Manuscript submitted to
AIMS' Journals
Volume **4**, Number **3**, August **2010**

Website: http://AIMsciences.org

pp. **405–417**

LDPC CODES ASSOCIATED WITH LINEAR REPRESENTATIONS OF GEOMETRIES

Peter Vandendriessche

Boudewijn Hapkenstraat 5
8820 Torhout, Belgium

(Communicated by Marcus Greferath)

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 any subset of $\mathrm{PG}(2,q)$, with $q = p^h$, $p \neq \mathrm{char}\,\mathbb{K}$. This includes the geometry $LU(3,q)^D$, the generalized quadrangle $T_2^*(\mathcal{K})$ with $\mathcal{K}$ a hyperoval, the affine space $\mathrm{AG}(3,q)$ and several partial and semi-partial geometries. In some cases the dimension and/or the code words of minimum weight are known. We prove an expression for the dimension and the minimum weight of the code. We classify the code words of minimum weight. We show that the code is generated completely by its words of minimum weight. We end with some practical considerations on the choice of $\mathcal{K}$.

1. **Introduction.** Originally introduced by Gallager [5], low density parity check (LDPC) codes are used frequently today due to their excellent empirical performance under belief-propagation/sum-product decoding [19]. In some cases, their performance is even near to the Shannon limit [19]. In general, a binary LDPC code C is a linear block code defined by a sparse parity check matrix H, this is a matrix that contains a lot more 0s than 1s.

To exploit structural properties, one usually wants an explicit construction rather than just random matrices. Lately there have been many different constructions: based on permutation matrices [4],[31], Ramanujan graphs [20],[24], expander graphs [28], q-regular bipartite graphs [14] or other incidence structures in discrete mathematics. In particular, one can take the incidence matrix of a finite geometry as the parity check matrix of a code.

Examples of such codes can be found in [9, 10, 11, 12], [18], [34]. Later, simulation results of Liu and Pados [16] 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. It would be interesting to perform the same simulations for the incidence geometries studied in this article, since all handled structures have a girth of at least 6 in their associated Tanner graph. If $\mathcal{K}$ is an arc, then the Tanner graph even has girth at least 8.

One class of geometries studied for this purpose are linear representations of geometries. One case that received a lot of attention lately is $T_2^*(\mathcal{K})$, with $\mathcal{K}$ a hyperoval [23, 33]. Here the minimum weight is known, the dimension is known

2000 *Mathematics Subject Classification.* Primary: 51E15, 51E20, 94B05.

Key words and phrases. LDPC codes, linear codes, finite geometry, linear representation, minimum distance, dimension.

when the characteristic of the code field $\operatorname{char}\mathbb{K} \neq 2$ and then we also know that the code is generated by its code words of minimum weight. Other structures studied in less detail in [23] include $T_2^*(\mathcal{B})$ with $\mathcal{B}$ a Baer subplane, $T_2^*(\mathcal{U})$ with $\mathcal{U}$ a unital, and $T_2^*(\mathcal{L})$ with $\mathcal{L}$ the pointwise union of two intersecting lines. One linear representation that has received a lot of attention is $LU(3,q)^D$ (and its dual $LU(3,q)$, which is not a linear representation) [13],[14],[27],[33]. In [14], the authors conjecture the binary dimension of the associated code to be

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

if q is odd. Here $q=p^h$ with p prime denotes the order of the finite field of the geometry. This conjecture was proven in [27]. Over all code fields with $\operatorname{char}\mathbb{K} \neq p$, this was proven in [33]. In this paper we present a uniform approach, having both of these as an immediate corollary.

In this article we study the general problem of codes associated with linear representations of geometries when $\operatorname{char}\mathbb{K} \neq p$. We generalize several of the above results, including the main theorems in [27] and [33]. When $\operatorname{char}\mathbb{K} \neq p$, we compute the minimum weight, the rank and the code rate of the code, we classify the code words of minimum weight and we prove that every code word is a linear combination of the code words of minimum weight. We end with some practical considerations on the choice of $\mathcal{K}$.

2. **Preliminaries.**

Let us begin by introducing some basic notations and definitions.

Notation 2.1. A finite field $\mathbb{F}_q$ of order q has $q=p^h$ elements, where p is a prime number and h is a positive integer. The characteristic of that field is p. We denote by $\operatorname{char}F$ the characteristic of the field F.

Notation 2.2. A linear $[n,k,d]$-code C over the field $\mathbb{F}_q$ can be defined by a parity check matrix H, this is an $(n-k)\times n$ matrix with the property that a word $c=(c_1,\ldots,c_n)$ is a code word of C if and only if $Hc^T=\vec{0}$ over $\mathbb{F}_q$, i.e. if and only if $Hc^T \equiv 0 \pmod p$, where $q=p^h$ with p prime and h a positive integer. In this paper we let H be the incidence matrix of a geometrical structure, hence its entries are only the elements 0 and 1 of the field $\mathbb{F}_q$.

Notation 2.3. We denote by $\mathrm{PG}(n,q)$ the n-dimensional projective space over the finite field $\mathbb{F}_q$. For $n=2$, we call this a projective plane and write $\mathrm{PG}(2,q)$. We denote by $\mathrm{AG}(n,q)$ the n-dimensional affine space over the finite field $\mathbb{F}_q$. A hyperoval is a set of $q+2$ points in $\mathrm{PG}(2,q)$ such that no three of them are collinear. Hyperovals exist if and only if q is even. More background on (substructures of) projective and affine spaces can be found in [7].

Definition 2.4. Let $\mathrm{PG}(3,q)$ be the 3-dimensional projective space over the field $\mathbb{F}_q$. Let $\Pi_0 := \mathrm{PG}(2,q)$ be a (hyper)plane in it and let $\mathcal{K}$ be an arbitrary subset of the points of that hyperplane. We define the geometry $T_2^*(\mathcal{K})$ as follows:

- the points of $T_2^*(\mathcal{K})$ are the affine points, being the points of $\mathrm{PG}(3,q)\backslash\mathrm{PG}(2,q)$,
- the lines are the affine lines of $\mathrm{PG}(3,q)$ which go through a point of $\mathcal{K}$,
- the incidence relation is inherited from $\mathrm{PG}(3,q)$.

Remark 2.5. Note that through every (affine) point we have $|\mathcal{K}|$ lines, one through each point of $\mathcal{K}$, while every line contains q points. In total there are q^3 points and $|\mathcal{K}|q^2$ lines: q^2 through each point of $\mathcal{K}$.

Remark 2.6. Let $N = |\mathcal{K}|$ and let H be the $q^3 \times Nq^2$ incidence matrix of $T_2^*(\mathcal{K})$, where points correspond to rows of H and lines correspond to columns of H and to the positions in the code. Let C be the linear code with H as its parity check matrix, over an arbitrary finite field $\mathbb{K}$. One can associate a *coefficient* to each line in a code word w, being its value at the corresponding position. A word $c = (c_1, \ldots, c_{Nq^2}) \in \mathbb{K}^{Nq^2}$ is in C if and only if $w \cdot H^T = \vec{0}$, hence (since $H_{ji} = 1 \Leftrightarrow \ell_i \ni p_j$) if and only if

$$\sum_{\ell_i} c_i H_{ji} = \sum_{\ell_i \ni p_j} c_i = 0$$

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 coefficients of the lines through every point equals 0 over $\mathbb{K}$.

Definition 2.7. Let $r_i, r_j \in \mathcal{K}$ with $i < j$ and let π be a projective two-dimensional plane through r_i, r_j different from Π_0. The *plane word through r_i and r_j in π* is the code word with

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

Notation 2.8. We denote by C' the code generated by all plane words. Given a plane word w through p_i and p_j, define $T(w)$ to be the plane π in Definition 2.7 and $L(w)$ to be the line $p_i p_j$ in Definition 2.7.

Definition 2.9. Let L be a line in Π_0 containing at least two points of $\mathcal{K}$. Let π be a projective (two-dimensional) plane through L different from Π_0, and let $p_0, \ldots, p_{k-1}$ be the points of $L \cap \mathcal{K}$. We define a *generalized plane word in π* to be a code word with

- a_0 in the positions corresponding to the lines of π through p_0,
- a_1 in the positions corresponding to the lines of π through p_1,
- $\ldots$
- a_{k-2} in the positions corresponding to the lines of π through p_{k-2},
- $-a_0 - a_1 - \ldots - a_{k-2}$ in the positions corresponding to the lines of π through p_{k-1},
- 0 in the positions corresponding to all other lines,

for some scalars $a_0, a_1, a_2, \ldots, a_{k-2} \in \mathbb{K}$.

Remark 2.10. Note that if a line would contain two points of $\mathcal{K}$, then it is not a line of $T_2^*(\mathcal{K})$, because is contained in the plane at infinty and hence not an affine line.

Remark 2.11. Note that a sum of plane words in a fixed plane π is a generalized plane word in π, and a sum of generalized plane words in π is still a generalized plane word in π. Moreover, if $\pi \cap \mathcal{K} = \{p_0, \ldots, p_{k-1}\}$, then the set of generalized plane words in π is spanned by the plane words through $(p_0, p_1), (p_0, p_2), \ldots, (p_0, p_{k-1})$. Hence, C' is also the code spanned by all generalized plane words and we need at most one generalized plane word per plane to obtain any word of C'.

Remark 2.12. It is known that $T_2^*(\mathcal{K})$ is a partial geometry if and only if $\mathcal{K}$ is a (maximal) $\{qn - q + n; n\}$-arc for some $n \geq 1$ (see [32]) and $T_2^*(\mathcal{K})$ is a semipartial geometry if and only if $\mathcal{K}$ is a Baer subplane or a unital (see [3]). A good general reference on $T_2^*(\mathcal{K})$ is [2].

3. Dimension of C'.

Notation 3.1. Denote by $\mathcal{L}$ the set of projective lines at infinity that contain at least one point of $\mathcal{K}$. Denote by L_N the size of $\mathcal{L}$, i.e.

$$\mathcal{L} = \{\ell_1, \ell_2, \ldots, \ell_{L_N}\}$$

and by L_S the summed size of $\mathcal{L}$, i.e.

$$L_S = \sum_{\ell \in \mathcal{L}} |\ell \cap \mathcal{K}|.$$

Let $\mathbb{K}$ be an arbitrary field with $\operatorname{char} \mathbb{K} \neq p$.

Lemma 3.2. *Let $\ell \in \mathcal{L}$ be a line in the plane at infinity, containing exactly k points of $\mathcal{K}$. Then there are exactly $q(k-1)$ linearly independent plane words among all plane words w with $L(w) = \ell$.*

Proof. Number the k points $p_0, p_1, \ldots, p_{k-1}$, then the pairs

$$(p_0, p_1), (p_0, p_2), \ldots, (p_0, p_{k-1})$$

each yield q different plane words, and these are linearly independent. Now for all other pairs, the plane words through (p_i, p_j), with $i < j$, can be written as the difference of the corresponding plane words through (p_0, p_i) and (p_0, p_j). Hence the result follows. □

Lemma 3.3. *Fix an arbitrary point $p_0 \in \mathcal{K}$. Let[1] $\sum_{i=1}^{n} \lambda_i v_i = \vec{0}$ be a linear combination of generalized plane words yielding the zero word, at most one generalized plane word per plane. If $L(v_i) = L(v_j)$ and this line contains p_0, then $\lambda_i = \lambda_j$.*

Proof. The q affine lines through p_0 in $T(v_i)$ each get a contribution of λ_i from v_i. The q affine lines through p_0 in $T(v_j)$ each get a contribution of λ_j from v_j. All other generalized plane words v_m contribute equally much to the sum of both sets of q lines (namely λ_m if $T(v_m)$ contains p_0 and 0 otherwise). Denote by R the total summed contribution to both sets of q lines.

Since the total sum of all contributions is 0 for every line (since we assumed that this linear combination yields the zero word) we have $q\lambda_i + R = 0 = q\lambda_j + R$, hence $q\lambda_i = -R = q\lambda_j$. Since $q \neq 0$ as an element of $\mathbb{K}$, it follows that $\lambda_i = \lambda_j$. □

Corollary 3.4. *We did not assume $\lambda_i \neq 0$, hence if one of the generalized plane words v_i appears in the linear combination with a nonzero λ_i, then all generalized plane words v_j through the same line at infinity should appear with $\lambda_j = \lambda_i$. Hence if we start from an empty code (considered as vector space), and we consider one by one all lines ℓ at infinity and we add the q generalized plane words through ℓ to this vector space, then each line increases the dimension with at least $(|\ell \cap \mathcal{K}| - 1)(q-1)$, since the codimension can be at most $|\ell \cap \mathcal{K}| - 1$.*

Theorem 3.5. *The dimension of C' is $(N-1) + (q-1)(L_S - L_N)$.*

Proof. Take any point p_0 and look at the lines $L_0, \ldots, L_q$ through p_0 in the plane at infinity. The plane words through (p_0, p), for $p \in \mathcal{K} \setminus \{p_0\}$, form a basis for the linear combinations of plane words on the lines $L_0, \ldots, L_q$. Starting from an empty vector space V as described in Corollary 3.4, these plane words contribute $(N-1)q$

[1] By convention, we choose all generalized plane words containing lines through p_0 in their support, to have coefficient $+1$ on the lines through p_0 (this can always be accomplished by scaling the λ_i).

to the dimension, because of Lemma 3.2. Now adding every other line $L \in \mathcal{L}$ at infinity, not through p_0, contributes

- at least $(|L \cap \mathcal{K}| - 1)(q-1)$ to $\dim V$, since Lemma 3.3 states that in any linear combination of generalized plane words yielding the zero word, all planes through L appear with the same coefficient, hence we miss at most $|L \cap \mathcal{K}| - 1$ degrees of freedom,
- and at most $(|L \cap \mathcal{K}| - 1)(q-1)$ to $\dim V$, since one can write the zero word as a linear combination of plane words through $(p_0, p), (p_0, p'), (p, p')$ for any two points $p, p' \in L \cap \mathcal{K}$ (note that all plane words through lines through p_0 are already in the code at this point).

Therefore, the dimension is exactly

$$(N-1)q + \sum_{p_0 \notin L \in \mathcal{L}} (|L \cap \mathcal{K}| - 1)(q-1).$$

Note that $\sum_{p_0 \in L \in \mathcal{L}} (|L \cap \mathcal{K}| - 1) = N - 1$ since both represent all points of $\mathcal{K}$ except for p_0. Hence

$$\begin{aligned} \dim(C') &= (N-1)q + \textstyle\sum_{p_0 \notin L \in \mathcal{L}} (|L \cap \mathcal{K}| - 1)(q-1) \\ &= (N-1) + (q-1) \textstyle\sum_{L \in \mathcal{L}} (|L \cap \mathcal{K}| - 1) \\ &= (N-1) + (q-1) \left(\left(\textstyle\sum_{L \in \mathcal{L}} |L \cap \mathcal{K}| \right) - |\mathcal{L}| \right) \\ &= (N-1) + (q-1)(L_S - L_N). \end{aligned}$$

□

Remark 3.6. We now know that C' is a linear $[Nq^2, N-1+(q-1)(L_S-L_N)]$-code. There is no general expression for $L_S - L_N$ in terms of q and N. However, there is an easy algorithm to compute $L_S - L_N$ for an arbitrary set $\mathcal{K}$:

Let $\mathcal{K}$ be an arbitrary subset of $\mathrm{PG}(2, q)$. Fix any point p_0 (inside $\mathcal{K}$ or outside $\mathcal{K}$). Call a line through p_0

- a *secant* if it contains two or more points of $\mathcal{K} \setminus \{p_0\}$,
- a *tangent* if it contains exactly one point of $\mathcal{K} \setminus \{p_0\}$, and
- a *passant* if it contains no points of $\mathcal{K} \setminus \{p_0\}$.

When adding/removing a point p_0,

- L_S increases/decreases by $q+1$, while
- L_N increases/decreases by the number of passants through p_0.

Hence, $L_S - L_N$ increases/decreases by the number of non-passant lines through p_0.

Some examples:

- If $\mathcal{K}$ is a k-arc, then adding the ith point increases $L_S - L_N$ by $i-1$. Hence,

$$L_S - L_N = \sum_{i=1}^{k} (i-1) = \frac{k(k-1)}{2}.$$

- If $\mathcal{K}$ is the pointwise union of two intersecting lines, then adding the points on the first line increases $L_S - L_N$ by 1 each time (except for the first point), while adding the other q points increases it by $q+1$ each time. Hence, in this case

$$L_S - L_N = q + q(q+1) = q^2 + 2q.$$

Remark 3.7. Since one has in general that $L_S = N(q+1)$, the dimension formula can be rewritten as $q^2N - q^3 + (q-1)(q^2+q+1-L_N)$. Since the parity check matrix is a $q^3 \times q^2N$ matrix, this means that the rank deficiency of the parity check matrix is $q-1$ times the number of lines at infinity, skew to $\mathcal{K}$. It may be interesting to find out if there exists a more direct way to obtain this formula.

4. **Dimension of C.** We will compute the dimension of $T_2^*(\mathcal{K})$ as follows. First we will compute $\dim C$ in the case $\mathcal{K} = \mathrm{PG}(2,q)$, and find that it equals $\dim C'$ in that case, hence $C = C'$. Then we will present a technique to keep this property valid while removing arbitrary points from $\mathcal{K}$. Every subset $\mathcal{K}$ of $\mathrm{PG}(2,q)$ can be obtained by removing a finite number of points from $\mathrm{PG}(2,q)$, so the conclusion will follow. As in the previous section, we will always assume that $q \neq 0$ over $\mathbb{K}$, i.e. $\mathrm{char}\,\mathbb{K} \neq p$.

4.1. **The case $\mathcal{K} = \mathrm{PG}(2,q)$.** In this case, we simply have $T_2^*(\mathcal{K}) = \mathrm{AG}(3,q)$. Remark that $\mathrm{AG}(3,q)$ is a 2-$(q^3,q,1)$ block design, using lines as blocks. Denote by A_n the incidence matrix of $\mathrm{AG}(n,q)$, with $n \geq 2$, where points correspond to rows and lines correspond to columns.

It is a classical result in design theory (see [1] for a proof and more general background on designs) that $A_nA_n^T = (q^{n-1} + q^{n-2} + \ldots + q^2 + q)I + J$, where J denotes the $q^n \times q^n$ matrix with all entries equal to 1. This has determinant

$$(q^n + q^{n-1} + \ldots + q^2 + q)(q^{n-1} + q^{n-2} + \ldots + q^2 + q)^{q^n-1}.$$

In fact, $A_nA_n^T$ has $q^n - 1$ eigenvalues equal to $q^{n-1} + q^{n-2} + \ldots + q^2 + q$ and one eigenvalue equal to $q^n + q^{n-1} + \ldots + q^2 + q$. The determinant is non-zero when $q^{n-1} + q^{n-2} + \ldots + q^2 + q \neq 0$ and $q^n + q^{n-1} + \ldots + q^2 + q \neq 0$. For most choices of the characteristic, this already shows that A_n has full rank, however in some cases further study is required:

- The case $q = 0$ has been excluded by our assumptions. In fact, A_n does not have full rank in this case. We will further assume $q \neq 0$.
- The case $q^{n-1} + q^{n-2} + \ldots + q + 1 = 0$ is easily solved. Note that this implies $q \neq 0$ and $q^{n-2} + q^{n-3} + \ldots + q + 1 \neq 0$, hence in this case only one of the eigenvalues of $A_nA_n^T$ equals zero over $\mathbb{K}$, and we see that $(1,\ldots,1)^T$ is an eigenvalue corresponding to this eigenvector. Hence $(1,\ldots,1)^T$ is (up to scalar multiples) the only eigenvector corresponding to this eigenvalue, but one can verify that

  $$(1,\ldots,1)A_n^T = q^n(1,\ldots,1) \neq \vec{0}$$

  since $q \neq 0$. Hence, this is not a code word, and hence there is no $v \neq \vec{0}$ such that $A_nv = \vec{0}$, i.e. A_n has full rank in this case.
- The case $q^{n-2} + q^{n-3} + \ldots + q + 1 = 0$ is more difficult. Purely combinatorial approaches seem to fail, but a geometric trick works. We will now develop this technique to find the rank of A_n over $\mathbb{K}$ for the case $q^{n-2}+q^{n-3}+\ldots+q+1 = 0$.

Lemma 4.1. *Let $k \geq 2$. If the incidence matrix of $\mathrm{AG}(k+1,q)$ is rank deficient over $\mathbb{K}$, then the incidence matrix of $\mathrm{AG}(k,q)$ is also rank deficient over $\mathbb{K}$.*

Proof. Assume that the incidence matrix of $\mathrm{AG}(k+1,q)$ is rank deficient. Since $k \geq 2$, there are more lines than points. Hence, rank deficiency means that there

exists a linear combination of points whose corresponding line-incidence vectors yield the zero vector (zero for each line):

$$\sum_{p_i \in \mathrm{AG}(k+1,q)} c_i p_i = \vec{0}$$

with not all $c_i = 0$ over $\mathbb{K}$.

Now consider any hyperplane $\Pi \cong \mathrm{AG}(k,q)$ in our $\mathrm{AG}(k+1,q)$ which contains at least one point with non-zero coefficient in the linear combination. For all lines contained in Π, the linear combination of the point-line incidence vectors has to be zero as well. I.e. in Π we also have $\sum_{p_i \in \Pi} c_i p_i = \vec{0}$. Since Π contains at least one point with non-zero coefficient in the linear combination, this linear combination is nontrivial. Hence, the incidence matrix of $\mathrm{AG}(k,q)$ is rank deficient as well. □

Theorem 4.2. *The incidence matrix of* $\mathrm{AG}(n,q)$ *(with* $n \geq 2$*),* $q = p^h$ *with* p *prime, has full rank over all fields with* $\operatorname{char}\mathbb{K} \neq p$.

Proof. We will prove this by induction on n. For $n = 2$ this is clear from the remarks in the beginning of this section, since the problematic case above ($q^{n-2}+\ldots+q+1 = 0$) reduces to $1 = 0$, hence can be excluded immediately. For each $n \geq 2$ it follows by contraposition of Lemma 4.1 that if the statement is true for n, then it is also true for $n+1$. Hence, by induction, it is true for all $n \geq 2$. □

We have proven thatf A_n has full rank for all $n \geq 2$ when $\operatorname{char}\mathbb{K} \neq p$. Explicitly, for the general setting of $\mathcal{T}_n^*(\mathcal{K})$ with $\mathcal{K} = \mathrm{PG}(n,q)$ (which yields exactly $\mathrm{AG}(n+1,q)$) the geometry has $q^n(q^n + q^{n-1} + \ldots + q + 1)$ lines and q^{n+1} points, hence Theorem 4.2 yields

$$\dim C = q^n(q^n + q^{n-1} + \ldots + q^3 + q^2 + 1).$$

Now compare this to the result of Section 3. In Section 3 we worked with $n = 2$, hence $N = q^2 + q + 1$ and $T_2^*(\mathcal{K}) = \mathrm{AG}(3,q)$. This gives us

$$\begin{aligned} \dim(C') &= (N-1) + (q-1)(L_S - L_N) \\ &= q^2 + q + (q-1)((q^2+q+1)(q+1) - (q^2+q+1)) \\ &= q^2(q^2+1) \\ &= \dim C. \end{aligned}$$

and hence $\dim C = \dim C'$. Hence, the code associated with $\mathrm{AG}(3,q)$ is spanned completely by its plane words. However, for $n > 3$, Theorem 3.5 is no longer valid. We finish with a conjecture for the higher dimensions:

Conjecture 4.3. *If* $\mathcal{K} = \mathrm{PG}(n,q)$ *with* $n > 2$ *and* $\operatorname{char}\mathbb{K} \neq p$ *then the code associated with* $\mathcal{T}_n^*(\mathcal{K}) = \mathrm{AG}(n+1,q)$ *over* $\mathbb{K}$ *also has* $\dim C' = \dim C$.

4.2. **The general case.** The main idea here is the following: if we remove a point from $\mathcal{K}$, we claim that the property that the code is spanned by its plane words remains valid. To distinguish between different point sets, we denote by $C_\mathcal{K}, C'_\mathcal{K}$ respectively the full code and the plane words code associated with $T_n^*(\mathcal{K})$. Similarly, we denote by $L_{N,\mathcal{K}}, L_{S,\mathcal{K}}$ the respective values of L_N and L_S for the set $\mathcal{K}$.

Theorem 4.4. *Let* $\mathcal{K}$ *be a nonempty subset of* $\mathrm{PG}(2,q)$ *and let* $\operatorname{char}\mathbb{K} \neq p$. *We have* $\dim C_\mathcal{K} = \dim C'_\mathcal{K}$ *(and hence* $C_\mathcal{K} = C'_\mathcal{K}$*).*

Proof. According to Theorem 3.5, if $T \subseteq \mathrm{PG}(2,q)$, then removing a point p_0 from T decreases $\dim C'$ by

$$\dim C'_T - \dim C'_{T\setminus\{p_0\}} = 1 + (q-1)((L_{S,T} - L_{N,T}) - (L_{S,T\setminus\{p_0\}} - L_{N,T\setminus\{p_0\}}))$$

and hence decreases $\dim C$ by at least this amount.

Fix one point $p_0 \in \mathcal{K}$. Now remove all other points of $\mathrm{PG}(2,q)$ one by one, first the points outside $\mathcal{K}$ then the points inside $\mathcal{K}$, except for p_0. Denote by T_i the set in the intermediary step with i points:

$$T_{|\mathrm{PG}(2,q)|} = \mathrm{PG}(2,q), \quad T_N = \mathcal{K}, \quad T_1 = \{p_0\},$$

and define $Q = |\mathrm{PG}(2,q)| - 1$. Here, $|\mathrm{PG}(2,q)|$ denotes the number of points in $\mathrm{PG}(2,q)$, which is q^2+q+1. Note that in any code word, every affine point lies on either 0 or at least 2 lines of the support of that code word, hence $C_{\{p_0\}} = \{\vec{0}\}$. Note that $C_{\{p_0\}}$ is simply $C_{\mathcal{K}}$ with $\mathcal{K} = \{p_0\}$. Hence, we have

$$\begin{aligned}
\dim C_{\mathrm{PG}(2,q)} &= \dim C_{\mathrm{PG}(2,q)} - \dim C_{\{p_0\}} \\
&= \sum_{i=1}^{Q} \left(\dim C_{T_{i+1}} - \dim C_{T_i}\right) \\
&\geq \dim C'_{T_2} - \dim C'_{T_1} + \sum_{i=2}^{Q} \left(\dim C_{T_{i+1}} - \dim C_{T_i}\right) \\
&\geq \cdots \\
&\geq \sum_{i=1}^{Q} \left(\dim C'_{T_{i+1}} - \dim C'_{T_i}\right) \\
&= \dim C'_{\mathrm{PG}(2,q)} - \dim C'_{\{p_0\}} \\
&= \dim C'_{\mathrm{PG}(2,q)}.
\end{aligned}$$

It was proven in the previous subsection that $\dim C_{\mathrm{PG}(2,q)} = \dim C'_{\mathrm{PG}(2,q)}$, hence we must have equality in each inequality "$\geq$". This means that

$$\dim C_{T_{i+1}} - \dim C_{T_i} = 1 + (q-1)((L_{S,T_{i+1}} - L_{N,T_{i+1}}) - (L_{S,T_i} - L_{N,T_i}))$$

for each i. A simple induction gives $\dim C_{T_i} = \dim C'_{T_i}$ for all i, in particular for $i = N$ we have $\dim C_{\mathcal{K}} = \dim C'_{\mathcal{K}}$. □

Hence for $T_2^*(\mathcal{K})$ in general it is now proven that $\dim C = \dim C'$ and the code C is generated completely by its plane words.

Remark 4.5. If Conjecture 4.3 is true, then Theorem 4.4 can be extended to arbitrary subsets of $\mathrm{PG}(n,q)$: then we have $\dim C = \dim C'$ for the code associated with $T_n^*(\mathcal{K})$ with arbitrary $\mathcal{K} \subseteq \mathrm{PG}(n,q)$.

Remark 4.6. As an immediate consequence, we get that in the binary code associated with $T_2^*(\mathcal{K})$ for q odd, all code words have even weight.

5. The minimum distance of C.

Now that the dimension and structure of C are known, we can attack another one of its key properties: the minimum distance. In some sporadic cases the minimum distance is known [13, 23], however in most cases one only has lower bounds from the tree bound [29], the bit-oriented bound and the parity-oriented bound [30].

Theorem 5.1. *For any finite field $\mathbb{K}$, all code words c with $w(c) < 2q$ must be contained in a single plane. If $w(c) = 2q$, then either c is a plane word, $supp(c)$ is the set of lines of a hyperbolic quadric with two intersecting lines contained in $\mathcal{K}$, or $\operatorname{char}\mathbb{K} = p = 2$.*

Proof. This follows from [23], Proposition 4. □

Now, we will use the structure of C to sharpen this result and classify the minimum weight code words.

Theorem 5.2. *If $\operatorname{char}\mathbb{K} \neq p$, there are no code words $c \in C$ with $w(c) < 2q$. If $w(c) = 2q$, then either c is a plane word or $supp(c)$ is the set of lines of a hyperbolic quadric with two intersecting lines contained in $\mathcal{K}$.*

Proof. The second part follows immediately from Theorem 5.1. For the first part, assume that there exists a code word with $w(c) < 2q$ and let U be its support. By Theorem 5.1, the support of this code word is contained in a plane T. Define $m = |\mathcal{K} \cap T|$ and write $\mathcal{K} \cap T = \{p_1, \ldots, p_m\}$. Since $\operatorname{supp}(c) \subset T$, we have that c is a generalized plane word in T. Hence, either each of the q lines through p_i appear in U, or none of them do. Since one needs either 0 or at least 2 lines through a point, it follows that $w(c) \geq 2q$, a contradiction. □

Remark 5.3. Note that it may actually happen that there are minimum weight code words other than plane words – Theorem 5.1 classfies them. However, it follows from Theorem 4.4 that these other minimum weight code words are also a linear combination of plane words. Hence, even in this case, the statements 'C is generated by its (generalized) plane words' and 'C is generated by its code words of minimum weight' are equivalent.

So far we have proven that C is a linear $[Nq^2, N-1+(q-1)(L_S-L_N), 2q]$-code and it is completely generated by its minimum weight code words.

Now, let us see what happens for $T_n^*(\mathcal{K})$ with $n > 2$, assuming Conjecture 4.3 is true.

Theorem 5.4. *If Conjecture 4.3 is true, then $d(C) = 2q$ is true for $T_n^*(\mathcal{K})$ with arbitrary $n \geq 2$ and arbitrary $\mathcal{K} \subseteq \mathrm{PG}(n,q)$ (still assuming $\operatorname{char}\mathbb{K} \neq p$).*

Proof. Assume there exists a non-zero code word c with $w(c) < 2q$ and let U be its support. Let T be any plane and define $t = |U \cap T|$. Since the sum of the coefficients has to be 0 in each point, any point on a line of the support lies on at least one other line of the support. Hence we have

$$t(q - t + 1) < 2q - t,$$

meaning $t > q$ or $t < 2$. Since t is an integer, this means $t \geq q+1$ or $t \leq 1$.

Now, if there are at least two planes for which $t \geq q+1$, then $w(c) \geq 2q+1$, contradiction. Hence there is at most one such plane and U is completely contained in this plane. The rest of the proof can be copied from Theorem 5.2. □

The following theorem summarizes the results obtained so far:

Theorem 5.5. *The code associated with $T_2^*(\mathcal{K})$ over any field $\mathbb{K}$, with $\operatorname{char}\mathbb{K} \neq p$, is a linear $[Nq^2, N-1+(q-1)(L_S-L_N), 2q]$-code and it is completely generated by its minimum weight code words. If Conjecture 4.3 is true, then for $\operatorname{char}\mathbb{K} \neq p$ the code associated with $T_n^*(\mathcal{K})$ also has $d(C) = 2q$ and it is also generated completely by its minimum weight code words.*

6. **Some practical considerations and further work.** A commonly used approach when constructing good LDPC codes is the maximization of the girth of its Tanner graph [8],[20],[35]. It is known that high girth decreases the dependence between passing messages in the belief-propagation sum-product algorithm. The Tanner graph of $T_2^*(\mathcal{K})$ always has a girth of at least 6. If $\mathcal{K}$ is an arc, then the girth is 8, as $T_2^*(\mathcal{K})$ contains no triangles.

Liu and Pados [16] mention an opposing objective: the minimization of the diameter of the Tanner graph, which brings them to generalized polygons. If $\mathcal{K}$ is not contained within a line, the diameter of $T_2^*(\mathcal{K})$ is at most 6. If $\mathcal{K}$ contains no tangents at infinity, the diameter is as low as 4. Examples of such choices of $\mathcal{K}$ include $\mathcal{K}$ a hyperoval and $\mathcal{K}$ a double blocking set.

There is a unique choice for $\mathcal{K}$ that combines both of the preceding objectives: the case where $\mathcal{K}$ is a hyperoval, which only exists for q even. Then $T_2^*(\mathcal{K})$ is a generalized quadrangle with girth 8 and diameter 4. From the above points of view, this is probably the most appealing case, however, the restriction $\operatorname{char}\mathbb{K} \neq p$ excludes the most important field for practical applications: $\mathbb{F}_2$.

Hyperoval	$q = p^h$	$\dim_{\mathbb{F}_2} C$	$\dim_{\mathbb{R}} C = \dim_{\mathbb{R}} C'$	$\dim_{\mathbb{F}_2} C'$
Regular hyperoval	$q = 2$	9	9	8
Regular hyperoval	$q = 4$	50	50	37
Regular hyperoval	$q = 8$	341	324	194
Regular hyperoval	$q = 16$	2670	2312	1105
Lunelli-Sce hyperoval	$q = 16$	2550	2312	1107
Regular hyperoval	$q = 32$	22248	17424	6578
Translation hyperoval	$q = 32$	21258	17424	6608
Cherowitzo hyperoval	$q = 32$	20358	17424	6613
Payne hyperoval	$q = 32$	20388	17424	6613
Segre hyperoval	$q = 32$	20553	17424	6613
O'Keefe-Penttila hyperoval	$q = 32$	20343	17424	6613
Regular hyperoval	$q = 64$	188665	135200	39937
Adelaide hyperoval	$q = 64$	169772	135200	40312
Subiaco I hyperoval	$q = 64$	169254	135200	40312
Subiaco II hyperoval	$q = 64$	169388	135200	40309

TABLE 1. Simulation results for the binary codes associated with $T_2^*(\mathcal{K})$ where $\mathcal{K}$ is a hyperoval.

For the binary code associated with $T_2^*(\mathcal{K})$ when $\mathcal{K}$ is a hyperoval, the results in this paper are no longer valid. Lemma 3.3 no longer guarantees the lower bound on $\dim C'$ and since the 2-rank is at most the real rank, the dimension of C could be larger than what we have derived in this paper. In Table 6, we have calculated the dimension (and hence the code rate, which is dimension over length) of the $\mathbb{F}_2$-code and $\mathbb{R}$-codes associated with $T_2^*(\mathcal{K})$ with $\mathcal{K}$ a hyperoval, by computer simulations, for multiple types of hyperovals in $\mathrm{PG}(2, 2^h)$. For $h \leq 5$, these are the only hyperovals (for proofs of these facts, see [6, 21, 22, 26]). For $h = 6$, it is commomly believed that these are the only hyperovals, but a proof has not been found yet. For $h > 6$, a classification is not even conjectured and the computations also become unfeasible.

We see that the results indeed deviate from the numbers in Theorem 5.5. In this case we get even better parameters: the dimension increases while no other

visible parameters change. However, we lose the structural property that the code is spanned by its code words of minimum weight. For other choices of $\mathcal{K}$ when q is even, even the minimum distance may decrease. This has been investigated more closely in [23]. In general, only a minimum weight of $q+1$ can be guaranteed.

If one wants to maintain the structure property and still use a binary code, then q must be odd. Some examples that are 'near-optimal' choices for $\mathcal{K}$ in these cases include:

- $\mathcal{K}$ is a $(q+1)$-arc (and hence a conic by [25]). Compared to the case where $\mathcal{K}$ is a hyperoval, the dimension and rate are slightly smaller, the code is a bit shorter in length and the girth remains 8. The diameter is 6 now, but there are only very few pairs of vertices where this distance is actually reached. It would be interesting to perform practical simulations to find out if this case still guarantees a fast average decoding speed.
- $\mathcal{K}$ has no tangents. For example, $\mathcal{K}$ is a dual double blocking set. Here the Tanner graph of $T_2^*(\mathcal{K})$ has girth 6 and the diameter is still 4, but the length of the code is necessarily longer compared to the case where $\mathcal{K}$ is a hyperoval, without an increase in minimum distance.

To finish, we take a look ahead on possible further work on this subject. Each of the following could be a significant contribution to the understanding of this class of codes.

- For $T_n^*(\mathcal{K})$, with $n > 2$, little is known. If Conjecture 4.3 is true, it could be interesting to find a generalization of the formula in Theorem 3.5 and to analyze its geometric interpretation. Regarding the optimal choice of $\mathcal{K}$, there is no n-dimensional equivalent of the hyperoval, so it would be interesting to know which choices for $\mathcal{K}$ yield interesting geometries (if any).
- Practical simulations on encoding/decoding speed or other performance measurements of $T_2^*(\mathcal{K})$ for some good choices for $\mathcal{K}$ would be very helpful. Especially the case $\mathcal{K}$ is a hyperoval, a conic or some of the examples studied in [23] or in this article, would be interesting. It would also be interesting to know if the case $\operatorname{char}\mathbb{K} = p$ generally performs better or worse than $\operatorname{char}\mathbb{K} \neq p$.
- If $\operatorname{char}\mathbb{K} = p$, few results in this paper remain valid. The only trick that works completely when $\operatorname{char}\mathbb{K} = p$ is that if the incidence matrix of $\mathrm{AG}(n+1,q)$ is rank deficient, then the incidence matrix of $\mathrm{AG}(n,q)$ is rank deficient. A suited structural property could potentially be preserved in a way similar to Section 4.2. This suggests that it may be a good help to first find out the structure of the base case $\mathrm{AG}(2,q)$, especially which code words remain valid if we remove certain classes of parallel lines. Another indication that this may be an interesting topic is the minimum distance. From Theorem 5.1, code words c with $w(c) < 2q$ are necessarily contained within a plane. If one knows the structure of the LDPC code associated with $\mathrm{AG}(2,q)$, one is likely to find a general result on the minimum weight. Until now, the only known lower bounds are the bounds in [23].
- If $\operatorname{char}\mathbb{K} = p$, it would be useful to find a structure or dimension result even just for special cases. Even for $T_2^*(\mathcal{K})$ with $\mathcal{K}$ a hyperoval this seems a lot harder. From the simulation results in Table 6, one can see that the dimensions are different between different types of hyperovals. This may be related to the approach in [23]: $\dim C$ may depend on how many points of a conic are contained in $\mathcal{K}$, while the difference in $\dim C'$ between regular/translation/other

hyperovals may be related to Remark 7 in [23], since other hyperovals are not known to have such special points.

Acknowledgements. The author wants to thank Leo Storme for his detailed proofreading and useful suggestions. The author also wants to thank the anonymous referees for their useful comments and suggestions.

REFERENCES

[1] P. J. Cameron and J. H. Van Lint, "Designs, Graphs, Codes and their Links," Cambridge University Press, 1991.
[2] F. De Clerck and H. Van Maldeghem, *On linear representations of* (α, β)*-geometries*, European J. Combin., **15** (1994), 3–11.
[3] I. Debroey and J. A. Thas, *Semi partial geometries in* $AG(2, q)$ *and* $AG(3, q)$, Simon Stevin, **51** (1978), 195–209.
[4] M. P. C. Fossorier, *Quasicyclic low-density parity check codes from circulant permutation matrices*, IEEE Trans. Inform. Theory, **50** (2004), 1788–1793.
[5] R. G. Gallager RG, *Low density parity check codes*, IRE Trans. Inform. Theory, **8** (1962), 21–28.
[6] M. Hall, *Ovals in the Desarguesian plane of order* 16, Ann. Mat. Pura Appl., **102** (1975), 159–176.
[7] J. W. P. Hirschfeld, "Projective Geometries over Finite Fields," 2^{nd} edition, Oxford University Press, 1998.
[8] X.-Y. Hu, E. Eleftheriou and D. M. Arnold, *Regular and irregular progressive edge-growth Tanner graphs*, IEEE Trans. Inform. Theory, **51** (2005), 386–398.
[9] S. J. Johnson and S. R. Weller, *Construction of low-density parity-check codes from combinatorial designs*, in "Proceedings of the IEEE Information Theory Workshop," Cairns, (2001), 90–92.
[10] S. J. Johnson and S. R. Weller, *Construction of low-density parity-check codes from Kirkman triple systems*, in "Proceedings of the IEEE Globecom Conference," San Antonio, 2001.
[11] S. J. Johnson and S. R. Weller, *Codes for iterative decoding from partial geometries*, in "Proceedings of the IEEE International Symposium on Information Theory," Switzerland, 2002.
[12] S. J. Johnson and S. R. Weller, *Regular low-density parity-check codes from oval designs*, Eur. Trans. Telecommun., **14** (2003), 399–409.
[13] J. L. Kim, K. Mellinger and L. Storme, *Small weight codewords in LDPC codes defined by (dual) classical generalized quadrangles*, Des. Codes Cryptogr., **42** (2007), 73–92.
[14] J. L. Kim, U. Peled, I. Perepelitsa, V. Pless and S. Friedland, *Explicit construction of families of LDPC codes with no* 4*-cycles*, IEEE Trans. Inform. Theory, **50** (2004), 2378–2388.
[15] Y. Kou, Y, 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, **47** (2001), 2711–2736.
[16] Z. Liu and D. A. Pados, *LDPC codes from generalized polygons*, IEEE Trans. Inform. Theory, **51**, 3890–3898.
[17] D. J. C. MacKay, *Good error correcting codes based on very sparse matrices*, IEEE Trans. Inform. Theory, **45** (1999), 399–431.
[18] D. J. C. MacKay and M. C. Davey, *Evaluation of Gallager codes for short block length and high rate applications*, in "Codes, Systems and Graphical Models" (eds. B. Marcus and J. Rosenthal), Springer-Verlag, New York, (2000), 113–130.
[19] D. J. C. MacKay and R. M. Neal, *Near Shannon limit performance of low density parity check codes*, Electron. Lett., **32** (1996), 1645–1646.
[20] G. A. Margulis, *Explicit constructions of graphs without short cycles and low density codes*, Combinatorica, **2** (1982), 71–78.
[21] C. M. O'Keefe and T. Penttila, *Hyperovals in* PG(2, 16), European J. Combin., **12** (1991), 51–59.
[22] T. Penttila and G. F. Royle, *Classification of hyperovals in* PG(2, 32), J. Geom., **50** (1994), 151–158.
[23] V. Pepe, L. Storme and G. Van de Voorde, *Small weight codewords in the LDPC codes arising from linear representations of geometries*, J. Combin. Des., **17** (2009), 1–24.

[24] J. Rosenthal and P. O. Vontobel, *Construction of LDPC codes using Ramanujan graphs and ideas from Margulis*, in "Proceedings of the 38th Allerton Conference on Communications, Control and Computing" (eds. P.G. Voulgaris and R. Srikant), Monticello, (2000), 248–257.
[25] B. Segre, *Ovals in a finite projective plane*, Canad. J. Math., **7** (1955), 414–416.
[26] B. Segre, *Sui k-archi nei piani finiti di caratteristica due* (in Italian), Rev. Math. Pures Appl., **2** (1957), 289–300.
[27] P. Sin and Q. Xiang, *On the dimension of certain LDPC codes based on q-regular bipartite graphs*, IEEE Trans. Inform. Theory, **52** (2006), 3735–3737.
[28] M. Sipser and D. A. Spielman, *Expander codes*, IEEE Trans. Inform. Theory, **42** (1996), 1710–1722.
[29] R. M. Tanner, *A recursive approach to low complexity codes*, IEEE Trans. Inform. Theory, **27** (1981), 533–547.
[30] R. M. Tanner, *Minimum distance bounds by graph analysis*, IEEE Trans. Inform. Theory, **47** (2001), 808–821.
[31] R. M. Tanner, D. Sridhara, A. Sridharan, T. E. Fuja and J. D. Costello Jr, *LDPC block codes and convolutional codes based on circulant matrices*, IEEE Trans. Inform. Theory, **50** (2004), 2966–2984.
[32] J. A. Thas, *Partial geometries in finite affine spaces*, Math. Z., **158** (1978), 1–13.
[33] P. Vandendriessche, *Some low-density parity-check codes derived from finite geometries*, Des. Codes Cryptogr., **54** (2010), 287–297.
[34] P. O. Vontobel and R. M. Tanner, *Construction of codes based on finite generalized quadrangles for iterative decoding*, in "Proceedings of 2001 IEEE International Symposium Information Theory," Washington, (2001), 233–233.
[35] H. Xiao and A. H. Banihashemi, *Improved progressive-edge-growth (PEG) construction of irregular LDPC codes*, IEEE Commun. Lett., **8**, 715–717.

Received December 2009; revised May 2010.

E-mail address: Peter.Vandendriessche@UGent.be