

## Research Letter

# Design of Short, High-Rate DVB-S2-Like Semi-Regular LDPC Codes

Luca Barletta and Arnaldo Spalvieri

Dipartimento di Elettronica e Informazione (DEI), Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy

Correspondence should be addressed to Luca Barletta, barletta@elet.polimi.it

Received 11 October 2008; Accepted 5 December 2008

Recommended by Rajesh Khanna

This work focuses on high-rate ( $R > 4/5$ ) moderate-length ( $1000 \leq n < 5000$ ) low-density parity-check codes. High-rate codes allow to maintain good quality of the preliminary decisions that are used in carrier recovery, while a moderate code length allows to keep the latency low. The interleaver of the LDPC matrix that we consider is inspired to the DVB-S2 standard one. A novel approach for avoiding short cycles is analyzed. A modified BP decoding algorithm is applied in order to deal with longer cycles. Simulations and results for the AWGN channel are presented, both for BPSK signalling and for coded modulation based on the partition  $\mathbb{Z}^2/4\mathbb{Z}^2$ .

Copyright © 2008 L. Barletta and A. Spalvieri. 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.

## 1. Introduction

An  $(n, k)$  LDPC code is a binary block code defined as the null space of the sparse “parity check” matrix  $\mathbf{H}_{(n-k) \times n}$ , where  $\mathbf{H}$  can be viewed as the adjacency matrix of the bipartite graph  $G(\{C, V\}, E)$ , the Tanner graph, with  $h_{ij} = 1$  meaning that the  $i$ th  $c$ -node is connected to the  $j$ th  $v$ -node. Richardson et al. [1] studied a method named *Density Evolution* (DE) for finding the pair of distributions  $(\lambda, \rho)$ , as  $n \rightarrow \infty$ , that allows to reach the fastest BER convergence at some probability error target  $\epsilon$ , under BP algorithm. For short and moderate-length LDPC codes, that is, for  $10^3 \leq n < 10^4$ , the design of  $\mathbf{H}$  via DE shows relatively high error floors ( $@P_b \approx 10^{-6}$ ), therefore a different approach is required. The performance can be improved by conditioning the construction of  $\mathbf{H}$  in such a way that short cycles are avoided. A technique for the elimination of short cycles in regular LDPC codes has been recently presented in [2]. However, regular codes are outperformed by irregular or semi-regular codes.

Eroz et al. [3] proposed a particular structure of  $\mathbf{H}$  that leads to irregular codes. This structure has been adopted for standardization of DVB-S2 applications (the DVB-S2 standard includes 2 versions with  $n = 64800$  and  $n = 16200$ ) and belongs to the family of IRA codes [4].

This paper addresses the problem of designing short DVB-S2-like codes, introducing a new technique for the elimination of cycles of length 4 and 6.

The remainder of the paper is outlined as follows. Section 2 introduces the system models taken under consideration and the matrix structure of the DVB-S2 standard. Section 3 explains the features adopted in designing the short-length DVB-S2-like LDPC codes. Section 4 shows simulation results and some comparisons with short codes designed via DE. Finally, some conclusions are drawn in Section 5.

## 2. DVB-S2-Like Parity-Check Matrix

Here follow some recalls of the parity-check matrix structure described in [3].

The LDPC matrix  $\mathbf{H}$  is composed of two submatrices  $\mathbf{H}_1^{(n-k) \times k}$  and  $\mathbf{H}_2^{(n-k) \times (n-k)}$ ,  $\mathbf{H} = [\mathbf{H}_1, \mathbf{H}_2]$ ; the rightmost one is a lower bidiagonal matrix which offers the advantage of simple systematic encoding. Specifically, taking modulo-2 additions of consecutive rows of  $\mathbf{H}_2$ , such that  $\mathbf{h}'_i = \mathbf{h}'_{i-1} \oplus \mathbf{h}_i$  for  $i = 1, \dots, n - k$  and  $\mathbf{h}'_0 = \mathbf{0}$ , one gets the identity matrix  $\mathbf{I}_{(n-k)}$ . Another advantage is to avoid short cycles in Tanner graph involving low-degree  $v$ -nodes, leading to improved performance as stated in [1].

Submatrix  $\mathbf{H}_1$  is built as follows: chosen an opportune value  $M$  that divides both  $n - k$  and  $k$ , randomly generate the first column of  $\mathbf{H}_1$ ,  $\mathbf{h}_1$ , and get the next  $M - 1$  columns by cyclic shifting of  $j((n - k)/M)$  positions, respectively, for  $j = 2, \dots, M$ . The second group of  $M$  columns is created in the same way, just after the first columns of the new group have been generated. This procedure allows to fill the entire submatrix  $\mathbf{H}_1$ .

### 3. Code Design

#### 3.1. Short Cycle Elimination

A cycle of length  $l$  in the Tanner graph is a set of  $l/2$   $v$ -nodes and  $l/2$   $c$ -nodes such that a path of consecutive  $l$  edges starts from a  $v$ -node ( $c$ -node) and ends at the same  $v$ -node ( $c$ -node).

Given two columns  $\mathbf{h}_i$  and  $\mathbf{h}_j$  of  $\mathbf{H}$ , a length-4 cycle is formed when the inner product between the 2 columns is greater than 1:

$$\mathbf{h}_i^T \cdot \mathbf{h}_j \geq 2 \quad (1)$$

for some  $i, j \in \{1, \dots, n : i \neq j\}$ . This condition suggests a practical way to check the presence of length-4 cycle in a parity-check matrix.

Focusing on a column of  $\mathbf{H}$ ,  $\mathbf{h}_i$ , let us define the set of the "cyclic-distances" between the nonzero entries of the  $i$ th column

$$D_i = \{d_{p,q}^{(i)} : \mathbf{h}_{p,i} = 1, \mathbf{h}_{q,i} = 1\}, \quad (2)$$

where  $p, q \in \{1, \dots, n - k : p \neq q\}$  and

$$d_{p,q}^{(i)} = \begin{cases} |p - q| & \text{if } 0 < |p - q| \leq \frac{n - k}{2}, \\ n - k - |p - q| & \text{if } \frac{n - k}{2} < |p - q| < n - k. \end{cases} \quad (3)$$

For example, if the nonzero entries of  $\mathbf{h}_i$  are located at the positions  $\{12, 23, 83\}$  with  $n - k = 100$ , then  $D_i = \{d_{12,23}^{(i)} = 11, d_{12,83}^{(i)} = 29, d_{23,83}^{(i)} = 40\}$ . Also we have  $D_i = \{1\}$  for all the columns of  $\mathbf{H}_2$ , briefly denoted with  $D_{\mathbf{H}_2} = \{1\}$ .

Given three columns  $\mathbf{h}_i$ ,  $\mathbf{h}_j$ , and  $\mathbf{h}_k$ , a length-6 cycle is formed when the sum of two cyclic-distances of two distinct columns  $i, j$  equals a cyclic-distance of the third column  $k$ . An example is shown below

$$\mathbf{H} = \begin{bmatrix} 0 & \overset{\cdot}{1} & \overset{\cdot}{0} & \overset{\cdot}{0} & \overset{\cdot}{1} \\ 1 & \overset{\cdot}{0} & \overset{\cdot}{0} & \overset{\cdot}{1} & \overset{\cdot}{1} \\ 0 & \overset{\cdot}{1} & \overset{\cdot}{0} & \overset{\cdot}{1} & \overset{\cdot}{0} \\ 1 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (4)$$

with  $D_2 = 2, D_4 = 1$ , and  $D_5 = 1$  a length-6 cycle is formed:  $d_{2,3}^{(4)} + d_{1,2}^{(5)} = d_{1,3}^{(2)}$ .

A complete search for the sets  $\{D_i : i = 1, \dots, n\}$  that avoid the presence of length-6 cycles is computationally expensive as  $n$  and the average weight of the columns grow.

The problem of finding the longest  $(n, k)$  LDPC code that is length-4 and length-6 cycle free is described as follows: let  $D_0 = D_{\mathbf{H}_2}$  be the initial set of admissible cyclic-distances (i.e., length-4 and length-6 cycle free); when a new column  $\mathbf{h}_c$  is added in the design of  $\mathbf{H}_1$  a new set of cyclic-distances  $D_c$  is added to the current set  $D_j$  according to the following update rule:

$$D_{j+1} = D_j \cup D_c \cup (D_j \oplus D_c), \quad (5)$$

as  $j$  runs in  $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ ; the objective is to find the maximum  $j$  s.t. the set  $D_j$  is cycle free. The binary operation  $A \oplus B$  between two sets  $A$  and  $B$  is defined as

$$A \oplus B = \{\min(|a_i - b_j|, n - k - |a_i - b_j|), \forall i \neq j\}. \quad (6)$$

A new column  $D_c$  is an eligible set if

$$D_j \cap (D_c \cup (D_j \oplus D_c)) = \emptyset. \quad (7)$$

A smart strategy is to choose the cyclic distances from a class of rest modulo a small integer  $m$ : in this way all cyclic distances will be eligible. For example, choosing the class of rest 1 modulo  $m = 4$ , namely  $[1]_4$ , the sum of two cyclic-distances drawn from two distinct columns is in the class  $[1]_4 + [1]_4 = [2]_4 \neq [1]_4$ , so a length-6 cycle involving 3 columns cannot be created. When the sum of two distances is greater than  $(n - k)/2$ , the sum is complemented to  $n - k$  as illustrated in (3); usually  $n - k \in [0]_4$  so  $n - k - ([1]_4 + [1]_4) = [0]_4 - [2]_4 = [2]_4 \neq [1]_4$ . This strategy automatically avoid the creation of length-4 cycles.

For some particular choices of  $k, M, m$ , and the average column weight, it is possible that the total number of eligible cyclic distances is not sufficient to build all the  $k/M$  columns. Actually the total number of original columns can be reduced to  $k/M \cdot M/(n - k) = k/(n - k)$ , where  $(n - k)/M$  is the cyclic offset between two consecutive columns in the same  $c$ -node  $M$ -subset. Without loss of generality, consider the first  $c$ -node  $M$ -subset  $\mathbf{H}_M = \{\mathbf{h}_i, i = 1, \dots, M\}$  and its shifted replica by 1 position denoted with  $\mathbf{H}_M^{\gg 1}$ . Obviously, no cycles are created between the two subsets  $\mathbf{H}_M$  and  $\mathbf{H}_M^{\gg 1}$ ; not only, but also the matrix  $[\mathbf{H}_M, \mathbf{H}_M^{\gg i}]$  for  $i \in \{1, 2, \dots, (n - k)/M - 1\}$  does not contain length-4 and length-6 cycles, so an original column can generate  $((n - k)/M) \cdot M - 1 = n - k - 1$  columns without any cycle.

#### 3.2. $v$ -Node Degree Distribution

In the semiregular construction the  $v$ -node degree distribution has the form

$$\lambda(x) = \frac{2(n - k)}{2(n - k) + w_v k} x + \frac{w_v k}{2(n - k) + w_v k} x^{w_v - 1}, \quad (8)$$

that is, the submatrix  $\mathbf{H}_1$  has a constant column degree  $w_v$ . The parameter  $w_v$  can assume only small values to keep the low-density nature of  $\mathbf{H}$  and hence few cycles in the factor graph.

It is known that there exists a tradeoff between the performance in the cliff region and the performance in the

floor region. In [5] it is shown that the floor improves when  $w_v$  increases. However the weight  $w_v$  cannot be increased above a maximum value that depends on the code length and the rate, in order to maintain the low-density nature of  $\mathbf{H}$ . This makes it difficult to obtain good performance in the floor region with codes of moderate length and high rate.

### 3.3. Modified BP Algorithm

The BP decoding algorithm is optimal with tree-like factor graph, a condition practically reached when  $n$  is very large. When the block length is finite, the BP becomes a suboptimal algorithm in the sense that it no longer performs MAP decoding [6, 7]. The main idea in order to outperform the BP decoder is to include a procedure of hard decisions on the  $v$ -nodes' posteriors, called *decimation*. The modified algorithm resembles that in [8]: in the generic iteration the mutual information from the  $v$ -nodes to  $c$ -nodes  $I_{vc}$  is estimated and some hard decisions are taken on the most reliable  $v$ -nodes to reach the desired value of  $I_{vc}$  for the current iteration. The increment  $\Delta I_{vc}$  between two consecutive iterations is assumed to be  $\Delta I_{vc} = 1/l_{\max}$ , such that the expected maximum number of iteration is  $l_{\max}$ .

## 4. Results

### 4.1. BPSK Signalling

In the BPSK system, the receiver computes the *a priori* metrics for the LDPC decoder

$$\mathbf{L}_0 = \frac{2}{\sigma^2} \mathbf{y}, \quad (9)$$

where  $\mathbf{y}$  is the received vector and  $\sigma^2 = N_0/2$  is the variance per dimension of the AWG noise. Several LDPC codes are compared, that are different for code rate ( $R = 9/10$  and  $R = 4/5$ ), code length ( $n = 1000$  and  $n = 5000$ ). Also, we considered LDPC codes designed via DE. The results obtained with DE are not reported here due to the poor performance in the floor region.

In Figure 1, there is a comparison between the original BP decoding algorithm and the modified version, working on the same semiregular code with  $n = 5000$ ,  $R = 0.8$ , and  $w_v = 8$ . The modified version of the BP algorithm exploits the original version with a constant gain in the cliff region.

In Figure 2 are plotted 8 BER curves versus  $E_b/N_0 = 1/8R\sigma^2$  in dB. The solid curves represent the performance of the codes designed with the short cycle elimination procedure, while the dashed curves represent the performance of the codes designed as described in Section 2. The code with  $n = 5000$  and  $R = 9/10$  is 1 dB from capacity and the code with  $R = 4/5$  is about 1.6 dB from capacity ( $@P_b = 10^{-5}$ ); the floor does not appear up to  $P_b = 10^{-7}$ .

### 4.2. Coded Modulation

Let us consider the coded modulation based on the 16-way partition  $\mathbb{Z}^2/4\mathbb{Z}^2$  where  $\log_2 16 = 4$  coded bits  $\mathbf{c} = [c_1, c_2, c_3, c_4]$  address one coset leader in  $[\mathbb{Z}^2/4\mathbb{Z}^2]$ , and

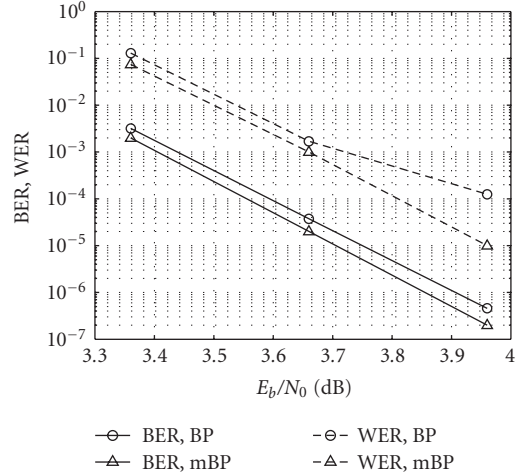


FIGURE 1: Word error rate (WER) and bit error rate (BER) of a semiregular LDPC code ( $n = 5000$ ,  $R = 0.8$ , and  $w_v = 8$ ) with BP decoding and with the modified BP algorithm (mBP).

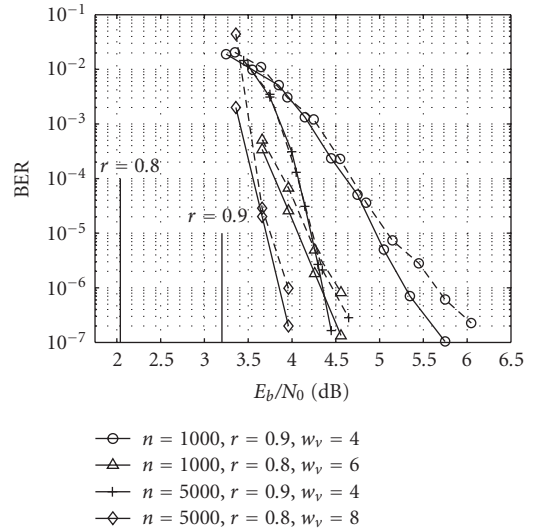


FIGURE 2: Bit error rate (BER) curves of some semiregular LDPC codes with DVB-S2-like structure. BPSK signalling.

$\log_2(128/16) = 3$  uncoded bits  $\mathbf{u} = [u_1, u_2, u_3]$  address one out of 8 subconstellations. The coded vector  $\mathbf{c}$  is drawn from an LDPC encoder of rate  $R = k/n$ ; the normalized redundancy is  $\rho = 4((n - k)/n)[b/2D]$  and the volume to noise ratio (VNR) is

$$\alpha^2 = \frac{2^\rho}{2\pi e \sigma^2}. \quad (10)$$

As suggested in [9] it has been adopted a partial Gray mapping rule.

At the receiver metrics for the  $j$ th received symbol and each of the 4 Gray-mapped bits are computed according to

$$I_{0,j}^{(i)} = \frac{1}{2\sigma^2} (|y_j - z_0^{(i)}|^2 - |y_j - z_1^{(i)}|^2) \quad (11)$$

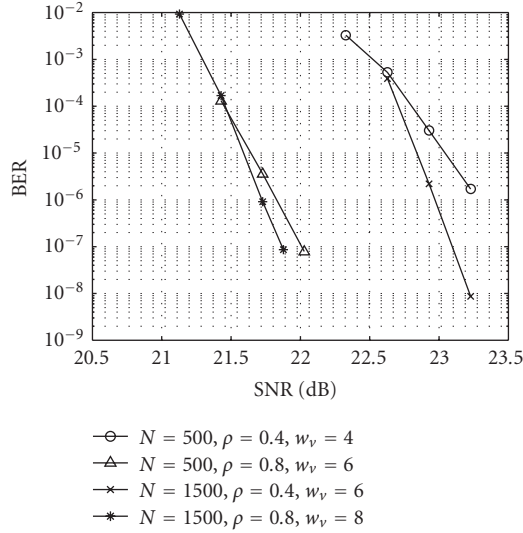


FIGURE 3: Bit error rate (BER) curves of some semiregular LDPC codes with DVB-S2-like structure. 128-CR coded modulation.

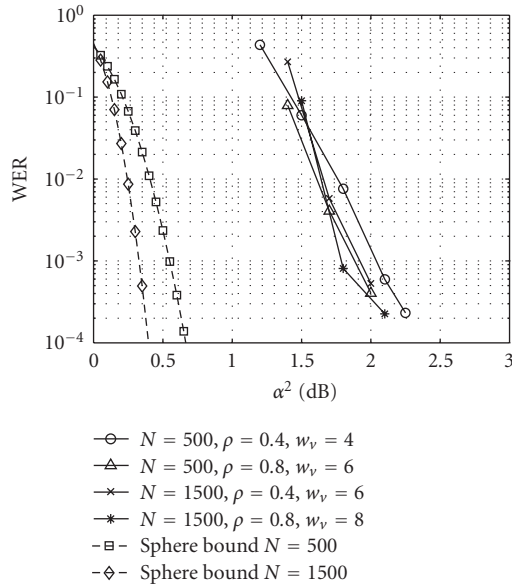


FIGURE 4: Word error rate (WER) curves versus the volume to noise ratio ( $\alpha^2$ ) of some semiregular LDPC codes. Coded modulation based on the partition  $\mathbb{Z}^2/4\mathbb{Z}^2$ .

for  $i = 1, \dots, 4$  and  $j = 1, \dots, n/4$ , such that  $4 \cdot n/4 = n$  metrics are provided to the LDPC decoder.

Four different LDPC codes are compared; they differ for code rate ( $R = 9/10$  and  $R = 4/5$  lead to overall redundancies of  $\rho = 0.4[b/2D]$  and  $\rho = 0.8[b/2D]$ , resp.), code length ( $n = 2000$  and  $n = 6000$  mean  $N = 500$  and  $N = 1500$  2D symbols, resp.). The LDPC decoder computes  $l_{\max} = 30$  iterations of the modified BP algorithm. In Figure 3 are plotted 4 BER performance curves versus SNR in dB. The longer semiregular code shows good performance and error floor below  $P_b = 10^{-8}$ . Figure 4 shows some WER curves versus the volume to noise ratio  $\alpha^2$ , which is recalled in

(10), and the sphere lower bound on the WER for the finite dimensionality of the lattice [10].

## 5. Conclusion

The design of the moderate length instance of the DVB-S2-like parity-check matrix has been considered. Several codes designed with a semiregular approach are worked out and their performance is obtained by running few iterations of the decoding algorithm. The proposed edge assignment and the modified BP algorithm, with a decimation procedure, allow to lower the error floor compared to the same codes designed via density evolution.

## References

- [1] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 619–637, 2001.
- [2] J. Fan, Y. Xiao, and K. Kim, "Design LDPC codes without cycles of length 4 and 6," *Research Letters in Communications*, vol. 2008, Article ID 354137, 5 pages, 2008.
- [3] M. Eroz, F.-W. Sun, and L.-N. Lee, "An innovative low-density parity-check code design with near-Shannon-limit performance and simple implementation," *IEEE Transactions on Communications*, vol. 54, no. 1, pp. 13–17, 2006.
- [4] H. Jin, A. Khandekar, and R. McEliece, "Irregular repeat-accumulate codes," in *Proceedings of the 2nd International Symposium on Turbo Codes and Related Topics*, pp. 1–8, Brest, France, September 2000.
- [5] M. Yang, W. E. Ryan, and Y. Li, "Design of efficiently encodable moderate-length high-rate irregular LDPC codes," *IEEE Transactions on Communications*, vol. 52, no. 4, pp. 564–571, 2004.
- [6] H. Pishro-Nik and F. Fekri, "On decoding of low-density parity-check codes over the binary erasure channel," *IEEE Transactions on Information Theory*, vol. 50, no. 3, pp. 439–454, 2004.
- [7] C. Measson, A. Montanari, and R. Urbanke, "Maxwell construction: the hidden bridge between iterative and maximum a posteriori decoding," *IEEE Transactions on Information Theory*, vol. 54, no. 12, pp. 5277–5307, 2008.
- [8] Q. Wang and C. He, "Approaching 1.53-dB shaping gain with LDGM quantization codes," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM '07)*, pp. 1571–1576, Washington, DC, USA, November 2007.
- [9] U. Wachsmann, R. F. H. Fischer, and J. B. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1361–1391, 1999.
- [10] V. Tarokh, A. Vardy, and K. Zeger, "Universal bound on the performance of lattice codes," *IEEE Transactions on Information Theory*, vol. 45, no. 2, pp. 670–681, 1999.





**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

