University of Tartu
Faculty of Mathematics and Computer Science
Institute of Computer Science
Cyber Security

YAUHEN YAKIMENKA

Optimization of Parity-Check Matrices of LDPC Codes

Master's Thesis (30 ECTS)

Supervised by Dr Vitaly Skachek

Tartu, Estonia 2014

Optimisation of parity-check matrices of LDPC codes

Abstract

Low-density parity-check (LDPC) codes are widely used in communications due to their excellent practical performance. Error probability of LDPC code under iterative decoding on the binary erasure channel is determined by a class of combinatorial objects, called stopping sets. Stopping sets of small size are the reason for the decoder failures. Stopping redundancy is defined as the minimum number of rows in a parity-check matrix of the code, such that there are no small stopping sets in it.

Han, Siegel and Vardy derive upper bounds on the stopping redundancy of general binary linear codes by using probabilistic analysis. For many families of codes, these bounds are the best currently known. In this work, we improve on the results of Han, Siegel and Vardy by modifying their analysis. Our approach is different in that we judiciously select the first and the second rows in the parity-check matrix, and then proceed with the probabilistic analysis. Numerical experiments confirm that the bounds obtained in this thesis are superior to those of Han, Siegel and Vardy for two codes: the extended Golay code and the quadratic residue code of length 48.

Keywords: Binary erasure channel, iterative decoding, low-density parity-check codes, stopping redundancy, stopping sets.

LDPC koodide paarsuskontrolli maatriksite optimiseerimine

Kokkuvõte

Madala tihedusega paarsuskontroll (LDPC) on laialdaselt kasutusel kommunikatsioonis tänu oma suurepärasele praktilisele võimekusele. LDPC koodi vigade tõenäosust iteratiivse dekodeerimise puhul binaarsel kustutuskanalil määrab klass kombinatoorseid objekte, nimega peatamise rühm. Väikese suurusega peatamise rühmad on dekodeerija vigade põhjuseks. Peatamise liiasust määratletakse kui minimaalset ridade arvu paarsuskontrolli koodi maatriksis, mille puhul pole selles väikesi peatuse rühmi.

Han, Siegel ja Vardy kasutavad üld binaarse lineaarkoodi ülemise piiri peatamiste liiasuse tuletamiseks tõenäosuslikku analüüsi. Need piirid on teadaolevalt parimad paljude koodi perekondade puhul. Selles töös me parendame Hani, Siegeli ja Vardy tulemusi modifitseerides selleks nende analüüsi. Meie lähenemine erineb sellepoolest, et me valime mõistlikult esimese ja teise rea paarsuskontrolli maatriksis ja siis läheme edasi tõenäosusliku analüüsiga. Numbrilised väärtused kinnitavad seda, et piirid mis on määratletud selles töös on paremad Hani, Siegeli ja Vardy omadest kahe koodi puhul: laiendatud Golay koodis ja kvadraatses jääk koodis pikkusega 48.

Võtmesõnad: binaarne kustutamise kanal, iteratiivne dekodeerimine, madala tihedusega paarsuskontrolli kood, peatamise liiasus, peatamise rühmad.

Acknowledgements

I would like to use the chance to thank those who helped me during my Master's studies in Estonia.

First of all, I want to thank Cryptography and Security research group in the Institute of Computer Science at the University of Tartu. Special thanks should be addressed to my supervisor, Dr Vitaly Skachek, who introduced me to the world of scientific research, was patient and supportive.

Tallinn University of Technology coordinated my studies and the staff there was of great help. Courses I took there showed me a different perspective of IT and computer science. The University also gave me a chance to spend one exchange semester in Chalmers University of Technology in Gothenburg, Sweden, which definitely widened my horizons.

DoRa-9 scholarship, provided by Archimedes Foundation, was a big support. It also gave me a conclusive argument to decide on studying in Estonia. Another important source of funding for me was IT Academy scholarships sponsored by Skype. I would hardly be able to study without this support.

And I am also thankful to all the descent people from around the world whom I met in Estonia and who became good friends of mine. Atko Kadakas' help to prepare the part of this thesis in Estonian is appreciated. I also thank Jaan Kroon for inspiration to finish the thesis. Deepest thanks are addressed to all my friends around the world and it goes without saying that I am really grateful to my parents.

This thesis is supported by the Norwegian-Estonian Research Cooperation Programme through the grant EMP133 and by the Estonian Research Council through the grant IUT2-1.

Yauhen Yakimenka, Tartu
Wednesday 28th May, 2014

Contents

List of Symbols

$\mathbb{F}_2$	Binary Galois field
$\mathbb{F}_2^n$	Space of vectors of length n over $\mathbb{F}_2$
$\mathsf{w}(\mathbf{c})$	Hamming weight of vector $\mathbf{c}$
$\mathsf{d}(\mathbf{c}_1, \mathbf{c}_2)$	Hamming distance between two vectors
$\mathcal{C}$	Linear code of length n, dimension k and minimum distance d
$[n,k,d]$	Linear code of length n, dimension k and minimum distance d
$\mathcal{C}^\perp$	Linear code dual to $\mathcal{C}$
r	Dimension of code $\mathcal{C}^\perp$, i.e. $n-k$
$d^\perp$	Minimum distance of code $\mathcal{C}^\perp$
$\mathcal{C}_0^\perp$	Dual code without all-zero vector, i.e. $\mathcal{C}^\perp \setminus \{\mathbf{0}\}$
$\mathfrak{I}_i$	The set of all i-sets, i.e. $\{\mathcal{S} \subset \{1,2,\dots,n\} \ : \ \lvert\mathcal{S}\rvert = i\}$
$\mathfrak{I}$	The set of all i-sets for $1 \leqslant i \leqslant d-1$, i.e. $\bigcup_{i=1}^{d-1} \mathfrak{I}_i$
$\mathsf{P}\{\cdot\}$	Probability measure
$\mathsf{E}\{\xi\}$	Expectation of random variable ξ
$\mathsf{I}\{\cdot\}$	Indicator function
$\lfloor a \rfloor$	Floor function, i.e the largest integer not greater than a
$\binom{m}{j}$	Binomial coefficient
$\operatorname{rank} H$	Rank of matrix H
$(\mathbf{h}_1^\mathsf{T},\mathbf{h}_2^\mathsf{T},\dots,\mathbf{h}_t^\mathsf{T})^\mathsf{T}$	Matrix consisting of rows $\mathbf{h}_1,\mathbf{h}_2,\dots,\mathbf{h}_t$

1

Introduction

Low-density parity-check (LDPC) codes were introduced in [1], but were almost forgotten until mid-1990s, when they were rediscovered [2] and became a focus of intensive research and numerous practical implementations.

Gallager in his work [1] introduced iterative decoding algorithms for LDPC codes. For binary symmetric channel he showed that for all code rates below a certain bound (which is lower than the Shannon capacity), the decoding error probability for these algorithms decays exponentially with the square root of the code length.

For binary erasure channel (BEC) the failure events during *edge removal* iterative decoding are completely characterised by one class of combinatorial objects, called *stopping sets*. It is known that by adding redundant rows to the parity-check matrix of the code it is possible to reduce number of stopping sets of small size. In this thesis we investigate existing upper bounds on the number of such redundant rows and improve them.

The contents of the thesis is as follows. Chapter 1 introduces all the required definitions and the statement of the problem. In Chapter 2 we survey existing upper bounds on the stopping redundancy and then present new improvements. The improved bounds are the main result of this thesis. Next, in Chapter 3 the techniques developed in Chapter 2 are applied to *stopping redundancy hierarchy*. Chapter 4 illustrates the new improvements with numerical examples. Finally, in Chapter 5 we summarise the results of the thesis and discuss open problems.

1.1 Binary erasure channel

The following communications model was first studied by Claude Shannon in 1948. The communication system consists of several components. Figure 1.1 shows the connection between the different components.

The information is generated by source and is transferred over the channel. The (memoryless) channel is defined by the triple $(\Sigma_{in}, \Sigma_{out}, \mathsf{Prob})$. Here Σ_{in} and Σ_{out} are input and output alphabets, respectively. Probability function $\mathsf{Prob} : \Sigma_{in} \times \Sigma_{out} \rightarrow [0{,}1]$ is defined on pairs of symbols as

$$\mathsf{Prob}(a{,}b) = \mathsf{P}\{b \text{ received} \mid a \text{ transmitted}\},$$

where $\mathsf{P}\{\cdot \mid \cdot\}$ denotes conditional probability.

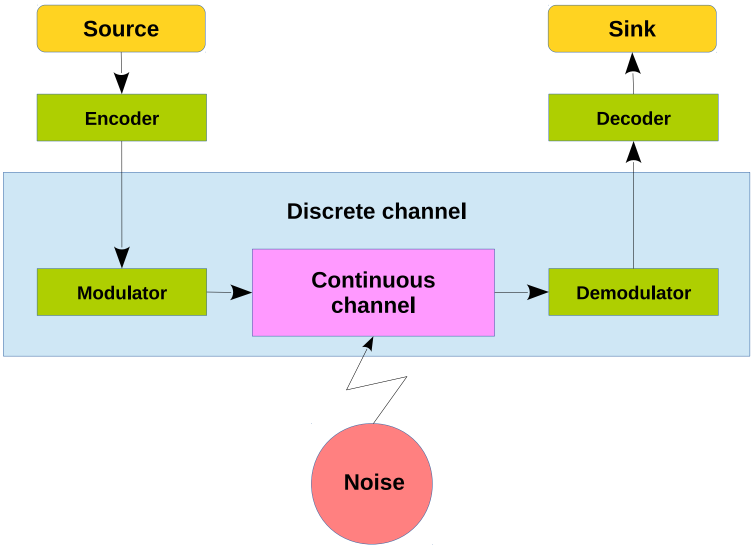


Figure 1.1: Noisy channel transmission

In this thesis we consider only binary erasure channel (BEC). BEC is a binary-input channel which imposes only one type of error — it erases each bit with probability $p \in [0,1]$ (we mark erased position with ε). More precisely, $\Sigma_{in} = \{0,1\}$, $\Sigma_{out} = \{0,1,\varepsilon\}$, and the erasure probability function is illustrated in Figure 1.2.

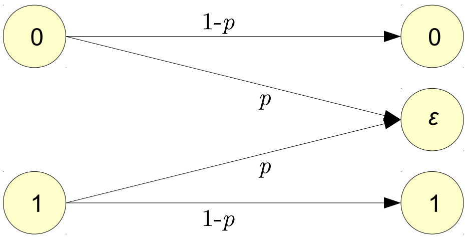


Figure 1.2: Binary erasure channel

1.2 Low-density parity-check codes

Let $\mathbb{F}_2$ denote the binary finite field and $\mathsf{w}(\mathbf{x})$ denote the Hamming weight of $\mathbf{x} \in \mathbb{F}_2^n$, i.e. the number of non-zero coordinates of $\mathbf{x}$. By $\mathsf{d}(\mathbf{x},\mathbf{y})$ we denote the Hamming distance between $\mathbf{x},\mathbf{y} \in \mathbb{F}_2^n$, which is defined as $\mathsf{d}(\mathbf{x},\mathbf{y}) = \mathsf{w}(\mathbf{x}+\mathbf{y})$.

A subspace $\mathcal{C} \subset \mathbb{F}_2^n$ is called a *linear* $[n,k,d]$ *code*[1], where n is the length of the code, $k = \log_2 |\mathcal{C}|$ is the dimension of the code and

$$d = \min\{\mathsf{d}(\mathbf{x},\mathbf{y}) \mid \mathbf{x},\mathbf{y} \in \mathcal{C}, \mathbf{x} \neq \mathbf{y}\}$$

is the minimum distance of the code. The *dual code* $\mathcal{C}^\perp$ is $[n,n-k,d^\perp]$ code over $\mathbb{F}_2$ that is the orthogonal complement of $\mathcal{C}$ in $\mathbb{F}_2^n$.

[1] Since in this thesis we talk only about linear codes, we usually omit the word "linear".

A linear code can be represented as the null space of a parity-check matrix H:

$$\mathcal{C} = \{\mathbf{x} \in \mathbb{F}_2^n \mid H\mathbf{x}^\intercal = \mathbf{0}^\intercal\},$$

where H is the binary matrix with n columns and rank $H = n - k$. It is easy to see that rows of H are codewords from $\mathcal{C}^\perp$. Moreover, any matrix, whose rows are codewords from $\mathcal{C}^\perp$ and rank is $n - k$, is a parity-check matrix for code $\mathcal{C}$.

For example, consider the following parity-check matrix:

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

It defines a $[7, 4, 2]$ code over $\mathbb{F}_2$ with length 7, dimension 4 and minimum distance 2.

The *Tanner graph* of a parity-check matrix H is a bipartite graph whose biadjacency matrix is H. Figure 1.3 shows the Tanner graph yielded by the matrix from the example above. There $v_1, v_2, \dots, v_7$ ("variable nodes") correspond to columns of H and c_1, c_2, c_3 ("check nodes") correspond to rows of H. v_i and c_j are connected if the corresponding element of the parity-check matrix is 1.

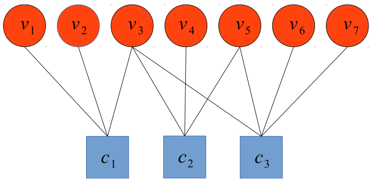


Figure 1.3: Tanner graph example

A linear code is called a *low-density parity-check* code if it has a sparse parity-check matrix. Typically, the number of non-zero elements in each row or column is upper bounded by a small constant. LDPC codes are good in practice and they asymptotically achieve the capacity of the binary erasure channel [3].

1.3 Iterative decoding on BEC

Iterative decoding is the most widely used decoding technique for LDPC codes. Here we describe how it works for the BEC.

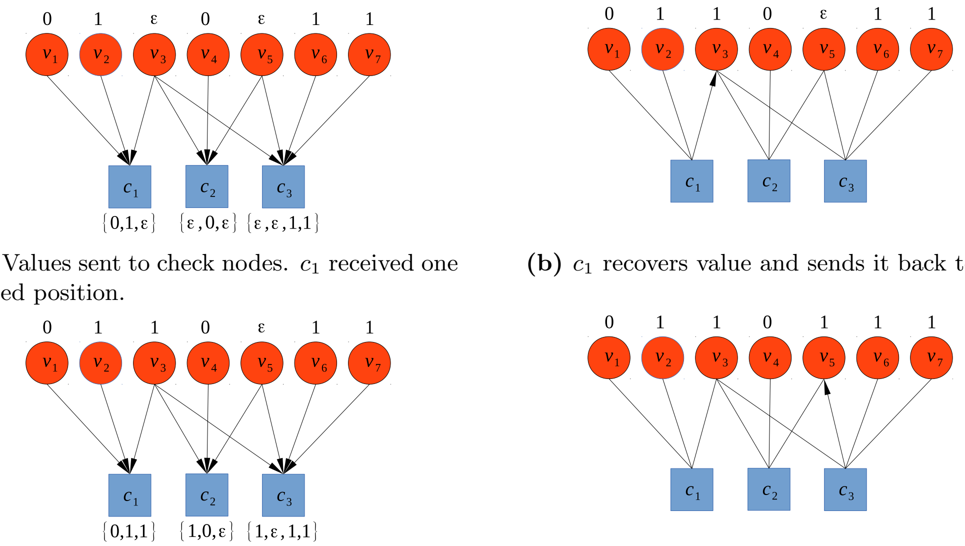
Consider the Tanner graph at the Figure 1.3. Assume that the encoder received the codeword (0,1,ε,0,ε,1,1) where ε denotes the erased positions. We put these values into corresponding variable nodes.

One iteration consists of the following:

1. Current values from variable nodes are sent to check nodes.

2. If there is no check node which has received exactly one ε (erased position), decoding stops with the error message "Decoding impossible".

3. Otherwise let c_j be the check-node that received exactly one ε (if there are several such nodes, choose any of them) and let v_i be the variable node, that has sent ε to c_j. Then from parity-check equation we deduce the value in v_i — it is equal to the sum of the other values received by c_j.



(a) Values sent to check nodes. c_1 received one erased position.

(b) c_1 recovers value and sends it back to v_3.

(c) Updated values are sent back to check nodes. Both c_2 and c_3 receive one erased position each.

(d) c_3 recovers value and sends it back to v_5. All the values are recovered.

Figure 1.4: Iterative decoding on BEC

4. The deduced value is sent back to v_i.

The algorithm continues until either all the erasures are restored or decoding error happens. The full decoding process is illustrated in Figure 1.4.

1.4 Stopping redundancy

It was shown in [4] that the performance of a code under iterative decoding on the BEC is completely determined by the stopping sets of its parity-check matrix. *Stopping set* in Tanner graph is a subset $\mathcal{S}$ of variable nodes such that all the check nodes that are neighbours of a node in $\mathcal{S}$ are connected to *at least two nodes* in $\mathcal{S}$.

For the Tanner graph on the Figure 1.3, the set $\{v_4,v_5,v_6,v_7\}$ is a stopping set (see Figure 1.5).

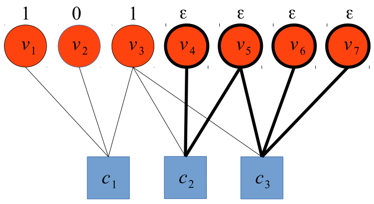


Figure 1.5: Stopping set in Tanner graph

The stopping sets introduce an ambiguity into the decoding process: if all the corresponding symbols are erased or corrupted, then all the neighbouring check nodes are connected to these erasures at least twice, and this form an under-constrained system of linear equations. In this case, an iterative decoder has no way of determining the erased values. Moreover, it could be shown that set of erased coordinates in undecodable under iterative decoder if and only if it includes some stopping set.

In this thesis we work mostly with the parity-check matrices, not with the Tanner graphs. Therefore we re-phrase the definition of the stopping sets in terms of the parity-check matrix. *Stopping set* is a set of columns of H with the property that the matrix, comprised of these columns, does not contain a row of weight one.

Example 1. Consider the matrix H:

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

The matrix consisting of the last 4 columns does not have a row of weight 1, therefore it is a stopping set.

It could be easily seen that the union of stopping sets is also a stopping set. We are interested in stopping sets of the minimum size since those correspond to more likely error events. We call the minimum size of a stopping set of H the *stopping distance* of H and denote it by $s(H)$. It is known (cf. [5, Corollary 1]) that stopping distance of any parity-check matrix is bounded from above by the minimum distance of the corresponding code.

Note that the stopping sets and the stopping distance are defined for a particular parity-check matrix, not for the code.

For the code $\mathcal{C}$ we want to find a parity-check matrix that maximises the stopping distance. On the other hand, among such matrices we want to choose the one with the minimum number of rows. This leads to the following characteristic of the code.

Definition 1. The *stopping redundancy* of an $[n,k,d]$ code $\mathcal{C}$ over $\mathbb{F}_2$ is the smallest integer $\rho(\mathcal{C})$ such that there exists a parity-check matrix H for $\mathcal{C}$ with $\rho(\mathcal{C})$ rows and $s(H) = d$.

The following theorem shows that the stopping redundancy is, indeed, well-defined.

Theorem 1 ([5, Theorem 2])**.** *Let H^* denote the parity-check matrix for $\mathcal{C}$ consisting of all the non-zero codewords of the dual code $\mathcal{C}^\perp$. Then $s(H^*) = d$.*

1.5 Example

Let us consider the [10,3,4] code over $\mathbb{F}_2$ with the following parity-check matrix:

$$H = \begin{array}{c} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \end{array} \begin{pmatrix} \mathbf{1} & 1 & \mathbf{0} & 1 & {\color{blue}1} & 0 & 1 & {\color{blue}1} & 1 & {\color{blue}\mathbf{1}} \\ \mathbf{1} & 1 & \mathbf{0} & 1 & {\color{blue}1} & 0 & 0 & {\color{blue}0} & 1 & {\color{blue}\mathbf{1}} \\ \mathbf{1} & 0 & \mathbf{1} & 0 & {\color{blue}0} & 1 & 0 & {\color{blue}1} & 0 & {\color{blue}\mathbf{1}} \\ \mathbf{0} & 0 & \mathbf{1} & 1 & {\color{blue}1} & 0 & 1 & {\color{blue}1} & 0 & {\color{blue}\mathbf{1}} \\ \mathbf{0} & 0 & \mathbf{1} & 0 & {\color{blue}1} & 0 & 0 & {\color{blue}1} & 1 & {\color{blue}\mathbf{1}} \\ \mathbf{0} & 0 & \mathbf{0} & 1 & {\color{blue}1} & 0 & 0 & {\color{blue}1} & 0 & {\color{blue}\mathbf{0}} \\ \mathbf{1} & 0 & \mathbf{0} & 1 & {\color{blue}0} & 0 & 1 & {\color{blue}1} & 0 & {\color{blue}\mathbf{1}} \end{pmatrix} \tag{1.1}$$

There are the following stopping sets of size less than 4: {1,3,10}, {1,5,8}, {4,8,10}, {5,8,10}. For example, we marked the first of these stopping sets with bold and the last of these stopping sets with blue colour. This corresponds to the Tanner graph shown in Figure 1.6.

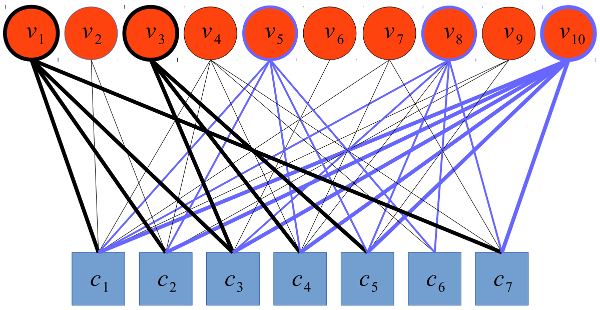


Figure 1.6: Tanner graph for matrix (1.1)

We could add redundant rows to this matrix in order to eliminate these stopping sets (see (1.2)).

$$H' = \begin{array}{l} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \\ c_7 \\ c_1 + c_2 + c_3 \\ c_2 + c_3 \end{array} \begin{pmatrix} \mathbf{1} & 1 & \mathbf{0} & 1 & 1 & 0 & 1 & 1 & 1 & \mathbf{1} \\ \mathbf{1} & 1 & \mathbf{0} & 1 & 1 & 0 & 0 & 0 & 1 & \mathbf{1} \\ \mathbf{1} & 0 & \mathbf{1} & 0 & 0 & 1 & 0 & 1 & 0 & \mathbf{1} \\ \mathbf{0} & 0 & \mathbf{1} & 1 & 1 & 0 & 1 & 1 & 0 & \mathbf{1} \\ \mathbf{0} & 0 & \mathbf{1} & 0 & 1 & 0 & 0 & 1 & 1 & \mathbf{1} \\ \mathbf{0} & 0 & \mathbf{0} & 1 & 1 & 0 & 0 & 1 & 0 & \mathbf{0} \\ \mathbf{1} & 0 & \mathbf{0} & 1 & 0 & 0 & 1 & 1 & 0 & \mathbf{1} \\ \mathbf{1} & 0 & \mathbf{1} & 0 & 0 & 1 & 1 & 0 & 0 & \mathbf{1} \\ \mathbf{0} & 1 & \mathbf{1} & 1 & 1 & 1 & 0 & 1 & 1 & \mathbf{0} \end{pmatrix} \tag{1.2}$$

These two redundant rows are enough to eliminate all the stopping sets of size 3. Therefore we conclude that stopping redundancy of the code is not more than 9. The same in terms of Tanner graph could be seen in Figure 1.7.

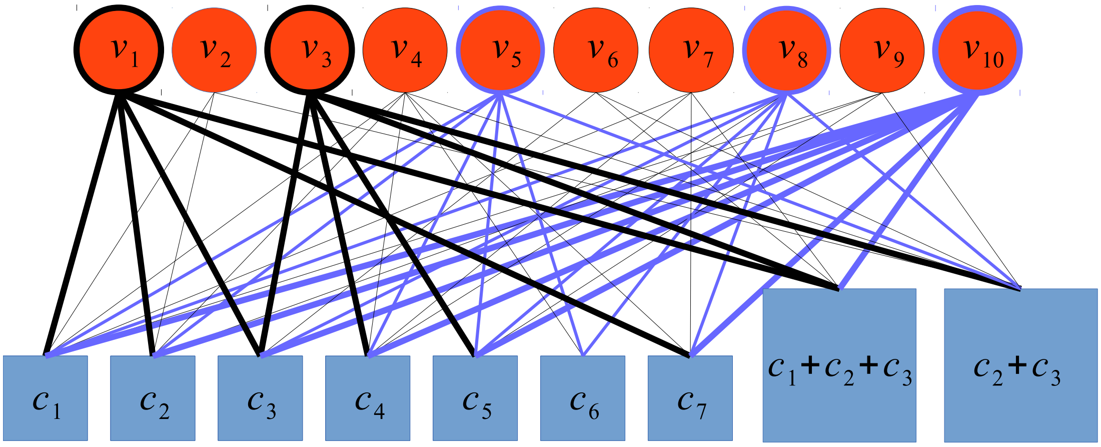


Figure 1.7: The Tanner graph for matrix with redundant rows

2

Upper Bounds on the Stopping Redundancy

The stopping redundancy was introduced by Schwartz and Vardy in [5] and subsequently studied in a number of papers. Existing results on stopping redundancy are of two types: bounds on the stopping redundancy of specific families of codes (e.g. cyclic codes [6], MDS codes [5], [7], Reed-Muller codes [5], [8] and Hamming codes [8], [9]) as well as bounds on the stopping redundancy of general binary linear codes. This thesis studies the latter type of settings.

This chapter contains the main results of the thesis. We start with previous results obtained in [5] and [10], which are the best known upper bounds for the general case. Then the new improvements are presented.

Further in this chapter, if not stated opposite, we will consider an $[n,k,d]$ code $\mathcal{C}$ over $\mathbb{F}_2$, as well as an $[n,r,d^{\perp}]$ code $\mathcal{C}^{\perp}$ which is dual to $\mathcal{C}$. As it was mentioned before, $r = n - k$. We denote the dual code $\mathcal{C}^{\perp}$ without all-zero vector (i.e. $\mathcal{C}^{\perp} \setminus \{\mathbf{0}\}$) by $\mathcal{C}_0^{\perp}$.

We call any subset of $\{1,2,\ldots,n\}$ of cardinality i an *i-set*. The set of all i-sets is denoted by $\mathfrak{I}_i$:

$$\mathfrak{I}_i = \{\mathcal{S} \subset \{1,2,\ldots,n\} : |\mathcal{S}| = i\}$$

We are interested in i-sets for $1 \leqslant i \leqslant d-1$, therefore we also use the following notation:

$$\mathfrak{I} = \bigcup_{i=1}^{d-1} \mathfrak{I}_i$$

We say that some row $\mathbf{h} \in \mathbb{F}_2^n$ *covers* i-set $\mathcal{S}$ if projection of $\mathbf{h}$ on coordinates indexed by $\mathcal{S}$ has Hamming weight 1. We also say that the matrix $(\mathbf{h}_1^{\intercal},\mathbf{h}_2^{\intercal},\ldots,\mathbf{h}_t^{\intercal})^{\intercal}$ *covers* $\mathcal{S}$ if any of its rows covers $\mathcal{S}$.

2.1 Known bounds

As it was shown in [5, Theorem 3], if $d \leqslant 3$ then any parity-check matrix achieves maximum stopping distance and therefore $\rho(\mathcal{C}) = r$. Therefore in all subsequent derivations we assume that $d \geqslant 4$.

The first upper bound was established in [5, Theorem 4]. It is given by

$$\rho(\mathcal{C}) \leqslant \sum_{i=1}^{d-2} \binom{r}{i}$$

We briefly present the idea of the proof. Having any $r \times n$ parity-check matrix H with linearly independent rows, we construct another parity-check matrix H' whose rows are all the sums of t rows of H, for all $t = 1,2,\dots,d-2$. It was shown in the proof of the theorem that the stopping distance of H' is d.

This constructive approach was further developed in [7], [6], [11] and [12].

An entirely different probabilistic argument was used by Han and Siegel in [7, Theorem 3]. With some small additional idea and more precise calculation of probabilities, this resulted in [10, Theorem 3]:

$$\rho(\mathcal{C}) \leqslant \min_{t\in\mathbb{N}} \left\{ t + \left\lfloor \sum_{i=1}^{d-1} \binom{n}{i} \prod_{j=1}^{t} \left(1 - \frac{i2^{r-i}}{2^r - j}\right) \right\rfloor \right\} + (r - d + 1)$$

This bound is based on the following technique. The expression

$$\mathcal{F}_{n,k,d}(t) = \sum_{i=1}^{d-1} \binom{n}{i} \prod_{j=1}^{t} \left(1 - \frac{i2^{r-i}}{2^r - j}\right) \tag{2.1}$$

is an average number of non-covered i-sets for $i = 1,2,\dots,d-1$ if $t \times n$ matrix is composed from the rows chosen uniformly at random without repetition from the codewords of $\mathcal{C}^\perp$. It was further shown that amongst such matrices there exists at least one with number of non-covered i-sets not more than floor of (2.1). Adding one row per each i-set left uncovered, every i-set gets covered. Finally, $r - d + 1$ rows are added to guarantee that the rank of the matrix is r.

However the best result was obtained in [10, Theorem 7] if r and d satisfy $(r-1)(d-1) \leqslant 2^{d-1}$:

$$\rho(\mathcal{C}) \leqslant \min_{t\geqslant r} \left\{ t + \min\left\{ i \in \mathbb{N} : Q_i\left(\lfloor \mathcal{G}_{n,d,k}(t) \rfloor\right) = 0 \right\} \right\}$$

where

$$\mathcal{G}_{n,d,k}(t) = \sum_{i=1}^{d-1} \binom{n}{i} \prod_{j=1}^{t} \left(1 - \frac{i2^{r-i}}{2^r - j}\right) + \frac{1}{2^{t-r}} \left(1 + \frac{2/3}{2^{t-r+1} - 1}\right)$$

$$Q_i(x) = P_i(P_{i-1}(\dots P_2(P_1(x))\dots))$$

$$P_j(x) = \left\lfloor x \left(1 - \frac{(d-1)2^{r-d+1}}{2^r - (t+j)}\right) \right\rfloor$$

The technique starts with the same probabilistic argument but now $\mathcal{G}_{n,d,k}(t)$ is a sum of the number of uncovered i-sets and the rank deficiency[1]. Further, one application of P_j describes the guaranteed decrease of this characteristic of matrix when a new specially chosen row is added.

This bound is the best known at the moment. In the following section we provide suggestions on how existing bounds could be improved.

[1] I.e. the difference between the rank of a parity-check matrix, r, and the actual.

2.2 Improved bounds

The new bounds we derive in this section are based on the known bound by Han, Siegel and Vardy [10]. The difference between this work and [10] is that first we judiciously select the first and the second rows in the parity-check matrix. Then we continue similarly to the analysis in [10]. This technique allows to cover many stopping sets already in the first step and therefore subsequent use of the technique in [10] requires less rows.

Lemma 1. Let ξ be an integer discrete random variable with expected value $\mathsf{E}\{\xi\} = a$. Then there is a realisation of ξ not larger than $\lfloor a \rfloor$.

Proof. Assume contrary, i.e. that ξ can only possess values $a_1, a_2, \dots$ with probabilities $p_1, p_2, \dots$, and $a_i \geqslant \lfloor a \rfloor + 1$ for all i. Then expectation is

$$\mathsf{E}\{\xi\} = \sum_i a_i p_i \geqslant (\lfloor a \rfloor + 1) \sum_i p_i = \lfloor a \rfloor + 1 > a.$$

This contradiction proves the lemma. □

Lemma 2. Let $\mathbf{h} \in \mathcal{C}_0^\perp$ and $\mathsf{w}(\mathbf{h}) = w \leqslant n - i + 1$. Then $\mathbf{h}$ covers exactly $w\binom{n-w}{i-1}$ i-sets from $\mathfrak{I}_i$, $i = 1,2,\dots,d-1$.

Proof. It's the number of i-sets whose projection on $\mathbf{h}$ have a Hamming weight 1. □

Lemma 3. Assume we have a (non-random) matrix $(\mathbf{h}_1^\mathsf{T}, \mathbf{h}_2^\mathsf{T}, \dots, \mathbf{h}_\tau^\mathsf{T})^\mathsf{T}$ whose rows are τ different codewords from $\mathcal{C}_0^\perp$. For $i = 1,2,\dots,d-1$ let us denote by $\mathfrak{U}_i$, $|\mathfrak{U}_i| \leqslant u_i$, the set of i-sets not covered by $(\mathbf{h}_1^\mathsf{T}, \mathbf{h}_2^\mathsf{T}, \dots, \mathbf{h}_\tau^\mathsf{T})^\mathsf{T}$.

Let us add t more rows $\mathbf{h}_{1+\tau}, \mathbf{h}_{2+\tau}, \dots, \mathbf{h}_{t+\tau}$ drawing them uniformly at random without repetitions from $\mathcal{C}_0^\perp \setminus \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_\tau\}$ and let H denotes the resulting matrix. If we denote by ξ the number of sets from $\mathfrak{I}$ that are not covered by H, then

$$\mathsf{E}\{\xi\} \leqslant \sum_{i=1}^{d-1} u_i \prod_{j=\tau+1}^{\tau+t} \left(1 - \frac{i2^{r-i}}{2^r - j}\right)$$

Proof. ξ is the integer discrete random variable:

$$\begin{aligned}\xi = \sum_{\mathcal{S}\in\mathfrak{I}} \mathsf{I}\{\mathcal{S} \text{ is not covered by } H\} &= \sum_{i=1}^{d-1} \sum_{\mathcal{S}\in\mathfrak{I}_i} \mathsf{I}\{\mathcal{S} \text{ is not covered by } H\} \\ &= \sum_{i=1}^{d-1} \sum_{\mathcal{S}\in\mathfrak{U}_i} \mathsf{I}\{\mathcal{S} \text{ is not covered by } (\mathbf{h}_{1+\tau}^\mathsf{T}, \mathbf{h}_{2+\tau}^\mathsf{T}, \dots, \mathbf{h}_{t+\tau}^\mathsf{T})^\mathsf{T}\}\end{aligned}$$

Then

$$\begin{aligned}\mathsf{E}\{\xi\} &= \sum_{i=1}^{d-1} \sum_{\mathcal{S}\in\mathfrak{U}_i} \mathsf{E}\left\{\mathsf{I}\{\mathcal{S} \text{ is not covered by } (\mathbf{h}_{1+\tau}^\mathsf{T}, \mathbf{h}_{2+\tau}^\mathsf{T}, \dots, \mathbf{h}_{t+\tau}^\mathsf{T})^\mathsf{T}\}\right\} \\ &= \sum_{i=1}^{d-1} \sum_{\mathcal{S}\in\mathfrak{U}_i} \mathsf{P}\left\{\mathcal{S} \text{ is not covered by } (\mathbf{h}_{1+\tau}^\mathsf{T}, \mathbf{h}_{2+\tau}^\mathsf{T}, \dots, \mathbf{h}_{t+\tau}^\mathsf{T})^\mathsf{T}\right\} \quad (2.2)\end{aligned}$$

To find the probability in the sum, we recall (cf. [13, p. 139]) that $2^r \times n$ matrix, consisting of all codewords of $\mathcal{C}^\perp$, is an orthogonal array of strength $d-1$. This means that for any $i = 1,2,\ldots,d-1$ projection of this matrix on any i-set $\mathcal{S}$ contains every vector of length i appearing 2^{r-i} times. From this it follows that there are exactly $i2^{r-i}$ codewords in $\mathcal{C}_0^\perp$ that cover $\mathcal{S}$. Then

$$\begin{aligned}
&\mathsf{P}\left\{\mathcal{S} \text{ is not covered by } (\mathbf{h}_{1+\tau}^\mathsf{T}, \mathbf{h}_{2+\tau}^\mathsf{T}, \ldots, \mathbf{h}_{t+\tau}^\mathsf{T})^\mathsf{T}\right\} \\
&\qquad = \frac{\sharp \text{ of the ways to choose } \mathbf{h}_{1+\tau}, \mathbf{h}_{2+\tau}, \ldots, \mathbf{h}_{t+\tau} \text{ that do not cover } \mathcal{S}}{\sharp \text{ of all the ways to choose } \mathbf{h}_{1+\tau}, \mathbf{h}_{2+\tau}, \ldots, \mathbf{h}_{t+\tau}} \\
&\qquad\qquad = \binom{(2^r-\tau-1) - i2^{r-i}}{t} \Big/ \binom{2^r-\tau-1}{t} = \prod_{j=\tau+1}^{\tau+t} \left(1 - \frac{i2^{r-i}}{2^r - j}\right)
\end{aligned}$$

By substituting this into (2.2), we have the result of the lemma. □

Lemma 4. In the settings of Lemma 3 for all $t \geqslant r$

$$\mathsf{E}\left\{r - \operatorname{rank} H\right\} \leqslant \frac{1}{2^{t-r}}\left(1 + \frac{2/3}{2^{t-r+1}-1}\right)$$

Proof. It was shown in [10, Lemma 6] that for any t

$$\mathsf{E}\left\{r - \operatorname{rank}(\mathbf{h}_1^\mathsf{T}, \mathbf{h}_2^\mathsf{T}, \ldots, \mathbf{h}_t^\mathsf{T})^\mathsf{T}\right\} \leqslant \frac{1}{2^{t-r}}\left(1 + \frac{2/3}{2^{t-r+1}-1}\right)$$

Since $(\mathbf{h}_1^\mathsf{T}, \mathbf{h}_2^\mathsf{T}, \ldots, \mathbf{h}_t^\mathsf{T})^\mathsf{T}$ is a submatrix of H, then

$$\operatorname{rank}(\mathbf{h}_1^\mathsf{T}, \mathbf{h}_2^\mathsf{T}, \ldots, \mathbf{h}_t^\mathsf{T})^\mathsf{T} \leqslant \operatorname{rank} H$$

Which proves the lemma. □

For any matrix H (not necessary random) with rows from $\mathcal{C}_0^\perp$ we could define the following characteristic[2]

$$\delta(H) = \Big|\{\mathcal{S} \in \mathfrak{I} \mid \mathcal{S} \text{ is not covered by } H\}\Big| + (r - \operatorname{rank} H).$$

$\delta(H) = 0$ means that $\operatorname{rank} H = r$ and all the i-sets for $i = 1,2,\ldots,d-1$ are covered, i.e. H is the parity-check matrix and its stopping distance is d. Therefore our goal is to construct a matrix H, such that $\delta(H) = 0$.

Corollary 1. In the settings of Lemma 3 there exist $\mathbf{h}_{1+\tau}, \mathbf{h}_{2+\tau}, \ldots, \mathbf{h}_{t+\tau}$ such that

$$\delta(H) \leqslant \left\lfloor \sum_{i=1}^{d-1} u_i \prod_{j=\tau+1}^{\tau+t}\left(1 - \frac{i2^{r-i}}{2^r-j}\right) + \frac{1}{2^{t-r}}\left(1 + \frac{2/3}{2^{t-r+1}-1}\right)\right\rfloor$$

Proof. From Lemmas 3 and 4

$$\mathsf{E}\{\delta(H)\} = \mathsf{E}\{\xi\} + \mathsf{E}\{r - \operatorname{rank} H\} \leqslant \sum_{i=1}^{d-1} u_i \prod_{j=\tau+1}^{\tau+t}\left(1 - \frac{i2^{r-i}}{2^r-j}\right) + \frac{1}{2^{t-r}}\left(1 + \frac{2/3}{2^{t-r+1}-1}\right)$$

[2] The matrix H is not necessary the parity-check matrix, since its rank can be less than r.

Since $\delta(H)$ is an integer discrete random variable, then from Lemma 1 we have that there is a realisation of it (which means there are $\mathbf{h}_{1+\tau}, \mathbf{h}_{2+\tau}, \ldots, \mathbf{h}_{t+\tau}$) such that

$$\delta(H) \leqslant \left\lfloor \sum_{i=1}^{d-1} u_i \prod_{j=\tau+1}^{\tau+t} \left(1 - \frac{i2^{r-i}}{2^r - j}\right) + \frac{1}{2^{t-r}} \left(1 + \frac{2/3}{2^{t-r+1} - 1}\right) \right\rfloor$$

□

Next, we present the new upper bounds on the stopping redundancy.

Lemma 5. There always exists the codeword $\mathbf{h}$ in $\mathcal{C}_0^\perp$ such that $\mathsf{w}(\mathbf{h}) \leqslant n - d + 2$.

Proof. We write down the Singleton bounds for both $\mathcal{C}$ and $\mathcal{C}^\perp$:

$$k + d \leqslant n + 1,$$
$$r + d^\perp \leqslant n + 1,$$

Then $d^\perp = d^\perp - 1 + 1 \leqslant n - r + 1 = k + 1 \leqslant n - d + 2$ and therefore at least the codeword from $\mathcal{C}^\perp$ of minimum weight $d^\perp$ satisfies the lemma. □

Theorem 2. Let $\mathbf{h}_1$ be any codeword in $\mathcal{C}_0^\perp$ with $\mathsf{w}(\mathbf{h}_1) = w \leqslant n - d + 2$. If we denote

$$\mathcal{D}_{n,k,d}(w,t) = \sum_{i=1}^{d-1} \left(\binom{n}{i} - w\binom{n-w}{i-1} \right) \prod_{j=2}^{t+1} \left(1 - \frac{i2^{r-i}}{2^r - j}\right) + \frac{1}{2^{t-r}} \left(1 + \frac{2/3}{2^{t-r+1} - 1}\right)$$

then the stopping redundancy is bounded from above by

$$\rho(\mathcal{C}) \leqslant 1 + \min_{t \geqslant r} \{t + \lfloor \mathcal{D}_{n,k,d}(w,t) \rfloor\}$$

Proof. From Lemma 5 such a codeword $\mathbf{h}_1$ always exists.

In Corollary 1 set $\tau = 1$ and use $\mathbf{h}_1$ as stated above. From Lemma 2 we know that $u_i = |\mathfrak{U}_i| = \binom{n}{i} - w\binom{n-w}{i-1}$. Therefore from Corollary 1 it follows that there exist $\mathbf{h}_2, \mathbf{h}_3, \ldots, \mathbf{h}_{t+1}$, such that

$$\delta(H) \leqslant \lfloor \mathcal{D}_{n,k,d}(w,t) \rfloor,$$

where $H = \left(\mathbf{h}_1^\intercal, \mathbf{h}_2^\intercal, \ldots, \mathbf{h}_{t+1}^\intercal\right)^\intercal$.

From the definition of $\delta(H)$ we have that the matrix H covers all but $(\delta(H) - (r - \operatorname{rank} H))$ sets from $\mathfrak{I}$.

We can easily add $\delta(H) - (r - \operatorname{rank} H)$ rows to cover all the rest of the sets in $\mathfrak{I}$ (one row per one set) and also add $r - \operatorname{rank} H$ rows to make the rank equal to r. That is $\delta(H)$ rows in total.

Altogether we have a matrix with $1 + t + \delta(H)$ rows which has the rank r and covers all the sets in $\mathfrak{I}$. The sum of these values is, as we stated before, not more than $1 + t + \lfloor \mathcal{D}_{n,k,d}(w,t) \rfloor$. □

Now we apply the same ideas to Theorem 4 in [10]. We use the following intermediate result.

Lemma 6. Let b be some number, $1 \leqslant b \leqslant r - 2$, and

$$(r-2)(d-1) \leqslant 3 \cdot 2^{d-3} \tag{2.3}$$

Then

$$b - \left(1 - \frac{2^b - x}{2^r - x}\right) \leqslant b\left(1 - \frac{(d-1)2^{r-d+1}}{2^r - x}\right)$$

for any $x < 2^r$.

Proof. 2.3 can be re-written in the following form

$$(d-1)2^{r-d+1} \leqslant \frac{2^r - 2^{r-2}}{r-2}$$

Right-hand side of this inequality is $f(r-2)$, where $f(b) = \left(2^r - 2^b\right)\big/\, b$. Since $f(b)$ is decreasing in b for $1 \leqslant b \leqslant r-2$, $f(b) \geqslant f(r-2)$ for all $1 \leqslant b \leqslant r-2$ and therefore it follows that

$$(d-1)2^{r-d+1} \leqslant \frac{2^r - 2^b}{b}$$

We continue with a sequence of implications:

$$\begin{aligned}
b(d-1)2^{r-d+1} &\leqslant 2^r - 2^b = 2^r - x - 2^b + x \\
\frac{b(d-1)2^{r-d+1}}{2^r - t} &\leqslant 1 - \frac{2^b - t}{2^r - t} \\
b - \left(1 - \frac{2^b - x}{2^r - x}\right) &\leqslant b - \frac{b(d-1)2^{r-d+1}}{2^r - x} \\
b - \left(1 - \frac{2^b - x}{2^r - x}\right) &\leqslant b\left(1 - \frac{(d-1)2^{r-d+1}}{2^r - x}\right)
\end{aligned}$$

□

Theorem 3. Let $\mathcal{D}_{n,k,d}(w,t)$ be defined as in Theorem 2 and $(r-2)(d-1) \leqslant 3 \cdot 2^{d-3}$. Then

$$\rho(\mathcal{C}) \leqslant 1 + \min_{t \geqslant r}\left\{t + \min\left\{i \in \mathbb{N} \;:\; Q_i(\lfloor \mathcal{D}_{n,k,d}(w,t) \rfloor) = 0\right\}\right\}$$

where

$$\begin{aligned}
Q_i(x) &= P_i\left(P_{i-1}\left(\ldots P_1(x)\ldots\right)\right) \\
P_j(x) &= \left\lfloor x\left(1 - \frac{(d-1)2^{r-d+1}}{2^r - (t+1+j)}\right)\right\rfloor
\end{aligned}$$

Proof. Again, as in Theorem 2, we construct a $(t+1) \times n$ matrix (denote it by H_0), such that

$$\delta(H_0) \leqslant \lfloor \mathcal{D}_{n,k,d}(w,t) \rfloor$$

Let $\mathfrak{U}_0 \subset \mathfrak{I}$ be the set of all i-sets $(1 \leqslant i \leqslant d-1)$ not covered by H_0. Denote $\xi_0 = |\mathfrak{U}_0|$ and $\eta_0 = r - \operatorname{rank} H_0$ $(\xi_0 + \eta_0 = \delta(H_0))$.

Add one more new row $\mathbf{h}_1$ randomly chosen from $\mathcal{C}_0^\perp \setminus \{\text{rows of } H_0\}$ and call the resulting $(t+2) \times n$ matrix H_1. Analogously to ξ_0 and η_0 for H_0, we define ξ_1 and η_1 for H_1. Then

$$\begin{aligned}\mathsf{E}\{\xi_1\} = \sum_{\mathcal{S}\in\mathfrak{U}_0} \mathsf{P}\{\mathcal{S} \text{ is not covered by } H_1\} = \sum_{\mathcal{S}\in\mathfrak{U}_0} \mathsf{P}\{\mathcal{S} \text{ is not covered by } \mathbf{h}_1\} \\ \leqslant |\mathfrak{U}_0| \cdot \max_{\mathcal{S}\in\mathfrak{U}_0} \mathsf{P}\{\mathcal{S} \text{ is not covered by } \mathbf{h}_1\} = \xi_0 \cdot \max_{\mathcal{S}\in\mathfrak{U}_0} \left(1 - \frac{|\mathcal{S}|\, 2^{r-|\mathcal{S}|}}{2^r - (t+2)}\right) \\ \leqslant \xi_0 \cdot \max_{1\leqslant i\leqslant d-1} \left(1 - \frac{i2^{r-i}}{2^r-(t+2)}\right) = \xi_0 \left(1 - \frac{(d-1)2^{r-d+1}}{2^r-(t+2)}\right)\end{aligned}$$

Here we used the fact that $f(i) = i2^{r-i}$ is decreasing in i for $i \geqslant 2$, and that $f(1) = f(2)$.

We note that two different non-zero binary vectors are always linearly independent. And H_0 consists of $t+1 \geqslant r+1 \geqslant 2$ different rows, therefore $\operatorname{rank} H_0 \geqslant 2$, i.e. $\eta_0 \leqslant r-2$.

Next, if $\operatorname{rank} H_0 \leqslant r-1$ (which is equivalent to $\eta_0 \geqslant 1$) then $\operatorname{rank} H_1$ could either stay unchanged ($\eta_1 = \eta_0$) or increase by one (and, equivalently, $\eta_1 = \eta_0 - 1$). To calculate the probabilities of these events, we note that any l linearly independent rows from $\mathcal{C}_0^\perp$ span in total 2^l rows (including $\mathbf{0}$). Then

$$\mathsf{P}\{\eta_1 = \eta_0\} = \frac{2^{\eta_0} - (t+2)}{2^r - (t+2)} = 1 - \mathsf{P}\{\eta_1 = \eta_0 - 1\}$$

and therefore

$$\mathsf{E}\{\eta_1\} \; = \; \eta_0 \cdot \frac{2^{\eta_0} - (t+2)}{2^r - (t+2)} + (\eta_0 - 1)\left(1 - \frac{2^{\eta_0} - (t+2)}{2^r - (t+2)}\right) \; = \; \eta_0 - \left(1 - \frac{2^{\eta_0} - (t+2)}{2^r - (t+2)}\right)$$

Next, we apply Lemma 6 with $b = \eta_0$ and $x = t+2$. Indeed, as we stated before $\eta_0 \leqslant r-2$, and we consider the case $\eta_0 \geqslant 1$. Additionally, $t+2 < 2^r$ because $t+2$ is the number of the rows in the matrix H_1 and 2^r is the maximum number of rows in any parity-check matrix for $\mathcal{C}$. Therefore

$$\mathsf{E}\{\eta_1\} = \eta_0 - \left(1 - \frac{2^{\eta_0} - (t+2)}{2^r - (t+2)}\right) \leqslant \eta_0 \left(1 - \frac{(d-1)2^{r-d+1}}{2^r - (t+2)}\right)$$

In case $\eta_0 = 0$ it necessary follows that $\eta_1 = 0$ and

$$\mathsf{E}\{\eta_1\} \leqslant \eta_0 \left(1 - \frac{(d-1)2^{r-d+1}}{2^r - (t+2)}\right)$$

also holds. Thus

$$\mathsf{E}\{\delta(H_1)\} = \mathsf{E}\{\xi_1\} + \mathsf{E}\{\eta_1\} \leqslant \delta(H_0) \left(1 - \frac{(d-1)2^{r-d+1}}{2^r - (t+2)}\right)$$

Therefore, by Lemma 1 there exists H_1 such that $\delta(H_1) \leqslant P_1(\delta(H_0))$. We iterate this process and after i steps obtain $(t+1+i) \times n$ matrix H_i with $\delta(H_i) \leqslant Q_i(\delta(H_0)) \leqslant Q_i\left(\lfloor \mathcal{D}_{n,k,d}(w,t) \rfloor\right)$.

Iterations should be stopped when $Q_i\left(\lfloor \mathcal{D}_{n,k,d}(w,t) \rfloor\right) = 0$.

□

Corollary 2. In the settings of Theorem 3

$$\rho(\mathcal{C}) \leqslant 1 + \min_{t\geqslant r} \left\{ t + \min\left\{ i \in \mathbb{N} \; : \; Q_i(\lfloor \mathcal{D}_{n,k,d}(d^\perp,t) \rfloor) = 0 \right\} \right\}$$

We also modify the analysis in Theorem 3 by using two words in $\mathcal{C}_0^\perp$ of weight $d^\perp$. For that we modify the result of Lemma 2.

Lemma 7. Let $\mathbf{h}_1$ and $\mathbf{h}_2$ be two different codewords from $\mathcal{C}_0^\perp$, such that $\mathsf{w}(\mathbf{h}_1) = \mathsf{w}(\mathbf{h}_2) = d^\perp$. Then the matrix consisting of these two rows covers at least

$$2d^\perp \binom{n-d^\perp}{i-1} - \left\lfloor \frac{d^\perp}{2} \right\rfloor \binom{n - 2d^\perp + \left\lfloor \frac{d^\perp}{2} \right\rfloor}{i-1}$$

i-sets for $i = 1,2,\ldots,d-1$.

Proof. Let $\Delta = |\operatorname{supp}(\mathbf{h}_1) \cap \operatorname{supp}(\mathbf{h}_2)| \leqslant \lfloor d^\perp/2 \rfloor$. Then the matrix $(\mathbf{h}_1, \mathbf{h}_2)^\intercal$ covers the following number of i-sets:

$$\begin{aligned} &\sharp \text{ covered by } \mathbf{h}_0 + \sharp \text{ covered by } \mathbf{h}_1 - \sharp \text{ covered by both} \\ &= d^\perp \binom{n-d^\perp}{i-1} + d^\perp \binom{n-d^\perp}{i-1} - \Delta \binom{n-2d^\perp+\Delta}{i-1} \\ &\geqslant 2d^\perp \binom{n-d^\perp}{i-1} - \left\lfloor \frac{d^\perp}{2} \right\rfloor \binom{n - 2d^\perp + \left\lfloor \frac{d^\perp}{2} \right\rfloor}{i-1} \end{aligned}$$

Here we used the fact that $\Delta\binom{n-2d^\perp+\Delta}{i-1}$ is strictly increasing in Δ, when n, i and $d^\perp$ are fixed. □

Now we incorporate this result into Corollary 2.

Theorem 4. If $(r-2)(d-1) \leqslant 3 \cdot 2^{d-3}$ and $\mathcal{C}^\perp$ contains at least two codewords of weight $d^\perp$ then

$$\rho(\mathcal{C}) \leqslant 2 + \min_{t \geqslant r} \left\{ t + \min \left\{ i \in \mathbb{N} \ : \ Q_i(\lfloor \mathcal{D}^{(2)}_{n,k,d}(d^\perp,t) \rfloor) = 0 \right\} \right\}$$

where

$$\begin{aligned} \mathcal{D}^{(2)}_{n,k,d}(d^\perp,t) = \sum_{i=1}^{d-1} \left(\binom{n}{i} - 2d^\perp \binom{n-d^\perp}{i-1} + \left\lfloor \frac{d^\perp}{2} \right\rfloor \binom{n - 2d^\perp + \left\lfloor \frac{d^\perp}{2} \right\rfloor}{i-1} \right) \times \\ \times \prod_{j=3}^{t+2} \left(1 - \frac{i2^{r-i}}{2^r - j}\right) + \frac{1}{2^{t-r}} \left(1 + \frac{2/3}{2^{t-r+1}-1}\right) \end{aligned}$$

$$\begin{aligned} Q_i(x) &= P_i\left(P_{i-1}\left(\ldots P_1(x)\ldots\right)\right) \\ P_j(x) &= \left\lfloor x \left(1 - \frac{(d-1)2^{r-d+1}}{2^r - (t+2+j)}\right) \right\rfloor \end{aligned}$$

Proof. These two words cover the number of i-sets given in Lemma 7. Then we proceed as in the proof of Theorem 3. □

Although the condition (2.3) is not very restrictive, it is desirable to obtain a bound that is applicable to all the codes. We note that the condition is required in the proof only to guarantee uniform decrease of ξ and η. Therefore we could repeat the argument from the proof of Theorem 4 for ξ only and then make sure that we have the matrix of the required rank, r.

Theorem 5. If $\mathcal{C}^{\perp}$ contains at least two codewords of weight $d^{\perp}$ then

$$\rho(\mathcal{C}) \leqslant 2 + \min_{t} \left\{ t + \min\left\{ i \in \mathbb{N} \ : \ Q_i(\lfloor \mathcal{B}^{(2)}_{n,k,d}(d^{\perp},t) \rfloor) = 0 \right\} \right\} + (r - d + 1)$$

where

$$\mathcal{B}^{(2)}_{n,k,d}(d^{\perp},t) = \sum_{i=1}^{d-1} \left(\binom{n}{i} - 2d^{\perp}\binom{n-d^{\perp}}{i-1} + \left\lfloor \frac{d^{\perp}}{2} \right\rfloor \binom{n - 2d^{\perp} + \left\lfloor \frac{d^{\perp}}{2} \right\rfloor}{i-1} \right) \times$$
$$\times \prod_{j=3}^{t+2} \left(1 - \frac{i2^{r-i}}{2^r - j} \right)$$

$$Q_i(x) = P_i\left(P_{i-1}\left(\ldots P_1(x)\ldots\right)\right)$$
$$P_j(x) = \left\lfloor x\left(1 - \frac{(d-1)2^{r-d+1}}{2^r - (t+2+j)}\right)\right\rfloor$$

Proof. The proof repeats all the steps in the proof of Theorem 4, yet at the end we are not guaranteed that the matrix we have constructed has rank r. However, as we have covered all the i-sets for $i = 1,2,\ldots,d-1$, the rank of the matrix is at least $d-1$. Hence, by adjoining at most $r-(d-1)$ rows, we finally obtain the required parity-check matrix. □

The methods of Theorems 2, 3, 4, 5 can be summarized in the diagram presented at Figure 2.1.

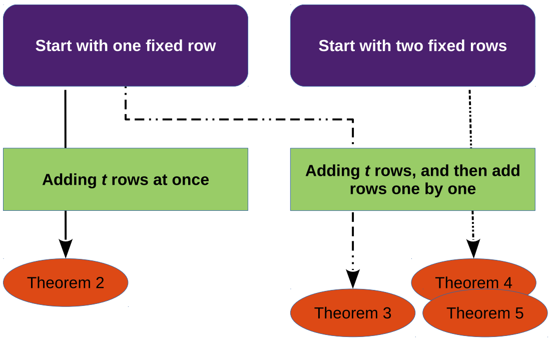


Figure 2.1: Summary of methods to find upper bound on stopping redundancy

For some codes deterministically adding three or more rows in the beginning could lead to even better bounds. Currently, we try to derive bounds of more general form, and that is still work in progress.

3

Stopping Redundancy Hierarchy

A weaker requirement on the parity-check matrix of the code can be considered. For instance, as it was suggested in [6] one could require the stopping distance of the code to be not less than l, for some $1 \leqslant l \leqslant d$. In that case , the number of rows in the parity-check matrix can be smaller than the stopping redundancy of the code.

Definition 2 ([6, Definition 2.4])**.** For $l \leqslant d$, the l-th stopping redundancy of $\mathcal{C}$ is the smallest nonnegative integer $\rho_l(\mathcal{C})$ such that there exists a (possibly redundant) parity-check matrix H of $\mathcal{C}$ with $\rho_l(\mathcal{C})$ rows and stopping distance at least l. The ordered set of integers

$$(\rho_1(\mathcal{C}), \rho_2(\mathcal{C}), \ldots, \rho_d(\mathcal{C}))$$

is called the *stopping redundancy hierarchy* of $\mathcal{C}$.

Note that the (conventional) stopping redundancy $\rho(\mathcal{C})$ is equal to $\rho_d(\mathcal{C})$ by definition. For codes with the minimum distance $d \geqslant 3$, neither two columns of the parity-check matrix are identical nor any of the columns equal to the all-zero vector. Therefore, $\rho_1(\mathcal{C}) = \rho_2(\mathcal{C}) = \rho_3(\mathcal{C}) = n - k$. Consequently, only stopping redundancies of order larger than three are considered.

It is interesting to apply the ideas in Chapter 2 to the stopping redundancy hierarchy.

Theorem 6. For $4 \leqslant l \leqslant d$, if $\mathcal{C}^\perp$ contains at least two codewords of minimum weight, then

$$\rho_l(\mathcal{C}) \leqslant 2 + \min_t \left\{ t + \min \left\{ i \in \mathbb{N} \ : \ Q_i(\lfloor \mathcal{B}^{(2)}_{n,k,l}(d^\perp, t) \rfloor) = 0 \right\} \right\} + (r - l + 1)$$

Moreover, if $(r-2)(l-1) \leqslant 3 \cdot 2^{l-3}$ then

$$\rho_l(\mathcal{C}) \leqslant 2 + \min_{t \geqslant r} \left\{ t + \min \left\{ i \in \mathbb{N} \ : \ Q_i(\lfloor \mathcal{D}^{(2)}_{n,k,l}(d^\perp, t) \rfloor) = 0 \right\} \right\}$$

where

$$\mathcal{B}^{(2)}_{n,k,l}(d^\perp,t) = \sum_{i=1}^{l-1} \left(\binom{n}{i} - 2d^\perp \binom{n-d^\perp}{i-1} + \left\lfloor \frac{d^\perp}{2} \right\rfloor \binom{n-2d^\perp+\left\lfloor \frac{d^\perp}{2} \right\rfloor}{i-1} \right) \times \\ \times \prod_{j=3}^{t+2} \left(1 - \frac{i2^{r-i}}{2^r-j} \right)$$

$$\mathcal{D}^{(2)}_{n,k,l}(d^\perp,t) = \sum_{i=1}^{l-1} \left(\binom{n}{i} - 2d^\perp \binom{n-d^\perp}{i-1} + \left\lfloor \frac{d^\perp}{2} \right\rfloor \binom{n-2d^\perp+\left\lfloor \frac{d^\perp}{2} \right\rfloor}{i-1} \right) \times \\ \times \prod_{j=3}^{t+2} \left(1 - \frac{i2^{r-i}}{2^r-j} \right) + \frac{1}{2^{t-r}} \left(1 + \frac{2/3}{2^{t-r+1}-1} \right)$$

$$Q_i(x) = P_i\left(P_{i-1}\left(\dots P_1(x)\dots\right)\right),$$
$$P_j(x) = \left\lfloor x \left(1 - \frac{(l-1)2^{r-l+1}}{2^r-(t+2+j)} \right) \right\rfloor$$

Proof. The proof is analogous to the proofs of Theorem 4 and Theorem 5, where instead of d we use l. □

The paper [6] studied the stopping redundancy hierarchy and for general codes they obtained three different upper bounds.

For $l \leqslant \left\lfloor \frac{d+1}{2} \right\rfloor$ it was shown in [6, Theorem 3.8] that

$$\rho_l(\mathcal{C}) \leqslant \left\lceil \frac{1+\log \sum_{j=1}^{l-1} \left(\binom{n}{j} - \binom{n-j}{j} \right)}{-\log\left(1-\frac{l-1}{2^{l-1}}\right)} \right\rceil + (r-l+1)$$

For the other values of l the result from [5, Theorem 4] was adapted to [6, Theorem 3.11], which is known to be loose:

$$\rho_l(\mathcal{C}) \leqslant \sum_{i=1}^{l-2} \binom{r}{i}$$

Another bound was obtained [6, Theorem 3.12]. Let Θ be the set of all subcodes of the dual code $\mathcal{C}^\perp$ of a linear $[n,k,d]$ code $\mathcal{C}$ that have support weight[1] n and dual distance l. Furthermore, let the dimensions of the subcodes in Θ be K_i, $i = 1,2,\dots,|\Theta|$, and define $K = \min_i K_i$. Then it was shown that

$$\rho_l(\mathcal{C}) \leqslant \sum_{i=1}^{l-2} \binom{K}{i}$$

However, additional bounds, which improve on Theorems 3.11 and 3.12, were developed in [6] for cyclic codes. Since this thesis studies only bounds on general codes, we omit those results.

[1] The support weight of a subcode of a code is defined as the number of positions for which at least one of the codewords of the subcode is nonzero.

4

Numerical Experiments

In this chapter we discuss the numerical comparison of bounds on the stopping redundancy obtained in [5], [7], [10] with the new results obtained in the previous chapter. We consider two codes: the extended [24,12,8] binary Golay code and the extended [48,24,12] binary Quadratic Residue code. Both of them are known to be self-dual (cf. [14]).

The extended [24,12,8] binary Golay code is arguably a remarkable binary block code. It is often used as a benchmark in studies of code structure and decoding algorithms. The code is self-dual, therefore $d^{\perp} = 8$. Moreover, it is known [13, p. 67] that there are 759 codewords of the minimum weight.

The example of (conventional) parity-check matrix of the code is shown in Table 4.1, where the blank spaces denote zeroes.

Table 4.1: Parity-check matrix of the extended [24,12,8] Golay code

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

Schwartz and Vardy in [5] used a greedy (lexicographic) computer search and found that the actual stopping redundancy of the extended [24,12,8] binary Golay code is at most 34.

It is known [13, p. 604] that there are 17296 codewords of minimum weight in the extended [48,24,12] binary Quadratic Residue (QR) code. The example of its (conventional) parity-check matrix is given in Table 4.2.

Therefore the bounds in Theorems 2, 3 and 4 are applicable to both of the codes. The comparison of upper bounds on stopping redundancy summary is given at Table 4.3.

We also compare the bounds on stopping redundancy hierarchy in the previous chapter with the results for general codes, obtained in [6][1]. The numerical results are presented in Table 4.4 and Table 4.5.

[1] The bounds in [6] for cyclic codes are not applicable because neither of the codes is cyclic.

Table 4.2: Parity-check matrix of the extended [48,24,12] QR code

$$\begin{pmatrix}
111111 & 111111 & & & & & & \\
 & 111111 & 111111 & & & & & \\
 & & 111111 & 111111 & & & & \\
 & & & & 111111 & 111111 & & \\
 & & & & & 111111 & 111111 & \\
 & & & & & & 111111 & 111111 \\
000111 & 111000 & 000111 & 111000 & & & & \\
 & & 100001 & 111001 & 010111 & 010010 & & \\
 & & 010010 & 111010 & 100111 & 100100 & & \\
 & & & & 000111 & 111000 & 000111 & 111000 \\
011100 & 011111 & 011010 & 011001 & 011000 & & & \\
001110 & 110111 & 010011 & 110010 & 110000 & & & \\
000011 & 011011 & 010001 & 001100 & 000011 & & & \\
000001 & 001110 & 001101 & 101001 & 000110 & & & \\
 & 010111 & 000101 & 100111 & 101110 & 110000 & & \\
 & 001111 & 000110 & 010111 & 110011 & 011000 & & \\
 & & 001001 & 010111 & 010100 & 010100 & 011000 & \\
 & & 000011 & 011110 & 001001 & 110011 & 010111 & \\
 & & & 110000 & 001100 & 010100 & 011011 & 110000 \\
 & & & 011000 & 011010 & 110010 & 001110 & 100000 \\
 & & & 000101 & 001010 & 110011 & 101000 & 010010 \\
 & & & 000011 & 010100 & 010001 & 001111 & 100100 \\
 & 000110 & 001010 & 010111 & 100111 & 000110 & 010100 & \\
 & 000011 & 100111 & 100111 & 111100 & 010111 & 000110 &
\end{pmatrix}$$

Table 4.3: Upper bounds on the stopping redundancy

	[24, 12, 8] Golay	[48, 24, 12] QR
[5, Theorem 4]	2509	4 540 385
[10, Theorem 1]	198	3655
[10, Theorem 3]	194	3655
[10, Theorem 4]	187	3577
[10, Theorem 7]	182	3564
Theorem 2	187	3616
Theorem 3	180	3538
Theorem 4	176	3509

Table 4.4: Bounds on the stopping redundancy hierarchy, ρ_l, for the extended [24,12,8] Golay code

l	[6, Theorem 3.8]	[6, Theorem 3.11]	[6, Theorem 3.12]	Theorem 6
4	26	78	—	24
5	—	298	—	34
6	—	793	385	58
7	—	1585	—	102
8	—	2509	—	176

Table 4.5: Bounds on the stopping redundancy hierarchy, ρ_l, for the extended [48,24,12] QR code

l	[6, Theorem 3.8]	[6, Theorem 3.11]	Theorem 6
4	42	300	40
5	62	2324	56
6	105	12950	90
7	—	55454	156
8	—	190050	284
9	—	536154	511
10	—	1 271 625	974
11	—	2 579 129	1851
12	—	4 540 385	3509

5

Conclusion

The Han-Siegel-Vardy probabilistic bounds [10] are the best currently known bounds on the stopping redundancy of general binary linear codes. In this thesis we presented several improvements upon these bounds based on additional preliminary step of judiciously selecting the first and the second rows of the parity-check matrix. By using similar technique, we also improve on the stopping redundancy hierarchy results for general binary linear codes in [6]. The improvements were tested/verified numerically for the extended [24,12,8] binary Golay code and the extended [48,24,12] binary QR code.

The following research questions are still open.

1. The presented bounds are not explicit. They involve solving minimisation problems. Hence their asymptotic behaviour is not obvious. It would be interesting to derive asymptotic results for these bounds.

2. The bounds developed in this thesis are not constructive. Tight constructive bounds are still not known.

3. As it is illustrated by the numerical results for the extended [24,12,8] binary Golay code, there is still a significant gap between the theoretical bounds on the stopping redundancy and the experimental results found by the computer search. On the other hand the introduced bounds are tighter than the bounds by Han-Siegel-Vardy [10].

Bibliography

[1] R. Gallager, Low-density parity-check codes, Information Theory, IRE Transactions on 8 (1) (1962) 21–28.

[2] R. J. McEliece, D. J. C. MacKay, J.-F. Cheng, Turbo decoding as an instance of Pearl's "belief propagation" algorithm, Selected Areas in Communications, IEEE Journal on 16 (2) (1998) 140–152.

[3] P. Oswald, A. Shokrollahi, Capacity-achieving sequences for the erasure channel, Information Theory, IEEE Transactions on 48 (12) (2002) 3017–3028.

[4] C. Di, D. Proietti, I. E. Telatar, T. J. Richardson, R. L. Urbanke, Finite-length analysis of low-density parity-check codes on the binary erasure channel, Information Theory, IEEE Transactions on 48 (6) (2002) 1570–1579.

[5] M. Schwartz, A. Vardy, On the stopping distance and the stopping redundancy of codes, Information Theory, IEEE Transactions on 52 (3) (2006) 922–932.

[6] T. Hehn, O. Milenkovic, S. Laendner, J. B. Huber, Permutation decoding and the stopping redundancy hierarchy of cyclic and extended cyclic codes, Information Theory, IEEE Transactions on 54 (12) (2008) 5308–5331.

[7] J. Han, P. H. Siegel, Improved upper bounds on stopping redundancy, Information Theory, IEEE Transactions on 53 (1) (2007) 90–104.

[8] T. Etzion, On the stopping redundancy of Reed-Muller codes, Information Theory, IEEE Transactions on 52 (11) (2006) 4867–4879.

[9] J. H. Weber, K. A. Abdel-Ghaffar, Stopping set analysis for Hamming codes, in: Information Theory Workshop, IEEE, 2005, pp. 244–247.

[10] J. Han, P. H. Siegel, A. Vardy, Improved probabilistic bounds on stopping redundancy, Information Theory, IEEE Transactions on 54 (4) (2008) 1749–1753.

[11] H. D. Hollmann, L. M. Tolhuizen, Generic erasure correcting sets: Bounds and constructions, Journal of Combinatorial Theory, Series A 113 (8) (2006) 1746–1759.

[12] H. D. Hollmann, L. M. Tolhuizen, On parity-check collections for iterative erasure decoding that correct all correctable erasure patterns of a given size, Information Theory, IEEE Transactions on 53 (2) (2007) 823–828.

[13] F. J. MacWilliams, N. J. A. Sloane, The theory of error-correcting codes, Elsevier, 1977.

[14] S. K. Houghten, C. W. Lam, L. H. Thiel, J. A. Parker, The extended quadratic residue code is the only (48, 24, 12) self-dual doubly-even code, Information Theory, IEEE Transactions on 49 (1) (2003) 53–59.

Licence

I, Yauhen Yakimenka, born 24.04.1986,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:

   (a) reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and

   (b) make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

   Optimization of Parity-Check Matrices of LDPC Codes,
   supervised by Dr Vitaly Skachek

2. I am aware of the fact that the author retains these rights.

3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, Wednesday 28th May, 2014