

# THE DESIGN OF STRUCTURED REGULAR LDPC CODES WITH LARGE GIRTH

Haotian Zhang and José M. F. Moura

Department of Electrical and Computer Engineering  
Carnegie Mellon University, Pittsburgh, PA 15213  
{haotian, moura}@ece.cmu.edu

## ABSTRACT

This paper introduces three new classes of structured regular  $(n, 2, k)$  LDPC codes with girth 12, 16, and 20, respectively. These codes are systematically constructed, well structured, and have uniform row and column weights, which make them able to greatly simplify the implementation of LDPC coders. Their large girth improves their decoding performance. Simulation results compare their bit error rate (BER) performance over additive white Gaussian noise (AWGN) channels with randomly constructed LDPC codes. When concatenated with error-correcting codes such as Reed-Solomon codes, LDPC codes with  $j = 2$  are promising for data storage and other applications.

## 1. INTRODUCTION

LDPC codes, i.e., Gallager codes [1], can achieve near-optimum performance in AWGN channels [2] when iteratively decoded by the sum-product algorithm [3]. Since short cycles in a Tanner graph prevent the sum-product algorithm from converging, the girth of the graph is an important factor to consider in the design of LDPC codes.

Gallager proved in his original work, [1], that the typical  $(n, j, k)$  LDPC codes with column weight  $j \geq 3$  have a minimum distance that grows linearly with block length  $n$  for given  $j$  and row weight  $k$ , and that the minimum distance of an  $(n, j, k)$  LDPC code with  $j = 2$  grows logarithmically with  $n$ . However, when compared with  $j \geq 3$ , LDPC codes with column weight  $j = 2$  have several advantages: their encoders and decoders are simpler to implement since they have lower computation complexity and storage complexity; and they have better block error statistics properties as pointed out recently by Song, Liu, and Kumar [4]. When concatenating them with error-correcting codes such as Reed-Solomon codes, these properties make LDPC codes with  $j = 2$  promising for data storage and other applications.

---

Research supported by the Data Storage Systems Center (DSSC) at Carnegie Mellon University.

In reference [4], Song, Liu, and Kumar have constructed 4-cycle-free and 6-cycle-free  $(n, 2, k)$  LDPC codes based on Disjoint Difference Sets (DDS). Kou, Lin, and Fossorier [5] construct four classes of (quasi-)cyclic LDPC codes with girth 6 based on finite geometries of lines and points of Euclidean and projective geometries over finite fields. References [6, 7] present LDPC codes with column weight 3 based on balanced incomplete block designs (BIBD) with girth 6, i.e., with no cycles of length 4. Here we construct codes with girth up to 20, much larger than all these codes described in the literature.

This paper introduces three new classes of structured regular  $(n, 2, k)$  LDPC codes with girth 12, 16, and 20, respectively. These methods serve as a basis for constructing codes of large girth and  $j \geq 3$ .

**Preliminaries** We will use the expression “ $K$ -cycle-free” code to refer to LDPC codes with no cycles of length  $L \leq K$ , i.e., to codes whose girth is  $K + 2$ . In contrast, “free of  $K$ -cycles” codes will describe an LDPC code with no cycles of length  $K$ . Note that “free of  $K$ -cycles” codes may have cycles of length shorter than  $K$ .

Let  $H$  be the parity check matrix of an LDPC code with  $v$  parity check equations, i.e.,  $H$  is  $v \times n$ . We represent these parity check equations by a set  $X$  of  $v$  points. We call  $X$  the *point set* of the LDPC  $H$  matrix. For LDPC codes with column weight  $j = 2$ , each column of the  $H$  matrix is represented by an edge between two points in the set  $X$  that correspond to the two nonzero elements in this column. We call the resulting graph the *structure graph* for the LDPC  $H$  matrix and refer to the graph by  $\mathcal{G}_H$  or simply  $\mathcal{G}$ . Figure 1 shows a 6-cycle in both the  $H$  matrix of an  $(n, 2, k)$  LDPC code and its structure graph.

The structure graph helps to identify cycles in the LDPC  $H$  matrices. Two distinct edges between two nodes in a structure graph stand for a 4-cycle. A 6-cycle is a triangle comprising three points and three edges between any two points. An 8-cycle is a loop composed of four points and four tail-biting edges. The pattern is similar for cycles with length greater than 8. Figure 2 shows examples of cycles

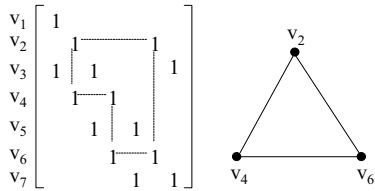


Fig. 1. A 6-cycle in an  $H$  matrix and its structure graph.

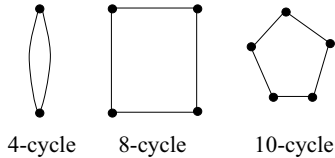


Fig. 2. Examples of cycles with length 4, 8 and 10.

with length 4, 8, and 10, where the positions of the points are flexible, i.e., changing their positions will not affect the relationship among them.

In sections 2, 3, and 4, we describe the construction of  $(n, j, k)$  LDPC codes with girth 12, 16, and 20, respectively. Simulation results and analysis are shown in section 5. Finally, section 6 concludes the paper.

## 2. $(n, 2, k)$ LDPC CODES WITH GIRTH 12

The basic idea in our design of  $(n, 2, k)$  LDPC codes with girth 12 is to partition the points in the point set  $X$  of a structure graph into two subsets such that no edge connects points in the same subset, and only edges between points in different subsets remain. To automate the construction, we introduce constraints that guide the selection of edges so that the resulting LDPC codes have girth 12.

**Fact 1** Let  $v = 2p$  be even and partition the point set  $X$  of  $H$  into two subsets  $X_1$  and  $X_2$  of equal size. If edges connect only points in different subsets then the corresponding  $(n, 2, k)$  LDPC codes are free of  $(4m + 2)$ -cycles, for  $m = 1, 2, \dots$

*Proof:* A  $(4m + 2)$ -cycle can be represented by a loop comprising of  $(2m+1)$  points and  $(2m+1)$  tail-biting edges. For any specific  $m > 1$ , assume that there exists a  $(4m + 2)$ -cycle in an LDPC code constructed by Fact 1, and let  $A$  be a component point in this  $(4m + 2)$ -cycle with neighboring points  $B$  and  $C$  in the loop, respectively, as shown in figure 3. Without loss of generality, assume point  $A$  to be in subset  $X_1$ . There are two paths between  $A$  and  $C$ : The

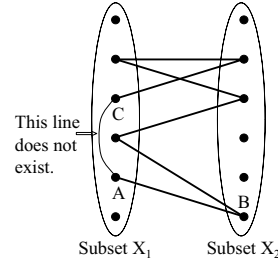


Fig. 3. Absence of  $(4m + 2)$ -cycles.

shorter path with length 1 and the longer path with length  $2m$ . Along the longer path,  $C$  must be in the same subset  $X_1$  where  $A$  is since the longer path has an even number of edges and each edge is between two points in different subsets. We know that there is no edge between two points in the same subset, hence the shorter path between  $A$  and  $C$  cannot exist. This contradicts the assumption, so  $(4m + 2)$ -cycles do not exist.

LDPC codes constructed by Fact 1 have no cycles with length 6 and 10, and they are also free of 4-cycles, since there is at most one edge between any two points. By adding additional constraints on the selection of edges to avoid the existence of 8-cycles, we obtain LDPC codes with girth 12.

With reference to figure 4, we consider the point set  $X = \{a_1, \dots, a_8, b_1, \dots, b_8\}$  of an LDPC  $H$  matrix and divide it into subsets  $X_1 = \{a_1, \dots, a_8\}$  and  $X_2 = \{b_1, \dots, b_8\}$ . The points in each subset are aligned in a vertical line. The reader should note that figure 4 repeats twice the set  $X$  (left and right of the figure). The points in subsets  $X_1$  and  $X_2$  are sequentially referred to as  $a_i$  and  $b_j$ , respectively,  $1 \leq i, j \leq v/2 = p$ . By Fact 1, each edge is between two points  $a_i \in X_1$  and  $b_j \in X_2$ . We introduce next concepts needed in our construction: slope of an edge, scissors, scissors width, base, admissible slope pair and admissible slope set.

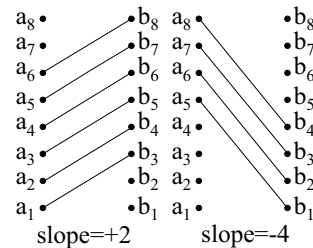


Fig. 4. Edges with specific slopes.

**Definition 1 (slope)** The “slope”  $s$  of an edge between points  $a_i \in X_1$  and  $b_j \in X_2$  is defined as  $s = j - i$ .

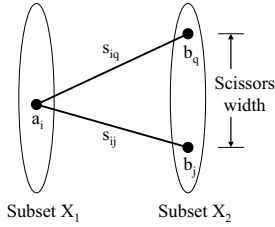


Fig. 5. Scissors structure

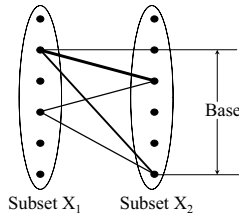


Fig. 6. The double scissors structure.

Slopes take values in the range  $-(p-1) \leq s \leq (p-1)$ , where  $p$  is the number of points in each subset. Hence, there are  $(2p-1)$  possible slopes. Note that we have assigned the points in subset  $X_1$  as reference points when calculating the slopes. The number of possible edges with slope  $s$  is  $(p-|s|)$ . Figure 4 gives examples of edges with slope  $+2$  and  $-4$  when  $p=8$ .

**Definition 2 (scissors)** Two edges with a common end point in subset  $X_1$  compose a “scissors.”

Figure 5 illustrates the scissors structure. It is composed of one point ( $a_i$ ) in  $X_1$  and two points ( $b_j$  and  $b_q$ ) in  $X_2$ .

**Definition 3 (scissors width)** The “scissors width”  $w$  is the distance between  $b_j$  and  $b_q$ , i.e.,  $w = |j - q|$ .

Let the slope of the edge between  $a_i$  and  $b_j$  be  $s_{ij}$ , and the slope of the edge between  $a_i$  and  $b_q$  be  $s_{iq}$ . Then the scissors width is also equal to  $|s_{ij} - s_{iq}|$ , as shown in figure 5.

In a graph, 8-cycles are composed of 4 points and 4 tail-biting edges. Since there is no edge between any two points in the same subset, 8-cycles are only of a single structural form: two scissors with a common width, which is called “double scissors structure,” as shown in figure 6.

**Definition 4 (base)** The “base” of a double scissors structure is defined as the vertical distance between the highest point and the lowest point among these four component points.

Figure 6 illustrates graphically the concept of base. It follows immediately from its definition that the base of a double scissors structure must be smaller than  $p = v/2$ .

**Definition 5 (admissible slope pair)** A slope pair  $(s_i, s_j)$  is an “admissible slope pair” iff

$$s_i = -\text{sgn}(s_j) \cdot (p - |s_j|).$$

**Definition 6 (admissible slope set)** A set  $\mathcal{A}$  of slopes is an “admissible slope set” iff all the slopes in  $\mathcal{A}$  can be partitioned into admissible slope pairs.

Each admissible slope pair  $(s_i, s_j)$  corresponds to  $(p - |s_i|) + (p - |s_j|) = p$  edges, and increases the degree of each point in  $X$  by 1. This means that each admissible slope pair corresponds to  $p$  columns of the  $H$  matrix, and increases the weight of each row by 1. Note that the slope 0 itself is equivalent to an admissible slope pair. This way, we are able to guarantee the consistence of the row weights, as well as the column weights.

**Construction 1 ( $(n, 2, k)$  LDPC codes with girth 12)** To construct regular  $(n, 2, k)$  LDPC codes with girth 12 choose only edges with slopes from an admissible slope set  $\mathcal{A}$ . The slopes in  $\mathcal{A}$  are carefully selected to avoid the existence of 8-cycles, i.e., when selecting arbitrarily two different pairs of slopes,  $(s_1, s_2)$  and  $(s_3, s_4)$  from  $\mathcal{A}$ , one of the following two conditions must be satisfied:

- (1)  $|s_1 - s_2| \neq |s_3 - s_4|$ ;
- (2) If  $|s_1 - s_2| = |s_3 - s_4|$ , i.e., if they form a double scissors structure, then the base of the double scissors structure must be greater than or equal to  $p$ .

Hence, if we find a desired admissible slope set  $\mathcal{A}$ , we obtain an LDPC code with girth 12.

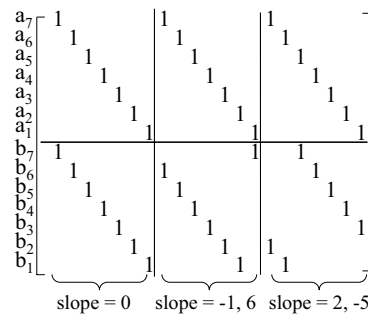


Fig. 7. The  $H$  matrix of an  $(n, 2, k)$  LDPC code with girth 12 when  $v=14$ .

$v$	150	210	260	350	430	520	610	680	830	1000	1090
$n$	600	945	1300	1925	2580	3380	4270	5100	6640	8500	9810
$r$	3/4	7/9	4/5	9/11	5/6	11/13	6/7	13/15	7/8	15/17	8/9

Fig. 8.  $(n, 2, k)$  LDPC codes with girth 12: code rates and block lengths under different  $v$ .

We provide a simple example. For  $v = 14$ , we can obtain  $\mathcal{A} = \{0, -1, 6, 2, -5\}$ . The corresponding  $H$  matrix is shown in figure 7 where the points  $a_i, 1 \leq i \leq 7$ , represent the first 7 rows in the  $H$  matrix, and the points  $b_j, 1 \leq j \leq 7$ , represent the last 7 rows. The  $H$  matrix is well structured and is completely determined by  $v$  and  $\mathcal{A}$ . Figure 8 gives the available code rates and block lengths under different values of  $v$ , where  $v, n$ , and  $r$  are the number of parity check equations, the block length, and the code rate, respectively.

### 3. $(n, 2, k)$ LDPC CODES WITH GIRTH 16

This section describes briefly the construction of a class of  $(n, 2, k)$  LDPC codes with girth 16 and code rate  $1/2$ . Please refer to [8] for details.

Assume the number of parity check equations  $v = 8p$ . We divide these points into 8 subsets of equal size, and the points in each subset are aligned in a vertical line. These subsets,  $X_0, X_1, \dots, X_7$ , comprise a loop, and each point within subset  $X_i$  can only connect to points in the previous or next subset, i.e.,  $X_{mod(i-1,8)}$  or  $X_{mod(i+1,8)}$ , and cannot connect to points in the same subset  $X_i$ . The overall structure graph looks like a cylinder, and we call it a *cylinder structure*. Figure 9 gives an example of the cylinder structure.

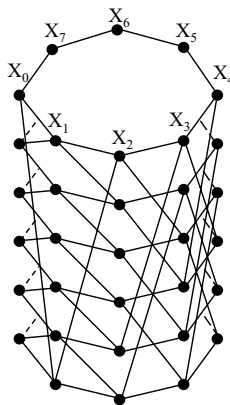


Fig. 9. A cylinder structure.

There are two types of cycles in a cylinder structure.

Type I includes the cycles that pass all the 8 subsets around the cylinder structure. The minimal length of this type of cycles is 16, so they are of no interest to us. Type II includes cycles with points in some of the 8 subsets, but these subsets must be consecutive. In a cylinder structure, since each point in a subset can only connect to points in the next subset or the previous subset, any cycles of type II must be composed of an even number of  $2m$  of edges, which means that the length of the cycles must be  $4m$ , where  $m$  is any integer greater than 1. Therefore, we need not consider  $4m+2$ -cycles. Figure 10 gives an example of type II cycles. If we can avoid the existence of 8-cycles and 12-cycles, we get LDPC codes with girth 16.

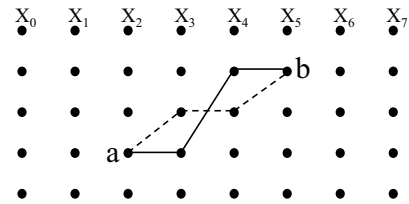


Fig. 10. Type II cycle in a cylinder structure.

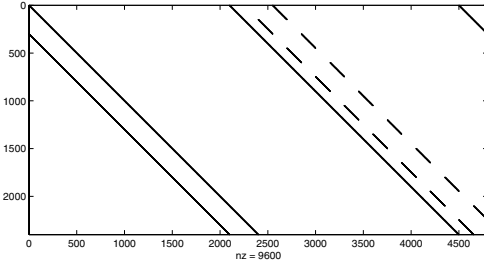
Section  $S_i$  represents all edges between two neighboring subsets  $X_i$  and  $X_{mod(i+1,8)}$ , and corresponds to an admissible slope set  $\mathcal{A}_i$ . Each edge in section  $S_i$  must have a slope in  $\mathcal{A}_i$ . To avoid the existence of 8-cycles and 12-cycles, we can find only two admissible slope pairs, i.e., 4 slopes, for each admissible slope set  $\mathcal{A}_i$ . In this case, the degree of each point in the structure graph is 4, i.e., the row weight of the  $H$  matrix is 4, so the code rate is  $1/2$ .

We develop a “Viterbi-like” algorithm to find out the 8 admissible slope sets. Please refer to [8] for details of the algorithm. With a different value of  $v$ , we can obtain a class of LDPC codes with girth 16. To make sure there is a solution,  $v$  must be large enough. Simulations point out that the minimal value of  $v$  is 160, i.e.,  $p = 20$ . Each solution corresponds to an  $H$  matrix of LDPC codes with girth 16. These  $H$  matrices are well structured and completely determined by  $p$  and the admissible slope set  $\mathcal{A}_0, \dots, \mathcal{A}_7$ . Figure 11 gives the  $H$  matrix when  $v = 2400, n = 4800$ , and  $r = 1/2$ .

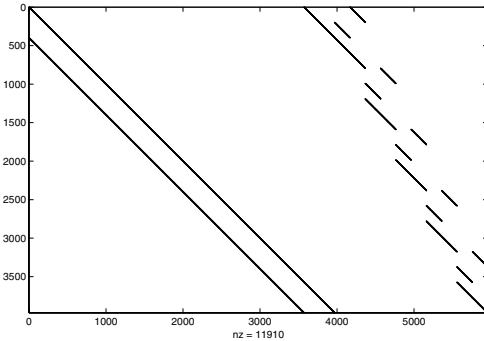
### 4. $(n, 2, k)$ LDPC CODES WITH GIRTH 20

By a modified version of the previous method, we can obtain  $(n, 2, k)$  LDPC codes with girth 20 and code rate  $1/3$ .

Assume  $v = 10p$ , and divide the  $v$  points into 10 subsets of equal size. Then we can establish a cylinder structure with 10 sections. Now our task is to find 10 admissible slope sets, i.e.,  $\mathcal{A}_0, \dots, \mathcal{A}_9$ , to avoid cycles with length smaller



**Fig. 11.** The  $H$  matrix for an  $(n, 2, k)$  LDPC code with girth 16 when  $n = 4800$  and  $r = 1/2$ .



**Fig. 12.** The  $H$  matrix for an  $(n, 2, k)$  LDPC codes with girth 20 when  $n = 5955$  and  $r = 1/3$ .

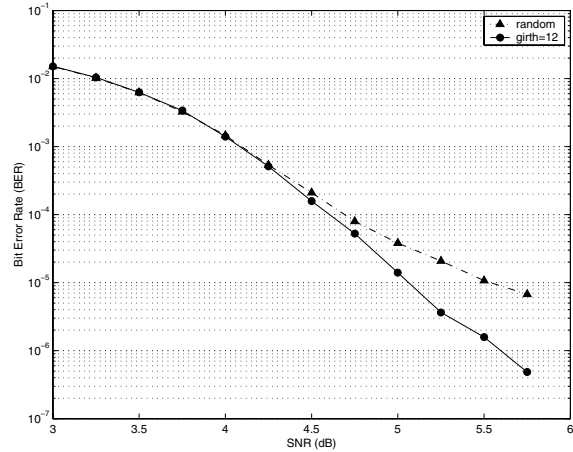
than 20. The admissible slope sets with even index include two admissible slope pairs, and the admissible slope sets with odd index include only one admissible slope pair. In this case, the degree of each point is 3, i.e., the row weight of  $H$  matrix is 3, so the code rate is  $1/3$ .

We use the same “Viterbi-like” algorithm to find out the 10 admissible slope sets. The number of points  $v$  should also be large enough. Simulations point out that the minimal value of  $v$  is 130, i.e.,  $p = 13$ . Figure 12 gives the  $H$  matrix when  $v = 3970$ ,  $n = 5955$ , and  $r = 1/3$ .

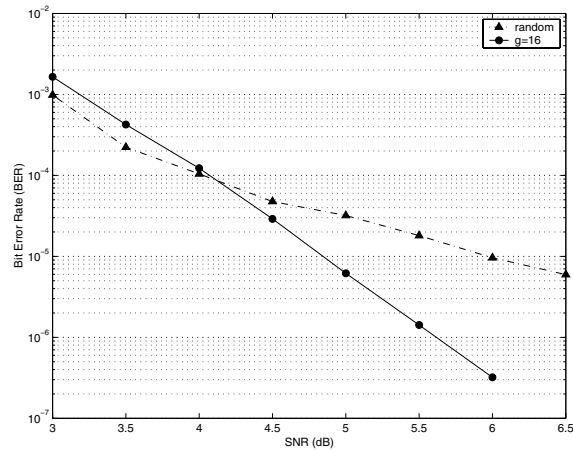
## 5. SIMULATION RESULTS

In simulations, we employed sum-product decoding, as presented in [3], over AWGN channels, and we adopted the rate-adjusted signal to noise ratio as shown in [9].

Figure 13 compares the bit error rate (BER) performance of an LDPC code with girth 12 with that of a randomly constructed  $(n, 2, k)$  code. Both codes have block length 4270 and code rate  $6/7$ . The two codes have similar performance in low SNR region. In high SNR region, the codes with girth 12 outperforms the randomly constructed code by 0.5dB at  $\text{BER} = 10^{-5}$ .



**Fig. 13.** BER performance comparison between LDPC codes with girth 12 and randomly constructed LDPC codes.

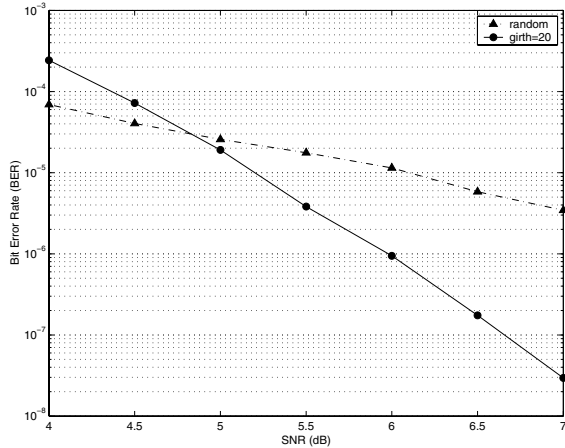


**Fig. 14.** BER performance comparison between LDPC codes with girth 16 and randomly constructed LDPC codes.

Figure 14 compares the BER performance of an LDPC code with girth 16 with that of a randomly constructed  $(n, 2, k)$  code. Both codes have block length 4368 and code rate  $1/2$ . The randomly constructed code has better performance in the low SNR region. In the high SNR region, the code with girth 16 outperforms the randomly constructed code by 1.1dB at  $\text{BER} = 10^{-5}$ .

Figure 15 compares the BER performance of an LDPC code with girth 20 with that of a randomly constructed  $(n, 2, k)$  code. Both codes have block length 4395 and code rate  $1/3$ . The randomly constructed code has better performance in the low SNR region. In the high SNR region, the code with girth 20 outperforms the randomly constructed code by 0.8dB at  $\text{BER} = 10^{-5}$ .

For LDPC codes with column weight 2, the minimum



**Fig. 15.** BER performance comparison between LDPC codes with girth 20 and randomly constructed LDPC codes.

distance  $d_{min} = g/2$ , where  $g$  is the girth. In the high SNR region,  $d_{min}$  is the dominant factor for BER performance. Therefore, LDPC codes with large girth outperform randomly constructed codes. For low SNR, the performance depends not only on the girth but also on the cycle length distribution. We speculate that, in figure 14, the randomly constructed code does have cycles with length lower than 16, but possibly the number of cycles of length 16 is smaller than the number of the cycles of length 16 of the girth 16 code that we designed. This may overwhelm the benefit it gains from its large girth. This is an interesting phenomenon that deserves further study and shows that when designing a code careful consideration should be given to the SNR region where they are to operate.

## 6. CONCLUSION

In this paper, we present three classes of regular  $(n, 2, k)$  LDPC codes with girth 12, 16, and 20. These codes are systematically constructed, and their  $H$  matrices are well structured and completely represented by a set of integers (parity check numbers and slopes). In the high SNR region, the codes with large girth have better BER decoding performance than that the performance of randomly constructed codes over AWGN channels.

## 7. REFERENCES

[1] R. G. Gallager, *Low-Density Parity Check Codes*, MIT Press, Cambridge, MA, 1963.  
 [2] D. J. C. Mackay and R. M. Neal, "Good codes based on very sparse matrices," in *Cryptography and Coding, 5th*

*IMA Conference, in Lecture Notes in Computer Science*, C. Boyd, Ed., 1995, vol. 1025, pp. 110–111.

[3] 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, Feb 2001.  
 [4] H. Song, J. Liu, and B. V. K. Vijaya Kumar, "Low complexity ldpc codes for magnetic recording," in *IEEE Globecom 2002*, Taipei, Taiwan, R.O.C., Nov 2002.  
 [5] Yu Kou, Shu Lin, and Marc P. C. Fossorier, "Low-density parity-check codes based on finite geometries: a rediscovery and new results," *IEEE Tans. Inform. Theory*, vol. 47, no. 7, pp. 2711–2736, Nov 2001.  
 [6] S. J. Johnson and S. R. Weller, "Regular low-density parity-check codes from combinatorial design," in *Inf. Tech. Workshop (ITW) 2001*, Cairns, Australia, Sep 2001.  
 [7] B. Vasic, "Structured iteratively decodable codes based on Steiner systems and their application in magnetic recording," in *IEEE Globecom 2001*, San Antonio, TX, Nov 2001.  
 [8] H. Zhang and José M. F. Moura, "Large-girth ldpc codes based on graphical models," in *IEEE SPAWC 2003*, Rome, Italy, Jun 2003.  
 [9] D. J. C. Mackay, "Good error-correcting codes based on very sparse matrices," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 399–431, March 1999.