New low-density parity-check codes with large girth based on hypergraphs

Bocharova, Irina; Hug, Florian; Johannesson, Rolf; Kudryashov, Boris; Satyukov, Roman

Published in: [Host publication title missing]

DOI: 10.1109/ISIT.2010.5513639

2010

Citation for published version (APA):
https://doi.org/10.1109/ISIT.2010.5513639

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
New Low-Density Parity-Check Codes with Large Girth Based on Hypergraphs

Irina E. Bocharova¹, Florian Hug², Rolf Johannesson², Boris D. Kudryashov¹, and Roman V. Satyukov¹

¹ Dept. of Information Systems
St. Petersburg Univ. of Information Technologies,
Mechanics and Optics
St. Petersburg 197101, Russia
Email: {irina, boris}@eit.lth.se
Email: satyukov@gmail.com

² Dept. of Electrical and Information Technology,
Lund University
P. O. Box 118, SE-22100 Lund, Sweden
Email: {florian, rolf}@eit.lth.se

Abstract—The relation between low-density parity-check (LDPC) codes and hypergraphs supports searching for powerful LDPC codes based on hypergraphs. On the other hand, coding theory methods can be used in searching for hypergraphs with large girth. Moreover, compact representations of hypergraphs based on convolutional codes can be found. Algorithms for iteratively constructing LDPC codes with large girth and for determining their minimum distance are introduced. New quasi-cyclic (QC) LDPC codes are presented, some having both optimal girth and optimal minimum distance.

I. INTRODUCTION

An important consequence of the relation between low-density parity-check (LDPC) codes and hypergraphs (see, for example [1]–[3]) is the new possibilities in searching for powerful LDPC codes. Moreover, coding theory methods can be used in searching for hypergraphs better than previously known. In particular, compact representations of hypergraphs based on convolutional codes can be found.

Typically, LDPC codes have a minimum distance which is less than those for the best known linear codes, but due to their structure they are suitable for low-complexity iterative decoding. In general, however, belief propagation [4] works better if the minimum distance of an LDPC code is large.

A parity-check matrix of a quasi-cyclic (QC) LDPC code can be regarded as the incidence matrix of a regular hypergraph. Although QC LDPC codes are not asymptotically optimal, they can outperform pseudo-random (asymptotically optimal) LDPC codes of short or moderate lengths [5]. This motivated our search for good short QC LDPC codes.

The problem of finding QC LDPC codes with large girth were considered in [6]–[10]. Best known examples are a (155, 64, 20) code with girth 8, a (305, 124) code with girth 10, and a (905, 364) code with girth 12; the latter two codes with previously unknown minimum distance. In [10], rate $R = 1/2$ codes of lengths 970 and 2534 with girth 10 and 12, respectively, were presented.

We present new examples of regular QC LDPC codes in the form of tailbiting (TB) LDPC codes with girth 8, 10, and 12; some of them having optimal girth and/or optimal minimum distance. This representation is compact and we can apply low-complexity encoding, as well as efficient searching and decoding procedures developed for convolutional and TB codes [11]. Moreover, the girth and free distance of the parent convolutional code upper-bound the girth and the minimum distance of the corresponding TB LDPC code [12].

Relations between the girth of the basic Tanner graph and the hypergraph corresponding to the parent LDPC convolutional code are derived. We introduce new algorithms for iteratively constructing LDPC codes with large girth and for determining their corresponding minimum distance. Examples of newly found QC LDPC codes and best known examples are tabulated together with their girth and minimum distance.

II. PARITY-CHECK MATRICES

A rate $R = b/c$ LDPC convolutional code $C$ is determined by its parity-check matrix of memory $m$

$$H(D) = \begin{pmatrix} h_{11}(D) & h_{12}(D) & \cdots & h_{1c}(D) \\ h_{21}(D) & h_{22}(D) & \cdots & h_{2c}(D) \\ \vdots & \vdots & \ddots & \vdots \\ h_{(c-b)1}(D) & h_{(c-b)2} & \cdots & h_{(c-b)c}(D) \end{pmatrix}$$

(1)

where the parity-check polynomials $h_{ij}(D) = D^{w_{ij}}$ are monomials of degree $w_{ij}$. If each column and each row contain exactly $J$ and $K$ nonzero elements, respectively, we call $C$ a regular $(J, K)$ LDPC convolutional code. Denoting the degree of 0 by $-\infty$, such a parity-check matrix can be represented by its degree matrix

$$W = \{w_{ij}\}$$

with $i = 1, 2, \ldots, c - b$ and $j = 1, 2, \ldots, c$ [13]. Expressing the $(c-b) \times c$ parity-check matrix $H(D)$ in terms of its binary matrices $H_i$, $i = 0, 1, \ldots, m$, that is,

$$H(D) = H_0 + H_1 D + H_2 D^2 + \cdots + H_m D^m$$

we obtain the binary semi-infinite parity-check matrix $H$, which can be written as

$$H^T = \begin{pmatrix} H_0^T & H_1^T & \cdots & H_m^T \\ H_1^T & H_2^T & \cdots & H_m^T \\ H_2^T & H_3^T & \cdots & H_m^T \\ \vdots & \vdots & \ddots & \vdots \\ H_{c-b}^T & H_{c-b+1}^T & \cdots & H_m^T \end{pmatrix}$$

(2)

where $H^T$ denotes the transpose of $H$.

By TB the parent convolutional parity-check matrix (2) to length $M > m$, we obtain the following $M(c-b) \times Mc$ parity-
check matrix $H_{TB}$ of the corresponding tailbitten linear binary block code $B$ of block-length $Mc$ as

$$
H_{TB}^T = \begin{pmatrix}
H_0^T & H_1^T & \cdots & H_{m-1}^T & 0 \\
0 & H_0^T & H_1^T & \cdots & H_{m-1}^T \\
H_1^T & 0 & H_0^T & \cdots & H_{m-1}^T \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
H_{m-1}^T & H_{m-2}^T & \cdots & H_0^T & 0
\end{pmatrix}.
$$

Note that $H_{TB}$ is $(J, K)$ regular, that is, there are exactly $J$ ones in every column and exactly $K$ ones in every row. With $J$ and $K$ being much smaller than $M$, $H_{TB}$ is considered to be sparse. Furthermore, the first $c$ columns of $H_{TB}$ are repeated in a cyclically shifted manner throughout the whole matrix.

III. GRAPHS AND HYPERGRAPHS

Every parity-check matrix of an LDPC code can be interpreted as an incidence matrix of a graph $G$ or hypergraph $H_G$. A hypergraph is a generalization of a graph and is determined by a set of vertices $V = \{v_i\}$ and a set of hyperedges $E = \{e_i\}$, where each hyperedge is a subset of vertices and may connect (contain) any number of vertices. If each hyperedge connects not more than two vertices it is called an edge and we obtain an ordinary graph.

A hypergraph is called $s$-uniform if every hyperedge has cardinality $s$, that is, it connects $s$ vertices. For $s = 2$, a hypergraph is a simple graph. The degree of a vertex in a hypergraph is the number of hyperedges that are connected to (contain) it. If all vertices have the same degree $c$, then the hypergraph is $c$-regular, that is, $c$ is the degree of the hypergraph.

Let the set of vertices $V$ of an $s$-uniform hypergraph be partitioned into $t$ disjoint subsets $V_k$, $k = 1, 2, \ldots, t$. If no hyperedge connects (contains) two vertices from the same set $V_k$, $k = 1, 2, \ldots, t$, the hypergraph is said to be $t$-partite.

A path of length $L$ in a hypergraph is an alternating sequence of $L + 1$ vertices $v_i$, $i = 1, 2, \ldots, L$, and $L$ hyperedges $e_i$, $i = 1, 2, \ldots, L$, with $e_i \neq e_{i+1}$. If the first and the final vertex coincide, that is, $v_1 = v_L$, we obtain a cycle. A cycle is called simple if all its vertices and edges are distinct, except the first and final vertex which coincide. A simple cycle is also known as a Berge cycle [14]. Finally, the girth of a hypergraph is the length of its shortest simple cycle. For graphs it has been shown in [15] that their girth coincides with the minimum distance of the corresponding block code.

Example 1: Consider the rate $R = 1/4$ convolutional code $C$ with parity-check matrix

$$
H(D) = \begin{pmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & D & D \\
1 & D & 1 & D
\end{pmatrix}
$$

and degree matrix

$$
W = \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 \\
0 & 1 & 0 & 1
\end{pmatrix}.
$$

Fig. 1. A 3-partite, 3-uniform, 4-regular hypergraph $H_G$ with vertices $v_i$, $i = 1, 2, \ldots, 6$, and hyperedges $e_j$, $j = 1, 2, \ldots, 8$.

Tailbitting (3) via its dual generator matrix [11] to length $M = 2$, we obtain the tailbitten $6 \times 8$ parity-check matrix

$$
H_{TB} = \begin{pmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 1 & 0
\end{pmatrix}.
$$

Interpreting (4) as an incidence matrix, we obtain the 3-partite, 3-uniform, 4-regular hypergraph $H_G$ as illustrated in Fig. 1 with 6 vertices and 8 hyperedges. Every row of the incidence matrix corresponds to a vertex of the hypergraph, while the columns are represented by hyperedges (subsets of vertices).

It is easy to see that the girth of this hypergraph is $g = 2$.

IV. TANNER, VOLTAGE, AND BASIC GRAPHS

In order to construct hypergraphs with large girth we need to introduce some auxiliary graph representations. The Tanner graph $\bar{G}_T$ [16] of a regular $(J, K)$ convolutional parity-check matrix $H(D)$ (1) is determined by the $(2c - b) \times Jc$ incidence matrix

$$
H_T = \begin{pmatrix}
C_1 & C_2 & \cdots & C_c \\
J_1 & J_2 & \cdots & J_c
\end{pmatrix}
$$

where each column of the $(c - b) \times J$ submatrix $C_i$ contains not more than one of the $J$ nonzero elements of column $i$ of $H(D)$, $i = 1, 2, \ldots, c$. The elements of the $c \times J$ matrix $J_i$ are all zero except for the elements of the $i$th row, which are equal to one. In other words, every hyperedge is replaced by an additional vertex and $J$ new edges between the newly introduced vertex and each of the original vertices connected to the hyperedge.

In order to represent the monomials in the incidence matrix of the Tanner graph (5), every edge is labeled by an edge voltage, that is, the degree difference of the corresponding monomials. Note, the sign of the edge voltage depends on
the passing direction. The edge voltage $\mu_{uv}$ from vertex $u$ to vertex $v; u, v \in \{1, 2, \ldots, 2c - b\}; u < v$ is given by

$$\mu_{uv} = w_{vk} - w_{uk}$$

$$\mu_{uv} = -\mu_{vu}$$

(6)

where $w_{ij}$ is the degree of the $i$th row and $j$th column entry of the degree matrix $W$ corresponding to the incidence matrix $H_T$ (5) with entries $h_{ij}$ and $k$ is chosen such that both $h_{ve}$ and $h_{uk}$ $\neq 0$. (Note, that in general $k$ is not necessarily unique.) Hereinafter we will refer to a Tanner graph with its edges labeled according to (6) as a voltage Tanner graph $G_{VT}$.

We conclude that the girth of the corresponding parent convolutional code $C$, determined by its parity-check matrix $H(D)$ (3), follows as $g_{free} = 0.5g_{VT}$. However, if we neglect all labels, we would obtain the corresponding basic Tanner graph $G_{BT}$ with girth $g_{BT} \leq g_{VT}$.

V. BOUNDS ON Girth AND MINIMUM DISTANCE

**Theorem 1**: The minimum distance $d_{\min}$ and the girth $g$ of an $(n, k, d_{\min})$ QC LDPC block code $B$ obtained from a rate $R = b/c$ convolutional code $C$ with free distance $d_{\text{free}}$ and girth $g_{\text{free}}$ by TB to length $M$ are upper-bounded by the inequalities

$$d_{\min} \leq d_{\text{free}}$$

$$g \leq g_{\text{free}}$$

In [13] a lower bound on the girth of a voltage Tanner graph $g_{VT}$ was found via the girth of corresponding basic Tanner graph $g_{BT}$ for ordinary graphs. It is straightforward to generalize this bound:

**Theorem 2**: Consider a basic Tanner graph of a regular ($J \geq 3, K$) QC LDPC convolutional code with girth $g_{BT}$ and let $d_s$ denote the $s$th generalized minimum Hamming distance, that is, the number of nontrivial (not identically zero) positions of an $s$-dimensional linear subcode. Then there exist a TB length $M$ and a set of edge labels, such that the girth $g_T$ of the Tanner graph for the corresponding TB block code of length $N = Mc$ satisfies the inequality

$$g_T \geq 2 \max \{g_{BT} \left[ g_{BT}/2 \right], d_s \}$$

where $d_2$ is the second generalized minimum Hamming distance of the linear $(JMc, M((J - 2)c + b))$ block code determined by the Tanner graph.

Finally, we want to recall the following upper bounds on the achievable girth and minimum distance.

**Theorem 3 ([5], [12], [18])**: Let $H(D)$ be the parity-check matrix of a rate $R = b/c$ convolutional code with all its entries being nonzero monomials and free distance $d_{\text{free}}$. By TB to length $M$ we obtain a QC LDPC block code of block length $Mc$ and minimum distance $d_{\min}$, together with its Tanner graph representation with girth $g_T$, which satisfies the following inequalities:

$$g_T \leq 12$$

$$d_{\min} \leq d_{\text{free}} \leq (c - b + 1)!.$$ 

VI. SEARCH FOR CODES WITH LARGE Girth

Every QC LDPC rate $R = Mb/Mc$ block code $B$ can be obtained from a rate $R = b/c$ parent convolutional code $C$ using a TB length $M$. Limiting the parity-check matrix $H(D)$ of the parent convolutional code $C$ to only nonzero monomial entries, we can represent it by its basic Tanner graph $G_{BT}$. Then we can use the algorithm, as presented in [13], to find example, vertex 2 to vertex 6 is labeled according to $\mu_{26} = -\mu_{62} = w_{68} - w_{28} = -1$.

We conclude that the girth of the corresponding parent convolutional code $C$, determined by its parity-check matrix $H(D)$ (3), follows as $g_{free} = 0.5g_{VT}$. However, if we neglect all labels, we would obtain the corresponding basic Tanner graph $G_{BT}$ with girth $g_{BT} \leq g_{VT}$. 

V. BOUNDS ON Girth AND MINIMUM DISTANCE

**Theorem 1**: The minimum distance $d_{\min}$ and the girth $g$ of an $(n, k, d_{\min})$ QC LDPC block code $B$ obtained from a rate $R = b/c$ convolutional code $C$ with free distance $d_{\text{free}}$ and girth $g_{\text{free}}$ by TB to length $M$ are upper-bounded by the inequalities

$$d_{\min} \leq d_{\text{free}}$$

$$g \leq g_{\text{free}}.$$
permutation matrices as in [5] gives the freedom to choose it allows us to generalize this construction to parity-check matrices of convolutional codes with zero entries.

For the sake of completeness, this algorithm is given below.

1) Choose a desired girth \( g \), the rate of the parent convolutional code \( R = b/c \), and a TB length \( M \).
2) Construct a list of \( n \) equations describing all cycles of lengths less than \( 2g \) in the basic Tanner graph \( G_{BT} \) corresponding to the chosen convolutional code rate.
3) Represent these \( n \) cycles by an \( n \times Jc \) matrix, where each row \( a = (a_1, a_2, \ldots, a_{Jc}) \) corresponds to one cycle with the value \( a_i \) denoting the difference between the number of passes of the \( i \)-th edge in forward and backward directions.
4) Search randomly for a vector \( \mu = (\mu_1, \mu_2, \ldots, \mu_{Jc}) \) such that \( \mu A^T \neq 0 \mod M \).
5) Label the edges of the basic Tanner graph according to \( \mu \) to obtain a voltage Tanner graph. Tailbite the parity-check matrix of the corresponding parent convolutional code via its dual generator matrix to length \( M \); this yields the parity-check matrix of a block code of block length \( N = Mc \) corresponding to a hypergraph with girth not less than \( g \).

For each code, the degree matrix \( W \) of the parent convolutional code is given together with the TB length \( M \) needed to construct the corresponding \( (Mc, Mb) \) block code of block length \( N = Mc \) with girth \( g_T \) and minimum distance \( d_{\min} \).

Table I also includes the minimum distance for the best-known solutions of rate \( R \) and minimum distance \( d_{\min} \) which, for the two longer ones (codes \#2 and \#3), were previously unknown.

![Table I](attachment:image.png)

VII. MINIMUM DISTANCE

Clearly, every codeword \( v \) of a tailbitten linear block code \( B \) of block-length \( N = Mc \) with an \( M(c-b) \times Mc \) parity-check matrix \( H_{TB} \) fulfills \( vH_{TB}^T = 0 \).

The minimum distance for a linear block code is equal to the minimal number of columns of \( H_{TB} \) that sum up to zero.

Starting with each of the first \( c \) columns as a root, \( c \) separate trees can be constructed, where each node \( \xi \) at depth \( \ell \) is associated with a state column-vector \( \sigma(\xi) \).

Initially assign column \( h_i \) to the state of the root node \( \sigma(\xi_{root,i}) \) of the \( i \)-th tree, \( i = 1, 2, \ldots, c \). Then build up a tree in such a way, that every branch between any two nodes \( \xi \) and \( \xi' \) is labeled by a column \( h_j, j = 1, 2, \ldots, Mc, j \neq i \), such that \( \sigma(\xi') = \sigma(\xi) + h_j \), where every branch label on the path \( \xi_{root,i} \rightarrow \xi' \) does not occur more than once.

Consider now a certain node \( \xi \) with nonzero state \( \sigma(\xi) \).

Assuming that the \( k \)-th position of \( \sigma(\xi) \) is nonzero, there are at most \( K-1 \) columns which can cancel this nonzero position in \( \sigma(\xi) \) and have not been considered previously. Therefore, every node \( \xi \) has at most \( K-1 \) children nodes per nonzero position.

However, such a tree would grow until all possible linear combinations have been found. Therefore, we limit ourselves to linear combinations of at most \( t \) columns; that is, the maximum depth of the tree is \( t-1 \). Moreover, a node \( \xi \) at depth \( \ell \) will not be extended, if the number of nonzero positions of its state \( \sigma(\xi) \) is larger than \( J(t-\ell-1) \), since at most \( J \) ones can be canceled by each branch.

Remark: Initially reordering the rows of the tailbitten parity-check matrix \( H_{TB} \) such that each block of \( M \) rows contains not more than a single one per column, strengthens the stopping criterion as follows: A node \( \xi \) at depth \( \ell \) will not be extended, if the number of nonzero positions in each block of \( M \) rows in its state \( \sigma(\xi) \) is larger than \( (t-\ell-1) \), as at most one 1 in each block can be canceled by each branch.

VIII. RESULTS

Using the algorithms presented in Sections VI and VII, we have obtained new regular QC LDPC codes. In Table I, a few best-known rate \( R = 2/5 \) QC LDPC codes [6]–[8] are listed together with our newly found codes of rate \( R = 2/5 \) and \( R = 3/6 \) having almost optimum or optimum girth 8, 10 and 12 and optimum minimum distance 24 (except code \#7 with minimum distance 8).

For each code, the degree matrix \( W \) of the parent convolutional code is given together with the TB length \( M \) needed to construct the corresponding \((Mc, Mb)\) block code of block length \( N = Mc \) with girth \( g_T \) and minimum distance \( d_{\min} \).

Table I also includes the minimum distance for the best-known examples of rate \( R = 2/5 \) QC LDPC codes, which, for the two longer ones (codes \#2 and \#3), were previously unknown.

Note, all codes in Table I achieving either the upper bound on the minimum distance \( d_{\min} = 24 \) or on the girth \( g_T = 12 \) according to (8) and (7), respectively, are marked by *.
IX. BIT ERROR PERFORMANCE

Using belief propagation decoding [4] with 60 iterations, the bit error rate performance of all regular QC LDPC codes from Table I is simulated and shown in Figs. 3 and 4. The best previously known rate $R = 2/5$ QC LDPC codes are compared with our newly found ones of the same rate in Fig. 3 and with our newly found codes of rate $R = 3/6$ in Fig. 4.

As expected, larger girth leads in general to better bit error rate performance. However, by comparing codes #1 and #4 or codes #2 and #5 in Fig. 3, we can also conclude that a larger tailbiting length $M$, and thereby a larger resulting block length $N$, yields better performance.

However, comparing codes #1 and #7 in Fig. 4, both of approximately the same block length $N$ and girth $g = 8$ but of rates $R = 2/5$ and $R = 3/6$, respectively, as well as minimum distance $d_{\text{min}} = 20$ and $d_{\text{min}} = 8$, respectively, we conclude that the chosen monomials, and thereby the underlying edge voltages $\mu_{\text{erm}}$ seem to play the most important role in achieving good bit error rate performance. Comparing codes #2 and #8 supports this assumption, even though their corresponding block length differ slightly more.

X. CONCLUSION

Using the relation between hypergraphs and LDPC codes, new searching techniques have been presented. Starting from a hypergraph, any number of LDPC codes of different rates can be obtained by tailbiting the corresponding parent convolutional code via its dual generator matrix to different lengths.

By representing hypergraphs in different ways, lower and upper bounds on the girth as well as on the minimum distance of the corresponding tailbiting block code have been obtained.

Algorithms for finding hypergraphs with optimum or almost optimum girth and for determining their minimum distance have been presented. Their bit error rate performance has been compared using belief propagation decoding, verifying that a larger girth result in an overall better code performance.

ACKNOWLEDGEMENTS

This research was supported in part by the Swedish Research Council under Grant 621-2007-6281.

REFERENCES