Hybrid Elementary Linear Subspace codes

Ermes Franch, Chunlei Li

University of Bergen
ermes.franch@uib.no, chunlei.li@uib.no

Abstract. In this paper we introduce a new family of rank metric codes, which is essentially a random subcode of elementary linear subspace (ELS). The proposed codes allow for a polynomial-time probabilistic decoding algorithm with low failure probability.

1 Introduction

Rank metric codes, introduced independently by Delsarte [3], Gabidulin [4] and Roth [13], have found important applications in cryptography and network coding [5,6,11,14]. The application of rank metric codes typically requires efficient decoding of the codes in use. So far there have been a few families of rank metric codes that can be efficiently decoded in polynomial time, such as the Gabidulin codes [4] and the low rank parity-check (LRPC) codes [12]. This paper introduces a new family of random-like rank metric codes, named Hybrid ELS codes, that are randomly generated and allow for a probabilistic decoding algorithm with polynomial-time complexity. The decoding algorithm adopts a similar idea used for the LRPC codes but exhibits better performance in terms of error correction capability at a given rate. Decoding failure of the proposed codes is also investigated.

The paper will be structured as follows. In Section 2 we present basic notation and a brief introduction to rank-metric codes. Section 3 briefly recalls the decoding approach for LRPC codes [12] and some properties of the Elementary Linear Subspace (ELS) [10], which are relevant to the definition and the decoding algorithm of the Hybrid ELS codes. In Section 4 we propose the new family of Hybrid ELS codes, and discuss the upper bound on the rank weights of codewords in the Hybrid ELS codes and the upper bound on the minimum rank distance of Hybrid ELS codes. Section 5 presents a probabilistic polynomial decoding algorithm for Hybrid ELS codes, and also analyzes possible failures in the two major decoding steps.

2 Preliminaries

Let $\mathbb{F}_q$ be the finite field of q elements where q is a prime power. Elements of a finite field will be denoted by lower case letters. Given a vector $\mathbf{v}$, its i-th component is denoted by v_i. Matrices will be denoted by upper case letters and the transpose of a matrix M will be indicated by $M^\intercal$. We denote by $\mathbb{F}_q^{k\times n}$ the

set of $k \times n$ matrices over $\mathbb{F}_q$. An $\mathbb{F}_{q^m}$-linear subspace of $\mathbb{F}_{q^m}^n$ will be denoted by a calligraphic letter; and an $\mathbb{F}_q$-linear subspace of $\mathbb{F}_{q^m}$ will be denoted by a fraktur letter. The number of possible different bases of a t-dimensional subspace of $\mathbb{F}_q^m$ is given by

$$A_q(m,t) = \prod_{i=0}^{t-1}(q^m - q^i). \tag{1}$$

Then the Gaussian binomial is given by $\begin{bmatrix} m \\ k \end{bmatrix}_q = \frac{A_q(m,k)}{A_q(k,k)}$.

Definition 1. *For a subset $S = \{s_1, \ldots, s_t\} \subseteq \mathbb{F}_{q^m}$, its **support** is defined as the $\mathbb{F}_q$-linear space of $\mathbb{F}_{q^m}$ generated by the elements of S, i.e.,*

$$\langle S \rangle_{\mathbb{F}_q} = \langle s_1, \ldots, s_t \rangle_{\mathbb{F}_q}.$$

Similarly, the support of a vector $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{F}_{q^m}^n$ is defined as $\langle \mathbf{v} \rangle_{\mathbb{F}_q} = \langle v_1, \ldots, v_n \rangle_{\mathbb{F}_q}$ and the support of a matrix $H \in \mathbb{F}_{q^m}^{k \times n}$ is defined as $\langle H \rangle_{\mathbb{F}_q} = \langle H_{i,j} \rangle_{\mathbb{F}_q}$, namely, the $\mathbb{F}_q$-linear space generated by all the entries of H.

Using the notion of support we can equip $\mathbb{F}_{q^m}^n$ with the rank metric.

Definition 2. *The **rank weight** of a vector $\mathbf{v} \in \mathbb{F}_{q^m}^n$ is given by*

$$\mathrm{rw}(\mathbf{v}) = \dim(\langle \mathbf{v} \rangle_{\mathbb{F}_q}).$$

*We can define the **rank distance** between $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{F}_{q^m}^n$ as*

$$\mathrm{rd}(\mathbf{v}, \mathbf{w}) = \mathrm{rw}(\mathbf{v} - \mathbf{w}) = \dim(\langle \mathbf{v} - \mathbf{w} \rangle_{\mathbb{F}_q}).$$

*A code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ is said to be a **rank-metric code** when we consider it as a subset of $\mathbb{F}_{q^m}^n$ equipped with the rank distance. When $\mathcal{C}$ is a linear subspace of $\mathbb{F}_{q^m}^n$, it is called an $\mathbb{F}_{q^m}$-linear rank metric code.*

An $\mathbb{F}_{q^m}$-linear code can be also given in terms of a generator matrix or a parity-check matrix.

Definition 3. *Let $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ be an $\mathbb{F}_{q^m}$-linear code of dimension k. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be a group of $\mathbb{F}_{q^m}$-linearly independent vectors in $\mathcal{C}$. A matrix G having $\mathbf{v}_i$'s as its row vectors is a **generator matrix** of $\mathcal{C}$. Denote by $\mathcal{C}^\perp$ the linear space of dimension $n-k$ of the codewords orthogonal with $\mathcal{C}$, that is $\mathcal{C}^\perp = \{\mathbf{v} \in \mathbb{F}_{q^m}^n | \mathbf{v} \cdot \mathbf{c} = 0,\ \forall \mathbf{c} \in \mathcal{C}\}$. A generator matrix H of $\mathcal{C}^\perp$ is a **parity-check matrix** of $\mathcal{C}$.*

From the above definition it's clear that any element $\mathbf{c} \in \mathcal{C}$ can be uniquely obtained as $\mathbf{c} = (x_1, \ldots, x_k)G$ for some $(x_1, \ldots, x_k) \in \mathbb{F}_{q^m}^k$. In addition, given a parity-check matrix H of a linear code $\mathcal{C}$, for any $\mathbf{v} \in \mathbb{F}_{q^m}^n$, one has $\mathbf{v}H^\intercal = 0$ iff $\mathbf{v} \in \mathcal{C}$. Given a vector $\mathbf{v} \in \mathbb{F}_{q^m}^n$, *the syndrome of* $\mathbf{v}$ is the $(n-k)$-dimensional vector given by $\mathbf{s} = \mathbf{v}H^\intercal$. An important problem in coding theory is the syndrome decoding problem. In the context of rank metric, we recall the rank syndrome decoding (RSD) problem below.

Definition 4 (RSD). *Given an $(n-k)\times n$ parity-check matrix H over $\mathbb{F}_{q^m}$, a syndrome $\mathbf{s}$ and a positive integer r, the* ***rank syndrome decoding (RSD)*** *problem is to find $\mathbf{e}\in\mathbb{F}_{q^m}^n$ such that $\mathbf{e}H^\intercal=\mathbf{s}$ and $\mathrm{rw}(\mathbf{e})\leq r$.*

Recently the complexity of the above RSD problem was comprehensively investigated in [7]. Later Gaborit and Zémor proved the hardness of the RSD problem under unfaithful randomized reductions [9]. With the confidence on the hardness of the problem, several cryptosystems based on rank-metric codes have been proposed [5,12,8,6,2,11].

3 LRPC codes and Elementary Linear Subspace

3.1 LRPC codes

In 2013 Gaborit, Murat, Ruatta and Zémor introduced the family of LRPC codes that allows for a probabilistic decoding algorithm with polynomial-time complexity [12].

Definition 5. *Let H be an $(n-k)\times n$ parity-check matrix over $\mathbb{F}_{q^m}$. The* ***density*** *of H is defined as $d=\dim(\langle H\rangle_{\mathbb{F}_q})$. A code $\mathcal{C}$ having a parity-check matrix H of low density d ($\ll m$) is called an* ***LRPC code*** *of density d.*

The decoding algorithm of LRPC codes [12,2] starts with the observation that, if an error $\mathbf{e}$ of rank r occurs, the components of its syndrome have to lie in a subspace of dimension at most rd. This subspace is obtained as $\langle\mathfrak{E}\mathfrak{H}\rangle_{\mathbb{F}_q}$ where $\mathfrak{E}=\langle\mathbf{e}\rangle_{\mathbb{F}_q}$ is the error support, $\mathfrak{H}=\langle H\rangle_{\mathbb{F}_q}$ is the support of the parity-check matrix and $\mathfrak{E}\mathfrak{H}=\{eh|\quad e\in\mathfrak{E},h\in\mathfrak{H}\}$. When this product subspace has dimension no greater than $n-k$, with a high probability one can treat $\langle\mathfrak{E}\mathfrak{H}\rangle_{\mathbb{F}_q}$ as $\langle\mathbf{s}\rangle_{\mathbb{F}_q}$, the $\mathbb{F}_q$-linear span of the $n-k$ elements given by the syndrome $\mathbf{s}$. With $\mathfrak{H}$ and $\langle\mathbf{s}\rangle_{\mathbb{F}_q}$ one can recover the error support $\mathfrak{E}=\langle\varepsilon_1,\ldots,\varepsilon_r\rangle_{\mathbb{F}_q}$ with a good probability.

Once the error support is obtained, one can represent the error vector as $\mathbf{e}=(\varepsilon_1,\ldots,\varepsilon_r)X$, where $X\in\mathbb{F}_q^{r\times n}$ is the matrix obtained from the coordinates of $\mathbf{e}$ with respect to the basis $\varepsilon_1,\ldots,\varepsilon_r$. In order to recover the nr unknowns in X, one expands the $\mathbb{F}_{q^m}$-linear syndrome equations $\mathbf{s}=\mathbf{e}H^\intercal$ over $\mathbb{F}_q$. This expansion gives $(n-k)dr$ equations in nr variables over $\mathbb{F}_q$. To summarize, in order to recover the error support, one needs $d\leq\frac{n-k}{r}$, and in order to have a uniquely solvable system, one needs $d\geq\frac{n}{n-k}$. That is to say, in order to allow for an efficient probabilistic decoding algorithm, an LRPC code of length n over $\mathbb{F}_{q^m}$ should have its density in the following range:

$$1<\frac{n}{n-k}\leq d\leq\frac{n-k}{r}. \tag{2}$$

With the random-like behavior and efficient decoding algorithm, LRPC codes have been applied in several cryptosystems in recent years [12,8,6,11].

3.2 Elementary Linear Subspace

The family of *elementary linear subspaces* (ELS) [10] will be an essential building block for the proposed codes in this paper. Here we recall the definition of ELS and briefly summarize some of its properties.

Definition 6. *[10] A linear subspace of $\mathbb{F}_{q^m}^n$ is said to be **elementary** if it has a basis B consisting of n-dimensional row vectors over $\mathbb{F}_q$. We denote by $E_k(q^m, n)$ the set of all elementary linear subspaces with dimension k in $\mathbb{F}_{q^m}^n$.*

Similarly, we say an $[n, k]$ code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ is **elementary** if it can generated by a $k \times n$ matrix G over the base field $\mathbb{F}_q \subseteq \mathbb{F}_{q^m}$. That is to say, an $[n, k]$ code in $\mathbb{F}_{q^m}^n$ is elementary if an only if $\mathcal{C} \in E_k(q^m, n)$. The rank weight of an ELS code is limited by the dimension of the code.

Proposition 1. *[10, Prop. 2] For $\mathcal{C} \in E_k(q^m, n)$, all codewords of $\mathcal{C}$ have rank weight upper bounded by k. Moreover if $k \leq m$ the upper bound is tight.*

Note that in ELS codes, a codeword $\mathbf{v} = \mathbf{m}G$ has rank equal to $\mathrm{rw}(\mathbf{m})$ since G has full rank over $\mathbb{F}_q$. Therefore there are in total $A_q(m, r) * A_q(n, r)/A_q(r, r)$ codewords of rank r. Recall that LRPC codes are defined in terms of parity-check matrices with low density. An interesting connection between $LRPC$ codes and ELS codes is given below.

Proposition 2. *An LRPC code of density one is an ELS code. In other words, $\mathcal{C} \in E_k(q^m, n)$ iff $\mathcal{C}^\perp \in E_{n-k}(q^m, n)$.*

Proof. Let $\mathcal{C}$ be an $[n, k]_{q^m}$ LRPC code of density 1. We need to show that there exists a generator matrix $G \in \mathbb{F}_q^{k \times n}$ of $\mathcal{C}$. Let H be a parity-check matrix of $\mathcal{C}$ of density 1 and rank $n - k$. Since H has density 1, we have $\langle H \rangle_{\mathbb{F}_q} = \langle \mu \rangle_{\mathbb{F}_q}$ where $\mu \in \mathbb{F}_{q^m}^*$. Hence the matrix $H' = \mu^{-1} H \in \mathbb{F}_q^{(n-k) \times n}$ is another parity-check matrix of density 1 of $\mathcal{C}$.

Let $\mathcal{H}$ be the row-span of H' inside $\mathbb{F}_q^n$. Then its orthogonal space $\mathcal{H}^\perp \subseteq \mathbb{F}_q^n$ has dimension k. Take a basis of this space and build a matrix G of which each row is an element of this basis. Then G is a $k \times n$ matrix over $\mathbb{F}_q$ of rank k. Clearly each row g of G is in $\mathcal{C}$ since $H' g^\intercal = 0$. Thus $\mathrm{rowspan}_{\mathbb{F}_{q^m}}(G) \subseteq \mathcal{C}$. For dimension reasons we can also see that $\mathcal{C} \subseteq \mathrm{rowspan}_{\mathbb{F}_{q^m}}(G)$. Hence G is actually a generator matrix of $\mathcal{C}$. The other implication follows in a similar way. □

From Proposition 2 we can see that ELS and LRPC codes of density 1 are the same set. Although ELS codes are a special case of LRPC codes, they are not decodable because $d = 1 < \frac{n}{n-k}$ for any nontrivial choice of k.

4 The Hybrid ELS codes

Suppose $\mathcal{C}$ is an ELS code of dimension k and an error $\mathbf{e}$ of rank r occurs. Since its parity-check matrix H is in $\mathbb{F}_q^{(n-k) \times n}$, each component of $\mathbf{s} = \mathbf{e}H^\intercal$ belongs

directly to the error support. This implies $\langle \mathbf{s} \rangle_{\mathbb{F}_q} \subseteq \langle \mathbf{e} \rangle_{\mathbb{F}_q}$ where equality holds with a good probability when $n - k > r$. When expanding the system $\mathbf{e}H^\intercal = \mathbf{s}$ over $\mathbb{F}_q$, since the ranks of $\mathbf{e}$ and $\mathbf{s}$ both equal r, one obtains only $(n-k)r$ linearly independent equations in nr variables. Such a linear system has kr free variables over $\mathbb{F}_q$ and therefore q^{kr} possible solutions.

In the RSD problem the difficult part usually is to efficiently recover the error support. With ELS this task becomes trivial. Nevertheless, the price to pay is that the under-determinedness of the $\mathbb{F}_q$-linear system makes this recovery almost useless.

Next we propose a new family of rank metric codes that allow for efficient decoding. The essential idea is to add extra rows of maximal density to the parity-check matrix such that the number of linearly independent equations in the expanded system can be increased.

Definition 7. *Suppose* $H = \begin{pmatrix} B \\ R \end{pmatrix}$ *is a full-rank matrix, where* $B \in \mathbb{F}_q^{l \times n}$ *and* $R \in \mathbb{F}_{q^m}^{t \times n}$ *has density* $\min\{n, m\}$*. A rank-metric code* $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ *having* H *as a parity-check matrix is called a* ***Hybrid ELS*** *code.*

Let M be a matrix and $\mathcal{C}_M$ denote the linear code over $\mathbb{F}_{q^m}$ that admits M as its parity-check matrix. Notice that a Hybrid ELS code $\mathcal{C}_H$ having a parity-check matrix H as in Definition 7 is the intersection of $\mathcal{C}_B$ and $\mathcal{C}_R$, where $\mathcal{C}_B$ and $\mathcal{C}_R$ are the codes having B and R as their parity-check matrices, respectively. In particular we have $\mathcal{C}_H \subset \mathcal{C}_B$ where $\mathcal{C}_B$ is an ELS code. As we discussed, an ELS code contains a large number of codewords of small rank weight and has minimum rank distance 1. When the additional conditions represented by the random submatrix R are added to the parity-check matrix H, they filter out most of the codewords with small rank weights in $\mathcal{C}_B$.

Remark 1. Assume $\mathcal{C}_B$ is an ELS code of dimension $n - l$. By Proposition 1 the codewords of $\mathcal{C}_B$ have rank weight upper bounded by $n - l$. Since $\mathcal{C}_H \subseteq \mathcal{C}_B$, the same upper bound also applies to the codewords of $\mathcal{C}_H$.

The row-reduced echelon form of the generator matrix G of $\mathcal{C}_H$ is considered in the following lemma.

Lemma 1. *Let* H *be an* $(l + t) \times n$ *matrix of the form* $\begin{pmatrix} B \\ R \end{pmatrix}$ *with* $B \in \mathbb{F}_q^{l \times n}$ *of rank* l *and* $R \in \mathbb{F}_{q^m}^{t \times n}$ *of rank* t*. Then it can be be reduced to the form*

$$H' = THP = \left(\begin{array}{c|c|c} I_l & 0 & B_k - B_t R_k \\ \hline 0 & I_t & R_k \end{array}\right),$$

where $T \in \mathbb{F}_{q^m}^{(l+t)\times(l+t)}$ *is an invertible matrix,* $P \in \mathbb{F}_q^{n \times n}$ *is a permutation matrix,* $B_t \in \mathbb{F}_q^{l \times t}, B_k \in \mathbb{F}_q^{l \times k}$ *and* $R_k \in \mathbb{F}_{q^m}^{t \times k}$*.*

Proof. For the matrix H, we can apply row operations on B and R separately. Hence the row-reduced echelon form of B will still be a matrix over $\mathbb{F}_q$. Applying

a column permutation on the reduced form of B we will obtain a matrix in the form $\left(I_l|B_t|B_k\right)$. With the identity matrix I_l we can reduce to 0 the first l columns of R. We can then further reduce what is left to its row-reduced echelon form. Putting reduced B and reduced R together, up to permutations of the columns, we a reduced matrix of H with the following form

$$\left(\begin{array}{c|c|c} I_l & B_t & B_k \\ \hline 0 & I_t & R_k \end{array}\right).$$

By further applying row operations on the above matrix, one can easily obtain the desired matrix H'. □

With the transformed parity-check matrix in Lemma 1, we discuss the rank of codewords in Hybrid ELS codes below.

Proposition 3. *Let $\mathcal{C}_H$ be a Hybrid ELS code defined by a parity-check matrix $H = \begin{pmatrix} B \\ R \end{pmatrix}$. Then $\mathcal{C}_H$ is equivalent to a code $\mathcal{C}'$ that is defined by the following generator matrix*

$$G = \left(B_k^{\intercal} - R_k^{\intercal}B_t^{\intercal} \,\middle|\, R_k^{\intercal} \,\middle|\, I_k\right),$$

where matrices B_k, R_k, B_t, I_k are as given in Lemma 1. Moreover, given a nonzero message $\mathbf{m} \in \mathbb{F}_{q^m}^k$, the rank of its corresponding codeword satisfies

$$\mathrm{rw}(\mathbf{m}G) = \mathrm{rw}(\mathbf{m}) + \dim(\langle \mathbf{m}R_k^{\intercal}\rangle_{\mathbb{F}_q}) - \dim(\langle \mathbf{m}\rangle_{\mathbb{F}_q} \cap \langle \mathbf{m}R_k^{\intercal}\rangle_{\mathbb{F}_q}) \leq \mathrm{rw}(\mathbf{m}) + t. \quad (3)$$

Proof. The fact that G is a generator matrix of $\mathcal{C}_H$ is a direct consequence of Lemma 1. A generic codeword $\mathbf{v} \in \mathcal{C}'$ can be written as

$$\mathbf{v} = \mathbf{m}G = \left(\mathbf{m}B_k^{\intercal} - \mathbf{m}R_k^{\intercal}B_t^{\intercal} \,\middle|\, \mathbf{m}R_k^{\intercal} \,\middle|\, \mathbf{m}\right),$$

where $\mathbf{m} = (m_1, \ldots, m_k) \in \mathbb{F}_{q^m}^k$. Observe that since B_k, B_t are two matrices over $\mathbb{F}_q$ they will not expand the support of $\mathbf{m}$. Thus,

$$\langle \mathbf{m}B_k^{\intercal} - \mathbf{m}R_k^{\intercal}B_t^{\intercal}\rangle_{\mathbb{F}_q} \subseteq \langle \mathbf{m}R_k^{\intercal}\rangle_{\mathbb{F}_q}.$$

Hence the rank of $\mathbf{v}$ can be given as

$$\begin{aligned} \mathrm{rw}(\mathbf{v}) &= \dim(\langle \mathbf{v}\rangle_{\mathbb{F}_q}) = \dim(\langle \mathbf{m}\rangle_{\mathbb{F}_q} + \langle \mathbf{m}R_k^{\intercal}\rangle_{\mathbb{F}_q}), \\ &= \dim(\langle \mathbf{m}\rangle_{\mathbb{F}_q}) + \dim(\langle \mathbf{m}R_k^{\intercal}\rangle_{\mathbb{F}_q}) - \dim(\langle \mathbf{m}\rangle_{\mathbb{F}_q} \cap \langle \mathbf{m}R_k^{\intercal}\rangle_{\mathbb{F}_q}). \end{aligned} \quad (4)$$

Note that the vector $\mathbf{m}R_k^{\intercal}$ has t components its rank weight is upper bounded by t. Hence the desired result follows. □

Remark 1 *Observe that given a uniformly chosen random matrix $\hat{R} \in \mathbb{F}_{q^m}^{k\times(k+t)}$, its row reduced echelon form, up to column permutation, would be of the form $(I_k|R_t)$, where $R_t \in \mathbb{F}_{q^m}^{k\times t}$ is a random matrix. The expected weight distribution of a hybrid-ELS code with parameters l, t, k, n over $\mathbb{F}_{q^m}$ will be therefore the same as the expected weight distribution a random $\mathbb{F}_{q^m}$-linear code of length $k+t$ and dimension k.*

Given a random matrix R_t, we can assume the components of $\mathbf{m}R_t$ to be uniformly distributed random elements of $\mathbb{F}_{q^m}^t$. Under this assumption, if $t \ll m$, with a high probability the $\mathbb{F}_q$-vector space generated by the components of $\mathbf{m}R_t$ has dimension t for a nonzero message $\mathbf{m} \in \mathbb{F}_{q^m}^k$. Therefore,

$$\begin{aligned}\mathrm{rw}(\mathbf{m}(I_k|R_t)) &= \dim(\langle \mathbf{m} \rangle_{\mathbb{F}_q} + \langle \mathbf{m}R_t \rangle_{\mathbb{F}_q}) \\ &= \dim(\langle \mathbf{m} \rangle_{\mathbb{F}_q}) + \dim(\langle \mathbf{m}R_t \rangle_{\mathbb{F}_q} - \dim(\langle \mathbf{m} \rangle_{\mathbb{F}_q} \cap \langle \mathbf{m}R_t \rangle_{\mathbb{F}_q}).\end{aligned}$$

The intersection of two $\mathbb{F}_q$-linear subspaces in $\mathbb{F}_{q^m}$, when the sum of their dimensions is smaller than m, is in general null or very small. If a message $\mathbf{m}$ has rank $\mathrm{rw}(\mathbf{m}) \ll m$, then $\mathrm{rw}(\mathbf{m}(I_k|R_t)) \leq \mathrm{rw}(\mathbf{m}) + t$, where the equality holds in most of the cases, especially when $\mathrm{rw}(\mathbf{m}) + t$ is small compared to m.

This remark sets a rough upper bound of $t + 1$ on the minimum distance of an hybrid ELS code. Moreover, it provides a rough estimation on the coefficients of low-degree terms in the rank weight enumerator of the code.

5 Rank Syndrome Decoding of Hybrid ELS Codes

The syndrome decoding algorithms of Gabidulin codes and LRPC codes can be generally divided into two steps: the first is to recover the error support and the second is to establish the error components by solving an expanded system of $\mathbb{F}_q$-linear equations.

Consider a Hybrid ELS code $\mathcal{C}$ having parity-check matrix $H = \begin{pmatrix} B \\ R \end{pmatrix}$, where $B \in \mathbb{F}_q^{l \times n}$ and $R \in \mathbb{F}_{q^m}^{t \times n}$. Suppose we receive the vector $\mathbf{y} = \mathbf{c} + \mathbf{e}$, where $\mathbf{c} \in \mathcal{C}$ and $\mathbf{e}$ is an error of rank $\mathrm{rw}(\mathbf{e}) = r$. The support $\mathfrak{E}$ of $\mathbf{e}$ is generated by r elements as $\langle \mathbf{e} \rangle_{\mathbb{F}_q} = \langle \varepsilon_1, \ldots, \varepsilon_r \rangle_{\mathbb{F}_q}$.

Step 1 - Error Support Recovering

Let $B \in \mathbb{F}_q^{l \times n}$ and $R \in \mathbb{F}_{q^m}^{t \times n}$, the syndrome $\mathbf{s} = \mathbf{e}H^\intercal$ can be divided into two parts: one part comes from B and the other comes from R.

The first l components of $\mathbf{s}$ are the result of $\mathbf{e}B^\intercal$. Since the entries of B are all in F_q, it means $\langle s_1, \ldots s_l \rangle_{\mathbb{F}_q} \subseteq \mathfrak{E}$. Note that the components of $\mathbf{e}$ can be considered as uniformly random elements sampled from $\mathfrak{E}$ and B has rank l. The elements $s_1, \ldots, s_l$ can then be considered as l uniformly random elements of $\mathfrak{E}$. Since we know $\dim(\mathfrak{E}) = r$, when $l > r$ we have $\langle s_1, \ldots, s_l \rangle_{\mathbb{F}_q} = \mathfrak{E}$ with a good probability, which will be discussed in next section.

Step 2 - Error Vector Recovering

With the error support $\mathfrak{E} = \langle \varepsilon_1, \ldots, \varepsilon_r \rangle_{\mathbb{F}_q}$ obtained from Step 1, we can rewrite the error as $\mathbf{e} = (\varepsilon_1, \ldots, \varepsilon_r)X$. We can expand over $\mathbb{F}_q$ the system of $l+t$ equations given by $\mathbf{e}H^\intercal = \mathbf{s}$ as

$$s_i = \sum_{j=1}^{n} h_{i,j} e_j = \sum_{j=1}^{n} h_{i,j} \sum_{k=1}^{r} x_{j,k} \varepsilon_k, \qquad x_{j,k} \in \mathbb{F}_q. \tag{5}$$

From (5) one obtains a system of equations with nr variables $x_{j,k} \in \mathbb{F}_q$. In the first l equations the terms $h_{i,j} = b_{i,j} \in \mathbb{F}_q$. We can decompose s_i as $s_i = s_{i,1}\varepsilon_1 + \cdots + s_{i,r}\varepsilon_r$. Therefore (5), for $i \leq l$, can be rewritten as

$$\sum_{j=1}^{n}\sum_{k=1}^{r} b_{i,j}x_{j,k}\varepsilon_k = \sum_{k=1}^{r}\varepsilon_k\left(\sum_{j=1}^{n} b_{i,j}x_{j,k}\right) = s_i = s_{i,1}\varepsilon_1 + \cdots + s_{i,r}\varepsilon_r. \tag{6}$$

Hence, from equation (6), we derive r equations in the form $\sum_{j=1}^{n} b_{i,j}x_{j,k} = s_{i,k}$. As i ranges from 1 to l, we obtain lr equations. We define $B_{exp} = I_r \otimes B$ and the two vectors over $\mathbb{F}_q$:

$$\begin{aligned}\mathbf{x} =&(x_{1,1}, \ldots, x_{n,1}, \ldots, x_{1,r}, \ldots, x_{n,r}),\\ \tilde{\mathbf{s}}_l =&(s_{1,1}, \ldots, s_{l,1}, \ldots, s_{1,r}, \ldots, s_{l,r}).\end{aligned}$$

We can rewrite (6) as $B_{exp}\mathbf{x}^\intercal = \tilde{\mathbf{s}}_l^\intercal$. This system has only lr equations while we have nr variables. In order to have a unique solution, we need more equations and we can get them from the expansion of the t equations given by R.

Note that the system in (5) for $l < i \leq l + t$ is

$$\sum_{j=1}^{n}\sum_{k=1}^{r} r_{i,j}x_{j,k}\varepsilon_k = \sum_{j=1}^{n}\sum_{k=1}^{r} x_{j,k}(r_{i,j}\varepsilon_k) = s_i,$$

which, by the definition of $\mathbf{x}$, can be rewritten as

$$[(\varepsilon_1, \ldots, \varepsilon_r) \otimes (r_{i,1}, \ldots, r_{i,n})] \cdot \mathbf{x}^\intercal = s_i,$$

where $(\varepsilon_1, \ldots, \varepsilon_r)\otimes(r_{i,1}, \ldots, r_{i,n}) = (\varepsilon_1 r_{i,1}, \ldots, \varepsilon_1 r_{i,n}, \ldots, \varepsilon_r r_{i,1}, \ldots, \varepsilon_r r_{i,n})$. Furthermore, the t linear equations over $\mathbb{F}_{q^m}$ can be rewritten as

$$[(\varepsilon_1, \ldots, \varepsilon_r) \otimes R] \cdot \mathbf{x}^\intercal = \mathbf{s}_t^\intercal,$$

where $\mathbf{s}_t$ denotes the vector of the last t components of $\mathbf{s}$.

Let $\alpha_1, \ldots, \alpha_m$ be a basis of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$.

$$\sum_{j=1}^{n}\sum_{k=1}^{r} x_{j,k}(r_{i,j}\varepsilon_k) = \sum_{j=1}^{n}\sum_{k=1}^{r}\sum_{u=1}^{m} x_{j,k}(r_{i,j}\varepsilon_k)^{(u)}\alpha_u = \sum_{u=1}^{m} s_{i,u}\alpha_u, \tag{7}$$

where $(r_{i,j}\varepsilon_k)^{(u)}$ denotes the u-th coordinate with respect to $\alpha_1, \ldots, \alpha_m$. From each row of R we will derive m equations

$$\sum_{j=1}^{n}\sum_{k=1}^{r} x_{j,k}(r_{i,j}\varepsilon_k)^{(u)} = s_{i,u}.$$

Let $\tilde{\mathbf{s}}_t = (s_{l+1,1}, \ldots, s_{n-k,1}, \ldots, s_{l+1,m}, \ldots, s_{n-k,m}) \in \mathbb{F}_q^{tm}$. Denote by $R_k = \varepsilon_k R$ the scalar multiplication of all the entries of R by ε_k. For the basis $\alpha_1, \ldots, \alpha_m$ of

$\mathbb{F}_{q^m}$, we can decompose R_k as $R_k = \alpha_1 R_k^{(1)} + \ldots + \alpha_m R_k^{(m)}$. Then the expanded linear system given by R can be expressed as

$$R_{exp} \cdot \mathbf{x}^\intercal = \tilde{\mathbf{s}}_t^\intercal,$$

where the matrix R_{exp} is given by

$$R_{exp} = \begin{pmatrix} R_1^{(1)} & R_2^{(1)} & \ldots & R_r^{(1)} \\ R_1^{(2)} & R_2^{(2)} & \ldots & R_r^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ R_1^{(m)} & R_2^{(m)} & \ldots & R_r^{(m)} \end{pmatrix}. \tag{8}$$

The whole expanded system can be expressed as

$$\mathbf{x} \cdot H_{exp}^\intercal = \mathbf{x} \cdot (B_{exp}^\intercal | R_{exp}^\intercal) = (\tilde{\mathbf{s}}_l | \tilde{\mathbf{s}}_t). \tag{9}$$

Note that there are $lr + tm$ $\mathbb{F}_q$-linear equations in (9). When $lr + tm \geq nr$, with a good probability there will be nr linearly independent equations, which give a unique solution of the expanded system. Suppose only $u < nr$ of the above mentioned $lr + tm$ equations are linearly independent, it is still possible to perform a list decoding with q^{nr-u} elements in the list.

5.1 Necessary conditions on parameters

Let $H^\intercal = (B^\intercal | R^\intercal)$ be the transposed parity-check matrix of a hybrid ELS code $\mathcal{C}$ of dimension k. In the previous section we saw how to correct an error of small rank r starting from the syndrome. Here we will discuss some necessary conditions on the parameters l and t so as to uniquely correct such an error.

Consider $B \in \mathbb{F}_q^{l \times n}$ and $R \in \mathbb{F}_{q^m}^{t \times n}$, since H has $n - k$ rows we necessarily have $n - k = l + t$. In **Step 1**, in order to recover the error support of an error of rank r, we need $l \geq r$. In **Step 2** we need that $nr \leq lr + tm$. Since $n = l + t + k$ we have $(l + t + k)r \leq lr + tm$. This gives the following bounds on l:

$$r \leq l \leq (n - k) - \frac{kr}{m - r}. \tag{10}$$

5.2 Probability of decoding failure

In the decoding algorithm, **Step 1 and Step 2** both have non-null probabilities of failure. For each step we will try to model the probability of failure. We express the error $\mathbf{e}$ as $(\varepsilon_1, \ldots, \varepsilon_r)X$, where $X \in \mathbb{F}_q^{r \times n}$ and $\langle \varepsilon_1, \ldots, \varepsilon_r \rangle_{\mathbb{F}_q} = \langle \mathbf{e} \rangle_{\mathbb{F}_q} = \mathfrak{E}$.

Step 1. This step only involves $\mathbf{e}B^\intercal = (s_1, \ldots, s_l)$. We want to give a heuristic for the probability that $\langle s_1, \ldots, s_l \rangle_{\mathbb{F}_q} = \mathfrak{E}$. Since $B \in \mathbb{F}_q^{l \times n}$ is a random matrix and $\langle \mathbf{e} \rangle_{\mathbb{F}_q} = \mathfrak{E}$, we can consider $s_1, \ldots, s_l$ as random elements of $\mathfrak{E}$. We can expand each of these terms over a basis of $\mathfrak{E}$. In this way the vector $(s_1, \ldots, s_l)$ becomes a random $r \times l$ matrix S. The condition $\langle s_1, \ldots, s_l \rangle_{\mathbb{F}_q} = \mathfrak{E}$ is equivalent

to rank$(S) = r$. Let $N_q(k, n, r)$ be the number of $k \times n$ matrices over $\mathbb{F}_q$ of rank r. This number can be computed as [1]:

$$N_q(k, n, r) = \frac{A_q(n, r)A_q(k, r)}{A_q(r, r)} = \prod_{i=0}^{r-1} \frac{(q^n - q^i)(q^k - q^i)}{q^r - q^i}. \tag{11}$$

The success probability of this step can be modeled as

$$P(\langle s_1, \ldots, s_l \rangle = \mathfrak{E}) = \frac{N_q(r, l, r)}{q^{lr}} = \frac{\prod_{i=0}^{r-1}(q^l - q^i)}{q^{lr}} = \prod_{i=0}^{r-1}(1 - q^{i-l})$$

The probability of failure is then approximated as follows:

$$P(\langle s_1, \ldots, s_l \rangle \neq \mathfrak{E}) = 1 - \frac{N_q(r, l, r)}{q^{lr}} \approx 1 - \left(1 - \sum_{i=0}^{r-1} q^{i-l}\right) \approx \frac{1}{(q-1)q^{l-r}}. \tag{12}$$

Step 2. In this step we have to solve a system of $lr + tm \geq nr$ linear equations in nr variables. Recall that this system can be formulated as in (9), where $H_{exp} = (B_{exp}^{\intercal} | R_{exp}^{\intercal})$ is an $(lr + tm) \times nr$ matrix over $\mathbb{F}_q$. In order to have a unique solution, we require this matrix to have rank nr.

Here we treat B_{exp} and R_{exp} separately. It is easy to see that $B_{exp} = I_r \otimes B$ in $\mathbb{F}_q^{lr \times nr}$ has full rank lr. The matrix R_{exp} is fully determined by the values of R and $(\varepsilon_1, \ldots, \varepsilon_r)$. Hence it is challenging to describe its expected rank as it is not a truly random matrix. Indeed the rank of R_{exp} could depend partly on the error support. On the other hand, experimental results indicate that the ranks of R_{exp} behave similarly to the rank of random matrices of the same size. In other words, although the matrix H_{exp} is not random, the rank of H_{exp} appears to follow the probability distribution

$$P(\text{rank}(H_{exp}) = z) = N_q(tm + lr, nr, z)/q^{nr(tm+lr)}.$$

To illustrate this behavior we run a test in Magma with two sets of parameters. For each set of parameters, we generate 1000 matrices H_{exp}, 1000 random matrices of the same size and then compute their ranks. The experimental results are summarized in Table 1. In Table 1, for a given value of the parameters $\langle q^m, n, l, t, r \rangle$, the upper part of a row lists the number of matrices H_{exp} with ranks in the range $[nr - 4, nr]$ and the lower part lists the number of random matrices of the same rank. The first set of parameters in Table 1 represents a limit case that $lr + tm - nr = 0$. As shown in Table 1, even in the limit case, there were just relatively few cases where rank$(H_{exp}) = nr - 3$, for which we would still be able to use a list decoding with list size 8. In the second set of parameters, we have $lr + tm - nr = 4$. With this small difference, we see that the ranks of the H_{exp} instances tend to rapidly concentrate on the first column.

To conclude this section, we illustrate the overall error correction capability of the Hybrid ELS codes with the unique decoding algorithm for different parameters as in Table 2. The results in Table 2 were obtained by running 1000

Parameters	Ranks				
$\langle q^m, n, l, t, r\rangle$	nr	$nr-1$	$nr-2$	$nr-3$	$nr-4$
$\langle 2^{15}, 25, 10, 4, 4\rangle$	297	565	132	6	0
	285	587	123	5	0
$\langle 2^{14}, 20, 12, 2, 3\rangle$	930	69	1	0	0
	937	63	0	0	0

Table 1. Rank distribution of H_{exp} and random matrices

$\langle q^m, n, l, t, r\rangle$	Support Recovery	Error Recovery
$\langle 2^{15}, 20, 10, 4, 6\rangle$	95.2%	57.1%
$\langle 2^{15}, 20, 10, 4, 5\rangle$	97.4%	97.4%
$\langle 2^{15}, 20, 10, 4, 4\rangle$	98.3%	98.3%
$\langle 2^{23}, 30, 15, 4, 6\rangle$	99.9%	86.3%
$\langle 2^{23}, 30, 15, 5, 6\rangle$	99.8%	99.8%
$\langle 2^{23}, 30, 16, 5, 6\rangle$	100%	100%
$\langle 7^{10}, 19, 9, 4, 4\rangle$	100%	85.4%
$\langle 7^{11}, 19, 9, 4, 4\rangle$	100%	100%

Table 2. Success rate of decoding HELS codes

randomly generated examples for each set of parameters. Experimental results show that the proposed codes can outperform the LRPC codes in terms of error correcting rate. Nevertheless, our analysis indicates that the proposed codes are not as properly masked in the same way as LRPC codes for cryptographic applications.

5.3 Cryptographic weakness

Last point is not that surprising since also in the hamming weight it is pretty trivial to solve the shortest vector problem when the weight is 1.

The main concern in using hybrid ELS codes for cryptography is related to the structure of their parity-check matrix. Suppose $\mathcal{C}$ is a hybrid ELS codes having parity-check matrix $H = \begin{pmatrix} B \\ R \end{pmatrix}$. As a public key we have to share G a generator matrix for $\mathcal{C}$ from which it is difficult to reconstruct H. To succeed an attacker does not need to find exactly the same H, it would be sufficient to find a parity-check matrix $H' = \begin{pmatrix} B' \\ R' \end{pmatrix}$ such that $B' \in \mathbb{F}_q^{l \times n}$ in order to correct. Finding B' with this property is equivalent to find l linearly independent vectors of rank weight 1 in $\mathcal{C}^\perp$. We know that such vectors always exist by construction. Moreover the attacker can always compute a base of $\mathcal{C}^\perp$ from the generator matrix G of $\mathcal{C}$. The problem of finding a base in the desired form can be reduced to the problem of finding l linearly independent vectors in $\mathcal{C}^\perp$ having rank weight 1. Find a vector of rank 1 is not difficult. Suppose $\mathbf{v}\mathcal{C}^\perp$ is a vector of rank 1, so its support given by $\langle \mathbf{v} \rangle_{\mathbb{F}_q} = \langle \alpha \rangle_{\mathbb{F}_q}$ for some $\alpha \in \mathbb{F}_{q^m}$. Thanks to $\mathbb{F}_{q^m}$-linearity without loss of generality we can consider $\langle \mathbf{v} \rangle_{\mathbb{F}_q} = \mathbb{F}_q$. We just have to determine

the n coordinates of this vector. This can be done expanding over $\mathbb{F}_q$ the system given by G. Noticing that G is a parity-check matrix of $\mathcal{C}^{\perp}$. The expansion will give us mk equations over $\mathbb{F}_q$, the solution to this system will be exactly the space generated by the rows of B.

6 Conclusion

In the paper we propose a new random-like rank metric codes that can be efficiently decoded with a polynomial-time complexity. The proposed codes appear to have reasonably good error-correcting capability, owing to the special structure of their parity-check matrices.

References

1. G. E. Andrews. *The Theory of Partitions*. Cambridge University Press, 2003.
2. N. Aragon, P. Gaborit, A. Hauteville, O. Ruatta, and G. Zémor. Low rank parity check codes: New decoding algorithms and applications to cryptography. *IEEE Transactions on Information Theory*, 65(12):7697–7717, 2019.
3. P. Delsarte. Bilinear forms over a finite field, with applications to coding theory. *Journal of Combinatorial Theory, Series A*, 25(3):226 – 241, 1978.
4. E. M. Gabidulin. Theory of codes with maximum rank distance. *Problemy Peredachi Informatsii*, 21(1):3–16, 1985.
5. E. M. Gabidulin, A. V. Paramonov, and O. V. Tretjakov. Ideals over a non-commutative ring and their application in cryptology. In D. W. Davies, editor, *Advances in Cryptology – EUROCRYPT'91*, pages 482–489. Springer, 1991.
6. P. Gaborit, A. Hauteville, D. H. Phan, and J.-P. Tillich. Identity-based encryption from codes with rank metric. In J. Katz and H. Shacham, editors, *Advances in Cryptology – CRYPTO 2017*, pages 194–224. Springer International Publishing, 2017.
7. P. Gaborit, O. Ruatta, and J. Schrek. On the complexity of the rank syndrome decoding problem. *IEEE Transactions on Information Theory*, 62(2):1006–1019, 2016.
8. P. Gaborit, O. Ruatta, J. Schrek, and G. Zémor. Ranksign: an efficient signature algorithm based on the rank metric. In M. Mosca, editor, *Post-Quantum Cryptography*, pages 88–107. Springer International Publishing, 2014.
9. P. Gaborit and G. Zémor. On the hardness of the decoding and the minimum distance problems for rank codes. *IEEE Transactions on Information Theory*, 62(12):7245–7252, 2016.
10. M. Gadouleau and Z. Yan. On the decoder error probability of bounded rank-distance decoders for maximum rank distance codes. *IEEE Transactions on Information Theory*, 54(7):3202–3206, 2008.
11. C. A. Melchor, N. Aragon, M. Bardet, S. Bettaieb, L. Bidoux, O. Blazy, J.-C. Deneuville, P. Gaborit, A. Hauteville, A. Otmani, O. Ruatta, J.-P. Tillich, and G. Zémor. ROLLO (merger of Rank-Ouroboros, LAKE and LOCKER). In *Second round submission to the NIST post-quantum cryptography call*, April, 2020.
12. Philippe Gaborit, Gaétan Murat, Olivier Ruatta, and Gilles Zémor. Low rank parity check codes and their application to cryptography. in proceedings of the workshop on coding and cryptography WCC'2013 Bergen Norway 2013. available on www.selmer.uib.no/wcc2013/pdfs/gaborit.pdf.

13. R. M. Roth. Maximum-rank array codes and their application to crisscross error correction. *IEEE Transactions on Information Theory*, 37(2):328–336, 1991.
14. D. Silva, F. R. Kschischang, and R. Koetter. A rank-metric approach to error control in random network coding. *IEEE Transactions on Information Theory*, 54(9):3951–3967, Sept 2008.