Construction of LDPC codes over GF(q) with modified progressive edge growth

CHEN Xin (✉), MEN Ai-dong, YANG Bo, QUAN Zi-yi

School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract

A parity check matrix construction method for constructing a low-density parity-check (LDPC) codes over GF(q) (q>2) based on the modified progressive edge growth (PEG) algorithm is introduced. First, the nonzero locations of the parity check matrix are selected using the PEG algorithm. Then the nonzero elements are defined by avoiding the definition of subcode. A proof is given to show the good minimum distance property of constructed GF(q)-LDPC codes. Simulations are also presented to illustrate the good error performance of the designed codes.

Keywords  LDPC codes over GF(q), progressive edge growth, large minimum distance

1 Introduction

Binary LDPC [1] codes have displayed near Shannon limit performance when decoded using the belief propagation algorithm [2]. Davey and MacKay have investigated LDPC codes on finite fields GF(q) when q>2 in Ref. [3]. The results indicate that system performance is improved with the increase of q when the proper design of parity check matrix on GF(q)-LDPC codes is conducted. It has been shown in Ref. [3] that as q becomes large the best performances at finite length are obtained for ‘ultra-sparse’ LDPC codes, that is, with the minimum connectivity on the symbol nodes \(d_v = 2\). For GF(q)-LDPC codes, the optimization problem is generally solved in a disjoint manner. First, the positions of the nonzero entries of the parity check matrix \(H\) associated with the nonbinary code are optimized to achieve good girth properties and minimize the impact of cycles on the belief propagation (BP) decoding algorithm. Then the nonzero entries can be selected either randomly from a uniform distribution among nonzero elements of GF(q) [4] or carefully to meet design criteria as done in Refs. [5–6]. In Ref. [7], the author proposed a method to select nonzero elements so that the determinants of the partial matrices corresponding to the cycles in the parity check matrix did not become zero. The author declared that the nonbinary parity check matrixes constructed in that way were able to give large-weight codewords, whereas no proof was given to show the good property of the codes.

In this article, a proof is presented on why the design criteria presented in Ref. [7] can construct a nonbinary LDPC codes with good distance property. A proposition is presented to show the good minimum distance property of the constructed codes. Also, a combined method is proposed to construct GF(q)-LDPC codes using the design criteria introduced in Ref. [7] and the PEG algorithm described in Ref. [4].

The remainder of this article is organized as follows. Sect. 2 provides a proof of good minimum distance property of GF(q)-LDPC codes constructed using the design criteria proposed in Ref. [7]. Sect. 3 gives a combined algorithm for constructing GF(q)-LDPC codes with the modified PEG algorithm. Sect. 4 presents the performance comparison of the codes by different construction methods. Finally, concluding remarks are given in Sect. 5.

2 Minimum distance problem of \((2, d_v)\) GF(q)-LDPC codes

An LDPC code of code rate \(R\) and code length \(N\) is defined...
by an \( M \) row and \( N \) column parity check matrix \( H \). This LDPC code is called a GF(q)-LDPC code when each element \( h_{i,j} \) of \( H \) is selected from Galois field \( \text{GF}(q) \), with \( q = 2^{n} \) and \( q \) as the order of the field. A row vector \( c \) of length \( N \) is a codeword if \( Hc = 0 \). A regular \((d_a,d_c)\) LDPC codes have constant column weight \( d_a \) and row weight \( d_c \).

2.1 Cycles and subcode

Let \( H_i \) be the block matrix representation of a cycle of length \( 2L \) extracted from parity check matrix \( H \) through row and column transformation. Then, the parity check \( H \) can be represented as:

\[
H = \begin{bmatrix}
X_1 & H_L & X_2 \\
X_1 & 0 & X_1
\end{bmatrix}
\]

The sub-matrix \( H_L \) is given by the following \( L \times L \) block square matrix:

\[
H_L = \begin{bmatrix}
a_{1,1} & a_{1,2} & 0 & \cdots & 0 \\
0 & a_{2,2} & a_{2,3} & \cdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & a_{L-1, L-1} & a_{L-1, L} & \cdots & a_{L, L}
\end{bmatrix}
\]

A vector \( c = [0 \ c_{ab} \ 0] \) satisfying Eq. (1) is a codeword of the LDPC code defined by the parity check matrix \( H \). Here, \( c_{ab} \) is a vector of length \( L \).

\[
He_L^T \begin{bmatrix}
X_1 \\
X_2 \\
X_3
\end{bmatrix} = \begin{bmatrix}
0 \\
0 \\
0
\end{bmatrix}
\]

According to Eq. (1), it follows that:

\[
H_L \times e_{ab} = 0
\]

If \( \det(H_L) \neq 0 \) (det: determinant of a matrix), \( c_{ab} \) will be unambiguously determined as a zero vector. Otherwise, \( c_{ab} \) will have multiple choices of values and thus defined as a subcode of the global code.

**Proof** (All arithmetic is operated on Galois field) Through row and column transformation, \( H_L \) can be represented as:

\[
H_L = \begin{bmatrix}
a_{1,1} & a_{1,2} & 0 & \cdots & 0 \\
0 & a_{2,2} & a_{2,3} & \cdots & \vdots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & a_{L-1, L-1} & a_{L-1, L} & \cdots & a_{L, L}
\end{bmatrix}
\]

Here, \( A = \prod_{j=1}^{L} \prod_{i=1}^{L} a_{i,j} \). Because \( H_L \times e_{ab} = 0 \), two sub-equations are obtained:

\[
\begin{align*}
\sum_{i=1}^{L} a_{i,1} \sum_{j=1}^{L} e_{i,j} a_{i,1} &= 0 \\
\sum_{i=1}^{L} a_{i,2} \sum_{j=1}^{L} e_{i,j} a_{i,2} &= 0
\end{align*}
\]

If \( A \neq 0 \), then \( \det(H_L) \neq 0 \), and thus one can easily obtain \( c_{ab} (L-1) = 0 \) and \( c_{ab} (L) = 0 \). Input this to Eq. (2), one gets \( c_{ab} = (0, 0, \ldots, 0) \).

2.2 Minimum distance

The minimum distance problem of linear codes is to find the minimum distance \( d_{\text{min}} \) of the corresponding code defined by a parity check matrix \( H \). The error correction capability of a linear code is obviously related to its minimum distance. Therefore, finding \( d_{\text{min}} \) is a fundamental computational problem in coding theory. A polynomial time algorithm to compute the distance would be the ideal solution to the problem. Unfortunately, no such algorithm is known. The complexity of this problem is generally NP-hard [8], while for a \( d_a = 2 \) LDPC code, the minimum distance can be exactly computed in polynomial time.

**Proposition 1** Given a nonbinary low density parity check matrix with \( d_a = 2 \) nonzero elements of each column, assume the length of the shortest cycle is \( L \) in the condition of \( \det(H_L) = 0 \). Then the minimum distance \( d_{\text{min}} \) of the corresponding nonbinary LDPC codes satisfies

\[
d_{\text{min}} = \frac{L}{2}
\]

which can be computed in polynomial time.

**Proof** If \( \det(H_L) \neq 0 \), from Eq. (4), there will be a nonzero code \( c = [0 \ c_{ab} \ 0] \) satisfying \( Hc = 0 \). Because the length of vector \( c_{ab} \) is \( L/2 \), one can obtain an upper bound \( d_{\text{min}} \leq L/2 \). To show that \( d_{\text{min}} \) also satisfies \( d_{\text{min}} \geq L/2 \), it can be considered that there is an ‘active’ subgraph in the Tanner graph induced by a minimum weight codeword. By adopting the same notation in Ref. [9], a symbol node will be called an active symbol node if its associated value in the minimum weight codeword is nonzero. The edges connected with active symbol nodes will be called active edges, and the check nodes with at least one active incident edge will be called active check nodes. Because each active symbol node in a cycle code has only two edges, it can be deduced that, starting from any active edge of a symbol node, there must be
a close path (cycle) with the active subgraph coming back from the other edge of the same symbol node. As the length of the shortest cycle is \( L \) under the condition of \( \det(H_{ij}) = 0 \), one has \( d_{\text{min}} \geq L/2 \), therefore, \( d_{\text{min}} = L/2 \). The girth of a Tanner graph can be computed in time proportional to \( n^3 \) [10–11].

### 3 Construction of GF\((q)\)-LDPC codes using the PEG algorithm

A method is proposed for constructing parity check matrix \( H \) over GF\((q)\). In the progressive edge growth algorithm, after nonzero location \( h_{ij} \) is defined by certain criterion. The value of \( h_{ij} \) is selected so that the determinants of corresponding \( H_{ij} \) do not become zero. Here, \( H_L \) means that all of the sub-matrices defined by loops get through nonzero element \( h_{ij} \). Hence, it is possible to avoid several subcodes defined by \( H_L \), thus reducing the number of low-weight codewords of the global code. The decoding performance can also be improved.

#### 3.1 Definitions and notations

Adopting the notation in Ref. [4], a bipartite graph with \( M \) check nodes in one class and \( N \) symbol nodes in the other can be created using \( H \) as the integer valued incidence matrix for the two classes. A parity check matrix \( H \) can also be denoted as \((V,E)\) with \( V \) as the set of nodes, i.e. \( V = V_c \cup V_s \), where \( V_c = \{c_1, c_2, \ldots, c_m\} \) is the set of check nodes and \( V_s = \{s_1, s_2, \ldots, s_n\} \) the set of variable nodes. \( E \) is the set of edges with edge \((c_i, s_j, a) \in E = E_{ij} \cup E_{ik} \cup \ldots \cup E_{in} \) if and only if \( h_{ij} = a \), \( h_{ij} \in H \), \( 1 \leq i \leq m \), \( 1 \leq j \leq n \). A regular \((d_c, d_s)\) code has every variable node participates in \( d_s \) check nodes and every check node involves \( d_c \) symbol nodes. Also, let the set of edges \( E \) be partitioned in terms of \( V_s \) as \( E = E_{ij} \cup E_{ik} \cup \ldots \cup E_{in} \), with \( E_{ij} \) containing all edges incident on variable node \( v_j \). Finally, denote the \( k \)th edge incident on \( v_j \) by \( E_{ijk} \), \( 1 \leq k \leq d_s \), \( a \in \{1,2,\ldots,q-1\} \).

For a given symbol node \( v_j \), define its neighbor within depth \( l \), \( N^l_{v_j} \), as the set consisting of all check nodes reached by a tree spreading from symbol node \( v_j \) within depth \( l \). Its complementary set, \( \overline{N^l_{v_j}} \), is defined as \( V_c \setminus N^l_{v_j} \).

In graph theory, girth \( g \) refers to the length of the shortest cycle in a graph. For each symbol node \( v_j \), a local girth \( g_{v_j} \) is defined as the length of the shortest cycle passing through that symbol node. The set of local girth \( \{g_{v_j}\} \) is referred to as girth histogram, then one has \( g = \min_{v_j} \{g_{v_j}\} \).

#### 3.2 GF\((q)\)-PEG algorithm

The authors describe the GF\((q)\)-PEG algorithm for construction a GF\((q)\)-LDPC codes with \( n \) variable nodes and \( m \) check nodes as follows:

For every variable node \( v_j \) \((j=1,2,\ldots,n)\), if \( k=1 \), we have \( E_{ij} = \text{edge}(c_i, s_j, a) \), where \( E_{ij} \) is the first edge incident to variable node \( v_j \), and \( a \) is one check node such that it has the lowest check node degree under the current graph setting \( E_{ij} \cup E_{ik} \cup \ldots \cup E_{in} \). \( a_{ij} \) is selected randomly from the GF\((q)\) elements, that is, arbitrarily from set \( \{1,2,\ldots,q-1\} \). If \( 1 < k \leq d_s \), we expand a subgraph from symbol node \( v_j \) up to depth \( l \) under the current graph setting such that the cardinality of \( N^l_{v_j} \) stops increasing but is less than \( m \), or \( \overline{N^l_{v_j}} \neq \emptyset \) but \( \overline{N^l_{v_j}} = \emptyset \), then \( E_{ij} = \text{edge}(c_i, s_j, a_{ij}) \), where \( E_{ij} \) is the \( k \)th edge incident to \( v_j \), and \( a \) is one check node selected from the set \( \overline{N^l_{v_j}} \) having the lowest check node degree. The value of the \( a_{ij} \) is selected so that the determinant does not become zero for each newly added cycle (sub-matrix) get through \( \text{edge}(c_i, s_j, a_{ij}) \).

### 4 Simulation results

In this section, the performance results of three construction methods are compared. The three different construction methods are random method (RM), Davey Mackay method (DM) and the proposed method (PM). For random method, the nonzero entries are randomly selected from the nonzero elements in the fields GF\((q)\). For DM method, the rows of parity check matrix are generated randomly from the \( d_c \) nonzero entries or from the \( d_s \) nonzero entries multiplied by constants (the \( d_s \) nonzero entries are previously optimized according to Ref. [6]). For different frame lengths and field orders, the bit error rates (BER) is compared assuming a memory less binary input additive white Gaussian noise (BI-AWGN) transmission channel and an iterative BP decoder over GF\((q)\) at the receiver. The maximum number of the iterations is fixed to 50.
For different parity check matrices constructed by RM, DM and PM, the number of small loops with $\det(H_{L_2}) = 0$ are listed in Table 1. According to proposition 1, the authors are able to calculate the minimum distance of different codes. The minimum distance of PM codes is 5 (over GF(8)) and 6 (over GF(16)). And for RM and DM codes, these values are reduce to 3 (over GF(8)) and 4 (over GF(16)). Thus, one can expect a better performance for PM codes.

<table>
<thead>
<tr>
<th>Codes</th>
<th>Length</th>
<th>GF(8)</th>
<th>GF(16)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>RM-A</td>
<td>0</td>
<td>22</td>
<td>63</td>
</tr>
<tr>
<td>RM-B</td>
<td>1</td>
<td>37</td>
<td>69</td>
</tr>
<tr>
<td>DM</td>
<td>2</td>
<td>16</td>
<td>73</td>
</tr>
<tr>
<td>PM-A</td>
<td>0</td>
<td>0</td>
<td>30</td>
</tr>
<tr>
<td>PM-B</td>
<td>0</td>
<td>0</td>
<td>31</td>
</tr>
</tbody>
</table>

Fig. 1 shows the performance of the (2,6) regular LDPC codes over GF(8) by various construction methods. All of the codes correspond to a block length of 189 bits (63 symbols). The minimum distance properties are listed in Table 1 (over GF(8)). As can be seen, the codes constructed using PM have a gain about 0.7 dB compared to the codes constructed by RM and DM in waterfall region. Note that for different PM codes, they have very stable bit error rate performance.

Fig. 2 shows the performance of the (2,6) regular LDPC codes over GF(16). The codes correspond to a block length of 504 bits (126 symbols). The minimum distance properties are listed in Table 1 (over GF(16), except PM-C). It is observed that the PM codes provide BER performance improvements by approximately 0.5–1 dB in the high SNR region compared with RM and DM codes. The error performances of different codes have confirmed the advantages of the proposed method.

5 Conclusions

In this article, the method of design GF($q$)-LDPC codes is addressed based on the modified progressive edge growth algorithm with the design criteria proposed in Ref. [7]. The nonzero elements in a parity check matrix over GF($q$) is selected so that the sub-matrices corresponding to the cycles cannot define a subcode. A proof was given to show the good minimum distance property of designed codes. It is shown that the minimum distance of (2, $d_c$) GF($q$)-LDPC codes can be computed in polynomial time. The simulation results also show that the LDPC codes constructed by the proposed method have good error performance.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (60672087).

References