LDPC Codes from Finite Geometries

Valentina Pepe

October 2007

Contents

Introduction

The concept of $LDPC$ code was developed by Robert G. Gallager in his doctoral dissertation at MIT in 1960 ([15]). Impractical to implement, the $LDPC$ codes were largely overlooked for almost 35 years, but in the last years, with the development of the iterative decoding algorithms, the interest in $LDPC$ codes has increased dramatically, since their performance close to the theoretical maximum (i.e. the *Shannon Limit*), when decoded iteratively. The construction of a $LDPC$ code falls into two main types of techniques: pseudo-random techniques and combinatorial approaches. Construction by a pseudo-random approach builds on theoretical results stating that, for large block-size, a random construction gives good decoding performance. In general, pseudo-random codes have complex encoding and decoding, and difficulty in determining the minimum distance. Combinatorial approaches can be used to create codes with simple encoders. In this setting, in some recent studies ([31, 26, 16]), parity–check matrices H related to finite geometries are considered and such constructions present several advantages:

- the matrix H is regular and sparse;
- if the code derives from a partial linear space, then the Tanner graph (that coincides with the incidence graph of the finite geometry) does not have cycles of length 4;
- geometric properties may be immediately translated into structural properties of the code and allow to determine bounds (or the exact value, in some cases) for the rank of H and the minimum distance of

the code;

- the existence of some collineation groups is closely related to the cyclic or quasi–cyclic structure of the code.

I was introduced to this topic by my supervisor, Prof. G.Lunardon, and by Dr. Valentino Toschi of *NEC Electronics*, during his visit at the *Dipartimento di Matematica e Applicazioni "R.Caccioppoli"*. He showed us how the companies providing network solutions and communications services are interested in the $LDPC$ codes arising from finite geometries, since most of them seems to have very good parameters, and for their code–rate, which is often greater or equal than $239/255 \approxeq 0.937255$, the lower bound required for optical fiber communication.

Coding theory involves a lot of topics, such as geometry, statistic, information theory, engineering, and I have made an effort to give my contribute through my mathematical knowledge, always looking at the developments of the other fields of research, trying to take a direction consistent with them. So far, we have directed our research towards two directions.

In the second chapter, we have studied the quasi-cyclic structures of the codes arising from finite geometries, hence the structure of the circulant matrices, the "double" quasi-cyclic structure of some codes and how we can determine the starters, that is those few rows by which we can construct matrices of large dimension. Most of the results of this chapter are published in [44].

In the third chapter, we have studied, through geometrical means, the minimum distance of the codes deriving from finite geometries, improving known lower bounds, finding a new lower bound specific to the codes we have considered, determining when some of them is sharp and, finally, for some classes of codes, we have characterized some low weight codewords as linear combination of codewords of minimum weight. Part of this research was done during my visit at the Department of Pure Mathematics and Computer Algebra at Gent University and it is presented in [45].

Acknowledgements

After four years of Ph.D., there are many people I would like to thank. First of all, my supervisor, Professor Guglielmo Lunardon, for being my guide, for continuously teaching me a lot about Finite Geometry, for conveying me his love for Mathematics and for creating a so comfortable working environment around me. Then, I would like to thank all the people working in his research group at Università di Napoli Federico II and Seconda Università di Napoli, especially Laura Bader, Olga Polverino, Giuseppe Marino, for always supporting and encouraging me, but a special thank goes to Rocco Trombetti. I thank Luca Giuzzi and Giovanni Cutolo for the big help with *GAP*, but not only for that. I wish to thank Professor Leo Storme, for being my guide during my stay in Department of Pure Mathematics and Computer Algebra, in Gent. Finally, I would like to thank all the Ph.D. students and researchers of "Studio 39", but above all, my friends Alessandro and Michelangelo, for the nice time spent together during the last four years.

Chapter 1

Preliminaries

1.1 A brief review of Coding Theory

Let $V(n,q)$ be the numerical vector space of dimension n over the finite field $GF(q)$. A $[n,k]$ *linear code* $\mathbf{C}$ over $GF(q)$ is a k–dimensional vector subspace of $V(n,q)$. The main goal in coding theory is transmitting information over *noisy channel*. A simple model for this is the following:

$$x = \text{sent codeword} \rightarrow \boxed{\text{Noisy channel}} \rightarrow y = x + e, e = \text{error vector} \rightarrow$$

$$\rightarrow \widehat{x} = \text{decoded vector.}$$

A simple class of channel models is the class of *Discrete Memoryless Channel*, DMC for short, and consists of

- a discrete input alphabet X,
- a discrete output alphabet Y,
- a conditional probability mass function $P_{Y_i/X_i}(y_i/x_i)$ that tell us the probability of observing the output symbol y_i when the input symbol x_i has been sent,
- the fact that the transmission at different times indices is statistically independent, i.e., using $x := (x_1, \ldots, x_n)$ and $y := (y_1, \ldots, y_n)$ we

have

$$P_{Y/X}(y/x) = \prod_{i=1}^{n} P_{Y_i/X_i}(y_i/x_i).$$

Simple models in this class are the *Binary Symmetric Channel*(BSC) and the *Binary Erasure Channel* (BEC), but a more realistic model is the *Binary–Input Additive Gaussian Noise Channel* ($AWGNC$), which is a memoryless channel model such that

- X=$\{0, 1\}$,
- Y=$\mathbb{R}$, hence it is strictly speaking not a DMC,
- the conditional probability density function is

$$P_{Y_i/X_i}(y_i/x_i) = \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{(y_i - \overline{x_i})^2}{2\sigma^2})$$

where

$$\overline{x_i} := 1 - 2x_i := \begin{cases} +1, & \text{if } x_i = 0; \\ -1, & \text{if } x_i = 1. \end{cases}$$

Roughly speaking, we consider a $k-$dimensional space embedded in a $n-$vector space just for error correction; the integer $n - k$ is called the *redundancy* of the code and $\frac{k}{n}$, that is the bits necessary to send the messages over the bits actually used in the communication, is the *code–rate* of $\mathbf{C}$. A concept strictly related to the error correction capacity of the code is that of *minimum distance*. Let $\mathbf{x}$ and $\mathbf{y}$ be two codewords of $\mathbf{C}$; the *Hamming distance* $d(\mathbf{x}, \mathbf{y})$ is the number of the components in which they differ. The *minimum distance* d of $\mathbf{C}$ is the minimum of the set $\{d(\mathbf{x}, \mathbf{y}), \mathbf{x}, \mathbf{y} \in \mathbf{C}, \mathbf{x} \neq \mathbf{y}\}$. The *weight* $w(\mathbf{x})$ of a codeword $\mathbf{x}$ is $d(\mathbf{x}, \mathbf{0})$, where $\mathbf{0}$ is the zero of the vector space $V(n, q)$.

Lemma 1.1. *([19]) For a linear code* $\mathbf{C}$, $\min\{w(\mathbf{x}), \mathbf{x} \in \mathbf{C}\backslash\{\mathbf{0}\}\} = \min\{d(\mathbf{x}, \mathbf{y}), \mathbf{x}, \mathbf{y} \in \mathbf{C}, \mathbf{x} \neq \mathbf{y}\}$.

Therefore, suppose that codeword x has been transmitted and we receive the vector y which may have been distorted by noise; for an assumption that all the codewords are equally likely to be transmitted, the best

decision (*hard decision*) is to decode y as that codeword $\widehat{x}$ such that $d(\widehat{x}, y)$ is the smallest possible. Such decoding is called *Maximum–Likelihood decoding* ($ML-$ decoding). There is the following result that shows how many errors can be corrected in hard-decision, if we know the minimum distance.

Theorem 1.2. *([19]) Let d be the minimum distance of the code C; then C can correct up to t errors if $d \geq 2t + 1$.*

So far we have discussed hard decision schemes, but modern coding theory is based on *soft decision* algorithms, i.e. on iterative probability decoding algorithm, for example the *Sum–Product Algorithm* (SPA), which is based on belief propagation and assumes that the channel is $AWGNC$. We will not describe in details this aspect of the coding theory (we remind to [9],[15],[30],[40]), but we observe that

- most of the iterative decoding algorithms are based on probabilistic arguments, and in fact also the ML–decoding can be reformulated as follows:
$$\widehat{x} = \arg\max_{x \in C} P_{Y/X}(y/x) = \arg\min_{x \in C} \sum_{i=1}^{n} \lambda_i x_i$$
where we have used the *log–likelihood ratios* (LLRs)
$$\lambda_i := \log\left(\frac{P_{Y_i/X_i}(y_i/0)}{P_{Y_i/X_i}(y_i/1)}\right)$$
- if we assume that more than t errors may occur (with $d = 2t + 1$ or $d = 2t + 2$), then any deciding algorithm can not guarantee perfect transmission, but, from a theoretical point of view, the probability of lost information can be made as small as desired (see *Shannon's first theorem*, [55]). Anyway, the performance of the probabilistic algorithms is strictly related to the minimum weight codewords, but also to low weight–codewords in general.

The decoding procedures for linear codes are valuable. It is well known that any vector subspace of $V(n, q)$ can be determined by a finite set of homogenous equations, in other words by a (m, n)–matrix over $GF(q)$ H.

A matrix H such that $\mathbf{x}^T H = 0 \iff \mathbf{x} \in \mathbf{C}$ is called a *parity–check* matrix of $\mathbf{C}$; we usually refer to the components of $\mathbf{x}$ and to the rows of H as the *code bits* and the *check sums*, respectively. We say that a check sum h_i checks the code bit x_j if the $j-ht$ component of h_i is different than zero.

The parity–check matrix plays a fundamental role in decoding. First of all, we want recall a well known result about the connection between the parity–check matrix H and the minimum distance of a linear code.

Theorem 1.3. *([19]) Let $\mathbf{C}$ be a linear $[n,k]$–code over $GF(q)$ with parity–check matrix H. Then the minimum distance of C is d if and only if any $d-1$ columns of H are linearly independent but there are some d columns linearly dependent.*

This result is connected to the *One–step Majority Logic decoding*(see [34], [46]. A set of parity–check sums given by a parity–check matrix of a code is said to be orthogonal on a given code bit if each of the parity check–sums include the code bit but no other code bit is included in more than one of these parity–check sums. If for each code bit there are γ parity–check sums that are orthogonal on it, then the code is *majority-logic decodable* up to $\lfloor \frac{\gamma}{2} \rceil$ bit errors. Hence the minimum distance is at least $2\lfloor \frac{\gamma}{2} \rceil + 1$. A *Low Density Parity Check Code*, *LDPC* for short, is a binary code such that

1. each row has ρ non–zero components,
2. each column has γ non–zero components,
3. ant two columns have at most one non–zero common components,
4. H is a sparse matrix .

The *LDPC* codes have a remarkable importance in coding theory because they were the first codes to allow data transmission rates close to the theoretical maximum, i.e. the *Shannon limit* (see [55]).

A linear code $\mathbf{C}$ is *cyclic* if the right shift of $\mathbf{x} = (x_1, x_2, \ldots x_n) \in \mathbf{C}$, that is the codeword $(x_n, x_1 \ldots x_{n-1})$, is also a codeword of $\mathbf{C}$.

Let $F[x]$ be the ring of polynomials in the indeterminate x over the field $F := GF(q)$ and let R be the ring of polynomials of $F[x]$ modulo $x^n - 1$, that is $R \cong \frac{F[x]}{(x^n-1)}$, where $(x^n - 1)$ is the ideal generated by $x^n - 1$. Recall that every ideal I of R has the following form: $I \cong \frac{(f(x))}{(x^n-1)}$, where $f(x)$ is a divisor of $x^n - 1$.

Theorem 1.4. *A $[n, k]$ code $\mathbf{C}$ is isomorphic, as a vector space, to an ideal of R generated by a polynomial $f(x)$ of degree $n - k$; such polynomial is said to be the generator or the characterization polynomial of the code.*

A matrix H is *circulant* if every row is the right shift of the previous one.

Proposition 1.5. *A cyclic code $\mathbf{C}$ has a circulant parity check matrix and, conversely, if a code $\mathbf{C}$ has a circulant parity check matrix, then $\mathbf{C}$ is a cyclic code.*

A linear code $\mathbf{C}$ is said to be *quasi–cyclic* of index l if the right shift of l position of a codeword of $\mathbf{C}$ is also a codeword of $\mathbf{C}$; if $l = 1$, $\mathbf{C}$ is cyclic.

Proposition 1.6. *A code $\mathbf{C}$ with a parity–check matrix*

$$H = \begin{pmatrix} H_{11} & \dots & \dots & H_{1t} \\ \dots & \dots & \dots & \dots \\ H_{s1} & \dots & \dots & H_{st} \end{pmatrix}$$

such that every H_{ij} is circulant, is quasi–cyclic.

The encoding of quasi–cyclic codes can be implemented with linear feedback shift register based on their generator polynomials (see [46],[34]).

In order to study the relationship between the code bits and the check sums, especially referred to the iterative decoding algorithm, Tanner ([50]) introduced a bipartite graph, that is indeed called the *Tanner graph*. The Tanner graph consists of two levels of vertices: one level consists of vertices that represent the code bits, the other one consists of vertices that represent the check sums. No two vertices of the same level are connected and a code

bit–vertex x_i is connected by an edge to a check sum–vertex h_j if and only if h_j checks x_i. A *cycle* in a graph is a sequence of connected edges which starts and ends at the same vertex and no vertex, except the first and the last one, appears more than once. The number of the edges on a cycle is called the *length* of the cycle and the length of the shortest cycle is the *girth* of the graph. It is clear that a Tanner graph has girth at least 4. The performance of an iterative decoding algorithm, like the SPA decoding, is decreased by short cycles, especially the ones of length 4.

1.2 Finite-Geometry LDPC Codes

Most of the currently used $LDPC$ codes have been generated by computer with a random–matrix approach; encoding and decoding of these long computer–generated $LDPC$ codes, or determining the minimum distance, is quite complex due to the lack of code structure. In some recent studies [31, 26, 16] parity–check matrices related to finite geometries are considered. Such constructions present several advantages: geometric properties — for example the axioms of linear spaces — may be immediately translated into structural properties of the code, and these might be used to implement more efficient decoding algorithms; likewise, the nature of the automorphism group of the geometry itself is closely related to the cyclic or quasi–cyclic structure of the code. Furthermore, a geometric approach allows an easier description of the characteristics of the code and a direct estimates on the minimum distance.

In order to introduce $LDPC$ codes from finite geometries, a brief overview of the geometrical concepts we are going to use is needed. An incidence structure is a triple $\mathcal{I} = (\mathcal{P}, \mathcal{B}, I)$, where $\mathcal{P}$ is a set of *points*, $\mathcal{B}$ is a set of *blocks* and $I \subseteq \mathcal{P} \times \mathcal{B}$ is an *incidence relation*. We say that a point P and a block ℓ are incident (or, equivalently, that P is on ℓ or ℓ passes through P) when $(P, \ell) \in I$. A finite incidence structure $\mathcal{I}$ with $|\mathcal{P}| = n$ and $|\mathcal{B}| = m$ may be represented by a $(m \times n)$–matrix, say $H = [h_{ij}]$ over $GF(2)$ (or, equivalently, by H^T) whose rows are indexed by the blocks and

whose columns are indexed by the points, and such that $h_{ij} = 1$ when the i–th block is incident with the j–th point and $h_{ij} = 0$ otherwise.

If any two distinct blocks of $\mathcal{I}$ have at most one common point, $\mathcal{I}$ is called a *partial linear space* and the blocks are usually called *lines*. A *partial geometry* $pg(s, t, \alpha)$ is a partial linear space such that:

1. each line is incident with a constant number $s + 1$ of points;
2. each point is incident with a constant number $t + 1$ of lines;
3. for any non–incident point–line pair (P, ℓ) the number of lines incident with P and intersecting ℓ is exactly α.

The incidence matrix of a partial geometry is regular, in the sense that any row has constant weight $s + 1$ and any column has constant weight $t + 1$. Let $rank_2(H)$ be the rank of the incidence matrix $\mathbf{H}$ of a partial geometry $pg(s, t, \alpha)$ over the field $GF(2)$; in [26], it is shown that

$$rank_2(H) \leq \frac{st(t+1)(s+1)}{\alpha(t+s+1-\alpha)} + 1 \tag{1.1}$$

and if $t + s + 1 - \alpha \equiv 1 \bmod 2$, then

$$rank_2(H) \geq \frac{st(t+1)(s+1)}{\alpha(t+s+1-\alpha)}. \tag{1.2}$$

The *incidence graph* $\mathcal{G}$ of an incidence structure $\mathcal{I} = (\mathcal{P}, \mathcal{B}, I)$, is a graph having as vertices $\mathcal{P} \cup \mathcal{B}$, such that never two vertices both in $\mathcal{P}$ or in $\mathcal{B}$ are connected and two vertices $x \in \mathcal{P}$ and $y \in \mathcal{B}$ are connected if and only if $(x, y) \in I$. A *cycle* in $\mathcal{G}$ is a sequence of connected vertices which starts and ends at the same vertex and does not contain any other vertex more than once. The length of a cycle is the number of its vertices and the *girth* of $\mathcal{G}$ is the length of its shortest cycle (for more details, see [3]). The graph of a partial linear space does not have cycles of length 4; if α is greater than one, then $\mathcal{G}$ has $N_6 = \frac{1}{3}m(n-s-1)\binom{\alpha}{2}$ cycles of length 6 (see [26]). If $\alpha = 1$, then the partial geometry is called *generalized quadrangle* and the graph $\mathcal{G}$ is devoid of cycles of length 4 and 6, while it contains $N_8 = \frac{1}{4}m(n-s-1)ts$ cycles of length 8.

We say that a $LDPC$ code $\mathbf{C}$ derives from a finite geometry if $\mathbf{C}$ has a parity check matrix $\mathbf{H}$ that is the incidence matrix of a finite incidence structure $\mathcal{I} = (\mathcal{P}, \mathcal{B}, I)$, mostly of a linear space. If $\mathbf{H}$ is the incidence matrix of a partial geometry $pg(s, t, \alpha)$, then we can express some parameters of the code as function of the parameters of the partial geometry. The code $\mathbf{C}$ turns out to be a $[n, n - rank_2(\mathbf{H})]$ binary code, with $n = |\mathcal{P}|$ or $n = |\mathcal{B}|$, according to the fact that we have labeled the columns by the points or by the lines, respectively. Also we have that $rank_2(\mathbf{H})$ is given by (1.1) and (1.2). Furthermore, the following lemma is proved in [26].

Lemma 1.7. *Let* $\mathbf{H}$ *be the incidence matrix of a partial geometry* $pg(s, t, \alpha)$ *and let* $\mathbf{C}$ *be the code which has* $\mathbf{H}^T$ *as parity–check matrix; then,*

$$d_{\min} \geq \max \left\{ \frac{(t+1)(s+1-t+\alpha)}{\alpha}, \frac{2(s+\alpha)}{\alpha} \right\}. \tag{1.3}$$

The following also holds true.

Lemma 1.8. *Let* $\mathbf{H}$ *be the incidence matrix of a partial geometry* $pg(s, t, \alpha)$ *and* $\mathbf{C}$ *be the code which has* $\mathbf{H}$ *as parity–check matrix, then we have*

$$d_{\min} \geq \max \left\{ \frac{(s+1)(t+1-s+\alpha)}{\alpha}, \frac{2(t+\alpha)}{\alpha} \right\}. \tag{1.4}$$

Proof. The matrix $\mathbf{H}^T$ is the incidence matrix of the dual geometry of the partial geometry $pg(s, t, \alpha)$, that is a partial geometry $pg(t, s, \alpha)$. □

Finally, if any two columns shares at most one non–zero component (for example, if $\mathbf{H}$ is the incidence matrix of a partial linear space and we have labeled the columns by the points), then we have the *Massey's bound* ([40]); precisely

$$d_{min} \geq \gamma + 2$$

where γ is the weight of a column.

These results are expressed in terms of the parameters of the partial geometry, but they basically derive from formulas that use the values and the multiplicities of the eigenvalues of HH^T; using geometric arguments,

we can obtain stronger results about the minimum distance of the code and about the small weight codewords in general (see Chapter3).

It is easy to see that the Tanner graph of a code from a partial geometry $pg(s,t,\alpha)$ coincides with the incidence graph $\mathcal{G}$. Hence if $\alpha > 1$, then the Tanner graph has girth 6 and if $pg(s,t,\alpha)$ is a generalized quadrangle, then the Tanner graph has girth 8 and we know exactly how many cycles of minimum length it has.

One of the most important proprieties of these codes is that they are either cyclic or quasi–cyclic and this is due to the action of some cyclic collineation groups. A map of $\mathcal{P} \cup \mathcal{B}$ into itself is called *collineation* if it maps points into points, lines into lines and preserves the incidence relation. Let $G = \langle g \rangle$ be a cyclic collineation group of $\mathcal{I} = (\mathcal{P}, \mathcal{B}, I)$, acting semi–regularly on the points and on the blocks; that is the only element of G that fixes a point or a block is the identity. Let $G(P)$ (respectively $G(\ell)$) be a point orbit (respectively a line orbit) under the action of G and label the elements of $G(P)$ (respect. $G(\ell)$) so that $P_i = P^{g^{i-1}}$ (respectively $\ell_i = \ell^{g^{i-1}}$). In this way, we obtain an incidence matrix

$$H = \begin{pmatrix} H_{11} & \dots & \dots & H_{1t} \\ \dots & \dots & \dots & \dots \\ H_{s1} & \dots & \dots & H_{st} \end{pmatrix}$$

such that any H_{ij} (the incidence matrix of $G(P) \cup G(\ell)$ for some P and ℓ) is a circulant matrix. It is clear that, in order to construct the matrix $\mathbf{H}$, it is enough to know the first row of every H_{ij}, that is, it is enough to know the incidence of t lines ℓ_i, $i = 1, \dots, s$, such that $G(\ell_i) \neq G(\ell_j) \forall\ i \neq j$; we usually call such lines *starters*. The importance of the quasi–cyclic codes is well known: they have a linear time encoding (based on their characterization polynomials, see [46],[34]) and this is not shared by other $LDPC$ codes in general. But in the $LDPC$ codes deriving from finite geometries, studying the action of collineation groups we can also find, in some sense, a regular display of the submatrices H_{ij} and we can find a geometrical characterizations of the line–starters. This is, essentially, the aim of the Chapter2.

An example (see [31]) of LDPC codes deriving from a finite geometry is the following. Let V be a $(d+1)$–dimensional vector space over the finite field $GF(q)$, with q a prime power. The lattice of the subspaces of V forms the d–dimensional *Projective Geometry* $\Sigma = PG(d,q)$. The Singer group of Σ is a cyclic group $S = \langle \sigma \rangle$ of collineations of Σ acting regularly on points and hyperplanes, that is the only element of S that fixes a point or a hyperplane is the identity and S is transitive on points and hyperplanes of Σ; hence, its order is $\frac{q^{d+1}-1}{q-1}$.

Choose as $(d+1)$–dimensional vector space over $GF(q)$ the field $GF(q^{d+1})$ and let ξ be a primitive element of $GF(q^{d+1})$; the collineation σ is induced by the map

$$x \in GF(q^{d+1}) \rightarrow \xi x \in GF(q^{d+1}).$$

The action of S is, in general, not regular on the k–dimensional subspaces of Σ (see [10]), with $k \notin \{0, d-1\}$; more precisely, if d is even, then Σ acts regularly on the lines, while if d is odd, the subgroup of order $q+1$ fixes the *regular spread* $\mathcal{S}$ line-wise and the subgroup of order $\frac{q^{d+1}-1}{q^2-1}$ acts regularly on the lines of $\mathcal{S}$. Let $\mathbf{H}$ be the incidence matrix of the set of points and lines of Σ; labeling the points and the lines as discussed above, we see that $\mathbf{H}$ can be written in the following way:

$$H = \begin{pmatrix} H_1 \\ \dots \\ H_t \end{pmatrix}.$$

If d is even, then $t = \frac{q^d-1}{q^2-1}$ and H_i is a circulant square matrix of order $\frac{q^{d+1}-1}{q-1}$. If d is odd, then $t = \frac{q^d-q}{q^2-1}$, H_i, $i > 1$, is the incidence matrix of the set of points of Σ and the lines not in $\mathcal{S}$ (i.e. H_i is a circulant square matrix of order $\frac{q^{d+1}}{q^2-1}$) and H_1 is the incidence matrix of the points of Σ and the lines in $\mathcal{S}$ (i.e. $H_1 = (I_1 \dots I_{q+1})$ and I_j is the identity matrix of order $\frac{q^{d+1}-1}{q^2-1}$). Hence, the $LDPC$ code arising from the points and lines of Σ is a quasi–cyclic code. In Chapter 2, we go further and investigate the action of some subgroups of the Singer group on the points and lines of other incidence structures.

Chapter 2

Quasi–cyclic codes from Hermitian Varieties

2.1 Codes from the Hermitian Curve

Let $PG(2,q^2)$ the Desarguesian projective plane of order q^2 and represent it via $GF(q^6)^* mod GF(q^2)^*$, where $GF(q)^*$ is $GF(q)\backslash\{0\}$: we denote a point of $PG(2,q^2)$ by (a), where $a \in GF(q^6)^*$ is a representative of the equivalence class. Let $Tr : x \in GF(q^6) \mapsto x + x^{q^2} + x^{q^4} \in GF(q^2)$ be the trace function; in this representation, a line ℓ has equation $Tr(ux) = 0$, for some $u \in GF(q^3)^*$ and we denote ℓ by $[u]$.

The function

$$(x,y) \in GF(q^6) \times GF(q^6) \mapsto Tr(x^{q^3}y) \in GF(q^2)$$

is a non–degenerate *Hermitian* sesquilinear form, hence it induces a *unitary* polarity π of the plane such that

$$P^\pi = (u)^\pi = [u^{q^3}]$$

$$\ell^\pi = [u]^\pi = (u^{q^3}).$$

The absolute points of such polarity , that is the points of the Hermitian curve $\mathcal{H}(2,q^2)$, have equation $Tr(x^{q^3+1}) = 0$ and $|\mathcal{H}(2,q^2)| = q^3 + 1$. Every line of the plane is tangent or $q+1$–secant to $\mathcal{H}(2,q^2)$: in the former case

the line are said to be *totally isotropic*, in the latter one they are said *non–isotropic*. From now on, we denote the set of non isotropic lines by $\mathcal{L}$ and we recall that $|\mathcal{L}| = q^2(q^2 - q + 1)$.

Let ξ be a primitive element for $GF(q^6)$; the Singer group $S = \langle \sigma \rangle$ of $PG(2, q^2)$ has order $q^4 + q^2 + 1$ and it is the direct sum of two subgroups: $S_1 = \langle \sigma_1 \rangle$ of order $q^2 + q + 1$, with σ_1 given by

$$x \in GF(q^6) \to \xi^{q^2-q+1} x \in GF(q^6)$$

and $S_2 = \langle \sigma_2 \rangle$ of order $q^2 - q + 1$, with σ_2 given by

$$x \in GF(q^6) \to \xi^{q^2+q+1} x \in GF(q^6).$$

In [2] it is shown that the point orbits of these two subgroups give rise to two different cyclic partitions of $PG(2, q^2)$: a cyclic partition of $PG(2, q^2)$ into Baer subplanes

$$Baer(u) := \left\{ \xi^{u+i(q^2-q+1)} \mid \ i = 0, \ldots, q^2 + q \right\}$$

for $u = 0, \ldots, q^2 - q$ and a cyclic partition of $PG(2, q^2)$ into complete arcs

$$Arc(t) := \left\{ \xi^{t+i(q^2+q+1)} \mid \ i = 0, \ldots, q^2 - q \right\}$$

for $t = 0, \ldots, q^2 + q$. We recall that an arc A of $PG(2, q^2)$ is a set of points of $PG(2, q^2)$ no three collinear; A is said to be complete if it is not properly contained in any other arc. A Baer subplane B is a subplane of $PG(2, q^2)$ which is incident with every line of $PG(2, q^2)$, or, dually, such that every point of $PG(2, q^2)$ is incident with a line of B. It is well known that when the order of $PG(2, q^2)$ is q^2, then the order of B is q. Thus, for every line ℓ of $PG(2, q^2)$, there exists one and only one subplane $Baer(u)$ such that $|\ell \cap Baer(u)| = q + 1$; in this case we say that ℓ contains a Baer subline of $Baer(u)$; for any other $Baer(v) \neq Baer(u)$, we have $|\ell \cap Baer(v)| = 1$. Let $S_i(\ell)$ be the orbit of ℓ under the action of S_i; we note that if ℓ contains a subline of $Baer(u)$, then any other line of $S_1(\ell)$ does; likewise, if ℓ is tangent (respect. secant, external) to $Arc(t)$, then any other line of $S_2(\ell)$ is. Let $PGU(3, q)$ be the *Unitary Group*, that is the group of

linear collineations of the plane fixing$\mathcal{H}(2,q^2)$; in [7], it's shown that S_2 is a subgroup of $PGU(3,q)$, hence, there is a partition of $\mathcal{H}(2,q^2)$ into $q+1$ complete arcs and a partition of $\mathcal{L}$ into q^2 orbits under the action of S_2.

We are interested in the LDCP code arising from the *Unital design* (or *Classical Unital*) $\mathbf{D} = (\mathcal{H}(2,q^2), \mathcal{L}, I)$, where I is the natural incidence relation (for more details about the Unital design see [10] or [21]).

Proposition 2.1. *The code* $\mathbf{C}$ *arising from* $\mathbf{D}$ *is a* $[q^2(q^2-q+1), k]$ *LDPC code, with* $k = (q^2-q-1)(q^2-q+1)$ *if* q *is even, or* $k = (q^2-q)(q^2-q+1)$*, if* q *is odd. The code* $\mathbf{C}$ *is capable of correcting any error pattern with* $\lfloor \frac{q+1}{2} \rfloor$ *or fewer errors using one–step majority logic decoding, its minimum distance is at least* $q+2$*, its Tanner graph has girth* 6 *and precisely it contains* $\frac{1}{6}q^4(q^3+1)(q^2-1)$ *cycles of length* 6*.*

Proof. Let $\mathbf{H}$ be the incidence matrix of $\mathbf{D}$: $\mathbf{H}$ is a $q^3+1 \times q^2(q^2-q+1)$ matrix over $GF(2)$, such that every row has weight q^2 and every column has weight $q+1$, hence the density of $\mathbf{H}$ is $\frac{1}{q^2-q+1}$. Hiss in [19] proved that when q is even, then $rank_2(\mathbf{H}) = q^3+1$; if q is odd, then $rank_2(\mathbf{H}) = q(q^2-q+1)$, where $rank_2(\mathbf{H})$ is the rank of $\mathbf{H}$ over $GF(2)$. Let $\mathbf{C}$ be the code having $\mathbf{H}$ as parity check matrix: the first part of the assert is hence clear. Since any two rows of $\mathbf{H}$ have exactly one non–zero component in common (that is there is exactly one line through two points), the check sums can be used for majority–logic decoding of the code to correct any error pattern with at most $\lfloor \frac{\gamma}{2} \rfloor$ errors, where γ is the weight of the columns (see [31]). On the other hand, the fact that any two columns have at most one non–zero component in common (that is any two lines have at most one point in common), implies that the Tanner graph doesn't have cycles of length 4, but it has cycles of length 6. Indeed, geometrically, a cycle of length 6 is equivalent to the existence of triangle in the design $\mathbf{D}$. For every two points in a line ℓ there are $q^3+1-(q+1) = q^3-q$ triangles, hence for ℓ pass $\frac{q(q+1)}{2}$ triangles; in this way we have counted every triangle tree times, hence in $\mathbf{D}$ there are $\frac{1}{6}q^4(q^3+1)(q^2-1)$ triangles.

□

Now we show how the code $\mathbf{C}$ is quasi–cyclic and to find starters of such a code.

Proposition 2.2. *The code* $\mathbf{C}$ *is quasi–cyclic.*

Proof. Let $P_1, \ldots, P_{q+1}$ be points of $\mathcal{H}(2, q^2)$ and $\ell_1, \ldots, \ell_{q^2}$ be lines of $\mathcal{L}$ that have distinct orbits under the action of S_2. Label the elements of $\mathcal{H}(2, q^2)$ in the following way: $P^{((i-1)(q^2-q+1)+j+1)} := P_i^{\sigma_2^j}$, $i = 1, \ldots, q+1$, $j = 0, \ldots, q^2 - q$; analogously, for the lines let $\ell^{((i-1)(q^2-q+1)+j+1)} := \ell_i^{\sigma_2^j}$, $i = 1, \ldots, q^2$, $j = 0, \ldots, q^2 - q$. Hence, the incidence matrix may be written as

$$\mathbf{H} = \begin{pmatrix} H_{1,1} & \ldots & H_{1,q+1} \\ \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots \\ H_{q^2,1} & \ldots & H_{q^2,q+1} \end{pmatrix},$$

where $H_{i,j}$ is a circulant square matrix of order q^2-q+1 for any $i = 1, \ldots, q^2$ and $j = 1, \ldots, q+1$. □

Let now B_0 be $Baer(0)$ and $C = B_0 \cap \mathcal{H}(2, q^2)$. The following result shows how to find points of $\mathcal{H}(2, q^2)$ and lines of $\mathcal{L}$ which have distinct orbits under the action of S_2, that is, respectively, *point–starters* and *line–starters*.

Proposition 2.3. *The points of* $\mathbf{C}$ *are point–starters and the lines* $\{[u] \mid (u) \in B_0 \backslash C\}$ *are line–starters; in particular, if* q *is even, then the set of line–starters* $\{[u] \mid (u) \in B_0 \backslash C\}$ *is* $\{[u]$ not passing through (1) and $(u) \in B_0\}$.

Proof. In [2], it is proved that $|Arc(0) \cap Baer(u)| = 1$ for any $u = 0, \ldots, q^2 - q$; the same proof can also be used to show that $|Arc(t) \cap Baer(u)| = 1$, for any $u = 0, \ldots, q^2-q$ and $t = 0, \ldots, q^2+q$, hence the points of $\mathbf{C}$ are starters. Using duality, we see that $S_1(\ell) \cap S_2(m)$ consists of one line, for every two lines ℓ and m of $\mathcal{L}$, hence the lines of $S_1(\ell) \cap \mathcal{L}$ are starters. The non–isotropic lines are $\{\ell = P^\pi \mid P \text{ non–absolute}\} = \{[u^{q^3}] \mid (u) \notin \mathcal{H}(2, q^2)\}$. If $\ell = [1]$, then $S_1(\ell) \cap \mathcal{L} = \{[u] \mid (u) \in B_0\} \cap \{[u^{q^3}] \mid (u) \notin \mathcal{H}(2, q^2)\} =$

$\{[u] \mid (u) \in B_0 \setminus C\}$. If q is even, then $\{[u] \mid (u) \in B_0 \setminus C\} = \{[u] \mid (u) \in B_0$ and $Tr(u^{q^3+1}) \neq 0\} = \{[u] \mid (u) \in B_0$ and $Tr(u^2) \neq 0\} = \{[u] \mid (u) \in B_0$ and $Tr(u) \neq 0\} = \{[u] \mid (u) \in B_0$ and$(1) \notin [u]\}$. □

A circulant matrix $H_{i,j}$ is the incidence matrix of an arc $Arc(t) \subseteq \mathcal{H}(2, q^2)$ with respect to a $S_2(\ell)$, with $\ell \in \mathcal{L}$, hence every row of $H_{i,j}$ has weight at most 2. Our next step is to describe how the $q+1$ points of any given line are arranged in the $q+1$ arcs, hence providing description of the submatrices H_{ij}'s. This kind of information is useful in case of truncation of the matrix $\mathbf{H}$: for applicative purpose, in some cases, not the whole matrix $\mathbf{H}$, but a part of it can be used as parity check matrix for a code. The truncation may consists in throwing away some lines or rows of the matrix (see the next subsection) or a few string of $H_{i,j}$'s; in the latter case, we preserve a quasi–cyclic structure.

The following two lemmas are proved in [12].

Lemma 2.4. *If q is even, then $Baer(u) \cap \mathcal{H}(2, q^2)$ is a subline of $Baer(u)$; if q is odd, then $Baer(u) \cap \mathcal{H}(2, q^2)$ is a subconic of $Baer(u)$, for $u = 0, \ldots, q^2 - q$.*

Remark 2.5. *A* conic *of the plane $PG(2, q)$ is, for q odd, the set of the points represented by vectors v in the underlying vector space, say $V(3, q)$, which annihilate a non–degenerate quadratic form $Q(v)$. All conics of $PG(2, q)$ are projectively equivalent and the group of collineations fixing any given one is the* Projective Orthogonal Group *$PO(3, q)$. We observe that $PO(3, q)$ is isomorphic to the group $PGL(2, q)$ of linear collineations of the line in its 3–transitive permutation representation.*

Lemma 2.6. *If ℓ contains a subline of $Baer(u)$, then ℓ is tangent to the $q+1$ arcs containing the $q+1$ points of $\ell \cap Baer(u)$.*

We can now prove the following theorem.

Theorem 2.7. *If q is even, then there exists a unique orbit $S_2(\ell)$ such that any line of $S_2(\ell)$ is tangent to $Arc(t)$, for any $Arc(t) \in \mathcal{H}(2, q^2)$;*

any other orbit $S_2(m)$ is such that any of its lines is tangent to the same, unique, arc of $\mathcal{H}(2,q^2)$ and secant to the same $\frac{q}{2}$ arcs of $\mathcal{H}(2,q^2)$. If q is odd, then there exist $\frac{1}{2}q(q+1)$ orbits $S_2(\ell)$ such that any line of $S_2(\ell)$ is tangent to the same two arcs of $\mathcal{H}(2,q^2)$ and secant to the same $\frac{(q-1)}{2}$ arcs of $\mathcal{H}(2,q^2)$; furthermore, there exist $\frac{1}{2}q(q-1)$ orbits $S_2(m)$ such that any line of $S_2(m)$ is secant to the same $\frac{q+1}{2}$ arcs of $\mathcal{H}(2,q^2)$.

Proof. Consider in any $S_2(\ell)$, the unique line that contains a subline of B_0, namely ℓ. If q is even, than $C = B_0 \cap \mathcal{H}(2,q^2)$ is a subline of B_0 and two cases may occur:

1. $\ell \cap C = C$; hence, by Lemma 2.6, the line ℓ is tangent to any arc of $\mathcal{H}(2,q^2)$ as well as any other line of $S_2(\ell)$;

2. $|\ell \cap C| = 1$; hence, ℓ is tangent to a unique arc of $\mathcal{H}(2,q^2)$ and any other line of $S_2(\ell)$ is tangent to the same arc.

When q is odd, $C = B_0 \cap \mathcal{H}(2,q^2)$ is a subconic and we have, again, two possible cases:

1. $|\ell \cap C| = 2$; thus, ℓ is tangent to two arcs of $\mathcal{H}(2,q^2)$ and any other line of $S_2(\ell)$ is tangent to the same arcs;

2. otherwise, $\ell \cap C = \emptyset$; hence, ℓ is not tangent to any arc of $\mathcal{H}(2,q^2)$ as well as any other line of $S_2(\ell)$.

□

We can finally prove the double cyclic structure of the code $\mathbf{C}$; that is, we write a circulant display of submatrices H_{ij}.

Theorem 2.8. *If q is even, then*

$$\mathbf{H} = \begin{pmatrix} I_1 & I_2 & \dots & \dots & \dots & I_{q+1} \\ A_{1,1} & A_{1,2} & \dots & \dots & \dots & A_{1,q+1} \\ A_{1,q+1} & A_{1,1} & \dots & \dots & \dots & A_{1,q} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A_{1,2} & A_{1,3} & \dots & \dots & \dots & A_{1,1} \\ A_{2,1} & A_{2,2} & \dots & \dots & \dots & A_{2,q+1} \\ A_{2,q+1} & A_{2,1} & \dots & \dots & \dots & A_{2,q} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A_{2,2} & A_{2,3} & \dots & \dots & \dots & A_{2,1} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A_{q-1,1} & A_{q-1,2} & \dots & \dots & \dots & A_{q-1,q+1} \\ A_{q-1,q+1} & A_{q-1,1} & \dots & \dots & \dots & A_{q-1,q} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A_{q-1,2} & A_{q-1,3} & \dots & \dots & \dots & A_{q-1,1} \end{pmatrix},$$

where I_i is the identity matrix of order q^2-q+1, for $i = 1, \dots, q+1$, while A_{ij} is a square circulant matrix of order $q^2 - q + 1$, for $i = 1, \dots, q-1$ and $j = 1, \dots, q+1$.

If q is odd and $r = q+1$, then

$$\mathbf{H} = \begin{pmatrix}
A_1 & A_2 & \dots & \dots & \dots & \dots & \dots & A_r \\
B_{1,1} & B_{1,2} & \dots & B_{1,\frac{r}{2}} & B_{1,\frac{r}{2}+1} & B_{1,\frac{r}{2}+2} & \dots & B_{1,r} \\
B_{1,\frac{r}{2}} & B_{1,1} & \dots & B_{1,\frac{r}{2}-1} & B_{1,r} & B_{1,\frac{r}{2}+1} & \dots & B_{1,r-1} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
B_{1,2} & B_{1,3} & \dots & B_{1,1} & B_{1,\frac{r}{2}+2} & B_{1,\frac{r}{2}+3} & \dots & B_{1,\frac{r}{2}+1} \\
B_{2,1} & B_{2,2} & \dots & B_{2,\frac{r}{2}} & B_{2,\frac{r}{2}+1} & B_{2,\frac{r}{2}+2} & \dots & B_{2,r} \\
B_{2,\frac{r}{2}} & B_{2,1} & \dots & B_{2,\frac{r}{2}-1} & B_{2,r} & B_{2,\frac{r}{2}+1} & \dots & B_{2,r-1} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
B_{2,2} & B_{2,3} & \dots & B_{2,1} & B_{2,\frac{r}{2}+2} & B_{2,\frac{r}{2}+3} & \dots & B_{2,\frac{r}{2}+1} \\
C_{1,1} & C_{1,2} & \dots & \dots & \dots & \dots & \dots & C_{1,r} \\
C_{1,r} & C_{1,1} & \dots & \dots & \dots & \dots & \dots & C_{1,r-1} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
C_{1,2} & C_{1,3} & \dots & \dots & \dots & \dots & \dots & C_{1,1} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
C_{q-2,1} & C_{q-2,2} & \dots & \dots & \dots & \dots & \dots & C_{q-2,r} \\
C_{q-2,r} & C_{q-2,1} & \dots & \dots & \dots & \dots & \dots & C_{q-2,r-1} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
C_{q-2,2} & C_{q-2,3} & \dots & \dots & \dots & \dots & \dots & C_{q-2,1}
\end{pmatrix},$$

where A_i, $i = 1, \dots, r$, B_{ij}, $i = 1, 2$ and $j = 1, \dots, r$, C_{ij}, $i = 1, \dots, q-2$ and $j = 1, \dots, r$ are all suitable square circulant matrices of order $q^2 - q + 1$.

Proof. The set $\mathbf{C}$ is either a line or a conic. Since $PO(3,q) \simeq PGL(2,q)$, there exists in both cases a cyclic group $T = \langle \tau \rangle$ fixing $\mathbf{C}$ and isomorphic to the Singer group of the line (that is $|T| = q+1$ and T acts regularly on the points of $\mathbf{C}$). Since $\mathbf{C}$ is contained in $\mathcal{H}(2,q^2)$, the group T is a subgroup of $PGU(3,q)$. If $\mathbf{C}$ is a line, then T fixes $\mathbf{C}$ and also the point $C^\pi = (1) \in B_0$. In this case, the other line–orbits in B_0 have all order $q+1$. On the other hand, when $\mathbf{C}$ is a conic of B_0, we may, by Remark 2.5, assume it to have equation $y^2 = xz$. We consider the isomorphism ϕ (see [21]) between $PO(3,q)$ and $PGL(2,q)$ which associates to $f \in PGL(2,q)$,

represented by the matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the collineation $\phi(f) \in PO(3,q)$ represented by

$$\phi'(A) = \begin{pmatrix} d^2 & 2cd & c^2 \\ bd & ad+bc & ac \\ b^2 & 2ab & a^2 \end{pmatrix}.$$

Let now ξ be a primitive element of $GF(q^2)$ with minimal polynomial $x^2 - x + \eta$, where η is a primitive element of $GF(q)$. Such an element always exists, see [6]. Then, a collineation spanning the Singer group of the line may be represented by

$$S = \begin{pmatrix} 0 & -\eta \\ 1 & 1 \end{pmatrix}.$$

Hence, τ is represented, via ϕ, by

$$S' = \phi'(S) = \begin{pmatrix} 1 & 2 & 1 \\ -\eta & -\eta & 0 \\ \eta^2 & 0 & 0 \end{pmatrix}.$$

Observe that $\det(S' - \lambda I) = (\eta - \lambda)(\lambda^2 + (2\eta - 1)\lambda + \eta^2)$. The irreducibility of $x^2 - x + \eta$ over $GF(q)$ implies that the polynomial $x^2 + (2\eta - 1)x + \eta^2$ is irreducible over $GF(q)$; hence, S' has one eigenvalue $\lambda = \eta$ with eigenspace $V(\eta) = \langle e \rangle = \langle (2, -1, 2\eta) \rangle$. In particular, the group T fixes the point $P = \langle e \rangle$ and the line $\ell = P^\pi$, that is the line of equation $\eta x + y + z = 0$. Let now $Q = \langle v \rangle$ where $v = (x, y, -\eta x - y)$ is a point on ℓ; then, $S'v^T = ((1-\eta)x + y, -\eta(x+y), \eta^2 x)^T$; hence, S' induces on ℓ a collineation represented by the matrix

$$\begin{pmatrix} 1-\eta & 1 \\ -\eta & -\eta \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -\eta & 0 \end{pmatrix}^2;$$

on the other hand,

$$\begin{pmatrix} 1 & 1 \\ -\eta & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} S \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This means that T induces on ℓ a group of collineations isomorphic to the subgroup of order $\frac{q+1}{2}$ of the Singer group of the line — hence, there are two point orbits on ℓ of order $\frac{q+1}{2}$. Any point not on ℓ and distinct from P can be written in unique way as $\langle v+e\rangle$, with $\langle v\rangle \in \ell$; we have that $\tau^k(\langle v+e\rangle) = \langle \xi^{2k}v + \eta^k e\rangle = \langle \xi^{2k}v + \xi^{k(q+1)}e\rangle = \langle v+e\rangle$ if $k \geq q+1$; hence, the orbit of any such point has order $q+1$. By duality, there are two line orbits of order $\frac{q+1}{2}$ and $q-1$ line orbits of order $q+1$. One of these is the orbit of the tangent lines to the conic, which do not occur in the incidence structure we are considering. Let now P be a fixed point of $\mathbf{C}$ and label the point starters in the following way: $P_{i+1} := P^{\tau^i}$, $i = 0, \ldots, q$. If q is even, then there is a line starter, say ℓ_0 that contains the subline $\mathbf{C}$ and there are $q-1$ line starters, say $\ell_1, \ldots, \ell_{q-1}$ that contain sublines with distinct orbits under the action of T. In this case label the line starters as follows: $\ell^{(0)} := \ell_0$, and $\ell^{((i-1)(q+1)+j+1)} := \ell_i^{\tau^j}$, $i = 1, \ldots, q-1$ and $j = 0, \ldots, q$. If q is odd, then there is one line starter, say ℓ_0, that contains the unique subline fixed by T; two further line starters, say ℓ_1,ℓ_2, that contain sublines with distinct orbits of order $\frac{q+1}{2}$ under the action of T; and, finally, there are $q-2$ line starters, say $\ell_3, \ldots, \ell_q$, that contain sublines with distinct orbits of order $q+1$ under the action of T; hence, we may label the line starters as follows: $\ell^{(0)} := \ell_0$, $\ell^{((i-1)(\frac{q+1}{2})+j+1)} := \ell_i^{\tau^j}$, $i = 1, 2$ and $j = 0, \ldots, \frac{q-1}{2}$, $\ell^{((i-1)(q+1)+j+1)} := \ell_i^{\tau^j}$, $i = 3, \ldots, q$ and $j = 0, \ldots, q$. □

Remark 2.9. *We explicitly observe that if q is even, then it is enough to know the incidence of $q-1$ lines of $S_1([1]) \setminus \{[1]\}$ passing through a fixed point of $\mathbf{C}$ and not passing through (1) in order to construct the matrix $\mathbf{H}$.*

2.1.1 Extended and shortened codes

The *code–rate* of an $[n,k]$ linear code is the number $\frac{k}{n}$; hence the code $\mathbf{C}$ has code–rate $\frac{q^2-q-1}{q^2}$, for q even, or $\frac{q-1}{q}$, for q odd. Observe that, in general, the higher is q the higher the code–rate turns out to be; however, high values of q imply high complexity of calculus and, sometimes, overly long codes.

It is possible to obtain new good LDPC codes by extending and shortening in suitable ways finite–geometry codes. As it has been shown in [16], these new codes will have the same Tanner graph girth as the ones given before and have higher code–rate. However, they will lack some of the regularity and the quasi–cyclic structure.

Column (or row) splitting is one technique employed in [16] for modifying codes from finite geometries; in particular it has been applied in [25] to the classical unital $(\mathcal{H}(2,q^2),\mathcal{L},I)$. If we split any column of $\mathbf{H}$ in s columns of lower weight, then we obtain a new code from $\mathbf{C}$, say $\mathbf{C}_{ext}$, with length sn and the same number of linearly independent check–sums; hence, in this case, $\mathbf{C}_{ext}$ has code–rate $\frac{sq^2-q-1}{sq^2}$ for q is even, or $\frac{sq-1}{sq}$ for q is odd.

Let H' be the matrix obtained by $\mathbf{H}$ deleting a row which corresponds to a point P of $\mathcal{H}(2,q^2)$. The Tanner graph of the code, say $\mathbf{C}^1_{sh}$, which has H' as parity–check matrix, has $N'_6 = N_6 - \frac{1}{2}(n-1)(n-q-1)$ cycles of length 6, with $n = q^3+1$. We can also delete a column and $q+1$ rows of $\mathbf{H}$, that is a line and the $q+1$ points on it; we thus obtain a code, say $\mathbf{C}^2_{sh}$, whose Tanner graph has $N''_6 = \frac{1}{3}(n-q-1)(m-q^2-1)\binom{q+1}{2}$ cycles of length 6, with $n = q^3+1$ and $m = q^2(q^2-q+1)$. Finally, deleting a row and q^2 columns of $\mathbf{H}$, that is a point and the q^2 lines through it, we obtain a code, say $\mathbf{C}^3_{sh}$, whose Tanner graph has $N'''_6 = \frac{1}{3}(n-1)(m-2q^2+1)\binom{q}{2}$ cycles of length 6 with $n = q^3+1$ and $m = q^2(q^2-q+1)$.

Actually, if we are looking for gains in the code–rate, just the codes $\mathbf{C}_{ext}$ and $\mathbf{C}^1_{sh}$ are interesting. In the following tables, we compare the code–rate of the code $\mathbf{C}$ and the code–rate of the codes obtained by $\mathbf{C}$ either shortening or extending $\mathbf{H}$:

	$\mathbf{C}$	$\mathbf{C}_{ext}$	$\mathbf{C}^1_{sh}$	$\mathbf{C}^2_{sh}$	$\mathbf{C}^3_{sh}$
q even	$\frac{q^2-q-1}{q^2}$	$\frac{sq^2-q-1}{sq^2}$	$\frac{(q-1)^2}{q^2-q+1}$	$\geq \frac{q^4-2q^3-q^2-q-1}{(q-1)(q^3+q+1)}$	$\geq \frac{q-2}{q-1}$
q odd	$\frac{q-1}{q}$	$\frac{sq-1}{sq}$	$\geq \frac{q-1}{q}$	$\geq \frac{(q^2-q+1)(q^2-q-1)}{(q^3+q+1)(q-1)}$	$\geq \frac{q^3-2q^2+q-1}{q^2(q-1)}$

	$\mathbf{C}$	$\mathbf{C}_{ext}$	$\mathbf{C}^1_{sh}$	$\mathbf{C}^2_{sh}$	$\mathbf{C}^3_{sh}$
$q=4, s=2$	0.6875	0.84375	0.69231	≥ 0.51691	≥ 0.66667
$q=8, s=2$	0.85937	0.92969	0.85965	≥ 0.82232	≥ 0.85714
$q=5, s=2$	0.8	0.9	≥ 0.8	≥ 0.76140	≥ 0.79
$q=7, s=2$	0.85714	0.92857	≥ 0.85714	≥ 0.83713	≥ 0.85374

2.1.2 The construction of H for q even

In this subsection we shall confine ourselves to the case when $q \neq 4$ is even, which features a geometric characterization of the line starters, determining a strongly simplification in computing the matrix $\mathbf{H}$; the case when q is odd can be dealt with by means of standard calculations.

By [6], there is a primitive polynomial over $GF(q^2)$, with $q \neq 4$, of the form $f(t) = t^3 + \beta t + \alpha$; let ξ be a root of $f(t)$ in $GF(q^6)$; then, ξ is a primitive element of $GF(q^6)$ and $Tr(\xi) = 0$. Choose the set $\{1, \xi, \xi^2\}$ as a basis of the three–dimensional vector space $GF(q^6)$ over $GF(q^2)$. Let $P = (x)$ be a point of $PG(2, q^2)$ and take $\ell = [u]$ as a line of the same plane, with $x = x_0 + x_1\xi + x_2\xi^2$ and $u = u_0 + u_1\xi + u_2\xi^2 \in GF(q^6)$; from now on we suppose $P = \langle (x_0, x_1, x_2) \rangle$ and we denote by ℓ the set of points (x_0, x_1, x_2) of equation $u_0x_0 + \alpha u_2x_1 + \alpha u_1x_2 = 0$. Take the Hermitian curve $\mathcal{H}(2, q^2)$ of equation $Tr(x^{q^3+1}) = x_0^{q+1} + Tr(\xi^{q^3+1})x_1^{q+1} + Tr(\xi^{2q^3+1})x_1x_2^q + Tr(\xi^{q^3+2})x_1^q x_2 + Tr(\xi^{q^3+1})^2 x_2^{q+1} = 0$; suppose ξ^{q^3} to be $(0, \lambda, \mu)$; then, by a straightforward calculation, we get $Tr(\xi^{q^3+1}) = \mu\alpha$ and $Tr(\xi^{q^3+2}) = \lambda\alpha$; hence, the equation of $\mathcal{H}(2, q^2)$ is

$$x_0^{q+1} + \mu\alpha x_1^{q+1} + (\lambda\alpha)^q x_1x_2^q + \lambda\alpha x_1^q x_2 + (\mu\alpha)^2 x_2^{q+1} = 0. \qquad (2.1)$$

The collineation σ that spans the Singer group of $PG(2, q^2)$ is induced by the non–singular linear application

$$x = x_0 + x_1\xi + x_2\xi^2 \in GF(q^6) \mapsto x\xi = \alpha x_2 + (x_0 + x_2\beta)\xi + x_1\xi^2 \in GF(q^6)$$

and may be represented by the matrix

$$S = \begin{pmatrix} 0 & 0 & \alpha \\ 1 & 0 & \beta \\ 0 & 1 & 0 \end{pmatrix}$$

Let $S_i = \langle \sigma_i \rangle$, for $i = 1, 2$, be the two subgroups of the Singer group described in Section 2.1; then, σ_1 is represented by S^{q^2-q+1} and σ_2 by S^{q^q+q+1}.

The set $C = \mathcal{H}(2, q^2) \cap Baer(0)$, is a Baer subline of the line $[1]$; hence, $\mathbf{C}$ has to satisfy the following conditions:

$$\begin{cases} x_0 = 0 \\ x_0^{q+1} + \mu\alpha x_1^{q+1} + (\lambda\alpha)^q x_1 x_2^q + \lambda\alpha x_1^q x_2 + (\mu\alpha)^2 x_2^{q+1} = 0. \end{cases} \tag{2.2}$$

We can assume $x_2 = 1$; so x_1 must satisfy:

$$\mu\alpha x_1^{q+1} + (\lambda\alpha)^q x_1 + \lambda\alpha x_1^q + (\mu\alpha)^2 = 0 \Leftrightarrow$$

$$(\mu\alpha x_1 + \lambda\alpha)^{q+1} + (\lambda\alpha)^{q+1} + (\mu\alpha)^3 = 0.$$

Since (2.1) is a non–degenerate Hermitian curve, $(\lambda\alpha)^{q+1} + (\mu\alpha)^3 \in GF(q) \setminus \{0\}$. Furthermore, α is a primitive element of $GF(q^2)$; hence, we can assume $(\lambda\alpha)^{q+1} + (\mu\alpha)^3 = \alpha^{k(q+1)}$, for a suitable $k \in \{1, \ldots, q-1\}$. Thus, we have

$$x_1 = \lambda\mu^{-1} + \mu^{-1}\alpha^{k-1+i(q-1)}, i = 1, \ldots, q-1.$$

If τ spans the group T (see the proof of Theorem 2.8), then the action of τ on $\mathbf{C}$ is given by:

$$x_1 \mapsto \lambda\mu^{-1} + \alpha^{q-1}(\lambda\mu^{-1} + x_1);$$

thus, τ is represented by the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \alpha^{q-1} & \lambda\mu^{-1}(1+\alpha^{q-1}) \\ 0 & 0 & 1 \end{pmatrix}.$$

We deduce that the line starters, which have distinct orbits under the action of T, are exactly those lines ℓ which, fixed a point $P = \langle (0, \overline{x}_1, 1) \rangle$ in $\mathbf{C}$, satisfy the following conditions:

1. $\ell \in S_1([1]) \setminus \{[1]\}$;
2. ℓ passes through P and not $\langle (1, 0, 0) \rangle$, that is a line ℓ of equation of the form $x_0 + ux_1 + u\overline{x}_1 x_2 = 0$, with $u \neq 0$.

2.2 Codes from the Hermitian Surface

Let $PG(3,q^2)$ be the tree–dimensional projective space over the finite field $GF(q^2)$ and represent it via the vector space $V = GF(q^6) \bigoplus GF(q^2) = \{(u,u_0), u \in GF(q^6), u_0 \in GF(q^2)\}$; let $Tr : x \in GF(q^6) \mapsto x + x^{q^2} + x^{q^4} \in GF(q^2)$ be the trace function, hence a plane has equation $Tr(ux)+u_0x_0 = 0$ for some $(u,u_0) \in V \backslash \{0\}$.

The function

$$((x,x_0),(y,y_0)) \in V \times V \mapsto Tr(x^{q^3}y) + x_0^q y_0 \in GF(q^2). \qquad (2.3)$$

is a non–degenerate *Hermitian* sesquilinear form, hence it induces a *unitary* polarity of $PG(3,q^2)$, such that the *polar plane* of a point $P = (u,u_0)$ has equation $Tr(u^{q^2}x)+u_0^q x_0 = 0$ and the absolute points, that is the points of the Hermitian surface $\mathcal{H}(3,q^2)$, have equation

$$Tr(x^{q^{3+1}}) + x_0^{q+1} = 0.$$

A line ℓ of $PG(3,q^2)$ may be either tangent, $q+1$–secant or contained in $\mathcal{H}(3,q^2)$; from now on, we denote the set of the lines contained in $\mathcal{H}(3,q^2)$ by $\mathcal{L}$. We recall that $|\mathcal{H}(3,q^2)| = (q^3+1)(q^2+1)$ and $|\mathcal{L}| = (q^3+1)(q+1)$. The incidence structure $(\mathcal{H}(3,q^2),\mathcal{L},I)$, with I the natural incidence relation, is a *generalized quadrangle* (for more details about generalized quadrangles see [43]) with parameters (q^2,q), that is every line ℓ contains q^2+1 points and every point P is contained in $q+1$ lines, which are the lines for P contained in the polar plane of P.

The group of linear collineations preserving $\mathcal{H}(3,q^2)$ is the *Unitary Group* $\Gamma = PGU(4,q)$. The group Γ is transitive on the points and the lines of $\mathcal{H}$, that is for every couple of points P and Q (respect. for every couple of lines l and m) of $\mathcal{H}$, there is an element γ of Γ, such that $P^\gamma = Q$ (respect. $l^\gamma = m$).

Proposition 2.10. *Let* $\mathbf{H}$ *be the incidence matrix of the generalized quadrangle* $\mathcal{H}(3,q^2)$. *The code* $\mathbf{C}$ *having* $\mathbf{H}$ *as parity check matrix is a* $[(q^3+1)(q^2+1),k]$ *LDPC code, with* $k \geq (q^2-q)(q^3+1)$ *if* q *is even, or*

$k = (q^2 - q)(q^3 + 1)$ if q is odd. The minimum distance is $2(q + 1)$ and its Tanner graph has girth 8.

Proof. The matrix $\mathbf{H}$ is a regular $(q^3+1)(q+1) \times (q^3+1)(q^2+1)$ matrix over $GF(2)$ with density $\frac{1}{q^3+1}$; the rank of $\mathbf{H}$ over $GF(2)$ is at most $q^4 + q^2 + 1$ and it is $q^4 + q^2 + 1$ for odd q (see [1]). The Hermitian surface $\mathcal{H}(3, q^2)$ is a generalized quadrangle and this means that it does not contain triangles; this is equivalent to have a Tanner graph without cycles of length 3, hence its girth is at least 4. Finally, in [45] it is proven by geometrical arguments that the minimum distance is $2(q + 1)$. □

Like in the previous section, we are interested in the partition of $\mathbf{H}$ in circulant submatrices, that is in the cyclicity of the code $\mathbf{C}$ and in finding a geometrical characterization of the point–starters.

Let P be the point $(0, 1)$ and π_∞ the plane of equation $x_0 = 0$, that is the polar plane of P. The sesquilinear form (2.3) induces on π_∞ a non–degenerate unitary polarity, hence $\mathcal{H}(3, q^2)$ intersects π_∞ in the Hermitian curve $\mathcal{H}_0$ of equation $Tr(x^{q^{3+1}}) = 0$. A line joining P and a point Q of π_∞ is either tangent or $q + 1$–secant, according to the fact that Q is in $\mathcal{H}_0$ or not, respectively.

Let ξ be a primitive element for $GF(q^6)$, let M be $(q^2 + q + 1)$, N be $(q^2 - q + 1)$ and $\eta = \xi^{MN}$ a primitive element of $GF(q^2)$. Let S be the cyclic group of collineation spanned by:

$$(u, u_0) \in V \mapsto (u\xi^{(q-1)M}, u_0) \in V,$$

hence S has order N, it is a subgroup of the Unitary Group Γ and it acts semi–regularly on the points and lines of $\mathcal{H}$, that is the only collineation of S that fixes a point or a line of $\mathcal{H}$ is the identity. The group S fixes the point P and the plane π_∞, on which induces a cyclic partition in Kestenband arcs (for the definitions and results on this topic see the previous section), $q + 1$ of which are contained in $\mathcal{H}_0$. Let $\mathcal{K}(\mathcal{A})$ be a cone with vertex P and base the Kestenband arc $\mathcal{A}$ of the plane π_∞: if $\mathcal{A}$ is contained in $\mathcal{H}_0$, then $\mathcal{K}(\mathcal{A}) \cap \mathcal{H}(3, q^2) = \mathcal{A}$; if $\mathcal{A} \cap \mathcal{H}_0 = \emptyset$, then $\mathcal{K}(\mathcal{A}) \cap \mathcal{H}(3, q^2)$ consists in $(q^3 + 1)$ points not in π_∞.

Lemma 2.11. *If ℓ is a line of $\mathcal{H}(3,q^2)$ and Q is a point of ℓ, then $|\ell \cap S(Q)| \leq 2$.*

Proof. If Q is a point of $\mathcal{H}_0$, then $S(Q)$ is contained in π_∞; since the line ℓ can intersect π_∞ in at most one point, we have $Q = S(Q) \cap \ell$. If $Q = (u, u_0)$ is a point of $\mathcal{H} \setminus \pi_\infty$, then a line ℓ through Q intersects π_∞ in the point $R = (v, 0)$ and $\ell =< (u, u_0), (v, 0) >$. The orbit $S(Q)$ is $\{(u\xi^{(q-1)Mi}, u_0), i = 0, \ldots, q^2 - q\}$, hence finding the intersection $\ell \cap S(Q)$ is equivalent to find a λ in $GF(q^2)\backslash\{0\}$ such that $(u, u_0) + \lambda(v, 0) \cong (u\xi^{(q^2-1)Mi}, u_0)$ for some i. In the plane π_∞, this is equivalent to find the intersection the line $\ell' =< u, v >$ and the Kestenband arc through u, and we know that this intersection consists in at most two points. □

In this way, we have proved the following result:

Proposition 2.12. *The incidence matrix $\mathbf{H}$ of the generalized quadrangle $\mathcal{H}(3,q^2)$ has a decomposition in circulant submatrices of order $N = q^2 - q + 1$, whose rows have weight at most 2.*

Proof. Label the points and the lines in the same way of Theorem (3.16) of the previous section. □

The point–starters for the group S can be chosen in the following way: let $\mathcal{B}$ be the Baer subplane $\{(\xi^{i(q^2-q+1)}, 0), i = 0, \ldots, q^2 + q\}$ of π_∞, then we may take as point–starters $\mathcal{B} \cap \mathcal{H}_0$ and the $q + 1$ absolute points of $\ell = \langle Q, P \rangle$, for $Q \in \mathcal{B}\backslash\mathcal{H}_0$. As line–starters we may take the $q + 1$ lines through every point of $\mathcal{B} \cap \mathcal{H}_0$.

Our goal is to find one or more cyclic subgroups of Γ that map a point–starters into another, in order to find a circulant display of the of the circulant submatrices and construct the incidence matrix $\mathbf{H}$ via the incidence of the smallest number possible of lines, as it is done in [44].

In the previous section we have described a cyclic group of collineations $PG(2,q^2)$, namely T, such that (for more details see the proof of Theorem (2.8) of the previous section):

- T has order $q + 1$;

- T fixes the point (1) and no other one;
- T fixes $\mathcal{B} \cap \mathcal{H}(2, q^2)$ set–wise and acts regularly on its points;
- T is a subgroup of $PGU(3, q)$.

Therefore, there exists a subgroup of Γ isomorphic to T and we keep on denoting this group by T by notation abuse. We have $q^2 + q + 1$ point–starters with different orbits under the action of T, namely Q, R_{0j}, $j = 1, \ldots, q+1$, and R_{ij}, $i = 1, \ldots, q-1, j = 1, \ldots, q+1$, where Q is a point of $\mathcal{B} \cap \mathcal{H}_0$, R_{0j} is the $j-th$ point of the line $< P, (1,0) >$, R_{ij} is the $j-th$ point of the line $< P, R_i >$ and $\{R_1, \ldots, R_{q-1}\}$ are $q-1$ points of $\mathcal{B} \backslash \{(1,0), \mathcal{H}_0\}$ with different orbits under the action of T. Label the other points as it is done in the proof of Theorem (3.16) of the previous section. The line–starters with different orbits under the action of T are $q+1$, namely $\ell_1, \ldots, \ell_{q+1}$, which are the $q+1$ lines through Q.

Finally, let U be the cyclic group of collineations spanned by:

$$(u, u_0) \mapsto (u, u_0 \eta^{q-1}).$$

The group $U =< v >$ has order $q+1$, it is a subgroup of Γ and it fixes point–wise π_∞ (hence it fixes P). Let Q be a point of $\pi_\infty \backslash \mathcal{H}_0$: the group U acts regularly on the absolute points of $\ell =< P, Q >$. Finally, let $Q = (u, 0)$ be a point of $\mathcal{H}_0$; the polar plane π of Q has equation $Tr(u^{q^2} x) = 0$, therefore the point P belongs to π; consider a $q+1$–secant through P in π, say ℓ: since the group U fixes Q and acts regularly on the absolute points of ℓ, the group U acts regularly on the lines through Q. Consider the before described point–starters $\{Q, R_{0j}, j = 1, \ldots, q+1, R_{ij}, i = 1, \ldots, q-1, j = 1, \ldots, q+1\}$: the point Q is fixed by U, while $R_{01}, R_{i1}, i = 1, \ldots, q-1$ have different orbits under the action of U, therefore label the other ones in the following way:

$$R_{0(j+1)} := R_{01}^{v^j}, j = 0, \ldots, q,$$

$$R_{i(j+1)} := R_{i1}^{v^j}, j = 0, \ldots, q, i = 1, \ldots, q-1.$$

The line-starters $\ell_1, \ldots, \ell_{q+1}$ form a unique orbit under the action of U, therefore label them in the usual way.

We are now able to state the following

Theorem 2.13. *If we order the points and the lines of $\mathcal{H}(3, q^2)$ in a suitable way, then the incidence matrix $\mathbf{H}$ presents a "double" circulant display of the circulant submatrices of order $N = q^2 - q + 1$ and $\mathbf{H}$ can be constructed by means of the incidence of just one line ℓ with respect to the points.*

Chapter 3

Small weight codewords of codes from linear representation

Let $\mathcal{K}$ be a set of points in the Desarguesian projective plane $PG(2,q)$ and embed $\pi_\infty = PG(2,q)$ as a hyperplane in $PG(3,q)$. The linear representation $T_2^*(\mathcal{K})$ has as point set the set of points of $AG(3,q) = PG(3,q) \setminus \pi_\infty$, as line set the set of lines of $AG(3,q)$ intersecting π_∞ in a point of $\mathcal{K}$, and it has the natural incidence relation. Each point of $T_2^*(\mathcal{K})$ is incident with $|\mathcal{K}|$ lines and each line is incident with q points. The type of incidence structure we obtain changes according to the choice of $\mathcal{K}$. Let α and β be two strictly positive integers. In the following three cases, $T_2^*(\mathcal{K})$ is a well-known combinatorial structure. References [43],[4] and [5] give more details about the following geometries.

1. If $|l \cap \mathcal{K}| \in \{0, \alpha+1\}$ for all lines l of Π_∞, then $T_2^*(\mathcal{K})$ is a partial geometry
   $pg(q-1, |\mathcal{K}|-1, \alpha)$. For example, if $\mathcal{K}$ is a hyperoval, then $T_2^*(\mathcal{K})$ is a generalized quadrangle.

2. If $|l \cap \mathcal{K}| \in \{1, \alpha+1\}$ for all lines l of Π_∞, then $T_2^*(\mathcal{K})$ is a semipartial geometry $spg(q-1, |\mathcal{K}|-1, \alpha)$. For example, let $\mathcal{K}$ be the Hermitian

curve in $PG(2,q^2)$, then we get a $spg(q^2-1,q^3,q)$.

3. If $|l \cap \mathcal{K}| \in \{\alpha+1, \beta+1\}$ for all lines l of Π_∞, then $T_2^*(\mathcal{K})$ is an (α,β)-geometry.

The incidence matrix of a linear representation $T_2^*(\mathcal{K})$ is a matrix $\mathbf{H}$ with rows labeled by the points of $T_2^*(\mathcal{K})$, columns labeled by the lines of $T_2^*(\mathcal{K})$. For $q=p^h$, p prime, $h \geq 1$, the p-ary linear code with $\mathbf{H}$ as parity check matrix is a LDPC code of length $q^2|\mathcal{K}|$. On the other hand, the code arising from the dual geometry $T_2^*(\mathcal{K})^D$ of $T_2^*(\mathcal{K})$ is a LDPC code of length q^3.

Our main goal is to find the minimum distance of these LDPC codes and to characterize their small weight codewords. The techniques used here are taken from [27] and they are valid for binary LDPC codes, LDPC codes over $GF(q)$, or over $GF(p)$.In Table 1, we have denoted by Θ, $\mathcal{B}$, $\mathcal{U}$ and $\mathcal{L}$ a hyperoval, a Baer subplane, a Hermitian curve, and two intersecting lines, respectively, and we have presented the lower bounds on the minimum distance d_{min} of the LDPC codes arising from their linear representations due to the *bit-oriented bound*, the *parity oriented bound* and *Massey's bound* [51].

LDPC code	Order (ρ,γ)	$d_{\min}$
$T_2^*(\Theta)$	$(q+2,q)$	$\geq 2q$
$T_2^*(\Theta)^D$	$(q,q+2)$	$\geq 4q$
$T_2^*(\mathcal{B})$	$(q+\sqrt{q}+1,q)$	$\geq q+1$
$T_2^*(\mathcal{U})$	$(q\sqrt{q}+1,q)$	$\geq q+1$
$T_2^*(\mathcal{L})$	$(2q+1,q)$	$\geq q+1$

Table 1: Known results

Using the following geometrical property, we can prove that some of the lower bounds of Table 1 are sharp, we can find larger lower bounds and in some cases prove their sharpness.

Let C be the LDPC code defined by $T_2^*(\mathcal{K})$. A codeword $c=(c_1,\ldots,c_n)$ of C is such that cH^T equals 0 and $supp(c)$, which is the set of all non-zero positions of c, defines

(∗) a set S of lines of $T_2^*(\mathcal{K})$ such that every point of $T_2^*(\mathcal{K})$ lies on zero or on at least two lines of S.

If we are considering the dual setting of $T_2^*(\mathcal{K})$, $supp(c)$ defines

(∗∗) a set S of points of $T_2^*(\mathcal{K})$ such that every line of $T_2^*(\mathcal{K})$ contains zero or at least two points of S.

The conditions (∗) and (∗∗) are necessary conditions, hence we look for codewords of C among the subsets of lines (or points) of $T_2^*(\mathcal{K})$ that satisfy these conditions.

Example 3.1. *Let π be an affine plane of $PG(3,q)$ that intersects $\mathcal{K}$ in at least two points P and Q, and let S be the set of all the affine lines of π through P and all the affine lines of π through Q. Clearly, every affine point lies in 0 or exactly two lines of S, hence S satisfies condition (∗). Take the vector c with 1 in the coordinate positions corresponding to the lines through P, with -1 in the coordinate positions corresponding to the lines through Q, and zero in the other positions. The vector c is orthogonal to every row of $\mathbf{H}$, and hence, it is a codeword of weight $2q$ of the code arising from $T_2^*(\mathcal{K})$, with $\mathcal{K}$ arbitrary. Therefore, if q is even and $\mathcal{K}$ is a hyperoval, the lower bound in the first row of Table 1 is sharp.*

Example 3.2. *Let q be even and let Θ be a regular hyperoval of $PG(2,q)$, $q = 2^h$, $h \geq 1$. We construct a set S of points that satisfies the condition (∗∗) using the construction introduced by Segre in [48]. Here we use the coordinate description by Pambianco and Storme [41] to construct complete caps. Suppose that the plane at infinity has equation $x_2 = x_3$ and let Θ be the set $\{(t^2,t,1,1)|t \in GF(q)\} \cup \{(1,0,0,0),(0,1,0,0)\}$. Let S be the set $C_1 \cup C_2 \cup C_1' \cup C_2'$, where $C_1 = \{(t^2,t,1,0)|t \in GF(q)\}$, $C_2 = \{(t^2,t,0,1)|t \in GF(q)\}$, $C_1' = \{(t^2+\mu,t+\mu,1,0)|t \in GF(q)\}$ and $C_2' = \{(t^2+\mu,t+\mu,0,1)|t \in GF(q)\}$, with $\mu \neq 0,1$. Then every affine line through Θ contains zero or two points of S. More precisely, there are four possibilities for a line that intersects S. A line can intersect*

C_1 and C_2, or C_1' and C_2', or C_1 and C_1', or C_2 and C_2'. The lines through a point of C_1 and a point of C_2', and the lines through a point of C_1' and a point of C_2 are not in the geometry $T_2^*(\Theta)$.

Let c be the vector with 1 in the coordinates corresponding to the points of $C_1 \cup C_2 \cup C_1' \cup C_2'$, and zero in the other positions. Clearly, the vector c is a vector of the code arising from $T_2^*(\Theta)^D$ of weight $4q$, hence the lower bound of the second row of Table 1 is sharp for Θ a regular hyperoval.

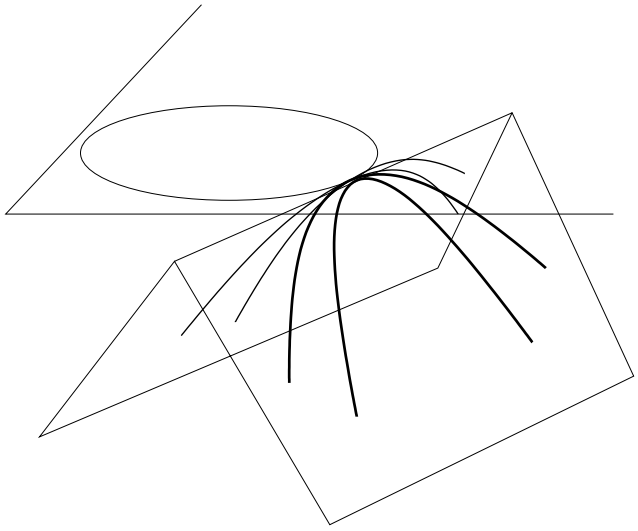

Figure 1: The configuration of Example 2.

If we replace t^2 in the descriptions of $\Theta, C_1, C_2, C_1', C_2'$ by t^{2^v}, $\gcd(v,h)=1$, we obtain similar codewords of weight $4q$ by using translation hyperovals instead of regular hyperovals. In Example 3.2, we use a set of three translation ovals C_1, C_2, C, through a same point at infinity, and having the same nucleus at infinity, with the property that any line that intersects two of them, intersects the third one. From now on, we denote $(q+1)$-arcs C_1, C_2 satisfying this condition with respect to $\Theta = C$, by *corresponding* $(q+1)$-arcs w.r.t. $T_2^*(\Theta)$.

Example 3.3. *Suppose that $\mathcal{K}$ contains a conic $\mathbf{C}$ and let S be the set of lines of a hyperbolic quadric $\mathcal{Q}^+(3,q)$ intersecting the plane at infinity in $\mathbf{C}$. Then the set S satisfies condition $(*)$. Let the quadric $\mathcal{Q}^+(3,q)$*

be the union of the two reguli $\mathcal{R}_1$ and $\mathcal{R}_2$, and take the vector c with 1 in the coordinate positions corresponding to the lines of $\mathcal{R}_1$, -1 in the coordinate positions corresponding to the lines of $\mathcal{R}_2$, and zero in the other positions. Then the vector c is a codeword of weight $2(q+1)$.

Remark 3.4. *In Example 3.1, we have shown that a codeword of weight $2q$ exists whatever be $\mathcal{K}$, hence the next step is to determine in which cases can exist codewords of weight lower than $2q$.*

Proposition 3.5. *Let $\mathcal{K}$ be an arbitrary set of points at infinity, let $\mathcal{C}$ be the LDPC code arising from $T_2^*(\mathcal{K})$, c a codeword of $\mathcal{C}$ and let S be the set of lines defined by $supp(c)$. If $wt(c) < 2q$, then S is contained in a plane. If $wt(c) = 2q$, then either:*

1. *S consists of $2q$ lines of a hyperbolic quadric having two lines at infinity contained in $\mathcal{K}$,*

or

2. *$S = S_1 \cup S_2$, where S_i, $i = 1, 2$, is a dual q-arc contained in the affine plane π_i, extended by the line at infinity to a dual $(q+1)$-arc. Let l be $\pi_1 \cap \pi_2$, then S_i, the line at infinity of π_i, and l form a dual hyperoval, $i = 1, 2$, and q is even. If l is not a line of $T_2^*(\mathcal{K})$, then S gives rise to a minimal codeword. On the other hand, if l is a line of $T_2^*(\mathcal{K})$, then $c = c' - c''$, where c' is the codeword derived from the dual $(q+1)$-arc $S_1 \cup \{l\}$, c'' is the codeword derived from the dual $(q+1)$-arc $S_2 \cup \{l\}$ and $wt(c') = wt(c'') = q+1$, where c' and c'' have the same symbol in their support.*

3. *S consists of $2q$ lines in a plane.*

Proof. Let S be the set of lines defined by $supp(c)$, with c a codeword of weight $\leq 2q$ in the LDPC code defined by $T_2^*(\mathcal{K})$. Let π be an affine plane and let $X = \{l_1, \ldots, l_i\}$ be the set of lines of S contained in π. In order to satisfy condition $(*)$, every line of X has at least $q - i + 1$ affine points that lie on a line of $S \setminus X$, hence

$$i(q - i + 1) \leq 2q - i$$

from which we get: $i \geq q$ or $i \leq 2$.

If $i = q$, then the line l_k of X has at least one affine point contained in a line of $S \setminus X$, $\forall k = 1, \ldots, q$, and $|S \setminus X| \leq q$, hence, l_k has exactly one affine point contained in a line of $S \setminus X$ and intersects the lines l_j, $j \neq k$, in different affine points, $\forall j, k = 1, \ldots, q$. If l_∞ is the line at infinity of π, then the set $\{l_\infty, l_1, \ldots, l_q\}$ is a dual $(q+1)$-arc of π. The lines of $S \setminus X$, say $m_1, \ldots, m_q$, must intersect each other in an affine point, hence, they all lie in the same plane π_1 and, using the same arguments, $m_1, \ldots, m_q$, and the line at infinity of π_1, say m_∞, form a dual $(q+1)$-arc. Let l be $\pi \cap \pi_1$; if l is a line of $T_2^*(\mathcal{K})$, then $\{l_\infty, l_1, \ldots, l_q, l\}$ and $\{m_\infty, m_1, \ldots, m_q, l\}$ are two dual hyperovals and they give rise to two codewords of weight $q+1$, say c' and c'', such that $c = c' - c''$. Hence, c' has a scalar α in the coordinate positions of the lines $l_1, \ldots, l_q, l$, and c'' has the same scalar α in the coordinate positions of the lines $m_1, \ldots, m_q, l$. If l is not a line of $T_2^*(\mathcal{K})$ (this happens when $\mathcal{K}$ contains at most q points of a line, for example when $\mathcal{K}$ is a maximal arc of degree q), then the set $\{l_1, \ldots, l_q, m_1, \ldots, m_q\}$ gives rise to a minimal codeword of weight $2q$. Therefore, if $i = q$, then we obtain case 2 of the proposition.

Now suppose that in the case $i \geq q+1$, there exists a line of S, say l, that is not contained in π. This line l has at least $q-1$ affine points that must lie on a second line of S not contained in π, hence $S \setminus X$ contains at least $q-1$ lines different from l. This yields

$$|S| = |X| + |S \setminus X| \geq q + 1 + 1 + q - 1 \geq 2q + 1,$$

a contradiction. We conclude that all lines of S are contained in π.

Suppose now that $i = 2$ and let $X = \{l_1, m_1\}$. We also assume that every plane contains at most two lines of S, since otherwise we are forced to the previous case. On the lines l_1 and m_1, there are $2(q-1)$ points that must lie on a line of $S \setminus X$; let $\{l_2, \ldots, l_q\}$ be the lines intersecting m_1 and $\{m_2, \ldots, m_q\}$ be the lines intersecting l_1. Until now, we have already counted $2q$ lines. If there are two lines of $\{l_2, \ldots, l_q\}$, say l_2 and l_3, intersecting in a point, then there exists a plane, say π', that contains $\{m_1, l_2, l_3\}$, but we excluded this possibility.

Now suppose that both in the set $\{l_1, \ldots, l_q\}$ and $\{m_1, \ldots, m_q\}$, the lines are pairwise skew, hence, l_i intersects m_j, $\forall i, j = 1, \ldots, q$; in other words, the lines of the set $\{l_1, \ldots, l_q, m_1, \ldots, m_q\}$ form a hyperbolic quadric intersecting $\mathcal{K}$ in two lines, so we get the case 1 of the proposition.

Finally, if $X = \{l\}$ and P is a point on l, then P is contained in at least a second line of S, say m. It follows that the plane $\pi' = \langle l, m \rangle$ contains at least two lines of S and so we get again one of the previous cases. □

Using Proposition 3.5, we can derive a new lower bound on the minimum distance of the LDPC code arising from $T_2^*(\mathcal{K})$, with $\mathcal{K}$ arbitrary.

Proposition 3.6. *Let c be a codeword of weight smaller than or equal to $2q$ in the LDPC code arising from $T_2^*(\mathcal{K})$ and let S be the set of lines defined by $supp(c)$. Suppose that the lines of S all lie in the same plane π and let x be the number of points of $\pi \cap \mathcal{K}$; then we have $wt(c) \geq q + q/(x-1)$.*

Proof. Let $wt(c) = q + k$, with $1 \leq k \leq q$, and let $\pi \cap \mathcal{K} = \{P_1, \ldots, P_x\}$; the average number of lines of S through a point of $\pi \cap \mathcal{K}$ is $(q+k)/x$, hence there exists a point of $\mathcal{K}$, say P_1, through which there pass at least $(q+k)/x$ lines of S. Let l be a line of S through P_1; every affine point of l is contained in at least another line of S, hence, there are at least q lines of S not through P_1. This implies that the following inequality must hold:

$$\frac{q+k}{x} + q \leq q + k,$$

from which we derive $k \geq q/(x-1)$. □

If x is the minimum integer, greater than one such that $x = |\pi \cap \mathcal{K}|$, then the lower bound of the previous proposition is sharp in a number of cases.

1. If $x = 2$, then $wt(c) \geq 2q$. The lower bound is sharp because of Example 3.1 which always occurs, whatever $\mathcal{K}$ is.

2. If $x = q + 1$, then $wt(c) \geq q + 1$. Let q be even and let S be a dual $(q+1)$-arc of the plane extended by the line at infinity to a

dual $(q+2)$-arc. The codeword c having a constant symbol α in the positions of S has weight $q+1$.

3. If $x = \sqrt{q}+1$, then $wt(c) \geq q+\sqrt{q}$. If q is even, then let S be a dual $(q+\sqrt{q})$-arc of type $(0,2,\sqrt{q})$ and let the line at infinity be the dual nucleus of the arc. The following example of a $(q+\sqrt{q})$-arc of type $(0,2,\sqrt{q})$ is based on a construction due to Korchmáros and Mazzocca (see [29]). Then

$$\{(z^2+z^{2\sqrt{q}},z,1)|z \in GF(q)\} \cup \{(1,z',0)|z' \in \mathbb{F}_{\sqrt{q}}\}$$

is a $(q+\sqrt{q})$-arc of type $(0,2,\sqrt{q})$ with $(0,1,0)$ as $\sqrt{q}$-nucleus. The points $(z^2+z^{2\sqrt{q}}=\rho,z,1)$, with $z \in GF(q)$, belong to $X=\rho Z$, for some $\rho \in \mathbb{F}_{\sqrt{q}}$. The points $(1,z',0)$, with $z' \in \mathbb{F}_{\sqrt{q}}$, are on $Z=0$. So the $\sqrt{q}$-secants through $(0,1,0)$ are $X=\rho Z$, with $\rho \in \mathbb{F}_{\sqrt{q}}$, and $Z=0$, and these $\sqrt{q}+1$ lines l_i form a dual Baer subline. When we dualize, this gives a line l_∞ with $P_1,\ldots,P_{\sqrt{q}+1}$ the $\sqrt{q}+1$ points of a Baer subline, where we denoted the dual of the line l_i by P_i.

There are $\sqrt{q}$ lines of S through every point P_i intersecting all the lines with a different direction in an affine point. Take a vector c with in all these $q+\sqrt{q}$ lines the same symbol, then c is a codeword of $T_2^*(\mathcal{K})$ with weight $q+\sqrt{q}$.

4. In general, the lower bound $q+q/(x-1)$ is sharp if we find a set of lines S that is a dual $(0,2,t)$-arc of size $q+t$ in $PG(2,q)$ such that the line at infinity is the dual t-nucleus and $t=q/(x-1)$. A result of Gács and Weiner [14] shows that a $(0,2,t)$-arc of size $q+t$ always has a t-nucleus. If such an arc exists, then q is even, unless $x=2$.

The following table summarizes the results obtained in the previous part.

LDPC code	Order (ρ,γ)	d_{min}
$T_2^*(\Theta)$	$(q+2,q)$	$2q$ (see Example 3.1)
$T_2^*(\Theta)^D$, Θ a translation hyperoval	$(q,q+2)$	$4q$ (see Example 3.2)
$T_2^*(\mathcal{B})$, q even	$(q+\sqrt{q}+1,q)$	$q+\sqrt{q}$ (see Proposition 3.6)
$T_2^*(\mathcal{U})$, q even	$(q\sqrt{q}+1,q)$	$q+\sqrt{q}$ (see Proposition 3.6)
$T_2^*(\mathcal{L})$, q even	$(2q+1,q)$	$q+1$ (see Proposition 3.6)

Table 2: New results

So far, we have results for even q. In the general case $T_2^*(\mathcal{K})$, with odd $q=p^h$, we have not been able to determine the minimum weight of the p-ary linear code of $T_2^*(\mathcal{K})$.

In the following proposition, we present a codeword of weight $2q-2$ in the LDPC code of $T_2^*(\mathcal{K})$, where $\mathcal{K}$ contains a Baer subline, which shows that most likely in general, the minimum weight of the p-ary linear LDPC code of $T_2^*(\mathcal{K})$ is smaller than $2q$. Note that the following construction is also valid for q an even square.

Proposition 3.7. *When $\mathcal{K}$ contains a Baer subline, there exists a codeword of weight $2q-2$ in the p-ary linear code of $T_2^*(\mathcal{K})$, with $q=p^h$ odd, q square.*

Proof. Let L be the Baer subline $PG(1,\sqrt{q})$ at infinity contained in $\mathcal{K}$. In $PG(2,q)$, there exist Baer subplanes B_1 and B_2 which share the Baer subline L and one extra point P_1 (not on the line $\bar{L}$ of $PG(2,q)$, extending L). Then B_1 and B_2 share $\sqrt{q}+2$ lines; namely the line L and the lines through a point of L and P_1. So B_1 has $q-1$ lines not lying in B_2, and B_2 has $q-1$ lines not lying in B_1.

Give all the lines of B_1, not in B_2, symbol 1, and all the lines of B_2, not in B_1, symbol -1. All other lines have symbol zero. We show that this vector gives a codeword of weight $2q-2$.

An affine point not lying in $B_1 \cup B_2$ lies on one line of B_1 and one line of B_2. If these lines are different, they have respectively symbols 1 and -1, so the sum is zero. If these lines coincide, they pass through P_1, so they have symbol zero.

The point P_1 only lies on lines with symbol zero.

A point R of $B_1 \setminus B_2$ lies on $\sqrt{q}$ lines of B_1 not in B_2. The only line through R lying in B_2 is the line RP_1 with symbol zero. So the sum of the symbols is $\sqrt{q} \equiv 0 \bmod p$. Similarly, a point R of $B_2 \setminus B_1$ lies on $\sqrt{q}$ lines of B_2 not in B_1. So the sum of the symbols is $-\sqrt{q} \equiv 0 \bmod p$.

This shows that for any point, the sum of the symbols of lines passing through it equals zero, hence, we have found a codeword. □

In general, if q is odd and if we are considering the binary code arising from $T_2^*(\mathcal{K})$, we know that every codeword has an even weight in virtue of the following result.

Proposition 3.8. *Let $\mathfrak{I} = (\mathcal{P}, \mathcal{B}, I)$ be a finite incidence structure such that every block contains $s+1$ points, let $\mathbf{H}$ be the incidence matrix of $\mathfrak{I}$ (labeling the columns by blocks) and let $\mathbf{C}$ be the binary linear code having $\mathbf{H}$ as parity check matrix. If $s+1$ is odd, then every codeword of $\mathbf{C}$ has an even weight.*

Proof. Let c be a codeword of $\mathbf{C}$ and let B be the set of blocks defined by $supp(c)$. Since the code is binary and regarding $(*)$, every point P_i of $\mathcal{P}$ is contained in zero or in an even number of elements of B, say x_i. A double counting argument yields that

$$|B|(s+1) = \sum_i x_i.$$

The right hand side is even, and $s+1$ is odd, so $|B| = wt(c)$ is even. □

3.1 Small weight codewords in $T_2^*(\Theta)$

In this section, we take a closer look at the case $T_2^*(\Theta)$, with Θ a hyperoval, hence q even. The linear representation $T_2^*(\Theta)$ is known to be (see [43]) a generalized quadrangle of order $(s,t) = (q-1, q+1)$. Let C be the q-ary LDPC code arising from $T_2^*(\Theta)$; in Example 3.1, we have already showed that the minimum distance of C is $2q$ and we have given the geometrical

description of a codeword of such a weight; if Θ contains a conic, we have a geometrical description of a codeword of weight $2(q+1)$ (see Example 3.3). Using the same arguments as in the proof of Proposition 3.5, yields that this is the only possibility for a codeword c with $2q < wt(c) \leq 2(q+1)$, hence we have the following Proposition.

Proposition 3.9. *Let* $\mathbf{C}$ *be the LDPC code defined by* $T_2^*(\Theta)$*, and let* S *be the set of lines defined by* $supp(c)$*. If* $wt(c) \leq 2(q+1)$*, then either:*

1. *S defines a set of $2q$ lines in a plane*

or

2. *S defines a set of $2(q+1)$ lines of a hyperbolic quadric $\mathcal{Q}$, intersecting Θ in a conic.*

For weights larger than $2(q+1)$, $q = 2^h$, $h \geq 7$, we will characterize the codewords of C, up to weight $2\sqrt[3]{q}(q+1)/3$, as linear combinations of codewords of weight $2q$ and $2(q+1)$ in a similar way as the authors do in [27], that is using geometrical arguments.

From now on, let c be a codeword of the code C arising from $T_2^*(\Theta)$, $q = 2^h$, $h \geq 7$, let $wt(c) = 2\delta(q+1) \leq 2\sqrt[3]{q}(q+1)/3$ and let S be the set of lines defined by $supp(c)$.

Proposition 3.10. *For every line l of S, there exists an affine plane π containing l such that π contains at least $2(q-2\delta+1)$ lines of S, or there exists a hyperbolic quadric $\mathcal{Q} \cong \mathcal{Q}^+(3,q)$ containing l and intersecting Θ in a conic, such that each regulus of $\mathcal{Q}$ contains at least $q-4\delta+2$ lines of S.*

Proof. Let l_1 be a line of S. In order to fulfill condition $(*)$, every affine point of l_1 needs to lie on a second line of S; let these lines be $m_1, \ldots, m_q$. The lines m_i do not intersect each other affinely (since $T_2^*(\Theta)$ is a GQ), hence there are $q(q-1)$ affine points on them that must lie on a second line of S. The average number of points of $m_1 \cup \cdots \cup m_q$ on one of the remaining lines is

$$y = \frac{q(q-1)}{(2\delta-1)(q+1)} > \frac{q-2}{2\delta}.$$

Hence, there exists a line l_2 in S that intersects at least y of the lines m_i, say $m_1, \ldots, m_k$, with $k \geq y > (q-2)/(2\delta)$. The lines l_1 and l_2 can be either skew or can intersect at infinity.

Case 1: Assume that the lines l_1 and l_2 intersect at infinity.

Then, the lines $l_1, l_2, m_1, \ldots, m_k$ all lie in the same plane π. Let x and t be the number of the lines of S through the two points of Θ in π, with $x \leq t$. Then there are $t(q-x) + x(q-t)$ affine points on these $x+t$ lines that still must lie on a second line of S. A line not in π can contain at most one affine point of π, so, in order to avoid a contradiction, we must have that

$$t(q-x) + x(q-t) \leq 2\delta(q+1) - x - t, \tag{3.1}$$

which implies that

$$x + t \leq 2\delta + \frac{2xt}{q+1} < 2\delta + 2x.$$

Let i be $t-x$, then $i < 2\delta$. Replacing t by $x+i$ in (3.1) yields:

$$2x^2 - 2x(q+1-i) + (2\delta - i)(q+1) \geq 0.$$

Recall that $\delta \leq \sqrt[3]{q}/3$ and that $i < 2\delta$. This implies that $x < \delta + 1/2$ or $x > q - 2\delta + 1$. Since t is at least $k \geq (q-2)/(2\delta)$, $x = t - i$ must be at least $q - 2\delta + 1$. So there exists a plane π through l_1 containing at least $2(q - 2\delta + 1)$ lines of S.

Case 2: Assume that the lines l_1 and l_2 are skew.

Hence, there are $k(q-2)$ affine points on the lines $m_1, \ldots, m_k$ that must lie on a second line of S, and the average number of these points on the remaining lines of S is

$$z = \frac{k(q-2)}{(2\delta-1)(q+1) - 1} > \frac{(q-2)^2}{4\delta^2(q+1)},$$

hence, there exists a line l_3 of S that intersects $h \geq z > (q-2)^2/(4\delta^2(q+1))$ lines m_i, say $m_1, \ldots, m_h$. The lines l_1, l_2 and l_3 are pairwise skew and they intersect $m_1, \ldots, m_h$ in different points, hence they define a hyperbolic quadric $\mathcal{Q} \cong \mathcal{Q}^+(3,q)$. Suppose that there are x lines of S in the first

regulus of $\mathcal{Q}$ and t lines of S in the opposite regulus, with $x \leq t$. On these lines, there are $t(q-x)+x(q-t)$ affine points that must lie on a second line of S. A line not contained in $\mathcal{Q}$ can meet the quadric $\mathcal{Q}$ in at most two points, hence

$$t(q-x)+x(q-t) \leq 4\delta(q+1)-2(x+t) \tag{3.2}$$

which yields that

$$x+t < 4\delta+2x. \tag{3.3}$$

Replacing t by $x+i$ in (3.2) gives the following inequality

$$2x^2-2x(q+2-i)+4\delta(q+1)-i(q+2) \geq 0.$$

Recall that $i < 4\delta$ from (3.3), $\delta \leq \sqrt[3]{q}/3$ and t must be at least $h > (q-2)^2/(4\delta^2(q+1))$, so the inequality (3.2) is only satisfied if $x > q-4\delta+2$. This implies that there exists a hyperbolic quadric $\mathcal{Q}^+(3,q)$ that contains at least $q-4\delta+2$ lines of S in each of its reguli. □

Proposition 3.10 implies that the lines of S are contained in planes and hyperbolic quadrics with "many" lines of S in it. Let S be contained in h of those planes and k of those hyperbolic quadrics. Two planes have at most one line in common, and a plane and a hyperbolic quadric have at most two lines in common. Two such hyperbolic quadrics containing at least $2(q-4\delta+2)$ lines of S share the same conic contained in Θ; and therefore share at most two lines. So we obtain the following inequality:

$$2h(q-2\delta+1)-\frac{h(h-1)}{2}+2k(q-4\delta+2)-(h+k-1)(h+k) \leq 2\delta(q+1),$$

which implies that

$$-2(h+k)^2+(h+k)(4q-8\delta+7)-h^2+3k-8\delta k \leq 4\delta(q+1). \tag{3.4}$$

Substituting λ for $h+k$, the inequality (3.4) becomes

$$-2\lambda^2+2\lambda(2q-8\delta+5)-h^2+h(8\delta-3) \leq 4\delta(q+1). \tag{3.5}$$

The inequality (3.5) is satisfied when λ is at most $\lceil \delta \rceil$, where $\lceil x \rceil$ denotes the smallest integer larger than or equal to x. Hence, we have proven the following proposition.

Proposition 3.11. *The set S is contained in at most $\lceil \delta \rceil$ planes or hyperbolic quadrics sharing at least $2(q-2\delta+1)$ or $2(q-4\delta+2)$ lines with S, respectively.*

We use Proposition 3.11 to prove the following result.

Proposition 3.12. *If $X = \{X_1, \ldots, X_k\}$, $k \leq \lceil \delta \rceil$, is the set of planes and hyperbolic quadrics containing S, then each X_i contains at least $2(q-2k)$ lines of S which are not contained in any X_j, $j \neq i$.*

Proof. Let X_1 be a plane and let l be a line of S in X_1 not contained in $X_2 \cup \cdots \cup X_k$. Then any of the planes or hyperbolic quadrics $X_2, \ldots, X_k$ intersects l in at most two points, hence there are at least $q-2k$ affine points of l that must be contained in a line of X_1 not contained in X_i, $i \neq 1$. If the line l goes through the point at infinity P of X_1, then these $q-2k$ lines of X_1 intersecting l go through the other point at infinity of $\Theta \cap X_1$. So in X_1, there are at least $q-2k$ lines of S not contained in $X_i, i \neq 1$, for any one of the two points at infinity of X_1 in Θ.

If X_1 is a hyperbolic quadric and l a line of S in X_1 not contained in X_i, $i \neq 1$, then, since any plane or hyperbolic quadric intersects l in at most two points, the same arguments show that for every regulus in X_1, there are at least $q-2k$ lines of S not contained in $X_i, i \neq 1$. □

Remark 3.13. *It follows from the proof of Proposition 3.12 that a line l_1 in S, contained in X_i and not contained in X_j, $j \neq i$, contains at least $q - 2\lceil \delta \rceil$ points that lie on exactly one other line l_2 of S, and that this line l_2 is contained in X_i, but not in X_j, $j \neq i$.*

Using the same techniques as in [27], we will characterize the codewords of small weight as being linear combinations of codewords of weight $2q$ and $2(q+1)$.

Proposition 3.14. *In the LDPC code defined by $T_2^*(\Theta)$, $q = 2^h$, $h \geq 7$, every codeword of weight at most $2\sqrt[3]{q}(q+1)/3$ is a linear combination of codewords of weight $2q$ or $2(q+1)$.*

Proof. We will prove this by induction on the weight of the codewords. Let c be a codeword of C of weight $2\delta(q+1)$, $\delta \leq \sqrt[3]{q}/3$, and assume that all the codewords of C of weight smaller than $wt(c)$ have already been characterized as being linear combinations of codewords of weight $2q$ and $2(q+1)$.

Let S be the set of lines defined by $supp(c)$ and let l_1 be a line of S contained in X_1 and not contained in X_j, $j \neq 1$. According to Remark 3.13, there exist h points $R_1, \ldots, R_h$, with $h = q - 2\lceil\delta\rceil$, on l_1 lying on exactly two lines of S, the line l_1 and another line of X_1. Denote the second line of S through R_i by l_{i+1}.

Every point R_i, $i \leq h$, defines a row of the parity check matrix $\mathbf{H}$ and since a codeword has to be orthogonal to every row of $\mathbf{H}$, the codeword c has (up to a scalar multiple) 1 in the position corresponding to l_1 and 1 in the positions corresponding to the lines $l_{i+1}, i = 1, \ldots, h$. The lines l_{i+1} intersect l_1, hence, if X_1 is a plane, then they are lines through an other point at infinity with respect to l_1. If X_1 is a hyperbolic quadric, then the lines l_{i+1} belong to the opposite regulus of the one containing l_1. Therefore, there are m lines, $l_{q-2\lceil\delta\rceil+k}, k = 2, \ldots, m+1$, with $m = q-2\lceil\delta\rceil-1$, through the same point at infinity as l_1 or in the same regulus of l_1 that belong only to X_1. A line $l_{q-2\lceil\delta\rceil+k}$ can intersect the $X_j, j > 1$, in at most $2\lceil\delta\rceil$ points (see Remark 3.13), hence, there exists a line among $l_2, \ldots, l_{q-2\lceil\delta\rceil+1}$ that intersects $l_{q-2\lceil\delta\rceil+k}$ in a point not belonging to $X_i, i > 1$. Repeating the same arguments yields that the codeword c has 1 in the positions corresponding to $l_1, l_{q-2\lceil\delta\rceil+k}$, with $k = 2, \ldots, m+1$, and 1 in the positions corresponding to $l_{i+1}, i = 1, \ldots, q-2\lceil\delta\rceil$. If X_1 is a hyperbolic quadric, then Θ is a regular hyperoval since it contains already at least $q - 4\delta + 2$ points of a conic [18, Lemma 8.9]. Let now c' be the codeword defined by taking all symbols in the positions corresponding to lines of X_1 equal to 1, then c and c' share at least $2q - 2\lceil\delta\rceil$ non-zero positions and symbols, so $wt(c-c') < wt(c)$. The induction hypothesis states that $c-c'$ is a linear combination of codewords of weight $2q$ and $2(q+1)$. Hence, $c = (c-c') + c'$ is a linear combination of such codewords too. □

3.2 $T_2^*(\Theta)^D$, with Θ a non-regular translation hyperoval

In this section, we characterize codewords of small weight of the LDPC code of $T_2^*(\Theta)^D$ with Θ a translation hyperoval. Therefore, we first give a detailed description of this dual generalized quadrangle and we distinguish between the cases Θ a non-regular translation hyperoval (Section 3.2) and Θ a regular hyperoval, i.e. a conic and its nucleus (Section 3.3). This description relies on the results of Payne and Thas [43].

Let Θ be the translation hyperoval

$$\left\{(1, x, x^\beta) | x \in GF(q)\right\} \cup \{(0,0,1),(0,1,0)\},$$

with β a generator of $\mathrm{Aut}(GF(q))$, embedded in the plane $X_0 = 0$ of $PG(3,q)$.

Proposition 3.15. *$T_2^*(\Theta)^D$ can be described as an incidence structure $(\mathcal{P}, \mathcal{L}, \mathcal{I})$ with*

$$\mathcal{P} = \begin{cases} \text{Affine points of } T_2^*(\Theta). \\ \text{Affine planes through } (0,0,1,0) \text{ and } (0,1,a,a^\beta), a \in GF(q). \\ \text{Affine planes through } (0,0,0,1) \text{ and } (0,1,a,a^\beta), a \in GF(q). \end{cases}$$

$$\mathcal{L} = \quad \text{Affine lines through the points } (0,1,a,a^\beta) \text{ of } \Theta.$$

$$\mathcal{I} = \begin{cases} \text{An affine point lies on an affine line if the point lies on that line.} \\ \text{An affine plane } \pi \text{ through } (0,1,a,a^\beta), \text{ and } (0,0,0,1) \text{ or } (0,0,1,0), \\ \text{is incident with the affine lines of } \pi \text{ through } (0,1,a,a^\beta). \end{cases}$$

Proof. Consider the mapping ϕ with $\phi(1,a,b,c) = \langle (1,0,c,b^\beta), (0,1,a,a^\beta)\rangle$. Then ϕ is obviously a bijection that maps points onto objects that will be the lines of the geometry $T_2^*(\Theta)^D$. From this definition, we get that $\mathcal{L}$ consists of all affine lines through the points $(0,1,u,u^\beta), u \in GF(q)$.

We determine the image of the lines of $T_2^*(\Theta)$ under ϕ, since this will be the points of $T_2^*(\Theta)^D$.

A line $\langle (0,0,0,1),(1,a,b,c)\rangle$ through $R = (0,0,0,1)$ corresponds to the set

$$\left\{\langle (1,0,c+\lambda,b^\beta),(0,1,a,a^\beta)\rangle | \lambda \in GF(q)\right\}.$$

All lines of this set are contained in a plane π_1 through $(0,0,1,0)$ and $(0,1,a,a^\beta)$, so we can identify this set of lines with π_1.

A line $\langle(0,0,1,0),(1,a,b,c)\rangle$ through $N=(0,0,1,0)$ corresponds to the set

$$\{\langle(1,0,c,b^\beta+\lambda^\beta),(0,1,a,a^\beta)\rangle|\lambda\in GF(q)\}.$$

All lines of this set are contained in a plane π_2 through $(0,0,0,1)$ and $(0,1,a,a^\beta)$, so we can identify this set of lines with π_2.

A line through $(1,a,b,c)$ and $(0,1,u,u^\beta)$ corresponds to the set

$$\left\{\langle(1,0,c+\lambda u^\beta,b^\beta+\lambda^\beta u^\beta),(0,1,a+\lambda,a^\beta+\lambda^\beta)\rangle|\lambda\in GF(q)\right\}.$$

Note that the lines of this set all pass through the point P with coordinates $(1,u^\beta,c+au^\beta,b^\beta+u^\beta a^\beta)$. So we can identify this set of lines with the point P.

Using these relations, it is clear that ϕ maps collinear points to intersecting lines, and intersecting lines to collinear points. □

We investigate the small weight codewords of the dual generalized quadrangle $T_2^*(\Theta)^D$ using property $(*)$, Hence, we are able to use the methods developed in Section3.1.

Theorem 3.16. *The minimum weight of the LDPC code of $T_2^*(\Theta)^D$, with Θ a translation hyperoval, is equal to $4q$. The minimum weight vectors correspond to the scalar multiples of incidence vectors of a set of all lines of $T_2^*(\Theta)^D$ in two planes, where these two planes pass through the same line at infinity.*

We immediately present the proof for codewords of weight $\leq 2\delta q$, with $\delta\leq\sqrt[3]{q}/3$, to avoid a too detailed repetition of the techniques of Section 3, and to build up already to Theorem 3.17.

Throughout this proof, we use $R=(0,0,0,1)$ and $N=(0,0,1,0)$. Let S be the set of lines defined by $supp(c)$, with c a codeword of the LDPC code of $T_2^*(\Theta)^D$.

Proof. Codewords of the LDPC code of $T_2^*(\Theta)^D$ satisfy Property $(*)$, hence the codeword corresponds to a set S of lines such that every point lies on zero or on at least two of them. There is only one kind of lines in $T_2^*(\Theta)^D$, the affine lines through the points with coordinates $(0,1,u,u^\beta)$, and there are three kinds of points of $T_2^*(\Theta)^D$ that have to lie on zero or on at least two lines of S.

A: The affine points.

When we only use the condition that every affine point has to lie on zero or on at least two lines, we can copy the proof for the LDPC code of $T_2^*(\Theta)$. In that case, the minimum weight of the code equals $2q$ and this weight occurs when taking all lines of $T_2^*(\Theta)$ in a fixed plane.

For every line l of S, there are two possibilities: either there exists a plane through l with at least $2(q-2\delta+1)$ lines of S, or there is a hyperbolic quadric through l with at least $2(q-4\delta+1)$ lines of S. But in this case, there are no codewords consisting of hyperbolic quadrics, since there is no conic lying at infinity in Θ. So the initial description of the codewords becomes: *Every possible codeword of weight* $\leq 2\delta q$, *with* $\delta \leq \sqrt[3]{q}/3$, *in* $T_2^*(\Theta)^D$ *is a linear combination of codewords of* $T_2^*(\Theta)$ *of weight* $2q$, *consisting of the* $2q$ *lines of* $T_2^*(\Theta)$ *in a plane containing two points* $(0,1,u,u^\beta)$ *and* $(0,1,v,v^\beta)$. All lines in such a plane have a fixed symbol in the corresponding codeword.

We still need to investigate which extra conditions the other two kinds of points of $T_2^*(\Theta)^D$ impose.

B: The points coming from tangent planes to Θ (planes through $(0,0,1,0)$).

Each tangent plane through a point $(0,1,u,u^\beta)$ has to contain zero or at least two lines. Case A implies that the possible codewords of $T_2^*(\Theta)^D$ of weight $\leq 2\delta q$ are linear combinations of codewords of weight $2q$ of $T_2^*(\Theta)$ in planes through two points $(0,1,u,u^\beta)$ and $(0,1,v,v^\beta)$. Take a codeword of weight $2q$, lying in the plane π, then the tangent planes at $\pi \cap \Theta$ contain only one line of S. So at least two codewords of $T_2^*(\Theta)$ (in planes π_1 and

π_2) are needed to construct a codeword. Now there are three possibilities.

- The intersection of $\pi_1 \cap \Theta$ and $\pi_2 \cap \Theta$ is empty. In this case, in each of the points of $\pi_1 \cap \Theta$ and $\pi_2 \cap \Theta$, their tangent planes through N contain only one line, a contradiction.
- There is exactly one intersection point in common in $\pi_1 \cap \Theta$ and $\pi_2 \cap \Theta$. In this case, for the two non-common intersection points, a tangent plane through them contains only one line, a contradiction.
- The two intersection points of $\pi_1 \cap \Theta$ and $\pi_2 \cap \Theta$ coincide.

The only possibility for a codeword consisting of two codewords of $T_2(\Theta)$, hence a codeword of weight $4q$, is a codeword arising from two planes π_1 and π_2 through the same points at infinity of $\Theta \backslash \{R, N\}$.

C: The points coming from planes through $(0,0,0,1) = R$ and a point of $\Theta \backslash \{R, N\}$.

Take a possible codeword found in Case B. Then S has two lines in common with the planes through R and the intersection points of the planes π_1 and π_2 with Θ. Furthermore, S has zero lines in common with planes through R and a different point of Θ. So the possible codeword of weight $4q$ does occur if we take the same symbol for the lines in the two planes π_1 and π_2. □

Theorem 3.17. *The codewords of the LDPC code of $T_2^*(\Theta)^D$, $q = 2^h$, $h \geq 7$, of weight $2\delta q$, with $\delta \leq \sqrt[3]{q}/3$, are linear combinations of codewords of $T_2^*(\Theta)$, with weight $2q$, which are coming from $2q$ lines in planes through two points of $\Theta \backslash \{R, N\}$, where the sum of the symbols of the lines through a point of $\Theta \backslash \{R, N\}$ has to be zero.*

Proof. From Case A, we derive that every codeword of weight at most $2\delta q$, with $\delta \leq \sqrt[3]{q}/3$, is a linear combination of codewords of $T_2^*(\Theta)$ with weight $2q$. Cases B and C yield the second condition, so that the sum of the symbols in the coordinate positions corresponding to the lines in each tangent plane to the q-arc $\Theta \backslash \{R, N\}$ equals zero. □

Remark 3.18. *Even though we use linear combinations of codewords of weight $2q$ of $T_2^*(\Theta)$, there are no codewords of weight $2q$ in $T_2^*(\Theta)^D$ (see Theorem 3.16).*

We will observe that it is sufficient to make the assumption that the sum of the symbols in the coordinate positions corresponding to the lines in one kind of tangent planes, i.e. either through R or through N, equals zero. In the proof, we use the planes through N.

By Example 3.2, we have already showed that the minimum weight of the code of $T_2^*(\Theta)^D$ is $4q$, giving the description of the geometrical configuration in the original setting of $T_2^*(\Theta)$that gives arise to a codeword of such a weight. Therefore, we want to describe also the codewords of small weight of Theorem3.17, in terms of points and lines of $T_2^*(\Theta)$. Let ϕ be the bijection between $T_2^*(\Theta)^D$ and $T_2^*(\Theta)$ defined in the proof of Proposition 3.15. When $\phi(x) = y$, or $\phi(x) = y$, then x and y are called *corresponding*.

Proposition 3.19. *The duality ϕ^{-1} maps lines of $T_2^*(\Theta)^D$ through the same point at infinity to points in the same plane in $T_2^*(\Theta)$.*

Proof. The line passing through $(1, 0, x, y^\beta)$ and $(0, 1, a, a^\beta)$ is mapped by ϕ^{-1} to the point $(1, a, y, x)$. So all lines of $T_2^*(\Theta)^D$ through $(0, 1, a, a^\beta)$ are mapped to points lying in the plane $aX_0 + X_1 = 0$. □

Proposition 3.20. *All planes in $T_2^*(\Theta)$ with points corresponding to lines in $T_2^*(\Theta)^D$ contain the points $R = (0, 0, 0, 1)$ and $N = (0, 0, 1, 0)$.*

Proof. As seen in Proposition 3.19, all these planes have equation $\alpha X_0 + X_1 = 0$, hence contain the points R and N. □

Proposition 3.21. *The duality ϕ^{-1} maps q coplanar lines of $T_2^*(\Theta)^D$ through a point $(0, 1, u, u^\beta), u \in GF(q)$, to a q-arc in a plane through R and N.*

Proof. All points of the plane Π through $(0, 1, u, u^\beta)$, $(0, 1, v, v^\beta)$ and $(1, 0, a, b^\beta)$ have coordinates $(1, \lambda + \mu, a + \lambda u + \mu v, b^\beta + \lambda u^\beta + \mu v^\beta)$.

It follows that the affine lines through $(0, 1, u, u^\beta)$ and the q points $(1, 0, a + \lambda(u + v), b^\beta + \lambda(u^\beta + v^\beta))$, $\lambda \in \mathbb{F}_q$, in Π are mapped to the q points

$(1, u, b + \lambda^{\beta^{-1}}(u+v), a + \lambda(u+v))$, with $\lambda \in \mathbb{F}_q$. It is easy to see that this set forms a q-arc. From Proposition 3.20 and 3.21, we get that this q-arc lies in the plane $uX_0 + X_1 = 0$ through R and N. □

Proposition 3.22. *Under the duality ϕ^{-1}, $2q$ coplanar lines in $T_2^*(\Theta)^D$ correspond to two corresponding q-arcs in two planes through RN.*

Proof. Consider $2q$ lines of $T_2^*(\Theta)^D$ lying in the same plane, say π. The preceding propositions tell us that these $2q$ lines correspond to a set B which is the union of two sets of q points, each set lying in a plane through R and N. Proposition 3.21 states that these sets form q-arcs. The duality ϕ^{-1} gives us the following correspondences.

By Proposition 3.15, a line through R in $T_2^*(\Theta)$ corresponds to a tangent plane in $T_2^*(\Theta)^D$ (which is a point of $T_2^*(\Theta)^D$); a line through N in $T_2^*(\Theta)$ corresponds to a plane through R in $T_2^*(\Theta)^D$ (which is a point of $T_2^*(\Theta)^D$).

A tangent plane in $T_2^*(\Theta)^D$ through one of the intersection points of π with Θ, say P, contains only one line. So a line through R in the plane defined by P in $T_2^*(\Theta)$ contains only one point of B. The affine planes through R and P contain only one line of π of $T_2^*(\Theta)^D$, so, applying ϕ^{-1}, every line through N in $T_2^*(\Theta)$ in the plane defined by P contains only one point of B. The same holds for the plane defined by the other intersection point of π with Θ.

We are taking q points in the two planes through RN containing the set B that form q-arcs. The points R and N only lie on tangents to these q-arcs, so R and N extend these q-arcs to $(q+2)$-arcs.

The two sets of coplanar concurrent lines in $T_2^*(\Theta)^D$ are such that a point lies on zero or exactly two lines of this set. So a line of $T_2^*(\Theta)$ not through R or N contains zero or exactly two points of B. Connecting a point of $\Theta \backslash \{R, N\}$ with the q points of one q-arc gives rise to the q-arc in the second plane through RN and vice versa. So the two q-arcs are corresponding ones (see Remark 3). □

Notation: If the points of a set X (e.g. a q-arc) all have the same symbol α in the corresponding codeword of the LDPC code of $T_2^*(\Theta)^D$, then we say briefly that this set X has symbol α.

Theorem 3.23. *The codewords of the LDPC code of $T_2^*(\Theta)^D$, $q = 2^h$, $h \geq 7$, described in terms of points and lines of $T_2^*(\Theta)$, with weight $\leq 2\sqrt[3]{q}q/3$, are linear combinations of incidence vectors of 2 corresponding q-arcs with the same symbol, each in a plane through RN, such that the sum of the symbols on a line of $T_2^*(\Theta)$ is zero. In particular, the sum of symbols of q-arcs in a fixed plane through RN, is zero. The minimum weight is equal to $4q$, corresponding to 2 sets of corresponding q-arcs, lying in 2 planes through RN.*

Proof. This is the dual of Theorems 3.16 and 3.17, using Propositions 3.19, 3.20, 3.22 to dualize. □

3.3 $T_2^*(\Theta)^D$, with Θ a regular hyperoval

In this section, we use the same strategy as in Section 3.2 to characterize codewords of small weight in the LDPC code of $T_2^*(\Theta)^D$, with Θ a regular hyperoval, i.e. the union of a conic and its nucleus. We deal with this case separately since in this case we are able to characterize codewords up to a larger upper bound, i.e. up to weight $4q^{3/2}/5$. Nevertheless, the arguments of the proofs are more complicated than in the case of the non-regular translation hyperoval and this seems to come from the fact that for the non–regular translation hyperoval $\left\{(1, t, t^{2^v}) | t \in GF(q)\right\} \cup \{(0,0,1),(0,1,0)\}$, with $q = 2^h$, $\gcd(v,h) = 1$, there is a point, namely $R = (0,0,1)$ or $N = (0,1,0)$, playing a special role in Proposition 3.20 (see also comments subsequent that Proposition), while we have no such point in the case of the regular hyperoval. Again, this may come from the fact that a non-regular translation hyperoval is stabilized by a group of order $2q(q-1)$ fixing $\{R, N\}$ while the regular hyperoval is stabilized by a group of order $q^3 - q$ only fixing $N = (0,1,0)$.

We first describe the structure of $T_2^*(\Theta)^D$ by using the following construction by Payne and Thas [42],[43].

Let $S = GQ(s) = (\mathcal{P}, \mathcal{L}, \mathcal{I})$ and let x be a regular point, i.e. a point

for which $|\{x,y\}^{\perp\perp}| = s+1$, for all points $y \neq x$. Then the following incidence structure $(\mathcal{P}', \mathcal{B}', \mathcal{I}')$ is a $GQ(s-1, s+1)$.

$\mathcal{P}' = \mathcal{P} \backslash x^{\perp}$

$\mathcal{L}' = \begin{cases} \text{The lines of } \mathcal{L} \text{ not through } x. \\ \text{The hyperbolic lines } \{x,y\}^{\perp\perp}, x \nsim y. \end{cases}$

$\mathcal{I}' =$ Natural incidence.

Applying the preceding construction on $T_2(\Theta')$, Θ' a conic, gives $T_2^*(\Theta)$ for Θ the regular hyperoval containing Θ'. Note that $T_2(\Theta')$ is isomorphic to $Q(4,q)$, so we can describe $T_2^*(\Theta)$ on $Q(4,q)$. Then $T_2^*(\Theta)$ is the following incidence structure $(\mathcal{P}, \mathcal{L}, \mathcal{I})$. Let P be a fixed point of $Q(4,q)$.

$\mathcal{P} =$ The points of $Q(4,q)$ not on $P^{\perp}$.

$\mathcal{L} = \begin{cases} \text{The lines of } Q(4,q) \text{ not through } P. \\ \text{The conics } C = \pi \cap Q(4,q) \text{ where } \pi \text{ is a plane through } \langle N, P \rangle, \\ \text{ with } N \text{ the nucleus of } Q(4,q). \end{cases}$

$\mathcal{I} =$ Natural incidence.

We want to characterize small weight codewords of the LDPC code of $T_2^*(\Theta)^D$, so the problem is again to find sets S of lines such that every point of $T_2^*(\Theta)^D$ lies on zero or on at least two lines of S in $T_2^*(\Theta)^D$.

We dualize the incidence structure of $T_2^*(\Theta)$ described on $Q(4,q)$. Since $Q(4,q)$, with q even, is self-dual (see e.g. [43]), the point P becomes a line L, and conics become reguli. So $T_2^*(\Theta)^D$ described on $Q(4,q)$ is an incidence structure $(\bar{\mathcal{P}}, \bar{\mathcal{L}}, \bar{\mathcal{I}})$ with

$\bar{\mathcal{P}} = \begin{cases} \text{The points of } Q(4,q) \text{ not on } L. \\ \text{The reguli through } L. \end{cases}$

$\bar{\mathcal{L}} =$ The lines of $Q(4,q)$ not in $L^{\perp}$.

$\bar{\mathcal{I}} =$ Natural incidence.

From now on, let c be a codeword of the LDPC code $\mathbf{C}$ arising from $T_2^*(\Theta)^D$, let $wt(c) \leq 4\delta q$, with $\delta \leq \sqrt{q}/5$, and with $q = 2^h$, $h \geq 5$, and let S be the set of lines defined by $supp(c)$.

Proposition 3.24. *For every line l of S, there exists a hyperbolic quadric $\mathcal{Q} \cong \mathcal{Q}^+(3,q)$ of $Q(4,q)$, containing l and such that each regulus*

of $\mathcal{Q}$ contains at least $q - 4\delta + 5/2$ lines of S.

Proof. The proof is similar to that of Proposition 3.10. □

Proposition 3.25. *The set S is contained in at most $2\delta + 1$ hyperbolic quadrics of $Q(4,q)$, each one containing at least $2(q-4\delta+5/2)$ lines of S.*

Proof. Two hyperbolic quadrics of $Q(4,q)$ share at most two lines and a hyperbolic quadric contains at most q lines of S in each regulus because the lines of S do not intersect L. In case there are J distinct hyperbolic quadrics, each one containing at least $2(q-4\delta+5/2)$ lines of S, we see that

$$|S| \geq \sum_{i=0}^{J-1}(2q - 8\delta + 5 - 2i).$$

Filling in $J = 2\delta + 2$ and using $\delta \leq \sqrt{q}/5$ yields a contradiction. So it follows that $J \leq 2\delta + 1$. □

Theorem 3.26. *The minimum weight of the LDPC code of $T_2^*(\Theta)^D$, Θ a regular hyperoval, is $4q$ and the codewords of weight $4q$ correspond to two hyperbolic quadrics of $Q(4,q)$, intersecting in two lines m_1, m_2, such that m_1, m_2 intersect L in the same point, where the hyperbolic quadrics have the same symbol in the corresponding codeword.*

Proof. We immediately present the proof for codewords of weight $\leq 4q^{3/2}/5$, to avoid a too detailed repetition of the techniques of Section 3, and to build up already to Theorem 3.27.

The lines of S lie in at most $2\delta + 1$ hyperbolic quadrics of $Q(4,q)$, with in each regulus at least $q - 4\delta + 5/2$ lines of S. If $X = \{\mathcal{Q}_1, \ldots, \mathcal{Q}_k\}$, $k \leq 2\delta + 1$, is the set of hyperbolic quadrics of $Q(4,q)$ containing S, then each $\mathcal{Q}_i$ contains at least $2q - 12\delta + 4$ lines of S which are not contained in any $\mathcal{Q}_j$, $j \neq i$. In particular, $\mathcal{Q}_1$ contains at least $2q - 8\delta + 4 - 4\delta$ lines of S not contained in $\mathcal{Q}_j$, $j \neq 1$. Each of these lines contains at least $q - 4\delta$ points of $Q(4,q) \cap \bar{\mathcal{P}}$ lying only in $\mathcal{Q}_1$. Take such a line l_1 and suppose that it has symbol 1 in the codeword c. Then there are at least $q - 4\delta$ lines of

the opposite regulus of $\mathcal{Q}_1$ having symbol 1. Note that we are not using the secant to L of this opposite regulus.

The other $q - 6\delta + 1$ lines of S only lying in the regulus of $\mathcal{Q}_1$ through l_1 can intersect already chosen lines of S in the opposite regulus of $\mathcal{Q}_1$ with symbol 1 in a point not only lying on $\mathcal{Q}_1$, but this can happen at most 4δ times for a line. Since $q - 4\delta > 4\delta$, there is for each of these $q - 6\delta + 1$ lines a point only lying on this line, and on an already chosen line of S in the opposite regulus of $\mathcal{Q}_1$ with symbol 1. We can conclude that all these $q - 6\delta + 1$ lines of S only lying in the regulus of $\mathcal{Q}_1$ through l must have symbol 1 in the codeword. So we have in total already $2q - 10\delta + 2$ lines in $\mathcal{Q}_1$ with symbol 1.

The line L intersects $\mathcal{Q}_1$ in a point P, and the sum of the symbols of the lines of $T_2^*(\Theta)^D$ through the points on the two lines l_1 and l_2 of $\mathcal{Q}_1$ through P has to be zero. There are points on l_1 and l_2 that lie only on a line with symbol 1 of $\mathcal{Q}_1$, so these points lie on at least one other quadric $\mathcal{Q}_i$, $i > 1$.

This shows that to obtain a codeword of minimal weight, we have to take at least two hyperbolic quadrics in $Q(4, q)$. Then the second hyperbolic quadric has also symbol 1 in most of its lines and passes through l_1 and l_2, and since we are only using two quadrics, they both have symbol 1 in all of their lines not lying in $L^\perp$. Since in every point, the sum of the symbols of the lines of S through it is zero, it is possible that this set is a codeword of $T_2^*(\Theta)^D$. But to make sure this set is a codeword, we have to check the other kind of points, the reguli through L.

The reguli of $Q(4, q)$ through L have to contain zero or at least two lines of the set S. So suppose that a regulus of $Q(4, q)$ through L contains the line L' of S in $\mathcal{Q}_1$ belonging to the regulus of l_2. Then the 3-dimensional space $\langle L, L' \rangle$ intersects the 3-space spanned by $\mathcal{Q}_2$ in a plane through the line l_1, so there has to lie a second line of $S \cap \mathcal{Q}_2$ intersecting l_1 in $\langle L, L' \rangle$. In order to have a codeword, the sum of the symbols of the lines of each regulus through L has to be equal to zero. This is here the case since the two lines of S in this regulus of $\mathcal{Q}(4, q)$ through L have the symbol 1. □

Theorem 3.27. *In the LDPC code defined by $T_2^*(\Theta)^D$, with $q = 2^h$,*

$h \geq 5$, Θ a regular hyperoval, every codeword c with $wt(c) \leq 4q^{3/2}/5$ is a linear combination of incidence vectors of hyperbolic quadrics such that the symbols corresponding to the coordinate positions of the lines intersecting L are zero, and such that the sum of the symbols of the lines in each regulus through L equals zero.

Proof. As seen in the proof of Proposition 3.26, we find a set of $2q-10\delta+2$ lines with constant symbol a lying in $\mathcal{Q}_i$, $i = 1, \ldots, k$.

Let l_1' be a line in $S \cap \mathcal{Q}_1$. The 'point' $\mathcal{R}(l_1', L)$ of $T_2^*(\Theta)^D$ which is the regulus through l_1' and L has to contain a second line l_2' of S. Then the line l_2' lies on a hyperbolic quadric $\mathcal{Q}'$ with $2q - 10\delta + 2$ lines of S only lying in $\mathcal{Q}'$. Suppose that one of these lines has the symbol b, then the preceding arguments lead to $2q - 10\delta + 2$ lines in $\mathcal{Q}'$ with symbol b.

We conclude that to every hyperbolic quadric $\mathcal{Q}_i$, there corresponds a value α_i which is the symbol of the lines of $\mathcal{Q}_i$, not intersecting L and not lying in an other quadric $\mathcal{Q}_j$, $j \neq i$, in the codeword.

Consider a point P lying in exactly one hyperbolic quadric $\mathcal{Q}_i$, where P does not lie on the lines of $\mathcal{Q}_i$ intersecting L. Then both lines of $\mathcal{Q}_i$ through P have symbol α_i, so the sum of the symbols of the lines of S through P is zero.

The same arguments prove that for a second point P lying in s hyperbolic quadrics $\mathcal{Q}_1, \ldots, \mathcal{Q}_s$, but not lying on any of the lines of $\mathcal{Q}_i$ intersecting L, that the sum of the symbols of the lines of $\mathcal{Q}_1, \ldots, \mathcal{Q}_s$ through P is zero.

Consider a point P of $\mathcal{Q}_1$ lying on a line of $\mathcal{Q}_1$ intersecting L. Denote this line by l_1. Since l_1 is not a line of $T_2^*(\Theta)^D$, but the sum of the symbols of the lines of S through P is zero, P lies in at least a second hyperbolic quadric $\mathcal{Q}_j$, $j > 1$. Since this must be valid for all q points of $l_1 \setminus \{P\}$, in fact, l_1 lies completely in at least a second hyperbolic quadric $\mathcal{Q}_j$, $j > 1$.

Suppose that l_1 lies in the hyperbolic quadrics $\mathcal{Q}_1, \ldots, \mathcal{Q}_r$. Using the same arguments as in [27, Lemma 6.4], we can find a point P on l_1, lying on r distinct lines in the opposite reguli of l_1 in $\mathcal{Q}_1, \ldots, \mathcal{Q}_r$. So their symbols are respectively $\alpha_1, \ldots, \alpha_r$. Since the sum of the symbols of the lines of S through P is zero, necessarily $\alpha_1 + \cdots + \alpha_r = 0$.

Consider a regulus $\mathcal{R}(l_1', L)$, where $l_1' \in S$. Suppose that l_1' lies in the hyperbolic quadrics $\mathcal{Q}_1, \ldots, \mathcal{Q}_u$, having symbols $\alpha_1, \ldots, \alpha_u$. Each hyperbolic quadric $\mathcal{Q}_i, i = 1, \ldots, u$, has a unique line l_i'' intersecting l_1' and L. For each such l_i'', the sum of symbols α_i of the hyperbolic quadrics $\mathcal{Q}_i$ in which l_i'' lies is equal to zero.

Vice versa, consider such a line l_i'' of a hyperbolic quadric $\mathcal{Q}_i$ lying in the complementary regulus of $\mathcal{R}(l_1', L)$. This hyperbolic quadric $\mathcal{Q}_i$ shares already one line l_i'' with the hyperbolic quadric containing $\mathcal{R}(l_1', L)$, so contains also a line of $\mathcal{R}(l_1', L) \setminus \{L\}$.

Consider all lines of hyperbolic quadrics in $\mathcal{Q}_1, \ldots, \mathcal{Q}_k$ lying in the regulus $\mathcal{R}(l_1', L)$.

These hyperbolic quadrics contain one line l_i'' of the opposite regulus of $\mathcal{R}(l_1', L)$. Each such hyperbolic quadric has a corresponding symbol α_i, and again for such a line l_i'', the sum of symbols α_i of the hyperbolic quadrics $\mathcal{Q}_i$ in which l_i'' lies is equal to zero. This implies that if we add up all the symbols of the lines of S in $\mathcal{R}(l_1', L)$, then this sum is zero.

We have found a linear combination of hyperbolic quadrics of $Q(4, q)$ which defines a codeword of the LDPC code of $T_2^*(\Theta)^D$.

This codeword coincides with the original codeword in all positions corresponding to the lines of S lying on exactly one of the hyperbolic quadrics of $\mathcal{Q}_1, \ldots, \mathcal{Q}_k$.

Since two hyperbolic quadrics share at most two lines, they differ in at most $k(k-1) \leq (2\sqrt{q}/5)^2 = 4q/25$ positions. So the difference of the two codewords has at most weight $8q/25$. Since the minimum distance is $4q$, the two codewords coincide. □

We have found the codewords of small weight of the LDPC code of $T_2^*(\Theta)^D$. Now we want to dualize these results to find the codewords of the LDPC code of $T_2^*(\Theta)^D$, described in terms of points and lines of $T_2^*(\Theta)$.

In Section 5.1, it is proven that $T_2^*(\Theta)^D$ can be described on $Q(4, q)$ by taking all points not on a special line L of $Q(4, q)$, and all lines not in $L^\perp$, with as special points the reguli through L.

From $T_2^*(\Theta)^D$ on $Q(4, q)$, we dualize and get $T_2^*(\Theta)$ on $Q(4, q)$ since

$Q(4,q)$, q even, is self-dual. We say that a set and its image under this duality are *corresponding*.

The minimum weight codewords of $T_2^*(\Theta)^D$ come from two hyperbolic quadrics of $Q(4,q)$, intersecting in two lines, which intersect L, so a minimum weight codeword of the code of $T_2^*(\Theta)^D$ is in the original setting $(\mathcal{P},\mathcal{L},\mathcal{I}) = T_2^*(\Theta)$ (see beginning of Section 5.1) a set of $4q$ points such that every line of $Q(4,q)$ not through P contains zero or at least two of them.

The line L corresponds to the point P, the intersection lines s_1 and s_2 of $\mathcal{Q}_1$ and $\mathcal{Q}_2$, the two hyperbolic quadrics defining a codeword of minimum weight in $T_2^*(\Theta)^D$, correspond to two points S_1 and S_2. Since s_1 and s_2 are not an element of $T_2^*(\Theta)^D$, S_1 and S_2 have to be points of $T_P(Q(4,q)) \cap Q(4,q)$, which is the set of all points of $Q(4,q)$ collinear with P. Since s_1, s_2 and L are concurrent, S_1, S_2 and P are collinear. The first hyperbolic quadric $\mathcal{Q}_1$ consists of two reguli $\mathcal{R}_1$ and $\mathcal{R}_2$, one through s_1 and one through s_2. The reguli $\mathcal{R}_1$ and $\mathcal{R}_2$ correspond to conics C_1 and C_2, respectively, in $Q(4,q)$, with $S_1 \in C_1$ and $S_2 \in C_2$.

Each line of $\mathcal{R}_1$ intersects each line of $\mathcal{R}_2$. So dually, every point on C_1 is collinear with every point on C_2. Projecting $Q(4,q)$ from its nucleus N onto a 3-dimensional space $PG(3,q)$, gives $W(3,q)$, a symplectic generalized quadrangle defined by a symplectic polarity η. In this projection, C_1 and C_2 become two lines M and M^η, because then every point of M is collinear with every point on M^η. The only conics that are projected onto a line under this projection are the conics of $Q(4,q)$ in a plane through N, and all conics in such a plane have N as their nucleus. So C_1 and C_2 are conics each lying in a plane through N, with N as nucleus.

To go from $T_2^*(\Theta)$ described on $Q(4,q)$ to the original setting $T_2^*(\Theta)$ as a linear representation in $PG(3,q)$, we project from the point P onto a 3-dimensional space $PG(3,q)$. The points S_1 and S_2 project onto the same point $S_1' = S_2'$ on a conic at infinity with nucleus N', so the conics C_1' and C_2' which are the projected conics C_1 and C_2 go through the same point $S_1' = S_2'$ at infinity.

The regulus of $\mathcal{Q}_2$ containing s_1 corresponds to a conic C_3' lying in a plane through S_1' and N'. This has to be the same plane as the one containing C_1', otherwise the lines through N' containing points of C_1' cannot have a second point in that plane. The other regulus of $\mathcal{Q}_2$ corresponds to a conic C_4' lying in the same plane as C_2'.

In order to get a codeword of the LDPC code of $T_2^*(\Theta)^D$, the conics have to be corresponding, which means that a line connecting a point of $\Theta \setminus \{S_1' = S_2', N'\}$ with a point of C_1' (C_3' resp.) intersects in C_2' (C_4' resp.).

Using this and dualizing Theorem 3.27 gives the following theorem.

Theorem 3.28. *In the LDPC code defined by $T_2^*(\Theta)^D$, with $q = 2^h$, $h \geq 5$, Θ a regular hyperoval, described in terms of points and lines of the linear representation $T_2^*(\Theta)$, every codeword c, with $wt(c) \leq 4q^{3/2}/5$, is a linear combination of incidence vectors of two by two corresponding conics with the same symbol, lying in tangent planes to the conic in Θ, such that the sum of the symbols on a line of $T_2^*(\Theta)$ is zero. In particular, the sum of symbols of the conics in one tangent plane is equal to zero.*

Bibliography

[1] B.Bagchi, A.E.Brouwer and H.A.Wilbrink, Notes on binary codes related to the $O(5,q)$ generalized quadrangle for odd q, *Geom.Dedicata*, **39** (1991), 339-355

[2] E. Boros and T. Szönyi, On the sharpness of a theorem of B. Segre, *Combinatorica*, **6** (1986), 261-268

[3] R.C. Bose, Strongly regular graphs, partial geometries and partially balanced designs, *Pacific J. Math.*, **13** (1963), No. 2, 389-419

[4] F. Buekenhout, *Handbook of incidence geometry: buildings and foundations*, North-Holland, Amsterdam (1995)

[5] S. Cauchie, F. De Clerck, and N. Hamilton, Full Embeddings of (α, β) -Geometries in Projective Spaces. *Eur. J. Comb.* 23(6) (2002) pp. 635–646.

[6] S.D. Cohen, Primitive elements and polynomials with arbitrary trace, *Disc. Math.*, **83** (1990), No. 1, 1-7

[7] A. Cossidente, On Kestenband-Ebert partition, *J. Comb. Des.*, **5** (1997), No. 5, 367-375

[8] J. Coykendall and J. Dover, Sets with few intersection numbers from Singer subgroup orbits, *European J. Combin.*, **22** (2001), No. 4, 455-464

[9] T.M. Cover, and J.A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications, (1991)

[10] P. Dembowski, *Finite Geometries*, Springer (1968)

[11] K. Drudge, On the orbits of Singer groups and their subgroups, *Elec. J. Comb.*,**9** (2002), R15 (eletronic)

[12] J.C. Fisher, J.W.P. Hirschfeld, J.A. Thas, Complete arcs in planes of square order, *Ann. Discr. Math.*, **30** (1986), 243-250

[13] M.P.C. Fossorier, Quasicyclic low-density parity-check codes from circulant permutation matrices, *IEEE Trans. Inform. Theory*, Vol. 50 (2004), 1788–1793.

[14] A. Gács and Zs. Weiner, On $(q+t)$-arcs of type $(0, 2, t)$. *Des. Codes Cryptogr.*, 29 (2003) pp. 131–139

[15] R.G. Gallager, Low-density parity check codes, *IRE Trans. Inform. Theory*, **IT-8** (1962), No.1, 21-28

[16] H. Hang, J. Xu, S. Lin and Khaled A.S. Abdel Ghaffar, Codes on Finite Geometries, *IEEE Trans. Inf. Theory*, **51** (2005), No. 2, 572-596

[17] R. Hill, *A first course in coding theory*, Oxford Applied Mathematics and Computing Science Series (1986)

[18] J.W.P. Hirschfeld, *Projective Geometries over Finite Fields*, Oxford University Press, Oxford, second edition (1998)

[19] G. Hiss, Hermitian function fields, classical unitals and representation of 3-dimensional unitary groups, *Indag. Math. (N.S.)*, **15** (2004), 223-243

[20] X.Y. Hu, M.P.C. Fossorier and E. Eleftheriou, On the computation of the minimum distance of low-density parity-check codes, *2004 IEEE International Conference on Communications*, Vol. 2 (2004), 767–771.

[21] D.R. Hughes and F.C.Piper, *Projective Planes*, Springer (1973)

[22] S.J. Johnson and S.R. Weller, Construction of low-density parity-check codes from Kirkman triple systems, *Proc. IEEE Globecom Conf., San Antonio, TX, Nov. 2001*, available at http://www.ee.newcastle.edu.au/users/sta/steve/

[23] S.J. Johnson and S.R. Weller, Regular low-density parity-check codes from combinatorial de- signs, *Proc. IEEE Inform. Theory Workshop, Cairns, Australia, Sep. 2001*, 90–92.

[24] S.J. Johnson and S.R.Weller, Codes for iterative decoding from partial geometries, *Proc. IEEE Int. Symp. Inform. Theory, Switzerland, June 30 - July 5*, (2002), 6 page, extended abstract, available at http://murray.newcastle.edu.au/users/staff/steve/

[25] S.J. Johnson and S.R. Weller, High-rate LDPC codes from unital designs. In Proceedings of the IEEE Globecom Conference, San Francisco, CA, 1-5 December 2003.

[26] S.J. Johnson and S.R. Weller, Codes for iterative decoding from partial geometries, *IEEE Trans. Comm.*, **52** (2004), No. 2, 236-243

[27] J.L. Kim, K. Mellinger, L. Storme. Small weight codewords in LDPC codes defined by (dual) classical generalized quadrangles. *Des. Codes Cryptogr.*, 42(1) (2007), 73–92.

[28] 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*, Vol. 50 (2004), 2378–2388.

[29] G. Korchmáros and F. Mazzocca. On $(q+t)$-arcs of type $(0, 2, t)$ in a desarguesian plane of order q. *Math. Proc. Camb. Phil. Soc.*, 108 (1990),445–459.

[30] V.D. Kolesnik, Probabilty deciding of majority codes, *Probl. Pered. Inform.*, **7** (1971), No. 7, 3-12

[31] Y. Kou, S. Lin and M.P.C. Fossorier, Low-density parity check codes based on finite geometries: a rediscovery and new results, *IEEE Trans. Inf. Theory*, **47** (2001), No. 7, 2711-2736

[32] R. Lidl, H. Niederreiter, *Finite fields*, Cambridge University Press (1997)

[33] Z. Liu, D.A. Pados, LDPC codes from generalized polygons. *IEEE Trans. Inform. Theory*, 51(11) (2005), 3890–3898.

[34] S. Lin, D.J. Costello Jr., *Error Control Coding: Fundamentals and Applications*, Prentice–Hall, (2004).

[35] D.J.C. MacKay, Good error correcting codes based on very sparse matrices, *IEEE Trans. Inform. Theory*, Vol. 45 (1999),399–431.

[36] D.J.C. MacKay and M.C. Davey, Evaluation of Gallager codes for short block length and high rate applications, *Codes, Systems and Graphical Models, B. Marcus and J. Rosenthal, editors,* Vol. 123, IMA in Math. and its Appl., Springer-Verlag, New York, (2000), 113–130.

[37] D.J.C. MacKay and R.M. Neal, Near Shannon limit performance of low density parity check codes, *Electron. Lett.*, Vol. 32, No. 18 (1996), 1645–1646.

[38] F.J. MacWilliams and N.J.A. Sloan, *The Theory of Error–Correcting Codes*, New York: North Holland (1977).

[39] G.A. Margulis, Explicit constructions of graphs without short cycles and low density codes, *Combinatorica*, Vol. 2 (1982), 71–78.

[40] J.L. Massey, *Thre shold Decoding*, MIT Press, (1963).

[41] F. Pambianco and L. Storme, Small Complete Caps in Spaces of Even Characteristic. *J. Comb. Theory, Ser. A* 75(1) (1996) pp. 70–84

[42] S.E. Payne, Quadrangles of order $(s-1, s+1)$. *J. Algebra* 22 (1972) pp. 97–119

[43] S.E. Payne and J.A. Thas, *Finite Generalized Quadrangles*, Pitman Advanced Publishing Program, (1984).

[44] V. Pepe, *LDPC Codes from the Hermitian Curve*, *Des. Codes Cryptogr.*, 42(3) (2007), 303–315.

[45] V. Pepe, L.Storme, G.Van de Voorde, *Small weight codewords in the LDPC codes arising from linear representations of geometries* , submitted to *Journal of Combinatorial Designs*.

[46] W.W. Peterson, E.J. Weldon Jr., *Error Correcting Codes*, MIT Press (1972).

[47] J. Rosenthal and P.O. Vontobel, Construction of LDPC codes using Ramanujan graphs and ideas from Margulis. *Proc. 38th Allerton Conf. on Communications, Control, and Computing, Monticello, IL, Coordinated Science Lab., P.G Voulgaris and R. Srikant, Eds., Oct. 4-6*, (2000), 248–257.

[48] B. Segre, On complete caps and ovaloids in three-dimensional Galois spaces of characteristic two. *Acta Arith.* 5 (1959),315–332.

[49] M. Sipser and D.A. Spielman, Expander codes, *IEEE Trans. Inform. Theory*, Vol. 42 (1996), 1710–1722.

[50] R. M. Tanner. A recursive approach to low complexity codes, *IEEE Trans. Inform. Theory*, IT-27: 533–547, 1981.

[51] R. M. Tanner. Minimum distance bounds by graph analysis. *IEEE Trans. Inform. Theory*, 47(2): 808–821, 2001.

[52] R.M. Tanner, D. Sridhara, A. Sridharan, T.E. Fuja and D.J. Costello, Jr., LDPC block and convolutional codes based on circulant matrices, *IEEE Trans. Inform. Theory*, Vol. 50 (2004), 2966–2984.

[53] P. O. Vontobel and R M. Tanner, Construction of codes based on finite generalized quadrangles for iterative decoding, *Proceedings of 2001 IEEE Intern. Symp. Inform. Theory, Washington, DC*, (2001), 223.

[54] S.R. Weller and S.J. Johnson, Regular low-density parity-check codes from oval designs, *European Transactions on Telecommunications*, Vol. 14, No. 5 (2003),399–409.

[55] J.H. van Lint, *Coding Theory*, Springer (1971)