

# Euclidean Geometry Codes, minimum weight words and decodable error-patterns using bit-flipping

Tom Høholdt

Department of Mathematics  
Technical University of Denmark  
Dk-2800, Lyngby, Denmark  
Email: T.Hoeholdt@mat.dtu.dk

Jørn Justesen

COM  
Technical University of Denmark  
DK-2800 Lyngby, Denmark  
Email: jju@com.dtu.dk

Bergtor Jonsson

Department of Mathematics  
Technical University of Denmark  
DK-2800 Lyngby, Denmark  
Email: beggi@beggi.com

**Abstract**—We determine the number of minimum weight codewords in a class of Euclidean Geometry codes and link the performance of bit-flipping decoding to the geometry of the error patterns.

## I. INTRODUCTION

Euclidean geometry codes are classical in coding theory and it is well known that they can be decoded up to half the minimum distance using majority logic decoding [1]. In [2] the authors realized that these and related codes were indeed LDPC-codes and that they performed remarkably well under various iterative bit-flipping decoding algorithms. The present paper is an attempt to give a better understanding of the reasons for this, in the sense that we use the geometry to explain the performance of a bit-flipping algorithms.

We have chosen to investigate a special class of codes from Euclidean geometry, this is given in Section II, which also contains a theorem giving the number of minimum weight codewords. In Section III we link the performance of bit-flipping decoding to the geometry of the error patterns.

## II. THE CODES AND THE NUMBER OF MINIMUM WEIGHT WORDS

Let  $EG(2, q)$  where  $q = 2^s$  denote the Euclidean plane over the finite field  $\mathbb{F}_q$ . Remove the point at the origin and all the lines through the origin. This leaves  $n = q^2 - 1$  points and  $b = q^2 - 1$  lines. Choose a numbering of the points  $P_1, P_2, \dots, P_n$  and of the lines  $L_1, L_2, \dots, L_b$  and define the matrix

$$H = (h_{uv}) \text{ where } h_{uv} = \begin{cases} 1 & \text{if } P_v \in L_u \\ 0 & \text{if } P_v \notin L_u \end{cases}$$

This means that we can number the positions in a word with the points of the geometry and that a word is a codeword iff it is orthogonal to all the incidence vectors of the lines.

It is well known [2] that the numbering of the points and lines can be chosen in such a way that the matrix  $H$  is circulant. Using  $H$  as the parity check matrix for a binary code  $K(s)$  we get a code of length  $n = 2^{2s} - 1$ , dimension  $k = n - 3^s + 1$  and minimum distance  $d = 2^s + 1$ . By construction it follows that the matrix  $H$  has  $j = 2^s$  ones in each row and  $i = 2^s$  ones in each column. Moreover any two rows have at most one 1 in common and any two columns have at most one 1 in common. With the proper numbering this is a cyclic code.

*Example 1:* With  $s = 2$  we get a  $(15, 7, 5)$  code with  $i = j = 4$ , this is the double error correcting quasiperfect BCH code.

We will now count the number of minimum weight codewords in the codes described above. The first observation is

*Lemma 1:* Let  $c$  be a codeword of minimum weight  $2^s + 1 = j + 1$ . The set  $\mathcal{C}$  of points corresponding to the nonzero positions satisfy

- 1) Any two points lie on a ( unique) line.
- 2) Any line has either 2 or 0 points in common with  $\mathcal{C}$

and if we have a set of  $q + 1$  points satisfying 1) and 2) the incidence vector corresponds to a minimum weight codeword. Proof: It follows immediately from the fact that  $Hc^T = 0$  that there must be an even number of ones in common between  $c$  and every row of  $H$ . From the cyclic structure of the matrix it is also clear that there is a line ( which by construction must be unique) with 2 or more points in common with  $c$ . To finish the proof let  $\rho > 1$  be the number of ones that  $c$  has in common with a fixed line  $L$ . For each one that  $L$  and  $c$  have in common there are  $j - 1$  other lines that have a one that has to be cancelled, and we have  $j + 1 - \rho$  remaining ones to do that. But for  $\rho > 2$  we have  $j + 1 - \rho < j - 1$  so this is impossible. Therefore  $\rho \leq 2$ . The reverse statement is obvious from the structure of  $H$ .  $\square$

We will now study the set  $\mathcal{C}$  of  $q + 1$  points from a minimum weight codeword a little closer.

In  $EG(2, 2^s)$  the set  $\mathcal{O} = \mathcal{C} \cup (0, 0)$  is a set of  $q + 2$  points with the property that any two points are on a ( unique) line and no three points are collinear. Moreover the lines through  $(0, 0)$  have exactly 1 point in common with  $\mathcal{C}$ . This means that  $\mathcal{O}$  is an *regular oval*. For the definition and results on these see [3]. It is also known that an oval in  $EG(2, 2^s)$  consists of a conic and its nucleus.

*Theorem 1:* The number of minimum weight codewords in  $K(s)$  is

- 1) 18 if  $s = 2$
- 2)  $(2^s + 2)(2^s - 1)^2 2^{s-1}$  if  $s > 2$

Proof: 1) If  $s = 2$  we have 6 points in  $EG(2, 4)$  no three on a line, it follows from ([3] p. 184) that these form a (nondegenerate) conic and a nucleus in 6 different ways, so we can choose the nucleus to be  $(0, 0)$ .

Then the equation for the conic is

$$ax^2 + bxy + cy^2 + 1 = 0 \text{ with } abc \neq 0 \text{ and } Tr\left(\frac{ca}{b}\right) = 1, \\ \text{where } Tr \text{ is the trace mapping from } \mathbb{F}_4 \text{ to } \mathbb{F}_2.$$

This can be seen from the fact that the general equation for a conic not passing through  $(0, 0)$  is  $ax^2 + bxy + cy^2 + dx + ey + 1 = 0$  and the condition that all lines through  $(0, 0)$  shall be tangents. This implies that the lines with equations  $x = 0$  and  $y = kx$  has exactly one point in common with the conic and from that one derives the conditions on the equation.

From this it is easily seen that the number of such conics is  $3^2 \times 2 = 18$ .

2) In the case  $s > 2$  every point of  $\mathcal{O}$  could be the nucleus so we have that the number of minimal weight codewords is  $(q + 2) \times$  (the number of regular ovals with nucleus  $(0, 0)$ ). But again this means that we have a conic with equation

$$ax^2 + bxy + cy^2 + 1 = 0 \text{ with } abc \neq 0 \text{ and } Tr\left(\frac{ca}{b}\right) = 1, \\ \text{where } Tr \text{ is the trace mapping from } \mathbb{F}_{2^s} \text{ to } \mathbb{F}_2.$$

and it is easily seen that the number of such conics is  $(2^s - 1)^2 \times 2^{s-1}$ .  $\square$

The above observation can also be used to generate all the minimum codewords.

### III. DECODING EUCLIDEAN GEOMETRY CODES USING BIT-FLIPPING

Euclidean Geometry codes can be decoded by versions of the iterative decoding method for LDPC codes known as bit-flipping. The decoder works on generalized syndromes  $s = Hr^T$ , where  $r$  is the vector at the output of a binary symmetric channel. (generalized because  $s$  has  $n$  bits rather than  $n - k$ ). The error pattern is described as a set of points in the Euclidean plane, and each parity check corresponds to a line in the plane. For each position in  $r$  we calculate the number of lines through the point where the parity check fails. In each step, one or more of the bits with a maximal number of parity failures is flipped, and a new syndrome is calculated. The performance of different variations of the bit-flipping algorithm may be analysed by considering the geometry of the error patterns. We introduce the concepts of the analysis by considering a small example.

*Example 2:* The  $(15, 7, 5)$  code.

The Euclidean Geometry code with  $s = 2$  has minimum distance 5, and all coset leaders have weight  $\leq 3$ . The decoding results for the bit-flipping algorithm can be explained

by considering the weights  $w(s)$  of the generalized syndromes, which are given in the table 2.

$w(s)$	number of cosets	description of error patterns
0	1	no errors
4	15	one error
6	100	90 double errors on a line and 10 cosets with 3 triple errors
8	75	15 double errors not on a line and 60 cosets with 3 triple errors
10	60	three errors on a line
12	5	three errors, no two on a line

Two errors are always decoded by majority decisions, and triple errors in those cosets are obviously not decoded. For syndrome weights 10 and 12 the error positions have a majority of parity failures, and they are decoded. In the cosets of weight 6 and 8 with coset leaders of weight 3, the result depends on the details of the algorithm. For  $w(s) = 6$  no position has more than 2 parity failures. In the cosets with  $w(s) = 8$  two of the errors are not on a line, and in this case the result may be one of the three closest codewords or a codeword further away depending on the scheduling of the algorithm.

For the class of codes we consider we have the following general observation.

*Lemma 2:* If  $t$  errors occur,  $t \geq \frac{d}{2}$  and  $w(s) > t^2$ , the errors are decoded by the bit-flipping algorithm, and hence there is a unique coset leader.

Proof: Let  $r_j$  denote the number of parity failures that involves position  $j$  of the received word. We then have

$$1) \quad w(s) \leq \sum_{\text{error positions}} r_j$$

2) If  $r_j > t$  then there is an error at position  $j$  and it is corrected by the bit-flipping algorithm.

Here 1) is obvious and 2) follows from the facts that  $r_j \leq q$  and  $t \geq \frac{d}{2} = \frac{q+1}{2}$ .

Since by assumption  $w(s) > t^2$  we have from 1) that  $\sum_{\text{error positions}} r_j > t^2$  and since there are  $t$  errors there is at least one position  $j$  with  $r_j > t$  and this is corrected.

After correction the new syndrome has weight  $w'(s) \geq w(s) - r_j \geq w(s) - q \geq t^2 + 1 - q \geq (t - 1)^2 + 2t - q > (t - 1)^2$  since  $2t > q$ . The proof now follows by repeating this argument.  $\square$

Thus for high syndrome weights, the errors are decoded. For low syndrome weights, there may be an error pattern of lower weight in the same coset or decoding may fail because no position has enough parity failures. For intermediate syndrome weights, there are more error patterns of the same weight in each coset, and the result depends on the number of bits that are flipped in each step and the order in which they are selected.

*Example 3:* The (63, 37, 9) code.

For  $s = 3$  we get the following distribution of syndrome weights ( this is the weight distribution of the dual code and is found by computer).

TABLE I

SYNDROME WEIGHT DISTRIBUTION OF COSETS FOR THE (63, 37, 9) CODE.

$w(s)$	#
0	1
8	63
14	1800
16	189
18	27048
20	243936
22	740880
24	2317035
26	4845456
28	9912672
30	11755296
32	13966659
34	10372320
36	7709856
38	3315312
40	1390221
42	388080
44	110880
46	10584
48	63
50	504
56	9

All error patterns with at most 4 errors are decoded by majority decisions. From Lemma 2 it follows that 5 errors are decoded when the syndrome weight is  $26 - 40$ , 6 errors for  $w(s) = 38 - 48$ , and 7 errors for  $w(s) = 50$  or  $56$ .

For 5 errors and  $w(s) = 20$ , the error pattern is in a coset with minimum weight 4 or the patterns are not decoded. For  $w(s) = 22$  and  $24$  the various cases can be described by their geometry and counted.

#### IV. CONCLUSION

We have determined the number of minimum weight words in a class of Euclidean Geometry codes and demonstrated how the performance of a bit-flipping algorithm can be explained by the geometry of the error patterns.

#### REFERENCES

- [1] I.F. Blake and R.C. Mullin *The Mathematical Theory of Coding* New York: Academic Press 1975 pp.112-116
- [2] Y. Kou, S. Lin and M. Fossorier: "Low-Density Parity-Check Codes Based on Finite Geometries: A Rediscovery and New Results" *IEEE Trans.Inform.Theory* vol.47/7 pp. 2711-2736, Nov. 2001.
- [3] J.W.P. Hirschfeld: *Projective Geometries over Finite Fields 2. ed.* Oxford: Oxford Science Publications 1998.