

# LDGM Codes for Channel Coding and Joint Source-Channel Coding of Correlated Sources

**Wei Zhong**

*Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716, USA  
Email: zhong@ece.udel.edu*

**Javier Garcia-Frias**

*Department of Electrical and Computer Engineering, University of Delaware, Newark, DE 19716, USA  
Email: jgarcia@ece.udel.edu*

*Received 2 October 2003; Revised 3 October 2004*

We propose a coding scheme based on the use of systematic linear codes with low-density generator matrix (LDGM codes) for channel coding and joint source-channel coding of multiterminal correlated binary sources. In both cases, the structures of the LDGM encoder and decoder are shown, and a concatenated scheme aimed at reducing the error floor is proposed. Several decoding possibilities are investigated, compared, and evaluated. For different types of noisy channels and correlation models, the resulting performance is very close to the theoretical limits.

**Keywords and phrases:** channel coding, LDPC codes, LDGM codes, iterative decoding, correlated sources, joint source-channel coding.

## 1. INTRODUCTION

The introduction of turbo codes [1] and low-density parity check (LDPC) codes [2, 3, 4, 5] has been one of the most important milestones in channel coding during the last years. Provided that the information block lengths are long enough, performance close to the Shannon theoretical limit can be achieved for different channel environments. However, in practical applications, complexity issues have to be carefully considered, since both schemes present either high encoding or high decoding complexity. Specifically, for the case of turbo codes the encoding complexity is very low, but the decoding complexity is high. Compared with turbo codes, standard LDPCs present a higher encoding complexity, but the decoder is simpler.

In this paper, we first show that it is possible to achieve a channel coding performance comparable to that of standard

LDPC and turbo codes by utilizing systematic linear codes with low-density generator matrices [6] (LDGM codes<sup>1</sup>). LDGM codes present a complexity advantage over standard LDPC and turbo codes. Specifically, because of the sparseness of the generator matrix, the amount of processing required in the encoder is linear with the block size and similar to that of turbo codes. Moreover, since the parity check matrix of systematic LDGM codes is also sparse, such codes are in fact a subset of LDPC codes and can be decoded in the same manner and with the same complexity as standard LDPC codes. Notice, however, that in order to facilitate the development of the paper, we will derive the decoding algorithm utilizing the graph corresponding to the generator matrix.

In the second part of the paper, we will focus on the use of LDGM codes to perform joint source-channel coding of correlated sources, which includes the case of pure source coding as a particular case. The problem of coding for multiterminal correlated sources has important practical applications (e.g., in the context of sensor networks [9]). Since compression and joint source-channel coding of correlated sources can be considered as a problem of channel coding with side information [10, 11], the use of powerful channel codes, such as LDGM codes, should produce very good results in this

---

This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This work was partially supported by NSF Grant CCR-0311014 and prepared through collaborative participation in the Communications and Networks Consortium sponsored by the US Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-2-0011. The US Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation thereon.

---

<sup>1</sup>As indicated in [7], concatenated single parity check codes [8] are particular examples of LDGM codes.

context. The first schemes for compression of multiterminal correlated sources using channel codes were proposed in [12, 13, 14], and the use of turbo-like codes in this problem was proposed in [15, 16]. However, although as shown in [17, 18] the separation principle between source and channel coding holds for transmission of correlated sources over separated noisy channels, it is not straightforward to implement a practical system based on this concept. One of the reasons is that, in spite of some previous work (see [14, 19] and the references therein), the problem of designing good practical codes for correlated sources from a source coding perspective is still open. Moreover, the separation between source and channel coding may lead to catastrophic error propagation. Previous work in joint source-channel coding using iterative decoding schemes (turbo and LDPC codes) for the cases in which only one source is corrupted by noise (and the special case of joint source-channel coding for single sources) can be found in [20, 21, 22, 23, 24]. The case in which both sources are transmitted through separated noisy channels has appeared only in [25, 26], where turbo codes were proposed to perform joint source-channel coding of correlated sources.

The contents of this paper are as follows: Section 2 introduces systematic LDGM codes and presents the decoding algorithm in relation with that of standard LDPC codes. The proposed concatenated scheme, required to eliminate the error floor in both channel coding and joint source-channel coding, is introduced in Section 3. Section 4 contains simulation results for the case of channel coding with LDGM codes over binary symmetric channels (BSC), and additive white gaussian noise (AWGN) and Rayleigh fading channels. The rest of the paper deals with the problem of joint source-channel coding of correlated sources. The theoretical limits for this problem are presented in Section 5, and the rationale for using turbo-like codes in this context is explained in Section 6. Section 7 specifies the proposed system and the correlation model utilized in this paper for joint source-channel coding of correlated sources. Simulation results are provided in Section 8. Finally, Section 9 concludes the paper, and a comparison between the decoding complexity of LDGM and turbo codes is provided in the appendix.

## 2. SYSTEMATIC LDGM CODES FOR CHANNEL CODING

### 2.1. Encoding

Systematic LDGM codes are linear codes with a sparse generator matrix,  $G = [I \ P]$  with  $P = [p_{km}]$  and thus the corresponding parity check matrix is  $H = [P^T \ I]$ . We denote the information message that we want to transmit as  $\underline{u} = [u_1 \ \dots \ u_K]$ . These bits, together with the coded bits generated as  $\underline{c} = \underline{u}P$ , with  $\underline{c} = [c_1 \ \dots \ c_M]$ , are transmitted through a noisy channel. The corrupted sequence at the decoder is denoted as  $[\underline{u}' \ \underline{c}']$ , where  $c'_m = c_m + e_m^1$  and  $u'_k = u_k + e_k^2$ , with  $e_m^1$  and  $e_k^2$  being the noise introduced by the channel. Notice that the proposed code is systematic with rate  $K/N$ , where  $N = K + M$ . In this paper, we will denote as regular  $(X, Y)$  LDGM codes those irregular systematic LDPC

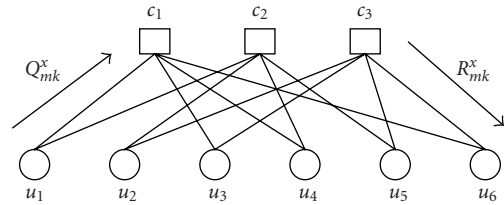


FIGURE 1: Bipartite graph representing an LDGM code.  $\{c_m\}$  represent the coded bits generated at the encoder (before being corrupted by the channel).  $\{u_k\}$  are the nodes corresponding to the systematic bits. The figure shows the two different types of messages that are propagated in the decoding process through the graph.

codes in which all the  $N - K$  check nodes have degree  $Y + 1$ , all the  $K$  systematic bit nodes have degree  $X$ , and each of the  $N - K$  coded bit nodes has degree 1 and is associated to its corresponding check node. In other words, the parity matrix  $P$  of an  $(X, Y)$  LDGM code has exactly  $X$  nonzero entries per row and  $Y$  nonzero entries per column. We will also abuse this notation to include irregular LDGM codes whose average degree distributions are also  $(X, Y)$  (i.e., the average number of nonzero entries per row and per column in their matrices  $P$  is  $X$  and  $Y$ , resp.) In any case, it is obvious that the relationship between the code rate and the degree distributions is given by

$$R_c = \frac{Y}{X + Y}. \quad (1)$$

### 2.2. Decoding algorithm

For the case of LDPC codes, the decoder goal is to iteratively estimate the most probable solution of the equation  $\underline{s} = \underline{e}H^T$ , where  $\underline{s}$  represents the syndrome calculated from the received sequence,  $H$  is the parity check matrix of the code, and  $\underline{e}$  is the error pattern that we are interested in calculating (the sparsest one satisfying the equation above). Analogously, we can see LDGM decoding as a method to find the most probable solution for the equation  $\underline{c} = \underline{u}P$ , where  $\underline{c}$  is the vector of coded bits generated at the encoder (i.e., before they are corrupted by the channel noise),  $G = [I \ P]$  is the generator matrix of the code, and  $\underline{u}$  is the information message that we want to calculate. Figure 1 shows the graph associated with the proposed code.

The decoding algorithm for LDGM codes can be derived by applying belief propagation [27] (or factor graph decoding [28]) over the graph described in Figure 1. The application of belief propagation for the proposed codes presents an important difference with respect to standard LDPC codes: in standard LDPC codes, the syndrome nodes are fixed to a deterministic value (i.e., for each position they are either 0 or 1). In the proposed codes, the coded bit nodes  $\underline{c}$  are random variables, with distribution calculated depending on the received corrupted coded bits  $\underline{c}'$  and the distribution of the noise vector  $\underline{e}^1$  (i.e.,  $c'_m = c_m + e_m^1$ ).

Although this decoding procedure is only exact for networks without cycles, we will see in our simulations that the

proposed codes achieve good performance. In order to facilitate the algorithm implementation, we present the decoding method indicating only the modifications with respect to the case of standard LDPC codes. Using the notation commonly utilized in the LDPC literature, we will denote by  $r_{mk}^x$ ,  $x \in \{0, 1\}$ , the message propagated from coded bit node  $c_m$  to systematic bit node  $u_k$  (and by  $q_{mk}^x$ ,  $x \in \{0, 1\}$  the message propagated from systematic bit node  $u_k$  to coded bit node  $c_m$ ), if the standard LDPC decoding algorithm were to be directly applied over the graph shown in Figure 1. We now indicate the modifications necessary to deal with LDGM codes. The messages passed from coded bit nodes to systematic bit nodes will be denoted by  $R_{mk}^x$ .  $Q_{mk}^x$  indicate the messages exchanged from systematic bit nodes to coded bit nodes.

- (1) *Initialization.* Fix the probability of the systematic bit nodes to its *a priori* value. For every  $(m, k)$  such that  $u_k$  and  $c_m$  are connected, let  $Q_{mk}^x = P_k^x = 1 - p_k$  if  $u'_k = x$ , and  $Q_{mk}^x = P_k^x = p_k$  if  $u'_k \neq x$ , with  $x \in \{0, 1\}$ .  $p_k$  denotes the probability that systematic bit  $u_k$  is received in error.
- (2) *Message passing from the coded bit nodes to the systematic bit nodes.*
  - (i) For every  $(m, k)$  pair such that  $u_k$  and  $c_m$  are connected, calculate  $r_{mk}^x$  as in standard LDPC decoding (but using the parameters  $Q_{mk}^x$  instead of  $q_{mk}^x$ ).
  - (ii) Calculate  $R_{mk}^x$  as

$$R_{mk}^0 = (1 - \gamma_m)r_{mk}^0 + \gamma_m r_{mk}^1, \quad (2)$$

$$R_{mk}^1 = (1 - \gamma_m)r_{mk}^1 + \gamma_m r_{mk}^0, \quad (3)$$

where  $\gamma_m$  denotes the probability that coded bit  $c_m$  is received in error. The intuitive idea behind the previous equations is that when the received coded bit is in error (which occurs with probability  $\gamma_m$ ) the message to pass,  $R_{mk}^x$ , is not  $r_{mk}^x$ , but  $r_{mk}^{1-x}$ . If the received coded bit is correct (which occurs with probability  $1 - \gamma_m$ ),  $R_{mk}^x = r_{mk}^x$ .

- (3) *Message passing from the systematic bit nodes to the coded bit nodes.* For every  $(m, k)$  pair such that  $u_k$  and  $c_m$  are connected, calculate  $Q_{mk}^x$  in the same way as  $q_{mk}^x$  is calculated in standard LDPC decoding. The only differences are that now parameters  $R_{mk}^x$  should be used instead of  $r_{mk}^x$ , and that the *a priori* probability for the systematic bit node  $u_k$  is  $P_k^x = 1 - p_k$  if  $u'_k = x$ , and  $P_k^x = p_k$  if  $u'_k \neq x$ , with  $x \in \{0, 1\}$ .

### 3. CONCATENATED LDGM SCHEMES

As shown in [29], and first recognized by MacKay [5], LDGM codes are bad codes, since they present error floors that are independent of the block length. Specifically, LDGM codes with small degrees have high error floors but good convergence thresholds, while the contrary occurs if the degrees are high [29]. However, these error floors can be reduced and practically eliminated. The reason is that, as explained in [29] for BSCs, the number of errors for the blocks in error decays very fast, provided that the crossover probability

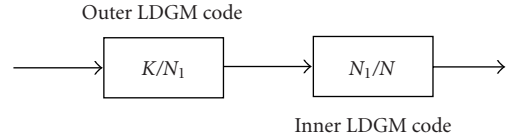


FIGURE 2: Concatenated scheme of LDGM codes to reduce the error floor. First, the information message is encoded by a high rate  $K/N_1$  outer LDGM code. The output is encoded by a rate  $N_1/N$  inner LDGM code to produce a rate  $K/N$  overall code.

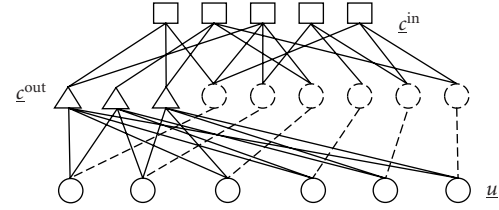


FIGURE 3: Graph associated with the concatenated LDGM codes with  $c^{\text{in}}$  as the inner coded bit nodes,  $c^{\text{out}}$  as the outer coded bit nodes, and  $u$  as the information (systematic) bit nodes.

is small enough (a similar behavior can be observed in other channels when their quality improves). Moreover, the LDGM decoder produces a good indication of where the residual errors are located. Most of the systematic bits in error will have a corresponding probability very close to 0.5, compared with a probability close to 1 for the bits that have been successfully decoded. This means that the results obtained from the decoding of an LDGM code can be seen as produced by an equivalent channel introducing a small amount of “erasures” at specific locations.

Notice however that this equivalent channel (consisting of the concatenation of the channel and the LDGM code) is not a standard erasure channel. The error locations are not known with certainty, but, on the other hand, there is *a priori* information about how the erasures should be filled. Specifically, the LDGM decoder provides this *a priori* probability for each of the systematic bits. This *a priori* probability can be easily exploited if those systematic bits for the LDGM code are in fact generated by another outer LDGM code. Then, the outer LDGM decoder would use the *a priori* probability to initialize its systematic and coded bit nodes in the decoding process, which would reduce the number of residual errors. The encoder diagram for the proposed concatenated scheme is shown in Figure 2.

Figure 3 shows the graph representing the concatenated scheme. Decoding can be easily performed by applying the belief propagation algorithm [27] over the graph. Since the network has cycles (more than for the case of single codes), different activation schedules can lead to different results. In order to define different activation schemes, we assume that message passing is performed according to a global clock. At each clock cycle, a group of nodes is activated. The *activation* of a node is defined as the process in which the node reads all its incoming messages from all its neighboring nodes,

performs computation, and outputs all its outgoing messages to its neighbors. We follow the convention that once a message is produced, it remains available to the corresponding nodes until the message gets updated.

We define two different scheduling schemes for the case of channel coding. The first one (decoding algorithm I) has already been defined in [29]. It first proceeds with the decoding of the inner code, and the last step is to reduce the residual errors by decoding the outer code. The second schedule (decoding algorithm II) iterates between the inner and outer code in each iteration. For notation purposes, we will assume that nodes are activated serially by order of appearance in the scheme definition, except when included in brackets (which means that activation for those nodes is performed in parallel within a clock cycle).<sup>2</sup> Then, the two different schedules can be expressed as follows.

- (i) Decoding algorithm I [29]:

$$\begin{aligned} & \underline{c}^{\text{out}}, \\ & \underline{u}, \underline{c}^{\text{in}}, \dots, \underline{u}, \underline{c}^{\text{in}}, \\ & \underline{u}, \underline{c}^{\text{out}}, \dots, \underline{u}, \underline{c}^{\text{out}}. \end{aligned}$$

- (ii) Decoding algorithm II: repeat

$$\underline{u}, \underline{c}^{\text{in}}, \underline{u}, \underline{c}^{\text{out}}.$$

It is possible to define other activation schedules by varying the number of times that the inner decoder is activated per iteration in the outer decoder. In general, we have observed that increasing the activation rate of the inner decoder leads to slightly higher error floors, while more frequent activations in the outer decoder slightly degrades the convergence threshold. This observation is consistent with the functioning of the concatenated scheme, since the inner code determines the convergence threshold and the outer code reduces the error floor by taking care of the small number of residual errors.

The concatenation of the two LDGM codes can be considered as a single irregular LDGM code of generator matrix  $G_{\text{joint}} = G_{\text{outer}}G_{\text{inner}}$ . Therefore, decoding can be performed using the graph corresponding to  $G_{\text{joint}}$ . Notice that this graph can be easily obtained from Figure 3 by eliminating the connections between the inner and outer coded bit nodes (connecting instead the systematic bit nodes directly to the corresponding inner coded bit nodes). In this case, a systematic bit node could be connected more than once to the same inner coded bit node. Then, if the number of connections is even, all connections cancel with each other and no connection appears in the final graph, while if the number of connections is odd, one of them is kept. However, as we illustrate in the next section, the performance of the “joint” scheme,  $G_{\text{joint}}$ , is much worse than that of schedules I and II.

#### 4. SIMULATION RESULTS OF LDGM CODES FOR CHANNEL CODING

In all our simulations, the matrices  $P_{\text{inner}}$  and  $P_{\text{outer}}$  are generated in a pseudorandom way without introducing cycles of length 4 or less. In all cases, at least 10 000 blocks are simulated, and, for the decoding of each block, the iterative process continues until 3 consecutive iterations produce the same result for the systematic bits or 100 iterations are run.

We first encoded 9500 information bits with a regular (4, 76) outer LDGM code to produce a total of 10 000 bits. These bits were encoded again by a regular (6, 6) inner LDGM code producing a total of 20 000 bits (i.e., overall rate  $R_c = 0.475$ ). For the joint scheme, the node-degree profiles (i.e., the percentage of nodes of a given degree) of the code resulting from the concatenation of the inner and outer codes ( $G_{\text{joint}} = G_{\text{outer}}G_{\text{inner}}$ ) are  $\lambda(x) = 0.0004x^{30} + 0.0277x^{32} + 0.9719x^{34}$  for the systematic bit nodes and  $\rho(x) = 0.697810x^5 + 0.047619x^{75} + 0.000095x^{76} + 0.009429x^{78} + 0.215524x^{80} + 0.000667x^{151} + 0.012190x^{153} + 0.015048x^{155} + 0.000190x^{224} + 0.000762x^{226} + 0.000571x^{228} + 0.000095x^{230}$  for the coded bit nodes. Considering an AWGN channel, even for very high signal to noise ratios (4 dB above the Shannon limit), the residual BER for each block is always higher than  $10^{-2}$ , and presents oscillations with the iteration number. This behavior can be explained by the existence in  $G_{\text{joint}}$  of a peculiar type of structure containing many short cycles of length 4, which are produced as described below.

Figure 4a shows all the connections for a given outer coded bit node (triangle) in the concatenated LDGM scheme shown in Figure 3. Following the rules described in Section 3, it is obvious that in the graph corresponding to  $G_{\text{joint}}$ , Figure 4a becomes the structure shown in Figure 4b. Figure 4b assumes that there are no other connections (either directly or through other outer coded bits) for the shaded systematic and inner coded bit nodes in Figure 4a. As described before, if an even number of connections between a given systematic bit and a given coded bit were to exist,  $G_{\text{joint}}$  would present no connection between these two nodes. The important point is that  $G_{\text{joint}}$  has as many of the structures shown in Figure 4b as outer coded bits (500). Even if some of the links in the structures are eliminated (in this example the probability of elimination of a link is less than 0.03), these structures are highly regular, and each of the 76 systematic bits presents many loops of length 4 (21, i.e., 7 choose 2, in the case of no link elimination), all of them involving the shaded bits in Figure 4b. The occurrence of this type of structure explains the poor performance of  $G_{\text{joint}}$  and the oscillating behavior in the decoding process. Notice that when decoding is performed in the graph containing both the inner code and the outer code (Figure 3), this type of cycles is broken by the outer coded bit nodes (see Figure 4a).

The results presented in this section for the joint scheme assume a code  $G'_{\text{joint}}$  with the same node-degree distributions as  $G_{\text{joint}}$  (in fact, with exactly the same number of nodes with a given degree), but with random connection assignments. In this way, the structures existing in  $G_{\text{joint}}$  disappear and the performance is expected to improve. Figure 5 shows

<sup>2</sup>In this section, both schedules consist of serial activations. Some parallel activation will occur in the case of joint source-channel coding of correlated sources.

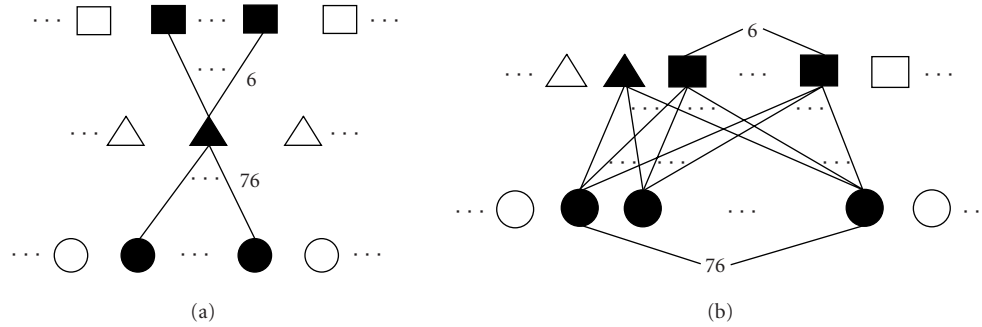


FIGURE 4: (a) All the connections for a given outer coded bit node (triangle) in the concatenated LDGM scheme when the graph is represented as the concatenation of the inner and outer codes as in Figure 3 are shown. They get converted into (b) in the graph corresponding to the joint scheme  $G'_{\text{joint}}$ .

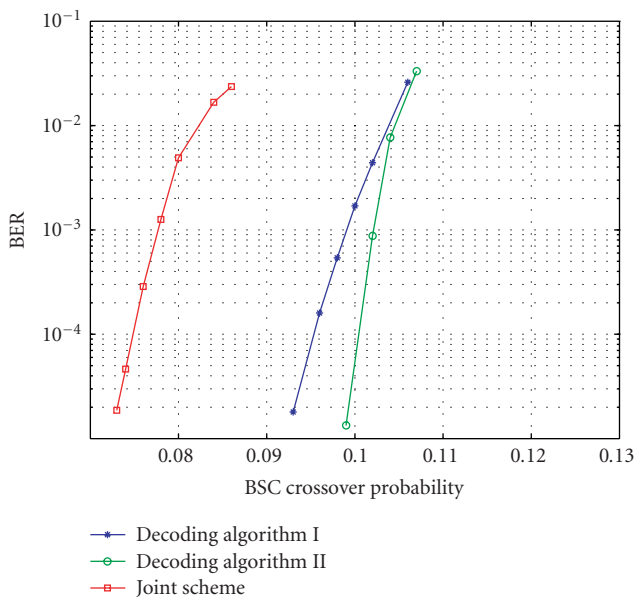


FIGURE 5: Performance of the joint scheme and of the decoding algorithms I and II when the proposed concatenated system is utilized over BSCs. The overall rate of the code is 0.475. The Shannon limit for this case is  $p = 0.118$ .

the performance of the proposed decoding algorithms over a BSC as a function of the crossover probability, while Figure 6 illustrates the performance for AWGN channels as a function of  $E_b/N_0$  (where  $E_b$  is the energy per information bit and  $N_0$  denotes the one-side spectral density). For schedules I and II, no error floor appeared after simulating more than 10 000 blocks, and the resulting performance is very close to the theoretical limits (which correspond to a crossover parameter 0.118 for the BSC case and  $E_b/N_0 = 0.08$  dB for the AWGN channel assuming binary signaling [30]). These results are comparable to those obtained with turbo and irregular LDPC codes [31, 32]. It is interesting to remark that both activation schedules lead to similar results for AWGN channels, while schedule II is slightly superior for BSCs. Notice that the performance of the “joint” scheme,  $G'_{\text{joint}}$ , is

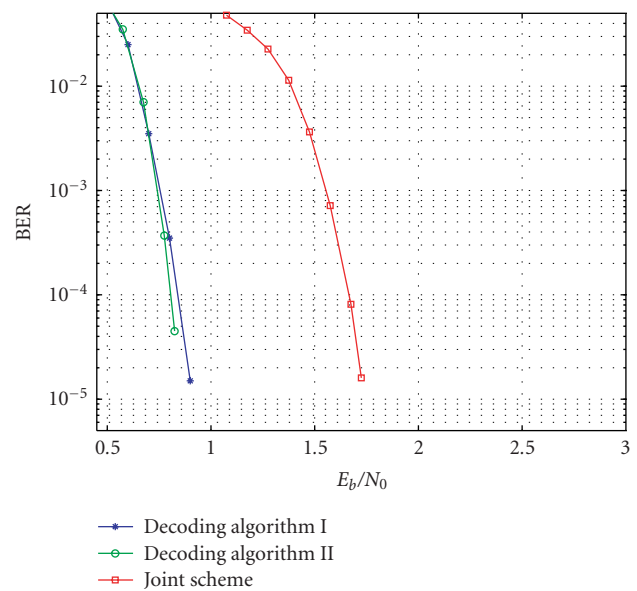


FIGURE 6: Performance of the joint scheme and of the decoding algorithms I and II when the proposed concatenated system is utilized over AWGN channels. The overall rate of the code is 0.475. The Shannon limit for this case (assuming binary signaling) is  $E_b/N_0 = 0.08$  dB.

still worse than that of decoding algorithms I and II, but it presents no oscillating behavior. This worse performance can be explained by the existence of some short cycles in  $G'_{\text{joint}}$ <sup>3</sup> and by noticing that random connection assignments lead to a graph structure different from the specific one resulting from the concatenation of the inner and outer codes (represented in Figure 3). This occurs because the graph in Figure 3 contains the outer coded bit nodes as additional variables that (due to the random connections) do not have equivalence in  $G'_{\text{joint}}$ .

<sup>3</sup>In the construction process, we do not allow cycles of length 4 in  $P_{\text{inner}}$  and  $P_{\text{outer}}$ . However, due to its non-low-density nature, it is not possible to eliminate all cycles of length 4 in  $G'_{\text{joint}}$ .

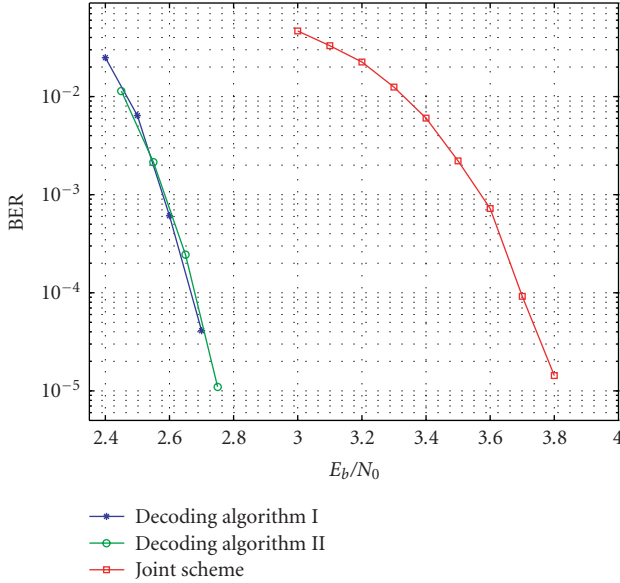


FIGURE 7: Performance of the joint scheme and of the decoding algorithms I and II when the proposed concatenated system is utilized over fully interleaved Rayleigh fading channels with perfect CSI at the receiver. The overall rate of the code (assuming binary signaling) is 0.475. The Shannon limit for this case is  $E_b/N_0 = 1.6$  dB.

We also investigated the performance of LDGM codes over Rayleigh fading channels. We assume that the received sequence can be expressed as  $r_k = c_k y_k + n_k$ , where  $\{y_k\}$  is the binary transmitted sequence,  $\{n_k\}$  is a set of statistically independent Gaussian random variables with zero-mean, and  $\{c_k\}$  is modeled as a Rayleigh process. We also assume an ideally interleaved channel, so that the sequence  $\{c_k\}$  is uncorrelated in time  $k$ , and perfect channel side information (CSI) is available at the decoder (i.e., the value of  $c_k$  is known). Figure 7 shows the performance of the concatenated LDGM code defined before in this Rayleigh fading environment. Notice that similar to the BSC and AWGN channels, no error floor appears here and decoding algorithms I and II outperform the “joint” scheme,  $G'_{\text{joint}}$  again. The theoretical limit in this case (assuming binary signaling) is  $E_b/N_0 = 1.6$  dB [30], and both schedules I and II achieve a performance within 1.3 dB from this limit.

## 5. SOURCE AND JOINT SOURCE-CHANNEL CODING OF CORRELATED SOURCES: THEORETICAL LIMITS

Figure 8 illustrates the system proposed in this paper for joint source-channel coding of correlated sources. For simplicity, we consider only two sources, but the approach can be easily extended to the case of more sources. The two sources are encoded independently from each other (i.e., for a given source neither the realization from the other source nor the correlation model are available at the encoder site) and transmitted through two different noisy channels to a common decoder. Since the correlation between the sources is exploited at the

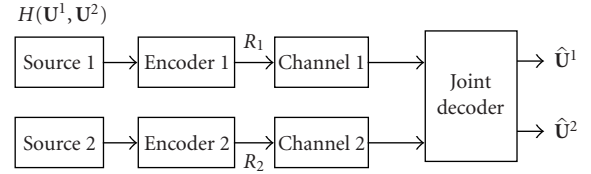


FIGURE 8: Proposed system for joint source-channel coding of correlated sources. Each source is encoded independently and transmitted through a different noisy channel.

common receiver, the value of  $E_b/N_0$  corresponding to the theoretical limit will be less than if the sources were independent. In this section, we review the theoretical limits for the case in which both channels are either noisy or noiseless (which corresponds to the case of compression of correlated sources).

### 5.1. Compression of correlated sources: Slepian-Wolf limit

It is well known [33, 34] that two jointly ergodic sources  $(\mathbf{U}^1, \mathbf{U}^2)$ , defined over countably infinite alphabets, can be compressed at rates  $(R_1, R_2)$ , provided that

$$\begin{aligned} R_1 &\geq H(\mathbf{U}^1 | \mathbf{U}^2), \\ R_2 &\geq H(\mathbf{U}^2 | \mathbf{U}^1), \\ R_1 + R_2 &\geq H(\mathbf{U}^1, \mathbf{U}^2). \end{aligned} \quad (4)$$

As explained before, compression is performed independently for each source and the decoder jointly acts over the compressed versions of the sources to recover the original sequences.

### 5.2. Transmission of correlated sources over independent noisy channels

It has been recently shown [17, 18] that the separation principle between source and channel coding applies to the case of transmission of correlated sources over separated noisy channels. In other words, the theoretical limit for the transmission of two sources generating i.i.d. random pairs can be obtained by performing first distributed data compression up to the Slepian-Wolf limit followed by channel coding. Therefore, assuming that both sources are encoded at the same rate ( $R_1 = R_2 = R/2$ ), the theoretical limit in communications for a fixed transmission rate of  $R/2$  information bits/channel use would then be achieved for each source when the two correlated sources are first compressed up to the joint entropy ( $H(\mathbf{U}^1, \mathbf{U}^2)$ ) and then a capacity achieving channel code of rate  $R'_c = R/2$  is used for each of them. By taking into account that the energy per generated source bit ( $E_{s_0}$ ) can be related with the energy per information bit ( $E_b$ ) by using the relation  $2E_{s_0} = H(\mathbf{U}^1, \mathbf{U}^2)E_b$ , the theoretical limit for  $E_{s_0}/N_0$  (for the case of two independent channels with capacity  $C$ ) can be obtained by solving the equation  $R/2 = C$  [18].

The previous separate source and channel coding approach would achieve the theoretical limit if two conditions

are met. On the one hand, optimum source coding for correlated sources should be utilized. On the other, capacity achieving channel codes are necessary. The problem with this approach in practical systems is twofold. First, it is necessary to design good practical source codes for correlated sources. Moreover, in practical systems, errors introduced by the channel decoder could be catastrophic for the source decoder. Besides, in our approach, it does not seem reasonable to first use LDGM codes to compress the sources and then some other LDGM codes to add redundancy, since the use of one LDGM code (per source) can perform the combined operation. In order to avoid these problems, we propose a joint source-channel coding scheme which in practical situations achieves performance very close to the theoretical limits. In our approach, each of the correlated binary sources is not source encoded, but directly channel encoded with a channel code of rate  $R_c$ . The information rate transmitted through the channel in this case is  $R_1 = R_2 = R/2 = H(\mathbf{U}^1, \mathbf{U}^2)R_c/2$  information bits/channel use. Notice that in order to keep the information rate per source ( $R/2$ ), the code used in our joint source-channel coding approach (of rate  $R_c$ ) has to be less powerful than in the separate source and channel coding scheme (code of rate  $R'_c = R/2$ ).

Specifically, the relation between  $R'_c$  and  $R_c$  to keep the same information rate through the channel,  $R/2 = R'_c$ , is given by  $R'_c = H(\mathbf{U}^1, \mathbf{U}^2)R_c/2$ . The “weakness” of the code in the joint source-channel coding approach will be compensated by exploiting the correlation between sources in the decoder. Notice that the proposed joint source-channel coding approach allows a channel code of a single rate to be used in combination with sources having arbitrary joint entropy rates, with the modifications to maintain efficient coding involving only processing in the decoder.

## 6. TURBO-LIKE CODES FOR CORRELATED SOURCES: RATIONALE

As indicated in the introduction, compression (or joint source-channel coding) of multiterminal correlated sources can be seen as a problem of channel coding with side information. This is illustrated in Figure 9. To simplify the description, we assume that the information from source 2 is perfectly available at the decoder and that  $\mathbf{U}^1$  is the sequence generated by source 1 that we want to compress. In order to do so,  $\mathbf{U}^1$  is encoded by a systematic channel encoder, so that the nonsystematic coded bits,  $\mathbf{C}^1$ , constitute the compressed sequence for source 1 (i.e., the systematic bits are eliminated). The decoder utilizes the compressed sequence for source 1 ( $\mathbf{O}^1 = \mathbf{C}^1$ ) plus the information proceeding from the other source ( $\mathbf{U}^2$ ). Notice that, because of the correlation between sources,  $\mathbf{U}^2$  can be used as a side information for source 1. In fact,  $\mathbf{U}^2$  can be thought as the corrupted version of the systematic bits in source 1 when  $\mathbf{U}^1$  is transmitted through a channel model defined by the correlation between sources. Therefore, recovery of  $\mathbf{U}^1$  can be interpreted as performing channel decoding over the corrupted version of  $\mathbf{U}^1$  ( $\mathbf{U}^2$ ) and the redundant/uncorrupted nonsystematic

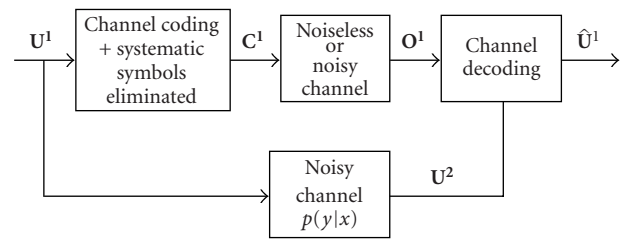


FIGURE 9: Source coding as a problem of channel coding with side information.  $\mathbf{U}^1$  is encoded by a systematic channel encoder, so that the nonsystematic bits,  $\mathbf{C}^1$ , constitute its compressed version.  $\mathbf{U}^2$  can be thought of as the corrupted version of  $\mathbf{U}^1$  when  $\mathbf{U}^1$  is transmitted through a channel model defined by the correlation between sources, and is used by the decoder together with  $\mathbf{O}^1 = \mathbf{C}^1$  to recover the original sequence  $\mathbf{U}^1$ .

bits  $\mathbf{O}^1 = \mathbf{C}^1$  proceeding from  $\mathbf{U}^1$ . In the case in which  $\mathbf{U}^2$  were not perfectly available at the receiver, the decoder would consist of two blocks as the one shown in Figure 9 (one for source 1 and the other for source 2). Then, decoding for  $\mathbf{U}^1$  would use an estimate of  $\mathbf{U}^2$  as side information, and provide the resulting estimate of  $\mathbf{U}^1$  as side information for the decoding of  $\mathbf{U}^2$ , with this process continuing iteratively. The interpretation of joint source-channel coding of correlated sources as a problem of channel coding with side information is also straightforward. The only differences with respect to the case of pure source coding described above are: (i) sequence  $\mathbf{C}^1$  may include systematic bits, and (ii) sequence  $\mathbf{C}^1$  is corrupted by the channel noise, producing sequence  $\mathbf{O}^1$ , which is available at the decoder.

Turbo-like codes are very well suited to be applied in the context described in Figure 9. The reason is two fold. First, they are pseudorandom codes, and therefore adequate to achieve the theoretical limits corresponding to random codes (by using one turbo-like code as encoder for each of the correlated sources). Second, turbo-like codes are very well prepared to exploit side information. In order to do so, the known probabilistic description of the different sequences available at the decoder can be easily incorporated in the decoding process, which will be (in general) represented by a graph. Moreover, even if the correlation model is not available at the decoder, it is still possible to estimate it jointly with the decoding process (in many occasions with little performance degradation). Sections 7 and 8 develop these ideas for the case of LDGM codes. Although the simulation results presented there focus on the case of joint source-channel coding, the particularization of the proposed schemes for the case of pure source coding results in performances very close to the Slepian-Wolf limit [16].

## 7. LDGM CODES FOR CORRELATED SOURCES: PROPOSED SYSTEM

As explained in Section 5, although the theoretical limits for transmission of correlated sources over noisy channels can be achieved by separation between source and channel coding,

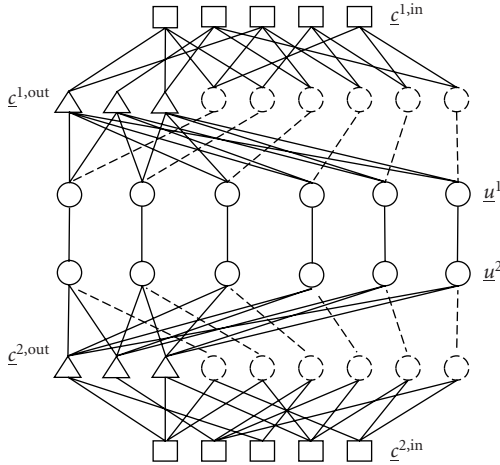


FIGURE 10: Graph representing the joint source-channel decoder for transmission of correlated sources over separated noisy channels.

it may be advantageous in practical applications to use a joint source-channel coding approach, such as the one presented in this paper and shown in Figure 8 [35, 36]. This approach can be particularized into some special cases such as source coding of a single source (by ignoring the other source and considering noiseless channels), distributed source coding (by considering noiseless channels) [37, 38], and joint source-channel coding of single sources (by ignoring the other source).

For the development contained in this paper, we denote the two correlated binary information sequences as  $\mathbf{U}^1 = u_1^1 u_2^1 \dots$  and  $\mathbf{U}^2 = u_1^2 u_2^2 \dots$  with  $u_k^j \in \{0, 1\}$ . The correlation model is established by first generating the symmetric i.i.d. sequence  $\mathbf{U}^1$  ( $P(u_k^1 = 0) = P(u_k^1 = 1) = 1/2$ ). Then, the sequence  $\mathbf{U}^2$  is defined as  $u_k^2 = u_k^1 \oplus e_k$ , where  $\oplus$  indicates modulus 2 addition and  $e_k$  is a random variable which takes value 1 with probability  $p$  and value 0 with probability  $1 - p$ . Each source is independently encoded with a system composed of a serial concatenation of two LDGM codes.<sup>4</sup> For source  $j$ , the coded bits generated by the outer encoder,  $\underline{c}^{j,\text{out}}$ , and the information bits,  $\underline{u}^j$ , constitute the systematic bit nodes for the inner encoder, which further generates the inner coded bits  $\underline{c}^{j,\text{in}}$ . After encoding, the resulting bits are sent through the corresponding noisy channel, and decoded in the common receiver by applying the belief propagation algorithm over the graph representing both decoders, which is shown in Figure 10.

Several activation schedules can be utilized in the decoding process, and, since the graph presents cycles, they can lead to different performance. We consider the five different activation schedules shown below, where each repetition constitutes one iteration. Notation is consistent with the channel coding case explained in Section 3.

(i) Schedule 1 (*flooding*): repeat

$$[\underline{u}^1, \underline{c}^{1,\text{in}}, \underline{c}^{1,\text{out}}, \underline{u}^2, \underline{c}^{2,\text{in}}, \underline{c}^{2,\text{out}}].$$

(ii) Schedule 2: repeat

$$\begin{aligned} &\underline{u}^1, [\underline{c}^{1,\text{in}}, \underline{c}^{1,\text{out}}], [\underline{c}^{1,\text{in}}, \underline{c}^{1,\text{out}}], \underline{u}^1, \\ &\underline{u}^2, [\underline{c}^{2,\text{in}}, \underline{c}^{2,\text{out}}], [\underline{c}^{2,\text{in}}, \underline{c}^{2,\text{out}}], \underline{u}^2. \end{aligned}$$

(iii) Schedule 3: repeat

$$\underline{u}^1, \underline{c}^{1,\text{in}}, \underline{c}^{1,\text{out}}, \underline{u}^1, \underline{u}^2, \underline{c}^{2,\text{in}}, \underline{c}^{2,\text{out}}, \underline{u}^2.$$

(iv) Schedule 4: repeat

$$\underline{u}^1, \underline{c}^{1,\text{in}}, \underline{u}^1, \underline{c}^{1,\text{out}}, \underline{u}^1, \underline{u}^2, \underline{c}^{2,\text{in}}, \underline{u}^2, \underline{c}^{2,\text{out}}, \underline{u}^2.$$

(v) Schedule 5 (see [35]):

$$[\underline{c}^{1,\text{out}}, \underline{c}^{2,\text{out}}].$$

$$\text{Repeat } \underline{u}^1, \underline{c}^{1,\text{in}}, \underline{u}^2, \underline{c}^{2,\text{in}}.$$

Repeat  $[\underline{u}^1, \underline{c}^{1,\text{out}}, \underline{u}^2, \underline{c}^{2,\text{out}}]$ , without exchanging information between  $\underline{u}^1$  and  $\underline{u}^2$ .

## 8. SIMULATION RESULTS FOR JOINT SOURCE-CHANNEL CODING OF CORRELATED SOURCES

In this section, we first analyze the performance of the five activation schedules introduced in last section when each source is independently channel encoded by the serial concatenation of a  $(6.5, 6.5)^5$  inner LDGM code and a regular  $(4, 76)$  outer LDGM code (i.e., overall rate  $R_c = 0.475$ ) and transmitted through a noisy channel. The length of the information sequences is assumed to be  $L = 9500$  and the correlation parameter is fixed to  $p = 0.1$ . Figure 11 shows the BER versus  $E_{so}/N_0$  when AWGN channels are considered. For schedules 1 to 4, no errors were observed at  $E_{so}/N_0 = -0.7$  dB after simulating more than 10 000 blocks. Since the theoretical limit in this case corresponds to  $E_{so}/N_0 = -1.85$  dB, the proposed system is within 1.15 dB from this limit. Similar results for the case of ideally interleaved Rayleigh fading channels with perfect CSI at the receiver are shown in Figure 12. In this case, the gap with respect to the theoretical limit is around 1.5 dB. Notice that, for both figures, schedules 1 to 4 have very similar performances and their curves basically overlap. For schedule 5, which decodes first the inner code and then tries to eliminate the error floor with the outer code (without further exchange of information between the outer and the inner codes), the gap from the theoretical limit is larger.

Table 1 shows the maximum, minimum, and average numbers of iterations required to achieve convergence at

<sup>4</sup>The use of a single LDGM code results in intolerable error floors.

<sup>5</sup>By a noninteger degree such as 6.5, we mean that half of the nodes have degree 6 and half of them have degree 7. All the fractional degrees utilized in this paper are generated from the proper combination of two consecutive integers.



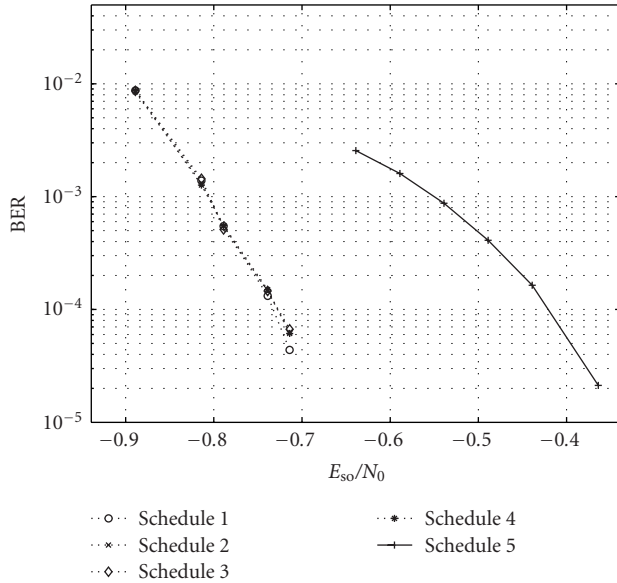


FIGURE 11: For the proposed joint source-channel coding scheme consisting of the serial concatenation of a (6.5,6.5) inner and a (4,76) outer LDGM code (overall rate  $R_c = 0.475$ ), performance of different activation schedules for correlation parameter  $p = 0.1$  and AWGN channels is presented. For schedules 1 to 4, no errors were observed at  $E_{so}/N_0 = -0.7$  dB after simulating more than 10 000 blocks.

some values of  $E_{so}/N_0$  in Figures 11 and 12. Notice that each iteration in schedule 2 corresponds roughly to two iterations in schedules 1, 3, and 4 (most of the complexity is produced in the activation of the coded bit nodes). Taking this into account, we can observe that in average schedules 1 to 4 require approximately the same number of iterations. Notice that the number of iterations required for schedule 5 are obtained at different values of  $E_{so}/N_0$  than those of schedules 1 to 4, which means that the comparison between schedule 5 and the other schedules is not very significant.

In order to further assess the performance of the proposed system, we consider different values of the parameter  $p$  and study the system performance utilizing schedule 1. As before, the length of the information sequence is fixed to  $L = 9500$ . For different values of  $p$ , we use the same (4,76) outer LDGM code as before, but we consider different inner codes in order to optimize performance. Simulation results are presented in Table 2 for AWGN channels and in Table 3 for ideally interleaved Rayleigh fading channels with perfect CSI at the receiver. In both cases, the optimum degree of the inner code decreases with parameter  $p$ . For all different values of  $p$ , at a bit error rate of  $10^{-5}$ , the gap between the theoretical limit and the proposed system is within 1.8 dB for the AWGN channel and within 2.2 dB for the Rayleigh fading channel. Notice that this gap increases when  $p$  gets smaller, which was already pointed out in previous related work [21, 25, 26]. The gain of the proposed system is evident if we realize that, when the source correlation is not exploited in the decoding process, the achievable theoretical limits for  $E_{so}/N_0$  are

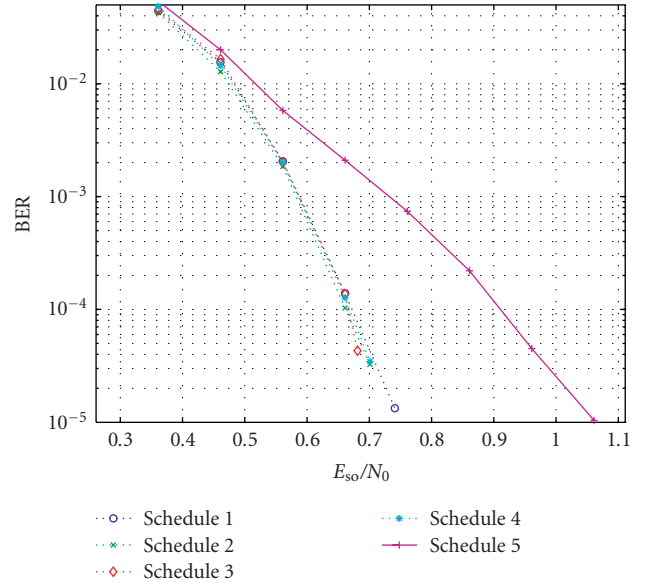


FIGURE 12: For the proposed joint source-channel coding scheme consisting of the serial concatenation of a (6.5,6.5) inner and a (4,76) outer LDGM codes (overall rate  $R_c = 0.475$ ), performance of different activation schedules for correlation parameter  $p = 0.1$  and ideally interleaved Rayleigh fading channels with perfect CSI at the receiver is presented.

0.08 dB and 1.6 dB for the AWGN and Rayleigh fading channel, respectively. The proposed approach achieves a performance (in terms of convergence threshold) similar to the system proposed in [25, 26] for joint source-channel coding of correlated sources over separated AWGN channels using turbo codes. Moreover, after simulating the same number of blocks as in [25, 26], no error floor could be observed here. As shown in the appendix, the use of LDGM codes instead of turbo codes leads to a lower decoding complexity.

## 9. CONCLUSION

We proposed the use of LDGM codes for channel coding and joint source-channel coding of correlated sources over noisy channels. In order to avoid error floors, it is necessary to utilize concatenated schemes. However, they should not be decoded utilizing the equivalent LDGM code resulting from the concatenation, but by combining the graphs of the constituent codes. In terms of encoding/decoding complexity, the proposed scheme presents complexity advantages with respect to turbo and standard LDPC codes. For channel coding, the performance over BSCs, AWGN channels and ideally interleaved Rayleigh fading channels with perfect CSI at the receiver is comparable to that of turbo codes and standard irregular LDPC codes, and close to the theoretical limits even without much code design optimization. In the case of correlated sources, where previous work is almost nonexistent, the proposed system also achieves a performance close to the theoretical limits and similar to those of turbo codes.

TABLE 1: Minimum, average, and maximum number of iterations required to achieve convergence for the schemes considered in Figures 11 and 12, consisting of the serial concatenation of a (6.5,6.5) inner and a (4,76) outer LDGM codes (overall rate  $R_c = 0.475$ ).

		Schedule 1	Schedule 2	Schedule 3	Schedule 4	Schedule 5
AWGN	$E_{so}/N_0$ (dB)	-0.7	-0.7	-0.7	-0.7	-0.3
	Min	18	11	18	15	37
	Average	30.9	17.3	29.5	25.7	59.5
	Max	77	59	97	75	97
Rayleigh	$E_{so}/N_0$ (dB)	0.8	0.8	0.8	0.8	1.1
	Min	18	11	17	16	19
	Average	27.8	15.7	26.3	23.4	29.0
	Max	54	31	49	50	90

TABLE 2: For AWGN channels and different correlation parameters  $p$ , theoretical limit for  $E_{so}/N_0$  in dB ( $[E_{so}/N_0]_l$ , taken in steps of 0.01 dB), value of  $E_{so}/N_0$  in dB for which the proposed system achieves a BER less than  $10^{-5}$  ( $[E_{so}/N_0]_s$ ), and gap (taken in steps of 0.05 dB) between the theoretical limit and the performance of the proposed system.

$p$	$[E_{so}/N_0]_l$	$[E_{so}/N_0]_s$	Gap	Inner code
0.2	-0.96	0.06	< 1.00	(6.5,6.5)
0.1	-1.84	-0.69	< 1.15	(6.5,6.5)
0.05	-2.56	-1.21	< 1.35	(6.25,6.25)
0.025	-3.07	-1.57	< 1.50	(6,6)
0.01	-3.47	-1.72	< 1.75	(5.75,5.75)

TABLE 3: For ideally interleaved Rayleigh fading channels with perfect CSI at the receiver and different correlation parameters  $p$ , theoretical limit for  $E_{so}/N_0$  in dB ( $[E_{so}/N_0]_l$ , taken in steps of 0.01 dB), value of  $E_{so}/N_0$  in dB for which the proposed system achieves a BER less than  $10^{-5}$  ( $[E_{so}/N_0]_s$ ) and gap (taken in steps of 0.05 dB) between the theoretical limit and the performance of the proposed system.

$p$	$[E_{so}/N_0]_l$	$[E_{so}/N_0]_s$	Gap	Inner code
0.2	0.41	1.76	< 1.35	(6.5,6.5)
0.1	-0.74	0.76	< 1.50	(6.5,6.5)
0.05	-1.62	-0.02	< 1.60	(6.25,6.25)
0.025	-2.23	-0.38	< 1.85	(6,6)
0.01	-2.71	-0.51	< 2.20	(5.75,5.75)

## APPENDIX

### CODING COMPLEXITY: LDGM VERSUS TURBO CODES

The encoding of a systematic LDGM code involves computation of the parity bits, each of which only depends on a finite number of systematic bits. Hence, similar to turbo codes, LDGM codes are encodable in linear time. From now on we will focus on the comparison between the two in terms of decoding complexity.

#### Complexity per decoding iteration

Reference [39] provides a detailed analysis on the decoding complexity of turbo codes. The main result is that for a turbo code with constituent encoders having rate  $k/n$  and  $S$  states, the total number of additions/subtractions (additions) and multiplications/divisions (multiplications) per information bit and per iteration are given by

- (i) additions [turbo ( $S, n$ )] =  $4(3S + n - 4)$ ,
- (i) multiplications [turbo ( $S, n$ )] =  $2(8S + 2n + 5)$ .

We now analyze the decoding complexity of an  $(X, Y)$  LDGM code by following the development in [5] and our definitions of  $Q_{mk}^x$  and  $R_{mk}^x$ . Because of their lower complexity, we will disregard operations consisting of additions/multiplications by constants (notice that [39] disregards table look ups and maximum operations). We proceed in two steps. First, we calculate the number of operations required in the processing of a coded bit node. Second, we

look at the complexity in an information bit node. The total number of operations per information bit and per iteration will be the sum of the operations required in all the coded bit nodes plus the operations performed in all the information bit nodes divided by the total number of information bit nodes.

In order to calculate the number of operations in each coded bit node, notice that in (2) and (3) in this paper  $R_{mk}^1 = 1 - R_{mk}^0$ . Therefore, once (2) is calculated, (3) can be obtained without any additional complexity. Following the notation in [5], we define  $\delta Q_{mk} = Q_{mk}^0 - Q_{mk}^1 = 1 - 2Q_{mk}^1$ . We also define  $D_m = (-1)^{\epsilon_m} \prod_{k \in L(m)} \delta Q_{mk}$ . Then, (49) in [5] can be calculated as  $\delta r_{mk} = D_m / (1 - 2Q_{mk}^1)$ . In this way, since as indicated in [5]  $r_{mk}^0 = (1 + \delta r_{mk})/2$  and  $r_{mk}^1 = (1 - \delta r_{mk})/2$ , (2) in this paper can be expressed as

$$\begin{aligned}
 R_{mk}^0 &= (1 - \gamma_m) \frac{1 + D_m / (1 - 2Q_{mk}^1)}{2} + \gamma_m \frac{1 - D_m / (1 - 2Q_{mk}^1)}{2} \\
 &= \frac{1}{2} + \frac{D_m (1 - 2\gamma_m)}{2(1 - 2Q_{mk}^1)}, \quad k = 1 \cdots Y.
 \end{aligned} \tag{A.1}$$

In order to calculate  $R_{mk}^0$ , we first calculate  $D_m$ , which requires  $Y - 1$  multiplications. Then, we utilize one more multiplication to obtain  $D_m(1 - 2\gamma_m)$ , and finally we calculate  $R_{mk}^0$  for  $k = 1 \cdots Y$ , which requires  $Y$  more multiplications.

Therefore, we just need  $2Y$  multiplications to perform all the processing required in a coded bit node. Since there are a total of  $N(1 - R_c)$  coded bit nodes and  $NR_c$  information bits (where  $R_c = Y/(X + Y)$  as indicated in this paper), the total amount of processing in the coded bits divided by the number of information bits is  $2Y(1 - R_c)/R_c = 2X$  multiplications.

In order to calculate the number of operations performed in an information bit node, we follow (50)–(53) in [5]. Notice that  $Q_{mk}^0 = \alpha_{mk} Q_k^0 / R_{mk}^0$ ,  $m = 1 \cdots X$ . By forcing  $Q_{mk}^0 + Q_{mk}^1 = 1$ ,  $Q_{mk}^0$  can be calculated as  $Q_{mk}^0 = 1 / (1 + (Q_k^1 / Q_k^0)(R_{mk}^0 / (1 - R_{mk}^0)))$ . Therefore, after calculating  $\alpha_k Q_k^0$  and  $\alpha_k Q_k^1$ , which requires  $2X$  multiplications, and  $Q_k^1 / Q_k^0$ , which requires another multiplication, the calculation of  $Q_{mk}^0$  for a fixed  $m$  (counting an inversion as a multiplication) can be performed with 2 divisions. Therefore, the total number of operations to calculate all  $Q_{mk}^0$ ,  $m = 1 \cdots X$ , is  $4X + 1$  multiplications/divisions. Since  $Q_{mk}^1 = 1 - Q_{mk}^0$ , no additional operations are required in an information bit node. Hence, the number of operations per information bit and per iteration in an  $(X, Y)$  LDGM code is as follows

- (i) additions [LDGM  $(X, Y)$ ] = 0,
- (ii) multiplications [LDGM  $(X, Y)$ ] =  $2X + 4X + 1 = 6X + 1$ .

For instance, a (6,6) LDGM code performs 37 multiplications per information bit and per iteration, while a serial concatenated LDGM scheme with codes (6,6) and (4,76) performs 62 multiplications. A turbo code with comparable performance ( $S = 8$  and  $n = 2$ ) requires 88 additions and 146 multiplications (plus the table look ups and maximum operations which are disregarded).

#### Total number of decoding iterations

The total number of iterations required for convergence cannot be predicted through analysis. Table 1 in this paper shows the number of iterations required to achieve convergence for different activation schedules of the concatenated LDGM scheme in the case of joint source-channel coding. This number is greater than the one usually required in turbo coding schemes, but it is not enough to compensate the advantage of LDGM codes in each iteration. Compensation does not occur in the channel coding case either. For instance, for the concatenated scheme used over AWGN channels ( $[(6,6)(4,76)]$  with block size 20 000), the average number of iterations at an  $E_b/N_0$  of 0.8 dB above the Shannon limit is 21.7, which is about twice the number of iterations required in a comparable turbo code.

#### DISCLAIMER

The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the US Government.

#### ACKNOWLEDGMENT

The material in this paper was presented in part at Asilomar '02 and ICIP '03.

#### REFERENCES

- [1] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Proc. IEEE International Communications Conference (ICC '93)*, vol. 2, pp. 1064–1070, Geneva, Switzerland, May 1993.
- [2] R. G. Gallager, "Low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 8, no. 1, pp. 21–28, 1962.
- [3] R. G. Gallager, *Low-Density Parity-Check Codes*, MIT Press, Cambridge, Mass, USA, 1963.
- [4] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electronic Letters*, vol. 33, no. 6, pp. 457–458, 1997.
- [5] D. J. C. MacKay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 399–431, 1999.
- [6] J.-F. Cheng and R. J. McEliece, "Some high-rate near capacity codes for the Gaussian channel," in *Proc. 34th Annual Allerton Conference on Communications, Control and Computing*, Allerton, Ill, USA, October 1996.
- [7] T. R. Oenning and J. Moon, "A low-density generator matrix interpretation of parallel concatenated single bit parity codes," *IEEE Trans. Magn.*, vol. 37, no. 2, pp. 737–741, 2001.
- [8] L. Ping, S. Chan, and K. L. Yeung, "Iterative decoding of multi-dimensional concatenated single parity check codes," in *Proc. IEEE International Communications Conference (ICC '98)*, vol. 1, pp. 131–135, Atlanta, Ga, USA, June 1998.
- [9] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Mag.*, vol. 19, no. 2, pp. 51–60, 2002.
- [10] S. Shamai (Shitz), S. Verdú, and R. Zamir, "Systematic lossy source/channel coding," *IEEE Trans. Inform. Theory*, vol. 44, no. 2, pp. 564–579, 1998.
- [11] A. D. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inform. Theory*, vol. 20, no. 1, pp. 2–10, 1974.
- [12] 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.
- [13] S. S. Pradhan and K. Ramchandran, "Distributed source coding: symmetric rates and applications to sensor networks," in *Proc. IEEE Data Compression Conference (DCC '00)*, pp. 363–372, Snowbird, Utah, USA, March 2000.
- [14] S. S. Pradhan, K. Ramchandran, and R. Koetter, "A constructive approach to distributed source coding with symmetric rates," in *Proc. IEEE International Symposium on Information Theory (ISIT '00)*, p. 178, Piscataway, NJ, USA, June 2000.
- [15] J. Garcia-Frias and Y. Zhao, "Data compression of unknown single and correlated binary sources using punctured turbo codes," in *Proc. 39th Annual Allerton Conference on Communication, Control, and Computing*, Allerton, Ill, USA, October 2001.
- [16] J. Garcia-Frias and Y. Zhao, "Compression of correlated binary sources using turbo codes," *IEEE Commun. Lett.*, vol. 5, no. 10, pp. 417–419, 2001.
- [17] J. Barros and S. D. Servetto, "On the capacity of the reachback channel in wireless sensor networks," in *Proc. IEEE Workshop on Multimedia Signal Processing (Special) Session on "Signal Processing for Wireless Networks"*, St. Thomas, Virgin Islands, USA, December 2002.
- [18] J. Barros and S. D. Servetto, "Network information flow with correlated sources. (Original title: the sensor reachback problem)," submitted to *IEEE Trans. Inform. Theory*, November 2003, <http://cn.ece.cornell.edu/publications/papers/20031112/>.

- [19] Q. Zhao and M. Effros, "Lossless and near-lossless source coding for multiple access networks," *IEEE Trans. Inform. Theory*, vol. 49, no. 1, pp. 112–128, 2003.
- [20] P. Mitran and J. Bajcsy, "Turbo source coding: a noise-robust approach to data compression," in *Proc. IEEE Data Compression Conference (DCC '02)*, p. 465, Snowbird, Utah, USA, April 2002.
- [21] 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 Multimedia Signal Processing Workshop*, pp. 53–56, St. Thomas, Virgin Islands, USA, December 2002.
- [22] J. Garcia-Frias and J. D. Villasenor, "Combining hidden Markov source models and parallel concatenated codes," *IEEE Commun. Lett.*, vol. 1, no. 4, pp. 111–113, 1997.
- [23] G.-C. Zhu and F. Alajaji, "Turbo codes for nonuniform memoryless sources over noisy channels," *IEEE Commun. Lett.*, vol. 6, no. 2, pp. 64–66, 2002.
- [24] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Proc. IEEE Data Compression Conference (DCC '02)*, pp. 252–261, Snowbird, Utah, USA, April 2002.
- [25] 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.
- [26] J. Garcia-Frias and Y. Zhao, "Near Shannon/Slepian-Wolf performance for unknown correlated sources over AWGN channels," *IEEE Trans. Comm.*, vol. 53, no. 4, pp. 555–559, 2005.
- [27] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, Morgan Kaufmann, San Mateo, Calif, USA, 1988.
- [28] 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.
- [29] J. Garcia-Frias and W. Zhong, "Approaching Shannon performance by iterative decoding of linear codes with low-density generator matrix," *IEEE Commun. Lett.*, vol. 7, no. 6, pp. 266–268, 2003.
- [30] S. G. Wilson, *Digital Modulation and Coding*, Prentice-Hall, Englewood Cliffs, NJ, USA, 1996.
- [31] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 619–637, 2001.
- [32] T. J. Richardson and R. L. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 599–618, 2001.
- [33] T. M. 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.
- [34] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inform. Theory*, vol. 19, no. 4, pp. 471–480, 1973.
- [35] J. Garcia-Frias, W. Zhong, and Y. Zhao, "Iterative decoding schemes for source and channel coding of correlated sources," in *Proc. 36th Asilomar Conference on Signals, Systems, and Computers (ASILOMAR '02)*, Pacific Grove, Calif, USA, November 2002.
- [36] W. Zhong, H. Lou, and J. Garcia-Frias, "LDGM codes for joint source-channel coding of correlated sources," in *Proc. ICIP'03*, Barcelona, Spain, September 2003.
- [37] T. Murayama, "Statistical mechanics of linear compression codes in network communication," *Europhysics Letters*, preprint, 2001.
- [38] 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.
- [39] M. Y. Alias, F. Guo, S. X. Ng, T. H. Liew, and L. Hanzo, "LDPC and turbo coding assisted space-time block coded OFDM," in *Proc. IEEE Vehicular Technology Conference (VTC '03)*, vol. 4, pp. 2309–2313, Jeju, Korea, April 2003.

---

**Wei Zhong** received the B.S. degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2001. He is currently working towards the Ph.D. degree at the University of Delaware, USA. His research interests are in communications, turbo codes, joint source-channel coding, and coding for multiterminal sources.



**Javier Garcia-Frias** received the Ingeniero de Telecomunicación degree from Universidad Politécnica de Madrid, Spain, in 1992, the Licenciado en Ciencias Matemáticas degree from UNED, Madrid, in 1995, and the Ph.D. degree in electrical engineering from UCLA, in 1999. In 1992 and from 1994 to 1996, he was with Telefonica I + D in Madrid. From September 1999 to August 2003, he was an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Delaware, where he is currently an Associate Professor. His research interests are in the area of information processing in communications and biological systems, with a focus on wireless communications, iterative decoding schemes for source and channel coding, coding for multiterminal sources, joint source-channel coding, and cellular regulatory networks. Javier Garcia-Frias is a recipient of a 2001 NSF CAREER Award and of a 2001 Presidential Early Career Award (PECASE) in support of his communications program. He is listed in the 2003 and 2004 editions of Who's Who in America, he is an Associate Editor of the *IEEE Transactions on Wireless Communications* and *IEEE Transactions on Signal Processing*, and a Member of the Signal Processing for Communications Technical Committee (SPCOM-TC) in the *IEEE Signal Processing Society*.

