21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

Decoding $\mathbb{F}_q$-linear codes over erasure channels*

Sara D. Cardell[1] and Joan-Josep Climent[1]

***Abstract*— We construct a new $\mathbb{F}_q$-linear code over $\mathbb{F}_q^b$ replacing the elements in the large field $\mathbb{F}_{q^b}$ by elements in $\mathbb{F}_q[C]$, where C is the companion matrix of a primitive polynomial of degree b. Then, we propose a decoding algorithm for the $\mathbb{F}_q$-linear code over the erasure channel, based on solving a linear system over $\mathbb{F}_q$.**

I. Introduction

An increasing number of applications in computer communications and storage systems uses erasure codes to deal with packet losses (see, for example, [1], [2]). Let us assume the data in a storage system is divided into packets of equal size. An (n, k) block code takes k information packets (symbols) and encodes them into a total of n packets (symbols) of the same size.

We consider as model of errors the erasure channel introduced by Elias [3]. Each codeword symbol is lost with a fixed independent probability. It is well-known that a code with minimum distance d can recover up to $d-1$ erasures. MDS codes are often the best adapted codes, since given a fixed redundancy, MDS codes offer maximal reliability, that is, any k packets are sufficient to recover all the information.

Reed Solomon codes [4], [5] are widely used in storage and communication applications. MDS array codes [6] are another class of MDS codes, for example the EVENODD [7]. In an array code, each packet consists of a column of elements, and the parities are computed by XORing some information bits. These codes have the advantage of low computation complexity over Reed Solomon codes because the encoding and decoding only involve XOR operations [8]. For these codes, operations can be implemented by high-speed hardware (see, for example, [9], [10]).

Since some Reed Solomon codes have high computational complexity, many research studies attempt to reduce the complexity of Reed Solomon codes [11], [12]. Here, we replace the elements in $\mathbb{F}_{q^b}$ by elements in $\mathbb{F}_q[C]$, where C is the companion matrix of a primitive polynomial of degree b and entries in $\mathbb{F}_q$ (see [13]).

In this work we consider block codes over $\mathbb{F}_q^b$, closed under addition and multiplication with elements in $\mathbb{F}_q$. These codes are called $\mathbb{F}_q$-linear codes [14]. These codes have practical applications in computer memory systems and deep-space communication [15], [16] and can be considered as array codes over the field $\mathbb{F}_q$ (see [17]).

*This work was partially supported by Spanish grant MTM2011-24858 of the Ministerio de Ciencia e Innovación of the Gobierno de España. The work of the first author was also supported by a grant for postdoctoral students from the Generalitat Valenciana with reference APOSTD/2013/081.

[1]Both authors are with the Departament d'Estadística i Investigació Operativa, Universitat d'Alacant, Ap. Correus 99, E-03080, Alacant, Spain s.diaz@ua.es, jcliment@ua.es

II. Preliminaries

Let $\mathbb{F}_q$ be the Galois field of q elements and consider b a positive integer. If $\mathcal{C}$ is a code of length n over $\mathbb{F}_q^b$, we can consider the codewords of $\mathcal{C}$ as codewords of length nb over $\mathbb{F}_q$. Then, a code $\mathcal{C}$ is said to be an $\mathbb{F}_q$**-linear code** of length n over $\mathbb{F}_q^b$ if it is a linear code of length nb over $\mathbb{F}_q$ (see [14]). The **dual code** $\mathcal{C}^\perp$ is the null space of $\mathcal{C}$ in $\mathbb{F}_q^b$. The parameter $k = \log_{q^b} |\mathcal{C}|$ is called the **normalized dimension** of $\mathcal{C}$ over $\mathbb{F}_q^b$. Recall that the linear space $\mathbb{F}_q^b$ can be viewed as the field $\mathbb{F}_{q^b}$; however, linearity is not assumed over the field $\mathbb{F}_{q^b}$.

The code $\mathcal{C}$ can be specified by either its parity-check matrix H of size $(n-k)b \times nb$ or its generator matrix G of size $kb \times nb$, both over $\mathbb{F}_q$. ¿From practical considerations, $\mathbb{F}_q$-linear codes are required to be systematic, that is, its parity-check (or generator) matrix has to be systematic. Recall that the matrix H (respectively, G) is said to be **systematic** if it contains the identity matrix of size $(n-k)b \times (n-k)b$ (respectively, $kb \times kb$).

To define the minimum distance of an $\mathbb{F}_q$-linear code we consider it as a code over the alphabet $\mathbb{F}_q^b$. Then, the minimum distance d is measured with respect to symbols of $\mathbb{F}_q^b$ (see [14]) and thus, the parameters of the code over $\mathbb{F}_q^b$ are $[n, k, d]$.

A code with parameters $[n, k, d]$ is MDS over $\mathbb{F}_q^b$ if the Singleton bound, $d \leq n - k + 1$, is attained (see [14], [18]).

Several constructions of MDS block codes based on superregular matrices have been proposed (see, for example, [19], [20]). A matrix A is said to be a **superregular matrix** if every square submatrix of A is nonsingular. As a generalization of this concept, we introduce the concept of block superregular matrix.

Definition 1: A matrix A of size $bm \times bl$ is said to be b**-block superregular matrix** over $\mathbb{F}_q$, if every square submatrix of A consisting of full blocks submatrices of size $b \times b$ is nonsingular over $\mathbb{F}_q$.

Due to the the previous definition and Proposition 3.2 of [14], we can introduce the following theorem. This theorem is useful to check whether an $\mathbb{F}_q$-linear code is MDS or not, without computing the minimum distance.

Theorem 1: *Let $H = \left(A \mid I_{(n-k)b}\right)$ be an $(n-k)b \times nb$ systematic parity-check matrix of an $\mathbb{F}_q$-linear code $\mathcal{C}$ with parameters $[n, k]$ over $\mathbb{F}_q^b$. Then $\mathcal{C}$ is an MDS $\mathbb{F}_q$-linear code if and only if A is a b-block superregular matrix over $\mathbb{F}_q$.*

ISBN: 978-90-367-6321-9

III. Construction

Consider the companion matrix C of a monic polynomial $p(\mathsf{x}) = p_0 + p_1\mathsf{x} + \cdots + p_{b-1}\mathsf{x}^{b-1} + \mathsf{x}^b \in \mathbb{F}_q[\mathsf{x}]$. If $p(\mathsf{x})$ is an irreducible polynomial, it is well known that $\mathbb{F}_{q^b} \approx \mathbb{F}_q[C]$ (see, for example, [21]). Moreover, if $p(\mathsf{x})$ is primitive, we have that $\mathbb{F}_q[C] = \left\{0, I, C, C^2, \ldots, C^{q^b-2}\right\}$.

The isomorphism $\psi : \mathbb{F}_{q^b} \to \mathbb{F}_q[C]$, can be defined as $\psi(\alpha) = C$, where $\alpha \in \mathbb{F}_{q^b}$ is a primitive element. It can be extended to a ring isomorphism

$$\Psi : \mathrm{Mat}_{m\times l}(\mathbb{F}_{q^b}) \longrightarrow \mathrm{Mat}_{m\times l}(\mathbb{F}_q[C]) \tag{1}$$

in the following way: if $A = (\alpha_{ij}) \in \mathrm{Mat}_{m\times l}(\mathbb{F}_{q^b})$, then $\Psi(A) = (\psi(\alpha_{ij})) \in \mathrm{Mat}_{m\times l}(\mathbb{F}_q[C])$. This isomorphism allows us to introduce the following result.

Theorem 2: *If $A \in \mathrm{Mat}_{m\times l}(\mathbb{F}_{q^b})$ is a superrregular matrix, then $H = (\Psi(A) \mid I_{bm})$ is the parity check-matrix of an $[m + l, l, m + 1]$ MDS $\mathbb{F}_q$-linear code $\mathcal{C}$ over $\mathbb{F}_q^b$.*

Next example helps us to understand this construction.

Example 1: Consider the companion matrix of the primitive polynomial $p(\mathsf{x}) = 1 + \mathsf{x}^2 + \mathsf{x}^3 \in \mathbb{F}_2[\mathsf{x}]$. Let $\alpha \in \mathbb{F}_{2^3}$ be a primitive element such that $1 + \alpha^2 + \alpha^3 = 0$. It is easy to check that $A = \begin{pmatrix} \alpha & \alpha^4 \\ 1 & \alpha^2 \end{pmatrix}$ is a superregular matrix over $\mathbb{F}_{2^3}$. So, according to Theorem 2, the matrix

$$H = \left(\begin{array}{cc|c} C & C^4 & \\ I_3 & C^2 & I_6 \end{array}\right)$$

is the parity-check matrix of an $\mathbb{F}_2$-linear code $\mathcal{C}$ over $\mathbb{F}_2^3$, whose parameters are $[4, 2, 3]$. Thus, the code is MDS.

If we consider the set of codewords of $\mathcal{C}$ as a linear code over $\mathbb{F}_2$, the code is not MDS, since only trivial codes are MDS over $\mathbb{F}_2$. In this case, the parameters of the code are $[12, 6, 3]$.

The dual code $\mathcal{C}^\perp$ is also an MDS $\mathbb{F}_2$-linear code over $\mathbb{F}_2^3$, whose parameters are $[4, 2, 3]$ as well and whose parity-check matrix is given by

$$G = H^\perp = \left(\begin{array}{c|cc} I_6 & C^T & I_3 \\ & (C^4)^T & (C^2)^T \end{array}\right). \qquad \square$$

The problem now is how to find superregular matrices over a finite field $\mathbb{F}_{q^b}$ (see, for example [19], [20], [22]).

IV. Decoding

Consider an MDS $\mathbb{F}_q$-linear code with length n and normalized dimension k over $\mathbb{F}_q^b$ and generator matrix G in systematic form. Assume we have k information symbols $\boldsymbol{u} = \left(\begin{array}{cccc} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots & \boldsymbol{u}_k \end{array}\right)$. The corresponding codeword is obtained via the generator matrix in the following way

$$\boldsymbol{v} = \boldsymbol{u}G = \left(\begin{array}{cccc|ccc} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_k & \boldsymbol{v}_{k+1} & \cdots & \boldsymbol{v}_n \end{array}\right).$$

Notice that $\boldsymbol{v}_i = \boldsymbol{u}_i$ for $i \in \{1, 2, \ldots, k\}$. Each transmitted symbol can be seen as a transmitted packet.

If $\mathcal{A}, \mathcal{B} \subseteq \{1, 2, \ldots, n\}$, we denote by $C_{\mathcal{A},\mathcal{B}}$ the submatrix of C with rows indexed by $\mathcal{A}$ and columns indexed by $\mathcal{B}$. Similarly, $\boldsymbol{v}_\mathcal{A}$ is the subvector of $\boldsymbol{v}$ with components indexed by $\mathcal{A}$. Moreover, given an integer t we define the set of indices $\mathcal{A} + t = \{a_i + t \mid a_i \in \mathcal{A}\}$.

Assume the codeword $\boldsymbol{v}$ is sent over an erasure channel and some symbols are lost. If t symbols are lost in the information part and t' are lost in the redundancy part, with $t \le k$ and $t' \le n - k$, the following algorithm helps us to recover the lost information symbols.

Algorithm 1:

1) *Let $\mathcal{I}$, with $|\mathcal{I}| = k - t$, be the set of the indices of the information symbols that have not been erased during transmission and $\mathcal{J}$, with $|\mathcal{J}| = n - k - t'$, the set of the indices of the redundancy symbols, starting by 1, that have not been erased during transmission. On the other hand, consider the set $\mathcal{I}^c$, with $|\mathcal{I}^c| = t$, as the set of the indices of the information symbols that have been erased.*
2) *If $t + t' > n - k$, it is not possible to recover the lost symbols. Go to step 7. Otherwise, go to next step.*
3) *Compute the matrices $C_{\mathcal{I},\mathcal{J}}$ and $C_{\mathcal{I}^c,\mathcal{J}}$.*
4) *Compute the vector $\boldsymbol{v}_{\mathcal{J}+k} - \boldsymbol{v}_\mathcal{I} \cdot C_{\mathcal{I},\mathcal{J}}$.*
5) *If $t + t' = n - k$, solve the system*

$$\boldsymbol{v}_{\mathcal{I}^c} \cdot C_{\mathcal{I}^c,\mathcal{J}} = \boldsymbol{v}_{\mathcal{J}+k} - \boldsymbol{v}_\mathcal{I} \cdot C_{\mathcal{I},\mathcal{J}} \tag{2}$$

   for $\boldsymbol{v}_{\mathcal{I}^c}$, and go to step 7. Otherwise, go to next step.
6) *In this case $t + t' < n - k$. Solve the system*

$$\boldsymbol{v}_{\mathcal{I}^c}\tilde{C} = \tilde{\boldsymbol{v}}, \tag{3}$$

   for $\boldsymbol{v}_{\mathcal{I}^c}$, where $\tilde{C}$ is the matrix composed by the first $t'b$ columns of $C_{\mathcal{I}^c,\mathcal{J}}$, and $\tilde{\boldsymbol{v}}$ is the vector formed by the first t' symbols of $\boldsymbol{v}_{\mathcal{J}+k} - \boldsymbol{v}_\mathcal{I} \cdot C_{\mathcal{I},\mathcal{J}}$. Go to step 7.
7) *End.*

The vector $\boldsymbol{v}_{\mathcal{I}^c}$, with length t, is the vector of lost information symbols. Once we compute $\boldsymbol{v}_{\mathcal{I}^c}$, we can obtain the complete vector $\boldsymbol{v}$ and thus, the information vector $\boldsymbol{u}$.

Next example helps us to understand the algorithm.

Example 2: Consider the MDS $\mathbb{F}_2$-linear code with parameters $[4, 2]$ over $\mathbb{F}_2^2$ whose generator matrix is given by

$$G = \left(\begin{array}{c|cc} I_4 & C_{11} & C_{12} \\ & C_{21} & C_{22} \end{array}\right) = \left(\begin{array}{cccc|cc|cc} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \end{array}\right).$$

Assume we receive the following word

$$\boldsymbol{v} = \left(\begin{array}{cc|cc} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \boldsymbol{v}_3 & \boldsymbol{v}_4 \end{array}\right) = \left(\begin{array}{cc|cc} 1\,0 & ?\,? & ?\,? & 0\,0 \end{array}\right).$$

We know symbols $\boldsymbol{v}_1$ and $\boldsymbol{v}_4$ and we have to recover at least $\boldsymbol{v}_2$. First of all, we construct the sets $\mathcal{I} = \{1\}$, $\mathcal{I}^c = \{2\}$ and $\mathcal{J} = \{2\}$. In this case, we have $t + t' = n - k = 2$ erasures so, according to step 5 in Algorithm 1, we consider the system in expression (2), that is,

$$\boldsymbol{v}_2 \cdot C_{2,2} = \boldsymbol{v}_4 - \boldsymbol{v}_1 \cdot C_{1,2}.$$

Solving the system for $\boldsymbol{v}_2$, we can complete the information symbols of the codeword, $\left(\begin{array}{cc} \boldsymbol{v}_1 & \boldsymbol{v}_2 \end{array}\right) = \left(\begin{array}{cc} 1\,0 & 1\,0 \end{array}\right)$.

Assume now, that we receive the word

$$\begin{aligned} \boldsymbol{w} &= \left(\begin{array}{cc|cc} \boldsymbol{w}_1 & \boldsymbol{w}_2 & \boldsymbol{w}_3 & \boldsymbol{w}_4 \end{array}\right) \\ &= \left(\begin{array}{cc|cc} ?\,? & 1\,0 & 0\,1 & 0\,0 \end{array}\right). \end{aligned}$$

In this case, we know symbols $\boldsymbol{w}_2$, $\boldsymbol{w}_3$ and $\boldsymbol{w}_4$, and we have to recover $\boldsymbol{w}_1$. It is easy to obtain the following sets $\mathcal{I} = \{2\}$, $\mathcal{I}^c = \{1\}$ and $\mathcal{J} = \{1, 2\}$. Here, we have $t + t' = 1 < n - k = 2$ erasures, so we consider step 6 in Algorithm 1 and solve the system in expression (3), that is,

$$\boldsymbol{w}_1 \tilde{C} = \tilde{\boldsymbol{w}},$$

where $\tilde{C}$ is the matrix composed by the first 2 columns of $\left(\begin{array}{cc} C_{11} & C_{12} \end{array}\right)$, that is, C_{11}, and $\tilde{\boldsymbol{w}}$ is the vector formed by the first symbol of $\left(\begin{array}{cc} \boldsymbol{w}_3 & \boldsymbol{w}_4 \end{array}\right) - \boldsymbol{w}_2 \left(\begin{array}{cc} C_{21} & C_{22} \end{array}\right)$, that is, $\boldsymbol{w}_3 - \boldsymbol{w}_2 C_{21}$. Then, solving the system

$$\boldsymbol{w}_1 C_{11} = \boldsymbol{w}_3 - \boldsymbol{w}_2 C_{21}$$

for $\boldsymbol{w}_1$, we can complete the information part of the codeword, $\left(\begin{array}{cc} \boldsymbol{w}_1 & \boldsymbol{w}_2 \end{array}\right) = \left(\begin{array}{cc} 1\,0 & 1\,0 \end{array}\right)$. □

V. Conclusions

We introduce a construction of MDS $\mathbb{F}_q$-linear codes over $\mathbb{F}_q^b$ based on the isomorphism between fields $\mathbb{F}_q[C]$ and $\mathbb{F}_{q^b}$, where C is the companion matrix of a primitive polynomial of degree b in $\mathbb{F}_q[\mathrm{x}]$. Given a superregular matrix over $\mathbb{F}_{q^b}$, we replace each element in the matrix by a power of C. Transmitting n packets of length b with elements in $\mathbb{F}_q$ through an erasure channel is the same as transmitting a codeword of length n of an $\mathbb{F}_q$-linear code over $\mathbb{F}_q^b$. Since the codes we construct are MDS, we can recover up to $n - k$ erasures, with k the normalized dimension of the code. Then, we propose an algorithm to recover the lost information symbols just solving a linear system with tb unknowns, where t in the number of unknown information symbols.

References

[1] Y. M. Chee, C. J. Colbourn, and A. C. H. Ling, "Asymptotically optimal erasure-resilient codes for large disk arrays," *Discrete Applied Mathematics*, vol. 102, pp. 3–36, 2000.

[2] M. A. Kousa and A. H. Mugaibel, "Cell loss recovery using two-dimensional erasure correction for ATM networks," in *Proceedings of the Seventh International Conference on Telecommunication Systems*, Mar. 1999, pp. 85–89.

[3] P. Elias, "Coding for noisy channels," in *IRE International Convention Record, pt. 4*, 1955, pp. 37–46.

[4] S. Roman, *Coding and Information Theory*. New York, NY: Springer, 1992.

[5] P. Sweeney, *Error Correcting Coding*. West Sussex, England: John Wiley & Sons, 2002.

[6] M. Blaum, P. G. Farrell, and H. C. A. van Tilborg, "Array codes," in *Handbook of Coding Theory*, V. S. Pless and W. C. Huffman, Eds. North-Holland: Elsevier, 1998, pp. 1855–1909.

[7] M. Blaum, J. Brady, J. Bruck, and J. Menon, "EVENODD: An efficient scheme for tolerating double disk failures in RAID architectures," *IEEE Transactions on Computers*, vol. 42, no. 2, pp. 192–202, 1995.

[8] Z. Wang, A. G. Dimakis, and B. Jehoshua, "Rebuilding for array codes in distributed storage systems," in *Proceedings of Workshop on the Application of Communication Theory to Emerging Memory Technologies (ACTEMT)*, 2012.

[9] M. Blaum and R. M. Roth, "New array codes for multiple phased burst correction," *IEEE Transactions on Information Theory*, vol. 39, no. 1, pp. 66–77, 1993.

[10] M. Blaum, J. Bruck, and A. Vardy, "MDS array codes with independent parity symbols," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 529–542, 1996.

[11] J. Blömer, M. Kalfane, R. Karp, M. Karpinski, M. Luby, and D. Zuckerman, "An XOR-based erasure-resilient coding scheme," 1995.

[12] J. Lacan and J. Fimes, "Systematic MDS erasure codes based on Vandermonde matrices," *IEEE Communications Letters*, vol. 8, no. 9, pp. 570–572, 2004.

[13] S. D. Cardell, "Constructions of MDS codes over extension alphabets," Ph.D. dissertation, Departamento de Ciencia de la Computación e Inteligencia Artificial, Universidad de Alicante, Alicante, España, Aug. 2012.

[14] M. Blaum and R. M. Roth, "On lowest density MDS codes," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 46–59, 1999.

[15] C. L. Chen, "Byte-oriented error-correcting codes for semiconductor memory systems," *IEEE Transactions on Computers*, pp. 646–648, 1986.

[16] C. Chen, Fellow, IEEE, and B. W. Curran, "Switching codes for delta-I noise reduction," *IEEE Transactions on Computers*, vol. 45, no. 9, pp. 1017–1021, 1996.

[17] M. Blaum, J. L. Fan, and L. Xu, "Soft decoding of several classes of array codes," 2002, extended version of [23].

[18] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, 6th ed. Amsterdam: North-Holland, 1988.

[19] R. M. Roth and G. Seroussi, "On generator matrices of MDS codes," *IEEE Transactions on Information Theory*, vol. 31, no. 6, pp. 826–830, 1985.

[20] R. M. Roth and A. Lempel, "On MDS codes via Cauchy matrices," *IEEE Transactions on Information Theory*, vol. 35, no. 6, pp. 1314–1319, 1989.

[21] R. Lidl and H. Niederreiter, *Introduction to Finite Fields and Their Applications*. New York, NY: Cambridge University Press, 1994.

[22] J. Lacan and J. Fimes, "A construction of matrices with no singular square submatrices," in *Finite Fields and Applications*, ser. Lecture Notes in Computer Science, G. L. Mullen, A. Poli, and H. Stichtenoth, Eds. Berlin: Springer-Verlag, 2003, vol. 2948, pp. 145–147.

[23] M. Blaum, J. L. Fan, and L. Xu, "Soft decoding of several classes of array codes," in *Proceedings of the 2002 IEEE International Symposium on Information Theory (ISIT 2002)*. Lausanne, Switzerland: IEEE, Jul. 2002, p. 368.