# Low Density Concatenated Codes with Reed-Solomon Component Codes*

Valentine B. Afanassiev
Institute for Information
Transmission Problems
Russian Academy of Sciences
Bol'shoi Karetnyi per. 19, GSP-4
Moscow, 127994, Russia
Email: afanv@iitp.ru

Alexander A. Davydov
Institute for Information
Transmission Problems
Russian Academy of Sciences
Bol'shoi Karetnyi per. 19, GSP-4
Moscow, 127994, Russia
Email: adav@iitp.ru

Victor V. Zyablov
Institute for Information
Transmission Problems
Russian Academy of Sciences
Bol'shoi Karetnyi per. 19, GSP-4
Moscow, 127994, Russia
Email: zyablov@iitp.ru

**Abstract**

We consider a special case of codes based on bipartite expander graphs. The code symbols are associated with the branches and the symbols connected to a given graph node have to be codewords in a Reed-Solomon component code. We give parameters of the code and algorithms of the code constructing and encoding.

## I. Introduction

We consider a special case of the codes based on bipartite expander graph described in [1] and [2]. The code symbols are associated with the branches, and the symbols connected to a given node have to be codewords of a Reed-Solomon (RS) code. In the known constructions of graph codes with RS component codes based on finite geometries [3], [4], [5], the same finite field is used for code defining and for constructing the geometry. We generalize the approach to the case when a finite field of RS code is not related to the finite field used for constructing the geometry. We also consider the case when a bipartite graph is coming from a stochastic search procedure.

Section 2 contains a stochastic procedure for a given graph enlargement with the given properties. Section 3 describes the modifications of expander graphs from geometries. In particular the algorithms of modification and the set of available parameters are derived. Section 4 gives a method for fast encoding based on a proper permutation on adjacency matrix of the graph.

## II. Expander graphs from a stochastic procedure

The bipartite graph with $m$ nodes in each set can be described by the adjacency matrix $A = \begin{pmatrix} 0 & M \\ M^T & 0 \end{pmatrix}$ where $M$ is $m \times m$ matrix with the property: each row and each column contains $n$ ones, $n \leq m$. The graph can be used to define a code by associating a code symbol with each edge. So all the ones in the adjacency matrix can be numbered as the code symbols and total number of ones in the $M$ matrix is equal to the code length $N = mn$. Next, each row of the adjacency matrix contains a codeword of $[n, k]$ RS code and each code symbol enters into codewords of two different RS codes associated with two nodes of the graph. It is important that we associate each node of the graph with the unique $[n, k]$ Generalized RS (GRS) code.

The rate the GRS code associated with the nodes is $\rho = k/n$, and the total rate is bounded by $R \geq 2\rho - 1$. The minimum distance is always lower bounded [3] by $D \geq d\left(d\left(d-1\right)+1\right)$ where $d$ is the component code distance. From expansion properties of the graph [1] it follows that the minimum distance is significantly larger when $n$ and $d$ are big itself. Most often values are $m = q^2+q+1$, $n = q+1$, and $m = q^2$, $n = q$, for the known constructions over $GF(q)$.

The first problem is construction of the $M$ matrix with given $m, n$. Let a matrix $U$ is given of a size $u < m$ with fixed $n$.

**Enlargement 1** procedure:

1. enlarge $U$ by bordering zeros filled row and column to the size $u + 1$;

* This work was supported by the Program of fundamental scientific researches of DITCS RAS (OITVS RAN), Project no. 3.1

2. choose at random an unused before *row* of $U$ and choose at random a one in the row (intersection of the row with unused before column);
3. clone the chosen one to bordering row and column and change the one to zero;
4. choose at random an unused before *column* of $U$ and choose at random a one in the column (intersection of the column with unused before row);
5. clone the chosen one to bordering row and column and change the one to zero;
6. perform steps 2-5 $n-1$ times and finish the process putting the last one to the corner of bordering row and column.

It is evident that starting from $n \times n$ ones filled matrix we get at random a $m \times m$ matrix with the same $n$ having used the procedure $m-n$ times. It is simple to calculate that the total number of the matrixes is $\prod_{i=1}^{n-1} \frac{(m-n+i)!}{(n-1)!i!}$. The resulting matrix contains a 4-cycle with probability that depends of the ratio $m$ to $n$. It is not very clear how is related existence of 4-cycle with the concatenated code distance. Nevertheless we give an enlargement procedure avoiding 4-cycle.

Let a matrix $U^*$ of a size $u^* < m$ with fixed $n$ having no one 4-cycle is given.

**Enlargement 2** procedure:
1. enlarge $U^*$ by bordering zeros filled row and column to the size $u^*+1$;
2. put the first one to the corner of bordering row and column;
3. choose at random an unmarked *row* of $U^*$ and choose at random a one (unmarked intersection with a column) in the row;
4. clone the chosen one to bordering row and column and change the one to zero;
5. mark the points in the bordering row and column that can complete a new 4-cycles if they will used on the next steps;
6. choose at random an unmarked *column* of $U^*$ and choose at random a one (unmarked intersection with a row) in the column;
7. clone the chosen one to bordering row and column and change the one to zero;
8. mark the points in the bordering row and column that can complete a new 4-cycles if they will used on the next steps;
9. perform steps 3-8 $n-1$ times and finish the process or finish the process with failure if there is no unmarked points.

**"Enlargement 2"** procedure not necessary can be applied to any initial matrix $U^*$. It is not obligatory that the procedure can be used as many times as we want. In the next section we discuss a geometric interpretation of the enlargement procedure with necessary conditions for its application. Finally we propose a hypothesis that **"Enlargement 2"** procedure can be certainly used when $u^* > n^3$.

## III. Expander graphs from geometries

The matrix $M$ as a constructive part of the adjacency matrix $A$ of the bipartite graph with $m$ nodes in each set can be viewed as an incidence matrix for Projective or Euclidian plane with $m$ points and $m$ lines free of 4-cycles [4], [5]. In a projective case we have the parameters: $m = q^2+q+1$, $n = q+1$, and for Euclidian plane we get the parameters: $m = q^2$, $n = q$ (we skip $q$ lines) [3]. Even when we are limited to use RS component code we need a freedom in choice of the code length. So, the problem, coming with geometries, is: *define available set of parameters of the adjacency matrix.*

It is known that the incidence matrix $M_{PG}\left(m,n\right), m = q^2+q+1,\ n = q+1,$ of the projective plane $PG(2,q)$ over $GF(q)$ can always be given in the following standard (block matrix) form

$$M_{PG} = \begin{bmatrix} G_{0,0} & G_{0,1} & G_{0,\alpha} & G_{0,\alpha^2} & \dots & G_{0,\alpha^{q-2}} & B_0 & V_0 \\ G_{1,0} & G_{1,1} & G_{1,\alpha} & G_{1,\alpha^2} & \dots & G_{1,\alpha^{q-2}} & B_1 & V_0 \\ G_{\alpha,0} & G_{\alpha,1} & G_{\alpha,\alpha} & G_{\alpha,\alpha^2} & \dots & G_{\alpha,\alpha^{q-2}} & B_\alpha & V_0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ G_{\alpha^{q-2},0} & G_{\alpha^{q-2},1} & G_{\alpha^{q-2},\alpha} & G_{\alpha^{q-2},\alpha^2} & \dots & G_{\alpha^{q-2},\alpha^{q-2}} & B_{\alpha^{q-2}} & V_0 \\ D_0 & D_1 & D_\alpha & D_{\alpha^2} & \dots & D_{\alpha^{q-2}} & Z & V_1 \\ H_0 & H_0 & H_0 & H_0 & \dots & H_0 & H_1 & 1 \end{bmatrix},$$

where all indices are elements of the finite field $GF(q)$. The $q \times q$ permutation matrix $G_{w,c}$ corresponds to the $q$ points $(1, c, x_2)$ and to the $q$ lines $x_2 = wx_1 + ux_0$ with free $u \in GF(q)$. The $q \times q$ matrix $D_c$ is generated by the line $x_1 = cx_0$ (one row filled by ones) and $q \times q$ matrix $B_w$ is generated by the point $(0, 1, w)$ (one column filled by ones). Matrix $V_0$ (or $H_0$) is the zero column (row) and $V_1$ (or $H_1$) is ones filled column (row). $Z$ is zero matrix and 1 is just one.

The incidence matrix $M_{EG}(m, n)$, $m = q^2$, $n = q$, for Euclidian plane$EG(2, q)$ is the $G$ part of the matrix$M_{PG}$. It is clear, if we need to change the parameters of bipartite graph we have to delete some elements of the standard incidence matrix or to add some other elements.

We consider the following modifications of incidence block matrix $M_{EG}$: **q-cancellation**, **$\Delta$-cancellation**, **$\theta$-extension.**

**q-cancellation:** given the block matrix $M_{EG}\,(m, n)$ and parameters $(v, s)$ delete the $v$-th *block row* and delete the $s-$th *block column*. The result is $M_{EG}\,(m - tq, n - t)$ incidence matrix after $t$ steps.

**$\Delta$-cancellation:** given the matrix $M_{EG}\,(m, n)$ and given a binary $\frac{m}{q} \times \frac{m}{q}$ matrix $S_0\,(\Delta)\,, 1 \leq \Delta < \frac{m}{q}$, of the weight $\Delta$ in each row and each column, remove from $M_{EG}\,(m, n)$ $q \times q$ blocks that corresponds to the ones in $S_0\,(\Delta)$. The result is $M_{EG}\,(m, n - \Delta)$ incidence matrix.

**$\theta$-extension:** given the matrix $M_{EG}\,(m, n)$ use $\theta$ times the procedure “enlargement 2”. The result is $M_{EG}\,(m + \theta, n)$ incidence matrix.

A geometrical interpretation of **$\theta$-extension** gives the following result. Given the matrix $M_{EG}\,(m, n)$ let say that a set of $n - 1$ rows having no intersections and $n - 1$ columns having no intersections too is the *extension set* if the intersection points of the chosen rows and columns form an $(n - 1) \times (n - 1)$ permutation matrix. In the other words, an extension set contains $n - 1$ parallel lines and $n - 1$ pairwise non collinear points. **$\theta$-extension** needs $\theta$ different disjoint extension sets. Due to the block structure of incidence matrixes we see that any set of $n - 1$ rows and $n - 1$ columns going through the block $G_{w,c}$ is the extension set. As $n \leq q$ then one block gives $\left\lfloor \frac{q}{n-1} \right\rfloor$ independent extension sets and in total $f_1 = (q - t) \left\lfloor \frac{q}{n-1} \right\rfloor$ possible extension sets. After the first $f_1$ extensions we get the new $\left\lfloor \frac{f_1}{n-1} \right\rfloor$ extension sets and so on. Finally, we get the iterative extension process with the following result:

$$\theta \leq F\,(n, q, 2, t) = f_1 + \sum_{i=1}^{f_1 < n-1} \left\lfloor \frac{f_1}{n-1} \right\rfloor, \quad f_i = f_{i-1} - \left\lfloor \frac{f_{i-1}}{n-1} \right\rfloor (n-2)\,, n = q - t - \Delta, \Delta \geq 0, t \geq 0.$$

If $F\,(n, q, 2, t) \geq q - 1$ for all $t \geq b$ then we can get incidence matrixes for continuous sequence of values of $m$ at least in the interval $n^2 + bn \leq m \leq q^2 + q + 1, q = n + b$.

Implementation of proposed modifications gives full freedom in the choice of component code parameters and the size of expander graph. Thus for the given $m, n$ we can find a proper finite field with $q \geq n$, $q^2 \geq m$, and transform the initial incidence matrix to the expected form by using cancellation-extension procedures. It is evident that the proposed manipulations with matrix can be generalized to Euclidian and projective space. For any $v \geq 2$ we have

$$F(n, q, v, t) = f_1 + \sum_{i=1}^{f_i < n-1} \left\lfloor \frac{f_i}{n-1} \right\rfloor, \quad f_1 = (q - t) \left\lfloor \frac{q^{v-1}}{n-1} \right\rfloor, \quad f_i = f_{i-1} - \left\lfloor \frac{f_{i-1}}{n-1} \right\rfloor (n-2).$$

If $F(n, q, v, t) \geq q^{v-1} - 1$ for all $t \geq b$ then we can get incidence matrixes for continuous sequence of values of $m$ at least in the interval $(n + b)^{v-1} n \leq m \leq q^{v-1} b + F(q, n, v, 0)$, $q = n + b$.

## IV. Encoding of low density concatenated codes

Encoding is the next of most important problems of implementation of general low density codes and graph codes. Since a graph code is defined by a large parity check matrix, it is not clear how to perform encoding in a simple way. One particular way for graph code with RS component code is considered in [3] based on evaluation of a polynomial from a subset of $F_q\,[X, Y, A, B]$ where quadruple $(x, y, a, b)$ represents an edge in the bipartite graph. We consider another way based on representation of a codeword as square array.

Given the square incidence matrix $M(m, n)$ we set the correspondence between $m$ rows and $m$ columns and $2m$ different parity check matrixes of $[n, k, r+1]$ GRS codes. Different GRS codes in the code construction give some guarantee that the concatenated code dimension is equal to the low bound $K = m\,(n - 2r)$. Encoding procedure consists of two steps: the first – encoding GRS codes from the *encoding sequence* of blocks, and the second – final parity check calculations.

**Generation of the *encoding sequence*:** take at random any row (just for example), fill any $k$ of $n$ with information symbols and calculate $r$ parity check symbols from the system parity check equations of GRS code that corresponds to the chosen row; choose any column that has an intersection with the first row; fill the column with $k-1$ new information symbols (one is defined by the first row) and calculate last $r$ parity check symbols from the system parity check equations of GRS code for the column; continue taking a row (column) that has the maximal number of intersections with previously chosen columns (rows) under the necessary condition – *there has to be at least $r$ free symbols in the chosen block for parity checks*. **Stop condition**: the number of uncalculated parity symbols is equal to the number of rows plus columns in the rest multiplied by $r$.

*Fact*: beginning from the $l$-th step we can find a column (row) having two (or more) intersections with already chosen rows (columns) if the incidence matrix contains $l$-cycles.

*Fact:* only parity symbols can be uncalculated after the first step of encoding procedure.

The final step of calculation of parity symbols in the rest has to be performed as solution of a system of linear equation. That is way, it is very important to generate as long encoding sequence as possible.

We can fill up the generation of encoding sequence with moving of chosen row to the up and columns to the left border of the incidence matrix. Let $L = 2\,(m - x)$ be length of the encoding sequence (number of GRS blocks). Then all uncalculated parity symbols will be collected in the low-right $x \times x$ corner (the rest) of the matrix.

From the model of uniform density of ones in a random incidence matrix we get the following estimate of $L$. Let the density of ones in $M(m, n)$ is $\delta = n/m$. Then we have equation: $2xr = \delta x^2$ and estimates: $x = 2r/\delta = 2rm/n$, $L_{ran} \approx 2m\,(1 - 2r/n)$. More over, from this estimate we see that a row (or column) in that corner contains $2r$ parity symbols in average. Two other empiric estimates follow from the distribution $Q_w$ of blocks with $w$ intersections in the given encoding sequence, $L = \sum_{w=0}^{k} Q_w$. It was found that empirical distribution of $Q_w$ can be approximated as uniform in the range $1 < w < k$. Let a *packet* be a collection of $(k-2)$ blocks with $w = 2, 3, ..., k-1$ from the encoding sequence. Then define the number of information symbols in a packet $V_k = \sum_{w=2}^{k-1} w = k\,(k-1)\,/2 - 1$ and the number of code symbols in a packet $V_n = \sum_{w=2}^{k-1} (n - w) = (k-2)\,n - V_k$. Now we have empiric (not the bounds) lower and upper estimates $\frac{K}{V_k} <\approx \frac{L}{k-2} <\approx \frac{N}{V_n}$ or, after a simple transform $L_{low} \approx 2m\frac{k-r}{k}$; $L_{up} \approx 2m\frac{n}{n+r}$.

A small piece of simulation results with proposed estimates is given in the table where $L_{\min}$, $L_{avr}$, and $L_{\max}$ are, respectively, the minimal, average, and maximal length of the encoding sequence obtained under the simulation.

| $r : m : n$ | Simulations $L_{\min} : L_{avr} : L_{\max}$ | $L_{ran}$ | $L_{low}$ | $L_{up}$ |
|---|---|---|---|---|
| **2**:256:16 | 477 : 490 : 503 | 384 | 438 | 455 |
| **3**:256:16 | 428 : 442 : 468 | 320 | 393 | 431 |
| **4**:256:16 | 354 : 370 : 399 | 256 | 341 | 409 |

**Complexity estimate:** the estimates of the encoding sequence length gives an estimate (in average) of encoding complexity. The complexity of the first step is $L_{ran}nr \approx 2m\,(1 - 2r/n)\,nr$. For the second step we have $\left(4r^2m/n\right)^2$. It is simple to see that for $r \approx \sqrt{n}$ we have the total complexity of order $O\left(m^2\right)$ or of order $O\left(z^2n^4\right)$ if $m$ grows as $zn^2$ for a fixed $n$.

*Fact:* fill all GRS blocks in the encoding sequence by zeros information symbols except the last one, put in the last block a one (only one nonzero) information symbol and calculate all parity symbols in the rest, then we get the upper estimate of the concatenated (graph) code distance. In average (for a random incidence matrix of fixed density of ones) we get the following estimate for graph code distance $D \approx 2mr^2/n$ or $D \approx 2m, r \approx \sqrt{n}$, because there are $x = 2mr/n$ GRS codewords of the weight $> r$ in the rest.

## REFERENCES

[1] G. Zemor, “On expander codes,” *IEEE Trans. Inform. Theory (special issue on codes on graphs and iterative algorithms)*, vol. IT-47, no. 2, pp. 835-837, Feb. 2001.

[2] R. M. Tanner, “A recursive approach to low complexity codes,” *IEEE Trans. Inform. Theory*, vol. IT-27, no. 5, pp. 533-547, Sept. 1981.

[3] T. Høholdt and J. Justesen, “Graph codes with Reed-Solomon Component codes,” ISIT 2006, Seattle, USA, July 9-14, 2006.

[4] 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. Inform. Theory*, vol. IT-47, no. 7, pp. 2711-2736, July 2001.

[5] J. Xu, L. Chen, I. Djurdjevic, and K. Abdel-Ghaffar, “Construction of regular and irregular LDPC codes: geometry decomposition and masking,” *IEEE Trans. Inform. Theory*, vol. IT-53, no. 1, pp. 121-134, Jan. 2007.