mu}{1-\mu}]\mathcal{E}_x[\omega_H(\mathbf{s})] + k\log[(1-\mu)] \\ &+ n\log[(1-\lambda-\varepsilon)]\end{aligned} \tag{10}$$

where $\mathcal{E}_x[\ ] = \mathcal{E}[\ ]$ denotes mathematical expectation over $\mathbf{x}$. Now let us define soft symbols, soft errors, and soft erasures to further simplify expression (10).

Soft source symbols $\tilde{s_j}$ are identified in the following expectation expression:

$$\mathcal{E}[\omega_H((s))] = \sum_{j=1}^{k} \mathcal{E}[s_j] = \sum_{j=1}^{k} \tilde{s}_j$$

The exact definitions of soft binary source symbols and soft binary code symbols are

$$\tilde{s_j} = \sum_{s_j} s_j \times APP_i(s_j) = APP_i(s_j = 1) \tag{11}$$

Similarly, $\tilde{x_j} = APP_i(x_j)$ and for errors and erasures we have

$$\begin{aligned}\tilde{e_j} &= \mathcal{E}[(1-2\times y_j)^2 (x_j)^{\overline{y_j}} \times (1-x_j)^{y_j}] \\ &= (1-2\times y_j)^2 (\tilde{x_j})^{\overline{y_j}} \times (1-\tilde{x_j})^{y_j}\end{aligned} \tag{12}$$

and

$$E_j = 4\times y_j \times (1-y_j) = \tilde{E_j}$$

Finally, thanks to the above definitions of soft symbols, the auxiliary function becomes

$$\begin{aligned}\mathcal{Q}(\theta|\theta^i) &= \log[\frac{\lambda}{1-\lambda-\varepsilon}]\tilde{e} + \log[\frac{\varepsilon}{1-\lambda-\varepsilon}]\tilde{E} \\ &+ \log[\frac{\mu}{1-\mu}]\tilde{\mathbf{s}} + N\log[(1-\mu)(1-\lambda-\varepsilon)]\end{aligned} \tag{13}$$

B. Maximization step on BECBSC

Derive the auxiliary function with respect to μ, and find the source distribution update rule

$$\mu^{i+1} = \frac{\sum_{j=1}^{k} \tilde{s_j}}{k} \tag{14}$$

Derive the auxiliary function with respect to λ to obtain the following second degree equation

$$n.\lambda^2 - (n + \sum_{j=1}^{n} \tilde{e_j}).\lambda + (1-\varepsilon)\sum_{j=1}^{n} \tilde{e_j} = 0$$

Let Δ denote the discriminator of the previous equation. The update rule for the cross-over probability is then given by

$$\lambda^{i+1} = \frac{n + \sum_{j=1}^{n} \tilde{e_j} \pm \sqrt{\Delta(\varepsilon^i)}}{2.n} \tag{15}$$

Derive the auxiliary function with respect to ε, then

$$n.\varepsilon^2 - (n + \sum_{j=1}^{n} \tilde{E_j}).\varepsilon + (1-\lambda)\sum_{j=1}^{n} \tilde{E_j} = 0$$

In a similar way, the update rule for the channel erasure probability is

$$\varepsilon^{i+1} = \frac{n + \sum_{j=1}^{n} \tilde{E_j} \pm \sqrt{\Delta(\lambda^i)}}{2.n} \tag{16}$$

C. Two simple special cases, BEC and BSC

For a binary erasure channel without errors (BEC), the update rule for the erasure probability reduces to

$$\varepsilon^{i+1} = \frac{\sum_{j=1}^{n} \tilde{E_j}}{n} \tag{17}$$

Also, for a binary symmetric channel without erasures (BSC), the update rule for the cross-over probability reduces to

$$\lambda^{i+1} = \frac{\sum_{j=1}^{n} \tilde{e_j}}{n} \tag{18}$$

Figures 4 and 5 illustrate joint EM estimation and decoding for a non-uniform i.i.d. source on an erasure channel. The source parameter is $\mu = 0.1$. The LDPC code is a systematic regular (3,6) with rate $R_c = 1/2$, and length $n = 2000$ bits. Fig. 4 shows the estimated source parameter versus the number of EM iterations. In all channel conditions, μ converges relatively fast to its exact value. Fig. 5 depicts the bit erasure/error rate performance of the same LDPC code with and without source *a priori* probability. The EM estimation induces no loss. Capacity limits are $\varepsilon_{max} = 0.5$ in absence of EM estimation and $\varepsilon_{max} = 0.68$ when estimating the source distribution.

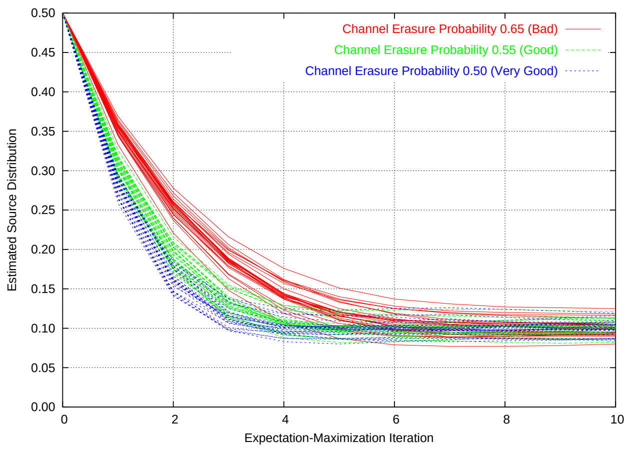

Fig. 4. Estimated source distribution μ versus EM iterations.

V. EM APPLICATION TO AWGN

In this section, we perform EM joint source-channel estimation to enhance decoding of LDPC codes over a complex AWGN channel defined by (5). The EM equations related to the source are identical to the BECBSC case. The observation for a codeword at the decoder input is

$$P(\mathbf{y}|\mathbf{x},\theta) = \frac{1}{(2\pi\sigma^2)^n} exp(\frac{-\sum_{j=1}^{n} \left|y_j - A\, e^{j\phi} x_j\right|^2}{2\sigma^2}) \tag{19}$$

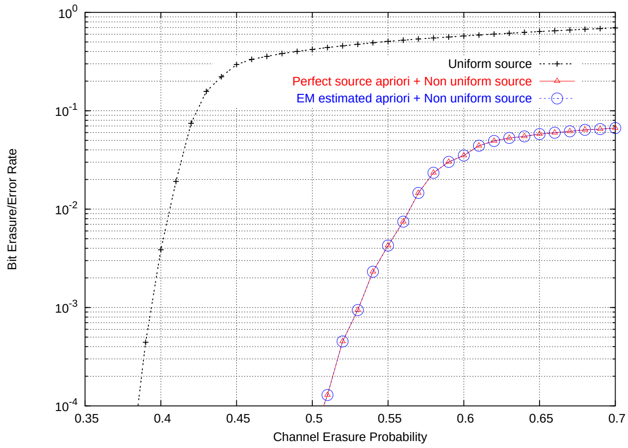

Fig. 5. Performance of a systematic LDPC code on BEC. In absence of source *a priori*, classical BEC decoders are non-probabilistic. Here we use a probabilistic decoder on the BEC in order to embed the EM algorithm.

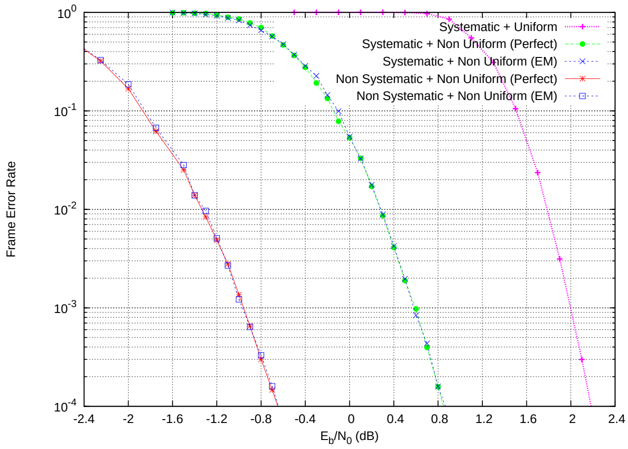

Fig. 6. Systematic and non-systematic LDPC codes on AWGN, $R_c = 1/2$.

A. Expectation step on AWGN

Combine (8) and (19). After some algebraic manipulations, the resulting expression of the auxiliary function on AWGN channel is

$$\begin{aligned}\mathcal{Q}(\theta|\theta^i) &= \log[\frac{\mu}{1-\mu}]\tilde{\mathbf{s}} + k\log[(1-\mu)] - n\log[2\pi\sigma^2] \\ &- \frac{1}{2\sigma^2}\sum_{j=1}^{n}|y_j|^2 - \frac{A^2}{2\sigma^2}\sum_{j=1}^{n}\widetilde{|x_j|^2} \\ &+ \frac{A}{\sigma^2}\sum_{j=1}^{n}\mathcal{R}\left\{\tilde{x_j}^* e^{-\Im\phi} y_j\right\}\end{aligned} \quad (20)$$

where soft complex symbols reduce to $2APP_i(x_j = 1) - 1$ for BPSK alphabet. The notation $\widetilde{|x_j|^2}$ means that conditional expectation must be performed after taking the squared module.

B. Maximization step on AWGN

Derive the auxiliary function with respect to μ to get the same EM update rule for the source distribution as for BECBSC in (14). Then, derive the auxiliary function with respect to the amplitude A, and evaluate high order moments with the product of marginals instead of the joint *a posteriori* probability, to get the amplitude update rule

$$A^{i+1} = \frac{\sum_{j=1}^{n}\mathcal{R}\{\tilde{x_j}y_j^* e^{\Im\phi^i}\}}{\sum_{j=1}^{n}\sum_{x_j} APP_i(x_j)|x_j|^2} = \frac{\sum_{j=1}^{n}\mathcal{R}\{\tilde{x_j}y_j^* e^{\Im\phi^i}\}}{\sum_{j=1}^{n}\widetilde{|x_j|^2}} \quad (21)$$

Finally, the EM update rule for noise variance and phase ambiguity are

$$\sigma^{2\,(i+1)} = \frac{1}{2n}\sum_{j=1}^{n}\widetilde{\left|y_j - A^i e^{\Im\phi^i} x_j\right|^2}$$

$$\phi^{i+1} = -\operatorname{Arg}\sum_{j=1}^{n}\tilde{x}_j\, y_j^*$$

The above expressions for the AWGN channel are well-known in the literature in the case of uncoded modulations and uniform sources.

Fig. 6 illustrates the word error rate performance on AWGN in three different situations for $n = 2000$. A systematic regular (3,6) LDPC code is compared to a non-systematic split-LDPC code with a splitter degree $d_s = 3$ [1]. The non-uniform source has $\mu = 0.1$. The AWGN channel has a uniformly distributed random phase and a random amplitude in the range $0.1 \ldots 10$. A small fraction of 2% pilots is used to initiliaze the phase ambiguity. As shown in Fig. 6, the EM algorithm is capable of attaining the performance of the perfect SSI and CSI knowledge scenario. Similar behavior is observed on BSC.

References

[1] A. Alloum, J.J. Boutros, G.I. Shamir, and L.Wang, "Non-systematic LDPC codes via scrambling and splitting," *43rd Annual Allerton's Conference*, Monticello, Illinois, Sept. 2005.

[2] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, New York: Wiley, 1991.

[3] A.P. Dempster, N.M. Laird and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.*, no. 39, pp. 1–38, 1977.

[4] M. Effros, K. Visweswariah, S.R. Kulkarni, and S. Verdú, "Universal lossless source coding with the Burrows Wheeler transform, "*IEEE Trans. on Inf. Theory*, vol. 48, no. 5, pp. 1061-1081, May 2002.

[5] R.G. Gallager, *Low-density parity-check codes*, MIT Press, 1963.

[6] J. Garcia-Frias and J. D. Villasenor, "Joint turbo decoding and estimation of hidden Markov sources," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 9, pp. 1671-1679, Sept 2001.

[7] F.R. Kschischang, "Codes defined on graphs," *IEEE Communications Magazine*, vol. 41, no. 8, pp. 118-125, Aug. 2003.

[8] D. J. C. MacKay, "Good Error-Correcting Codes based on Very Sparse Matrices, "*IEEE Trans. on Inf. Theory*, vol. 45, no. 2, pp. 399-431, March 1999.

[9] T.K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47-60, Nov. 1996.

[10] G.I. Shamir and J. J. Boutros, "Non systematic Low-Density Parity-Check Codes for Nonuniform Sources," in Proceedings of *IEEE International Symposium on Information Theory*, Adelaide, Australia, pp. 1898-1902, Sept. 2005.

[11] G.C. Zhu, F. Alajaji, J. Bajcsy, and P. Mitran, "Transmission of nonuniform memoryless sources via nonsystematic turbo codes," *IEEE Trans. Communications*, vol. 52, no. 8, pp. 1344-1354, August 2004.