Computing and Informatics, Vol. 44, 2025, 429–444, doi: 10.31577/cai_2025_2_429

NEW FAMILY OF LINEAR 3-ERASURE CORRECTING BLOCK CODES WITH POSSIBLE APPLICATION IN STORAGE SYSTEMS

Peter Farkaš

Institute of Multimedia Information and Communication Technologies
Faculty of Electrical Engineering and Information Technology
Slovak University of Technology in Bratislava
Ilkovičova 3, 812 19 Bratislava, Slovakia
&
Institute of Applied Informatics
Faculty of Informatics Pan-European University, Slovakia
Tematínska 10, 851 05 Bratislava, Slovakia
e-mail: p.farkas@ieee.org

Katarína Farkašová, Frederik Pavelka, Martin Rakús

Institute of Multimedia ICT, FEI STU
Ilkovičova 3, 812 19 Bratislava, Slovakia
e-mail: {katarina.farkasova, frederik.pavelka, martin.rakus}@stuba.sk

Abstract. A construction of a new family of three erasure correcting linear block codes over $GF(q)$ with characteristic two together with their syndrome decoding procedures are presented in this paper. The designed code distance of four was confirmed by demonstrating a decoding algorithm capable of correcting three erasures. The second confirmation was obtained from the weight spectra of selected codes, which were calculated using Krawtchouck polynomials derived from the weight spectra of their dual codes.

Keywords: Erasure correcting code, linear block code, Vandermonde matrix, parity check matrix, decoding

Mathematics Subject Classification 2010: 58F15, 58F17, 53C35

1 INTRODUCTION

In the past, the construction of practical erasure-correcting codes was mainly motivated by their applications in packet-switched networks and distributed storage systems. A practical family of codes for erasure correction was proposed by McAuley in [1]. It was named Weighted sum codes. Their main advantage is very efficient simple hardware and software realizations of corresponding encoders and decoders. In [2] Farkaš noted that these codes are equivalent to extended Reed Solomon codes and shortened versions of extended Reed Solomon codes, respectively. In [2], it was also demonstrated that it is possible to correct one error in each codeword of such codes.

The Weighted sum codes inspired further research of codes with very simple encoding and decoding procedures. In [3] Farkaš and Baylis proposed other families of related codes, namely t-information error correcting codes, single error correcting codes and conditionally double error correcting codes. In all these families the codes are distinguished by high coding rates and low complexity of decoding. Later in [4] following this direction of research a family of double error correcting codes was proposed. More recently, continuation of these efforts brought a surprising result, namely five times extended Reed Solomon codes constructed over finite fields $GF(2^{\xi})$ where ξ is an odd integer, that were discovered in [5]. In [6] it was shown that these codes can correct up to two errors and they could be decoded using syndrome decoding. In this paper a new family of three erasure correcting codes is proposed which is distinguished by simple implementation of decoding procedures.

In recent times DNA computing, DNA communication and especially DNA storage research have invoked fresh interest in constructing new coding schemes adapted to the needs of these areas [7, 8].

DNA storage has several advantages that may make it the preferred storage system of the future:

- it has high density,
- it has longevity,
- it has small CO_2 footprint,
- it is a universal storage medium used by nature.

However, storage systems need to use codes which correct the impairments that occur during synthesizes and reading the information stored in DNA strings [9, 10, 11, 12, 13].

Levick at al. showed in [14] that if the DNA multi-draw storage channel is modeled as an erasure channel, then its capacity could be achieved using linear codes. Reed Solomon codes, which belong to linear block codes, were used in [15] to protect information in DNA storage. In [16, 17] the codes correcting single or double deletions (erasures) were considered for DNA storage.

Data is stored in DNA strings with lengths of several hundred nucleotides. The state-of-the-art results indicate that erasure correcting block codes over finite fields

similar to Reed Solomon codes with some flexibility in the encoded block lengths could be useful for DNA storage.

The paper is organized as follows. In Section 2 the basic theoretical background is given. In Section 3 the new family of codes will be presented. In Section 4 the syndrome decoding algorithm for the new family of codes will be proposed. In Section 5 the encoding procedures will be described. The last section will give some concluding remarks on possible further research in this direction.

2 BASIC THEORETICAL BACKGROUND

A linear block code C is defined as a k-dimensional subspace of an n-dimensional vector space defined over a finite field $GF(q)$. In practical terms n can be interpreted as a codeword length and k as the number of symbols which contain the encoded information – the so called payload. $R_k = k/n$ is a code rate. The Hamming weight of the codeword $\mathbf{c} = (c_{n-1}, c_{n-2}, \ldots, c_0)$, is the number of its non-zero symbols or in other words non-zero coordinates in the vector used for its mathematical description.

Any linear block code can be defined using its parity check matrix $\mathbf{H}$ with dimensions: $(n-k) \times n$. The rows of $\mathbf{H}$ describe the so-called parity check equations valid for all $\mathbf{c} \in C$. In a compact way it is expressed by the following equation

$$\mathbf{c}.\mathbf{H}^T = \mathbf{0}. \tag{1}$$

Error correcting linear block code construction has particularly contradictory goals. On one hand the goal is to minimize their redundancy given by $n - k$ and on the other hand to maximize the minimal Hamming distance between their codewords. The Hamming distance between two codewords $\mathbf{c}_i \in C$ and $\mathbf{c}_j \in C$ denoted as $d(\mathbf{c}_i, \mathbf{c}_j)$ is the number of coordinates by which the two vectors differ. The minimal Hamming distance between any two codewords of a code is denoted as d_m. It is well known that the number of errors t which can be corrected in one codeword of a code with code distance d_m is

$$2t + 1 \leq d_m. \tag{2}$$

In non-binary error control codes, the error can only be corrected if its position and value is known, therefore each error is connected with two unknowns. In contrast, in case of erasure the position is known and only one unknown, namely the erasure value has to be obtained in order to correct it. Therefore the number of correctable erasures ε is connected with code distance by this expression

$$\varepsilon + 1 \leq d_m. \tag{3}$$

In some applications and for some codes it is also possible to use decoding of errors and erasures. In this case the following inequality must hold

$$2t + \varepsilon + 1 \leq d_m. \tag{4}$$

The basic parameters of linear block codes are given in a compact way using a triple $[n, k, d_m]_{GF(q)}$. Different tables give the state-of-the-art knowledge about the bounds on d_m as function of n and k [18] which we will denote $d_m(n, k)$.

It is worth noting that for some of the codes mentioned in the introduction it is known that they are reaching the upper bounds on $d_m(n, k)$ and so belong to the class of so-called best linear block codes. For example the five times extended Reed Solomon codes defined in [5] belong to such a class. However, the tables give data about the bounds and basic parameters of best-known linear block codes only up to certain values of codeword lengths and code dimensions, therefore it is not possible to decide whether all mentioned codes belong to such a class of the best linear block codes.

Some linear block codes belong to the class of cyclic codes. They have the property that each cyclic shift of codeword from a cyclic code is also a codeword from that code. For cyclic codes it is very useful to denote the codewords as polynomials

$$c(x) = c_{n-1}x^{n-1} + c_{n-2}x^{n-2} + \cdots + c_1x + c_0. \tag{5}$$

One of the advantages of such an expression is that some polynomials are directly connected to hardware realizations of different linear shift registers with feedback and digital filters which are used to realize the encoders and decoders of cyclic codes.

The other advantage is that many codes are defined by roots of the polynomials which represent codewords of cyclic codes. BCH codes and Reed Solomon codes are the most popular examples of it.

This allows for checking whether a given polynomial is a codeword by substituting all roots by which the code is defined into the corresponding polynomial. If, for some roots, the evaluation gives nonzero values, these values are known as syndromes and can be used in so-called syndrome decoding in order to correct the corresponding polynomial or vector in such a way that, as a result, an estimation of the codeword is obtained.

The model which will be used to describe erasures is depicted in Figure 1.

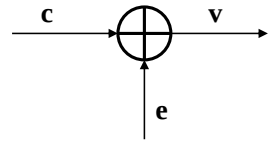

Figure 1. Erasure channel model where **c** and **v** and **e** denote transmitted codeword, received vector and erasures modelling vector respectively. In this model erasures (nonzero coordinates of **e**) are inverse elements with respect to addition in $GF(q)$.

Alternatively, polynomials $c(x)$, $v(x)$ and $e(x)$ can be used to represent these vectors as well, which will be more convenient for the decoding description in this paper.

3 THREE ERASURE CORRECTING LINEAR BLOCK CODES

The new family of codes proposed in this paper is defined by the parity check matrix of the codes defined by different finite fields $GF(q)$ with characteristic 2 in Theorem 1.

Theorem 1. The following control matrix:

$$\mathbf{H} = \begin{bmatrix} \mathbf{A} & \dots & \mathbf{A} & \mathbf{A} & \mathbf{A} & \mathbf{A} \\ \mathbf{B}_{q-2} & \dots & \mathbf{B}_3 & \mathbf{B}_2 & \mathbf{B}_1 & \mathbf{B}_0 \end{bmatrix}, \tag{6}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & \dots & 1 & 1 & 1 \\ \alpha^{q-2} & \dots & \alpha^2 & \alpha^1 & 1 \\ \alpha^{2(q-2)} & \dots & \alpha^4 & \alpha^2 & 1 \end{bmatrix}, \tag{7}$$

$$\mathbf{B}_0 = \begin{bmatrix} 1 & \dots & 1 & 1 & 1 \\ 1 & \dots & 1 & 1 & 1 \end{bmatrix}, \tag{8}$$

$$\mathbf{B}_1 = \begin{bmatrix} \alpha & \dots & \alpha & \alpha & \alpha \\ \alpha^2 & \dots & \alpha^2 & \alpha^2 & \alpha^2 \end{bmatrix}, \tag{9}$$

$$\mathbf{B}_2 = \begin{bmatrix} \alpha^2 & \dots & \alpha^2 & \alpha^2 & \alpha^2 \\ \alpha^4 & \dots & \alpha^4 & \alpha^4 & \alpha^4 \end{bmatrix}, \tag{10}$$

$$\mathbf{B}_3 = \begin{bmatrix} \alpha^3 & \dots & \alpha^3 & \alpha^3 & \alpha^3 \\ \alpha^6 & \dots & \alpha^6 & \alpha^6 & \alpha^6 \end{bmatrix}, \tag{11}$$

$$\dots$$

$$\mathbf{B}_{q-2} = \begin{bmatrix} \alpha^{(q-2)} & \dots & \alpha^{(q-2)} & \alpha^{(q-2)} & \alpha^{(q-2)} \\ \alpha^{2(q-2)} & \dots & \alpha^{2(q-2)} & \alpha^{2(q-2)} & \alpha^{2(q-2)} \end{bmatrix} \tag{12}$$

and α is a primitive element from $GF(q)$, defines a family of linear block codes over $GF(q)$, which could correct up to 3 erasures in each codeword.

Proof. In order to correct 3 erasures we need to obtain a system of three linearly independent equations from which the erasure values could be computed. By observing the structure of the matrix $\mathbf{H}$ defined in (6) it is obvious that it is possible to obtain at least 3 linearly independent equations from the rows of $\mathbf{H}$ for any possible combination of 1, 2 or 3 erasure positions. Therefore this control matrix defines a 3 erasure correcting code. □

(More details on decoding will be given in the next section). Let us, however, first introduce the basic parameters of the codes constructed using matrix $\mathbf{H}$ (6). The codeword lengths depend on q, namely $n = (q-1)^2$. On the other hand the number of parity check symbols is constant, namely $n - k = 5$. Therefore the code

rate of the proposed code is: $R_k = [(q-1)^2 - 5]/(q-1)^2$. For example, if the proposed code is constructed over $GF(16)$ then $n = 225$ and $k = 220$ with the code rate of $R_k = 0.977$. The standard Reed Solomon code correcting three erasures defined over $GF(16)$ has $n = 15$, $k = 12$ and $R_k = 0.8$. It can be seen that the proposed codes have higher code rates than Reed Solomon codes constructed over the same finite fields.

4 ERASURE DECODING IN CODEWORDS OF THE PROPOSED CODES

In this chapter a decoding algorithm, which could be used for any code from the proposed family of 3 erasure correcting codes will be described. It was inspired by the well-known syndrome decoding of Reed Solomon codes. Reed Solomon codes in their original form could be defined by an $\mathbf{H}$ matrix, which is equivalent to a single Vandermonde matrix. In contrast to the Reed Solomon codes the first 3 rows (from top) of the $\mathbf{H}$ matrix (6), which defines the new codes are composed of $q-1$ Vandermonde matrices $\mathbf{A}$ and the last 2 (bottom) rows are composed of matrices $\mathbf{B}_0$, $\mathbf{B}_1$, ..., $\mathbf{B}_{q-2}$. By closer inspection of $\mathbf{B}_j$; $j = 0, \ldots, q-2$ it could be seen that the last two rows contain columns which represent "block-interleaved" Vandermonde matrices. In particular, $\mathbf{B}_0$ is composed of columns containing the rightmost elements of the last two rows of $\mathbf{A}$ (7). $\mathbf{B}_1$ is composed of columns containing the second rightmost elements of the last two rows of $\mathbf{A}$ and so on.

This special structure guarantees that enough linearly independent equations for syndrome decoding are available to the decoder independently of the positions of the 3 erasures. The model in Figure 1 represents a transmission channel with erasures which can be described as

$$\mathbf{v} = \mathbf{c} + \mathbf{e}, \tag{13}$$

where vectors $\mathbf{c}$, $\mathbf{v}$ and $\mathbf{e}$ represent the transmitted codeword, received word and erasure vector, respectively.

In this model erasures (non-zero coordinates of vector $\mathbf{e}$) are inverse elements with respect to addition in $GF(q)$. Therefore, the goal of decoding is to calculate the values of erasures. One method is to use syndromes or, more specifically, syndrome equations.

In the syndrome decoding algorithm the *first step* is to calculate syndromes using equation

$$\mathbf{S} = \mathbf{v}.\mathbf{H}^T, \tag{14}$$

where $\mathbf{S} = (S_0, S_1, S_2, S_3, S_4)$ is a syndrome vector, and $\mathbf{H}^T$ is a transposed $\mathbf{H}$ matrix. Its coordinates are syndromes, corresponding to the five rows of control matrix $\mathbf{H}$ given by indices in order from the top to the bottom.

In order to get a more detailed description for calculating particular syndromes let's use the following notation for the received vector:

$$\begin{aligned}\mathbf{v} = (&v_{q-2,q-2}, \ldots, v_{q-2,1}, v_{q-2,0} \mid, \ldots \\ &\cdots \mid v_{q-2,1}, \ldots, v_{1,1}, v_{0,1}, \mid \cdots \mid v_{q-2,0}, \ldots, v_{1,0}, v_{0,0}).\end{aligned} \tag{15}$$

Using this notation:

$$\begin{aligned}\mathbf{S}_0 = &(v_{0,0} + v_{1,0} + v_{2,0} + \cdots + v_{q-2,0}) \\ &+ (v_{0,1} + v_{1,1} + v_{2,1} + \cdots + v_{q-2,1}) \\ &\vdots \\ &+ (v_{0,q-2} + v_{1,q-2} + v_{2,q-2} + \cdots + v_{q-2,q-2}),\end{aligned} \tag{16}$$

$$\begin{aligned}\mathbf{S}_1 = &(v_{0,0} + v_{0,1} + v_{0,2} + \cdots + v_{0,q-2})\alpha^0 \\ &+ (v_{1,0} + v_{1,1} + v_{1,2} + \cdots + v_{1,q-2})\alpha^1 \\ &\vdots \\ &+ (v_{q-2,0} + v_{q-2,1} + v_{q-2,2} + \cdots + v_{q-2,q-2})\alpha^{q-2},\end{aligned} \tag{17}$$

$$\begin{aligned}\mathbf{S}_2 = &(v_{0,0} + v_{0,1} + v_{0,2} + \cdots + v_{0,q-2})\alpha^0 \\ &+ (v_{1,0} + v_{1,1} + v_{1,2} + \cdots + v_{1,q-2})\alpha^2 \\ &\vdots \\ &+ (v_{q-2,0} + v_{q-2,1} + v_{q-2,2} + \cdots + v_{q-2,q-2})\alpha^{2(q-2)},\end{aligned} \tag{18}$$

$$\begin{aligned}\mathbf{S}_3 = &(v_{0,0} + v_{1,0} + v_{2,0} + \cdots + v_{q-2,0})\alpha^0 \\ &+ (v_{0,1} + v_{1,1} + v_{2,1} + \cdots + v_{q-2,1})\alpha^1 \\ &\vdots \\ &+ (v_{0,q-2} + v_{1,q-2} + v_{2,q-2} + \cdots + v_{q-2,q-2})\alpha^{q-2},\end{aligned} \tag{19}$$

$$\begin{aligned}\mathbf{S}_4 = &(v_{0,0} + v_{1,0} + v_{2,0} + \cdots + v_{q-2,0})\alpha^0 \\ &+ (v_{0,1} + v_{1,1} + v_{2,1} + \cdots + v_{q-2,1})\alpha^2 \\ &\vdots \\ &+ (v_{0,q-2} + v_{1,q-2} + v_{2,q-2} + \cdots + v_{q-2,q-2})\alpha^{2(q-2)}.\end{aligned} \tag{20}$$

If we introduce:

$$\rho_i = \sum_{j=0}^{q-2} v_{i,j}, \quad i = 0, \ldots, q-2, \tag{21}$$

$$\kappa_i = \sum_{i=0}^{q-2} v_{i,j}, \quad i = 0, \ldots, q-2 \tag{22}$$

then:

$$\mathbf{S}_0 = \sum_{i=0}^{q-2} \kappa_i, \tag{23}$$

$$\mathbf{S}_1 = \sum_{i=0}^{q-2} \rho_i \alpha^i, \tag{24}$$

$$\mathbf{S}_2 = \sum_{i=0}^{q-2} \rho_i \alpha^{2i}, \tag{25}$$

$$\mathbf{S}_3 = \sum_{i=0}^{q-2} \kappa_i \alpha^j, \tag{26}$$

$$\mathbf{S}_4 = \sum_{i=0}^{q-2} \kappa_i \alpha^{2j}. \tag{27}$$

This notation allows us to illustrate the "block interleaved" structure of the code in Figure 2. In Figure 2 the received vector is stored in a 2-dimensional storage. It is obvious that ρ_i; $i = 0, \ldots, q-2$ and κ_j; $j = 0, \ldots, q-2$ could be obtained by adding the values in corresponding rows and columns from the obtained table.

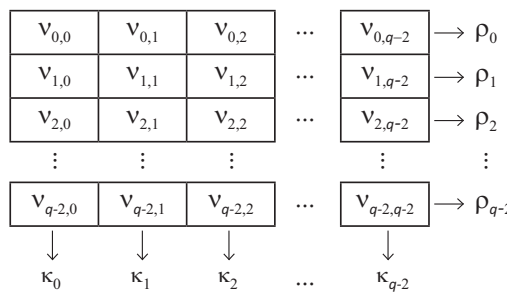

Figure 2. Received vector symbols stored in a 2-dimensional memory ("block interleaver")

In the second step the syndrome equations are formed. From (1), (13) and (14) it is obvious that

$$\mathbf{S} = \mathbf{e}.\mathbf{H}^T. \tag{28}$$

An appropriate subset of linearly independent equations has to be selected from which the estimations of erasure values are calculated. The following subsections analyse any possible combination of erasure location in the received vector for up to 3 erasures. In the following text, the erasures will be denoted as E_i, where i equals I, II, III.

4.1 One Erasure

If one erasure occurs, it is obvious that only one equation necessary for decoding always exists and can be obtained be multiplying $\mathbf{v}$ with the first row of $\mathbf{H}^T$. For example, an equation which corresponds to the first syndrome $\mathbf{S}_0$ could be used to compute the value of one erasure denoted as E_I. In this case from (13) and (15) it follows:

$$\mathbf{S}_0 = E_I. \tag{29}$$

4.2 Two Erasures

If two erasures occur in the received vector, a system of two linearly independent equations has to be formed. A more detailed analysis is required depending on the positions of the erasures.

Two erasures E_I and E_{II} will be in the positions which are inside the same matrix A (7) **or in two different matrices A localized in different columns.** In these cases, the two linearly independent equations that are necessary to calculate the values of the erasures could be formed using the first two syndromes.

$$\mathbf{S}_0 = E_I + E_{II}, \tag{30}$$

$$\mathbf{S}_1 = \alpha^{i_I} E_I + \alpha^{i_{II}} E_{II}. \tag{31}$$

where α^{i_I} and $\alpha^{i_{II}}$ are the known locators (since the positions of erasures are known) of the erasures inside the $\mathbf{A}$ matrices. In more detail, the first erasure is in the i_I^{th} column from the right and second erasure is in the i_{II}^{th} column from the right inside their $\mathbf{A}$ matrices.

In a practical decoding algorithm, in the first step the syndromes S_0 and S_1 are calculated using (23) and (24). Then E_I and E_{II} are calculated. From (30) and (31)

it follows:

$$E_I = \frac{S_1 + \alpha^{i_{II}} S_0}{(\alpha^{i_I} + \alpha^{i_{II}})}, \tag{32}$$

$$E_{II} = S_0 + \frac{S_1 + \alpha^{i_{II}} S_0}{(\alpha^{i_I} + \alpha^{i_{II}})}. \tag{33}$$

Two erasures E_I and E_{II} will be in the positions which are inside two different A matrices localized in the same columns of A or in the other words: $i_I = i_{II}$. In this case, the two linearly independent equations stem from multiplication of **v** with the first, fourth and the fifth (from top) transposed row of the **H** matrix (6).

Because the erasures are localized in two different matrices **A**, they are also located in different matrices **B**.

$$\mathbf{S}_0 = E_I + E_{II}, \tag{34}$$

$$\mathbf{S}_4 = \alpha^{2j_I} E_I + \alpha^{2j_{II}} E_{II}. \tag{35}$$

Using (30) and for example (34) we get

$$E_I = \frac{S_3 + \alpha^{j_{II}} S_0}{(\alpha^{j_I} + \alpha^{j_{II}})}, \tag{36}$$

$$E_{II} = S_0 + \frac{S_3 + \alpha^{j_{II}} S_0}{(\alpha^{j_I} + \alpha^{j_{II}})}. \tag{37}$$

4.3 Three Erasures

Three erasures E_I, E_{II} and E_{III} will be in the positions which are inside the same A matrix or in three different A matrices localized in three different columns. In this case, the first three syndromes could be exploited to form a system of 3 linearly independent equations necessary to compute E_I, E_{II}, and E_{III}. The syndromes S_0, S_1 and S_2 are calculated using (23), (24) and (25):

$$\mathbf{S}_0 = E_I + E_{II} + E_{III}, \tag{38}$$

$$\mathbf{S}_1 = \alpha^{i_I} E_I + \alpha^{i_{II}} E_{II} + \alpha^{i_{III}} E_{III}, \tag{39}$$

$$\mathbf{S}_2 = \alpha^{2i_I} E_I + \alpha^{2i_{II}} E_{II} + \alpha^{2i_{III}} E_{III}. \tag{40}$$

This system could be solved using the following determinants:

$$E_I = \frac{\begin{vmatrix} S_0 & 1 & 1 \\ S_1 & \alpha^{i_{II}} & \alpha^{i_{III}} \\ S_2 & \alpha^{2i_{II}} & \alpha^{2i_{III}} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ \alpha^{i_I} & \alpha^{i_{II}} & \alpha^{i_{III}} \\ \alpha^{2i_I} & \alpha^{2i_{II}} & \alpha^{2i_{III}} \end{vmatrix}}, \tag{41}$$

$$E_{II} = \frac{\begin{vmatrix} 1 & S_0 & 1 \\ \alpha^{i_I} & S_1 & \alpha^{i_{III}} \\ \alpha^{2i_{II}} & S_2 & \alpha^{2i_{III}} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ \alpha^{i_I} & \alpha^{i_{II}} & \alpha^{i_{III}} \\ \alpha^{2i_I} & \alpha^{2i_{II}} & \alpha^{2i_{III}} \end{vmatrix}}, \tag{42}$$

$$E_{III} = \frac{\begin{vmatrix} 1 & 1 & S_0 \\ \alpha^{i_I} & \alpha^{i_{II}} & S_1 \\ \alpha^{2i_I} & \alpha^{2i_{II}} & S_2 \end{vmatrix}}{\begin{vmatrix} 1 & 1 & 1 \\ \alpha^{i_I} & \alpha^{i_{II}} & \alpha^{i_{III}} \\ \alpha^{2i_I} & \alpha^{2i_{II}} & \alpha^{2i_{III}} \end{vmatrix}} \tag{43}$$

or by Gaussian elimination.

Three erasures E_I, E_{II} and E_{III} will be in the positions which are inside three different A matrices localized in the same columns. In this case, a system of 3 linearly independent equations necessary to compute E_I, E_{II}, and E_{III} can be formed using the first, fourth and fifth syndrome:

$$\mathbf{S}_0 = E_I + E_{II} + E_{III}, \tag{44}$$

$$\mathbf{S}_3 = \alpha^{j_I} E_I + \alpha^{j_{II}} E_{II} + \alpha^{j_{III}} E_{III}, \tag{45}$$

$$\mathbf{S}_4 = \alpha^{2j_I} E_I + \alpha^{2j_{II}} E_{II} + \alpha^{2j_{III}} E_{III} \tag{46}$$

This system could be solved using determinants or Gaussian elimination as in case 4.3.

Three erasures. Two erasures E_I, E_{II} are localized in the same column of A and the third erasure E_{III} is in a different column of A. This situation can occur in different sub cases.

Case A. Two erasures are localized within the same matrix **A** and the third one is in different matrix **A**. For example let: $i_I = i_{II} \neq i_{III}$, $j_I = j_{III} \neq j_{II}$, $i_I \neq j_{III}$.

In this case the decoder can form five equations for computing all five syndromes S_0, S_1, S_2, S_3, and S_4

$$\mathbf{S}_0 = E_I + E_{II} + E_{III}, \tag{47}$$

$$\mathbf{S}_1 = \alpha^{i_I} E_I + \alpha^{i_I} E_{II} + \alpha^{i_{III}} E_{III}, \tag{48}$$

$$\mathbf{S}_2 = \alpha^{2i_I} E_I + \alpha^{2i_I} E_{II} + \alpha^{2i_{III}} E_{III}, \tag{49}$$

$$\mathbf{S}_3 = \alpha^{j_I} E_I + \alpha^{j_{II}} E_{II} + \alpha^{j_I} E_{III}, \tag{50}$$

$$\mathbf{S}_4 = \alpha^{2j_I} E_I + \alpha^{2j_{II}} E_{II} + \alpha^{2j_I} E_{III}, \tag{51}$$

From (50) and (51) we get

$$E_{II} = \frac{\alpha^{j_I} S_4 + \alpha^{2j_I} S_3}{\alpha^{2j_I}\alpha^{j_{II}} + \alpha^{j_I}\alpha^{2j_{II}}} \tag{52}$$

and from (48) and (49) we get

$$E_{III} = \frac{\alpha^{i_I} S_2 + \alpha^{2i_I} S_1}{\alpha^{2i_I}\alpha^{i_{III}} + \alpha^{i_I}\alpha^{2i_{III}}} \tag{53}$$

and by using (47)

$$E_I = \mathbf{S}_0 + E_{II} + E_{III}. \tag{54}$$

Case B. Three erasures are localized in three different matrices **A** while two of them are localized in the same columns of two different **A** matrices. Then: $i_I = i_{II}, i_I \neq i_{III}, j_I \neq j_{II} \neq j_{III}$.

$$\mathbf{S}_0 = E_I + E_{II} + E_{III}, \tag{55}$$

$$\mathbf{S}_1 = \alpha^{i_I} E_I + \alpha^{i_I} E_{II} + \alpha^{i_{III}} E_{III}, \tag{56}$$

$$\mathbf{S}_2 = \alpha^{2i_I} E_I + \alpha^{2i_I} E_{II} + \alpha^{2i_{III}} E_{III}, \tag{57}$$

$$\mathbf{S}_3 = \alpha^{j_I} E_I + \alpha^{j_{II}} E_{II} + \alpha^{j_{III}} E_{III}, \tag{58}$$

$$\mathbf{S}_4 = \alpha^{2j_I} E_I + \alpha^{2j_{II}} E_{II} + \alpha^{2j_{III}} E_{III}. \tag{59}$$

In this case, (55), (58) and (59) could be solved using determinants or Gaussian elimination as in case 4.3.

The last step in the decoding algorithm is to correct the erasures in known positions, which is straightforward from (13).

The analysis of the decoding algorithm serves as a detailed proof of the previous theorem. Another confirmation that the presented family of codes has code distance 4 was obtained by calculating its weight spectra. An examples of the weight spectra of code constructed over $GF(4)$, $GF(8)$, and $GF(16)$ follows.

The weight spectrum of $[9, 4, 4]_{GF(4)}$ code: $a_0 = 1$, $a_4 = 27$, $a_6 = 54$, $a_7 = 108$, $a_8 = 54$, $a_9 = 12$.

The relevant part of the weight spectrum of $[49, 44, 4]_{GF(8)}$ code: $a_0 = 1$, $a_4 = 32\,585$, $a_5 = 806\,736$, $a_6 = 50\,853\,866$, $a_7, \ldots, a_{49} \neq 0$.

The relevant part of the weight spectrum of $[225, 220, 4]_{GF(16)}$ code: $a_0 = 1$, $a_4 = 10\,135\,125$, $a_5 = 3\,193\,835\,400$, $a_6 = 1\,834\,779\,161\,250$, $a_7, \ldots, a_{225} \neq 0$.

5 CONCLUSION

In this paper, a construction of new family of three erasure correcting linear block codes over $GF(q)$ together with their syndrome decoding procedures was presented. The designed code possesses code distance of four which enables correcting up to 3 erasures. This code distance was confirmed by weight spectra of their dual codes using Krawtchouck polynomials. These code could be potentially useful in DNA storage systems. However this will need further future research.

Acknowledgment

Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I05-03-V02-00051.

REFERENCES

[1] McAuley, A. J.: Weighted Sum Codes for Error Detection and Their Comparison with Existing Codes. IEEE/ACM Transactions on Networking, Vol. 2, 1994, No. 1, pp. 16–22, doi: 10.1109/90.282604.

[2] Farkaš, P.: Comments on "weighted Sum Codes for Error Detection and Their Comparison with Existing Codes". IEEE/ACM Transactions on Networking, Vol. 3, 1995, No. 2, pp. 222–223, doi: 10.1109/90.374122.

[3] Farkaš, P.—Baylis, J.: Modified Generalized Weighted Sum Codes for Error Control. Communication Theory and Applications, Vol. 4, 2000, pp. 62–72.

[4] Rakús, M.—Farkaš, P.: Double Error Correcting Codes with Improved Code Rates. Journal of Electrical Engineering, Vol. 55, 2004, No. 3-4, pp. 89–94.

[5] Rakús, M.—Farkaš, P.—Páleník, T.—Daniš, A.: Five Times Extended Reed-Solomon Codes Applicable in Memory Storage Systems. IEEE Letters of the Computer Society, Vol. 2, 2019, No. 2, pp. 9–11, doi: 10.1109/LOCS.2019.2911517.

[6] Farkaš, P.—Rakús, M.: Decoding Five Times Extended Reed Solomon Codes Using Syndromes. Computing and Informatics, Vol. 39, 2020, No. 6, pp. 1311–1335, doi: 10.31577/cai_2020_6_1311.

[7] Carmean, D.—Ceze, L.—Seelig, G.—Stewart, K.—Strauss, K.—Willsey, M.: DNA Data Storage and Hybrid Molecular-Electronic Computing. Proceedings of the IEEE, Vol. 107, 2019, No. 1, pp. 63–72, doi: 10.1109/JPROC.2018.2875386.

[8] Liu, Q.—Yang, K.—Xie, J.—Sun, Y.: DNA-Based Molecular Computing, Storage, and Communications. IEEE Internet of Things Journal, Vol. 9, 2022, No. 2, pp. 897–915, doi: 10.1109/JIOT.2021.3083663.

[9] Antkowiak, P. L.—Lietard, J.—Darestani, M. Z.—Somoza, M. M.—Stark, W. J.—Heckel, R.—Grass, R. N.: Low Cost DNA Data Storage Using Photolithographic Synthesis and Advanced Information Reconstruction and Error Correction. Nature Communications, Vol. 11, 2020, No. 1, Art. No. 5345, doi: 10.1038/s41467-020-19148-3.

[10] Meiser, L. C.—Antkowiak, P. L.—Koch, J.—Chen, W. D.—Kohll, A. X.—Stark, W. J.—Heckel, R.—Grass, R. N.: Reading and Writing Digital Data in DNA. Nature Protocols, Vol. 15, 2020, No. 1, pp. 86–101, doi: 10.1038/s41596-019-0244-5.

[11] Weber, J. H.—de Groot, J. A. M.—van Leeuwen, C. J.: On Single-Error-Detecting Codes for DNA-Based Data Storage. IEEE Communications Letters, Vol. 25, 2021, No. 1, pp. 41–44, doi: 10.1109/LCOMM.2020.3023826.

[12] Thanh Nguyen, T.—Cai, K.—Song, W.—Schouhamer Immink, K. A.: Optimal Single Chromosome-Inversion Correcting Codes for Data Storage in Live DNA. 2022 IEEE International Symposium on Information Theory (ISIT), 2022, pp. 1791–1796, doi: 10.1109/ISIT50566.2022.9834376.

[13] He, X.—Cai, K.: Basis-Finding Algorithm for Decoding Fountain Codes for DNA-Based Data Storage. IEEE Transactions on Information Theory, Vol. 69, 2023, No. 6, pp. 3691–3707, doi: 10.1109/TIT.2023.3241773.

[14] Levick, K.—Heckel, R.—Shomorony, I.: Achieving the Capacity of a DNA Storage Channel with Linear Coding Schemes. 2022 56$^{\text{th}}$ Annual Conference on Information Sciences and Systems (CISS), 2022, pp. 218–223, doi: 10.1109/CISS53076.2022.9751151.

[15] Grass, R. N.—Heckel, R.—Puddu, M.—Paunescu, D.—Stark, W. J.: Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes. Angewandte Chemie International Edition, Vol. 54, 2015, No. 8, pp. 2552–2555, doi: 10.1002/anie.201411378.

[16] Lu, X.—Kim, S.: Design of Nonbinary Error Correction Codes with a Maximum Run-Length Constraint to Correct a Single Insertion or Deletion Error for DNA Storage. IEEE Access, Vol. 9, 2021, pp. 135354–135363, doi: 10.1109/ACCESS.2021.3116245.

[17] Wang, S.—Sima, J.—Farnoud, F.: Non-Binary Codes for Correcting a Burst of at Most 2 Deletions. 2021 IEEE International Symposium on Information Theory (ISIT), 2021, pp. 2804–2809, doi: 10.1109/ISIT45174.2021.9517917.

[18] MinT, Dept. of Mathematics, University of Salzburg, `http://mint.sbg.ac.at/` (accessed in August 2023).

Peter Farkaš is with the Institute of Multimedia Information and Communication Technologies, Slovak University of Technology in Bratislava (STU) and also with the Institute of Applied Informatics, Faculty of Informatics, Pan European University in Bratislava as Professor. From 2002 until 2007 he was Visiting Professor at the Kingston University, UK and a senior researcher at SIEMENS PSE. In 2003 SIEMENS named him VIP for his innovations and patents. In 2004, he received the Werner von Siemens Excellence Award for his research on two-dimensional Complete Complementary Codes. From 2008 to 2009 he worked also as Consultant in the area of Software Defined Radio for SANDBRIDGE Tech. (USA). He was the leader of a team from STU in projects funded by the European Community under the 5FP and 6FP Information Society Technologies Programs: NEXWAY IST-2001-37944 (Network of Excellence in Wireless Applications and technology) and CRUISE (Creating Ubiquitous Intelligent Sensing Environments) FP6 IST-2005-4-027738, (2006–2007). His research interests include Coding, Communications Theory and sequences for CDMA. He has published 1 book, about 45 papers in reviewed scientific journals and about 100 papers in international conferences. He is currently serving on the TPC of approximately 60 international conferences and has presented 12 invited lectures. As an IEEE volunteer, he was serving in the IEEE Czechoslovakia Section Executive Committee in different positions from 1992 to 2014 and from 2005 to 2006 he served as a chair of the Conference Coordinator Subcommittee in IEEE Region 8. He organized the IEEE R8 Conference EUROCON 2001 and was a chairman of SympoTIC '03, SympoTIC '04, SympoTIC '06 and co-organizer of the Winter School on Coding and Information Theory 2005. Since 2016, he has been serving as a vice-chair of the Computer Chapter in the IEEE Czechoslovakia Section.

Katarína Farkašová has been with ESET s.r.o. as software tester since 2022, and as of 2025 with the Institute of Multimedia ICT at the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava as a researcher in the project Securing Data in Post-Quantum 6G Age: Advanced Coding Systems for Physical Layer Security, Distributed and DNA Storages, which is funded by European Union in NextGenerationEU framework. In 2016 she earned her Bachelor's degree in applied informatics. In 2019, she graduated with honors, obtaining her Master's degree in the same field. From 2019 to 2021 she worked as a software tester for VÚB Bank, Intesa Sanpaolo Group. From 2021 to 2022, she continued as IT tester and was leased to Shell Plc Corporation by Sanae Slovakia s.r.o. in order to work on international payment system for track drivers.

Frederik Pavelka is with the Institute of Multimedia ICT at the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava from 2025 as a Researcher in project Securing Data in Post-Quantum 6G Age: Advanced Coding Systems for Physical Layer Security, Distributed and DNA Storages, which is funded by the European Union in the NextGenerationEU framework. He earned his Bachelor's degree and Master's degree from the Pan-European University, Slovakia. On behalf of the Software foundation s.r.o. as a developer he developed hospital and ambulance information systems and general information systems for the Social Insurance and Medirex a.s.

Martin Rakús studied radio electronics and graduated at the Slovak University of Technology in 2001. In 2004 he received his Ph.D. from the Slovak University of Technology and in 2020 he became Associate Professor at the same institute. Since 1995 he has been with the Institute of Multimedia Information and Communication Technology, the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava, Slovakia. His primary research interests are error control coding and digital communication systems. He is a member of the IEEE.