On the generator matrix of array LDPC codes

Marco Baldi, Marco Bianchi,
Giovanni Cancellieri, Franco Chiaraluce
DIIT - Università Politecnica delle Marche
Ancona, Italy
Email: {m.baldi, m.bianchi, f.chiaraluce, g.cancellieri}@univpm.it

Torleiv Kløve
Department of Informatics - University of Bergen
Bergen, Norway
Email: Torleiv.Klove@ii.uib.no

Abstract—In this paper we present a general expression for the generator matrix of array low-density parity-check codes. This is a further contribution towards understanding the inner structure of these codes. Moreover, it represents a useful tool that can be used in the estimation and optimization of their minimum distance, which is an open problem. By using the new form of the generator matrix, we derive some necessary conditions on the maximization of the minimum distance.

I. INTRODUCTION

Array low-density parity-check (LDPC) codes (or LDPC array codes) represent a special case of array codes; the latter are a side class of codes, first studied by Mario Blaum and others [1]–[4]. In particular, array LDPC codes, first defined by Fan [5], are a special class of quasi-cyclic LDPC codes having a parity-check matrix formed by $M \times L$ circulant permutation matrices with fixed size. Since we deal with circulant permutation matrices, we indicate each of them through the value of its associated shift, that is, the column index (starting from zero) of the symbol 1 in its first row. In this way, we can define the binary parity-check matrix of a code by giving its corresponding base matrix, that contains the shift values. By using such notation, array LDPC codes have a base matrix that is similar to that of a Reed-Solomon (RS) code. For this reason, array LDPC codes are also sometimes called RS-LDPC codes.

The characteristic of the base matrix of an array LDPC code is to have the first row and the first column that are all-zero vectors. The second column (having index 1) characterizes the whole code. In fact, each column is obtained by multiplying that column by the corresponding column index. So, the third column is twice the second one, the fourth column is three times the second one and so on. Obviously, the shift values must be considered modulo $q$, where $q$ is a prime number. The original proposal of array LDPC codes only considered codes having the second column with elements $0, 1, 2, \ldots, M - 1$. These codes were later named proper array LDPC codes. The maximum value for $L$ is equal to $q$. In fact, if we consider a further column, it would be the all-zero vector, according to the code construction rule, and the code minimum distance would drop to 2. So, having a number of columns $L \leq q$ is necessary to design codes with good distance. The most common choice is to fix $L = q$, thus including all the possible shifts in each row of the base matrix. These codes are also called simple and full length array LDPC (SFA-LDPC) codes.

Estimating the minimum distance of array LDPC codes is not a trivial problem. Though a number of results are already available in the literature, the problem remains basically open. In this paper, we aim at providing a further tool for addressing such issue, that is, a general form for the generator matrix of an array LDPC code. Through the generator matrix, we are also able to obtain some new bounds on the minimum distance of the codes.

Rather tight upper bounds on the minimum distance of simple array LDPC codes can be found in [6]–[9]. In [6], in particular, the following upper bounds were found, which only depend on $M$: $d_{\text{min}} = 6$ for $M = 3$, $d_{\text{min}} = 12$ for $M = 4$, $d_{\text{min}} = 20$ for $M = 5$, $d_{\text{min}} = 32$ for $M = 6$. A first form of the generator matrix was also derived. These bounds were improved in [7], where the dependence on $q$ was also considered. For $M = 4$ it was shown in [7] that $d_{\text{min}} = 8$ for $q = 5, 7$, while $d_{\text{min}} \geq 10$ for $q > 7$. In addition, a new notation for describing the columns of the parity-check matrix was introduced, which helps finding low-weight codewords. The structure of some minimum weight codewords was also disclosed. These values were further improved in [8] and [9], where also multiplicities were studied. A fast procedure in searching for low weight codewords was also proposed, and it was found that these codes often have a rather poor minimum distance, that goes further and further from the upper bounds when $M$ increases. For example, it was found that, for many values of $q$, it is $d_{\text{min}} = 10$ for $M = 4$ and $d_{\text{min}} = 12$ for $M = 5$.

These results suggest that some variation of the array LDPC code construction method should exist, that is able to improve their minimum distance properties. A further indication is the fact that in [10] it is stated that a generic quasi-cyclic LDPC code formed by circular permutation matrices has minimum distance $d_{\text{min}} \leq (M + 1)!$ and a code with $M = 3$ and $d_{\text{min}} = 20$ was found.

A first "natural" modification consists in changing the second column of the base matrix from $0, 1, 2, \ldots, M - 1$ to $\Delta_0, \Delta_1, \Delta_2, \ldots, \Delta_{M-1}$, with $\Delta_0, \Delta_1, \Delta_2, \ldots, \Delta_{M-1}$ distinct integers $< q$. The other columns are still obtained by multiplying the second column by their column index. The minimum
distance of these codes was studied in [11] and [12], where they were called improper array codes. In those works, only full length codes are considered, that is, with $L = q$ and, therefore, length $n = q^2$.

Another solution for increasing the minimum distance is to shorten the code, that is, using a rather large value of $q$ and then limiting the number of columns to $L < q$. Moreover, a further degree of freedom can be obtained by rearranging the block columns before shortening the code. Some work in this direction was already done in [13] where, however, the main target was to increase the local cycles length. In fact, there is a strict relationship between the local cycles length and the minimum distance of array codes. The approach used in [13] consists in shortening an array code in such a way as to only retain those columns of its parity-check matrix whose indexes form a sequence that avoids solutions in some "cycle-governing" equations. This way, one can obtain array codes with a predetermined distribution of cycles of various lengths. The minimum distance is also studied, but mainly as a consequence of the cycle distribution, since the aim of the work [13] is not the maximization of the minimum distance.

Even by shortening the code, the maximum value of the minimum distance is $2^M$, as will be shown later. An open problem is to find the maximum value of $L$, $L_{\text{max}}(d_{\text{min}})$, which allows achieving the minimum distance $d_{\text{min}}$ for given $q$ and $M$. $L_{\text{max}}(d_{\text{min}})$ increases with increasing $q$, for a fixed $M$. Particular attention should be devoted to $L_{\text{max}}(2^M)$. Table I reports the values of $L_{\text{max}}(8)$, found heuristically, for the case of $M = 3$ and some increasing values of $q$. They have been found, for each $q$, through an exhaustive numerical analysis of the minimum distance, by testing several choices of $\Delta_0, \Delta_1, \Delta_2$, and progressively increasing the value of $L$, until the minimum distance drops to 6. The final values of $\Delta_0, \Delta_1, \Delta_2$ are listed in the table, although their choice seems not to be critical. The table also reports the nominal code rate, that is, $\frac{L - M}{L}$. This is slightly different from the actual code rate, due to the presence of $M - 1$ linearly dependent rows in the parity-check matrix $H$, as will be shown afterwards. We observe how the code rate increases as $q$ grows, together with the code length $n$.

The organization of the paper is as follows. In Section II we introduce some basic notation and definitions. In Section III we provide the general expression of the generator matrix of array LDPC codes. In Section IV we discuss the main issue of maximizing the minimum distance through a suitable choice of the design degrees of freedom.

**Table I**

<table>
<thead>
<tr>
<th>$q$</th>
<th>$\Delta_0, \Delta_1, \Delta_2$</th>
<th>$L_{\text{max}}(8)$</th>
<th>nominal rate</th>
<th>$[n, k]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>[0, 8, 10]</td>
<td>6</td>
<td>1/2</td>
<td>[66, 35]</td>
</tr>
<tr>
<td>17</td>
<td>[0, 12, 16]</td>
<td>8</td>
<td>5/8</td>
<td>[136, 87]</td>
</tr>
<tr>
<td>23</td>
<td>[0, 16, 22]</td>
<td>9</td>
<td>2/3</td>
<td>[207, 140]</td>
</tr>
<tr>
<td>29</td>
<td>[0, 17, 28]</td>
<td>10</td>
<td>7/10</td>
<td>[290, 205]</td>
</tr>
<tr>
<td>37</td>
<td>[0, 21, 36]</td>
<td>12</td>
<td>3/4</td>
<td>[444, 335]</td>
</tr>
</tbody>
</table>

II. NOTATION AND DEFINITIONS

Given the prime $q$, we denote as $I$ the $q \times q$ identity matrix, while $P$ is the circulant $q \times q$ matrix with first row $(0100 \cdots 0)$, that is

$$P = \begin{bmatrix} 0100 & \cdots & 0 \\ 0010 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0000 & \cdots & 01 \\ 1000 & \cdots & 00 \end{bmatrix}$$

All circulant permutation matrices can be expressed as powers of $P$, and we use the short-hand notation $t = P^t$. For example, $0 = I$ and $1 = P$.

Let $\{\Delta_0, \Delta_1, \Delta_2, \ldots, \Delta_{M-1}\}$ be a set of $M$ integers that are distinct modulo $q$. Without loss of generality, we assume that

$$\Delta_0 < \Delta_1 < \Delta_2 < \cdots < \Delta_{M-1} < q.$$ (1)

Let $0 < M \leq L \leq q$ and let $H$ be the $Mq \times Lq$ matrix given by

$$H = \begin{bmatrix} 0 & \Delta_0 & 2\Delta_0 & \ldots & (L-1)\Delta_0 \\ 0 & \Delta_1 & 2\Delta_1 & \ldots & (L-1)\Delta_1 \\ 0 & \Delta_2 & 2\Delta_2 & \ldots & (L-1)\Delta_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \Delta_{M-1} & 2\Delta_{M-1} & \ldots & (L-1)\Delta_{M-1} \end{bmatrix}$$ (2)

Diao et al. [14] considered a more general class of quasi-cyclic LDPC codes. We will use their method to determine the rank of $H$.

**Theorem 1**. The matrix $H$ has rank $Mq - M + 1$.

**Proof**. Let $\alpha$ be an element in $GF(2^{q-1})$ of order $q$. For $0 \leq t \leq q - 1$, let $B_t$ be the $M \times L$ matrix with elements $\alpha^{(t)\Delta_i}$, $0 \leq i \leq M - 1$, $0 \leq j \leq L - 1$. [14, Lemmas 2 and 3] describe linear reversible mappings that transform $H$ into the matrix

$$B = \begin{bmatrix} B_0 & O & O & \cdots & O \\ O & B_1 & O & \cdots & O \\ O & O & B_2 & \cdots & O \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ O & O & O & \cdots & B_{q-1} \end{bmatrix},$$

where $O$ is the all-zero matrix. We see that $B_0$ is the all-one matrix and so has rank 1. For $1 \leq t \leq q - 1$, $B_t$ is full rank. Indeed, the $M \times M$ matrix containing the first $M$ columns of $B_t$ is a Vandermonde matrix of full rank. Hence, the rank of $B$ (and $H$) is $1 + (q-1)M$.

Theorem 1 shows that $H$ can be used as the parity check matrix for an $[n, k; q]$ code $C(q, L, \{\Delta_0, \Delta_1, \Delta_2, \ldots, \Delta_{M-1}\})$, where

$$n = Lq \text{ and } k = q(L-M) + M - 1.$$ We remark that $C(q, L, \{\Delta_0, \Delta_1, \Delta_2, \ldots, \Delta_{M-1}\})$, is equivalent to

$$C(q, L, \{0, \Delta_1 - \Delta_0, \Delta_2 - \Delta_0, \ldots, \Delta_{M-1} - \Delta_0\}).$$
Hence one may choose $\Delta_0 = 0$ without loss of generality. If $L = q$, and, hence, $n = q^2$, the code is full length. For $0 \leq i \leq M$ and $0 \leq \ell \leq q - 1$, define the sets $\Omega_{i,\ell}$ by

$$
\Omega_{i,\ell} = \left\{ (x_0, x_1, \ldots, x_{M-1}) \in \mathbb{Z}_2^M \mid \sum_{j=0}^{M-1} x_j \equiv i, \sum_{j=0}^{M-1} x_j \Delta_j \equiv \ell \pmod{q} \right\}
$$

and let

$$
\Gamma_i = \sum_{\ell=0}^{q-1} |\Omega_{i,\ell}| \pmod{2}
$$

where $|\cdot|$ denotes the size of a set.

In particular,

$$
\begin{align*}
\Gamma_0 &= 0 = I, \\
\Gamma_1 &= \Delta_0 + \Delta_1 + \Delta_2 + \cdots + \Delta_{M-1}, \\
\Gamma_2 &= \Delta_0 + \Delta_1 + \Delta_0 + \Delta_2 + \Delta_0 + \Delta_3 + \Delta_1 + \Delta_2 + \Delta_0 + \Delta_3 + \Delta_2 + \Delta_3 + \Delta_{M-2} + \Delta_{M-1}, \\
& \vdots \\
\Gamma_M &= \Delta_0 + \Delta_1 + \Delta_2 + \cdots + \Delta_{M-1}.
\end{align*}
$$

**Example 1:** For $M = 4$, we get

$$
\begin{align*}
\Gamma_0 &= 0 = I, \\
\Gamma_1 &= \Delta_0 + \Delta_1 + \Delta_2 + \Delta_3, \\
\Gamma_2 &= \Delta_0 + \Delta_1 + \Delta_0 + \Delta_2 + \Delta_0 + \Delta_3 + \Delta_1 + \Delta_2 + \Delta_0 + \Delta_3 + \Delta_2 + \Delta_3, \\
\Gamma_3 &= \Delta_0 + \Delta_1 + \Delta_2 + \Delta_0 + \Delta_1 + \Delta_3 + \Delta_0 + \Delta_2 + \Delta_1 + \Delta_2 + \Delta_3, \\
\Gamma_4 &= \Delta_0 + \Delta_1 + \Delta_2 + \Delta_3.
\end{align*}
$$

**III. THE GENERATOR MATRIX OF ARRAY LDPC CODES**

Let $G_0$ be the $q \times (M + 1)q$ matrix defined by

$$
G_0 = \left[ \Gamma_0 | \Gamma_1 | \Gamma_2 | \cdots | \Gamma_M \right].
$$

We see that the rows of $\Gamma_i$ have weight

$$
\sum_{\ell=0}^{q-1} |\Omega_{i,\ell}| \pmod{2} \leq \sum_{\ell=0}^{q-1} |\Omega_{i,\ell}| = \binom{M}{i} = 2^M.
$$

with equality if and only if $|\Omega_{i,\ell}| \leq 1$ for all $\ell$. Moreover, any row of $G_0$ has weight at most $\sum_{i=0}^{M} \binom{M}{i} = 2^M$.

**Theorem 2:** The $[q(L - M) + M - 1] \times Lq$ matrix $G$ defined by (3), where $I$ is the all-one vector of length $q$ and all the missing elements are zero, is a generator matrix for $C(q, L, \{\Delta_0, \Delta_1, \Delta_2, \ldots, \Delta_{M-1}\})$.

**Proof:** We will show that the inner product of a row $g$ in $G$ and a row $h$ in $H$ (considered as real vectors) is even. Let us consider one of the first $(L - M)q$ rows of $G$. The row meets a copy of $G_0$ in $G$, and it is zero outside this copy of $G_0$. Because the code is quasi-cyclic (cyclic within each block) it is sufficient to consider the top rows of the $G_0$. There are integers $m \in [0, M - 1]$ and $t \in [0, q - 1]$ such that the row from $H$ has a one in position $p_i = (i\Delta_m + t) \pmod{q}$ within the $i$th block of $G_0$ (that is $\Gamma_i$). Then

$$
g \cdot h \equiv \sum_{i=0}^{M-1} |\Omega_{i,p_i}| \pmod{2}. \quad (4)
$$

By definition, $(x_0, x_1, \ldots, x_{M-1}) \in \Omega_{i,p_i}$ if and only if

$$
\sum_{j=0}^{q-1} x_j \Delta_j \equiv p_i \equiv t + \Delta_m \sum_{j=0}^{q-1} x_j \pmod{q},
$$

that is

$$
(x_0, x_1, \ldots, x_{M-1}) \in \Omega_{i,p_i} \Leftrightarrow \sum_{j=0}^{q-1} x_j (\Delta_j - \Delta_m) \equiv t \pmod{q}. \quad (5)
$$

For $(x_0, x_1, \ldots, x_{M-1}) \in \Omega_{i,p_i}$, define $(y_0, y_1, \ldots, y_{M-1})$ by $y_m = 1 - x_m$ and $y_j = x_j$ for $j \neq m$. By (5),

$$(y_0, y_1, \ldots, y_{M-1}) \in \Omega_{i+\epsilon,p_{i+\epsilon}},$$

where

$$
\epsilon = 1 \text{ if } x_m = 0 \text{ and } \epsilon = -1 \text{ if } x_m = 1.
$$

Hence the elements in $\bigcup_{i=0}^{M-1} \Omega_{i,p_i}$ appear in pairs, that is, $\bigcup_{i=0}^{M-1} \Omega_{i,p_i}$ is even. By (4),

$$
g \cdot h \equiv \sum_{i=0}^{M-1} |\Omega_{i,p_i}| = \sum_{i=0}^{M-1} \left| \bigcup_{i=0}^{M-1} \Omega_{i,p_i} \right| \equiv 0 \pmod{2}.
$$

Hence, the inner product of any of the first $q(L - M)$ rows in $G$ and any row in $H$ is even. The inner product of any of the last $M - 1$ rows in $G$ and any row in $H$ is clearly 2. This completes the proof that $GH^T$ (where $H^T$ denotes the transpose of $H$) has only even elements, that is, $G$ is a generator matrix. \[\blacksquare\]
IV. LINKS WITH THE MINIMUM DISTANCE

In the previous section we have shown that the matrix \( G \) defined by (3) is a valid generator matrix for the array LDPC codes \( C(q, L; \{\Delta_0, \Delta_1, \Delta_2, \ldots, \Delta_{M-1}\}) \). The first \( q(L-M) \) rows of \( G \) have Hamming weight \( 2^M \), while its last \( M-1 \) rows have weight \( 2q \). So, the minimum distance of \( C \) is \( d_{\text{min}} \leq \min \{2^M, 2q\} \). Starting from this observation, the following questions arise:

1) The main general problem is: given \( q, L, \) and \( M, \) how do we choose \( \{\Delta_0, \Delta_1, \Delta_2, \ldots, \Delta_{M-1}\} \) to maximize the minimum distance of \( C \)?

2) A particular problem is the following: given \( M, \) for which (if any) \( q \) and \( L \) can we find \( \{\Delta_0, \Delta_1, \Delta_2, \ldots, \Delta_{M-1}\} \) such that \( C(q, L; \{\Delta_0, \Delta_1, \Delta_2, \ldots, \Delta_{M-1}\}) \) has minimum distance \( 2^M \) (that is, minimal maximum distance)?

3) In \( G_0 \), all the rows have the same weight. For the \( [(M+1)q, q, q] \) code generated by \( G_0 \), is the minimum distance always equal to the row weight of \( G_0 \)?

The first problem is still unsolved, and it represents the main result toward which work is still in progress. Some comments on the second problem are provided next.

We say that a set \( X = \{\Delta_0, \Delta_1, \Delta_2, \ldots, \Delta_{M-1}\} \) is an \((M, q, q)\)-basis if \( |\Omega_{i,\ell}| \leq 1 \) for all \( i \) and \( \ell \). As noted above, this is equivalent to saying that the rows of \( G_0 \) all have weight \( 2^M \) which is clearly a necessary condition for the minimum distance to be \( 2^M \). Since the last \( M-1 \) rows of \( G \) have weight \( 2q \), another necessary condition for the minimum distance to be \( 2^M \) is \( q \geq 2^{M-1} \). In this case, the set \( \{0, 1, 2, \ldots, 2^{M-4}, 2^{M-3}, 2^{M-2}\} \) is clearly an \((M, q, q)\)-basis for all \( q \geq 2^{M-1} \).

However, it may be of interest to determine \((M, q, q)\)-bases also for \( q < 2^{M-1} \), for example when we consider the code generated by \( G_0 \) or when we consider codes with minimum distance less than \( 2^M \). We will now consider such sets and we show the following result.

The Conway-Guy sequence

\[
a(0) = 0, a(1) = 1, a(2) = 2, a(3) = 4, a(4) = 7, \ldots
\]

is defined recursively by

\[
a(n) = 2a(n-1) - a\left(n - 1 - \left\lfloor \frac{1 + \sqrt{8n - 15}}{2} \right\rfloor \right).
\]

The first few elements in the Conway-Guy sequence are:

\[
0, 1, 2, 4, 7, 13, 24, 44, 84, 161, 309, 594, 1164, 2284, 4484, 8807, 17305, 34301, 68008, 134852, 267420, 530356, 1051905, 2095003, 4172701, 8311101, 16554194, 32973536, 65679652, 130828948, 261127540, 521203175, 104031347, 2076449993.
\]

It is known that \( a(n)/2^n \rightarrow 0.23513 \ldots \) when \( n \rightarrow \infty \), see [15].

Conway and Guy [16] conjectured and Bohman [15] proved the following property of the Conway-Guy sequence: all the \( 2^M \) sums

\[
\sum_{j=0}^{M-1} x_j (a(M) - a(j)), \text{ where } x_j \in \{0, 1\},
\]

are distinct.

Theorem 3: If

\[
q > \left\lceil \sum_{n=M-M/2}^{M-1} a(n) \right\rceil - \left\lceil \sum_{n=1}^{M-2M/2} a(n) \right\rceil, \quad (6)
\]

then \( \{a(0), a(1), a(2), \ldots, a(M-1)\} \) is an \((M, q, q)\)-basis.

Proof: Let \( q \) satisfy (6). Suppose \( |\Omega_{i,\ell}| \geq 2 \) and let \( (x_0, x_1, \ldots, x_{M-1}) \) and \( (y_0, y_1, \ldots, y_{M-1}) \) be distinct sequences in \( \Omega_{i,\ell} \). Let

\[
S_1 = \sum_{j=0}^{M-1} x_j a(j) \text{ and } S_2 = \sum_{j=0}^{M-1} y_j a(j).
\]

Then \( S_1 \equiv S_2 \equiv \ell \pmod q \). Without loss of generality, we may assume that \( S_1 \geq S_2 \). If \( S_1 = S_2 \), then \( \sum_{j=0}^{M-1} x_j (a(M) - a(j)) = a(M) i - S_1 = \sum_{j=0}^{M-1} y_j (a(M) - a(j)) \). By the property of the Conway-Guy sequence quoted above, this implies that \( x_j = y_j \) for \( 0 \leq j \leq M-1 \), a contradiction.

If \( S_1 > S_2 \), then \( S_1 \geq S_2 + q \), and so

\[
q \leq S_1 - S_2 \leq \sum_{j=M-1}^{M-1} a(j) - \sum_{j=0}^{M-1} a(j) \leq \left\lceil \sum_{n=M-M/2}^{M-1} a(n) \right\rceil - \left\lceil \sum_{n=0}^{M-2M/2} a(n) \right\rceil < q, \quad (7)
\]

again a contradiction. Hence \( |\Omega_{i,\ell}| \geq 2 \) is not possible.

In Table II we list the \((M, q, q)\)-bases given by the theorem for \( M \leq 15 \). Note that we use the bound on \( q \) given in the theorem, but also the fact that \( q \) must be a prime. We remark that the bound \( q \geq \left\lceil \sqrt{M/2} \right\rceil \) is always weaker.

V. CONCLUSION

Array LDPC codes have been considered in this paper. The main result found is a general structure for the generator matrix, that is valid for improper, even shortened, codes. The
highly structured shape of such matrix has permitted us to formulate a set of necessary conditions that are useful to bound and maximize the code minimum distance. This is another step towards a full comprehension of the properties of this important, but yet largely unexplored, class of codes.

REFERENCES