ALBANIAN JOURNAL OF MATHEMATICS
Volume 10, Number 1, Pages 37–45
ISSN: 1930-1235; (**2016**)

GRAPH BASED LINEAR ERROR CORRECTING CODES

MONIKA POLAK
University of Maria Curie–Sklodowska
Lublin, Poland
Email: monika.katarzyna.polak@gmail.com

EUSTRAT ZHUPA
University of Maria Curie–Sklodowska
Lublin, Poland
Email: e.zhupa@gmail.com

ABSTRACT. In this article we present a construction of error correcting codes, that have representation as very sparse matrices and belong to the class of Low Density Parity Check Codes. LDPC codes are in the classical Hamming metric. They are very close to well known Shannon bound. The ability to use graphs for code construction was first discussed by Tanner in 1981 and has been used in a number of very effective implementations. We describe how to construct such codes by using special a family of graphs introduced by Ustimenko and Woldar. Graphs that we used are bipartite, bi-regular, very sparse and do not have short cycles C_4. Due to the very low density of such graphs, the obtained codes are fast decodable. We describe how to choose parameters to obtain a desired code rate. We also show results of computer simulations of BER (bit error rate) of the obtained codes in order to compare them with other known LDPC codes.

1. INTRODUCTION

All information in a computer is represented as a sequence of binary digits. Data are stored and shared with others. Both in the case of data storage and during data transfer, we need protection against transmission errors or data loss. In the first scenario, unreliable or faulty hardware (computer memories, compact discs, QR Code) can seriously corrupt data. Coding techniques are one of the important measures for improving reliability.

As to the second scenario, very often digital data are sent over unreliable communication channels (air, a telephone line, a beam of light or a cable). Because of the channel, noise errors may be introduced during transmission from the source to the receiver. It is very important for the recipient to receive the correct message as intended. In order to minimize the number of errors during transmission, we can use error correcting codes.

Coding of information with the use of error correcting codes consists of adding to sequences of K elements some extra bits in a certain way. Such additional bits don't carry any information and they only have check purposes. An error correcting code is $A \subset \mathbb{F}_2^N$, where $\mathbb{F}_2 = \{0, 1\}$ and codewords follow the classical Hamming metric:

$$d(x, y) = |\{i : x_i \neq y_i\}|.$$

2000 *Mathematics Subject Classification.* 94B05 and 94B05 and 05C50 .
Key words and phrases. LDPC codes and sparse graph and graph based algorithm and quality of transmission.

We denote with $[N, K]$ the code with code words length N and K information bits. In such a code there are $R = N - K$ parity check equations. The ratio K/N is called *code rate* and is denoted by R_C. It is interesting to look at codes with the best correction properties and the biggest code rate for cost reasons.

Linear error correcting code can be represented in three ways: by the generator matrix G, by parity check matrix H or by Tanner graph $\Gamma(V, E)$. Parity check matrix H for $[N, K]$ code is $R \times N$ matrix whose words are zeros or ones. Rows of such matrix correspond to the parity checks and the columns correspond to codeword bits. If bit number j in the codeword is checked by parity check number i then in position (i, j) in matrix H there is a 1, if not there is a 0. Switching column does not change code properties and provides an equivalent code. A simple example of linear error correcting code is Hamming code $[7, 4]$ with matrix H of the form:

$$H = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}$$

To encode vector of 4 bits by using $[7, 4]$ Hamming code we add 3 extra bits. Each bit is checked by a unique set of control equations. We assume that every codeword is from the set

$$C = \{y \in \mathbb{F}_2^N : Hy^T = 0\}.$$

Let's recall few simple facts from Graph Theory that can be found for example in (Biggs, 1993). The *distance* between vertices v_1 and v_2 of the graph is the length of minimal path from v_1 and v_2. The graph is connected if for arbitrary pair of vertices v_1, v_2 there is a path from v_1 to v_2. The *girth* of simple graph is the length of the shortest cycle in graph. A *bipartite* graph is a graph $\Gamma(V, E)$, in which a set of nodes V can be divided into two subsets $V = V_1 \cup V_2$ ($V_1 \cap V_2 = \emptyset$) in such a way that no two vertices from each set V_i , $i = 1, 2$ are connected with an edge. We refer to bipartite graph $\Gamma(V, E)$ with partition sets V_i, $i = 1, 2$, $V = V_1 \cup V_2$ as bi-regular one if the number of neighbours for representatives of each partition set are constants s and r (bi-degrees); we say that we have bi-regularity (s, r). We call the graph *regular* in the case $s = r$.

Tanner in 1981 introduced an effective graphical representation for LDPC Tanner codes, i.e. Tanner graph. *Tanner graph* is the bipartite graph in which one subset V_1 corresponds to the codeword bits and second V_2 to the parity checks. Vertex from the set V_1 is connected to a vertex from the set V_2 if and only if a bit corresponding to the vertex from V_1 is controlled by the parity check corresponding to the vertex from V_2.

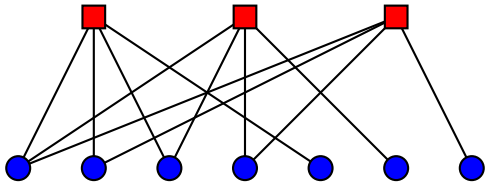

FIGURE 1. Tanner graph for $[7, 4]$ Hamming code

Tanner graph represents parity check equations. There is a standard way to create error correcting code codes depending on adjacency matrix of bipartite, bi-regular graph. Parity

check matrix H is a part of the adjacency matrix A of the graph with desired properties used to create the code:

$$A = \begin{pmatrix} 0 & H \\ H^T & 0 \end{pmatrix}$$

Determination of the matrix H is equivalent to code designation. However parity check matrix is not unique. Switching columns doesn't change code properties and gives us an equivalent code.

One famous class of error correcting codes is Low Density Parity Check Codes introduced in 1962 by Robert G. Gallager (Gallager, 1962). Such codes have a wide range for selection of parameters, enabling creation of codes with a large block size and excellent correction properties. In the present work we show new results on applications of Computer Algebra and Theory of Algebraic Graphs in constructions of new LDPC error correcting codes. Only specific graphs are suitable for such purpose. Usually, simple graphs are used. The graph should be bipartite, sparse, without small cycles and bi-regular or regular with the possibility to obtain bi-regularity. Codes related to graph defined below have all such important properties.

A code with a representation as a sparse matrix or a sparse Tanner graph is a Low-Density Parity-Check Code, (Gallager, 1962). A matrix is sparse if number of ones in it is small compared to number of zeros. Low Density Parity Check Codes have a very sparse parity check matrix. A *sparse* graph has a small number of edges in relation to the number of vertices. A simple relationship describing the density of the graph $\Gamma(V, E)$ is

$$D = \frac{2|E|}{|V|(|V|-1)}, \tag{1}$$

where $|E|$ is the number of edges of graph Γ and $|V|$ the number of vertices.

2. DESCRIPTION OF THE GRAPH

Let q be a prime power and let $\mathbb{F}_{q^2}$ be the quadratic extension of $\mathbb{F}_q$. The family of graphs $F = F(\mathbb{F}_q, \mathbb{F}_{q^2})$ was introduced in (Ustimenko and Woldar, 2003). In fact, the described graphs are affine part of generalized quadrangles. The following representation can be found in (Ustimenko, 2011), where they are denoted as $I4_q$. Those graphs are bipartite with set of vertices $V = V_1 \cup V_2$, where $V_1 \cap V_2 = \emptyset$. They have girth at least 8 and very different bi-regularity (q, q^2). Traditionally because of geometric construction, one partition set $V_1 = P$ is called set of points and other $V_2 = L$ is called the set of lines:

$$P = \{(a, b, c) : a \in \mathbb{F}_q, b \in \mathbb{F}_{q^2}, c \in \mathbb{F}_q\},$$

$$L = \{(x, y, z) : x \in \mathbb{F}_{q^2}, y \in \mathbb{F}_{q^2}, z \in \mathbb{F}_q\}.$$

Two types of brackets are used to distinguish points and lines. Let $x \longrightarrow x^q$ be the Frobenius automorphism of $\mathbb{F}_{q^2}$. We say point (p) is *incident* to line $[l]$ in graph $F(\mathbb{F}_q, \mathbb{F}_{q^2})$, and we write $(p)I[l]$, if the following relations on their coordinates hold:

$$\begin{cases} y - b = ax \\ z - c = ay + ay^q \end{cases} \tag{2}$$

The set of vertices is $V(F) = P \cup L$ and the set of edges consists of all pairs $((p), [l])$, for which $(p)I[l]$. Because $a \in \mathbb{F}_q, b \in \mathbb{F}_{q^2}, c \in \mathbb{F}_q, x \in \mathbb{F}_{q^2}, y \in \mathbb{F}_{q^2}, z \in \mathbb{F}_q$ we have $|P| = q^4$ and $|L| = q^5$ ($|V(F)| = q^5 + q^4 = q^4(q+1)$). This is a family of simple graphs.

Proposition 2.1. *$F(\mathbb{F}_q, \mathbb{F}_{q^2})$ is a family of sparse graphs.*

Proof. If we set point $(p) = (a, b, c)$ and x coordinate of line $[l]$ ($x \in \mathbb{F}_{q^2}$) then we can calculate y and z from 2. There is just one solution. So it's easy to see that each point has exactly q^2 neighbours. We have $|P| = q^4$ so set of edges has $|E| = q^4 \cdot q^2 = q^6$ elements. Using formula 1 the density of described graphs is

$$D = \frac{2q^6}{q^4(q+1)(q^4(q+1)-1)} = \frac{2q^2}{2q^6+2q^5+q^4-1} \approx \frac{2}{2q^4+2q^3+q^2}.$$

□

Instead of using elements of fields $\mathbb{F}_{q^2}$ and $\mathbb{F}_q$ as coordinates, we propose to use two rings $\mathbb{Z}_{n^2}, \mathbb{Z}_n$ and modulo operations. In such case sets P and L for the graph $F(\mathbb{Z}_n, \mathbb{Z}_{n^2})$ are the following :

$$P = \{(a, b, c) : a \in \mathbb{Z}_n, b \in \mathbb{Z}_{n^2}, c \in \mathbb{Z}_n\},$$
$$L = \{(x, y, z) : x \in \mathbb{Z}_{n^2}, y \in \mathbb{Z}_{n^2}, z \in \mathbb{Z}_n\}.$$

We say point (p) is *incident* to line $[l]$, and we write $(p)I[l]$, if the following relations on their coordinates hold:

$$\begin{cases} (y-b) \mod n^2 = (ax) \mod n^2 \\ (z-c) \mod n = (ay + ay^n) \mod n \end{cases} \tag{3}$$

Graphs defined in terms of finite rings as coordinates are bipartite, have girth at least 6 (probably 8, but not tested) and bi-regularity (n, n^2). In this case they are not affine part of generalized quadrangles. Set of lines L has n^5 elements and $|P| = n^4$. Set of vertices has $V = n^4(n+1)$ elements and set of edges consists of n^6 elements. By analogy with 2.1, the density of graphs $F(\mathbb{Z}_n, \mathbb{Z}_{n^2})$ is

$$\frac{2}{nq^4+nq^3+n^2}.$$

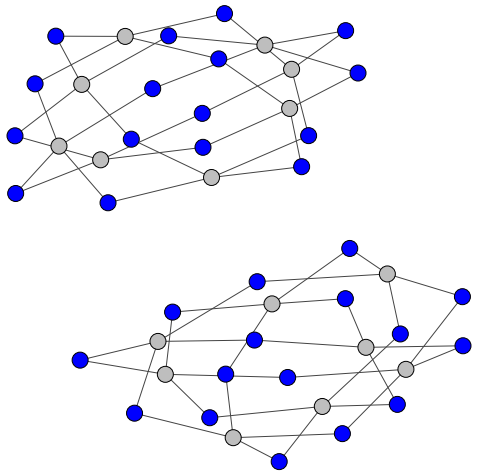

FIGURE 2. Graph $F(\mathbb{Z}_2, \mathbb{Z}_4)$

Table 1 contains some properties of small representatives of described families. Figure 2 shows the smallest example: $F(\mathbb{Z}_2, \mathbb{Z}_4)$ with $|P| = 16$ (grey colour) and set of lines with $|L| = 32$ elements. Each point has exactly 4 neighbours and each line has two neighbours.

TABLE 1. Properties of described graphs, for example number fields and finite rings

Graph	Graph density	biregularity	$\lvert V\rvert$	$\lvert E\rvert$
$F(\mathbb{F}_3,\mathbb{F}_9)$, $F(\mathbb{Z}_3,\mathbb{Z}_9)$	≈ 0.014	(3,9)	324	729
$F(\mathbb{F}_4,\mathbb{F}_{16})$, $F(\mathbb{Z}_4,\mathbb{Z}_{16})$	≈ 0.005	(4,16)	1280	4096
$F(\mathbb{F}_5,\mathbb{F}_{25})$, $F(\mathbb{Z}_5,\mathbb{Z}_{25})$	≈ 0.002	(5,25)	3750	15625
$F(\mathbb{Z}_6,\mathbb{Z}_{36})$	≈ 0.001	(6,36)	9072	46656
$F(\mathbb{F}_7,\mathbb{F}_{49})$, $F(\mathbb{Z}_7,\mathbb{Z}_{49})$	≈ 0.0006	(7,49)	19208	117649

The advantages of using finite rings (and modulo operations) instead of prime powers is that graphs $F(\mathbb{Z}_n,\mathbb{Z}_{n^2})$ are defined for all $n \geq 2$.

2.1. **Code construction.** Our graphs F are already biregular but they also have a structure that enables us to remove points and lines in such a way as to obtain a subgraph with different bidegree. This operation yields a code with better error correcting properties, but it reduces code rate and makes the code less convenient. We can construct LDPC codes in two ways:

(1) In graphs $F(\mathbb{F}_q,\mathbb{F}_{q^2})$ or $F(\mathbb{Z}_n,\mathbb{Z}_{n^2})$z set of lines is bigger than set of points: $|L| > |P|$, so lines correspond to code words bits and points correspond to parity checks. We decide to put one or zero in parity check matrix by checking if relations 2 between corresponding points and lines hold. Every bit of the codeword is checked by q or n parity checks. In this case parity check matrix H is simply a part of adjacency matrix of used graph. $|L| = q^5$ and $|P| = q^4$ (or $|L| = n^5$ and $|P| = n^4$) so the size of H is $q^4 \times q^5$ ($n^4 \times n^5$) and

$$R_C = \frac{q^5 - q^4}{q^5} = \frac{q^5(q-1)}{q^5} = \frac{(q-1)}{q}$$

($R_C = \frac{n-1}{n}$).

(2) Recalling that L is the set of all lines and P is the set of all points in graph F, in order to obtain a bi-degree (q,r) different from (q,q^2), where $q < r < q^2$ we propose to put restrictions on the coordinates ((n,r) different from (n,n^2), where $n < r < n^2$) in the following way. Let $R \subset \mathbb{F}_{q^2}$ ($R \subset \mathbb{Z}_{n^2}$) be an r-element subset and let V_L be the set of lines in a new bipartite graph. This is the following set:

$$V_L = \{[l] \in L | x \in R\},$$

where lines $[l]$ are represented by vectors $[x,y,z]$. Bi-degree reduction can only increase the girth. After reduction, the bi-degree graph can be disconnected and we use only one component to create a parity check matrix H. In this case parity check matrix H is a part of adjacency matrix of a subgraph of the used graph. $|V_L| = r \cdot q^3$ and $|P| = q^4$ (or $|V_L| = r \cdot n^3$ and $|P| = n^4$) so the size of H is $q^4 \times r \cdot n^3$ ($n^4 \times r \cdot n^3$) and

$$R_C = \frac{r \cdot q^3 - q^4}{r \cdot q^3} = 1 - \frac{q}{r} = \frac{(r-q)}{q}$$

($R_C = \frac{r-n}{n}$).

In first case $r = q^2$ and $s = q$ ($r = n^2$ and $s = n$). In the second case we choose r and $s = q$ ($s = n$). In regular LDPC code every row has the same constant weight r and every

TABLE 2. Properties of described $[N, K]$ codes for example number fields and finite rings

Graph	$N = \|L\|$	$R = \|P\|$	$K = N - R$	R_C	Girth
$F(\mathbb{F}_2, \mathbb{F}_4)$	2^5	2^4	16	0.5	≥ 8
$F(\mathbb{Z}_2, \mathbb{Z}_4)$	2^5	2^4	16	0.5	≥ 6
$F(\mathbb{F}_3, \mathbb{F}_9)$	3^5	3^4	162	≈ 0.67	≥ 8
$F(\mathbb{Z}_3, \mathbb{Z}_9)$	3^5	3^4	162	≈ 0.67	≥ 6
$F(\mathbb{F}_4, \mathbb{F}_{16})$	4^5	4^4	768	0.75	≥ 8
$F(\mathbb{Z}_4, \mathbb{Z}_{16})$	4^5	4^4	768	0.75	≥ 6
$F(\mathbb{F}_5, \mathbb{F}_{25})$	5^5	5^4	2500	0.8	≥ 8
$F(\mathbb{Z}_5, \mathbb{Z}_{25})$	5^5	5^4	2500	0.8	≥ 6
$F(\mathbb{Z}_6, \mathbb{Z}_{36})$	6^5	6^4	6480	≈ 0.83	≥ 6
subgraph of $F(\mathbb{Z}_5, \mathbb{Z}_{25})$	2000	625	1375	≈ 0.69	≥ 6
subgraph of $D(6, 7)$	2401	686	1715	≈ 0.71	≥ 10
subgraph of $D(8, 5)$	625	250	375	0.6	≥ 12
subgraph of $D(10, 3)$	243	162	81	≈ 0.33	≥ 14

column has the same constant weight s. Despite the fact that graphs F have very different bidegree, they give us regular LDPC codes. Every column of parity check matrix H has the same weight.

2.2. **Corresponding LDPC codes.** Transmission quality depends mainly on code, on decoding algorithm and level of noise in the communication channel. Code error correcting properties are often tested by determining the relationship between noise level and bit error rate. *Bit error rate* (BER) is the ratio of number of error bits to the total number of transferred bits. The generated codes are based on the described graphs and we performed computer simulation of a noisy channel. We encoded vectors, added White Gaussian Noise and decoded. The lack of short cycles guaranty convergence of the decoding algorithm. Presented LDPC codes work well with existing decoding algorithms, so there is no need do implement dedicated decoding technique. Table 2 contains properties of described $[N, K]$ codes and other previously known codes in order to compare thein properties. Fig. 3 and Fig. 4 present bit error rate for LDPC codes corresponding to described graphs. The code rates R_C of such codes are bigger than 0.5, so they are very convenient and have a small number of redundant bits. Tab. 2 contains properties of codes, for which results of the simulation are presented in this article, and parameters for a code based on graphs $D(n, q)$, $A(n, q)$ with similar code rate.

Codes corresponding to graphs from special family $D(n, q)$ were presented in (Guinand and Lodge, 1997a). The presented codes have as good error correcting properties as codes constructed by Guinand and Lodge in (Guinand and Lodge, 1997b), whose construction is based on infinite family of graphs $D(n, q)$ constructed by Lazebnik and Ustimenko (Lazebnik and Ustimenko, 1993). In the case $q = p$ is a prime number the MATLAB code to generate LDPC codes is available in (Polak, 2016).

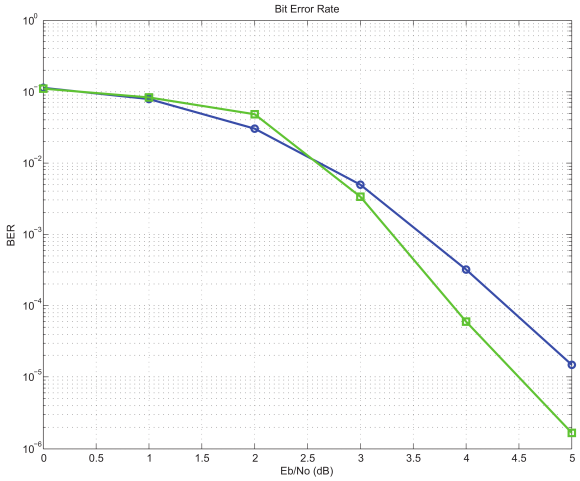

FIGURE 3. Bit error rate for $[243, 162]$ code (blue) corresponding to graph $F(\mathbb{F}_3, \mathbb{F}_9)$ and $[1024, 768]$ code (green) corresponding to graph $F(\mathbb{F}_4, \mathbb{F}_{16})$

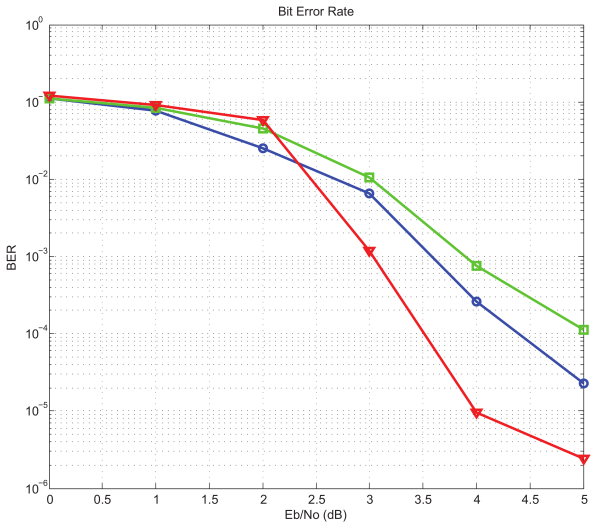

FIGURE 4. Bit error rate for $[243, 162]$ code (blue) corresponding to graph $F(\mathbb{Z}_3, \mathbb{Z}_9)$, $[1024, 768]$ code (green) corresponding to graph $F(\mathbb{Z}_4, \mathbb{Z}_{16})$ and $[2000, 1375]$ code (red) corresponding to subgraph of $F(\mathbb{Z}_5, \mathbb{Z}_{25})$

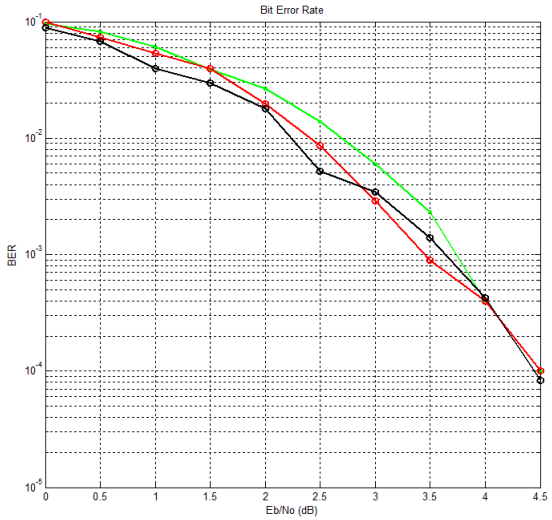

FIGURE 5. Bit error rate for $[2401, 1715]$ code (green) corresponding to graph $D(6, 7)$, $[625, 375]$ code (red) corresponding to graph $D(8, 5)$ and $[243, 81]$ code (black) corresponding to graph $D(10, 3)$

Fig. 5 presents bit error rate for example LDPC codes corresponding to graphs $D(n, q)$. Green and red codes have similar code rate like codes based on graphs F. Comparing figures Fig. 3 and Fig. 4 with Fig. 5 we can see that our codes perform very well and have better error correcting properties.

3. CONCLUSION

The present work shows a modification of the family of graphs introduced by Ustimenko and Woldar by changing the finite field to a finite ring. LDPC codes are constructed based on the original family and on the modified one. An analysis is performed on such codes as well as a BER simulation. A comparison of the results with former graph based codes shows an improvement of the error correcting properties with the graphs introduced here. It's also important to mention that they can easily be combined and work with existing decoding techniques.

REFERENCES

Biggs, N. 1993. *Algebraic graph theory*, Cambridge Mathematical Library, Cambridge University Press.

Bollobs, Bla. 1978. *Extremal graph theory*, London Mathematical Society Monographs, vol. 11, Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London-New York. MR506522

Brouwer, A. E., A. M. Cohen, and A. Neumaier. 1989. *Distance-regular graphs*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 18, Springer-Verlag, Berlin. MR1002568

Gallager, R. G. 1962. *Low-density parity-check codes*, IRE Trans. **IT-8**, 21–28. MR0136009

Guinand, P. and J. Lodge. 1997a. *Graph theoretic construction of generalized product codes*, Proc. IEEE International Symposium on Information Theory ISIT'97.

Guinand, P. and J Lodge. 1997b. *Tanner type codes arising from large girth graphs*, Proc.Canadian Workshop on Information Theory CWIT'97.

Huffman, W. Cary and Vera Pless. 2003. *Fundamentals of error-correcting codes*, Cambridge University Press, Cambridge. MR1996953

Lazebnik, Felix and Vasiliy A. Ustimenko. 1993. *New examples of graphs without small cycles and of large size*, European J. Combin. **14**, no. 5, 445–460. Algebraic combinatorics (Vladimir, 1991). MR1241911

Polak, Monika. 2016. *Some implementations of a(n,q) and d(n,q) graphs*.

Polak, Monika and Vasyl Ustimenko. 2011. *On LDPC codes corresponding to affine parts of generalized polygons*, Ann. Univ. Mariae Curie-Skłodowska Sect. AI-Inform. **11**, no. 2, 143–152. MR3164287

Seress, kos. 2000. *Large families of cospectral graphs*, Des. Codes Cryptogr. **21**, no. 1-3, 205–208. Special issue dedicated to Dr. Jaap Seidel on the occasion of his 80th birthday (Oisterwijk, 1999). MR1801201

Shannon, C. E. 1948. *A mathematical theory of communication*, Bell System Tech. J. **27**, 379–423, 623–656. MR0026286

Ustimenko, V. A. 2009. *Algebraic groups and small world graphs of high girth*, Albanian J. Math. **3**, no. 1, 25–33. MR2487018

———. 2011. *Algebraic graphs and security of digital communications*, Maria Curie-Skodowska University. Institute of Computer Science.

Ustimenko, V. A. and A. J. Woldar. 2003. *Extremal properties of regular and affine generalized m-gons as tactical configurations*, European J. Combin. **24**, no. 1, 99–111. MR1957968

Ustimenko, Vasiliy A. 2005. *Maximality of affine group, and hidden graph cryptosystems*, Algebra Discrete Math. **1**, 133–150. MR2148826