EURASIP Journal on Wireless Communications and Networking 2005:4, 483–492

Asymmetric Joint Source-Channel Coding for Correlated Sources with Blind HMM Estimation at the Receiver

Javier Del Ser
Centro de Estudios e Investigaciones Técnicas de Gipuzkoa (CEIT), Parque Tecnologico de San Sebastián, Paseo Mikeletegi, N48, 20009 Donostia, San Sebastián, Spain
Email: jdelser@ceit.es

Pedro M. Crespo
Centro de Estudios e Investigaciones Técnicas de Gipuzkoa (CEIT), Parque Tecnologico de San Sebastián, Paseo Mikeletegi, N48, 20009 Donostia, San Sebastián, Spain
Email: pcrespo@ceit.es

Olaia Galdos
Centro de Estudios e Investigaciones Técnicas de Gipuzkoa (CEIT), Parque Tecnologico de San Sebastián, Paseo Mikeletegi, N48, 20009 Donostia, San Sebastián, Spain
Email: ogaldos@ceit.es

Received 25 October 2004; Revised 17 May 2005

We consider the case of two correlated sources, S_1 and S_2. The correlation between them has memory, and it is modelled by a hidden Markov chain. The paper studies the problem of reliable communication of the information sent by the source S_1 over an additive white Gaussian noise (AWGN) channel when the output of the other source S_2 is available as side information at the receiver. We assume that the receiver has no *a priori* knowledge of the correlation statistics between the sources. In particular, we propose the use of a turbo code for joint source-channel coding of the source S_1. The joint decoder uses an iterative scheme where the unknown parameters of the correlation model are estimated jointly within the decoding process. It is shown that reliable communication is possible at signal-to-noise ratios close to the theoretical limits set by the combination of Shannon and Slepian-Wolf theorems.

Keywords and phrases: distributed source coding, hidden Markov model parameter estimation, Slepian-Wolf theorem, joint source-channel coding.

1. INTRODUCTION

Communication networks are multiuser communication systems. Therefore, their performance is best understood when viewed as resource sharing systems. In the particular centralized scenario where several users intend to send their data to a common destination (e.g., an access point in a wireless local area network), the receiver may exploit the existing correlation among the transmitters, either to reduce power consumption or gain immunity against noise. In this context, we consider the system shown in Figure 1. The output of two correlated binary sources $\{X_k, Y_k\}_{k=1}^{\infty}$ are separately encoded, and the encoded sequences are sent through two different channels to a joint decoder. The only requirement imposed on the random process $\{X_k, Y_k\}_{k=1}^{\infty}$ is to be ergodic. Notice that this includes the situation where the process $\{X_k, Y_k\}_{k=1}^{\infty}$ is modelled by a hidden Markov model (HMM); this is the case analyzed in this paper.

If the channels are noiseless, the problem is reduced to one of distributed data compression. The Slepian-Wolf theorem [1] (proven to be extensible to ergodic sources in [2]) states that the achievable compression region (see Figure 2) is given by

$$\begin{aligned} \mathcal{R}_1 &\geq \mathcal{H}(S_1 \mid S_2), \\ \mathcal{R}_2 &\geq \mathcal{H}(S_2 \mid S_1), \\ \mathcal{R}_1 + \mathcal{R}_2 &\geq \mathcal{H}(S_1, S_2), \end{aligned} \tag{1}$$

where $\mathcal{R}_1$ and $\mathcal{R}_2$ are the compression rates for sources S_1

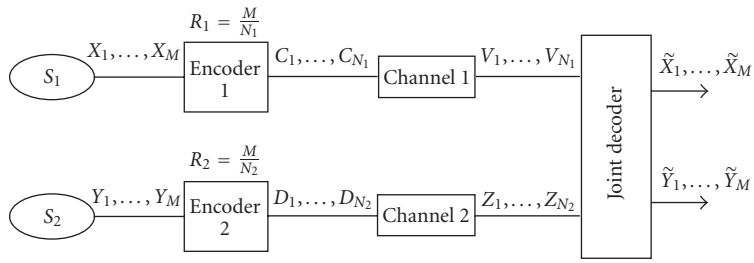

Figure 1: Block diagram of a typical distributed data coding system.

and S_2 (bits per source symbol), and

$$\mathcal{H}(S_1 \mid S_2) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n, \mid Y_1 \ldots, Y_n),$$
$$\mathcal{H}(S_1, S_2) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n; Y_1, \ldots, Y_n), \tag{2}$$

their respective conditional and joint entropy rates. In the particular case where the joint sequence $\{X_k, Y_k\}_{k=1}^{\infty}$ is i.i.d., the above entropy rates are replaced by their corresponding entropies.

As already mentioned, we assume that the output of the multiterminal source $\{X_k, Y_k\}_{k=1}^{\infty}$ can be modelled by a HMM, and we analyze a more general problem of reliable communication when channels 1 and 2 in Figure 1 are additive white Gaussian noise (AWGN) and noiseless, respectively. The main goal is to minimize the energy per information bit E_b sent by the source S_1 for a given encoding rate $R_1 < 1$ and binary phase-shift keying (BPSK) modulation (i.e., the system operates in the power-limited regime). When the complexity of both encoder and decoder is not an issue, the minimum theoretical limit $(E_b/N_0)^*$ is achieved when the source S_1 is compressed at its minimum rate, namely, $\mathcal{H}(S_1 \mid S_2)$. This can be done if the compression rate $\mathcal{R}_2$ of the source S_2 is greater than or equal $\mathcal{H}(S_2)$ (marked point in Figure 2). Without any loss of generality, we can assume that the source S_2 is available as side information at the decoder $(\mathcal{R}_2 = \mathcal{H}(S_2))$.

From the source-channel separation theorem with side information [3], the limit $(E_b/N_0)^*$ is inferred from the condition $C \geq \mathcal{H}(S_1 \mid S_2)R_1$, where $C = (1/2)\log_2(1 + 2E_bR_1/N_0)$ is the capacity of the AWGN channel in bits per channel use.[1] The above condition yields

$$\left(\frac{E_b}{N_0}\right)^* = \frac{2^{2R_1\mathcal{H}(S_1|S_2)} - 1}{2R_1}. \tag{3}$$

Referring to Figure 1, the encoder 1 has been implemented using a binary turbo encoder [4] with coding rate R_1.

[1]Since the modulation scheme used is BPSK, the capacity of the constrained AWGN channel with a binary input constellation should be used instead of the unconstrained channel capacity. However, since the system operates in the power-limited regim the difference between both capacities is small.

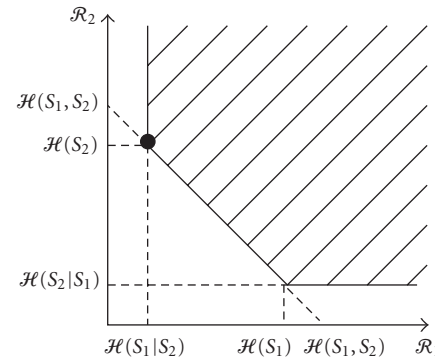

Figure 2: Diagram showing the achievable region for the coding rates. The displayed point $[\mathcal{R}_1 = \mathcal{H}(S_1 \mid S_2),\ \mathcal{R}_2 = \mathcal{H}(S_2)]$ shows the asymmetric compression pair selected in our system.

However, with the corresponding decoding modifications, other type of probabilistic channel codes could have been employed, for example, low-density parity-check (LDPC) codes. The joint decoder bases its decision on both the output of the channel V_k and the side information $Z_k = Y_k$ coming from the source S_2.

The first practical scheme of distributed source compression exploiting the potential of the Slepian-Wolf theorem was introduced by Pradhan and Ramchandran [5]. They focused on the asymmetric case of compression of a source with side information at the decoder and explored the use of simple channel codes like linear block and trellis codes. If this asymmetric compression pair can be reached, the other corner point of the Slepian-Wolf rate region can be approached by swapping the roles of both sources and any point between these two corner points can be realized by time sharing. For that reason, most of the recent works reported in the literature regarding distributed noiseless data compression consider the asymmetric coding problem, although they use more powerful codes such as turbo [6, 7] and LDPC [8, 9] schemes. An exception is [10] that deals with symmetric source compression. In all the above references, except in [9], the correlation between the sources is very simple because they assume that this correlation does not have memory (i.e., $\{X_k, Y_k\}_{k=1}^{\infty}$ is i.i.d. and $P(X_k \neq Y_k) = p\ \forall k$). In [10], the correlation parameter p is estimated iteratively. However, Garcia-Frias and Zhong in [9] consider a much

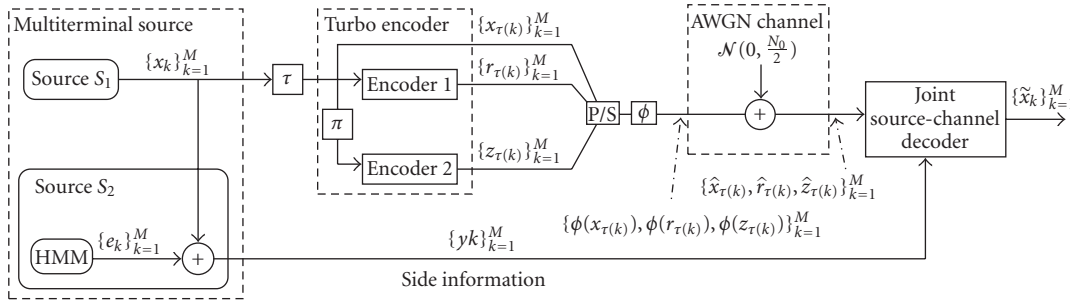

Figure 3: Proposed communication system for the joint source-channel coding scheme with side information. The decoder provides an estimate $\widetilde{x}_k$ of x_k with the help of the side information sequence $\{y_k\}_{k=1}^M$ and the redundant data $\{r_k, z_k\}_{k=1}^M$ computed in the turbo encoder. The interleaver τ decorrelates the output of the sources.

more general model with hidden Markov correlation and assumes that its parameters are known at the decoder.

When one of the channels is noisy, the authors in [11] (for a binary symmetric channel, BSC) and in [12] (for a BSC, AWGN and Rayleigh channel) have proposed a joint source-channel coding scheme based on turbo and irregular repeat accumulate (IRA) codes, respectively. In both cases, the correlation among the sources is again assumed to be memoryless and known at the receiver. Under the same correlation assumptions, the case of symmetric joint source-channel coding when both channels are noisy (AWGN) has been studied using turbo [13] and low-density generator-matrix (LDGM) [14] codes. Both assume that the memoryless correlation probability is known at the decoder.

In this paper, we take a further step and consider that the correlation between the sources follows a hidden Markov model like the correlation proposed in [9] for distributed source compression. However, unlike what is assumed in [9], our proposed scheme does not require any previous knowledge of the HMM parameters. It is based on an iterative scheme that jointly estimates, within the turbo-decoding process, the parameters of the HMM correlation model. It is an extension of the estimation method presented by Garcia-Frias and Villasenor [15] (for point-to-point data transmission over an AWGN of a single HMM source) to the mentioned distributed joint source-channel coding scenario. As we show in the simulation results, the loss in BER performance that results from the blind estimation of the HMM parameters when compared to their perfect knowledge is negligible.

The rest of this paper is organized as follows. In the next section, the proposed system is introduced and the iterative source-channel joint decoder is described. Section 3 discusses the simulation results of the joint decoding scheme. Finally, in Section 4, some concluding remarks are given.

2. SYSTEM MODEL

In this section, we present the proposed joint source-channel encoder shown in Figure 3. It uses an iterative decoding scheme that exploits the hidden Markov correlation between sources based on the side information available at the decoder. After describing the model assumed for the correlated sources, the encoding and decoding process is analyzed. We place a special emphasis on the description of the iterative decoding algorithm by means of factor graphs and the sum-product algorithm (SPA). For an overview about graphical models and the SPA, we refer to [16].

2.1. Joint source model

We assume the following model for the multiterminal source (MS) sequence $\{X_k, Y_k\}_{k=1}^{\infty}$.

(i) The X_k are i.i.d. binary random variables with probability distribution $P(x_k = 1) = P(x_k = 0) = 0.5$.

(ii) The output Y_k from the source S_2 is expressed as $Y_k = X_k \oplus E_k$, where $\bigoplus$ denotes modulus 2 addition, and E_k is a binary random process generated by an HMM with parameters $\{\mathbf{A}, \mathbf{B}, \mathbf{\Pi}\}$. The model is characterized by [17]
   (1) the number of states P;
   (2) the state-transition probability distribution $\mathbf{A} = [a_{s,s'}]$, where $a_{s,s'} = P_{S_k^{\mathrm{MS}} | S_{k-1}^{\mathrm{MS}}}(s' \mid s)$, $s, s' \in \{0, \ldots, P-1\}$;
   (3) the observed symbol probabilities distribution $\mathbf{B} = [b_{s,e}]$, where $b_{s,e} = P_{E_k | S_k^{\mathrm{MS}}}(e \mid s)$, $s \in \{0, \ldots, P-1\}$, and $e \in \{0, 1\}$;
   (4) the initial-state distribution $\mathbf{\Pi} = \{\pi_s\}$, where $\pi_s = P_{S_0^{\mathrm{MS}}}(s)$ and $s \in \{0, \ldots, P-1\}$.

We may note that for this model, the outputs of both sources S_1 and S_2 are i.i.d. and equiprobable. Thus, $\mathcal{H}(S_1) = H(X_1) = 1$ and $\mathcal{H}(S_2) = H(Y_1) = 1$. On the contrary, the correlation between sources does have memory since

$$\begin{aligned} \mathcal{H}(S_1 \mid S_2) &= \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n \mid Y_1, \ldots, X_n) \\ &= \lim_{n\to\infty} \frac{1}{n} H(E_1, \ldots, E_n) = \mathcal{H}(E) < H(E_1), \end{aligned} \tag{4}$$

where $\mathcal{H}(E)$ denotes the entropy rate of the random sequence E_k generated by the HMM. By changing the parameters of the HMM, different values of $\mathcal{H}(S_1 \mid S_2)$ can be obtained. Also notice that, for the particular case where $P = 1$, the correlation is memoryless, resulting in $\mathcal{H}(S_1 \mid S_2) = H(E_1) = h(b_{0,1})$; that is, the entropy of a binary random variable with distribution $(b_{0,1}, 1 - b_{0,1})$.

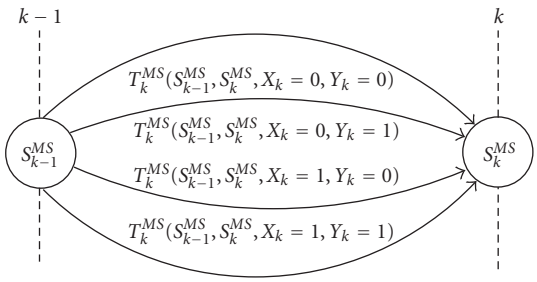

Figure 4: Branch transition probabilities from the generic state S_{k-1}^{MS} to S_k^{MS} of the trellis describing the HMM multiterminal source.

Using the fact that $Y_k = X_k \oplus E_k$, the above model can be reduced to an equivalent HMM that outputs directly the joint sequence $\{X_k, Y_k\}_{k=1}^{\infty}$ without any reference to the variable E_k. Its trellis diagram has P states and 4 parallel branches between states, one for each possible output (X_k, Y_k) combination (see Figure 4). The associated branch *a priori* probabilities are easily obtained from the original HMM model and the X_k *a priori* probabilities $P(x_k)$. For instance, the branch probability of going from state s to state s', associated with outputs $X_k = q$ and $Y_k = v$, $q \neq v$ $(q = v)$, is given by the probability of the following three independent events $\{S_{k-1} = s,\ S_k = s'\}$, $\{E_k = 1,$ when being in state $s\}$ ($\{E_k = 0,$ when being in state $s\}$), and $\{X_k = q\}$; that is, $a_{s,s'} \cdot b_{s,1} \cdot P(x_k = q)$. Therefore,

$$T^{\mathrm{MS}}(S_{k-1}^{\mathrm{MS}} = s,\ S_k^{\mathrm{MS}} = s',\ X_k = q,\ Y_k = v) = \begin{cases} a_{s,s'} \cdot b_{s,0} \cdot 0.5 & \text{if } q = v, \\ a_{s,s'} \cdot b_{s,1} \cdot 0.5 & \text{if } q \neq v, \end{cases} \tag{5}$$

where $q, v \in \{0, 1\}$ and $s, s' \in \{0, \ldots, P-1\}$. The MS label for the trellis branch transitions T_k^{MS} and state variables S_k^{MS} stands for multiterminal source.

2.2. Turbo encoder

The block sequence $\{x'_k\}_{k=1}^{M} = \{X_1 = x'_1, \ldots, X_M = x'_M\}$ produced by a realization of the source S_1 is first randomized by the interleaver τ before entering to a turbo code, with two identical constituent convolutional encoders C_1 and C_2. The encoded binary sequence is denoted by $\{x'_{\tau(k)}, r'_{\tau(k)}, z'_{\tau(k)}\}_{k=1}^{M}$, where we assume that the coding rate is $R_1 = 1/3$, and $r'_{\tau(k)}, z'_{\tau(k)}$ are the redundant symbols produced by C_1, C_2, respectively. The input to the AWGN channel is $\{\phi(x'_{\tau(k)}), \phi(r'_{\tau(k)}), \phi(z'_{\tau(k)})\}_{k=1}^{M}$, where $\phi : \{0, 1\} \to \mathbb{R}$ denotes the BPSK transformation performed by the modulator. Finally, the received corresponding sequence will be denoted by $\{\hat{x}_{\tau(k)}, \hat{r}_{\tau(k)}, \hat{z}_{\tau(k)}\}_{k=1}^{M}$.

2.3. Joint source-channel decoder

To better understand the joint source-channel decoder with side information, we begin analyzing a simplified decoder that bases its decisions only on

(i) the received systematic symbols $\{\hat{x}_k\}_{k=1}^{M}$;

(ii) the side information sequence $\{\hat{y}_k\}_{k=1}^{M}$ generated by a realization of the source S_2.

The decoder will decide for the $X_k \in \{0, 1\}$ that maximizes the *a posteriori* probability $P(x_k \mid \{\hat{x}_j, \hat{y}_j\}_{j=1}^{M})$ (MAP decoder). This is done via the forward-backward algorithm, also known as MAP or BCJR [18]. This algorithm is a particularization of the SPA applied to factor graphs derived from an HMM or a trellis diagram, and it is an efficient marginalization procedure based on message-passing rules among the nodes in a factor graph.

From the trellis description of our source model (see Figure 4), the joint probability distribution function of the random variables $\{X_k\}_{k=1}^{M}$ conditioned by the observations $\{\hat{x}_j\}_{j=1}^{M}$ and the side information $\{\hat{y}_j\}_{j=1}^{M}$, that is, $P(x_1, \ldots, x_M \mid \{\hat{x}_j, \hat{y}_j\}_{j=1}^{M})$, can be decomposed in terms of factors, one for each time instant k. In turn, this factorization may be represented by a factor graph [16], like the one shown in Figure 5. We keep the same convention used in [16], representing in lower case the variables involved in a factor graph. There should be no confusion from the context whether x denotes an ordinary variable taking on values in some finite alphabet $\mathcal{X}$, or the realization of some random variable X.

Since the channel is AWGN, the local functions of x_k, $P(\hat{x}_k \mid x_k)$, are given by the Gaussian distribution $\mathcal{N}(\phi(x_k), N_0/2)$. On the other hand, the local functions $I_{\hat{y}_k}(y_k)$ are indicator functions taking value 1 when $y_k = \hat{y}_k$ and 0 otherwise. This shows the fact that the output of the source S_2 is known with certainty at the decoder.

Based on this factor graph, the decoder can now efficiently compute the *a posteriori* probability $P(x_k \mid \{\hat{x}_j, \hat{y}_j\}_{j=1}^{M})$ by marginalizing $P(x_1, \ldots, x_M \mid \{\hat{x}_j, \hat{y}_j\}_{j=1}^{M})$ via the SPA which, in this case, reduces to the forward-backward algorithm.

In particular, the forward and backward recursion parameters $\alpha_{k-1}^{\mathrm{MS}}(s_{k-1}^{\mathrm{MS}})$ and $\beta_k^{\mathrm{MS}}(s_k^{\mathrm{MS}})$ defined in the forward-backward algorithm are the messages passed from the state variable node s_{k-1}^{MS} to the factor node T_k^{MS} and from the state variable node s_k^{MS} to T_k^{MS}, respectively. From the sum-product update rules, the following expressions are obtained for these messages:

$$\alpha_k^{\mathrm{MS}}(s_k) = \sum_{\sim\{s_k\}} \alpha_{k-1}^{\mathrm{MS}}(s_{k-1}) \cdot T_k^{\mathrm{MS}}(s_{k-1}, s_k, x_k, y_k) \cdot P(\hat{x}_k \mid x_k) \cdot I_{\hat{y}_k}(y_k), \quad k = 1, \ldots, M, \tag{6}$$

$$\beta_k^{\mathrm{MS}}(s_k) = \sum_{\sim\{s_k\}} \beta_{k+1}^{\mathrm{MS}}(s_{k+1}) \cdot T_{k+1}^{\mathrm{MS}}(s_k, s_{k+1}, x_{k+1}, y_{k+1}) \cdot P(\hat{x}_{k+1} \mid x_{k+1}) \cdot I_{\hat{y}_{k+1}}(y_{k+1}), \quad k = M-1, \ldots, 1, \tag{7}$$

where $x_k, y_k \in \{0, 1\}$, $s_{k-1}, s_k \in \{0, \ldots, P-1\}$, and $\sum_{\sim\{s_k\}}$ indicates that all variables are being summed over except variable s_k. The subindex MS in the state variables has been omitted for clarity's sake. The initialization is done by setting $\alpha_0^{\mathrm{MS}}(j) = \pi_j$ and $\beta_M^{\mathrm{MS}}(j) = 1/P$, for all $j \in \{0, \ldots, P-1\}$. Once the $\alpha_k^{\mathrm{MS}}(s_k)$ and $\beta_k^{\mathrm{MS}}(s_k)$ have been computed, the messages $\delta_k^{\mathrm{MS}}(x_k)$, passed from the factor nodes

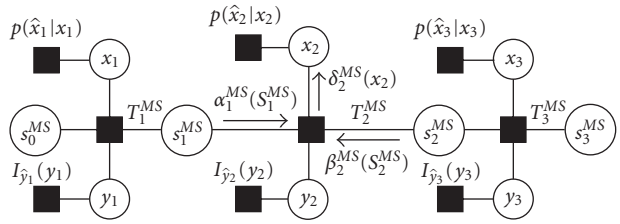

Figure 5: Simplified factor graph defined by the trellis of Figure 4. For simplicity, only $M = 3$ stages has been drawn.

$T_k^{\mathrm{MS}}(s_{k-1}^{\mathrm{MS}}, s_k^{\mathrm{MS}}, x_k, y_k)$ to the variable nodes x_k, are obtained by the SPA update rules as

$$\delta_k^{\mathrm{MS}}(x_k) = \sum_{\sim\{x_k\}} \alpha_{k-1}^{\mathrm{MS}}(s_{k-1}) \cdot T_k^{\mathrm{MS}}(s_{k-1}, s_k, x_k, y_k) \cdot \beta_k^{\mathrm{MS}}(s_k) \cdot I_{\hat{y}_k}(y_k), \quad k = 1, \ldots, M. \tag{8}$$

The *a posteriori* probability $P(x_k \mid \{\hat{x}_j, \hat{y}_j\}_{j=1}^M)$ is now calculated as the product of all the messages arriving at variable node x_k. In our case, the message passed from the local function node $P(\hat{x}_k \mid x_k)$ to the variable node x_k is simply the probability function itself, whereas the message passed from the local function node $T_k^{\mathrm{MS}}(s_{k-1}^{\mathrm{MS}}, s_k^{\mathrm{MS}}, x_k, y_k)$ to the variable node x_k is $\delta_k^{\mathrm{MS}}(x_k)$ (see Figure 5). Therefore,

$$P\left(x_k \mid \{\hat{x}_j, \hat{y}_j\}_{j=1}^M\right) \propto P(\hat{x}_k \mid x_k) \cdot \delta_k^{\mathrm{MS}}(x_k). \tag{9}$$

The problem we want to solve in this paper is an extension of what we have just analyzed. The joint decoder must compute the *a posteriori* probability of the symbol X_k by observing not only the corresponding received symbols $\{\hat{x}_j\}_{j=1}^M$ and the side information $\{\hat{y}_j\}_{j=1}^M$ as described before, but also the additional outputs of the channel $\{\hat{r}_j\}_{j=1}^M$ and $\{\hat{z}_j\}_{j=1}^M$, that is, $P(x_k \mid \{\hat{x}_j, \hat{r}_j, \hat{z}_j, \hat{y}_j\}_{j=1}^M)$. The global factor graph results by properly attaching, through interleaver τ, the factor graph describing a standard turbo decoder to the graph in Figure 5.

Figure 6 shows this arrangement. Observe that the three sub-factor graphs have the same topology since each models a trellis (with different parameters); namely, the trellis of the two constituent convolutional decoders and the trellis of the multiterminal source.

Similarly to what happens with the standard factor graph of a turbo decoder, the compound factor graph has cycles and the message sum-product algorithm has no natural termination. To overcome this problem, the following schedule has been adopted. During the ith iteration, a standard SPA is separately applied to each of the three factor graphs describing the decoders $D1$, $D2$, and the multiterminal source, in this order: MS $\rightarrow$ $D1$ $\rightarrow$ $D2$. Since these subfactor graphs do not have cycles, the corresponding SPAs will terminate. Notice, however, that the updating rules for the SPA, when applied to one of the subfactor graphs, require incoming messages from the other two subfactor graphs (called *extrinsic* information in turbo-decoding jargon), since all share the same variable nodes $x_{\tau(k)}$. The messages computed in the previous steps are used for that purpose.

For example, referring to Figure 6, the former SPA update expressions (see (6)–(8)) are now modified to include the *extrinsic* information $\xi_{k,i}^{\mathrm{MS}}(x_k)$ coming from $D1$ and $D2$ (i.e., from the turbo-decoding iteration), instead of $P(\hat{x}_k \mid x_k)$. That is,

$$\alpha_{k,i}^{\mathrm{MS}}(s_k) = \sum_{\sim\{s_k\}} \alpha_{k-1,i}^{\mathrm{MS}}(s_{k-1}) \cdot T_k^{\mathrm{MS}}(s_{k-1}, s_k, x_x, y_k) \cdot \xi_{k,i}^{\mathrm{MS}}(x_k) \cdot I_{\hat{y}_k}(y_k), \quad k = 1, \ldots, M, \tag{10}$$

$$\beta_{k,i}^{\mathrm{MS}}(s_k) = \sum_{\sim\{s_k\}} \beta_{k+1,i}^{\mathrm{MS}}(s_{k+1}) \cdot T_{k+1}^{\mathrm{MS}}(s_k, s_{k+1}, x_{k+1}, y_{k+1}) \cdot \xi_{k+1,i}^{\mathrm{MS}}(x_{k+1}) \cdot I_{\hat{y}_{k+1}}(y_{k+1}), \quad k = M-1, \ldots, 1, \tag{11}$$

$$\delta_{k,i}^{\mathrm{MS}}(x_k) = \sum_{\sim\{x_k\}} \alpha_{k-1,i}^{\mathrm{MS}}(s_{k-1}) \cdot T_k^{\mathrm{MS}}(s_{k-1}, s_k, x_x, y_k) \cdot \beta_{k,i}^{\mathrm{MS}}(s_k) \cdot I_{\hat{y}_k}(y_k), \quad k = 1, \ldots, M, \tag{12}$$

where the subindex i denotes the current iteration. The *extrinsic* information $\xi_{k,i}^{\mathrm{MS}}(x_k)$ is the message passed from the variable node x_k to the factor node T_k^{MS} through interleaver τ (see Figure 6). Using the SPA update rules, this is given by

$$\xi_{k,i}^{\mathrm{MS}}(x_k) = \delta_{k,i-1}^{D1}(x_k) \cdot \delta_{k,i-1}^{D2}(x_k) \cdot P(\hat{x}_k \mid x_k), \quad k = 1, \ldots, M. \tag{13}$$

With the obvious modifications, the same set of recursions also holds for the factor graphs $D1$ and $D2$. Observe that the SPA applied to $D1$ and $D2$ is nothing more than the standard turbo-decoding procedure modified to include the *extrinsic* information $\delta_{k,i-1}^{\mathrm{MS}}(x_k)$ coming from the MS.

After L iterations, the *a posteriori* probabilities $P(x_{\tau(k)} \mid \{\hat{x}_j, \hat{r}_j, \hat{z}_j, \hat{y}_j\}_{j=1}^M)$ are calculated as the product of all messages arriving at variable node $x_{\tau(k)}$, that is,

$$P\left(x_{\tau(k)} \mid \{\hat{x}_j, \hat{r}_j, \hat{z}_j, \hat{y}_j\}_{j=1}^M\right) \propto \delta_{\tau(k),L}^{D1}(x_{\tau(k)}) \cdot \delta_{\tau(k),L}^{D2}(x_{\tau(k)}) \cdot \delta_{\tau(k),L}^{\mathrm{MS}}(x_{\tau(k)}) \cdot P(\hat{x}_{\tau(k)} \mid x_{\tau(k)}), \quad k = 1, \ldots, M. \tag{14}$$

Finally, the estimated source symbol at $\tau(k)$ is given by $\arg\max_{x_{\tau(k)} \in \{0,1\}} P(x_{\tau(k)} \mid \{\hat{x}_j, \hat{r}_j, \hat{z}_j, \hat{y}_j\}_{j=1}^M)$.

If the local functions $I_{\hat{y}_k}(y_k)$ in the factor nodes of Figure 6 were substituted by $P(y_k) = 0.5$ (i.e., if no side information was available at the decoder or the sources were not correlated), the resulting normalized messages from the SPA would be $\delta^{\mathrm{MS}}_{k,i}(x_k) = 0.5$ for all k, i and all values of variables x_k (showing the fact that the source S_1 is i.i.d. and equiprobable). In other words, the subfactor graph of the MS would be superfluous and the decoder would be reduced to a standard turbo decoder. Should we assume for S_1 (see Figure 3) a two-state HMM source, like the one considered in [15] instead of i.i.d., the resulting MS overall HMM, combining both HMM models (for $\{E_k\}$ and $\{X_k\}$), would have $2P$ states with 4 branches between states. The corresponding branch probabilities in (5) would have to be modified accordingly. In the lack of side information, the MS factor graph would be reduced to that describing the HMM of the source S_1. As a result, our decoding process would coincide with the scheme studied in [15].

2.4. Iterative estimation of the HMM parameters of the multiterminal source model

The updating equations (10)–(12) require the knowledge of the HMM parameters $\{\mathbf{A}, \mathbf{B}, \mathbf{\Pi}\}$, since they appear in the definition of the branch transition probabilities in (5). However, in most cases, this information is not available. Therefore, the joint decoder must additionally estimate these parameters. The proposed estimation method is based on a modification of the iterative Baum-Welch algorithm (BWA) [17], which was first applied in [15] to estimate the parameters of hidden Markov source in a point-to-point transmission scenario. The underlying idea is to use the BWA over the trellis associated with the multiterminal source by reusing the SPA messages computed at each iteration.

For the derivation of the reestimation formulas, it is convenient to define the functions $a_i(s, s')$, $b_i(s, e)$, and $\pi_i(s)$, where s, s', and e are variables taking on values in $\{0, \ldots, P-1\}$ and $\{0, 1\}$, respectively. The index i denotes the iteration number and the values taken by these functions at iteration i are the reestimated distributions of the probability of going from state s to state s', the probability that the HMM outputs the symbol e when being in state s, and the probability that the initial state of the HMM is s, respectively. With this new notation, the local functions $T^{\mathrm{MS}}_k(s_{k-1}, s_k, x_k, y_k, e_k)$ (5) in the MS factor graph will now depend on i, yielding

$$T^{\mathrm{MS}}_{k,i}(s_{k-1}, s_k, x_k, y_k, e_k) = \begin{cases} a_{i-1}(s_{k-1}, s_k) \cdot b_{i-1}(s_k, 0) \cdot 0.5 & \text{if } x_k = y_k,\ e_k = 0, \\ a_{i-1}(s_{k-1}, s_k) \cdot b_{i-1}(s_k, 1) \cdot 0.5 & \text{if } x_k \neq y_k,\ e_k = 1, \\ 0 & \text{elsewhere.} \end{cases} \tag{15}$$

Notice that the variable e_k is explicitly included in the argument of $T^{\mathrm{MS}}_{k,i}$ since the access to this variable is required when obtaining the reestimation formula for $b_i(s, e)$ (17).

Having said that, the reestimation expressions for these functions are easily derived by realizing that the conditional probability $P(s_{k-1}, s_k, x_k, y_k, e_k \mid \{\hat{x}_j, \hat{r}_j, \hat{z}_j, \hat{y}_j\}^M_{j=1})$ at iteration i is proportional to the product $\alpha^{\mathrm{MS}}_{k-1,i}(s_{k-1}) \cdot T^{\mathrm{MS}}_{k,i}(s_{k-1}, s_k, x_k, y_k, e) \cdot \beta^{\mathrm{MS}}_{k,i}(s_k) \cdot \xi^{\mathrm{MS}}_{k,i}(x_k) \cdot I_{\hat{y}_k}(y_k)$. Using this fact on the BWA, the following reestimation equations are obtained:

$$a_i(s, s') = \frac{\sum_{k=1}^{M} \sum_{\sim\{s,s'\}} \alpha^{\mathrm{MS}}_{k-1,i}(s) \cdot T^{\mathrm{MS}}_{k,i}(s, s', x_k, y_k, e) \cdot \beta^{\mathrm{MS}}_{k,i}(s') \cdot \xi^{\mathrm{MS}}_{k,i}(x_k) \cdot I_{\hat{y}_k}(y_k)}{\sum_{k=1}^{M} \sum_{\sim\{s\}} \alpha^{\mathrm{MS}}_{k-1,i}(s) \cdot T^{\mathrm{MS}}_{k,i}(s, s', x_k, y_k, e) \cdot \beta^{\mathrm{MS}}_{k,i}(s') \cdot \xi^{\mathrm{MS}}_{k,i}(x_k) \cdot I_{\hat{y}_k}(y_k)}, \tag{16}$$

$$b_i(s, e) = \frac{\sum_{k=1}^{M} \sum_{\sim\{s,e\}} \alpha^{\mathrm{MS}}_{k-1,i}(s) \cdot T^{\mathrm{MS}}_{k,i}(s, s', x_k, y_k, e) \cdot \beta^{\mathrm{MS}}_{k,i}(s') \cdot \xi^{\mathrm{MS}}_{k,i}(x_k) \cdot I_{\hat{y}_k}(y_k)}{\sum_{k=1}^{M} \sum_{\sim\{s\}} \alpha^{\mathrm{MS}}_{k-1,i}(s) \cdot T^{\mathrm{MS}}_{k,i}(s, s', x_k, y_k, e) \cdot \beta^{\mathrm{MS}}_{k,i}(s') \cdot \xi^{\mathrm{MS}}_{k,i}(x_k) \cdot I_{\hat{y}_k}(y_k)}, \tag{17}$$

$$\pi_i(s) = \frac{\sum_{\sim\{s\}} \alpha^{\mathrm{MS}}_{0,i}(s) \cdot T^{\mathrm{MS}}_{1,i}(s, s', x_1, y_1, e) \cdot \beta^{\mathrm{MS}}_{1,i}(s') \cdot \xi^{\mathrm{MS}}_{1,i}(x_1) \cdot I_{\hat{y}_1}(y_1)}{\sum_{\sim\{\varnothing\}} \alpha^{\mathrm{MS}}_{0,i}(s) \cdot T^{\mathrm{MS}}_{1,i}(s, s', x_1, y_1, e) \cdot \beta^{\mathrm{MS}}_{1,i}(s') \cdot \xi^{\mathrm{MS}}_{1,i}(x_1) \cdot I_{\hat{y}_1}(y_1)}. \tag{18}$$

The $\sum_{\sim\{\varnothing\}}$ in the denominator of (18) indicates that all variables are summed over. At iteration i, the above expressions are computed after the SPA has been applied to MS, $D1$, and $D2$. We have noticed that (18) may be omitted whenever the block length is large enough (the initial $\alpha^{\mathrm{MS}}_{0,i}(j)$ can be set to $1/P$ for all $j \in \{0, \ldots, P-1\}$). We now give a brief summary of the proposed iterative decoding scheme.

(i) Phase I: $i = 0$.

(1) Perform the SPA over the factor graphs that describe the decoders $D1$ and $D2$ without considering the *extrinsic* information coming from the MS block (i.e., with $\delta^{\mathrm{MS}}_{k,0}(x_k) = 0.5$, for all $k \in \{1, \ldots, M\}$). For each k, obtain an initial estimate $\widetilde{x}_k$ of the source symbol x_k by $\widetilde{x}_k = \arg\max_{x_k \in \{0,1\}} P(x_k \mid \{\hat{x}_j, r_j, z_j\}^M_{j=1})$. Notice that this is equivalent to considering only the turbo decoder.

(2) Based on the observation $\widetilde{e}_k = \widetilde{x}_k \oplus \hat{y}_k$, apply the standard BWA [17] to obtain an initial estimate of the Markov parameters $a_0(s, s')$, $b_0(s, e)$, and $\pi_0(s)$, $e \in \{0, 1\}$, $s, s' \in \{0, \ldots, P-1\}$.

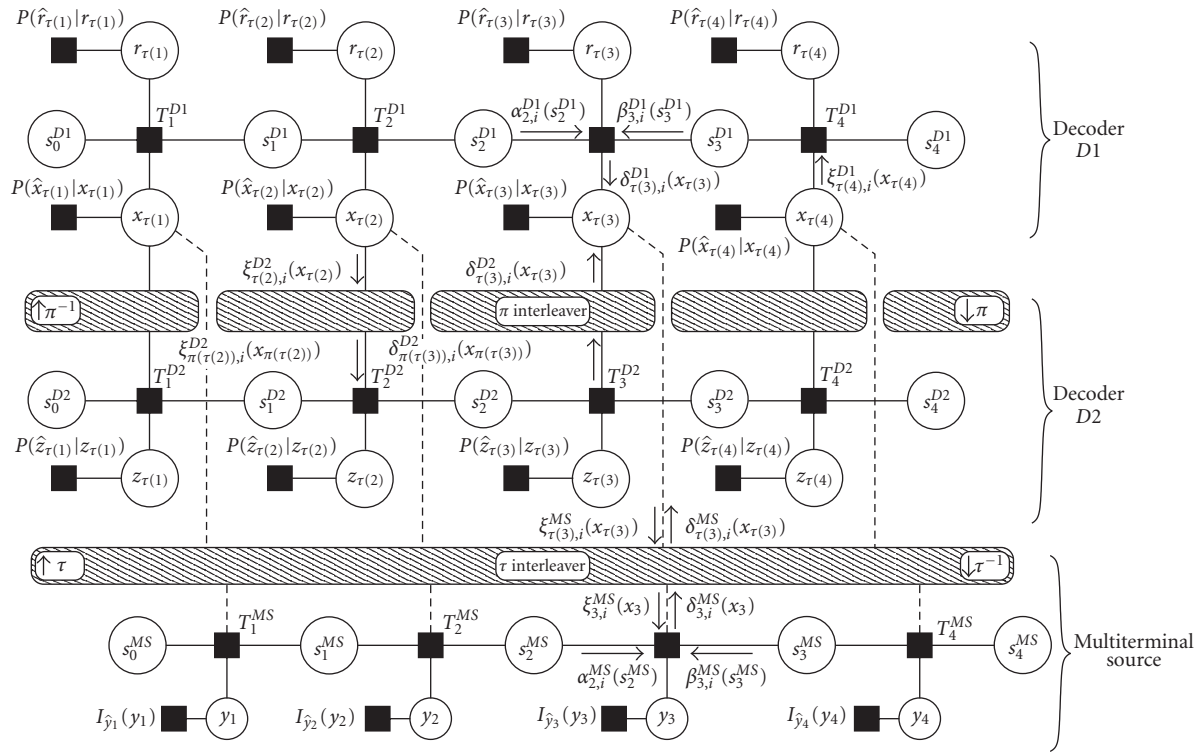

Figure 6: Assembly of the standard turbo decoder to the factor graph in Figure 5. For simplification purposes, the data length has been fixed to $M = 4$.

(ii) Phase II: $i \geq 1$.
  (3) $i = i + 1$.
  (4) Perform the SPA over the MS factor graph using the functions $T^{\mathrm{MS}}_{k,i}$ in (15) as factor nodes. This will produce the set of messages $\delta^{\mathrm{MS}}_{k,i}(x_k)$.
  (5) Perform the SPA over the factor graphs $D1$ and $D2$ with messages $\delta^{\mathrm{MS}}_{k,i}(x_k)$ as *extrinsic* information coming from the factor graph MS.
  (6) Reestimate the HMM parameters using (16)–(18), and go back to step 3.

3. SIMULATION RESULTS

In order to assess the performance of the proposed joint decoding/estimation scheme, a simulation has been carried out using different values of the conditional entropy rate $\mathcal{H}(S_1 \mid S_2)$. The two constituent convolutional encoders C_1 and C_2 of the turbo code are characterized by the polynomial generator $g(Z) = [1, (Z^3 + Z^2 + Z + 1)/(Z^3 + Z^2 + 1)]$. In all simulated cases, the number of states P for the HMM characterizing the joint source correlation has been set to 2. Performance comparisons with and without the decoder having *a priori* knowledge of the hidden Markov parameters are presented.

The simulation uses 2000 blocks of 16384 binary symbols each, and the maximum number of iterations is fixed to 35. Figure 7 displays the bit error ratio (BER) versus E_b/N_0 for two different values of the conditional entropy rate, $\mathcal{H}(S_1 \mid S_2) = 0.45$ and 0.73, and for the rate 1/3 standard turbo decoder. The HMM model that generates the stationary random process E_k, giving raise to $\mathcal{H}(S_1 \mid S_2) = 0.45$ (0.73), has transition probabilities $a_{0,0} = 0.97$ (0.9), $a_{1,1} = 0.98$ (0.85) and output probabilities $b_{0,0} = 0.05$ (0.05), $b_{1,0} = 0.95$ (0.92). In both cases, the initial-state distribution Π is the corresponding stationary distribution of the chain.

As opposed to what happens to the joint probability distribution of $(E_1, \ldots, E_n)$, the marginal distribution $P_{E_k}(e_k)$ is easily computed by $P_{E_k}(e_k) = \pi_1 \cdot b_{1,e_k} + \pi_0 \cdot b_{0,e_k}$, for all k. It can be checked that in both models this distribution is nearly equiprobable, giving a value for the entropy $H(E_k)$ of approximately 0.98. Since $H(X_k \mid Y_k) = H(E_k) \approx H(X_k)$, we have that $P_{X_k|Y_k}(x_k \mid y_k) \approx P_{X_k}(x_k)$, that is, the random variables X_k and Y_k are practically independent. Therefore, the correlation between the processes $\{X_k\}_{k=1}^{\infty}$ and $\{Y_k\}_{k=1}^{\infty}$ is embedded in the memory of the joint process $\{X_k, Y_k\}_{k=1}^{\infty}$ (see (4)).

The standard turbo-decoder curve has been included in Figure 7 for reference. It shows the performance degradation that the proposed joint decoder would incur, should the side information not be used in the decoding algorithm (or, equivalently, if no correlation exists between both sources, i.e., $\mathcal{H}(S_1 \mid S_2) = \mathcal{H}(S_1) = 1$).

For comparison purposes, the three theoretical limits -0.55, -2.2, and -4.6 dB given in (3) corresponding to $\mathcal{H}(S_1 \mid S_2) = 1$, 0.73, and 0.45, respectively, are also shown as vertical lines in Figure 7. For $\mathcal{H}(S_1 \mid S_2) = 0.73$ and

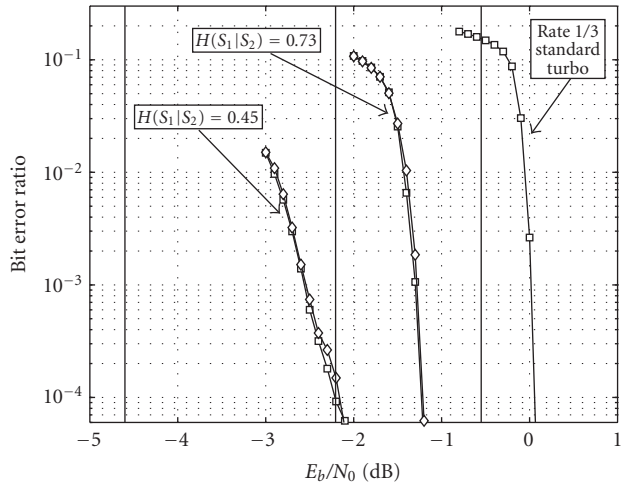

FIGURE 7: BER versus E_b/N_0 for entropy values $H(S_1 \mid S_2) = 1.0, 0.73$, and 0.45 after 35 iterations. The results for known and unknown HMM are depicted with □ and ◇ markers, respectively. The theoretical Shannon limits are represented by the vertical solid lines. The BER range is bounded at $1/M$ (less than one error in $M = 16384$ bits).

$\mathcal{H}(S_1 \mid S_2) = 0.45$, the BER curves with □ markers represent the performance when perfect knowledge of the joint source parameters is available at the decoder. On the other hand, the curves with ◇ display the performance when no initial knowledge is available at the joint decoder. In this case, the estimation of the HMM parameters is run afresh for each input block, that is, without relying on any previous reestimation information.

Observe that the degradation in performance due to the lack of *a priori* knowledge in the source correlation statistics is negligible. Also we may note that at a given BER, the gap between the required E_b/N_0 and their corresponding theoretical limits widens as the conditional entropy rate decreases (i.e., the amount of correlation between sources increases). In particular, at BER $= 10^{-4}$, the gaps are 0.65, 1 and 2.4 dB, respectively. As mentioned in [13] for the memoryless case, when the correlation between the sequences is very strong the side information can be interpreted as an additional systematic output of the turbo decoder. As it is well known in the turbo-code literature, this repetition involves a penalty in performance.

The set of curves in Figure 8 illustrates the BER performance versus E_b/N_0 as the number of iterations increases. Plots 8a and 8b are for the conditional entropy rates $\mathcal{H}(S_1 \mid S_2) = 0.45$ and $\mathcal{H}(S_1 \mid S_2) = 0.73$, respectively. Although the BER performance is similar in both cases, the convergence rate when the decoder estimates the parameters of the HMM is slower, as expected.

Finally, suppose that the joint decoder is implemented assuming that the correlation between sources is memoryless (like in [13]), that is, the state variables in the MS factor graph can only take a single value $s_k = 0$, and the factor nodes T_k^{MS} in (5) have $a_{0,0} = 1$ and $b_{0,0} = P_{E_k}(0)$. As a result,

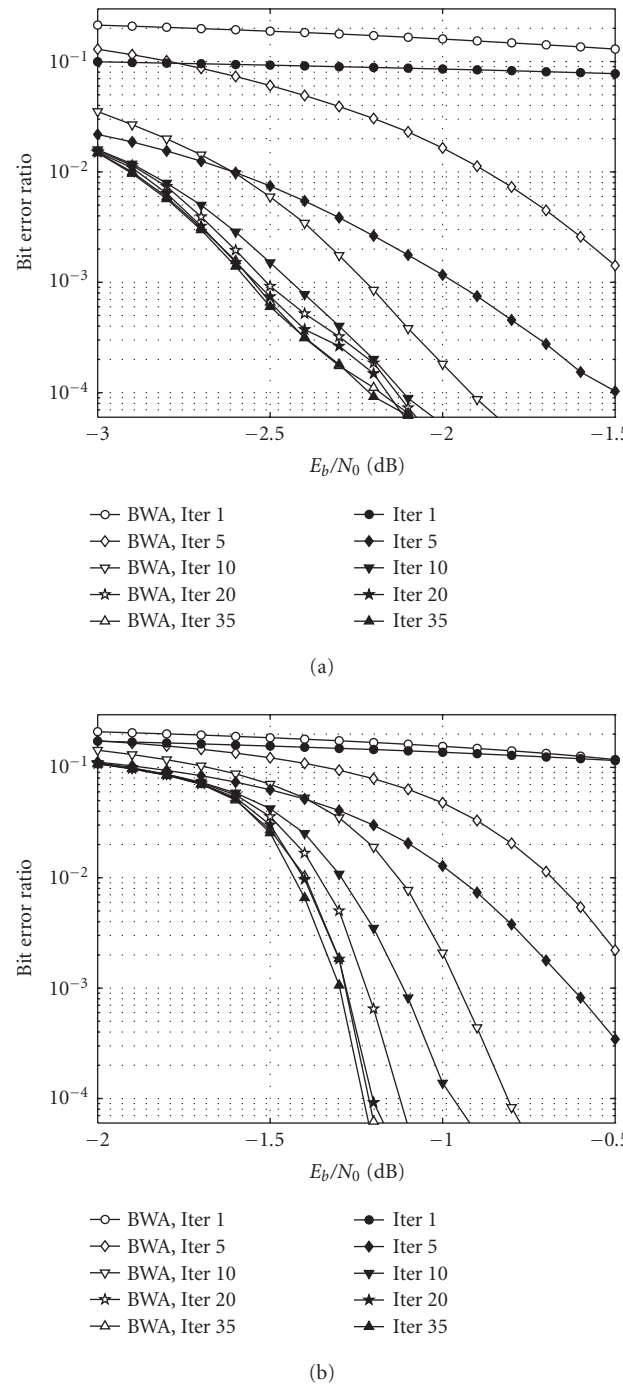

FIGURE 8: BER versus E_b/N_0 (dB) for several iteration numbers: (a) $H(S_1 \mid S_2) = 0.45$ and (b) $H(S_1 \mid S_2) = 0.73$. The label BWA stands for the case where the HMM parameters are iteratively estimated.

we would not achieve any performance improvement with respect to the case of no side information. As previously mentioned, the reason is that with this decoder, the rate compression for source S_1 would be limited to $H(X_k \mid Y_k) = H(E_1) \approx H(X_k)$, implying that there is practically no correlation (of depth $n = 1$) between S_1 and S_2.

4. CONCLUSIONS

Given two binary correlated sources with hidden Markov correlation, this paper proposes an asymmetric distributed joint source-channel coding scheme for the transmission of one of the sources over an AWGN. We assume that the other source output is available as side information at the receiver. A turbo encoder and a joint decoder are used to exploit the Markov correlation between the sources. We show that, when the correlation statistics are not initially known at the decoder, they can be estimated jointly within the iterative decoding process without any performance degradation. Simulation results show that the performance of this system achieves signal to noise ratios close to those established by the combination of Shannon and Slepian-Wolf theorems.

REFERENCES

[1] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[2] T. Cover, "A proof of the data compression theorem of Slepian and Wolf for ergodic sources (Corresp.)," *IEEE Trans. Inform. Theory*, vol. 21, no. 2, pp. 226–228, 1975.

[3] S. Shamai and S. Verdú, "Capacity of channels with uncoded side information," *European Transactions on Telecommunications*, vol. 6, no. 5, pp. 587–600, 1995.

[4] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: turbo-codes," *IEEE Trans. Commun.*, vol. 44, no. 10, pp. 1261–1271, 1996.

[5] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): design and construction," in *Proc. IEEE Data Compression Conference (DCC '99)*, pp. 158–167, Snowbird, Utah, USA, March 1999.

[6] J. Bajcsy and P. Mitran, "Coding for the Slepian-Wolf problem with turbo codes," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM '01)*, vol. 2, pp. 1400–1404, San Antonio, Tex, USA, November 2001.

[7] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "Distributed compression of binary sources using conventional parallel and serial concatenated convolutional codes," in *Proc. IEEE Data Compression Conference (DCC '03)*, pp. 193–202, Snowbird, Utah, USA, March 2003.

[8] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "Compression of binary sources with side information at the decoder using LDPC codes," *IEEE Commun. Lett.*, vol. 6, no. 10, pp. 440–442, 2002.

[9] J. Garcia-Frias and W. Zhong, "LDPC codes for compression of multi-terminal sources with hidden Markov correlation," *IEEE Commun. Lett.*, vol. 7, no. 3, pp. 115–117, 2003.

[10] J. Garcia-Frias, "Compression of correlated binary sources using turbo codes," *IEEE Commun. Lett.*, vol. 5, no. 10, pp. 417–419, 2001.

[11] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. IEEE Data Compression Conference 2002 (DCC '02)*, pp. 252–261, Snowbird, Utah, USA, April 2002.

[12] A. D. Liveris, Z. Xiong, and C. N. Georghiades, "Joint source-channel coding of binary sources with side information at the decoder using IRA codes," in *Proc. IEEE International Workshop on Multimedia Signal Processing (MMSP '02)*, pp. 53–56, St. Thomas, US Virgin Islands, December 2002.

[13] J. Garcia-Frias, "Joint source-channel decoding of correlated sources over noisy channels," in *Proc. IEEE Data Compression Conference (DCC '01)*, pp. 283–292, Snowbird, Utah, USA, March 2001.

[14] W. Zhong, H. Lou, and J. Garcia-Frias, "LDGM codes for joint source-channel coding of correlated sources," in *Proc. IEEE International Conference on Image Processing (ICIP '03)*, vol. 1, pp. 593–596, Barcelona, Spain, September 2003.

[15] J. Garcia-Frias and J. D. Villasenor, "Joint turbo decoding and estimation of hidden Markov sources," *IEEE J. Select. Areas Commun.*, vol. 19, no. 9, pp. 1671–1679, 2001.

[16] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 498–519, 2001.

[17] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[18] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (Corresp.)," *IEEE Trans. Inform. Theory*, vol. 20, no. 2, pp. 284–287, 1974.

Javier Del Ser was born on March 13, 1979, in Barakaldo, Spain. He studied telecommunication engineering from 1997 to 2003 at the Technical Engineering School of Bilbao (ETSI), Spain, where he obtained his M.S. degree in 2003. As a Member of the Signal and Communication Group at the Department of Electronics and Telecommunications of the University of the Basque Country (EHU/UPV), he developed a signal processing system for the measurement of quality parameters of the power line supply. Currently, he is working toward the Ph.D. degree at the Centro de Estudios e Investigaciones Técnicas de Gipuzkoa (CEIT), San Sebástian, Spain. He is also a Teaching Assistant at TECNUN (University of Navarra). His research interests are focused on factor graph theory, distributed source coding, and both turbo-coding and turbo-equalization schemes, with a special interest in their practical application in real scenarios.

Pedro M. Crespo was born in Barcelona, Spain. In 1978, he received the Engineering degree in telecommunications from Universidad Politécnica de Barcelona, and the M.S. degree in applied mathematics and Ph.D. degree in electrical engineering from the University of Southern California (USC), in 1983 and 1984, respectively. From September 1984 to April 1991, he was a Member of the technical staff in the Signal Processing Research Group at Bell Communications Research, New Jersey, USA, where he worked in the areas of data communication and signal processing. He actively contributed in the definition and development of the first prototypes of digital subscriber lines transceivers (xDSL). From May 1991 to August 1999, he was a District Manager at Telefónica Invéstigacion y Desarrollo, Madrid, Spain. From 1999 to 2002, he was the Technical Director of the Spanish telecommunication operator Jazztel. At present, he is the Department Head of the Communication and Information Theory Group at Centro de Estudios Investigaciones Técnicas de Gipuzkoa (CEIT), San Sebástian, Spain. He is also a Full Professor at TECNUN (University of Navarra). Pedro Crespo is a Senior Member of the Institute of Electrical and Electronic Engineers (IEEE) and he is a recipient of the Bell Communication Researchs Award of Excellence. He holds seven patents in the areas of digital subscriber

lines and wireless communications. His research interests currently include space-time coding techniques for MIMO systems, iterative coding and equalization schemes, bioinformatics, and sensor networks.

Olaia Galdos was born on April 20, 1976, in Legazpi, Spain. She studied mathematics from 1994 to 1999 at Sciences Faculty of the University of the Basque Country, Leioa, Spain. Currently she is a Ph.D. candidate at TECNUN (University of Navarra, Spain). Her research topics are in Slepian-Wolf distributed source coding with turbo and LDPC codes, factor graph theory and its application to coding and decoding algorithms.