LowMS: a new rank metric code-based KEM without ideal structure

Nicolas Aragon[1], Victor Dyseryn[2 ✉], Philippe Gaborit[2], Pierre Loidreau[3], Julian Renner[4], and Antonia Wachter-Zeh[4]

[1] NAQUIDIS Center, France.
[2] XLIM, Université de Limoges, France.
[3] Université de Rennes, DGA-MI, CNRS, IRMAR, France.
[4] Technical University of Munich, Institute for Communications Engineering, Germany.

✉ victor.dyseryn_fostier@unilim.fr

Abstract. We propose and analyze LowMS, a new rank-based key encapsulation mechanism (KEM). The acronym stands for *Loidreau with Multiple Syndromes*, since our work combines the cryptosystem of Loidreau (presented at PQCrypto 2017) together with the multiple syndromes approach, that allows to reduce parameters by sending several syndromes with the same error support in one ciphertext.
Our scheme is designed without using ideal structures. Considering cryptosystems without such an ideal structure, like the FrodoKEM cryptosystem, is important since structure allows to compress objects, but gives reductions to specific problems whose security may potentially be weaker than for unstructured problems. For 128 bits of security, we propose parameters with a public key size of 4.8KB and a ciphertext size of 1.1KB. To the best of our knowledge, our scheme is the smallest among all existing unstructured post-quantum lattice or code-based algorithms, when taking into account the sum of the public key size and the ciphertext size. In that sense, our scheme is for instance about 4 times shorter than FrodoKEM. Our system relies on the hardness of the Rank Support Learning problem, a well-known variant of the Rank Syndrome Decoding problem, and on the problem of indistinguishability of distorted Gabidulin codes, i.e., Gabidulin codes multiplied by a homogeneous matrix of given rank. The latter problem was introduced by Loidreau in his paper.

Keywords: Rank-based cryptography, code-based cryptography, post-quantum cryptography, rank support learning.

1 Introduction and previous work

Quantum resistant cryptography (or post-quantum cryptography) aims at replacing currently-used number theorectic systems like RSA or Diffie-Hellman, which were shown vulnerable against quantum computer attacks [47].

This paper deals with one family of post-quantum-secure public-key cryptographic algorithms – code-based cryptography – for which two metrics can be considered. The most famous one is the Hamming metric and was used in the seminal work of McEliece [40]. The second one is the rank metric [22], where words are embedded in $\mathbb{F}_{q^m}$, the degree-m extension of the field $\mathbb{F}_q$. In the rank metric, the weight of a word is defined as the rank of the matrix computed by unfolding the word using a basis of $\mathbb{F}_{q^m}$ on $\mathbb{F}_q$.

Rank-metric codes are a promising candidate for code-based cryptography since generic decoding in the rank metric appears to be much harder than generic decoding in the Hamming metric for the same length and alphabet size. Hence, they provide significantly smaller key sizes at the same level of security against generic decoding.

Among the different cryptographic primitives, rank-based cryptography literature is mainly focused on encryption schemes. Note that the rank metric is also relevant to produce small size and general purpose digital signatures, such as Durandal [8]. The first rank-based cryptosystem was the Gabidulin–Paramonov–Tretjakov (GPT) [24] system, a McEliece-like cryptosystem in the rank metric using Gabidulin codes. GPT and most of its variants [20,12,33,31] were broken by attacks which exploit the particular structure of Gabidulin codes [43,23,32].

Some alternative rank-metric cryptosystems are based on other code classes, such as LRPC codes [26], which are easier to mask. Another possibility is to design schemes without masking, such as RQC [1]. Both approaches are less efficient than GPT and, in order to remain competitive, authors introduced structure in the underlying algebraic objects, such as quasi-cyclic or ideal structure. Adding structure comes at the cost of losing reductions to difficult problems in the more general form; it is a potential weakness.

In this paper we do not require any ideal structure. We build upon the only Gabidulin-code-based GPT variant that has not been broken so far; the one by Loidreau [37]. For this cryptosystem the masking consists in multiplying the Gabidulin parity-check matrix by a homogeneous matrix of rank λ. The inconvenience of this approach is that the error weight is multiplied by λ, which strongly increases parameters. However, whenever λ is chosen sufficiently high, it seems to resist against structural attacks, which makes sense since this type of homogeneous structure is also used for the LRPC cryptosystem which is also resistant (still depending on the value of λ). Notice that Loidreau's cryptosystem is also known as DRANKULA and was implemented in [4]. The multiple syndromes approach, which inherently increases the decoding capacity of the code, permits to drastically reduce its parameters.

The multiple syndromes technique consists of sending several syndromes $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_\ell$ of same error support. The idea was introduced in [25] and further developed in [49]. Even more recently, the multiple syndromes technique was applied to LRPC-based cryptosystems and gave birth to LRPC-MS [3]. Multi-UR-AG [14] is the combination of RQC with multiple syndromes and no quasi-cyclic structure. These systems were the most efficient unstructured code-based KEMs so far. The decoding of multiple syndromes that result from errors sharing the same support can also be referred to as the decoding of interleaved codes. Interleaved Gabidulin codes and their decoding were studied in [35]. Furthermore, the idea of using interleaved codes for Hamming-based cryptosystems was introduced in [19,30].

As written earlier, in this paper we apply the multiple syndromes technique to Loidreau's cryptosystem [37]. The resulting cryptosystem is presented in Section 3 and is related to [45] which combines the ideas of interleaved codes with Loidreau's cryptosystem. In Section 4 we provide a security analysis with an IND-CPA proof and a review of known attacks. The crucial difference between [45] and our work is the error model. In [45], the rank weight of an error matrix $\boldsymbol{E}$ is the rank of the matrix obtained by the *vertical* concatenation of matrices given by unfolding each row of $\boldsymbol{E}$ using a public basis of $\mathbb{F}_{q^m}$ on $\mathbb{F}_q$. In this paper, the rank weight of an error matrix $\boldsymbol{E}$ is the rank of the matrix obtained by the *horizontal* concatenation of the unfoldings. The difference is fundamental and constitutes the main contribution of this paper; full details are given in Section 3.3. In the first vertical case, the error must be drawn as a product of two matrices $\boldsymbol{AB}$, which generates contraints. In the second horizontal case – presented in this very paper – the error is drawn naturally by choosing an error support E and picking each coordinate at random in E. This alleviates many constraints and the public keys and ciphertexts of the resulting cryptosystem are more than five times shorter than [45]. Our new approach also strongly outperforms all previous approaches based on a masking of Gabidulin codes, shortening the public key with at least the same factor. Moreover, in the horizontal case, the security relies on the RSL problem which is believed to be hard [25].

More generally, it makes our cryptosystem the shortest KEM without ideal structure, outperforming the sizes of LRPC-MS [3] and Multi-UR-AG [14]. With a public key size of 4.77KB and ciphertext size of 1.14KB for 128 bits of security, our scheme is more than three times shorter than FrodoKEM and 45 times shorter than Classic McEliece. When comparing to structured lattice and code-based proposals like CRYSTALS-Kyber [15] or HQC [2], our size is about twice longer. We consider it a small price to pay for an additional guarantee of security granted by the removal of the underlying structure. More details about parameters and comparison with other schemes can be found in Section 5.

2 Background on rank metric codes

2.1 General definitions

Let $\mathbb{F}_q$ denote the finite field of q elements where q is the power of a prime and let $\mathbb{F}_{q^m}$ denote the field of q^m elements i.e., the extension field of degree m of $\mathbb{F}_q$. $\mathbb{F}_{q^m}$ is also an $\mathbb{F}_q$-vector space of dimension m; we denote by capital letters the $\mathbb{F}_q$-subspaces of $\mathbb{F}_{q^m}$ and by lower-case letters the elements of $\mathbb{F}_{q^m}$. The Grassmannian $\mathbf{Gr}(\mathbb{F}_{q^m}, k)$ represents the set of all subspaces of $\mathbb{F}_{q^m}$ of dimension k.

Let $X \subset \mathbb{F}_{q^m}$. We denote by $\langle X \rangle$ the $\mathbb{F}_q$-subspace generated by the elements of X:

$$\langle X \rangle = \mathrm{Vect}_{\mathbb{F}_q}(X).$$

If $X = \{x_1, \ldots, x_n\}$, we simply use the notation $\langle x_1, \ldots, x_n \rangle$.

Vectors are denoted by bold lower-case letters and matrices by bold capital letters (e.g., $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{F}_{q^m}^n$ and $\boldsymbol{M} = (m_{ij})_{\substack{1 \leqslant i \leqslant k \\ 1 \leqslant j \leqslant n}} \in \mathbb{F}_{q^m}^{k \times n}$).

If S is a finite set, we denote by $x \stackrel{\$}{\leftarrow} S$ when x is chosen uniformly at random from S.

The number of $\mathbb{F}_q$-subspaces of dimension r of $\mathbb{F}_{q^m}$ is given by the Gaussian coefficient

$$\begin{bmatrix} m \\ r \end{bmatrix}_q = \prod_{i=0}^{r-1} \frac{q^m - q^i}{q^r - q^i}.$$

Definition 1 (Rank metric over $\mathbb{F}_{q^m}^n$). *Let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{F}_{q^m}^n$ and $(\gamma_1, \ldots, \gamma_m) \in \mathbb{F}_{q^m}^m$ be a basis of $\mathbb{F}_{q^m}$ viewed as an m-dimensional vector space over $\mathbb{F}_q$. Each coordinate x_j is associated to a vector of $\mathbb{F}_q^m$ in this basis by $x_j = \sum_{i=1}^m m_{ij}\gamma_i$. The $m \times n$ matrix associated to $\boldsymbol{x}$ is given by $\boldsymbol{M}(\boldsymbol{x}) = (m_{ij})_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}}$.*

The rank weight $\|\boldsymbol{x}\|$ of $\boldsymbol{x}$ is defined as

$$\|\boldsymbol{x}\| \stackrel{\text{def}}{=} \operatorname{rank} \boldsymbol{M}(\boldsymbol{x}).$$

The associated distance $d(\boldsymbol{x}, \boldsymbol{y})$ between two elements $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{F}_{q^m}^n$ is defined by $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$.

Definition 2 ($\mathbb{F}_{q^m}$-linear code). *An $\mathbb{F}_{q^m}$-linear code $\mathcal{C}$ of dimension k and length n is a subspace of dimension k of $\mathbb{F}_{q^m}^n$. The notation $\mathcal{C}[n, k]$ is used to denote its parameters. Its minimal distance d is the minimum weight of non-zero vectors in $\mathcal{C}$.*

The code $\mathcal{C}$ can be represented by two equivalent ways:

- *by a generator matrix $\boldsymbol{G} \in \mathbb{F}_{q^m}^{k \times n}$. Each row of $\boldsymbol{G}$ is an element of a basis of $\mathcal{C}$,*

$$\mathcal{C} = \{\boldsymbol{x}\boldsymbol{G}, \boldsymbol{x} \in \mathbb{F}_{q^m}^k\}.$$

- *by a parity-check matrix $\boldsymbol{H} \in \mathbb{F}_{q^m}^{(n-k) \times n}$. Each row of $\boldsymbol{H}$ determines a parity-check equation verified by the elements of $\mathcal{C}$:*

$$\mathcal{C} = \{\boldsymbol{x} \in \mathbb{F}_{q^m}^n : \boldsymbol{H}\boldsymbol{x}^T = \boldsymbol{0}\}.$$

We say that $\boldsymbol{G}$ (respectively $\boldsymbol{H}$) is in systematic form if and only if it is of the form $(\boldsymbol{I}_k|\boldsymbol{A})$ (respectively $(\boldsymbol{I}_{n-k}|\boldsymbol{B})$).

We also need to define the support of a word and homogeneous matrices, which play a key role in our cryptosystem.

Definition 3 (Support of a word). *Let* $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{F}_{q^m}^n$*. The support of* $\boldsymbol{x}$*, denoted* $\mathrm{Supp}(\boldsymbol{x})$*, is the* $\mathbb{F}_q$*-subspace of* $\mathbb{F}_{q^m}$ *generated by the coordinates of* $\boldsymbol{x}$*:*

$$\mathrm{Supp}(\boldsymbol{x}) := \langle x_1, \dots, x_n \rangle_{\mathbb{F}_q}.$$

This definition is coherent with the definition of the rank weight since $\dim(\mathrm{Supp}(\boldsymbol{x})) = \|\boldsymbol{x}\|$*.*

We extend the previous definition to matrices.

Definition 4 (Support of a matrix). *Let* $\boldsymbol{A} = (A_{i,j}) \in \mathbb{F}_{q^m}^{\ell \times n}$*. The support of* $\boldsymbol{A}$ *is defined as:*

$$\mathrm{Supp}(\boldsymbol{A}) := \langle A_{1,1}, \dots, A_{1,n}, A_{2,1}, \dots, A_{2,n}, \dots, A_{\ell,1}, \dots, A_{\ell,n} \rangle_{\mathbb{F}_q}.$$

We also need to define homogeneous matrices.

Definition 5 (Homogeneous matrices of given support/weight). *Let* $\boldsymbol{M} \in \mathbb{F}_{q^m}^{k \times n}$ *be a matrix over* $\mathbb{F}_{q^m}$ *and let* E *be an* $\mathbb{F}_q$*-subspace of* $\mathbb{F}_{q^m}$*. The matrix* $\boldsymbol{M}$ *is said to be homogeneous of support* E *if* $\mathrm{Supp}(\boldsymbol{M})$ *is equal to* E*. If* $d = \dim E$*, then* $\boldsymbol{M}$ *is also said to be homogeneous of weight* d*.*

2.2 Interleaved Gabidulin codes and their decoding

Gabidulin codes [22] are a well-known class of rank-metric codes and can be seen as the rank-metric analogs of Reed–Solomon codes.

Definition 6 (Gabidulin Code). *A Gabidulin code* $\mathcal{G}[n,k]$ *over* $\mathbb{F}_{q^m}$ *of length* $n \leq m$ *and dimension* k *is defined by its* $k \times n$ *generator matrix*

$$\boldsymbol{G} = \begin{pmatrix} g_1 & g_2 & \cdots & g_n \\ g_1^{[1]} & g_2^{[1]} & \cdots & g_n^{[1]} \\ \vdots & \vdots & \ddots & \vdots \\ g_1^{[k-1]} & g_2^{[k-1]} & \dots & g_n^{[k-1]} \end{pmatrix},$$

where $\boldsymbol{g} = (g_1, g_2, \dots, g_n) \in \mathbb{F}_{q^m}^n$*,* $\|\boldsymbol{g}\| = n$ *and* $[i] = q^i$*. The vector* $\boldsymbol{g}$ *is called the generator of the code* $\mathcal{G}$*.*

Proposition 1 ([36]). *A Gabidulin code* $\mathcal{G}[n,k]$ *generated by* $\boldsymbol{g}$ *admits as a parity-check matrix*

$$\boldsymbol{H} = \begin{pmatrix} h_1 & h_2 & \dots & h_n \\ h_1^{[1]} & h_2^{[1]} & \dots & h_n^{[1]} \\ \vdots & \vdots & \ddots & \vdots \\ h_1^{[n-k-1]} & h_2^{[n-k-1]} & \dots & h_n^{[n-k-1]} \end{pmatrix},$$

where $(h_1, ..., h_n) = (\alpha_1^{[n-k+1]}, ..., \alpha_n^{[n-k+1]})$ *with the* α_i *verifying*

$$\sum_{i=1}^{n} \alpha_i g_i^{[j]} = 0$$

for $j \in \{0, 1, ..., n-2\}$*. We note* $\mathcal{G}^\top_{(n,k)}$ *the set of all parity-check matrices of* $[n, k]$ *Gabidulin codes.*

In [22], it is shown that Gabidulin codes are called Maximum Rank Distance (MRD) codes, i.e., their minimum distance satisfies $d = n - k + 1$, and can be decoded uniquely up to $t \leq \lfloor \frac{d-1}{2} \rfloor$.

Interleaved Gabidulin codes are a code class containaing words of length ℓn in which each subword of length n is a Gabidulin codeword.

Definition 7 (Interleaved Gabidulin Codes [38]). *Let* $\mathcal{G}$ *be a Gabidulin code. An interleaved Gabidulin code* $\mathcal{IG}(\ell; \mathcal{G})$ *over* $\mathbb{F}_{q^m}$ *of interleaving order* ℓ *is defined by*

$$\mathcal{IG}(\ell; \mathcal{G}) := \left\{ \left(\boldsymbol{c}_{\mathcal{G},1} \dots \boldsymbol{c}_{\mathcal{G},\ell} \right) \mid \boldsymbol{c}_{\mathcal{G},i} \in \mathcal{G}, \forall i \in [1, \ell] \right\}.$$

An interleaved Gabidulin code is a rank metric code of length ℓn and dimension ℓk.

Proof. The interleaved code $\mathcal{IG}(\ell; \mathcal{G})$ is stable under linear combinations because each of the subwords $\boldsymbol{c}_{\mathcal{G},i}$ are stable under linear combinations. Hence $\mathcal{IG}(\ell; \mathcal{G})$ is a rank metric code of length ℓn. As for its dimension, $\mathcal{IG}(\ell; \mathcal{G})$ is the direct sum of ℓ subcodes $\left\{ \left(0 \dots \boldsymbol{c}_{\mathcal{G},i} \dots 0 \right) \mid \boldsymbol{c}_{\mathcal{G},i} \in \mathcal{G} \right\}$, each isomorphic to a Gabidulin code of dimension k, hence the total dimension of $\mathcal{IG}(\ell; \mathcal{G})$ is ℓk.

Remark 1. This corresponds to the so-called horizontal interleaving. Others authors considered vertical interleaving for different purposes, see for example [46].

Interleaved Gabidulin codes can be corrected with high probability beyond the $\lfloor \frac{d-1}{2} \rfloor$ bound. More precisely, efficient decoders are known that are able to correct $t \leq \lfloor \frac{\ell}{\ell+1}(n-k) \rfloor$ errors with high probability. We recall below the result of [48] regarding the decoding probability of an interleaved Gabidulin code.

Proposition 2 ([48], Equations (43) and (44)). *Let* $\mathcal{G}$ *be a Gabidulin code of parity check matrix* $\boldsymbol{H}$ *and* $\mathcal{IG}(\ell; \mathcal{G})$ *the corresponding interleaved code of order* ℓ*.*

Let $E \xleftarrow{\$} \mathbf{Gr}(\mathbb{F}_{q^m}, t)$ *be an error support of dimension* t *with* $\ell \leq t \leq \lfloor \frac{\ell}{\ell+1}(n-k) \rfloor$*. Let an error* $\boldsymbol{e} = (\boldsymbol{e}_1 \dots \boldsymbol{e}_\ell) \in E^{\ell n}$ *where for each* i*,* $\boldsymbol{e}_i \xleftarrow{\$} E^n$*. Let* $\boldsymbol{y} \in \mathbb{F}_{q^m}^{\ell(n-k)}$ *be the corresponding syndrome of the interleaved code* $\mathcal{IG}(\ell; \mathcal{G})$*:*

$$\boldsymbol{y} = (\boldsymbol{e}_1 \boldsymbol{H}^\top \dots \boldsymbol{e}_\ell \boldsymbol{H}^\top).$$

The decoding Algorithm 4 from [48], on input $\boldsymbol{y}$*, fails to output correctly the error* $\boldsymbol{e}$ *with a probability upper bounded by*

$$3.5 q^{-m\left\{(\ell+1)\left(\frac{\ell}{\ell+1}(n-k)-t\right)+1\right\}}.$$

We can then build a decoding algorithm for Interleaved Gabidulin codes that takes as input an $\ell \times (n-k)$ syndrome matrix and returns the error vector.

Algorithm 1 $\mathsf{InterleavedGab.Decode}$

Input: Received syndrome matrix $\boldsymbol{Y} \in \mathbb{F}_{q^m}^{\ell \times (n-k)}$
Output: Error matrix $\boldsymbol{E} \in \mathbb{F}_{q^m}^{\ell \times n}$ or decoding failure $\perp$

1: Flatten $\boldsymbol{Y}$ into $\boldsymbol{y} = (\boldsymbol{y}_1 \dots \boldsymbol{y}_\ell) \in \mathbb{F}_{q^m}^{\ell(n-k)}$ where $\boldsymbol{y}_i$ denotes the i-th row of $\boldsymbol{Y}$.
2: Apply Algorithm 4 from [48] to $\boldsymbol{y}$.
3: If it fails, return $\perp$.
4: Else, we get an error vector $\boldsymbol{e} = (\boldsymbol{e}_1 \dots \boldsymbol{e}_\ell)$ where each suberror $\boldsymbol{e}_i \in \mathbb{F}_{q^m}^n$.
5: **return** the matrix $\boldsymbol{E}$ whose rows are $\boldsymbol{e}_1, \dots, \boldsymbol{e}_\ell$.

Proposition 2 turns immediately into the following corollary which is adapted to $\mathsf{InterleavedGab.Decode}$ algorithm.

Corollary 1. *Let $\mathcal{G}$ be a Gabidulin code of parity-check matrix $\boldsymbol{H}$. Let $E \xleftarrow{\$} \mathbf{Gr}(\mathbb{F}_{q^m}, t)$ an error support of dimension t with $\ell \leq t \leq \lfloor \frac{\ell}{\ell+1}(n-k) \rfloor$. Let $\boldsymbol{Y} \in \mathbb{F}_{q^m}^{\ell \times n}$ be defined by $\boldsymbol{Y} = \boldsymbol{E}\boldsymbol{H}^\top$ where the error is a matrix $\boldsymbol{E} \xleftarrow{\$} E^{\ell \times n}$ whose coefficients are picked uniformly at random in the error support.*

Algorithm $\mathsf{InterleavedGab.Decode}$ (1), on input $\boldsymbol{Y}$, fails to output correctly the error matrix $\boldsymbol{E}$ with a probability upper bounded by

$$3.5 q^{-m\left\{(\ell+1)\left(\frac{\ell}{\ell+1}(n-k)-t\right)+1\right\}}.$$

2.3 Difficult problems in rank metric

We recall some hard problems for the rank metric.

Problem 1 (Rank Support Decoding (RSD)*).* Let $\boldsymbol{H}$ be an $(n-k)\times n$ parity-check matrix of an $[n,k]$ $\mathbb{F}_{q^m}$-linear code, $\boldsymbol{y} \in \mathbb{F}_{q^m}^{n-k}$ and r an integer. The $\mathrm{RSD}_{q,m,n,k,r}$ problem is to find $\boldsymbol{e}$ such that $\|\boldsymbol{e}\| = r$ and $\boldsymbol{H}\boldsymbol{e}^T = \boldsymbol{y}^T$.

We also define the Rank Support Learning (RSL) problem, on which the security of our cryptosystem will be based.

*Problem 2 (Rank Support Learning (*RSL*)).* Let $\boldsymbol{H}$ be a random full-rank $(n-k)\times n$ matrix over $\mathbb{F}_{q^m}$. Let $\mathcal{O}$ be an oracle which, given $\boldsymbol{H}$, gives samples of the form $\boldsymbol{H}\boldsymbol{e}_1^T, \boldsymbol{H}\boldsymbol{e}_2^T, \dots, \boldsymbol{H}\boldsymbol{e}_\ell^T$, with the vectors $\boldsymbol{e}_i$ randomly chosen from a space E^n, where E is a random subspace of $\mathbb{F}_{q^m}$ of dimension r. The $\mathrm{RSL}_{q,m,n,k,r}$ problem is to recover E given only access to the oracle.

We denote $\mathrm{RSL}_{q,m,n,k,r,\ell}$ the $\mathrm{RSL}_{q,m,n,k,r}$ problem where we are allowed to make exactly ℓ calls to the oracle, meaning we are given exactly ℓ syndrome values $\boldsymbol{He}_i^T$. By an instance of the RSL problem, we shall mean a sequence

$$(\boldsymbol{H}, \boldsymbol{He}_1^T, \boldsymbol{He}_2^T, \ldots, \boldsymbol{He}_\ell^T),$$

that we can also view as a pair of matrices $(\boldsymbol{H}, \boldsymbol{T})$, where $\boldsymbol{T}$ is the matrix whose columns are the $\boldsymbol{He}_i^T$.

Decisional problems. Both the RSD and RSL problems also have decisional variants for which the goal is to distinguish (for the example of RSD) between a random input $(\boldsymbol{H}, \boldsymbol{s})$ or an actual syndrome input $(\boldsymbol{H}, \boldsymbol{He}^\top)$. We denote these decisional versions DRSD and DRSL. The reader is referred to [9] for more details about decisional problems.

We define the problem of the indistinguishability of distorted Gabidulin codes.

Problem 3 (Distorted Gabidulin codes indistinguishability IND-Gab*).* Given a matrix $\boldsymbol{H'} \in \mathbb{F}_{q^m}^{(n-k)\times n}$, the problem $\mathsf{IND\text{-}Gab}_{q,m,n,k,\lambda}$ distinguish whether $\boldsymbol{H'}$ is random or the parity-check matrix of a distorted Gabidulin code, i.e. $\boldsymbol{H'} = \boldsymbol{SHP}$ with $\boldsymbol{S}$ an $(n-k)\times(n-k)$ matrix with entries in $\mathbb{F}_{q^m}$, $\boldsymbol{H}$ the parity-check matrix of an $[n,k]$ Gabidulin code, and $\boldsymbol{P}$ an $n \times n$ homogeneous matrix of weight λ.

This problem was studied in [34] and we give the complexity of the best known attack to solve this problem in Section 4.3.

3 LowMS: Loidreau's cryptosystem with Multiple Syndromes

3.1 Description of the scheme

The LowMS KEM scheme is given by three algorithms (LowMS.KeyGen, LowMS.Encaps, LowMS.Decaps) defined in Algorithms 2, 3, 4. LowMS KEM is parametrized by the following parameters:

- q the size of the base field $\mathbb{F}_q$
- m the degree of the field $\mathbb{F}_{q^m}$ used in rank metric
- (k, n) the dimension and length of a Gabidulin code
- r the rank weight of the error[1]
- λ the rank weight of the perturbation matrix
- ℓ the number of syndromes sent in the ciphertext (interleaving order)
- $\mathcal{H}$ is a hash function which outputs values $\in \mathbb{F}_2^{512}$, such as SHA-512

[1] In the comparison paper [45], r is noted t_{pub}.

We use the Niederreiter framework [42] instead of the McEliece one to define our scheme, i.e., we perform all operations using parity-check matrices instead of generator matrices. This allows to divide the size of the ciphertext by 2 (if $k = n/2$). Similarly to ROLLO [7] and other rank metric KEMs, using a Niederreiter system implies to compute the shared secret as a hashed value of the error support E.

Algorithm 2 LowMS.KeyGen

Input: None
Output: Keypair $(pk, sk) \in \left(\mathbb{F}_{q^m}^{(n-k)\times n}, \mathbb{F}_{q^m}^{(n-k)\times(n-k)} \times \mathbb{F}_{q^m}^{(n-k)\times n} \times \mathbb{F}_{q^m}^{n\times n}\right)$

1: Choose a parity-check matrix of an $[n, k]$ Gabidulin code $\boldsymbol{H} \xleftarrow{\$} \mathcal{G}_{(n,k)}^{\top} \in \mathbb{F}_{q^m}^{(n-k)\times n}$.
2: Choose an $\mathbb{F}_q$-subspace of $\mathbb{F}_{q^m}$, $F \xleftarrow{\$} \mathbf{Gr}(\mathbb{F}_{q^m}, \lambda)$.
3: Choose uniformly at random an $n \times n$ perturbation matrix with entries in F, $\boldsymbol{P} \xleftarrow{\$} F^{n\times n}$.
4: Compute $\boldsymbol{S} \in \mathbb{F}_{q^m}^{(n-k)\times(n-k)}$ such that $\boldsymbol{H}' = \boldsymbol{S}^{\top}\boldsymbol{H}\,\boldsymbol{P}^{\top}$ is in systematic form.
5: Define $pk := \boldsymbol{H}'$ and $sk := (\boldsymbol{S}, \boldsymbol{H}, \boldsymbol{P})$.
6: **return** (pk, sk).

Algorithm 3 LowMS.Encaps

Input: Public key $pk = \boldsymbol{H}' \in \mathbb{F}_{q^m}^{(n-k)\times n}$.
Output: Ciphertext $c \in \mathbb{F}_{q^m}^{\ell\times(n-k)}$, session key $K \in \mathbb{F}_2^{512}$.

1: Sample the error support $E \xleftarrow{\$} \mathbf{Gr}(\mathbb{F}_{q^m}, r)$.
2: Sample the error matrix $\boldsymbol{E} \xleftarrow{\$} E^{\ell\times n}$, such that $\mathrm{Supp}(\boldsymbol{E}) = E$.
3: Compute $\boldsymbol{C} = \boldsymbol{E}\,\boldsymbol{H}'^{\top}$.
4: Compute $K = \mathcal{H}(E)$.
5: **return** $c = \boldsymbol{C}, K$.

The decoding algorithm recovers the support E of the error matrix as long as conditions of Corollary 1 apply, i.e. $\ell \leqslant r\lambda \leqslant \lfloor \frac{\ell}{\ell+1}(n-k) \rfloor$.

Remark 2. In order to hash E and obtain the same value during encryption and decryption, we need a canonical representation for a subspace E of $\mathbb{F}_{q^m}$ of dimension r. We choose the unique matrix $\in \mathbb{F}_q^{r\times m}$ in reduced row echelon form such that its rows form a basis of E.

3.2 Decoding failure rate

We prove simultaneously the correctness of our KEM and its decoding failure rate.

Algorithm 4 LowMS.Decaps

Input: Ciphertext $c = \boldsymbol{C} \in \mathbb{F}_{q^m}^{\ell \times (n-k)}$ and secret key $sk = (\boldsymbol{S}, \boldsymbol{H}, \boldsymbol{P}) \in \mathbb{F}_{q^m}^{(n-k)\times(n-k)} \times \mathbb{F}_{q^m}^{(n-k)\times n} \times \mathbb{F}_{q^m}^{n \times n}$
Output: Session key $K \in \mathbb{F}_2^{512}$

1: Compute $\boldsymbol{C}' = \boldsymbol{C}\boldsymbol{S}^{-1}$.
2: Recover $\boldsymbol{E}' = \boldsymbol{E}\boldsymbol{P} = \mathsf{InterleavedGab.Decode}(\boldsymbol{C}')$.
3: Compute $\boldsymbol{E} = \boldsymbol{E}'\boldsymbol{P}^{-1}$ and $E = \mathrm{Supp}(\boldsymbol{E})$
4: **return** $K = \mathcal{H}(E)$.

Proposition 3 (DFR). *The decoding failure rate (DFR) of our scheme is upper bounded by*

$$3.5q^{-m((\ell+1)(\frac{\ell}{\ell+1}(n-k)-r\lambda)+1)}.$$

Proof. We have

$$\begin{aligned} \boldsymbol{C}' &= \boldsymbol{C}\boldsymbol{S}^{-1} \\ &= \boldsymbol{E}\boldsymbol{H}'^{\top}\boldsymbol{S}^{-1} \\ &= \boldsymbol{E}\boldsymbol{P}\boldsymbol{H}^{\top} \\ &= \boldsymbol{E}'\boldsymbol{H}^{\top}, \end{aligned}$$

with $\boldsymbol{E}' = \boldsymbol{E}\boldsymbol{P}$ being the error matrix decoded by $\mathsf{InterleavedGab.Decode}$ (Algorithm 1). Each of its coordinates E'_{ij} is such that $E'_{ij} \in EF$, where E is the support of the coordinates of $\boldsymbol{E}$ and F is the support of the coordinates of $\boldsymbol{P}$.

The behaviour of a product matrix $\boldsymbol{E}\boldsymbol{P}$ was previously studied in the context of LRPC decoding. In [7, Proposition 2.4.3], the decoding failure rate calculation, validated by simulations, relies on the fact that a product of a vector with entries in E by a matrix with entries in F is a random vector with entries in EF. In [3, Theorem 1], it is shown that the support of a product matrix $\boldsymbol{E}\boldsymbol{P}$ has the same probability, up to a constant factor, of being equal to EF then a random matrix $\boldsymbol{E}'$ with entries in EF. Therefore we can reasonably make the assumption that every coordinate of $\boldsymbol{E}'$ is a random element of EF. We can then apply Corollary 1, the dimension of the error support being $t = r\lambda$. In our parameter sets, we were careful enough to fulfill inequalities $\ell \leq r\lambda \leq \lfloor \frac{\ell}{\ell+1}(n-k) \rfloor$, so that the conditions of Corollary 1 are met. We thus obtain the upper bound on the decryption failure rate. $\square$

3.3 Analysis of the difference with [45]

In this subsection, we try to present in the most understandable manner the difference with the approach of [45] which also suggests to interleave Loidreau's cryptosystem. The fine comprehension of this difference led us to build this new system with much more efficient parameters. The two main differing points concern the DFR and the error model.

Decoding failure rate. In [45], Theorem 6, the DFR is given by a complex formula which can be approximated by

$$\frac{4}{q^m}.$$

To ensure a negligible DFR, the value $q = 16$ has been chosen in [45]. Our formula seems more natural because it takes the value of ℓ into account, and therefore we are able to choose $q = 2$. This results in significantly more competitive parameters and also takes into account that for implementation reasons, cryptographic systems are usually preferred to work over binary fields.

Error model. Another key difference between this scheme and the one from [45] stems from the error model. To be more precise about this difference, let us recall some definitions presented in [45].

Definition 8 (Vector and matrix extension from $\mathbb{F}_{q^m}$ to $\mathbb{F}_q$). *Let $\gamma = (\gamma_1, ..., \gamma_m)$ be an ordered basis of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$. By utilizing the vector space isomorphism $\mathbb{F}_{q^m} \approxeq \mathbb{F}_q^m$, we can relate each vector $\boldsymbol{a} \in \mathbb{F}_{q^m}^n$ to a matrix $\boldsymbol{A} \in \mathbb{F}_q^{m\times n}$ according to $ext_\gamma : \mathbb{F}_{q^m}^n \to \mathbb{F}_q^{m\times n}, \boldsymbol{a} = (\boldsymbol{a}_1, ..., \boldsymbol{a}_n) \mapsto \boldsymbol{A}$ where $\boldsymbol{a}_j = \sum_{i=1}^m A_{i,j}\gamma_i, \forall j \in 1, \ldots, n$.*

Further, we extend the definition of ext_γ to matrices by extending each row and then vertically concatenating the resulting matrices.

Definition 9 (Vertical rank norm [45]). *The (vertical) rank norm $\mathrm{rank}_q(\boldsymbol{A})$ of a matrix $\boldsymbol{A} \in \mathbb{F}_{q^m}^{\ell\times n}$ is the rank of the γ-extension of $\boldsymbol{A}$:*

$$\mathrm{rank}_q(\boldsymbol{A}) := \mathrm{rank}(ext_\gamma(\boldsymbol{A})) = \mathrm{rank}\begin{pmatrix} ext_\gamma(\boldsymbol{A}_1) \\ \hline ... \\ \hline ext_\gamma(\boldsymbol{A}_\ell) \end{pmatrix},$$

where $\boldsymbol{A}_1, ..., \boldsymbol{A}_\ell$ are the rows of $\boldsymbol{A}$. Note that $ext_\gamma(\boldsymbol{A})$ is an $\mathbb{F}_q^{\ell m\times n}$ matrix.

The following table shows the difference between the RSL problem and the Interleaved search RSD problem (used in [45]) for the same parameters (q, m, n, k, ℓ, r).

Interleaved RSD	RSL
Given $(\boldsymbol{H}, \boldsymbol{Y}) \in \mathbb{F}_{q^m}^{(n-k)\times n} \times \mathbb{F}_{q^m}^{\ell\times(n-k)}$, find $\boldsymbol{E} \in \mathbb{F}_{q^m}^{\ell\times n}$ such that $\boldsymbol{H}\boldsymbol{E}^\top = \boldsymbol{Y}^\top$ and $\underline{\mathrm{rank}_q(\boldsymbol{E}) = r}$.	Given $(\boldsymbol{H}, \boldsymbol{Y}) \in \mathbb{F}_{q^m}^{(n-k)\times n} \times \mathbb{F}_{q^m}^{\ell\times(n-k)}$, find $\boldsymbol{E} \in \mathbb{F}_{q^m}^{\ell\times n}$ such that $\boldsymbol{H}\boldsymbol{E}^\top = \boldsymbol{Y}^\top$ and $\underline{\dim_{\mathbb{F}_q}(\mathrm{Supp}(\boldsymbol{E})) = r}$.

Trying to reconcile the two definitions even further, we found that the RSL problem corresponds to finding a syndrome matrix $\boldsymbol{E}$ such that $\overline{\mathrm{rank}}_q(\boldsymbol{E}) = r$ where $\overline{\mathrm{rank}}_q$ is an alternative definition of the rank norm obtained by **horizontally** concatenating when extending the definition of ext_γ to matrices over $\mathbb{F}_{q^m}$.

Definition 10 (Horizontal rank norm). *The horizontal rank norm* $\overline{\mathrm{rank}}_q(\boldsymbol{A})$ *of a matrix* $\boldsymbol{A} \in \mathbb{F}_{q^m}^{\ell \times n}$ *is the rank of the horizontal concatenation of the* γ*-extensions of rows in* $\boldsymbol{A}$*:*

$$\overline{\mathrm{rank}}_q(\boldsymbol{A}) := \mathrm{rank}(ext_\gamma(\boldsymbol{A}_1)|...|ext_\gamma(\boldsymbol{A}_\ell)),$$

where $\boldsymbol{A}_1, ..., \boldsymbol{A}_\ell$ *are the rows of* $\boldsymbol{A}$*.*

There is an easy correspondence between the horizontal rank norm and the support.

Proposition 4. *Let* $\boldsymbol{A} \in \mathbb{F}_{q^m}^{\ell \times n}$ *be such that* $\mathrm{Supp}(\boldsymbol{A}) = E$*. Then*

$$\overline{\mathrm{rank}}_q(\boldsymbol{A}) = \dim_{\mathbb{F}_q}(E).$$

Proof. The horizontal concatenation of the γ-extensions of rows in $\boldsymbol{A}$ gives a matrix whose columns are exactly the unfoldings of coefficients in $\boldsymbol{A}$:

$$\mathrm{rank}(ext_\gamma(A_{1,1}), \dots, ext_\gamma(A_{1,n}), ext_\gamma(A_{2,1}), \dots, ext_\gamma(A_{\ell,1}), \dots, ext_\gamma(A_{\ell,n})).$$

This is exactly the matrix of the coefficients of $\boldsymbol{A}$ in the basis γ so the dimension of E is exactly the dimension of the column space of the above matrix. $\square$

This being said, the difference between Interleaved RSD and RSL is only a matter of norm, as shown in the following table.

<u>Interleaved RSD</u>	<u>RSL</u>
Given $(\boldsymbol{H}, \boldsymbol{Y}) \in \mathbb{F}_{q^m}^{(n-k) \times n} \times \mathbb{F}_{q^m}^{\ell \times (n-k)}$, find $\boldsymbol{E} \in \mathbb{F}_{q^m}^{\ell \times n}$ such that $\boldsymbol{H}\boldsymbol{E}^\top = \boldsymbol{Y}^\top$ and $\underline{\mathrm{rank}_q(\boldsymbol{E}) = r}$.	Given $(\boldsymbol{H}, \boldsymbol{Y}) \in \mathbb{F}_{q^m}^{(n-k) \times n} \times \mathbb{F}_{q^m}^{\ell \times (n-k)}$, find $\boldsymbol{E} \in \mathbb{F}_{q^m}^{\ell \times n}$ such that $\boldsymbol{H}\boldsymbol{E}^\top = \boldsymbol{Y}^\top$ and $\underline{\overline{\mathrm{rank}}_q(\boldsymbol{E}) = r}$.

We state that using the horizontal rank norm (and thus RSL) for the error instead of the vertical rank norm is a better choice for cryptographic applications. We state below a few elements in support for this claim:

- The technique presented in [46] allows to decode an interleaved code of interleaving order $\ell \geq t$ where t is the *vertical* rank norm of the error. It penalizes the parameters of [45] by forcing to choose $\ell < t$. The decoding algorithm in [46] succeeds with negligible probability in the horizontal rank norm (see next subsection for details), hence opening the possibility of a higher interleaving order.
- The RSL problem has been studied for several years, and the recent algebraic attacks from [10] allow us to precisely compute the complexity of the RSL instances resulting from the chosen parameters.

– Because of the vertical rank norm, the error matrix $\boldsymbol{E}$ must be chosen in [45] as a product matrix $\boldsymbol{AB}$, which in turn implies high constraints, such as

$$\frac{n-k}{2\lambda} < d_E \leq t - \ell + 1,$$

where d_E is the minimum rank distance of the code spanned by the rows of $\boldsymbol{E}$. These constraints are lifted when using the horizontal rank norm.

Choosing the horizontal rank norm therefore allows a higher interleaving order ℓ and leads to better parameter sets (see Section 5).

3.4 Avoiding the Metzner-Kapturowski approach

The algorithm in [46] is an adaptation to the rank metric of the Metzner-Kapturowski approach [41] and constitutes a polynomial-time algorithm for decoding arbitrary linear interleaved codes of high-interleaving order.

As said earlier, the decoding algorithm works when the interleaved order satistifies $\ell \geq t$ where t is the *vertical* rank norm of the error. We thus need to study the *vertical* rank norm of our error matrix $\boldsymbol{E}$ (which is of *horizontal* norm r) and show that it is larger than ℓ with great probability.

Proposition 5. *Let* $E \xleftarrow{\$} \mathbf{Gr}(\mathbb{F}_{q^m}, r)$ *and* $\boldsymbol{E} \xleftarrow{\$} E^{\ell \times n}$*. Let* t *be the vertical rank norm of* $\boldsymbol{E}$*. We have*

$$\mathrm{Prob}(t \leq \ell) < q^{\ell^2 r + n\ell(1-r)}.$$

In order to prove this proposition, we first need the following result on the rank of random $\mathbb{F}_q$-matrices.

Lemma 1. *For a uniformly random* $\mathbb{F}_q$ *matrix* $\boldsymbol{M}$ *of size* $m \times n$ *with* $m \leq n$ *and for* $0 \leq i \leq m$*,* $\mathrm{Prob}(\mathrm{rank}(\boldsymbol{M}) \leq i) \leq q^{im+(i-m)n}$*.*

Proof. Let S be a subspace of $\mathbb{F}_{q^m}$ of dimension i. The number of such possible subspaces is $\begin{bmatrix} m \\ i \end{bmatrix}_q \leq q^{im}$.

For a uniformly random q-ary $m \times n$ matrix $\boldsymbol{M}$, since the n columns of $\boldsymbol{M}$ are independent random variables, $\mathrm{Prob}(\mathrm{Supp}(\boldsymbol{M}) \subset S) = q^{(i-m)n}$. Then:

$$\begin{aligned} \mathrm{Prob}(\mathrm{rank}(\boldsymbol{M}) \leq i) &\leq \mathrm{Prob}(\bigcup_S \mathrm{Supp}(\boldsymbol{M}) \subset S) \\ &\leq \sum_S \mathrm{Prob}(\mathrm{Supp}(\boldsymbol{M}) \subset S) \\ &\leq q^{im+(i-m)n}. \end{aligned}$$

□

We can now prove Proposition 5.

Proof (Proof of Proposition 5). Let $(e_1, ..., e_r)$ be a basis of E. We can complete it into a basis $\gamma := (e_1, ..., e_r, x_1, ..., x_{m-r})$ of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$. We will use γ to calculate the vertical rank norm, since it does not depend on the choice of the basis.

It is clear that when one picks at random a matrix $\boldsymbol{E} \stackrel{\$}{\leftarrow} E^{\ell \times n}$, then $ext_\gamma(\boldsymbol{E})$ writes as follows:

$$ext_\gamma(\boldsymbol{E}) = \begin{pmatrix} \boldsymbol{A}_1 \\ \hline \boldsymbol{0} \\ \hline ... \\ \hline \boldsymbol{A}_\ell \\ \hline \boldsymbol{0} \end{pmatrix},$$

with $\boldsymbol{A}_i \stackrel{\$}{\leftarrow} \mathbb{F}_q^{r \times n}$ being the unfoldings of the ℓ rows of $\boldsymbol{E}$ in the basis of E and the $\boldsymbol{0}$ blocks being of size $(m-r) \times n$.

The probability distribution of the rank of $ext_\gamma(\boldsymbol{E})$ is therefore identical to the distribution of the rank of a matrix $\boldsymbol{A} \stackrel{\$}{\leftarrow} \mathbb{F}_q^{\ell r \times n}$.

Finally we conclude with Lemma 1 applied with parameters $i = \ell$ and $m = \ell r$. $\square$

For all parameter sets presented in Section 5, the probability obtained with Proposition 5 is less than 2^{-1000}. We can consider that the threat of the Metzner-Kapturowski approach is avoided by design, and we do not need to take additional precautions when sampling the error matrix $\boldsymbol{E}$.

4 Security

4.1 Definitions

We define the IND-CPA-security of a KEM formally via the following experiment, where Encap_0 returns a valid pair c^*, K^*, and Encap_1 returns a valid c^* and a random K^*.

Indistinguishability under Chosen Plaintext Attack: This notion states that an adversary should not be able to efficiently guess which key is encapsulated.

$\mathbf{Exp}_{\mathcal{E},\mathcal{A}}^{\mathsf{ind}-b}(\lambda)$

1. $\mathsf{param} \leftarrow \mathsf{Setup}(1^\lambda)$
2. $(\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{KeyGen}(\mathsf{param})$
3. $(c^*, K^*) \leftarrow \mathsf{Encap}_b(\mathsf{pk})$
4. $b' \leftarrow \mathcal{A}(\mathtt{GUESS} : c^*, K^*)$
5. $\mathtt{RETURN}\ b'$

Definition 11 (IND-CPA **Security**). *A key encapsulation scheme KEM is* IND-CPA*-secure if for every PPT (probabilistic polynomial time) adversary $\mathcal{A}$, we have that*

$$\mathsf{Adv}_{\mathsf{KEM}}^{indcpa}(\mathcal{A}) := |\Pr[\mathsf{IND\text{-}CPA}_{\mathsf{real}}^{\mathcal{A}} \Rightarrow 1] - \Pr[\mathsf{IND\text{-}CPA}_{\mathsf{rand}}^{\mathcal{A}} \Rightarrow 1]|$$

is negligible.

4.2 IND-CPA proof

Theorem 1. *Under the hardness of the distorted Gabidulin codes indistinguishability (Problem 3) and Rank Support Learning (Problem 2), the KEM presented in Section 3 is IND-CPA secure (Definition 11) in the Random Oracle Model (ROM).*

Proof. We proceed in a sequence of games. The simulator starts from the real scheme. First we replace the public key by a random code instead of a distorted Gabidulin code, and then we use the ROM to solve the Rank Support Learning problem.

– We start with the game G_0: in this game we generate $\boldsymbol{H}$, $\boldsymbol{H}'$, $\boldsymbol{E}$ and $\boldsymbol{C}$ honestly.
– In game G_1 we replace $\boldsymbol{H}'$ by a parity-check matrix of a random $[n, k]$ code. From an adversary point of view, everything is identical, except the distribution on $\boldsymbol{H}'$ which is either generated at random or from a distorted Gabidulin code. Distinguishing between the two is an instance of $\mathsf{IND\text{-}Gab}_{q,m,n,k,\lambda}$ (see Problem 3), hence

$$\mathsf{Adv}_{\mathcal{A}}^{G_0} \leq \mathsf{Adv}_{\mathcal{A}}^{G_1} + \mathsf{Adv}_{\mathcal{A}}^{\mathsf{IND\text{-}Gab}}.$$

– In game G_2 we now replace $\mathcal{H}(E)$ by a random value r. By monitoring the calls the adversary makes to the random oracle, we can prove that the difference between G_1 and G_2 is solving the DRSL problem:

$$\mathsf{Adv}_{\mathcal{A}}^{G_1} \leq \mathsf{Adv}_{\mathcal{A}}^{G_2} + \mathsf{Adv}_{\mathcal{A}}^{\mathrm{DRSL}}.$$

In game G_2 everything is sampled independently from the secret values, which leads to the conclusion. □

4.3 Known attacks

Attacks against the RSD problem. There are two main types of attacks for solving the generic RSD problem: combinatorial attacks and algebraic attacks. For cryptographic parameters the best attacks are usually the recent algebraic attacks, but it may also depend on parameters, sometimes combinatorial attacks can be better.

Combinatorial attacks against RSD. The best combinatorial attacks for solving the RSD problem on a random $[n, k]$ code over $\mathbb{F}_{q^m}$ for a rank weight d as described in [6] have complexity (for ω the linear algebra exponent):

$$\min\{(n-k)^\omega m^\omega q^{(d-1)(k+1)}, (km)^\omega q^{d\lceil \frac{km}{n} \rceil - m}\}. \tag{1}$$

The first term of the min typically corresponds to the case where $m \geq n$, the second term corresponds to the case where $m \leq n$, but still it can happen that this term is better than the first one, when $m \geq n$ but close to n. A detailed description of the complexity of the second term is given in [6].

Algebraic attacks against RSD. The general idea of algebraic attacks is to rewrite an RSD instance as a system of multivariate polynomial equations and to find a solution to this system.

For a long time, algebraic attacks were less efficient than combinatorial ones. Recent results improved the understanding of these attacks. The best algebraic attacks against RSD can be found in [11] and have complexity (for ω the linear algebra exponent):

$$q^{ar} m \binom{n-k-1}{r} \binom{n-a}{r}^{\omega-1} \tag{2}$$

operations in $\mathbb{F}_q$. a is defined as the smallest integer such that the condition $m\binom{n-k-1}{r} \geq \binom{n-a}{r} - 1$ is fulfilled.

Attacks against the RSL *problem.* The difficulty of solving an instance of the $\mathrm{RSL}_{q,n,k,r,\ell}$ problem depends on the number ℓ of samples. Clearly, for $\ell = 1$, the RSL problem is exactly the RSD problem with parameters (q, n, k, r), which is probabilistically reduced to the NP-hard syndrome decoding problem in the Hamming metric in [27]. When $\ell \geqslant nr$, the RSL problem is reduced to linear algebra, as stated in [25] where this problem was first introduced.

This raises the question of the security of the RSL problem in the case $1 < \ell < nr$. In [25] the authors relate this problem to the one of finding a codeword of rank r in a code of same length and dimension containing q^ℓ words of this weight, and conjecture that the complexity of finding such a codeword gets reduced by at most a factor q^ℓ compared to the case $\ell = 1$. They also observe that in practice, the complexity gain seems lower, likely due to the fact that said codewords are deeply correlated.

There have been recent improvements on the complexity of the RSL problem. In [18] the authors show that the condition $\ell \leqslant kr$ should be met in order to avoid a subexponential attack, which is further improved in [14]: the authors show that the case $\ell > kr$ actually leads to a polynomial attack. Our proposed parameters all fulfill the condition $\ell < kr$.

The best known attacks on the RSL problem in the $\ell < kr$ regime are described in [14], improving upon [10]. In our case, the value ℓ of multiple syndromes is too few (at most 6) for these attacks to apply on our parameters. The best known combinatorial attack of [14, Section 5.3] does not impact on the security of our given parameters, nor does the best algebraic attack, which needs at least $n-k-r$ multiple syndromes to be applied [14, p. 22].

Attacks against the masking of Gabidulin codes. One of the key-points in the security reduction presented in Section 4.2 is the complexity of distinguishing the public-key pk, a.k.a $\boldsymbol{G'}$ in Algorithm 2 from a randomly generated $[n, k]$ matrix over $\mathbb{F}_{q^m}$. This precise problem was addressed in the paper [34].

To sum up the results, there are two ways to investigate the problem:

- If $\lambda(n-k) < n$, there exists a polynomial-time distinguisher, see [17]. Moreover, a decryption algorithm can be recovered in polynomial-time for $\lambda = 2, 3$, see [17,28] and exponential time for $\lambda > 4$, but with a complexity much less than expected to be suitable for encryption purposes [39]. Since in our parameter sets, the rate k/n is $1/2$ and $\lambda \geq 3$, we are not in that case.
- If $\lambda(n-k) \geq n$, then the best distinguisher to date is the one published in [16]. The exponential part corresponds to the enumeration of some constrained vector spaces and the polynomial term consists of the use of Wiedemann's algorithm. This gives

$$\mathcal{W}_{\text{Mask}} \geq m^3 n^5 R^3 (1+R) q^{m(\lambda-1)-\lambda n R(1-R)},$$

  where $R = k/n$ is the rate of the code.

5 Parameters

We give six sets of parameters (see Table 1): two sets for each security level $\eta \in \{128, 192, 256\}$. For each security level, we give an efficient parameter set with a smaller value of λ and a conservative parameter set with a higher value of λ.

The parameters are chosen following these steps in order:

- q is always equal to 2;
- the parameter r is chosen in a way to avoid RSD and RSL attacks. We need $r = 7$ for 128-bit security, $r = 8$ for 192-bit security and $r = 9$ for 256-bit security;
- the parameters n and k are chosen such that $k = n/2$ and $n - k$ is slightly larger than $r\lambda$, so as to respect the condition $r\lambda \leqslant \lfloor \frac{\ell}{\ell+1}(n-k) \rfloor$ with a reasonably small ℓ;
- m is set as the next prime after n;
- if needed, m and n are increased in order to have a complexity large enough for MaxMinors (algebraic attack from [11]) and $\mathcal{W}_{\text{Mask}}$. We always keep $k = n/2$ and m prime[2] larger than n;
- finally, parameter ℓ is chosen large enough so that the DFR is at most $2^{-\eta}$.

[2] We traditionally choose m prime to avoid any potential attacks.

The sizes of the proposed parameters are expressed in kilobytes. The public key is an $(n-k)\times n$ parity-check matrix with entries in $\mathbb{F}_{q^m}$ given in systematic form, therefore

$$pk \text{ size} = \log_2(q)mk(n-k) \text{ bits.}$$

The ciphertext consists of ℓ syndromes of $n-k$ entries in $\mathbb{F}_{q^m}$ each, therefore

$$ct \text{ size} = \log_2(q)m\ell(n-k) \text{ bits.}$$

For the DFR, MaxMinors and $\mathcal{W}_{\text{Mask}}$ columns, we chose to put the base 2 logarithm.

Security level 128											
q	m	n	k	λ	r	ℓ	pk size	ct size	DFR	MaxMinors	$\mathcal{W}_{\text{Mask}}$
2	61	50	25	3	7	6	4.77KB	1.14KB	-242	139	131
2	67	66	33	4	7	6	9.12KB	1.66KB	-199	155	183
Security level 192											
q	m	n	k	λ	r	ℓ	pk size	ct size	DFR	MaxMinors	$\mathcal{W}_{\text{Mask}}$
2	101	74	37	3	8	2	17.28KB	0.93KB	-301	193	197
2	79	78	39	4	8	5	15.02KB	1.93KB	-314	218	209
Security level 256											
q	m	n	k	λ	r	ℓ	pk size	ct size	DFR	MaxMinors	$\mathcal{W}_{\text{Mask}}$
2	101	88	44	4	9	5	24.44KB	2.78KB	-503	278	267
2	107	106	53	5	9	6	37.57KB	4.25KB	-426	313	349

Table 1. Parameters for LowMS

Comparison with other KEMs We compare our cryptosystem to other GPT-based KEMs, as well as to unstructured proposals, either lattice-based or code-based. Our comparison metric is the usual TLS-oriented communication size (public key + ciphertext). Although our scheme is only proven IND-CPA at this stage, we believe that, since our DFR is negligible, it can be turned to an IND-CCA scheme using the Fujisaki-Osamoto transform [21]. Indeed, when applying the HHK framework [29], similarly to [7, §5.3.2], the difference of advantages between CPA and CCA adversaries is explained by a term being equal to the product of the number of queries to the random oracle, by the probability of generating an decipherable ciphertext in an honest execution. With a negligible DFR, the advantages are thus similar. This comes at the cost of adding only two 64-byte hashes to the ciphertext and would only be a negligible increase, hence we took the liberty to compare our work with other IND-CCA parameters.

For the original Loidreau cryptosystem, we consider the parameters presented in the conclusion of [44] which take into account the recent improvements on algebraic attacks. For this cryptosystem, parameters were not available (N/A) for 192 bits of security.

Instance	128 bits	192 bits
LowMS ($\lambda = 3$)	5.76KB	14.97KB
LowMS ($\lambda = 4$)	10.78KB	16.95KB
DRANKULA [4]	28.8KB	N/A
Interleaved Loidreau [45]	33.35KB	N/A
Original Loidreau [37]	36.30KB	N/A

Table 2. Comparison of sizes of other GPT-based KEMs. The sizes represent the sum of the public key and the ciphertext expressed in bytes.

Instance	128 bits	192 bits
LowMS ($\lambda = 3$)	5.76KB	14.97KB
NH-Multi-UR-AG [14]	7.12KB	12.60KB
LRPC-MS [3]	7.21KB	14.27KB
LowMS ($\lambda = 4$)	10.78KB	16.95KB
Multi-UR-AG [14]	11.03KB	21.08KB
FrodoKEM [5]	19.34KB	31.38KB
Classic McEliece [13]	261KB	524KB

Table 3. Comparison of sizes of unstructured post-quantum KEMs. The sizes represent the sum of public key and ciphertext expressed in bytes.

6 Conclusion and perspectives

In this paper we presented LowMS, the shortest unstructured post-quantum lattice or code-based KEM so far, considering the sum of the public key and the ciphertext. We provided an IND-CPA proof for our scheme, whose security relies on the hardness of the DRSL and IND-Gab distinguishing problems.

The size could be optimized even further by adding some structure – using ideal Gabidulin codes for instance. However, we decided not to go along this path since any additional structure can potentially lead to an attack.

Data availability

Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

References

1. Carlos Aguilar Melchor, Nicolas Aragon, Slim Bettaieb, Loïc Bidoux, Olivier Blazy, Jean-Christophe Deneuville, Philippe Gaborit, and Gilles Zémor. Rank quasi cyclic (RQC). First round submission to the NIST post-quantum cryptography call, November 2017.
2. Carlos Aguilar Melchor, Nicolas Aragon, Slim Bettaieb, Loïc Bidoux, Olivier Blazy, Jean-Christophe Deneuville, Philippe Gaborit, Edoardo Persichetti, Gilles Zémor, and Jurjen Bos. HQC. Round 3 Submission to the NIST Post-Quantum Cryptography Call, June 2021. https://pqc-hqc.org/.
3. Carlos Aguilar-Melchor, Nicolas Aragon, Victor Dyseryn, Philippe Gaborit, and Gilles Zémor. Lrpc codes with multiple syndromes: near ideal-size KEMs without ideals. In *International Conference on Post-Quantum Cryptography*, pages 45–68. Springer, 2022.
4. Ameera Salem Al Abdouli, Mohamed Al Ali, Emanuele Bellini, Florian Caullery, Alexandros Hasikos, Marc Manzano, and Victor Mateu. DRANKULA: a McEliece-like rank metric based cryptosystem implementation. *Cryptology ePrint Archive*, 2018.
5. Erdem Alkim, Joppe W Bos, Léo Ducas, Patrick Longa, and Ilya Mironov. FrodoKEM. 3rd round submission to the NIST. 2021.
6. N. Aragon, P. Gaborit, A. Hauteville, and J.-P. Tillich. A new algorithm for solving the rank syndrome decoding problem. In *Proc. IEEE ISIT*, 2018.
7. Nicolas Aragon, Olivier Blazy, Jean-Christophe Deneuville, Philippe Gaborit, Adrien Hauteville, Olivier Ruatta, Jean-Pierre Tillich, Gilles Zémor, Carlos Aguilar Melchor, Slim Bettaieb, Loïc Bidoux, Bardet Magali, and Ayoub Otmani. ROLLO (merger of Rank-Ouroboros, LAKE and LOCKER). Second round submission to the NIST post-quantum cryptography call, March 2019.
8. Nicolas Aragon, Olivier Blazy, Philippe Gaborit, Adrien Hauteville, and Gilles Zémor. Durandal: a rank metric based signature scheme. In *Advances in Cryptology - EUROCRYPT 2019*, pages 728–758. Springer, 2019.
9. Nicolas Aragon, Philippe Gaborit, Adrien Hauteville, Olivier Ruatta, and Gilles Zémor. Low rank parity check codes: New decoding algorithms and applications to cryptography. *IEEE Transactions on Information Theory*, 65(12):7697–7717, 2019.
10. Magali Bardet and Pierre Briaud. An algebraic approach to the rank support learning problem. In Jung Hee Cheon and Jean-Pierre Tillich, editors, *Post-Quantum Cryptography*, lncs. Springer International Publishing, 2021.
11. Magali Bardet, Maxime Bros, Daniel Cabarcas, Philippe Gaborit, Ray Perlner, Daniel Smith-Tone, Jean-Pierre Tillich, and Javier Verbel. Improvements of algebraic attacks for solving the rank decoding and minrank problems. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 507–536. Springer, 2020.
12. Thierry P. Berger and Pierre Loidreau. Designing an efficient and secure public-key cryptosystem based on reducible rank codes. In *Progress in Cryptology - INDOCRYPT 2004*, volume 3348 of *LNCS*, pages 218–229, 2004.
13. Daniel J Bernstein, Tung Chou, Tanja Lange, Ingo von Maurich, Rafael Misoczki, Ruben Niederhagen, Edoardo Persichetti, Christiane Peters, Peter Schwabe, Nicolas Sendrier, et al. Classic McEliece. 2017.
14. Loïc Bidoux, Pierre Briaud, Maxime Bros, and Philippe Gaborit. RQC revisited and more cryptanalysis for rank-based cryptography. *arXiv preprint arXiv:2207.01410*, 2022.

15. Joppe Bos, Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, John M Schanck, Peter Schwabe, Gregor Seiler, and Damien Stehlé. CRYSTALS-Kyber: a CCA-secure module-lattice-based KEM. In *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 353–367. IEEE, 2018.
16. Pierre Briaud and Pierre Loidreau. Cryptanalysis of rank-metric schemes based on distorted gabidulin codes. In Thomas Johansson and Daniel Smith-Tone, editors, *Post-Quantum Cryptography - 14th International Workshop, PQCrypto 2023*, volume 14154 of *Lecture Notes in Computer Science*, pages 38–56. Springer, 2023.
17. Daniel Coggia and Alain Couvreur. On the security of a Loidreau rank metric code based encryption scheme. *Des. Codes Cryptogr.*, 88(9):1941–1957, 2020.
18. Thomas Debris-Alazard and Jean-Pierre Tillich. Two attacks on rank metric code-based schemes: Ranksign and an identity-based-encryption scheme. In *Advances in Cryptology - ASIACRYPT 2018*, 2018.
19. Molka Elleuch, Antonia Wachter-Zeh, and Alexander Zeh. A public-key cryptosystem from interleaved Goppa codes. *arXiv preprint arXiv:1809.03024*, 2018.
20. Cédric Faure and Pierre Loidreau. A new public-key cryptosystem based on the problem of reconstructing p-polynomials. In *Coding and Cryptography, International Workshop, WCC 2005*, 2005.
21. Eiichiro Fujisaki and Tatsuaki Okamoto. Secure integration of asymmetric and symmetric encryption schemes. In *Advances in Cryptology—CRYPTO'99: 19th Annual International Cryptology Conference Santa Barbara, California, USA, August 15–19, 1999 Proceedings*, pages 537–554. Springer, 1999.
22. Ernest M. Gabidulin. Theory of codes with maximum rank distance. *Problemy Peredachi Informatsii*, 21(1):3–16, 1985.
23. Ernst M. Gabidulin. Attacks and counter-attacks on the GPT public key cryptosystem. *Des. Codes Cryptogr.*, 48(2):171–177, 2008.
24. Ernst M. Gabidulin, A. V. Paramonov, and O. V. Tretjakov. Ideals over a non-commutative ring and their applications to cryptography. In *Advances in Cryptology - EUROCRYPT'91*, Brighton, 1991.
25. P. Gaborit, A. Hauteville, D. H. Phan, and J.-P. Tillich. Identity-based encryption from rank metric. In *Advances in Cryptology - CRYPTO*, 2017.
26. 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*, volume 2013, 2013.
27. Philippe Gaborit and Gilles Zémor. On the hardness of the decoding and the minimum distance problems for rank codes. *IEEE Trans. Inform. Theory*, 62(12):7245–7252, 2016.
28. Anirban Ghatak. Extending Coggia-Couvreur attack on Loidreau's rank-metric cryptosystem. *Des. Codes Cryptogr.*, 90(1):215–238, 2022.
29. Dennis Hofheinz, Kathrin Hövelmanns, and Eike Kiltz. A modular analysis of the Fujisaki-Okamoto transformation. In *Theory of Cryptography Conference*, pages 341–371. Springer, 2017.
30. Lukas Holzbaur, Hedongliang Liu, Sven Puchinger, and Antonia Wachter-Zeh. On Decoding and Applications of Interleaved Goppa Codes. In *IEEE Int. Symp. Inf. Theory (ISIT)*, July 2019.
31. Jon-Lark Kim, Young-Sik Kim, Lucky Erap Galvez, and Myeong Jae Kim. A modified Dual-Ouroboros public-key encryption using Gabidulin codes. *Applicable Algebra in Engineering, Communication and Computing*, 32(2):147–156, 2021.
32. Terry Shue Chien Lau, Chik How Tan, and Theo Fanuela Prabowo. On the security of the modified Dual-Ouroboros PKE using Gabidulin codes. *Applicable Algebra in Engineering, Communication and Computing*, 32(6):681–699, 2021.

33. Matthieu Legeay. Permutation decoding : Towards an approach using algebraic properties of the σ-subcode. In Daniel Augot and Anne Canteaut, editors, *WCC 2011*, pages 193–202, 2011.
34. Pierre Loidreau. Analysis of a public-key encryption scheme based on distorted Gabidulin codes. In *Proceedings of the twelth international workshop on coding and cryptography WCC 2022*. https://www.wcc2022.uni-rostock.de/storages/uni-rostock/Tagungen/WCC2022/Papers/WCC_2022_paper_5.pdf.
35. Pierre Loidreau. Decoding rank errors beyond the error-correcting capability. In *ACCT 2010, Tenth international workshop on Algebraic and Combinatorial Coding Theory*, 2006.
36. Pierre Loidreau. *Metrique rang et cryptographie*. HDR thesis, Université Pierre et Marie Curie-Paris VI, 2007.
37. Pierre Loidreau. A new rank metric codes based encryption scheme. In *Post-Quantum Cryptography 2017*, volume 10346 of *LNCS*, pages 3–17. Springer, 2017.
38. Pierre Loidreau and Raphael Overbeck. Decoding rank errors beyond the error-correction capability. In *Proceedings of the Tenth International Workshop on Algebraic and Combinatorial Coding Theory, ACCT-10*, pages 168–190, 2006.
39. Pierre Loidreau and Ba-Duc Pham. An analysis of Coggia-Couvreur attack on Loidreau's rank-metric public key encryption scheme in the general case. *CoRR*, abs/2112.12445, 2021.
40. Robert J. McEliece. *A Public-Key System Based on Algebraic Coding Theory*, pages 114–116. Jet Propulsion Lab, 1978. DSN Progress Report 44.
41. John J. Metzner and Edward J. Kapturowski. A General Decoding Technique Applicable to Replicated File Disagreement Location and Concatenated Code Decoding. *IEEE Trans. Inf. Theory*, 36(4):911–917, 1990.
42. Harald Niederreiter. Knapsack-type cryptosystems and algebraic coding theory. *Problems of Control and Information Theory*, 15(2):159–166, 1986.
43. Raphael Overbeck. Structural attacks for public key cryptosystems based on Gabidulin codes. *J. Cryptology*, 21(2):280–301, 2008.
44. Ba Duc Pham. *Étude et conception de nouvelles primitives de chiffrement fondées sur les codes correcteurs d'erreurs en métrique rang*. PhD thesis, Rennes 1, 2021.
45. Julian Renner, Sven Puchinger, and Antonia Wachter-Zeh. Interleaving Loidreau's rank-metric cryptosystem. In *2019 XVI International Symposium" Problems of Redundancy in Information and Control Systems"(REDUNDANCY)*, pages 127–132. IEEE, 2019.
46. Julian Renner, Sven Puchinger, and Antonia Wachter-Zeh. Decoding high-order interleaved rank-metric codes. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 19–24. IEEE, 2021.
47. Peter W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In S. Goldwasser, editor, *FOCS*, pages 124–134, 1994.
48. Vladimir Sidorenko, Lan Jiang, and Martin Bossert. Skew-feedback shift-register synthesis and decoding interleaved Gabidulin codes. *IEEE transactions on information theory*, 57(2):621–632, 2011.
49. Li-Ping Wang. Loong: a new IND-CCA-secure code-based KEM. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 2584–2588. IEEE, 2019.