

Explicit Construction of Families of LDPC Codes with Girth at Least Six

Jon-Lark Kim

Dept. of Mathematics and Statistics
University of Nebraska - Lincoln
Lincoln, NE 68588-0323
jlkim@math.unl.edu

Uri N. Peled, Irina Perepelitsa, and Vera Pless

Dept. of Mathematics, Statistics, and Computer Science
University of Illinois at Chicago
Chicago, IL 60607-7045
uripeled@uic.edu
{irina, pless}@math.uic.edu

November 11

Abstract

LDPC codes are serious contenders to Turbo codes in terms of decoding performance. One of the main problems is to give an explicit construction of such codes whose Tanner graphs have known girth. For a prime power q and $m \geq 2$, Lazebnik and Ustimenko construct a q -regular bipartite graph $D(m, q)$ on $2q^m$ vertices, which has girth at least $2\lceil m/2 \rceil + 4$. We regard these graphs as Tanner graphs of binary codes $\text{LU}(m, q)$. We can determine all the parameters of $\text{LU}(2, q)$. We know that their girth is 6 and their diameter is 4. We know that $\text{LU}(3, q)$ has girth 8 and diameter 6 and we conjecture its dimension. We find some interesting LDPC codes by our partial row construction.

1 Introduction

Low density parity check (LDPC) codes were originally introduced by Gallager [4]. They have again become interesting because of the success of iterative decoding for Turbo codes. LDPC codes are competitors of these codes in performance of iterative decoding algorithms, as their performance approaches the Shannon limit [9]. Tanner's graphical representation of LDPC codes [10] influenced much of the current literature. Most of these codes are constructed randomly, but explicit constructions are needed for implementation purposes as well as for knowing the properties of these codes. We give such constructions based on constructions of graphs with good girth.

Let $m \geq 2$ be an integer and q a power of a prime. In [7] Lazebnik and Ustimenko construct a family $D(m, q)$ of q -regular bipartite graphs on $2q^m$ vertices, with q^m vertices called *points* and q^m vertices called *lines*. Points and lines are elements of $\text{GF}(q)^m$ and equations are given in [7], which determine incidence of points and lines. If a point is

incident to a line, an edge joins them in $D(m, q)$. Each graph is edge transitive [7]. It is further shown [7] that when m is odd, $D(m, q)$ has girth at least $m + 5$. It also follows from general graph homomorphism results of [8] that the girth of $D(m, q)$ is not less than the girth of $D(m - 1, q)$, so for m even, the girth of $D(m, q)$ is at least $m + 4$. Thus for all m , the girth of $D(m, q)$ is at least $2\lceil m/2 \rceil + 4$. We let $H(m, q)$ be the incidence matrix of lines and points of $D(m, q)$, where rows are indexed by lines and columns are indexed by points, and consider it and its transpose to be parity check matrices of binary codes of length q^m called $LU(m, q)$ codes. In other words, we take $D(m, q)$ to be the Tanner graph [10] of the LDPC code $LU(m, q)$ and investigate the properties of these codes. As the rows as well as the columns of $H(m, q)$ are linearly dependent, the dimensions of these codes need to be determined.

In [7] it is shown that

(*) any two rows (columns) of $H(m, q)$ have a 1 in at most one common column (row).

This implies that the girth of the graph is at least 6.

We show that $D(2, q)$ has girth 6 and diameter 4. We derive the parameters of all $LU(2, q)$. When q is even we obtain Euclidean geometry codes.

We have computed the parameters of $LU(3, q)$ codes through $q = 19$. We prove that $D(3, q)$ has girth 8 (already shown in [11]). This implies that the minimum weight of $LU(3, q)$ is at least $2q$ [10]. For $q \geq 3$ the diameter of $D(3, q)$ is 6 [11]. When q is odd we apparently have a family of codes whose rates approach $1/2$.

We examined some LU codes for $m = 4, 5, 6$ and 7 and we give our observations. We give a lower bound on the minimum weight of $LU(m, q)$ in terms of q and m for odd m using Tanner's bound [10].

We use a new technique, the partial row construction, to obtain codes with larger rate than LU codes but not smaller girth. We give lists of interesting codes found in this way.

Proposition 1 *For every even integer $m \geq 2$, the $LU(m, 2^s)$ codes obtained from $H(m, 2^s)$ and $H(m, 2^s)^T$ are equivalent.*

Proof. It is proven in [7] that for $q = 2^s$, $s \geq 1$, and any even integer $m \geq 2$, the graphs $D(m, q)$ are vertex-transitive. This means that there is a permutation of vertices that interchanges all point and line vertices. Hence, we can obtain $H(m, 2^s)$ from $H(m, 2^s)^T$ by permuting rows and columns, so that the codes are equivalent. \square

2 LU(2,q) Codes

(i) According to [7], in $D(2, q)$ a point (a, b) is on a line $[x, y]$ iff $y = ax + b$ where a, b, x and y are in $\text{GF}(q)$.

We label the rows and columns of $H(2, q)$ with the pairs $[x, y]$ and (a, b) ordered lexicographically under a fixed ordering of $\text{GF}(q)$. If q is a prime, this is the usual ordering; if q is a prime power, we order the elements of $\text{GF}(q)$ in some way, say as powers of a primitive element, with the element 0 first. It can be seen that $H(2, q)$ consists of q^2 $q \times q$ permutation matrices, where each permutation matrix corresponds to a fixed a and a fixed x . If q is a prime, these permutation matrices are circulants. So the first q rows of $H(2, q)$ consist of q permutation matrices, similarly for the next q rows, etc.

We call a *row block* the set of all rows with fixed x , and a *column block* the set of all columns with fixed a .

(ii) No two rows in a row block have a 1 in common, i.e., in the same column. Any two rows from different row blocks have exactly one 1 in common. This follows from (i). Similarly for columns.

Example.

$$H(2,3) = \{H_{ij}\} = \left(\begin{array}{ccc|ccc|ccc} 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \end{array} \right),$$

where i, j run over the index set $\{00, 01, 02, 10, 11, 12, 20, 21, 22\}$.

We also note that the code whose parity check matrix is $H(2,3)^T$ is the same as the one with parity check matrix $H(2,3)$.

Theorem 1 *For $q > 2$, all $D(2, q)$ have girth 6. Also all $D(2, q)$ have diameter 4.*

Proof. By (*), the girth of $D(2, q)$ is at least 6. We show that we can find a cycle of length 6. The first row r_1 of $H(2, q)$ meets the first row r_2 in the second row block in a column c_1 . There is another column c_2 with a one in r_2 . Column c_2 has a one in a unique row r_3 of the third row block. Row r_3 must meet row r_1 , but not in c_1 (or else r_2 and r_3 would meet twice) and not in c_2 (or else r_1 and r_2 would meet twice). So there is a third column c_3 meeting r_1 and r_2 . Then $r_1 - c_1 - r_2 - c_2 - r_3 - c_3 - r_1$ is a cycle of length 6 in $D(2, q)$.

Two rows in different row blocks and two columns in different column blocks have distance 2 from each other. A row and a column have distance 1 or 3, and two rows or columns in the same row or column block have distance 4. Hence the diameter of $D(2, q)$ is 4. \square

Theorem 2 *If q is odd, the two $LU(2, q)$ codes are the same $[q^2, q - 1, 2q]$ code, whose group has order $(q!)^{q+1}$.*

Proof. We construct a canonical spanning set of $LU(2, q)^\perp$. If we add all the rows in any row block of $H(2, q)$, we obtain the all-one vector. If we add up all the rows in $H(2, q)$ that have 1 in a fixed column, we will be adding one row from each row block. If for example the column is the first column, the resulting sum will be

$$\underbrace{10\dots 0}_q \underbrace{11\dots 1}_q \dots \underbrace{11\dots 1}_q.$$

This is so since no two rows in a row block have a 1 in common by (ii) and since q is

odd. Hence $\text{LU}(2, q)^\perp$ contains all the rows of the following matrix,

$$A = \begin{pmatrix} E & 0 & \dots & \dots & 0 \\ 0 & E & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & E \\ 1 & 1 & 1 & \dots & 1 \end{pmatrix},$$

where $E = I + J$ with I the $q \times q$ identity matrix and J the $q \times q$ all-one matrix.

As E has rank $q-1$, A generates a code of dimension $q(q-1)+1$ (the all-one vector of odd weight is not equal to any sum of previous rows, as all such sums have even weight). It is not hard to see that the rows of A span $\text{LU}(2, q)^\perp$, as we can express any row of $H(2, q)$ as a sum of these rows. Hence for q odd, $\dim(\text{LU}(2, q)) = q^2 - (q(q-1)+1) = q-1$.

From the generating set A of $\text{LU}(2, q)^\perp$, we see that the group of this code consists of $\text{Sym}(q)$ operating independently on each column block of q elements, and another $\text{Sym}(q)$ permuting the q column blocks. Hence for q odd, the group of $\text{LU}(2, q)$ has order $(q!)^{q+1}$.

We can also determine the minimum weight of $\text{LU}(2, q)$ by looking at A . The dual of each E is the all-one vector of length q . But as the all-one vector of length q^2 is in $\text{LU}(2, q)^\perp$, every vector in $\text{LU}(2, q)$ has even weight. So the minimum weight of $\text{LU}(2, q)$ is $2q$. $\text{LU}(2, q)$ can be regarded as all even-weight row vectors made out of all-0 and all-1 blocks of length q . As we also get A as above for $H(2, q)^T$, the two $\text{LU}(2, q)$ codes are in fact the same. \square

When q is even, we get interesting results.

Lemma 1 $H(2, 2^s)$ is the incidence matrix of 2^{2s} points and 2^{2s} lines consisting of parallel classes from the affine plane $\text{AG}(2, 2^s)$. Further, the code C generated by $H(2, 2^s)$ contains all the lines of this affine plane

Proof. Each row of $H(2, 2^s)$ has weight 2^s , the weight of a line in an affine geometry from a projective plane $\text{PG}(2, 2^s)$ of order 2^s . We regard these rows as lines of the geometry. By (ii), each row block is a parallel class of lines. There are 2^s such blocks in $H(2, 2^s)$. The affine plane has $2^s + 1$ parallel classes of lines. This last parallel class consists of the 2^s row vectors each of which is the all-one vector in a fixed column block and zero outside the block. We show as follows that these vectors are in C . If we add up all the rows of $H(2, 2^s)$ that have a 1 in their first position, we get $\underbrace{0 \dots 0}_{2^s} \underbrace{1 \dots 1}_{2^s} \dots \underbrace{1 \dots 1}_{2^s}$ by (ii) and since 2^s is even. Adding the all-one vector, which is the sum of the rows in any row block, we get $\underbrace{1 \dots 1}_{2^s} \underbrace{0 \dots 0}_{2^s} \dots \underbrace{0 \dots 0}_{2^s}$, a line in the missing parallel class. We can get the rest of the lines similarly. The fact that this affine plane comes from $\text{PG}(2, 2^s)$ follows from the equations in (i). \square

Theorem 3 $\text{LU}(2, 2^s)$ is a $[2^{2s}, 2^{2s} - 3^s, 2^s + 2]$ code.

Proof. It is known [1] that the incidence matrix of an affine plane of order 2^s generates a $[2^{2s}, 3^s]$ binary code C . The minimum weight vectors of C^\perp contain all the ovals of the corresponding projective plane [6], and as there exist ovals disjoint from the line at ∞ , the minimum weight of $\text{LU}(2, 2^s)$ is $2^s + 2$. Hence $\text{LU}(2, 2^s)$ is a $[2^{2s}, 2^{2s} - 3^s, 2^s + 2]$ code. \square

In [5] families of LDPC codes with girth 6 were constructed from finite geometries. One of these families of Euclidean geometry codes has parameters $[2^{2s} - 1, 2^{2s} - 3^s, 2^s + 1]$. We extended two of these codes for $s = 2$ and $s = 3$ and (using Magma [3]) found that they are equivalent to $LU(2, 4)$ and $LU(2, 8)$. This will be so in general since both families of codes are constructed from $PG(2, 2^s)$. However, the two families could have different decoding performance as the parity check matrices used are different.

3 $LU(3, q)$ Codes

In $D(3, q)$, a point (a, b, c) is incident with a line $[x, y, z]$ iff $y = ax + b$ and $z = ay + c$, where a, b, c, x, y and z are in $GF(q)$ [7].

We investigated the parameters of the $LU(3, q)$ codes for $q = 2$ up to $q = 19$ by Magma. By [7], all the Tanner graphs of the $LU(3, q)$ codes have girth at least $3 + 5 = 8$. We give a simple proof that the girth is exactly 8. By [11], $D(3, q)$ has diameter 6 for $q \geq 3$. $D(3, 2)$ is disconnected; it is a union of two 8-cycles. So $LU(3, 2)$ is the direct sum of two $[4, 1, 4]$ codes, each of which is an $LU(2, 2)$ code.

Theorem 4 ([11]) *$D(3, q)$ has girth 8. Its diameter is 6 if $q > 2$.*

Proof. Since by [7] we know that the girth of $D(3, q)$ is at least 8, finding one 8-cycle shows that the girth is 8. It is not hard to check that $(000) - [000] - (100) - [111] - (011) - [011] - (110) - [-100] - (000)$ is an 8-cycle in $D(3, q)$. $D(3, q)$ has diameter 6 for $q > 2$ [11, Thm. 3.9]. \square

Conjecture If q is odd, then $LU(3, q)$ is a $[q^3, (q^3 - 2q^2 + 3q - 2)/2]$ code.

This is so for all of the $LU(3, q)$ codes for odd q that we constructed. If this is true, then the rate of these codes approach $1/2$ as q gets large. We also noticed that for $q = 3$ and $q = 5$, the two $LU(3, q)$ codes we obtain from $H(3, q)$ and its transpose have different minimum weights. For $q = 4$ the two codes are equivalent by Proposition 1. See Table 1. For $q \geq 7$, we were unable to determine the minimum weight.

Table 1: Parameters of $LU(3, q)$ codes for $q = 3, 4, 5$.

q	3	4	5
H	[27,8,6]	[64,22,8]	[125,44,10]
H^T	[27,8,8]	[64,22,8]	[125,44,20]

When q is even, the rate of the $LU(3, q)$ codes seems higher than when q is odd.

Theorem 5 *$LU(3, q)$ has minimum weight at least $2q$.*

Proof. This follows from [10, Theorem 2] since we know that the girth of $LU(3, q)$ is 8. \square

4 The Partial Row Construction

In investigating the LU codes, we found many that have low rates. We decided to consider those codes whose parity check matrices consist of the first i rows of $H(m, q)$, where $i < q^m$ (we order the rows and columns lexicographically as in Section 2). We call this *the partial row construction*. If we consider a code C whose parity check matrix consists of the first i rows of $H(m, q)$, then the rate of C may stay the same or be higher than that of $\text{LU}(m, q)$, the girth of its Tanner graph may stay the same or go up, but the minimum weight might go down. We found a number of interesting LDPC codes by the partial row construction for $m = 2$, which we list in Table 2.

Table 2: LDPC codes obtained by the partial row construction from $\text{LU}(2, q)$ codes.

q	$[n, k, d]$	(girth, diameter)	# of rows of $H(2, q)$
3	[9,4,4]	(8,4)	6
4	[16,9,4]	(8,4)	8
5	[25,12,6]	(6,4)	14–15
7	[49,24,8]	(6,4)	27–28
8	[64,37,10]	(6,4)	57–64
9	[81,32,16]	(6,4)	53–54
11	[121,84,8]	(6,4)	39

When $q = 8$ and the number of rows is 64, this code is $\text{LU}(2, 8)$. Note that the [9,4,4], [16,9,4] and [64,37,10] codes are optimal, whereas the [25,12,6] and [81,32,16] codes are just 2 short of being optimal [2]. The other two codes have minimum weight 4 less than the optimal codes. The parity check matrix for the [9,4,4] code consists of the first 6 rows of $H(2, 3)$ given in the example in Section 2.

We also improve the rate while maintaining the minimum weight, girth and diameter for $\text{LU}(2, q)$ codes by the partial row construction. We list them below.

old	[25,4,10]	[49,6,14]	[81,8,18]
new	[25,6,10]	[49,10,14]	[81,14,18]

We obtain interesting LDPC codes from $\text{LU}(3, q)$ codes by the partial row construction. They are listed in Table 3. Many have larger girths than the $\text{LU}(3, q)$ code. We list only those where we were able to find the minimum distance.

Table 3: Codes from $\text{LU}(3, q)$ codes by the partial row construction.

q	$[n, k, d]$	(girth, diameter)	$H(3, q)$ or $H^T(3, q)$	# of rows
3	[27,12,4]	(16,10)	H	15
3	[27,10,6]	(12,8)	H	18
4	[64,35,4]	(8,10)	H^T	33
5	[125,54,14]	(8,6)	H^T	85
5	[125,47,20]	(8,6)	H^T	105

5 The Cases $m = 4, 5, 6, 7$

The equations for $D(m, q)$ for $m = 4, 5, 6, 7$ are considerably more complicated than for $m = 2, 3$. They can be found in [7].

$D(m, 2)$ is disconnected for $m = 3, 4, 5, 6$ and 7 . In [7], the authors state that they and A.J. Woldar proved that for $m \geq 6$, all $D(m, q)$ are disconnected. In fact, in [11, pg. 79] it is shown that for $q = 3$ and for $q > 4$, $D(m, q)$ has q^{t-1} connected components, where $t = \lfloor \frac{m+2}{4} \rfloor$, and for $m \geq 4$ $D(m, 4)$ has 4^t connected components. So even though the graphs $D(m, q)$ have large girth (at least $2\lceil m/2 \rceil + 4$), the large length of the code and the disconnectedness makes them more difficult to use as Tanner graphs of LDPC codes. We do know the following.

Theorem 6 *When $D(m, q)$ is disconnected, it is a union of isomorphic connected subgraphs. In this case $LU(m, q)$ is a direct sum of equivalent codes each of which has its parity check matrix from the incidence matrix of a connected component subgraph.*

Proof. When $D(m, q)$ is disconnected, it is a union of isomorphic connected subgraphs since the group of $D(m, q)$ is edge-transitive [7]. This is so because if an automorphism of a graph maps an edge e into an edge f , then it maps the connected component of e onto the connected component of f . So if for every two edges of $D(m, q)$ there is an automorphism mapping them into each other, then for every two connected components there is an automorphism mapping them onto each other.

We can reorder the rows and the columns of $H(m, q)$ by putting the rows and the columns of the first connected component first, the rows and the columns of the second connected component second, etc. From this we can see that $LU(m, q)$ is a direct sum of codes. Codes corresponding to distinct connected components are equivalent, since the connected components are isomorphic. \square

We found directly that $LU(4, 4)$, a $[256, 88, 8]$ code, is a direct sum of four $[64, 22, 8]$ codes of girth 8 and diameter 6; and that $LU(5, 4)$, a $[1024, 216]$ code, is a direct sum of four $[256, 54]$ codes of girth 10 and diameter 8.

Since we have a lower bound of $2\lceil m/2 \rceil + 4$ on the girth, a lower bound on the minimum distance can be obtained.

Theorem 7 *The minimum distance d of $LU(m, q)$ satisfies*

$$d \geq \begin{cases} 2 \frac{(q-1)^{\lfloor m/4 \rfloor + 1} - 1}{q-2}, & m \equiv 0 \pmod{4} \\ 2 \frac{(q-1)^{\lfloor m/4 \rfloor + 2} - 1}{q-2}, & m \equiv 3 \pmod{4} \\ 2 \frac{(q-1)^{\lfloor m/4 \rfloor + 1} - 1}{q-2} + \frac{2}{q} (q-1)^{\lfloor m/4 \rfloor + 1}, & m \equiv 1, 2 \pmod{4}. \end{cases}$$

When $q = 2$, the fraction $\frac{(q-1)^{\lfloor m/4 \rfloor + 1} - 1}{q-2}$ is understood to be $\lfloor m/4 \rfloor + 1$, and $\frac{(q-1)^{\lfloor m/4 \rfloor + 2} - 1}{q-2}$ to be $\lfloor m/4 \rfloor + 2$.

Proof. This follows from the proof of [10, Theorem 2], using the fact that the column sums of $H(m, q)$ are q . \square

In particular, $d \geq 2q$ for $m = 3, 4$; $d \geq 4q - 3$ for $m = 5, 6$; $d \geq 2(q^2 - q + 1)$ for $m = 7, 8$; and $d \geq 4(q - 1)^2 + 4$ for $m = 9, 10$ and $q > 2$.

Acknowledgments

We thank Keith Mellinger for calling our attention to [11].

References

- [1] E. F. Assmus Jr and J. D. Key, *Designs and Their Codes*, Cambridge University Press, Cambridge, 1992.
- [2] A. E. Brouwer, "Bounds on the size of linear codes," in *Handbook of Coding Theory*, V. S. Pless and W. C. Huffman, Editors, Elsevier, pp. 295–452, 1998.
- [3] J. Cannon and C. Playoust, *An Introduction to Magma*, University of Sydney, Sydney, Australia, 1994.
- [4] R. G. Gallager, "Low density parity check codes," *IRE Trans. Infom. Theory*, vol. IT-8, pp.21-28, Jan. 1962
- [5] Y. Kuo, S. Lin and M. P. C. Fossorier, "Low-density parity-check codes based on finite geometries: a rediscovery and new results," *IEEE Trans. Inform. Theory*, vol. 47, no. 7, pp. 2711–2736, Nov. 2001.
- [6] E. S. Lander, *Symmetric Designs: An Algebraic Approach*, Cambridge University Press, Cambridge, 1983.
- [7] F. Lazebnik and V. A. Ustimenko, "Explicit construction of graphs with arbitrary large girth and of large size," *Discrete Applied Math.*, vol. 60, pp. 275–284, 1997.
- [8] F. Lazebnik and A. J. Woldar, "General properties of some families of graphs defined by systems of equations," *J. Graph Theory*, vol. 38, pp. 65–86, 2001.
- [9] D. J. C. MacKay and R. M. Neal, "Near Shannon limit performance of low density parity check codes," *Electron. Lett.*, vol. 32, no. 18, pp. 1645-1646, 1996
- [10] R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. IT - 27, pp. 533–547, 1981.
- [11] R. Viglione, "Properties of some algebraically defined graphs," *Ph.D. Thesis*, Univ. of Delaware, 2002.