Des. Codes Cryptogr. (2010) 54:287–297
DOI 10.1007/s10623-009-9324-9

Some low-density parity-check codes derived from finite geometries

Peter Vandendriessche

Received: 1 May 2009 / Revised: 20 August 2009 / Accepted: 21 August 2009 /
Published online: 3 September 2009

Abstract We look at low-density parity-check codes over a finite field $\mathbb{K}$ associated with finite geometries $T_2^*(\mathcal{K})$, where $\mathcal{K}$ is a sufficiently large k-arc in PG$(2, q)$, with $q = p^h$. The code words of minimum weight are known. With exception of some choices of the characteristic of $\mathbb{K}$ we compute the dimension of the code and show that the code is generated completely by its code words of minimum weight.

Keywords LDPC · Error rate · Dimension · Linear representation · Generalized quadrangle · $LU(3, q)$

Mathematics Subject Classification (2000) 05B25 · 05C38 · 05C50 · 51E12 · 51E21 · 51E22 · 94B27

1 Introduction

The concept of low-density parity check (LDPC) codes was originally introduced by Gallager [5]. In its broader sense, a binary LDPC code C is a linear block code defined by a sparse parity check matrix H, a matrix that contains a lot more 0s than 1s.

When LDPC codes are decoded using Gallager's decoding method, their empirical performance is known to be excellent, even near the Shannon limit [16, 17]. Early known LDPC codes have been constructed randomly [5, 16], but lately several types of explicit constructions have been made. These have been based on permutation matrices [4, 26], Ramanujan graphs [19, 22], expander graphs [25] or on q-regular bipartite graphs [13].

In 2001, Kou et al. [14] studied classes of LDPC codes defined by incidence structures in finite geometries. Since then, many LDPC codes have been constructed based on various

Communicated by J.W.P. Hirschfeld.

P. Vandendriessche (✉)
Boudewijn Hapkenstraat 5, 8820 Torhout, Belgium
e-mail: Peter.Vandendriessche@UGent.be

incidence structures in discrete mathematics and finite geometry [8–11, 18, 27]. In particular, Vontobel and Tanner [27] considered the LDPC codes generated by generalized polygons, focussing on generalized quadrangles. They demonstrated that some generalized quadrangle LDPC codes perform well under the sum product algorithm from [16]. Later, simulation results of Liu and Pados [15] showed that several generalized polygon LDPC codes have powerful bit-error-rate performance when decoding is carried out via low-complexity variants of belief propagation [15]. It would be interesting to perform the same simulations for the incidence geometries studied in this article. This is because all handled structures have a girth of at least eight in their associated Tanner graph. The latter is because $\mathcal{K}$ is an arc, hence the geometry contains no triangles.

Two structures that have received a lot of attention lately are $LU(3,q)$ and its dual $LU(3,q)^D$ [12, 13, 24]. In [13] the authors conjecture the dimension of the associated code to be

$$\frac{q^3-2q^2+3q-2}{2}$$

when q is odd. Over the field $GF(2)$ this was proven in [24]. This result will be extended to all fields with char $\mathbb{K} \neq p$ and for all q, for both $LU(3,q)$ and $LU(3,q)^D$. In [12] the authors classify the code words of small weight of $LU(3,q)^D$ as linear combinations of code words of minimum weight. This result will be extended to all code words, regardless of the weight, when char $\mathbb{K} \neq p$.

Another structure that recently received a lot of attention is $T_2^*(\mathcal{K})$ with $\mathcal{K}$ a hyperoval. This structure is a generalized quadrangle and codes associated to generalized quadrangles have been studied thoroughly. For regular generalized quadrangles it has been proven in [1] that, excluding some choices of the characteristic of $\mathbb{K}$, the code is generated by its code words of minimum weight. For this generalized quadrangle the condition is char $\mathbb{K} \neq 2$. Under the same condition char $\mathbb{K} \neq 2$, the same result will be established for this nonregular generalized quadrangle and its dimension will be computed. In [21], the authors manage to classify the code words of small weight for sufficiently large q, as a linear combination of code words of minimum weight but in some cases also second-minimum weight. For char $\mathbb{K} \neq 2$ we show that all words in the code are a linear combination of the code words of minimum weight, even the code words of second minimum weight used in the classification of [21], regardless of weight and even for small q.

2 Preliminaries

Let PG$(3,q)$ be the 3-dimensional projective space over $GF(q)$. Let PG$(2,q)$ be a plane in it and let $\mathcal{K}$ be a set of points in that plane with no three points collinear; this is called an arc. Define the geometry $T_2^*(\mathcal{K})$ as follows: the points are the affine points, being the points of PG$(3,q)\backslash$PG$(2,q)$. The lines are the affine lines of PG$(3,q)$ which go through a point of $\mathcal{K}$. The incidence is inherited from PG$(3,q)$. Note that through every point there are k lines, one through each point of $\mathcal{K}$, while every line contains q points. In total there are q^3 points and kq^2 lines: q^2 through each point of $\mathcal{K}$.

Let $\mathcal{K} = \{r_1, \ldots, r_k\}$ be a k-arc and denote by $\mathcal{P} = \{p_1, \ldots, p_{q^3}\}$ the points and by $\mathcal{L}$ the lines of $T_2^*(\mathcal{K})$. Let H be its $q^3 \times kq^2$ incidence matrix, where points are rows and lines are columns. Let C be the linear code with H as its parity check matrix, over an arbitrary field $\mathbb{K}$.

One can associate a 'weight' to each line in a codeword w, being its value at the position corresponding to that line. A word $w \in \mathbb{K}^{kq^2}$ is in C if and only if $w \cdot H^{\mathrm{T}} = \bar{0}$, hence if and only if

$$\sum_{i=1}^{n} c_i H_{ji} = 0$$

as an element of $\mathbb{K}$, for every point p_j. Alternatively formulated: a word is a code word of C if and only if the sum of the weights of the lines through every point equals 0 over $\mathbb{K}$.

Definition 2.1 Let $r_i, r_j \in \mathcal{K}$ with $i < j$ and let π be a projective plane through r_i, r_j different from the fixed PG$(2, q)$. The 'plane word through r_i and r_j' is the codeword with

- $+1$ on the positions corresponding to the lines of π through r_i,
- -1 on the positions corresponding to the lines of π through r_j,
- 0 on all other positions.

Remark 2.2 It has been proven in [21] that the plane words are exactly the code words of minimum weight, up to a scalar factor.

Notation 2.3 Given a plane word w through p_i and p_j, define $T(w)$ to be the plane π in the definition above and by $L(w)$ the line $p_i p_j$ in Definition 2.1. Denote by C' the code generated by all plane words.

Remark 2.4 There are no triangles in this geometry. If there was a triangle, these three lines would be coplanar; hence their points on $\mathcal{K}$ must lie on the same line. This contradicts the fact that $\mathcal{K}$ is an arc.

3 Dimension of C'

Remark 3.1 Up to a scalar factor, there are exactly $\frac{qk(k-1)}{2}$ plane words.

Proof A plane word is determined up to a scalar factor by its corresponding plane. Through every two points of $\mathcal{K}$ there are exactly q planes and other planes do not allow plane words. This yields $q \binom{k}{2}$ as required. □

Let $\mathcal{K}$ be any k-arc and let $\mathbb{K}$ be an arbitrary field with char $\mathbb{K} \neq p$.

Lemma 3.2 *Assume that* $\sum \lambda_i w_i = \overrightarrow{0}$ *with* w_i *plane words and* $\lambda_i \in \mathbb{K}$. *Whenever* $L(w_i) = L(w_j)$, *then* $\lambda_i = \lambda_j$.

Proof Without loss of generality, by renumbering the indices as necessary, let $L(w_j) = L(w_i) = \langle p_{k-1}, p_k \rangle$. Define $L_i := \langle p_i, p_k \rangle$ and denote the weights of the q plane words through every L_i as $a_{i,1}, a_{i,2}, \ldots, a_{i,q}$. Denote by $\left(\Pi_j\right)_{j=1,\ldots,q}$ the planes through $\langle p_{k-1}, p_k \rangle$ and by $\left(\Pi_{j,t}\right)_{t=1,\ldots,q}$ the q lines trough p_k in the jth plane.

When a linear combination yields the zero word, this means that every line in the geometry has weight 0, in particular every line in Π_j through p_k. Hence for line $\Pi_{j,t}$ one has

$$a_{1,x_t} + a_{2,x_t} + \cdots + a_{k-1,x_t} = 0$$

with x_t a t-dependent index permutation.

For a given Π_j all a_{k-1,x_t} are equal, since they come from the same plane word. Hence we simply write $a_{k-1,j}$. By definition of $a_{i,l}$, every $a_{i,l}$ has to appear exactly once. This means that there are permutations $\sigma_{j,i}$ on $\{1, \ldots, q\}$ such that

$$\begin{cases} a_{1,\sigma_{j,1}(1)} + a_{2,\sigma_{j,2}(1)} + \cdots + a_{k-2,\sigma_{j,k-2}(1)} + a_{k-1,j} = 0 \\ a_{1,\sigma_{j,1}(2)} + a_{2,\sigma_{j,2}(2)} + \cdots + a_{k-2,\sigma_{j,k-2}(2)} + a_{k-1,j} = 0 \\ \vdots \\ a_{1,\sigma_{j,1}(q)} + a_{2,\sigma_{j,2}(q)} + \cdots + a_{k-2,\sigma_{j,k-2}(q)} + a_{k-1,j} = 0 \end{cases}$$

Now sum the rows above. Since summing over a permutation of a set is the same as summing over the set, this yields

$$\sum_{t=1}^{q}\sum_{i=1}^{k-2} a_{i,t} = \sum_{t=1}^{q}\sum_{i=1}^{k-2} a_{i,\sigma_{j,i}(t)} = -qa_{k-1,j}$$

for every j. Since $q \neq 0$, all $a_{k-1,j}$ are the same, regardless of j. □

Theorem 3.3 *The dimension of C' is $\frac{(k-1)(qk-k+2)}{2}$.*

Proof Let $S = \{(p_1, p_2)\}$ and let $d(S)$ be the dimension of the code generated by the plane words $\{w | T(w) \in S\}$. One by one, add the pairs

$$(p_1, p_3), (p_2, p_3), (p_1, p_4), (p_2, p_4), (p_3, p_4), (p_1, p_5), \ldots, (p_{k-1}, p_k)$$

to S. Every time a pair of the form (p_1, p_ℓ) is added, a new point is added, hence the dimension increases by q. Every time another pair is added, say (p_i, p_j) with $i > 1$, the dimension increases by:

- at least $q - 1$ because of Lemma 3.2,
- by strictly less than q since one can easily write the zero word as a linear combination of plane words through $(p_1, p_i), (p_1, p_j), (p_i, p_j)$.

Therefore, the dimension is exactly

$$q + \sum_{i=3}^{k} q + \sum_{i=3}^{k} (i-2)(q-1) = \frac{(k-1)(qk-k+2)}{2}$$

as claimed. □

4 Dimension of C

We start with some preliminaries from algebraic graph theory. Then we compute the dimension of C and find out that it is the same as the dimension of C'. Hence, since $C' \leq C$, C is generated completely by its code words of minimum weight, here the plane words. First this is done for $\mathcal{K}$ a hyperoval, in this case $T_2^*(\mathcal{K})$ is a $(q-1, q+1)$-generalized quadrangle [20]. Then is done for the case that $\mathcal{K}$ is a conic (q odd). Finally we show that these two cases are sufficient to prove $C = C'$ for any sufficiently large arc. We also compare with earlier known results.

4.1 Preliminaries

Lemma 4.1 *Let A be an adjacency matrix of a graph. Then $(A^k)_{ij}$ is the number of paths of length k from vertex i to vertex j.*

Proof This is lemma 2.5 in [2]. □

Lemma 4.2 *Let A be the adjacency matrix of a connected d-regular graph. Then d is an eigenvalue of A with real multiplicity* 1 *and corresponding eigenvector* $(1, 1, \ldots, 1)^{\mathrm{T}}$.

Proof This is Proposition 3.1 in [2]. □

Now, look at the matrix HH^{T} and note that

$$(HH^{\mathrm{T}})_{ij} = \begin{cases} k & \text{if } i = j \\ 1 & \text{if } p_i \text{ and } p_j \text{ are collinear}, \\ 0 & \text{otherwise} \end{cases}$$

hence the point adjacency matrix of the graph of $T_2^*(\mathcal{K})$ is $A := HH^{\mathrm{T}} - kI$. We will study the eigenvalues of A.

A codeword of C fulfils $cH^{\mathrm{T}} = 0$, hence $Hc^{\mathrm{T}} = 0$ and hence $H^{\mathrm{T}}Hc^{\mathrm{T}} = 0$. This means that c^{T} is a right eigenvector of $H^{\mathrm{T}}H$ with eigenvalue zero. Hence the dimension of the code cannot be larger than the dimension of the eigenspace corresponding to the eigenvalue 0. This is at most the algebraic multiplicity of the eigenvalue 0 over $\mathbb{K}$.

Now, from linear algebra it is known that the dimension of the null space of a matrix A is at most the number of zero eigenvalues of $H^{\mathrm{T}}H$. It is also known that, except for maybe zero, $H^{\mathrm{T}}H$ and HH^{T} have the same set of eigenvalues.

4.2 The case $\mathcal{K}$ is a hyperoval ($q = 2^h$)

Here $k = q + 2$, hence $\dim C' = \frac{q(q+1)^2}{2}$. We will now compute $\dim C$.

It may be interesting to note that $T_2^*(\mathcal{K})$ is a $(q-1, q+1)$-generalized quadrangle [20]. It is known from [1] that the code associated to a regular generalized quadrangle is generated by its code words of smallest weight under certain restrictions on the characteristic. For $T_2^*(\mathcal{K})$ this condition is $2q^2 \neq 0$, which would follow from our original assumptions. However, this generalized quadrangle $T_2^*(\mathcal{K})$ is not regular. The following indirectly shows that the regularity is not a necessary condition for the theorem of [1].

Lemma 4.3 *The matrices HH^{T} and $H^{\mathrm{T}}H$ only has eigenvalues* 0, $2q$ *and* $q(q+2)$.

Proof Let us first look at the eigenvalues of $A := HH^{\mathrm{T}} - (q+2)I$. Since there are only 3 projectively non-equivalent pairs of points in $T_2^*(\mathcal{K})$ (equal, collinear, non-collinear) the number of paths between two points only depends on whether they are equal, collinear or non-collinear. Hence Lemma 4.1 implies that each of the matrices I, A, A^2, A^3 has at most 3 different entries. The entries only depend on whether the two points are equal, collinear or non-collinear. Hence, the matrices I, A, A^2, A^3 must be linearly dependent: $c_3A^3 + c_2A^2 + c_1A + c_0I = 0$ for some $c_0, \ldots, c_3$. Therefore A has a minimal polynomial of degree at most 3 and hence it has at most 3 different eigenvalues.

As

$$\rho(H) = q^2(q+2) - \dim C \leq q^2(q+2) - \dim C' < q^3,$$

HH^{T} is singular and hence has an eigenvalue 0. Therefore A has an eigenvalue $-q-2$. Also, by Lemma 4.2, $(q-1)(q+2)$ is an eigenvalue of A. Now consider Lemma 4.1 with two identical points. These correspond to the diagonal entries of A^n; these are 1,0,$(q+2)(q-1)$ and $(q+2)(q-1)(q-2)$ for $n = 0, 1, 2, 3$.

Since A has at most 3 eigenvalues, there is a polynomial $f(x) = x^3 + ax^2 + bx + c$ with $f(A) = 0$ and $f(\lambda) = 0$ for all eigenvalues λ of A. Hence, one can find the coefficients a, b, c, d by solving the system of equations:

$$\begin{cases} f((q+2)(q-1)) = 0, \\ f(-q-2) = 0, \\ (q+2)(q-1)(q-2) + a(q+2)(q-1) + c = 0, \end{cases}$$

where the third equation represents the diagonal entries of the matrix equation

$$A^3 + aA^2 + bA + cI = 0.$$

This results in $f(x) = (x+2-q)(x+2+q)(x-q^2+2-q)$.

Hence A has eigenvalues $-q-2, q-2, (q-1)(q+2)$ and hence HH^{T} has eigenvalues $0, 2q$ and $q(q+2)$. □

Theorem 4.4 *The algebraic multiplicities of the eigenvalues* $0, 2q, q(q+2)$ *are* $\frac{q(q+1)^2}{2}$, $\frac{(q+2)(q^2-1)}{2}$, 1.

Proof The last multiplicity follows from Lemma 4.2. Denote by μ_1, μ_2 the other eigenvalues and expand the characteristic polynomial $x^{\mu_1}(x-2q)^{\mu_2}(x-q(q+2))$. Then linear algebra tells us that the degree of the polynomial is the sum of the multiplicities, yielding

$$\mu_1 + \mu_2 + 1 = q^2(q+2).$$

The coefficient of x^{n-1} is equal to

$$-\sum_{i=1}^{q^2(q+2)} (H^{\mathrm{T}}H)_{ii} = -\sum_{i=1}^{q^3+2q^2} q = -q(q^3+2q^2).$$

Solving the system

$$\begin{cases} \mu_1 + \mu_2 + 1 = q^3 + 2q^2 \\ -(2q\mu_2 + q(q+2)) = -q(q^3+2q^2) \end{cases}$$

in a computer algebra package gives the above values of μ_1, μ_2. □

If $q+2 \neq 0$ then $\dim C \leq \frac{q(q+1)^2}{2}$ since $2q \neq 0$ by assumption. Theorem 3.3 now becomes $\dim C' = \frac{q(q+1)^2}{2}$. Since $C' \leq C$ this yields

$$\frac{q(q+1)^2}{2} \geq \dim C \geq \dim C' = \frac{q(q+1)^2}{2};$$

hence $C = C'$ and C is generated completely by its plane words.

If $q+2 = 0$, the same conclusion holds. All plane words correspond to the eigenvalue 0 over a field of characteristic 0, the only words that possibly need to be added for a basis of C are the eigenvectors corresponding to an eigenvalue 0 over $\mathbb{K}$ which is not 0 over $\mathbb{R}$. In this case this is only $(1, 1, \ldots, 1)^{\mathrm{T}}$. But this vector is a linear combination of plane words: consider the sum of all plane words through one of $L_1, \ldots, L_k$ having weight -1 for the lines through p_k. All lines not through p_k have weight 1. All lines through p_k have weight $-(q+1) = 1-(q+2) = 1$ since $q+2 = 0$ over $\mathbb{K}$. Hence $(1, \ldots, 1) \in C'$, which means $C = C'$ and C is generated completely by its plane words.

4.3 The case $\mathcal{K}$ is a conic, q odd

For $\mathcal{K}$ a conic, every pair of equal or collinear points are projectively equivalent, but for non-collinear points there are two types: the line they span in PG$(3, q)$ can intersect the plane containing $\mathcal{K}$ at an internal or external point of the conic $\mathcal{K}$. They are called pairs of internal and external points. Every two pairs of internal points are projectively equivalent and every two pairs of external points are projectively equivalent.

Lemma 4.5 *The matrices HH^{T} and $H^{\mathrm{T}}H$ only have eigenvalues 0, q, $2q$ and $q(q+1)$.*

Proof Let us first look at the eigenvalues of $A := HH^{\mathrm{T}} - (q+1)I$. There are only 4 projectively non-equivalent pairs of points in $T_2^*(\mathcal{K})$: equal, collinear, non-collinear internal, non-collinear external. Hence, each of the matrices I, A, A^2, A^3, A^4 has at most 4 different entries and the entry only depends on the type of the corresponding pair of points. Hence they must be linearly dependent: $c_4A^4+c_3A^3+c_2A^2+c_1A+c_0I = 0$ for some $c_0, \ldots, c_4$. Hence A has a minimal polynomial of degree at most 4 and A has at most 4 different eigenvalues.

As

$$\rho(H) = q^2(q+1) - \dim C \leq q^2(q+1) - \dim C' < q^3,$$

HH^{T} is singular and hence has an eigenvalue 0. This means A has an eigenvalue $-q-1$. Also, by Lemma 4.2, q^2-1 is an eigenvalue of A. Now, consider Lemma 4.1 with twice the same point and two collinear points. The corresponding entries in A^k are:

- $k = 0$: 1 and 0,
- $k = 1$: 0 and 1,
- $k = 2$: $(q+1)(q-1)$ and $q-2$ as $T_2^*(\mathcal{K})$ contains no triangles,
- $k = 3$: $(q+1)(q-1)(q-2)$ and q^3+q^2-4q+3,
- $k = 4$: $(q+1)(q-1)(q^3+q^2-4q+3)$ and $q^5-4q^3-4q^2+10q-4$.

Since A has at most 4 eigenvalues, there is a polynomial $f(x) = x^4+ax^3+bx^2+cx+d$ with $f(A) = 0$ and $f(\lambda) = 0$ for all eigenvalues λ of A. Hence, one can find the coefficients a, b, c, d by solving the system of equations:

$$\begin{cases} f(q^2-1) = 0, \\ f(-q-1) = 0, \\ (q+1)(q-1)(q^3+q^2-4q+3) + a(q+1)(q-1)(q-2) + b(q+1)(q-1) + d = 0, \\ (q^5-4q^3-4q^2+10q-4) + a(q^3+q^2-4q+3) + b(q-2) + c = 0. \end{cases}$$

where the last two equations represent the diagonal entries and collinear points-entries in

$$A^4 + aA^3 + bA^2 + cA + dI = 0.$$

Solving this system results in $f(x) = (x+q+1)(x+1)(x-q+1)(x-q^2+1)$. Hence A has eigenvalues $-q-1, -1, q-1, q^2-1$ and by definition of A this proves the lemma. □

Theorem 4.6 *The multiplicities of $q(q+1), 0, q, 2q$ are $1, \frac{q(q^2+1)}{2}, q^2-1, \frac{q(q^2-1)}{2}$.*

Proof The first multiplicity follows from Lemma 4.2. Denote by μ_1, μ_2, μ_3 the other multiplicities and expand the characteristic polynomial $x^{\mu_1}(x-q)^{\mu_2}(x-2q)^{\mu_3}(x-q^2-q)$. Then linear algebra tells us that:

- that the degree of the polynomial is the sum of the multiplicities:

$$\mu_1 + \mu_2 + \mu_3 + 1 = q^3 + q^2.$$

- that the coefficient of x^{n-1} is equal to

$$-\sum_{i=1}^{q^3+q^2}(H^{\mathrm{T}}H)_{ii} = -\sum_{i=1}^{q^3+q^2} q = -q(q^3+q^2)$$

- that the coefficient of x^{n-2} of the characteristic polynomial of any matrix

$$B = (b_{ij})_{i=1,\ldots,n;\ j=1,\ldots,n}$$

equals the sum of the 2×2 diagonal subdeterminants

$$\sum_{i\neq j}\begin{vmatrix} b_{ii} & b_{ij} \\ b_{ji} & b_{jj} \end{vmatrix}.$$

In this case each subdeterminant equals q^2-1 if $b_{ij}=1$ and q^2 if $b_{ij}=0$. Hence, this coefficient equals $\binom{q^3+q^2}{2}q^2 - q^3\binom{q+1}{2}$.

Since the degree of $x^{\mu_1}(x-q)^{\mu_2}(x-2q)^{\mu_3}(x-q^2-q)$ is $\mu_1+\mu_2+\mu_3+1$ and the coefficients of x^{n-1} and x^{n-2} can formally be written as

$$-(q^2+q+2\mu_3 q+\mu_2 q)$$

and

$$q^2\binom{\mu_2}{2} + (2q)^2\binom{\mu_3}{2} + q(2q)\mu_2\mu_3 + (q^3+q^2)(\mu_2+2\mu_3),$$

yielding the system of equations

$$\begin{cases} \mu_1+\mu_2+\mu_3+1 = q^3+q^2, \\ -(q^2+q+2\mu_3 q+\mu_2 q) = -q(q^3+q^2), \\ q^2\binom{\mu_2}{2} + (2q)^2\binom{\mu_3}{2} + q(2q)\mu_2\mu_3 + (q^3+q^2)(\mu_2+2\mu_3) = \binom{q^3+q^2}{2}q^2 - q^3\binom{q+1}{2}. \end{cases}$$

Solving this system in a computer algebra package gives the values of μ_1, μ_2, μ_3 as suggested. □

If $q+1 \neq 0$ then $\dim C \leq \frac{q(q^2+1)}{2}$, since $q, 2q \neq 0$ over $\mathbb{K}$ by assumption. Theorem 3.3 becomes $\dim C' = \frac{q(q^2+1)}{2}$. Since $C' \leq C$ this yields

$$\frac{q(q^2+1)}{2} \geq \dim C \geq \dim C' = \frac{q(q^2+1)}{2};$$

hence $C = C'$ and C is generated completely by its plane words.

If $q+1=0$ over $\mathbb{K}$ the conclusion still holds as $(1,\ldots,1) \in C'$ by the same argument as in the previous subsection.

4.4 The case $\mathcal{K}$ is a sufficiently large arc

The main idea here is as follows: if one removes a point from $\mathcal{K}$, the property that the code is spanned by its minimum weight code words remains valid. This is shown in the following theorem.

Theorem 4.7 *If the code associated with $T_2^*(\mathcal{K}\cup\{r\})$ is spanned by its minimum weight code words, then so is $T_2^*(\mathcal{K})$.*

Proof In the proof of Theorem 3.3 we saw that adding an ith point to an $(i-1)$-arc increases $\dim C'$ with exactly $q+(i-2)(q-1)$ and hence $\dim C$ with at least $q+(i-2)(q-1)$. This means that if one removes a point, $\dim C$ decreases with at least $q+(i-2)(q-1)$.

Remove each of the k points one by one, except for the last one. This yields that the dimension is now 0 and has decreased by at least

$$\sum_{i=2}^{k}(q+(i-2)(q-1)) = \frac{(k-1)(qk-k+2)}{2} = \dim C' = \dim C.$$

Hence, in each step the dimension has decreased by exactly that much. Hence, in each step $\dim C = \dim C'$ is maintained. □

Now we want to examine the arcs that can be obtained by removing points from a hyperoval or a conic, leading to the theory of (in)complete arcs.

Remark If q is even, every arc has at most $q+2$ points and the arcs with exactly $q+2$ points are called hyperovals. If q is odd, every arc has at most $q+1$ points. Both of these are called maximum arcs.

Proof See Ref. [3].

Theorem 4.8 *If q is odd, every $(q+1)$-arc is a conic.*

Proof See Ref. [23].

Theorem 4.9 *Denote by $m'(2,q)$ the size of the largest arc which is not contained in a maximum arc of $PG(2,q)$ and let $q=p^h$ with p prime. Then Tables 1 and 2 provide upper bounds on $m'(2,q)$.*

Proof See Ref. [7] for an overview of results on this topic.

Remark Note that Table 2 does not cover $q=2,4$, since in those cases every arc in PG$(2,q)$ is contained in a $(q+2)$-arc. For $q=2$ this is trivial and for $q=4$ this follows from Lemma 9.2.1 in [6].

The following theorem summarizes the results obtained so far.

Table 1 Upper bounds on $m'(2,q)$, q odd

$q=p^h$	Upper bound on m'(2,q)
$q=p^{2e}, e\geq 1$	$m'(2,q)\leq q-\sqrt{q}/4+25/16$
$q=p^{2e+1}, e\geq 1$	$m'(2,q)\leq q-\sqrt{pq}/4+29p/16+1$
$q=p$	$m'(2,q)\leq 44q/45+8/9$
$q=p^h, p\geq 5$	$m'(2,q)\leq q-\sqrt{q}/2+5$
$q=p^h, q\geq 23^2, q\neq 5^5, 3^6$, h even if $p=3$	$m'(2,q)\leq q-\sqrt{q}/2+3$

Any larger arc is contained in a $(q+1)$-arc

Table 2 Upper bounds on $m'(2,q)$, q even

$q=p^h$	Upper bound on m'(2,q)
$q=2^{2e}, e>1$	$m'(2,q)=q-\sqrt{q}+1$
$q=2^{2e+1}, e\geq 1$	$m'(2,q)\leq q-\sqrt{2q}+2$

Any larger arc is contained in a $(q+2)$-arc

Theorem 4.10 *If $\mathcal{K}$ is a k-arc that can be extended to a maximum arc, then the code associated with $T_2^*(\mathcal{K})$ has dimension $\frac{(k-1)(qk-k+2)}{2}$ over any field $\mathbb{K}$ with char $\mathbb{K} \neq 2, p$.*

Now we will eliminate the restriction char $\mathbb{K} \neq 2$ as claimed. When q is even, this is trivial as $2 = p$. If q is odd, then every arc that is contained in a conic is either equal to a conic, in which case char $\mathbb{K} \neq 2$ is not claimed, or it is contained in a conic minus one point.

A conic minus one point is exactly the arc used in the construction of $LU(3, q)^D$. Since $LU(3, q)$ has a square incidence matrix, $LU(3, q)$ and $LU(3, q)^D$ have the same dimension. Hence

$$\dim C = \frac{q^3 - 2q^2 + 3q - 2}{2}$$

from [24]*, since the binary dimension is the same as over any other finite field of characteristic* 2*. From Theorem 3.3 it follows that this is also equal to* $\dim C'$*, hence* $\dim C = \dim C'$*. The rest of the reasoning works perfectly when char* $\mathbb{K} = 2$*.*

References

1. Bagchi B., Sastry N.S.N.: Codes associated with generalized polygons. Geometriae Dedicata **27**, 1–8 (1988).
2. Biggs N.: Algebraic graph theory, Cambridge tracts in mathematics 67. Cambridge University Press, London (1974).
3. Bose R.C.: Mathematical theory of the symmetric factorial designs. Sankhya **8**, 107–166 (1947).
4. Fossorier M.P.C.: Quasicyclic low-density parity check codes from circulant permutation matrices. IEEE Trans. Inform. Theory **50**, 1788–1793 (2004).
5. Gallager R.G.: Low density parity check codes. IRE Trans. Inform. Theory **8**, 21–28 (1962).
6. Hirschfeld J.W.P.: Projective geometries over finite fields 2nd edn. Oxford University Press, Oxford. (1998).
7. Hirschfeld J.W.P., Storme L.: The packing problem in statistics, coding theory and finite projective spaces: update 2001. In: Blokhuis A., Hirschfeld J.W.P., Jungnickel D., and Thas J.A.(eds.) Proceedings of the fourth Isle of thorns conference, developments in mathematics, vol. 3, pp. 201–246. Kluwer Academic Publishers, Dordrecht Finite Geometries, (Chelwood Gate, 16-21 July 2000), (2001).
8. Johnson S.J., Weller S.R.: Regular low-density parity-check codes from oval designs. Eur. Trans. Telecommun. **14**(5), 399–409 (2003).
9. Johnson S.J., Weller S.R.: Construction of low-density parity-check codes from Kirkman triple systems, In: Proceedings of the IEEE globecom conference, San Antonio TX, available at http://www.ee.newcastle.edu.au/users/staff/steve/ (2001).
10. Johnson S.J., Weller S.R.: Construction of low-density parity-check codes from combinatorial designs. In: Proceedings of the IEEE information theory workshop, pp. 90–92. Cairns, Australia (2001).
11. Johnson S.J., Weller S.R.: Codes for iterative decoding from partial geometries. In: Proceedings of the IEEE international symposium on information theory, p. 6. Switzerland, June 30–July 5, extended abstract, available at http://murray.newcastle.edu.au/users/staff/steve (2002).
12. Kim J.L., Mellinger K., Storme L.: Small weight code words in LDPC codes defined by (dual) classical generalized quadrangles. Des. Codes Cryptogr. **42**(1), 73–92 (2007).
13. Kim J.L., Pele U., Perepelitsa I., Pless V., Friedland S.: Explicit construction of families of LDPC codes with no 4-cycles. IEEE Trans. Inform. Theory **50**, 2378–2388 (2004).
14. Kou Y., Lin S., Fosserier M.P.C.: Low-density parity-check codes based on finite geometries: a rediscovery and new results. IEEE Trans. Inform. Theory **47**(7), 2711–2736 (2001).
15. Liu Z., Pados D.A.: LDPC codes from generalized polygons. IEEE Trans. Inform. Theory **51**(11), 3890–3898 (2005).
16. MacKay D.J.C., Neal R.M.: Near Shannon limit performance of low density parity check codes. Electron. lett. **32**(18), 1645–1646 (1996).
17. MacKay D.J.C.: Good error correcting codes based on very sparse matrices. IEEE Trans. Inform. Theory **45**(2), 399–431 (1999).
18. MacKay D.J.C., Davey M.C.: Evaluation of Gallager codes for short block length and high rate applications; Codes, systems and graphical models. In: Marcus B., Rosethal J. (eds.) IMA in Mathematics and its Applications, vol. 123, pp. 113–130. Springer-Verlag, New York (2000).

19. Margulis G.A.: Explicit constructions of graphs without short cycles and low density codes. Combinatorica **2**, 71–78 (1982).
20. Payne S.E., Thas J.A.: Finite Generalized Quadrangles. Pitman Advanced Publishing Program, MA (1984).
21. Pepe V., Storme L., Van de Voorde G.: Small weight code words in the LDPC codes arising from linear representations of geometries. J. Combin. Designs **17**, 1–24 (2009).
22. Rosenthal J., Vontobel P.O.: Construction of LDPC codes using Ramanujan graphs and ideas from Margulis. In: Voulgaris P.G., and Srikant R. (eds.) Proceedings of the 38th Allerton conference on communications, control and computing, Monticello, IL, pp. 248–257. Coordinated Science Laboratory. 4–6 Oct 2000.
23. Segre B.: Ovals in a finite projective plane. Canad J Math **7**, 414–416 (1955).
24. Sin P., Xiang Q.: On the dimension of certain LDPC codes based on q-regular bipartite graphs. IEEE Trans. Infom. Theory **52**, 3735–3737 (2006).
25. Sipser M., Spielman D.A.: Expander codes. IEEE Trans. Inform. Theory **42**, 1710–1722 (1996).
26. Tanner R.M., Sridhara D., Sridharan A., Fuja T.E., Costello Jr J.D.: LDPC block codes and convolutional codes based on circulant matrices. IEEE Trans. Inform. Theory **50**, 2966–2984 (2004).
27. Vontobel P.O., Tanner R.M.: Construction of codes based on finite generalized quadrangles for iterative decoding. In: Proceedings of 2001 IEEE international symposium information theory, p. 233. Washington DC (2001).