Design of High-Rate LDPC Codes Based on Matroid Theory
Guangfu Wu, Yijie Lv, Jiguang He

Abstract—In this letter, sufficient conditions for the determination of the girth are studied from the perspective of matroid theory. The girth of a Tanner graph is at least $2(t_1 + 2)$ if $t_1$ specific conditions are simultaneously met. A novel method of constructing high-rate low-density parity-check (LDPC) codes is proposed based on the matroid theory. The parity-check matrices of the constructed LDPC codes are in the form of $H = [I|H_2]$ with $H_2$ constructed under the conditions of a given girth and a fixed column weight (e.g., $W_c = 4$ or $W_c = 6$). Simulation results verify that the proposed LDPC codes outperform those in the literature over additive white Gaussian noise channels in terms of bit error rate performance.

Index Terms—Girth condition, high rate, matroid theory, parity-check matrix, Bit error rate

I. INTRODUCTION

In the pioneering paper A Mathematical Theory of Communication, Shannon proved the existence of a channel code which ensures reliable communications provided that the information rate for the given code does not exceed the capacity of the channel. Several decades since Shannon’s paper was published, great efforts were devoted to finding the practical capacity-achieving error correction codes. Until the 1990s, the invention of turbo codes and rediscovery of low-density parity-check (LDPC) codes make approaching Shannon’s theoretical performance limits practically feasible.

LDPC codes were initially invented by Gallager in 1960. Tanner further introduced a graphical representation for the LDPC codes, well-known as Tanner graph. LDPC codes belong to the class of linear block codes with implementable decoders, which bring near-capacity performance on a variety of channels, e.g., binary-input additive white Gaussian noise (AWGN) channels in [1]. In [2], a novel family of protograph LDPC codes has been proposed, which can achieve not only linear complexity encoding and high-speed decoding by means of a quasi-cyclic (QC) structure but also near-interruption performance over different block fading channels. LDPC codes was also used to design bit-interleaved coded modulation (BICM) with iterative demapping and decoding functions, which bring improved throughput under the constraints of limited bandwidth [3].

In the literature, there exist two major construction methods for LDPC codes: 1) random construction approach, e.g., random or semi-random methods in [4], 2) algebraic construction approach, e.g., algebraic construction of geometric LDPC codes and quasi-cyclic LDPC codes in [5]–[9]. However, when constructing short- or medium-length high-rate LDPC codes, the two aforementioned construction methods can not bring satisfactory performance in terms of bit error rate (BER). The research on short-length error correction codes primarily concentrates on the algebraic codes, such as Golay codes and quadratic residue (QR) codes. Their minimum Hamming distances can be equivalent to the theoretical maximum, but they are not tailored for low- and medium-complexity decoding algorithms, e.g., iterative message passing algorithm and its variants. These motivates the study of construction of short- or medium-length high-rate LDPC codes by leveraging other novel construction methods in this letter.

Matroid theory was first proposed by Whitney, and the relationship between matroids and graphs was built by. Edmonds and Fulkerson realized that matroids play an critical role in the transversal theory. Greene derived the MacWilliams identities from matroid theory, which are now widely used in the fields of combinatorial optimization, network theory [10], and coding theory [11]. In the letter, we study the construction of LDPC codes with a given girth, which is well interpreted by the matroid theory.

II. PRELIMINARIES

A. Fundamentals of LDPC Codes

LDPC codes fall into the category of linear block codes with their parity-check matrices being sparse. The parity-check matrix of a LDPC code $(n,k)$ is defined as $H \in \mathbb{B}^{(n-k) \times n}$ with $\mathbb{B} = \{0,1\}$. A regular LDPC code has fixed column weight $W_c$ and row weight $W_r$, where $W_r = W_c \cdot (n/(n-k))$, $W_c \ll n-k$, and $W_r \ll n$. The code rate $R$ for a regular LDPC code is

$$R = \frac{k}{n} = 1 - \frac{W_c}{W_r}. \quad (1)$$

Unlike regular LDPC codes, an irregular LDPC code has more than one column weights and row weights. Usually, the irregular LDPC code is characterized by the degree distribution (i.e., distribution of column weights and row weights) [12].
B. Graphical Representation of LDPC Codes

The parity-check matrix of LDPC codes can be represented by a bipartite Tanner graph, composed of two kinds of nodes, i.e., variable nodes (VNs) and the check nodes (CNs), denoted by $V_i$ and $C_j$, respectively, for $i = 1, \cdots, n - k$, and $j = 1, \cdots, n$. If $h_{ij} = 1$, the $i$-th check node ($C_j$) is connected to the $j$-th variable node ($V_i$) in the associated Tanner graph. For the purpose of better illustration, an example of an LDPC code with $n = 8$ and $k = 3$ is provided as

$$
H = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix},
$$

with its corresponding Tanner graph shown in Fig. 1.

C. Cycles in Tanner Graphs

A cycle in a Tanner graph refers to a trajectory that begins with one of the vertices, traverses through several vertices, and finally returns to the original vertex [13]. For instance, the four tandem connected thick edges in Fig. 1 form a cycle of length 4. The girth is defined as the length of the shortest cycle in the Tanner graph (e.g., the girth is 4 for Fig. 1), which has a great impact on the performance of iterative message passing decoding. Therefore, in the construction process of a parity-check matrix, great efforts should be put on maximizing the girth. In this letter, we construct short-length high-rate LDPC codes with enhanced performance through the matroid theory.

III. NOVEL CONSTRUCTION OF HIGH-RATE LDPC CODES

In principle, short cycles should be avoided in the process of constructing the parity-check matrix. In the literature, the theorems on the girth are proposed from the perspective of matrix or graph theory. In this letter, matrix and graph theory are transformed into set problem via matroid theory, and the theorems on determining the girth are provided from the perspective of matroid theory.

A. Definition and Theorems

Each column of the parity-check matrix can be represented by a set. The support of non-zero elements in vector $h_i$ (i.e., the $i$-th column of $H$) is represented by the set $L_i$, for $i = 1, \cdots, n$. For example, the parity-check matrix in (2) can be represented by the following sets: $L_1 = \{1\}$, $L_2 = \{2\}$, $L_3 = \{3\}$, $L_4 = \{4\}$, $L_5 = \{5\}$, $L_6 = \{1, 2, 3\}$, $L_7 = \{1, 4, 5\}$, $L_8 = \{2, 3, 4\}$, $L_9 = \{1, 2, 5\}$, $L_{10} = \{3, 4, 5\}$. Equivalently, the corresponding parity-check matrix can be constructed according to $L_i$, for $i = 1, \cdots, n$.

Definition 1. A matroid $M$ is an ordered pair of $(E, I)$, where $E$ is a finite set, and $I$ is a collection of subsets of $E$. They satisfy the following three conditions [14]:

1. $\emptyset \in I$.
2. If $S \subseteq T \subseteq E$, then $S \in I$.
3. If $S \subseteq T \subseteq E$ and $|S| < |T|$, then there exists $e \in T \setminus S$ such that $S \cup \{e\} \in I$.

Following the matroid theory, we first construct a matroid with all sets whose cardinalities are less than or equal to $W_e$, and then select parts of the sets which satisfy the predetermined conditions to form a parity-check matrix. The matroid theory is leveraged for analyzing the relationship among sets, which represent the columns of the corresponding parity-check matrix.

A series of theorems on the girth of the LDPC code, listed in the sequel, are obtained by transforming the relationship among columns into that among sets based on the matroid theory [6], [15].

Theorem 1. If $\exists |L_i \cap L_j| \geq 2$ for $i, j$, where $i \neq j$, then the girth of the LDPC code is 4.

Proof. It can be easily seen from the relationship between the parity-check matrix and its corresponding Tanner graph shown in the previous example. If $|L_i \cap L_j| = 2$, e.g., $|L_6 \cap L_8| = 2$ for $H$ in (2), a length-4 cycle exists.

Theorem 2. If $L_i$, for $i = 1, \cdots, n$, satisfies the following two conditions, then the girth of the LDPC code is greater than or equal to 8.

1. $|L_i \cap L_j| < 2, i \neq j, \forall i, j \in \{1, 2, \cdots, n\}$.
2. $|\{L_i \cap L_j\} \cup (L_i \cap L_k) \cup (L_j \cap L_k)| < 3, i \neq j \neq k, \forall i, j, k \in \{1, 2, \cdots, n\}$.

Proof. If there exists a Tanner graph with girth 6, there exist 3 sets and the unions of intersections of every 2 sets out of the 3 sets must have no less than 3 elements, e.g., $|L_6 \cap L_8| \cup |L_5 \cap L_6| \cup |L_8 \cap L_6| = 3$ for $H$ in (2). Therefore, the girth of the LDPC code is at least 8.

Theorem 3. Let $L(n) = \{1, 2, \cdots, n\}$ denote a set having $n$ elements, and $T(t)$ represent a subset of $L(n)$ with $t$ elements ($t \leq n$). If $L_i$, for $i = 1, \cdots, n$, satisfies the following $t_1$ conditions, the girth of the Tanner graph is at least $2(t_1 + 2)$.

Conditions: $|\cup_{i,j \in T(t) \subseteq L(n)} (L_i \cap L_j)| < t, t = 2, 3, \cdots, t_1 + 1$.

Proof. The proof can be done by the method proof by contradiction. If there exists a length-4 cycle, $\exists i, j \in T(2) \subseteq L(n)$, $|L_i \cap L_j| = 2$, which contradicts with the first condition when $t = 2$. If there exists a length-6 cycle, $\exists i, j, k \in T(3) \subseteq L(n)$, $|L_i \cap L_j| \cup (L_i \cap L_k) \cup (L_j \cap L_k)| = 3$, which contradicts with the second condition when $t = 3$. By analogy, if there exists a length-2($t_1 + 1$) cycle, $\exists i_1, \cdots, i_{t_1 + 1} \in T(1) \subseteq L(n)$, $|\cup_{i_1, i_2 \in T(1)} (L_{i_1} \cap L_{i_2}) \cup (L_{i_2} \cup L_{i_3}) \cup \cdots \cup L_{i_{t_1 + 1}}| = t_1 + 1$, which contradicts with the last condition when $t = t_1 + 1$. In other words, if all the $t_1$ conditions in Theorem 3 are simultaneously satisfied, there does not exist any cycles with length less than or equal to $2(t_1 + 1)$, i.e., the girth of the Tanner graph is at least $2(t_1 + 2)$.
B. Construction of High-Rate LDPC Codes

Algorithm 1 CONSTRUCTION OF H2

Input:
1. r, Wc; \( \triangleright \) row number and column weight of H2.

Output:
2. H2: initialized as H2 = [ ].
3. Begin
4. \( \mathbb{L}_1 = \{ l_1, l_2, \ldots, l_i, W_c \} \subset \{ 1, \ldots, r \} \); \( \triangleright \) \( \mathbb{L}_1 \) is the first column indexed by \( \mathbb{L}_1 \).
5. for \( j = 2 : 1 : \lfloor \frac{r}{W_c} \rfloor \) do
6. \( \triangleright \) \( \mathbb{L}_j \) is the column number of current H2.
7. \( \mathbb{L}_j = \{ j_1, j_2, \ldots, j_i, W_c \} \subset \{ 1, \ldots, r \} \);
8. if \( \bar{c}_H \leq 2 \) then
9. \( \triangleright \) \( q(m) = \text{cal}(\mathbb{L}_j, \mathbb{L}_m); \triangleright \) \( \text{cal}(\mathbb{L}_j, \mathbb{L}_m) \) calculates \( |\mathbb{L}_j \cap \mathbb{L}_m| \).
10. end for
11. if \( \max(q) < 2 \) then
12. \( \triangleright \) \( \max(q) \) returns the maximum value in the vector q.
13. \( H_2 = [ H_2 | \mathbb{L}_{i_j} ]; \)
14. else
15. \( H_2 = [ H_2 ]; \)
16. end if
17. else
18. for \( s = 1 : 1 : \bar{c}_H_2 \) do
19. \( q(s) = \text{cal}(\mathbb{L}_j, \mathbb{L}_s); \)
20. \( \triangleright \) \( \text{cal}(\mathbb{L}_j, \mathbb{L}_s) \) calculates \( |\mathbb{L}_j \cap \mathbb{L}_s| \).
21. end for
22. end for
23. if \( \max(Q) < 3 \) \&\& \( \max(q) < 2 \) then
24. \( \triangleright \) \( \max(Q) \) returns the maximum value in the matrix Q.
25. \( H_2 = [ H_2 | \mathbb{L}_{i_j} ]; \)
26. else
27. \( H_2 = [ H_2 ]; \)
28. end if
29. end if
30. end for
31. End

Given \( W_c \) and \( r \) (i.e., \( r = n - k \) ), the maximum number of columns of \( H_2 \) can be roughly determined by Theorem 4, shown in the sequel.

Theorem 4. The parity-check matrix of the LDPC code is assumed to be in the form of \( H = [ I | H_2 ] \), where \( H_2 \) is a submatrix with column weight being \( W_c \) without length-4 and length-6 cycles. The column number of \( H_2 \) is upper bounded as

\[
\bar{c}_H_2 \leq \lfloor \frac{r - 1}{W_c - 1} \rfloor \cdot \frac{r}{W_c} \quad (3)
\]

The structure is favorable since the generator matrix can be easily obtained. Moreover, we only need to focus on the construction of a smaller matrix \( H_2 \) instead of \( H \).

Proof. If \( |\mathbb{L}_j^H \cap \mathbb{L}_j^H| < 2, i \neq j, \forall i, j \in \{ 1, 2, \ldots, W_c \} \), the weight of each row in \( H_2 \) is less than or equal to \( \lfloor \frac{r - 1}{W_c - 1} \rfloor \), where \( \mathbb{L}_j^H \) denotes the set representation of the \( i \)-th column of \( H_2 \) and \( |x| \) returns the greatest integer less than or equal to \( x \). The total number of 1’s in \( H_2 \) can be expressed as \( W_c \cdot \bar{c}_H_2 \), so we get:

\[
| r - 1 |
\leq \frac{r}{W_c - 1} \cdot \bar{c}_H_2 \quad (4)
\]

This letter focuses on the construction of high-rate LDPC codes, which avoid length-4 and length-6 cycles in the Tanner graphs. The algorithm for constructing \( H_2 \) is presented as follows:

Step 1: \( \text{List } \lfloor \frac{r}{W_c} \rfloor \text{ sets with each set containing } W_c \text{ elements out of } \{ 1, \ldots, r \}. \)

Step 2: \( \text{Randomly select one set out of the } \lfloor \frac{r}{W_c} \rfloor \text{ sets as the first column of } H_2. \)

Step 3: \( \text{Select another set as the second column of } H_2 \) by following Condition 1 of Theorem 2.

Step 4: \( \text{Construct the remaining columns of } H_2 \text{ according to Condition 1 and Condition 2 of Theorem 2.} \)

The pseudocode for constructing \( H_2 \) is provided in Algorithm 1.

The algorithm does not have a pre-fixed row weight and column number for \( H_2 \). Nevertheless, the proposed algorithm can guarantee the girth of constructed LDPC codes is at least 8.

IV. SIMULATION RESULTS

In the simulations, we set \( W_c = 4 \) and \( W_c = 6 \) and evaluate the BER performance over AWGN channels with binary phase shift keying (BPSK) modulation and sum-product iterative decoding algorithm. The maximum number of iterations is set to 100 and the maximum erroneous bits are set 3000 for all the signal-to-noise ratios (SNRs). Simulation results for the LDPC codes constructed by Algorithm 1 are shown in Fig. 2 with different code rates and lengths. The code (774, 672) with \( W_c = 4 \) has the best BER performance due to its lowest code rate (even though it also has the shortest length), while the three LDPC codes with \( W_c = 6 \) have similar BER performance.

The construction methods from [16]–[18] are taken as the benchmark schemes. Fig. 3 shows the BER performance between our constructed LDPC codes and those from [16]–[18]. The code rates of our constructed LDPC codes are slightly lower than those from [16]–[18], while the code lengths are (almost) the same. Our constructed LDPC codes outperform those from [16]–[18] in terms of BER performance. When the BER is at the level of \( 10^{-4} \), there exist an approximate 0.25 dB gain between our constructed LDPC code and that from [16] and an approximate 0.7 dB gain between our constructed LDPC code and that from [17]. When the BER is at the level of \( 10^{-7} \), there exists an approximate 0.25 dB gain between our constructed LDPC code and that from [18].

1In practice, the number of columns constructed by the proposed algorithm is difficult to reach the theoretical upper bound, shown in (3).
The degree distributions of our constructed LDPC codes are those from [16]–[18].

**TABLE I: THE DEGREE DISTRIBUTION OF OUR CONSTRUCTED (2121, 1974) LDPC CODE**

<table>
<thead>
<tr>
<th>Degree of CNs</th>
<th>25</th>
<th>26</th>
<th>27</th>
<th>29</th>
<th>30</th>
<th>31</th>
<th>33</th>
<th>34</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of CNs</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Degree of CNs</td>
<td>35</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>40</td>
<td>41</td>
<td>42</td>
</tr>
<tr>
<td>Number of CNs</td>
<td>2</td>
<td>3</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Degree of CNs</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
<td>48</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>Number of CNs</td>
<td>5</td>
<td>9</td>
<td>11</td>
<td>15</td>
<td>15</td>
<td>11</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Degree of CNs</td>
<td>51</td>
<td>52</td>
<td>53</td>
<td>54</td>
<td>55</td>
<td>56</td>
<td>57</td>
<td>58</td>
</tr>
<tr>
<td>Number of CNs</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>9</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>

**TABLE II: THE DEGREE DISTRIBUTION OF OUR CONSTRUCTED (243, 158) LDPC CODE**

<table>
<thead>
<tr>
<th>Degree of CNs</th>
<th>6</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of CNs</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>20</td>
<td>24</td>
</tr>
<tr>
<td>Degree of CNs</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Number of CNs</td>
<td>11</td>
<td>9</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE III: THE DEGREE DISTRIBUTION OF OUR CONSTRUCTED (674, 520) LDPC CODE**

<table>
<thead>
<tr>
<th>Degree of CNs</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of CNs</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>5</td>
<td>15</td>
</tr>
<tr>
<td>Degree of CNs</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>Number of CNs</td>
<td>11</td>
<td>12</td>
<td>27</td>
<td>31</td>
<td>10</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Degree of CNs</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Number of CNs</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

V. Conclusions

In the paper, a novel high-rate LDPC code (including short- and medium-length) construction scheme has been proposed based on the matroid theory. Simulation results have shown that the BER performance of the constructed LDPC codes is better than those constructed in the literature over AWGN channels with avoidance of length-4 and length-6 cycles. It would be of great interest to study the avoidance of length-8 cycles in order to construct even better LDPC codes as our future investigation.

References


