LDPC Codes for Non-Uniform Memoryless Sources and Unequal Energy Allocation

Idoia Ochoa, *Student Member, IEEE*, Pedro M. Crespo, *Senior Member, IEEE*, and Mikel Hernaez, *Student Member, IEEE*

***Abstract*—In this paper, we design a new energy allocation strategy for non-uniform binary memoryless sources encoded by Low-Density Parity-Check (LDPC) codes and sent over Additive White Gaussian Noise (AWGN) channels. The new approach estimates the *a priori* probabilities of the encoded symbols, and uses this information to allocate more energy to the transmitted symbols that occur less likely. It can be applied to systematic and non-systematic LDPC codes, improving in both cases the performance of previous LDPC based schemes using binary signaling. The decoder introduces the source non-uniformity and estimates the source symbols by applying the SPA (Sum Product Algorithm) over the factor graph describing the code.**

***Index Terms*—Systematic and non-systematic LDPC codes, non-uniform memoryless sources, unequal energy allocation.**

I. Introduction

WE consider the transmission of the information generated by a non-uniform memoryless source with probability distribution (p_0, $p_1 = 1-p_0$). The standard approach to tackle this problem has been to separate the encoding process in two parts: first, a source encoder capable of compressing the source up to its theoretical limit (which is given by its entropy $H(p_0)$), and second, a capacity achieving channel code. Consequently, the lower limit E_{br}, average energy per source symbol, is given by

$$\frac{E_{br}}{N_0} = \frac{2^{2R_c H(p_0)} - 1}{2R_c} \quad (1)$$

where N_o is the one-sided noise power spectral density of the additive gaussian noise, and R_c is the code rate (source symbols per channel symbol).

The justification for this partition lies on Shannon's Separation Theorem. However, when complexity is an issue, the overall performance can be improved if the tasks of source and channel coding are blend together by means of a joint source-channel encoder. In this way, the joint decoder can employ some of the inherent redundancy of the source to alleviate the requirements of the channel encoder. For this reason, in the proposed scheme source and channel coding are jointly performed by an LDPC encoder.

Point-to-point communication schemes including non-uniform memoryless sources and joint source-channel coding have been well-studied for both Turbo and LDPC codes. As to mention, systematic Turbo codes are considered in [1] and [2], where significant improvements are obtained by using Unequal Energy Allocation (UEA) rather than binary signaling (BPSK). In [5] they extend this scheme to the case of sources with memory. In [4] they show that non-systematic Turbo codes perform better than systematic Turbo codes for non-uniform sources. In [3] they combine non-systematic Turbo codes and unequal energy allocation, improving the performance over previous schemes. In [6] and [7] they use systematic and non-systematic LDPC codes, respectively, with binary signaling. To the best of our knowledge, no schemes with LDPC codes and unequal energy allocation are found in the literature.

The remainder of the paper is organized as follows. Section II introduces the proposed system. Then, Section III and IV describe the encoding and decoding process with UEA for systematic and non-systematic LDPC codes, respectively. Simulation results are presented in Section V and some concluding remarks are given in Section VI.

II. System Description

Let $\mathbf{u} \doteq (u_1, \ldots, u_K) \in \{0,1\}^K$ be a binary sequence of length K generated by an i.i.d. source with probability distribution (p_0, p_1). The sequence $\mathbf{u}$ is then encoded by an LDPC code of rate $R_c = K/N$ into the code sequence $\mathbf{c} \doteq (c_1, \ldots, c_N)$, and then modulated by an unequal energy allocation technique, producing an amplitude signal $\mathbf{x} \doteq (x_1, \ldots, x_N)$ of length N. The destination receives $\mathbf{y}$, which is a version of $\mathbf{x}$ corrupted by Additive White Gaussian Noise (AWGN). The decoder estimates $\mathbf{u}$ by applying the SPA over the factor graph describing the corresponding LDPC code, and introducing the source statistics (p_0, p_1).

III. Proposed System with Systematic LDPC Codes

An LDPC code is defined by its low-density parity-check matrix $\mathbf{H}$ of dimension $(N-K) \times N$, which can be regular or irregular. We will consider only regular codes for explanation purposes. The extension to irregular codes is straightforward. We will denote by (d_b, d_c) LDPC codes those codes in which the matrix $\mathbf{H}$ has exactly d_b non-zero entries per column and d_c non-zero entries per row. The encoding is done by simply multiplying the source sequence $\mathbf{u}$ by the systematic generator matrix $\mathbf{G}$ of dimensions $K \times N$, which satisfies $\mathbf{GH}^T = \mathbf{0}$.

The new approach modulates the encoded symbols depending on their *a priori* probability p_{c_i}, for $i = 1, \ldots, N$. Hereafter, we will consider that the first K bits of $\mathbf{c}$ correspond to the source symbols. Therefore, $(p_{c_i}(0), p_{c_i}(1)) = (p_0, p_1)$, for $i = 1, \ldots, K$. For the remaining $N-K$ parity symbols, we estimate the *a priori* probabilities p_{c_i}, for $i = (K+1), \ldots, N$. These probabilities differ from 0.5 when the next two conditions hold:

Manuscript received April 22, 2010. The associate editor coordinating the review of this letter and approving it for publication was V. Stankovic.

The authors are with the Centro de Estudios e Investigaciones Técnicas de Gipuzkoa (CEIT) and TECNUN (University of Navarra), 20009 San Sebastian, Spain (e-mail: {iochoa, pcrespo, mhernaez}@ceit.es).

Digital Object Identifier 10.1109/LCOMM.2010.09.100662

1) The source symbols are generated by a non-uniform source.
2) The generator matrix **G** has only a few ones per column.

The *a priori* probabilities of the parity symbols can be estimated off-line, as follows. Since $\mathbf{c} = \mathbf{u}\mathbf{G}$, the value of the parity bit c_i is given by multiplying the source sequence **u** by the column i of **G**. Let us denote by $w(i)$ the number of non-zero elements of column i. Then, the value of the parity bit c_i is given by the sum of the $w(i)$ source symbols located at the same position as the $w(i)$ ones of column i. Therefore, c_i will be 0 when the considered source symbols have an even number of 1's, and 1, otherwise. We propose a simple algorithm based on this principle to estimate these probabilities. For $i = (K+1), \ldots, N$ we perform the following pseudo program:

1) Initialize the variables $p_{c_i}(0)$ and j to 0.
2) Update $p_{c_i}(0)$ as $p_{c_i}(0) = p_{c_i}(0) + \binom{w(i)}{j} p_1^j p_0^{(w(i)-j)}$, where the binomial coefficient indicates the number of sequences of length $w(\cdot)$ with j 1's and $(w(\cdot) - j)$ 0's.
3) Increment variable j by 2, so the number of ones is still even, and perform the algorithm from step 2 until $j \leqslant w(i)$.
4) $p_{c_i}(1) = 1 - p_{c_i}(0)$

Notice that the actual position of the $w(\cdot)$ ones is irrelevant, since the source is stationary and all the source symbols have the same a priori probability (p_0, p_1).

Finally, the coded symbols are modulated following the UEA strategy given by

$$\begin{cases} -\sqrt{\frac{p_{c_i}(1)}{p_{c_i}(0)}E} \text{ , if } c_i = 0 \\ +\sqrt{\frac{p_{c_i}(0)}{p_{c_i}(1)}E} \text{ , if } c_i = 1 \end{cases} \quad (2)$$

which allocates more energy to the less likely symbols, as in [2]. This results in asymmetric 2-PAM constellations with the same average energy per coded symbol (i.e., $E[c_i^2] = E$, $\forall i$), and maximized distance between the two points. This makes the symbols with higher distance (those with probability close to 0 or/and 1) be more protected against noise than those with lower distance (those with probability close to 0.5).

At the destination, the decoder estimates **u** by applying the SPA over the factor graph describing the LDPC code. Due to the UEA, the channel probabilities are proportional to

$$\begin{aligned} p(y_i|c_i = 0) &\propto e^{\frac{-\left(y_i + \sqrt{\frac{p_{c_i}(1)}{p_{c_i}(0)}E}\right)^2}{N_0}} \\ p(y_i|c_i = 1) &\propto e^{\frac{-\left(y_i - \sqrt{\frac{p_{c_i}(0)}{p_{c_i}(1)}E}\right)^2}{N_0}}. \end{aligned} \quad (3)$$

Furthermore, the decoder also introduces the source statistics (p_0, p_1) for the systematic symbols. However, it is not required to introduce the estimated *a priori* probabilities for the parity bits since at the first iteration, the parity bit nodes receive that information from the parity check nodes, which are connected to the systematic bit nodes.

IV. Proposed System with Non-Systematic LDPC Codes

We consider the scrambler-LDPC and the splitter-LDPC non-systematic codes proposed in [6] and [7] since they present good performance for channel coding of nonuniform

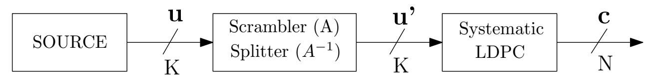

Fig. 1. General scheme of scrambler-LDPC and splitter-LDPC codes.

sources. As shown in Fig. 1, the encoders are implemented by the concatenation of a pre-coder matrix modeled either by a low density squared matrix **A** of size K and column and row degree d_s (scrambler-LDPC), or by its inverse $\mathbf{A}^{-1}$ (splitter-LDPC), followed by a systematic LDPC encoder.

Due to the non-systematic nature of the codes, the *a priori* probabilities p_{c_i} for $i = 1, \ldots, N$ are unknown, and they have to be estimated. We distinguish between the first K symbols of **c**, which correspond to $\mathbf{u}'$, and the remaining $N - K$ symbols, which will be referred to as parity symbols. For the first K symbols, we follow the steps introduced in the previous section, but referring to the pre-coder matrix instead of to the generator matrix **G**. On the other hand, since $\mathbf{c} = \mathbf{u}'\mathbf{G}$ instead of $\mathbf{c} = \mathbf{u}\mathbf{G}$, in order to estimate the *a priori* probabilities of the remaining $N - K$ parity symbols, we have to replace the *a priori* probabilities of the source symbols, i.e., (p_0, p_1), by the *a priori* probabilities of $\mathbf{u}'$, i.e., $(p_{c_i}(0), p_{c_i}(1))$ for $i = 1, \ldots, K$. However, since these probabilities are in general non-stationary, i.e., $p_{c_i} \neq p_{c_j}$, $\forall i \neq j$, one has to modify the steps introduced in the previous section. Instead, for the sake of simplicity, we estimate the *a priori* probabilities of the parity symbols by simulation using

$$\tilde{p}_{c_i}(0) = \frac{1}{M} \sum_{k=1}^{M} \mathbb{I}(c_i(k) = 0), \text{ for } i = (K+1), \ldots, N, \quad (4)$$

where M is the number of sequences simulated, $\mathbb{I}(P)$ is the indicator function that takes the value 1 when the proposition P is true and 0, otherwise, and $c_i(k)$ corresponds to the coded symbol c_i of sequence k. Notice that as $M \longrightarrow \infty$, $\tilde{p}_{c_i} \longrightarrow p_{c_i}$. Observe that due to the pre-coder matrix, the symbols of $\mathbf{u}'$ become more uniformly distributed than those of **u**, making the latter symbol probabilities closer to 0.5. This symbol uniformity will manifest more strongly in the case of splitter-LDPC non-systematic codes since in this case the pre-coder matrix $\mathbf{A}^{-1}$ is not of low density. Therefore, although the encoded symbols are modulated following expression (2), the non-systematic codes are expected to perform worse than their systematic counterpart, since the UEA strategy will have less effect.

At the destination, the decoder estimates **u** by running the SPA over the factor graph describing the non-systematic LDPC code. The decoding graphs of scrambler-LDPC and splitter-LDPC codes are detailed in [[6], Fig. 1-bottom] and [[7], Fig. 4-right], respectively. Both graphs contain K variable nodes corresponding to the source symbols, which allows to include the source non-uniformity (p_0, p_1), and N variable nodes corresponding to the channel symbols, which introduce only the channel probabilities calculated by expression (3).

V. Simulation Results

To assess the performance of the proposed LDPC-UEA scheme over previous LDPC based systems for non-uniform

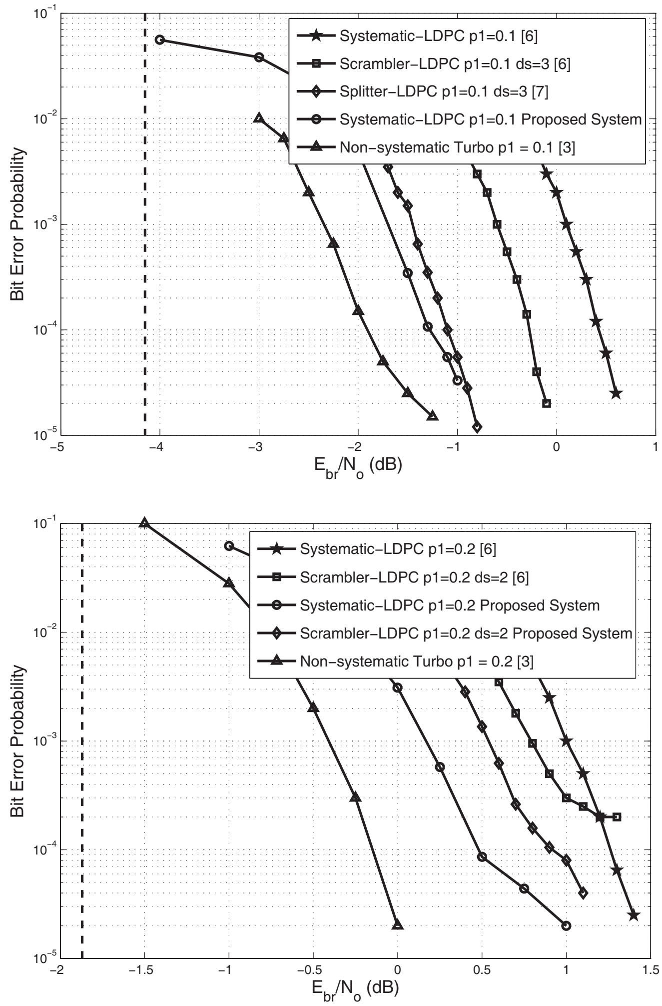

Fig. 2. Bit error probabilities vs. signal-to-noise ratio for $R = 1/2$ LDPC and Turbo codes and distributions $p_1 = 0.1$ (top) and $p_1 = 0.2$ (bottom).

memoryless sources, Montecarlo simulations have been done. To that end, the proposed system is compared with the results shown in [6] and [7], where a regular-(3,6) LDPC systematic code of rate $R = 1/2$ and block length $K = 1000$ is used for the systematic code and for the design component of the non-systematic codes. Simulations have been performed with the same LDPC codes, $10,000$ input blocks and a maximum number of decoding iterations of 100. As shown in Fig. 2 (top), for $p_1 = 0.1$ and BER $= 10^{-4}$ the proposed scheme with the systematic LDPC code is 2.87 dB away from the Shannon limit (1), denoted by a vertical line. A gain of 1.71 dB is obtained with respect to the same systematic LDPC code but now using binary signaling. Furthermore, our system performs 1 dB and 0.18 dB better than the scrambler-LDPC code and the splitter-LDPC code considered in [6] and [7], respectively, both with optimum degree $d_s = 3$. The results for $p_1 = 0.2$ are presented in Fig. 2 (bottom), where the curves of the proposed scheme (systematic and scrambler-LDPC), the systematic LDPC code from [6] and the scrambler-LDPC code with optimum degree $d_s = 2$ from [6] are shown. As suggested in the previous section, the systematic LDPC proposed scheme outperforms all the others and is 2.35 dB away from the Shannon limit. The scrambler-LDPC code has a gain of 0.28 dB with respect to the same code proposed in [6] for equal energy allocation and does not present an error floor. The asymptotic performance of the proposed scheme can be calculated by density evolution. However, since it is not straightforward to apply it for non-uniform sources and no-symmetric constellations, we give an upperbound by simulating with $K = 20,000$. The gap to the Shannon limit is reduced to 1.93 dB ($p_1 = 0.1$) and 1.53 dB ($p_1 = 0.2$).

The proposed system has also proved to have good performance for extremely non-uniform sources. For $p_1 \in \{0.005, 0.01\}$ and a rate $R = 1/2$ irregular LDPC code with $K = 20,000$, the gap to the theoretical limit for the systematic case is 3.91 dB and 3.81 dB ,respectively, for a BER $= 10^{-4}$.

Although the improvement in performance of the proposed LDPC-UEA scheme is substantial when compared to existing LDPC based schemes for non-uniform sources, this is not in general the case when compared to their counterpart, i.e., Turbo code-UEA schemes [1]-[3]. For example, as shown in Fig. 2, an optimized non-systematic Turbo code with the same rate and block length ($K = 1000$) as before, outperforms our scheme by 0.63 dB and 0.62 dB for $p_1 = 0.1$ and $p_1 = 0.2$, respectively, and a BER of 10^{-4} [3].

VI. CONCLUSION

We have proposed a source-controlled LDPC coding scheme for the transmission of non-uniform memoryless sources over AWGN channels. The novel scheme is based on the UEA strategy which allocates more energy to the less likely encoded symbols. This strategy can be applied to systematic and non-systematic LDPC codes, improving in both cases the performance of existing LDPC based schemes using binary signaling. Besides, the proposed system has proved to have good performance even for strongly non-uniform sources.

ACKNOWLEDGMENT

This work was supported in part by the Spanish Ministry of Education and Science and by the European Regional Development Fund through the MultiMIMO TEC2007-68020-C04-03 and the CONSOLIDER-INGENIO 2010 CSD200800010.

REFERENCES

[1] F. Cabarcas, R. D. Souza, and J. Garcia-Frias, “Source-controlled turbo coding of non-uniform memoryless sources based on unequal energy allocation,” in *Proc. IEEE Intern. Symp. on Inform. Theory*, p. 166, June 2004.

[2] F. Cabarcas, R. D. Souza, and J. Garcia-Frias. “Turbo coding of strongly non-uniform memoryless sources with unequal energy allocation and PAM signaling,” *IEEE Trans. Signal Process.*, vol. 54, no. 5, pp. 1942-1946, May 2006.

[3] R. D. Souza, G. I. Shamir, J. Garcia-Frias, and K. Xie, “Non-systematic turbo coding with unequal energy allocation for nonuniform memoryless sources,” in *Proc. IEEE Intern. Symp. on Inform. Theory*, pp. 1893-1897, Sep. 2005.

[4] G. Zhu, F. Alajaji, J. Bajcsy, and P. Mitran, “Non-systematic turbo codes for non-uniform I.I.D. sources over AWGN channels,” *Conference on Information Sciences and Systems*, Princeton University, Mar. 2002.

[5] J. del Ser, P. Crespo, I. Esnaola, and J. Garcia-Frias, “Joint source-channel coding of sources with memory using turbo codes and the Burrows-Wheeler transform,” *IEEE Trans. Commun.*, accepted.

[6] G. I. Shamir and J. J. Boutros, “Non-systematic low-density parity-check codes for nonuniform sources,” in *Proc. IEEE Intern. Symp. on Inform. Theory*, pp. 1898-1902, Sep. 2005.

[7] A. Alloum, J. J. Boutros, G. I. Shamir, and L. Wang, “Non-systematic LDPC codes via scrambling and splitting,” in *Allerton Conference on Communication and Control*, Illinois, Sep. 2005.