RQC revisited and more cryptanalysis for Rank-based Cryptography

Loïc Bidoux, Pierre Briaud, Maxime Bros, Philippe Gaborit

RQC revisited and more cryptanalysis for Rank-based Cryptography

Loïc Bidoux[1], Pierre Briaud[2,3], Maxime Bros[4], and Philippe Gaborit[4]

[1] Cryptography Research Center, Technology Innovation Institute, Abu Dhabi, UAE
loic.bidoux@tii.ae
[2] Inria, 2 rue Simone Iff, 75012 Paris, France
[3] Sorbonne Universités, UPMC Univ Paris 06
pierre.briaud@inria.fr
[4] University of Limoges, CNRS, XLIM, UMR 7252, Limoges, France
{maxime.bros, philippe.gaborit}@unilim.fr

Abstract. We propose two main contributions: first, we revisit the encryption scheme Rank Quasi-Cyclic (RQC) [3] by introducing new efficient variations, in particular, a new class of codes, the Augmented Gabidulin codes; second, we propose new attacks against the Rank Support Learning (RSL), the Non-Homogeneous Rank Decoding (NHRSD), and the Non-Homogeneous Rank Support Learning (NHRSL) problems. RSL is primordial for all recent rank-based cryptosystems such as Durandal [7] or LRPC with multiple syndromes [1], moreover, NHRSD and NHRSL, together with RSL, are at the core of our new schemes.
The new attacks we propose are of both types: combinatorial and algebraic. For all these attacks, we provide a precise analysis of their complexity.
Overall, when all of these new improvements for the RQC scheme are put together, and their security evaluated with our different attacks, they enable one to gain 50% in parameter sizes compared to the previous RQC version. More precisely, we give very competitive parameters, around 11 KBytes, for RQC schemes with unstructured public key matrices. This is currently the only scheme with such short parameters whose security relies solely on pure random instances without any masking assumptions, contrary to McEliece-like schemes. At last, when considering the case of Non-Homogeneous errors, our scheme permits to reach even smaller parameters.

Keywords: Rank Metric · Encryption · Code-Based Cryptography · Gabidulin Codes.

1 Introduction

Background on rank metric code-based cryptography. In the last decade, rank metric code-based cryptography has evolved to become a real alternative to traditional code-based cryptography based on the Hamming metric. The original scheme based on rank metric was the GPT cryptosystem [18], an adaptation of

the McEliece scheme in a rank metric context where Gabidulin codes [17], a rank metric analogue of Reed-Solomon codes, were the masked codes. However, the strong algebraic structure of these codes was successfully exploited for attacking the original GPT cryptosystem and its variants with the Overbeck attack [34] (see [32] for the latest developments). This situation is similar to the Hamming metric where most of McEliece cryptosystems based on variants of Reed-Solomon codes have been broken.

Besides the McEliece scheme where a secret code is masked through using permutation, it is possible to generalize the approach by considering public key matrices with trapdoor. Examples of such an approach are NTRU [26] or MDPC [30] cryptosystems where the masking consists in knowing a very small weight vector of the given public matrix. Such an approach was adapted for rank metric through the introduction of LRPC codes [22], a rank metric analogue of MDPC.

The security of such type of cryptosystems relies on the general rank decoding problem together with the computational indistinguishability of the public key (a public matrix). The fact that the public matrix is used both for encryption and decryption, permits to obtain very efficient schemes, at the cost of an inversion. It is worth noticing that Loidreau's scheme, which uses homogeneous LRPC matrices in a McEliece context, seems to resist to structural attacks with an homogeneous matrix of sufficiently high enough rank [29].

The RQC scheme. Another approach, proposed by Aleknovich in [5], permits to rely solely on random instances of the Syndrome Decoding problem without any masking of a public key. However, such an approach is strongly inefficient in practice; a few years later a more optimized approach was proposed with the HQC scheme [4], relying on Quasi-Cyclic codes. It has been generalized to rank metric with the RQC scheme [3]. For these schemes, two type of codes are used: a first random double circulant code permits to ensure the security of the scheme when a second public code permits to decode/decrypt the ciphertext. In RQC, Gabidulin codes are used as public decryption codes. Besides RQC, some other variations were proposed in [21,37,20]. The main advantage of the RQC cryptosystem compared to the LRPC cryptosystem is the fact that its security reduction is done to random decoding instances whereas the LRPC approach requires another indistinguishability assumption; however, this advantage comes at a price since parameters are larger for RQC than for LRPC.

The RQC scheme was proposed to NIST Standardization Process with very competitive parameters but algebraic attacks of [11,12], which were published during the standardization process, had a dreadful impact on RQC parameters so that, in order not to increase too much RQC parameters, the introduction of non-homogeneous errors [3] permitted to limit the impact of these algebraic attacks.

The idea of non-homogeneous errors is to consider errors in three parts of length n such that the error weight is the same for the first two sets, but larger for the third one. Such an approach permits to limit the impact of the security reduction of RQC to decoding random $[3n, n]$ codes rather than $[2n, n]$ codes

in LRPC cryptosystems. The notion of non-homogeneous error led to the introduction of the Non Homogeneous RSD problem (NHRSD) and was a first approach to decrease the size of RQC parameters. At this point, it is meaningful to notice that for LRPC and RQC systems, the weight of the error to attack is structurally $\mathcal{O}(\sqrt{n})$ (where n is the length of the code), a type of parameters for which algebraic attacks are very efficient.

Besides the RSD problem, the RSL problem which consists in having N syndromes whose associated errors share the same support, was introduced in [21] to construct the RankPKE scheme and later in [37]. This problem which generalizes RSD is meaningful to give more margin in building cryptosystems; it has been recently used to improve on the LRPC schemes [1]. It permits, in particular, to increase the weight of the error to decode from $O(\sqrt{n})$ to a weight closer to the Rank Gilbert-Varshamov (RGV) bound; this is (of importance) since for that type of parameters, i.e. close to the RGV bound, algebraic attacks become relatively less efficient than combinatorial attacks.

Attacks and problems in rank metric. There are two types of attacks in rank metric. Combinatorial attacks which were the first to be introduced in the late 1990's then algebraic attacks ten years later. At first combinatorial attacks were the most efficient ones, but recently and especially for parameters where the error has weight $O(\sqrt{n})$ the seminal approaches of [11,12] permitted to have a strong impact on such parameters. Besides the RSD problem, the RSL problem was studied in [21] and [16] and more recently algebraic attacks were considered in [10]. In particular in the definition of the RSL problem in [21] it was shown that giving more than nr syndromes led to a combinatorial attack on the RSL problem. Moreover, the Non Homogeneous RSD (NHRSD) problem was introduced in [3] in which a first approach for algebraic attack was proposed.

Contributions. We saw in previous paragraphs, how before the present paper some new problem in rank metric had emerged (NHRSD, RSL) which permitted to improve on parameters both for RQC and LRPC systems.

In this paper our contributions are twofold: first we propose new variations on the RQC scheme in order to improve on parameters and second we study in details the new problems on which are based these approaches. All these problems NHRSD and RSL appear as natural variations on the RSD problems and are bound to be the future problems on which will be relying systems in rank metric.

New schemes. The new schemes we propose are based on three types of improvements:

Our first and main improvement, is the introduction of a new class decodable code, denoted by Augmented Gabidulin codes. These codes exploit the concept of *support erasure* in a rank-metric context. Compared to classical Gabidulin codes, the introduction of known support erasure permits to decrease the value m down to a value close to the weight of the error, whereas m had to be at least twice

bigger with classical Gabidulin codes. This comes at the cost of a probabilistic decoding; however the decryption failure rate (DFR) can be controlled very easily as it is done with LRPC in [22,1].

Second, as for the recent LRPC improvement [1] we consider the use of multiple syndromes in the RQC scheme. As for LRPC this approach permits to greatly improve the decoding capacity of the RQC scheme by increasing the information available for the decryption at a lower cost than directly increasing all parameters. This variation implies that the scheme relied on the RSL [21,37,7] rather than on the RSD problem. As for the new LRPC approach [22] this approach permits in particular to increase the weight of the error to decode, so that in practice reaching almost the RGV bound becomes possible (but with larger parameters).

Third, like pioneered in [3], we use a variant of the RSD problem by considering an error with non-homogeneous weight (w_1, w_2) which is the Non-Homogeneous Rank Syndrome Decoding problem (NHRSD). In short, this error contains a part of weight w_1 and a part of weight $w_1 + w_2$. This optimization allows one extra degree of freedom while choosing the target error weights, and this has a strong impact on the parameters.

In conclusion, we propose two types of scheme with very competitive sizes. First, Multi-RQC-AG has parameters similar to MS-LRPC [1], around 4.5 KBytes for the public key together with the ciphertext; its security relies on ideal-codes. Second, Multi-UR-AG has parameters a little bit larger than MS-LRPC, around 11 KBytes total, this time without any structure; more precisely, it relies only on pure random instances of the RSL problem. This is the most conservative security one could expect. For both of the aforementioned schemes, one could add a non-homogeneous structure in order to shorten the sizes down of 30%, this corresponds to our scheme: NH-Multi-RQC-AG and NH-Multi-UR-AG .

The scheme we propose without any ideal structure with small parameters of 11KBytes is meaningful, indeed since it is not proven that any ideal structure cannot be used to get faster attack with a quantum computer, scheme without any ideal structure may hence provide a better security. Of course in that case the size of parameters increases a lot but for rank metric our scheme shows that it remains small when for Hamming metric having no additional structure implies very large public key (see McEliece scheme for instance). Moreover our scheme does not necessitate any supplementary indistinguishability assumption. Moreover, our schemes compare very well with other code-based schemes.

New attacks and analysis. We saw that our new improvements on RQC relied on recent problems for rank metric, namely the NHRSD et RSL problems (and also a combination of the two latter problems the NHRSL problem). Although these schemes have begun to be considered we go deeper in their study by proposing new attacks and adaptation of known attacks for the security evaluation of these problems. The motivation comes both from the general interest of these problems for rank based cryptography and for the new schemes that we introduce in this paper.

More precisely, recall that an RSL (m, n, k, r, N) instance is like a rank syndrome decoding instance of parameters (m, n, k, r) where N instead of 1 syndromes are given, and all their associated errors share the same support. The security of RSL is inherent to the value of N, and it is known since [21] that the problem can be solved in polynomial time as long as $N > nr$. Our contributions are then the following:

- With our new combinatorial attack against RSL, first we improve on the most recent algebraic attack for some instances; most importantly, we improve the aforementioned bound, unchanged since 2017 [21], showing that RSL becomes polynomial as long as $N > kr\frac{m}{m-r}$.
- We also propose the first combinatorial attack against NHRSD, together with a precise complexity analysis of the algebraic attack, still against NHRSD, described in [3].
- Finally, we propose an attack against NHRSL. That it is to say that we were able to take advantage of two structure in the same attack: the fact that the error is non-homogeneous and that one is given several syndromes.

2 Preliminaries

2.1 Coding theory and rank metric

Let q be a prime power, let m a positive integer, let $\mathbb{F}_{q^m}$ an extension of degree m of $\mathbb{F}_q$ and let $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_m)$ be an $\mathbb{F}_q$-basis of $\mathbb{F}_{q^m}$. Any vector in $\mathbb{F}_{q^m}^n$ can naturally be viewed as a matrix in $\mathbb{F}_q^{m\times n}$ by expressing its coordinates in $\boldsymbol{\beta}$.

Definition 1 (Rank weight). *Let $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}_{q^m}^n$ be a vector. The rank weight of $\mathbf{x}$ denoted $\|\mathbf{x}\|$ is defined as the rank of the matrix $\mathrm{Mat}(\mathbf{x}) := (x_{ij})_{i,j} \in \mathbb{F}_q^{m\times n}$ where $x_j = \beta_1 x_{1j} + \cdots + \beta_m x_{mj}$ for $j \in \{1..n\}$. The set of vectors of weight w in $\mathbb{F}_{q^m}^n$ is denoted $\mathcal{S}_w^n(\mathbb{F}_{q^m})$.*

Definition 2 (Support). *The support of $\mathbf{x} \in \mathbb{F}_{q^m}^n$ is the $\mathbb{F}_q$-linear space generated by the coordinates of $\mathbf{x}$,* i.e. *$Supp(\mathbf{x}) := \langle x_1, \ldots, x_n\rangle_{\mathbb{F}_q}$. It follows from the definition that $\|\mathbf{x}\| = \dim_{\mathbb{F}_q}(Supp(\mathbf{x}))$.*

These notions can be extended to matrices. The support of a matrix $\mathbf{M} \in \mathbb{F}_{q^m}^{r\times c}$ denoted $\mathrm{Supp}(\mathbf{M})$ is the $\mathbb{F}_q$-vector space spanned by all its $r \times c$ entries, and the rank weight $\|\mathbf{M}\|$ is defined as the dimension of this support. Note that we always have $\|\mathbf{M}\| \leq \mathrm{Rank}(\mathbf{M})$, for example $\|\mathbf{I}_n\| = 1$ while $\mathrm{Rank}(\mathbf{I}_n) = n$.

Definition 3 ($\mathbb{F}_{q^m}$-linear code). *An $\mathbb{F}_{q^m}$-linear code $\mathcal{C}$ of length n and dimension k is an $\mathbb{F}_{q^m}$-linear subspace of $\mathbb{F}_{q^m}^n$ of dimension k. We say that it has parameters $[n, k]_{q^m}$. A generator matrix for $\mathcal{C}$ is a full-rank matrix $\boldsymbol{G} \in \mathbb{F}_{q^m}^{k\times n}$ such that $\mathcal{C} = \left\{\mathbf{m}\boldsymbol{G},\ \mathbf{m} \in \mathbb{F}_{q^m}^k\right\}$. A parity-check matrix is a full-rank matrix $\boldsymbol{H} \in \mathbb{F}_{q^m}^{(n-k)\times n}$ such that $\mathcal{C} = \left\{\mathbf{x} \in \mathbb{F}_{q^m}^n,\ \boldsymbol{H}\mathbf{x}^\intercal = 0\right\}$. Finally, the rowspace of $\boldsymbol{H}$ is a basis of the dual code $\mathcal{C}^\perp$.*

The use of $\mathbb{F}_{q^m}$-linear codes instead of standard $\mathbb{F}_q$-linear codes permits to obtain a more compact description for the public key in code-based cryptosystems. Another classical way to reduce the keysize is to use some type of cyclic structure, which leads to the notion of ideal codes. Let $P \in \mathbb{F}_q[X]$ denote a polynomial of degree n. The linear map $\mathbf{u} := (u_0, \ldots, u_{n-1}) \mapsto \mathbf{u}(X) := \sum_{i=0}^{n-1} u_i X^i$ is a vector space isomorphism between $\mathbb{F}_{q^m}^n$ and $\mathbb{F}_{q^m}[X]/\langle P \rangle$, and we use it to define a product between two elements $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{F}_{q^m}^n$ via $\mathbf{u} \cdot_P \mathbf{v} := \mathbf{u}(X)\mathbf{v}(X) \bmod P$. Note that we have

$$\mathbf{u} \cdot \mathbf{v} = \left(\sum_{i=0}^{n-1} u_i X^i \right) \mathbf{v}(X) \bmod P = \sum_{i=0}^{n-1} u_i \left(X^i \mathbf{v}(X) \bmod P \right),$$

so that the product by $\mathbf{v} \in \mathbb{F}_{q^m}^n$ can be seen as a matrix-vector product by the so-called ideal matrix generated by $\mathbf{x}$ and P.

Definition 4 (Ideal matrix). *Let $P \in \mathbb{F}_q[X]$ a polynomial of degree n and let $\mathbf{v} \in \mathbb{F}_{q^m}^n$. The ideal matrix generated by $\mathbf{v}$ and P, noted $\mathcal{IM}_P(\mathbf{v})$, is the $n \times n$ matrix, with entries in $\mathbb{F}_{q^m}$, and whose rows are the following: $\mathbf{v}(X) \bmod P$, $X\mathbf{v}(X) \bmod P$, ..., $X^{n-1}\mathbf{v}(X) \bmod P$.*

For conciseness, we use the notation $\mathcal{IM}(\mathbf{v})$ since there will be no ambiguity in the choice of P in the paper.

One can see that $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}\mathcal{IM}(\mathbf{v}) = \mathbf{v}\mathcal{IM}(\mathbf{u}) = \mathbf{v} \cdot \mathbf{u}$. An ideal code $\mathcal{C}$ of parameters $[sn, tn]_{q^m}$ is an $\mathbb{F}_{q^m}$-linear code which admits a generating matrix made of $s \times t$ ideal matrix blocks in $\mathbb{F}_{q^m}^{n \times n}$. A crucial point regarding the choice of the modulus P (see [3, Lemma 1]) is that if $P \in \mathbb{F}_q[X]$ is *irreducible* of degree n and if n and m are *prime*, then such a code $\mathcal{C}$ always admits a *systematic* generator matrix made of ideal blocks. Hereafter, we only consider $t = 1$.

Definition 5 (Ideal codes). *Let $P \in \mathbb{F}_q[X]$ a polynomial of degree n. An $[ns, n]_{q^m}$-code $\mathcal{C}$ is an ideal code if it has a generator matrix of the form $\boldsymbol{G} = \left(\boldsymbol{I}_n \;\; \mathcal{IM}(\mathbf{g}_1) \;\; \ldots \;\; \mathcal{IM}(\mathbf{g}_{s-1})\right) \in \mathbb{F}_{q^m}^{n \times ns}$, where $\mathbf{g}_i \in \mathbb{F}_{q^m}^n$ for $1 \leq i \leq s-1$. Similarly, $\mathcal{C}$ is an ideal code if it admits a parity-check matrix of the form*

$$\boldsymbol{H} = \begin{pmatrix} & \mathcal{IM}(\mathbf{h}_1)^\mathsf{T} \\ \boldsymbol{I}_{n(s-1)} & \vdots \\ & \mathcal{IM}(\mathbf{h}_{s-1})^\mathsf{T} \end{pmatrix} \in \mathbb{F}_{q^m}^{n(s-1) \times ns}.$$

2.2 Gabidulin codes

Gabidulin codes were introduced by Gabidulin in 1985 [17]. These codes can be seen as the rank metric analogue of Reed-Solomon codes [36], where standard polynomials are replaced by q-polynomials (also called Ore polynomials or linearized polynomials).

Definition 6 (q-polynomials). *The set of q-polynomials over $\mathbb{F}_{q^m}$ is the set of polynomials with the following shape:*

$$\left\{ P(X) = \sum_{i=0}^{r} p_i X^{q^i}, \text{ with } p_i \in \mathbb{F}_{q^m} \text{ and } p_r \neq 0 \right\}.$$

The q-degree of a q-polynomial P is defined as $\deg_q(P) = r$.

Definition 7 (Ring structure). *The set of q-polynomials over $\mathbb{F}_{q^m}$ is a non-commutative ring for $(+, \circ)$, where $\circ$ is the composition of $\mathbb{F}_q$-linear endomorphisms.*

Due to their structure, the q-polynomials are inherently related to decoding problems in the rank metric as stated by the following propositions.

Theorem 1 ([31]). *Any $\mathbb{F}_q$-subspace of $\mathbb{F}_{q^m}$ of dimension r is the set of the roots of a unique monic q-polynomial P such that* $\deg_q(P) = r$.

Corollary 1. *Let $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{F}_{q^m}^n$ and V be the monic q-polynomial of smallest q-degree such that $V(x_i) = 0$ for $1 \leq i \leq n$, then $\|\mathbf{x}\| = r$ if and only if* $\deg_q(V) = r$.

Finally, Gabidulin codes can be seen as the evaluation of q-polynomials of bounded degree on the coordinates of a fixed vector over $\mathbb{F}_{q^m}$.

Definition 8 (Gabidulin codes). *Let $k, n, m \in \mathbb{N}$ such that $k \leqslant n \leqslant m$ and let $\mathbf{g} = (g_1, \ldots, g_n)$ be an-$\mathbb{F}_q$ linearly independent family of elements of $\mathbb{F}_{q^m}$. The* Gabidulin code $\mathcal{G}_{\mathbf{g}}(n, k, m)$ *is the code of parameters $[n, k]_{q^m}$ defined by*

$$\mathcal{G}_{\mathbf{g}}(n, k, m) := \left\{ P(\mathbf{g}), \ \deg_q(P) < k \right\}, \text{where } P(\mathbf{g}) := (P(g_1), \ldots, P(g_v)).$$

2.3 Hard problems in rank-based cryptography

As in the Hamming metric, the main source of computational hardness for rank-based cryptosystems is a decoding problem. More precisely, it is the decoding problem in the rank metric setting restricted to $\mathbb{F}_{q^m}$-linear codes which is called the Rank Syndrome Decoding Problem (RSD).

Problem 1 (RSD Problem, Search) *Given $\boldsymbol{H} \in \mathbb{F}_{q^m}^{(n-k)\times n}$, a full rank parity-check matrix for a random $\mathbb{F}_{q^m}$-linear code $\mathcal{C}$, an integer $w \in \mathbb{N}$ and a syndrome $\mathbf{s} \in \mathbb{F}_{q^m}^{n-k}$, the Rank Syndrome Decoding problem RSD(m, n, k, w) asks to find $\mathbf{e} \in \mathbb{F}_{q^m}^n$ such that $\|\mathbf{e}\| = w$ and $\boldsymbol{H}\mathbf{e}^\mathsf{T} = \mathbf{s}^\mathsf{T}$.*

The decision version is denoted by DRSD. Even if RSD is not known to be NP-complete, there exists a randomized reduction from RSD to an NP-complete problem, namely to decoding in the Hamming metric [25]. Also, the average number of solutions for a fixed weight w is given by the following Gilbert-Varshamov bound for the rank metric:

Definition 9 (Rank Gilbert-Varshamov bound). *The Gilbert-Varshamov bound $w_{GV}(q,m,n,k)$ for $\mathbb{F}_{q^m}$-linear codes of length n and dimension k in the rank metric is defined as the smallest positive integer t such that $q^{m(n-k)} \leq B_t$, where $B_t := \sum_{j=0}^{t} \left(\prod_{\ell=0}^{j-1}(q^n - q^\ell)\right) \begin{bmatrix} m \\ j \end{bmatrix}_q$ is the size of the ball of radius t in the rank metric.*

In other words, it means that, with overwhelming probability, as long as $w \leq w_{GV}(q,m,n,k)$, a random RSD instance will have at most a unique solution. In this paper, we also focus on a slightly less standard assumption which is the NHRSD problem. This RSD variant was proposed in the Second Round update of RQC [3] in order to mitigate the impact of the recent algebraic RSD attacks [11,12] on the choice of the parameters. In NHRSD, the error $\mathbf{e}$ is no longer a random low weight vector but instead a vector with a *non-homogeneous* weight:

Problem 2 (NHRSD Problem, Search) *Given $\boldsymbol{H} \in \mathbb{F}_{q^m}^{(n+n_1)\times(2n+n_1)}$, a full rank parity-check matrix of a random $\mathbb{F}_{q^m}$-linear code $\mathcal{C}$ of parameters $[2n+n_1, n]_{q^m}$, integers $(w_1, w_2) \in \mathbb{N}^2$, and a syndrome $\mathbf{s} \in \mathbb{F}_{q^m}^{n+n_1}$, the Non-Homogeneous Rank Syndrome Decoding problem $\mathsf{NHRSD}(m,n,n_1,w_1,w_2)$ asks to find a vector $\mathbf{e} = (\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3) \in \mathbb{F}_{q^m}^{2n+n_1}$ such that $\|(\mathbf{e}_1, \mathbf{e}_3)\| = w_1$, $\|\mathbf{e}_2\| = w_1 + w_2$, $\mathrm{Supp}(\mathbf{e}_1, \mathbf{e}_3) \subset \mathrm{Supp}(\mathbf{e}_2)$, and such that $\boldsymbol{H}\mathbf{e}^\mathsf{T} = \mathbf{s}^\mathsf{T}$.*

We denote by DNHRSD the corresponding decisional version. Note that Definition 2 is slightly more general than the one of [3] where it is assumed that $n = n_1$. Finally, recall that one of our improvements on the RQC scheme uses multiple syndromes which are correlated since they correspond to errors which share the same support. This formulation exactly corresponds to the definition of the Rank Support Learning problem (RSL and DRSL for the decision version). This problem can be seen as the rank metric analogue of the Support-Learning problem in the Hamming metric [27,33].

Problem 3 (RSL Problem, Search) *Given $(\boldsymbol{H}, \boldsymbol{H}\boldsymbol{E}^\mathsf{T})$, where $\boldsymbol{H} \in \mathbb{F}_{q^m}^{(n-k)\times n}$ is of full-rank, and $\boldsymbol{E} \in \mathbb{F}_{q^m}^{N\times n}$ has all its entries lying in a subspace $\mathcal{V} \subset \mathbb{F}_{q^m}$ of dimension $w \in \mathbb{N}$, the Rank Support Learning problem $\mathsf{RSL}(m,n,k,w,N)$ asks to find the secret subspace $\mathcal{V}$.*

RSL may enable the construction of more advanced cryptographic primitives in the rank metric. It was introduced in [21], and is at the core of the Durandal signature scheme [7], and the recent Multi-LRPC proposal [1]. Naturally, it is possible to somehow combine the error distributions from Problems 2 and 3.

Problem 4 (NHRSL Problem, Search) *Given $(\boldsymbol{H}, \boldsymbol{H}\boldsymbol{E}^\mathsf{T})$, where $\boldsymbol{H}$ is a $(n+n_1)\times(2n+n_1)$ matrix of full rank, and $\boldsymbol{E} \in \mathbb{F}_{q^m}^{N\times(2n+n_1)}$ such that $\mathbf{e}_i = \boldsymbol{E}_{i,*} = (\mathbf{e}_{i,1}, \mathbf{e}_{i,2}, \mathbf{e}_{i,3}) \in \mathbb{F}_{q^m}^{(2n+n_1)}$, $\|(\mathbf{e}_{i,1}, \mathbf{e}_{i,3})\| = w_1$, $\|\mathbf{e}_{i,2}\| = w_1 + w_2$, and such that the supports $\mathcal{V} := \mathrm{Supp}(\mathbf{e}_{i,1}, \mathbf{e}_{i,3}) \subset \mathcal{W} := \mathrm{Supp}(\mathbf{e}_{i,2})$ are independent of i; the Non-Homogeneous Rank Support Learning problem $\mathsf{NHRSL}(m,n,n_1,w_1,w_2,N)$ asks to find the secret subspaces $\mathcal{V}$ and $\mathcal{W}$.*

Finally, both classic RQC and our Multi-RQC-AG proposal involve ideal codes, so that we have to consider the ideal versions of Problems 1, 2, 3, and 4 (denoted by, respectively, IRSD, NHIRSD, IRSL, and NHIRSL). For the sake of conciseness, we do not give a formal definition of these ideal variants.

2.4 RQC scheme

On Figure 1 we briefly recall the classical RQC scheme [3], for which one needs the following notation:

$$\mathcal{S}^{n}_{w,1}(\mathbb{F}_{q^m}) = \{\mathbf{x} \in \mathbb{F}^{n}_{q^m} : \|\mathbf{x}\| = w, 1 \in \mathrm{Supp}(\mathbf{x})\},$$
$$\mathcal{S}^{3n}_{(w_1,w_2)}(\mathbb{F}_{q^m}) = \{\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \in \mathbb{F}^{3n}_{q^m} : \|(\mathbf{x}_1, \mathbf{x}_3)\| = w_1, \|\mathbf{x}_2\| = w_1 + w_2,$$
$$\mathrm{Supp}(\mathbf{x}_1, \mathbf{x}_3) \subset \mathrm{Supp}(\mathbf{x}_2)\}.$$

$\underline{\mathsf{Setup}(1^\lambda)}$: Generates and outputs $\mathsf{param} = (n, k, \delta, w, w_1, w_2, P)$ where $P \in \mathbb{F}_q[X]$ is an irreducible polynomial of degree n.

$\underline{\mathsf{KeyGen}(\mathsf{param})}$: Samples $\mathbf{h} \xleftarrow{\$} \mathbb{F}^{n}_{q^m}$, $\mathbf{g} \xleftarrow{\$} \mathcal{S}^{n}_{n}(\mathbb{F}_{q^m})$ and $(\mathbf{x}, \mathbf{y}) \xleftarrow{\$} \mathcal{S}^{2n}_{w,1}(\mathbb{F}_{q^m})$, computes the generator matrix $\mathbf{G} \in \mathbb{F}^{k\times n}_{q^m}$ of a code $\mathcal{C}$, sets $\mathsf{pk} = (\mathbf{g}, \mathbf{h}, \mathbf{s} = \mathbf{x} + \mathbf{h} \cdot \mathbf{y} \mod P)$ and $\mathsf{sk} = (\mathbf{x}, \mathbf{y})$, returns $(\mathsf{pk}, \mathsf{sk})$.

$\underline{\mathsf{Encrypt}(\mathsf{pk}, \mathbf{m}, \theta)}$: Uses randomness θ to generate $(\mathbf{r}_1, \mathbf{e}, \mathbf{r}_2) \xleftarrow{\$} \mathcal{S}^{3n}_{(w_1,w_2)}(\mathbb{F}_{q^m})$, sets $\mathbf{u} = \mathbf{r}_1 + \mathbf{h} \cdot \mathbf{r}_2 \mod P$ and $\mathbf{v} = \mathbf{mG} + \mathbf{s} \cdot \mathbf{r}_2 + \mathbf{e} \mod P$, returns $\mathbf{c} = (\mathbf{u}, \mathbf{v})$.

$\underline{\mathsf{Decrypt}(\mathsf{sk}, \mathbf{c})}$: Returns $\mathcal{C}.\mathsf{Decode}(\mathbf{v} - \mathbf{u} \cdot \mathbf{y} \mod P)$.

Fig. 1. Description the RQC PKE scheme.

3 Augmented Gabidulin code: a new family of efficiently decodable codes for cryptography

In what follows, we introduce a new family of efficiently decodable codes, namely Augmented Gabidulin codes. The main idea behind these codes is to add a sequence of zeros at the end of the Gabidulin codes; by doing this, one directly gets elements of the support of the error, which correspond to *support erasure* in a rank metric context. The decoding of this code corresponds to the decoding of a classical Gabidulin code to which support erasures are added. In practice, this approach permits to decrease the size of m at the cost of having a probabilistic decoding. The probability of decoding failure can then be easily controlled at the cost of sacrifying only a few support erasures; indeed, for these codes, the decoding failure probability decreases exponentially fast, with a quadratic exponent, see Equation (1).

This approach is then especially suitable in the case where many errors have to be corrected which exactly corresponds to our case where the code we want to decode has a very low rate.

In what follows, we give a definition of augmented Gabidulin codes, and for didactic purpose, in Proposition 1, we recall a simple and natural way to decode Gabidulin code with support erasures and we give their decoding failure rate. For other or more efficient approaches, the reader may refer to [9,15,19].

Notice that this type of approach (adding zeros) is not relevant in Hamming metric since the errors are independent in a classical noisy canal, whereas in rank metric, errors located on different coordinates are linked since they share the same support.

Definition 10 (Augmented Gabidulin codes). *Let $(k, n, n', m) \in \mathbb{N}^4$ such that $k \leq n' < m < n$. Let $\mathbf{g} = (g_1, \ldots, g_{n'})$ be an $\mathbb{F}_q$- linearly independent family of n' elements of $\mathbb{F}_{q^m}$ and let $\overline{\mathbf{g}}$ be the vector of length n which is equal to $\mathbf{g}$ padded with $n - n'$ extra zeros on the right. The* Augmented Gabidulin code *$\mathcal{G}^+_{\overline{\mathbf{g}}}(n, n', k, m)$ is the code of parameters $[n, k]_{q^m}$ defined by*

$$\mathcal{G}^+_{\overline{\mathbf{g}}}(n, n', k, m) := \left\{ P(\overline{\mathbf{g}}),\ \deg_q(P) < k \right\},$$

where $P(\overline{\mathbf{g}}) := (P(g_1), \ldots, P(g_{n'}), 0, \ldots, 0)$.

Proposition 1 (Decoding capacity of Augmented Gabidulin codes). *Let $\mathcal{G}^+_{\overline{\mathbf{g}}}(n, n', k, m)$ be an augmented Gabidulin code, and let*

$$\varepsilon \in \{1, 2, \ldots, \min(n - n', n' - k)\}$$

be the dimension of the vector space generated by the support erasures.

Then, $\mathcal{G}^+_{\overline{\mathbf{g}}}(n, n', k, m)$ can uniquely decode an error of rank weight up to

$$t := \left\lfloor \frac{n' - k + \varepsilon}{2} \right\rfloor.$$

Proof. The minimal distance of $\mathcal{G}^+_{\overline{\mathbf{g}}}(n, n', k, m)$ is clearly $d = n' - k + 1$ since it is made of a Gabiudlin code augmented with zeros.

Let $x = c_1 + e_1$ be a noisy codeword where $c_1 \in \mathcal{G}^+_{\overline{\mathbf{g}}}$, and $\|e_1\| \leq t$.

Let us assume that x is not uniquely decodable to find a contradiction.

If x is not uniquely decodable, it means that there exists c_2 in $\mathcal{G}^+_{\overline{\mathbf{g}}}$ such that $c_2 \neq c_1$ and $x = c_2 + e_2$ where $\|e_2\| \leq t$.

Recall that we assume that one knows support erasures which span a vector space of dimension ε. These support erasures come from the $n - n'$ last coordinates of the code $\mathcal{G}^+_{\overline{\mathbf{g}}}$, thus these support elements are common to e_1 and e_2. Since $\mathrm{Supp}(e_1)$ and $\mathrm{Supp}(e_2)$ share ε elements, one has that

$$\mathrm{d}(e_1, e_2) \leq 2(t - \varepsilon) + \varepsilon = 2t - \varepsilon \leq n' - k.$$

Since $x = c_1 + e_1 = c_2 + e_2$, one clearly has that $\mathrm{d}(c_1, c_2) = \mathrm{d}(e_1, e_2)$, thus $\mathrm{d}(c_1, c_2) \leq n' - k$, which is a contradiction.

Thus, $\mathcal{G}^+_{\overline{\mathbf{g}}}$ can uniquely decode errors of rank weight up to $t := \left\lfloor \frac{n'-k+\varepsilon}{2} \right\rfloor$.

Finally, the condition $1 \leq \varepsilon \leq \min(n - n', n' - k)$ comes from the fact that the dimension of the vector space spanned by support erasures can not exceed

the maximum rank weight of the error nor the number of zero coordinates of the augmented Gabidulin code; in other words, on one hand ε is clearly smaller than $n - n'$, and on the other hand

$$\varepsilon \leq \left\lfloor \frac{n' - k + \varepsilon}{2} \right\rfloor \implies 2\varepsilon \leq n' - k + \varepsilon$$
$$\implies \varepsilon \leq n' - k.$$

□

Proposition 2 (Decoding Algorithm for Augmented Gabidulin codes). *Let $\mathcal{G}^{+}_{\mathbf{g}}(n, n', k, m)$ be an augmented Gabidulin code, and let*

$$\varepsilon \in \{1, 2, \ldots, \min(n - n', n' - k)\}$$

be the dimension of the vector space generated by the support erasures.

This code benefits from an efficient decoding algorithm correcting errors of rank weight up to $\delta := \left\lfloor \frac{n'-k+\varepsilon}{2} \right\rfloor$ with a decryption failure rate (DFR) of

$$1 - \frac{1}{\delta(n - n')} \sum_{i=\varepsilon}^{\delta} \prod_{j=0}^{\varepsilon-1} \frac{(q^{\delta} - q^{j})(q^{n-n'} - q^{j})}{q^{\varepsilon} - q^{j}}. \tag{1}$$

Proof. The proof gives the decoding algorithm. One is given a noisy encoded word $\overline{\mathbf{y}} = \overline{\mathbf{c}} + \overline{\mathbf{e}} \in \mathbb{F}_{q^m}^n$ where $\overline{\mathbf{c}} := \mathbf{x}\boldsymbol{G}$ belongs to $\mathcal{G}^{+}_{\mathbf{g}}(n, n', k, m)$ and $\|\mathbf{e}\| \leq \delta$.

Step 1: recovering a part of the error support. By construction we have $\overline{\mathbf{c}} = (*|0 \ldots 0)$, so that the last $n - n'$ coordinates of $\mathbf{y}$ are exactly the last coefficients of $\overline{\mathbf{e}}$. Thus, one may use these coefficients to recover ε elements in $E := \mathrm{Supp}(\overline{\mathbf{e}})$. This will be doable as long as these $n - n'$ coefficients contain at least ε linearly independent ones. The converse probability is the probability that a random $\delta \times (n - n')$ matrix with coefficients in $\mathbb{F}_q$ has rank less than ε. This yields to the probability given by Equation (1).

Step 2: recovering $\overline{\mathbf{c}}$. Assume now that ε elements in the support of $\overline{\mathbf{e}}$ are known and let E_2 be the vector space spanned by these elements. In what follows, we focus on the first n' coordinates of $\overline{\mathbf{y}}, \overline{\mathbf{c}}$, and $\overline{\mathbf{e}}$ which are denoted by $\mathbf{y}, \mathbf{c}$ and $\mathbf{e}$ respectively. By definition of $\mathcal{G}^{+}_{\mathbf{g}}(n, n', k, m)$, there exists a q-polynomial P of q-degree at most $k - 1$ such that for $1 \leq i \leq n'$:

$$y_i = P(g_i) + e_i. \tag{2}$$

Let also V and V_2 be the unique monic q-polynomials of q-degree δ and ε which vanish on the vector spaces E and E_2 respectively. The ring of q-polynomials being left Euclidean, there exists a unique monic q-polynomial W of degree $\delta - \varepsilon$ such that $V = W \circ V_2$. As E_2 is known, one can easily build the q-polynomial V_2, for instance using the iterative process described in [31,28]. Evaluating V at

both sides of Equation (2), one gets $V(y_i) = (V \circ P)(g_i) + V(e_i) = V \circ P(g_i)$. This secret polynomial can be written symbolically using $\delta - \varepsilon$ unknowns in $\mathbb{F}_{q^m}$, and similarly we view $R := V \circ P$ as a q-polynomial of q-degree $k - 1 + \delta$ with unknown coefficients. Thus, we can derive a linear equation containing $k + 2\delta - \varepsilon$ unknowns in $\mathbb{F}_{q^m}$ from

$$V(y_i) = R(g_i), \tag{3}$$

and the same goes for any $i \in \{1, 2, \dots n'\}$. Overall, this gives a linear system with n' equations in $k + 2\delta - \varepsilon$ variables. This linear system has more equations than unknowns as long as $\delta \leq \left\lfloor \frac{n'-k+\varepsilon}{2} \right\rfloor$, which is the case by assumption. Moreover, this system has a unique solution by Proposition 1. This means that exactly $k + 2\delta - \varepsilon$ equations are linearly independent, thus one can solve the system to recover V and R, so one finally gets P. □

4 New Rank-based Encryption Schemes

4.1 Multi-RQC-AG scheme

Our new encryption scheme denoted Multi-RQC-AG stands for RQC with multiple syndromes. Indeed, it uses several syndromes $\mathbf{U}$ and $\mathbf{V}$ which differ from the original RQC proposal that relies on unique syndromes $\mathbf{u}$ and $\mathbf{v}$. As a consequence, our new scheme is based on the IRSL problem which can be seen as a generalization of the IRSD problem used by RQC.

Notations. We start by introducing several sets and operators required to define the Multi-RQC-AG scheme. Let $\mathcal{S}^{2n}_{w,1}(\mathbb{F}_{q^m})$ and $\mathcal{S}^{n_2 \times 3n_1}_{(w_1,w_2)}(\mathbb{F}_{q^m})$ be defined as:

$$\begin{aligned}
\mathcal{S}^{2n}_{w,1}(\mathbb{F}_{q^m}) &= \{\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{F}^{2n}_{q^m} \mid \|\mathbf{x}\| = w,\ 1 \in \mathrm{Supp}(\mathbf{x})\},\\
\mathcal{S}^{n_2 \times 3n_1}_{(w_1,w_2)}(\mathbb{F}_{q^m}) &= \{\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) \in \mathbb{F}^{n_2 \times 3n_1}_{q^m} \mid \|(\mathbf{X}_1, \mathbf{X}_3)\| = w_1,\\
&\qquad \|\mathbf{X}_2\| = w_1 + w_2,\ \ \mathrm{Supp}(\mathbf{X}_1, \mathbf{X}_3) \subset \mathrm{Supp}(\mathbf{X}_2)\}.
\end{aligned}$$

Let n_1, n_2 be positive integers such that $n = n_1 \times n_2$, for a vector $\mathbf{v} \in \mathbb{F}^{n_2}_{q^m}$ and a matrix $\mathbf{M} \in \mathbb{F}^{n_2 \times n_1}_{q^m}$ whose columns are labelled $\mathbf{M}_1, \dots, \mathbf{M}_{n_1}$, we extend the aforementioned dot product such that:

$$\mathbf{v} \cdot \mathbf{M} \quad = \quad \left((\mathbf{v} \cdot \mathbf{M}^{\intercal}_1 \mod P)^{\intercal},\ \dots,\ (\mathbf{v} \cdot \mathbf{M}^{\intercal}_{n_1} \mod P)^{\intercal}\right) \quad \in \quad \mathbb{F}^{n_2 \times n_1}_{q^m},$$

Let $\mathbf{v} = (\mathbf{v}_1, \dots, \mathbf{v}_{n_1}) \in \mathbb{F}^{n}_{q^m}$ with $\mathbf{v}_i \in \mathbb{F}^{n_2}_{q^m}\ \forall i \in \{1, \dots, n_1\}$, the $\mathsf{Fold}()$ procedure turns the vector $\mathbf{v}$ into a $n_2 \times n_1$ matrix $\mathsf{Fold}(\mathbf{v}) = (\mathbf{v}^{\intercal}_1, \dots, \mathbf{v}^{\intercal}_{n_1}) \in \mathbb{F}^{n_2 \times n_1}_{q^m}$. The procedure $\mathsf{Unfold}()$ is naturally defined as the converse of $\mathsf{Fold}()$.

Protocol. The Multi-RQC-AG is described on Figure 2. It relies on two codes namely an augmented Gabidulin code $\mathcal{G}^{+}_{\mathbf{g}}(n, n', k, m)$ that can correct up to $\delta := \left\lfloor \frac{n'-k+\varepsilon}{2} \right\rfloor$ errors using the efficient decoding algorithm $\mathcal{G}^{+}_{\mathbf{g}}.\mathsf{Decode}(.)$ as well

as a random ideal $[2n_2, n_2]_{\mathbb{F}_{q^m}}$-code with parity check matrix $(\mathbf{I} \;\; \mathcal{IM}(\mathbf{h}))$. The correctness of the protocol follows from:

$$\begin{aligned}\mathbf{V} - \mathbf{y}\cdot\mathbf{U} &= \mathsf{Fold}(\mathbf{mG}) + (\mathbf{x} + \mathbf{h}\cdot\mathbf{y})\cdot\mathbf{R}_2 + \mathbf{E} - \mathbf{y}\cdot(\mathbf{R}_1 + \mathbf{h}\cdot\mathbf{R}_2)\\ &= \mathsf{Fold}(\mathbf{mG}) + \mathbf{x}\cdot\mathbf{R}_2 - \mathbf{y}\cdot\mathbf{R}_1 + \mathbf{E}.\end{aligned}$$

As a consequence, $\mathsf{Unfold}\,(\mathbf{V} - \mathbf{y}\cdot\mathbf{U}) = \mathbf{mG} + \mathsf{Unfold}\,(\mathbf{x}\cdot\mathbf{R}_2 - \mathbf{y}\cdot\mathbf{R}_1 + \mathbf{E}) \in \mathbb{F}_{q^m}^n$ which means that $\mathcal{G}_{\mathbf{g}}.\mathsf{Decode}\,(\mathsf{Unfold}\,(\mathbf{V} - \mathbf{y}\cdot\mathbf{U})) = \mathbf{m}$ as long as:

$$\|\,\mathsf{Unfold}\,(\mathbf{x}\cdot\mathbf{R}_2 - \mathbf{y}\cdot\mathbf{R}_1 + \mathbf{E})\,\| \leq \delta.$$

$\underline{\mathsf{Setup}(1^\lambda)}$

Generate and output the parameters $\mathsf{param} = (n', n_1, n_2, k, \epsilon, \delta, w, w_1, w_2, P)$ where $P \in \mathbb{F}_q[X]$ is an irreducible polynomial of degree n_2.

$\underline{\mathsf{KeyGen}(\mathsf{param})}$:

Sample $\mathbf{g} \xleftarrow{\$} \mathcal{S}_{n'}^{n'}(\mathbb{F}_{q^m})$, $\mathbf{h} \xleftarrow{\$} \mathbb{F}_{q^m}^{n_2}$ and $(\mathbf{x}, \mathbf{y}) \xleftarrow{\$} \mathcal{S}_{w,1}^{2n_2}(\mathbb{F}_{q^m})$

Compute $\mathbf{s} = \mathbf{x} + \mathbf{h}\cdot\mathbf{y} \mod P$

Output $\mathsf{pk} = (\mathbf{g}, \mathbf{h}, \mathbf{s})$ and $\mathsf{sk} = (\mathbf{x}, \mathbf{y})$

$\underline{\mathsf{Encrypt}(\mathsf{pk}, \mathbf{m}, \theta)}$:

Compute $\overline{\mathbf{g}} = (\mathbf{g}\,|\,0\ldots0) \in \mathbb{F}_{q^m}^{n_1 n_2}$

Compute the generator matrix $\mathbf{G} \in \mathbb{F}_{q^m}^{k\times(n_1 n_2)}$ of $\mathcal{G}_{\overline{\mathbf{g}}}^{+}(n_1 n_2, n', k, m)$

Sample $(\mathbf{R}_1, \mathbf{E}, \mathbf{R}_2) \xleftarrow{\$} \mathcal{S}_{w_1,w_2}^{n_2\times 3n_1}(\mathbb{F}_{q^m})$ using randomness θ

Compute $\mathbf{U} = \mathbf{R}_1 + \mathbf{h}\cdot\mathbf{R}_2$ and $\mathbf{V} = \mathsf{Fold}(\mathbf{mG}) + \mathbf{s}\cdot\mathbf{R}_2 + \mathbf{E}$

Output $\mathbf{C} = (\mathbf{U}, \mathbf{V})$

$\underline{\mathsf{Decrypt}(\mathsf{pk}, \mathsf{sk}, \mathbf{C})}$:

Output $\mathbf{m} = \mathcal{G}_{\overline{\mathbf{g}}}^{+}.\mathsf{Decode}\,(\mathsf{Unfold}\,(\mathbf{V} - \mathbf{y}\cdot\mathbf{U}))$

Fig. 2. Multi-RQC-AG encryption scheme

Theorem 2. *The Multi-RQC-AG scheme depicted in Figure 2 is* IND-CPA *under the DIRSD and the DNHIRSL assumptions.*

Proof. The proof of the Multi-RQC-AG scheme is similar to the proof from [3] with an $\mathsf{IRSD}(m, 2n_2, n_2, \omega)$ instance defined from a $[2n_2, n_2]$ code and an $\mathsf{NHIRSL}(m, n_2, n_2, \omega_1, \omega_2, n_1)$ instance defined from a $[3n_2, n_2]$ code. These instances are defined by the following products:

$$\begin{pmatrix}\mathbf{I}_{n_2} & \mathcal{IM}(\mathbf{h})\end{pmatrix} \times (\mathbf{x}, \mathbf{y})^{\mathsf{T}} = \mathbf{s}^{\mathsf{T}},$$

$$\begin{pmatrix}\mathbf{I}_{n_2} & \mathbf{0} & \mathcal{IM}(\mathbf{h})\\ \mathbf{0} & \mathbf{I}_{n_2} & \mathcal{IM}(\mathbf{s})\end{pmatrix} \times (\mathbf{R}_1, \mathbf{E}, \mathbf{R}_2)^{\mathsf{T}} = (\mathbf{U}, \mathbf{V} - \mathrm{Fold}(\mathbf{mG}))\,.$$

4.2 Multi-UR-AG scheme

Our new encryption scheme denoted Multi-UR-AG stands for Multiple syndromes Unstructured Rank with Augmented Gabidulin codes encryption scheme. It is particularly interesting security wise as it does not use structured codes contrarily to existing constructions such as ROLLO, RQC or our new proposal Multi-RQC-AG . Indeed, it only relies on the security of the RSL problem. Multi-UR-AG leverages multiple syndromes and augmented Gabidulin codes. In addition, it features two variants as it can be instantiated with either homogeneous or non-homogeneous errors.

Notations. Hereafter, Fold and Unfold refer to the procedure introduced in Section 4.1. Let $\mathcal{S}_{w,1}^{n\times 2n_1}(\mathbb{F}_{q^m})$ and $\mathcal{S}_{(w_1,w_2)}^{n_2\times(n+n_1+n)}(\mathbb{F}_{q^m})$ be defined as:

$$\begin{aligned}\mathcal{S}_{w,1}^{n\times 2n_1}(\mathbb{F}_{q^m}) &= \{\mathbf{X}=(\mathbf{X}_1,\mathbf{X}_2)\in\mathbb{F}_{q^m}^{n\times 2n_1} \,\big|\, \|\mathbf{X}\|=w,\ 1\in\mathrm{Supp}(\mathbf{X})\},\\ \mathcal{S}_{(w_1,w_2)}^{n_2\times(n+n_1+n)}(\mathbb{F}_{q^m}) &= \{\mathbf{X}=(\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3)\in\mathbb{F}_{q^m}^{n_2\times(n+n_1+n)} \,\big|\, \|(\mathbf{X}_1,\mathbf{X}_3)\|=w_1,\\ &\qquad \|\mathbf{X}_2\|=w_1+w_2,\quad \mathrm{Supp}(\mathbf{X}_1,\mathbf{X}_3)\subset\mathrm{Supp}(\mathbf{X}_2)\}.\end{aligned}$$

Protocol. The Multi-UR-AG is described on Figure 3. It relies on two codes namely an augmented Gabidulin code $\mathcal{G}_{\mathbf{g}}^{+}(n,n',k,m)$ that can can correct up to $\delta := \left\lfloor \frac{n'-k+\varepsilon}{2} \right\rfloor$ errors using the efficient decoding algorithm $\mathcal{G}_{\mathbf{g}}^{+}.\mathsf{Decode}(.)$ as well as a random $[2n,n]_{\mathbb{F}_{q^m}}$-code with parity check matrix $(\mathbf{I}\ \mathbf{H})$. The correctness of the protocol follows from:

$$\begin{aligned}\mathbf{V}-\mathbf{UY} &= \mathsf{Fold}(\mathbf{mG})+\mathbf{R}_2(\mathbf{X}+\mathbf{HY})+\mathbf{E}-(\mathbf{R}_1+\mathbf{R}_2\mathbf{H})\mathbf{Y}\\ &= \mathsf{Fold}(\mathbf{mG})+\mathbf{R}_2\mathbf{X}-\mathbf{R}_1\mathbf{Y}+\mathbf{E}.\end{aligned}$$

As a consequence, $\mathsf{Unfold}\,(\mathbf{V}-\mathbf{YU}) = \mathbf{mG}+\mathsf{Unfold}\,(\mathbf{XR}_2-\mathbf{YR}_1+\mathbf{E}) \in \mathbb{F}_{q^m}^n$ which means that $\mathcal{G}_{\mathbf{g}}.\mathsf{Decode}\,(\mathsf{Unfold}\,(\mathbf{V}-\mathbf{YU})) = \mathbf{m}$ as long as:

$$\|\,\mathsf{Unfold}\,(\mathbf{XR}_2-\mathbf{YR}_1+\mathbf{E})\,\| \le \delta.$$

Theorem 3. *The Multi-UR-AG scheme is* IND-CPA *under the DRSL and the DNHRSL assumptions.*

Proof. The proof of the Multi-UR-AG scheme is similar to the proof from [3] with an $\mathsf{RSL}(m,2n,n,\omega,n_1)$ instance defined from a $[2n,n]$ code and an $\mathsf{NHRSL}(m,n,n_1,\omega_1,\omega_2,n_2)$ instance defined from a $[2n+n_1,n]$ code. These instances are defined by the following products:

$$\begin{pmatrix}\mathbf{I}_n & \mathbf{H}\end{pmatrix}\times(\mathbf{X},\mathbf{Y})^{\intercal}=\mathbf{S},$$

$$(\mathbf{R}_1,\mathbf{E},\mathbf{R}_2)\times\begin{pmatrix}\mathbf{I}_n & \mathbf{0} & \mathbf{H}\\ \mathbf{0} & \mathbf{I}_{n_1} & \mathbf{S}\end{pmatrix}^{\intercal}=(\mathbf{U},\mathbf{V}-\mathrm{Fold}(\mathbf{mG}))\,.$$

$\underline{\mathsf{Setup}(1^\lambda)}$

Generate and output $\mathsf{param} = (n, n', n_1, n_2, k, \epsilon, \delta, w, w_1, w_2)$ where $n = n_1 n_2$.

$\underline{\mathsf{KeyGen}(\mathsf{param})}$:

Sample $\mathbf{g} \xleftarrow{\$} \mathcal{S}_{n'}^{n'}(\mathbb{F}_{q^m})$, $\mathbf{H} \xleftarrow{\$} \mathbb{F}_{q^m}^{n \times n}$ and $(\mathbf{X}, \mathbf{Y}) \xleftarrow{\$} \mathcal{S}_{w,1}^{n \times 2n_1}(\mathbb{F}_{q^m})$

Compute $\mathbf{S} = \mathbf{X} + \mathbf{HY}$

Output $\mathsf{pk} = (\mathbf{g}, \mathbf{H}, \mathbf{S})$ and $\mathsf{sk} = (\mathbf{X}, \mathbf{Y})$

$\underline{\mathsf{Encrypt}(\mathsf{pk}, \mathbf{m}, \theta)}$:

Compute $\overline{\mathbf{g}} = (\mathbf{g} \,|\, 0 \dots 0) \in \mathbb{F}_{q^m}^n$

Compute the generator matrix $\mathbf{G} \in \mathbb{F}_{q^m}^{k \times n}$ of $\mathcal{G}_{\overline{\mathbf{g}}}^{+}(n, n', k, m)$

Sample $(\mathbf{R}_1, \mathbf{E}, \mathbf{R}_2) \xleftarrow{\$} \mathcal{S}_{w_1, w_2}^{n_2 \times (n + n_1 + n)}(\mathbb{F}_{q^m})$ using randomness θ

Compute $\mathbf{U} = \mathbf{R}_1 + \mathbf{R}_2\mathbf{H}$ and $\mathbf{V} = \mathsf{Fold}(\mathbf{mG}) + \mathbf{R}_2\mathbf{S} + \mathbf{E}$

Output $\mathbf{C} = (\mathbf{U}, \mathbf{V})$

$\underline{\mathsf{Decrypt}(\mathsf{pk}, \mathsf{sk}, \mathbf{C})}$:

Output $\mathbf{m} = \mathcal{G}_{\overline{\mathbf{g}}}^{+}.\mathsf{Decode}\left(\mathsf{Unfold}\left(\mathbf{V} - \mathbf{UY}\right)\right)$

Fig. 3. Multi-UR-AG (with non-homogeneous errors) encryption scheme

5 Security analysis

In this section, we provide the complexity to solve some hard problems in rank-based cryptography.

5.1 Attacks on the RSD problem [23,8,12]

There are two general classes of attacks to RSD, based on combinatorial or algebraic techniques. On the one hand, combinatorial attacks can be seen as the equivalent of ISD-type attacks in the rank metric setting. Relying on [23,8], we estimate that the complexity of the best combinatorial attack is in

$$\min\left(2^{(w-1)\left\lfloor\frac{(k+1)m}{n}\right\rfloor}, 2^{w\left\lceil\frac{(k+1)m}{n}\right\rceil - m}\right) \tag{4}$$

$\mathbb{F}_q$-operations. On the other hand, algebraic attacks on the RSD problem are by modeling the decoding instance into a system of polynomial equations, and the overall cost is reduced to the one of solving this system. To design our parameters, we take into account the most recent algebraic attack, namely the MaxMinors attack [12]. Its complexity in $\mathbb{F}_q$ operations is estimated to be

$$\mathcal{O}\left(q^{aw} m \binom{n-k-1}{w} \binom{n-a}{w}^{\omega-1}\right), \tag{5}$$

where $a \geq 0$ the smallest integer such that $m\binom{n-k-1}{w} \geq \binom{n-a}{w} - 1$ and where ω is a linear algebra constant.

5.2 Attacks on the NHRSD problem

This section is dedicated to the first cryptanalysis of the NHRSD problem by proposing two attacks which exploit the inhomogeneous structure of the error.

A new combinatorial attack. In this section, we may assume for clarity that the n_1 leftmost coordinates of $\mathbf{e}$ correspond to the part of weight w_1+w_2, namely $\mathbf{e} = (\mathbf{e}_2, \mathbf{e}_1, \mathbf{e}_3)$, and we also adopt a systematic form for the parity-check matrix $\boldsymbol{H}_{\mathbf{e}} = \left(\boldsymbol{I}_{n+n_1-1} \ *\right)$ of the public code $\mathcal{C}_{\mathbf{e}} := \mathcal{C} \oplus \langle \mathbf{e} \rangle$. The parity-check equations for this code which are traditionally used in this type of attack are as follows:

1. those associated to the n first rows of $\boldsymbol{H}_{\mathbf{e}}$ provide n linear relations over $\mathbb{F}_{q^m}$ which can be mapped into nm relations over $\mathbb{F}_q$ between unknowns coming from $\mathbf{e}_2$, $\mathbf{e}_1$ and $\mathbf{e}_3$.
2. those associated to the $n_1 - 1$ last rows of $\boldsymbol{H}_{\mathbf{e}}$ give $(n_1 - 1)m$ equations over $\mathbb{F}_q$ in unknowns coming from the components of $\mathbf{e}_1$ and $\mathbf{e}_3$ only.

Before describing our attack, let us recall how [23,8] would solve a non-structured $\mathsf{RSD}(m, 2n, n, w_1)$ instance to recover $(\mathbf{e}_1, \mathbf{e}_3)$. The most enhanced version of [8] consists in guessing a subspace V of dimension $r_1 \geq w_1$ such that $\alpha S_1 \subset V$ for some element $\alpha \in \mathbb{F}_{q^m}^*$ instead of simply $S_1 \subset V$ as it provides a better success probability. Then, it aims at solving the linear system given by the parity-check equations from 2. The largest value of r_1 for which one may expect a unique solution is given by

$$r_1 := \left\lfloor \frac{m(n-1)}{2n} \right\rfloor = m - \left\lceil \frac{m(n+1)}{2n} \right\rceil . \tag{6}$$

The classical cost given in [8] is then roughly

$$\widetilde{\mathcal{O}}\left(q^{w_1(m-r_1)-m}\right) = \widetilde{\mathcal{O}}\left(q^{w_1\left\lceil \frac{m(n+1)}{2n} \right\rceil - m}\right), \tag{7}$$

where $\widetilde{\mathcal{O}}$ hides a polynomial factor corresponding to solving this linear system. To benefit from the inhomogeneous structure of $\mathbf{e}$ from NHRSD, our approach follows the natural path of making a guess on a random subspace V of dimension $r \geq w_1$ such that $S_1 \subset V$ *and* a random subspace $Z \subset \mathbb{F}_{q^m}/V$ of dimension $\rho \in \{w_2..m - r\}$ such that $S_2 \subset V \oplus Z$.

Theorem 4. *Our proposed combinatorial algorithm runs in time*

$$\widetilde{\mathcal{O}}\left(q^{(w_1+w_2)(m-r)-w_2\rho-m}\right). \tag{8}$$

The complexity given by Equation (8) is of the same shape as Equation (7) since the rest of our attack is totally similar to [23,8]: expressing the coordinates of $(\mathbf{e}_1, \mathbf{e}_3)$ in a fixed basis of V yields $2nr$ variables over $\mathbb{F}_q$, while we get $n_1(r+\rho)$ variables over $\mathbb{F}_q$ by writing the coordinates of $\mathbf{e}_2$ in a fixed basis of $V \oplus Z$. For the linear algebra step, $n_1(r+\rho)$ random equations from 1. are used in order to

express all the variables from $\mathbf{e}_2$ in terms of the other variables, and we are left with a linear system of $(n+n_1-1)m-n_1(r+\rho)$ equations over $\mathbb{F}_q$ in only $2nr$ variables. This leads to the condition

$$2nr \leq m(n+n_1-1)-n_1(r+\rho)$$

in order to expect at most one solution. Overall, the main task to prove Theorem 4 is to compute the success probability $\Pi := \Pr_{V,Z}[S_1 \subset V,\ S_2 \subset V \oplus Z]$, see Appendix A.1. Using [8], recall also that one may take advantage of $\mathbb{F}_{q^m}$-linearity by considering a greater probability of the form

$$\Pr_{V,Z}\left[\exists \alpha \in \mathbb{F}_{q^m}^*,\ \alpha S_1 \subset V,\ \alpha S_2 \subset V \oplus Z\right] \approx \frac{q^m-1}{q-1}\Pi, \tag{9}$$

and in case of success decoding the word $\alpha\mathbf{e}$ instead of $\mathbf{e}$. For clarity Appendix A.1 presents the plain version of the attack, but as this trick is compatible with our analysis the corresponding q^{-m} factor appears in Equation (8). Finally, one has to consider the couple (r,ρ) which leads to the best exponent in Equation (8). In other words, the goal will be to maximize the quantity $(w_1+w_2)r+w_2\rho$ under the constraints $(2n+n_1)r+n_1\rho \leq m(n+n_1-1)$, $w_1 \leq r$, $w_2 \leq \rho$, $r+\rho \leq m-1$, where $r,\rho \in \mathbb{N}$.

This is an example of integer linear program (ILP), and to solve this instance we have used dedicated tools.

Adaptation of the algebraic attack of [12] against NHRSD. A first approach of this attack was proposed in [3], we build upon this work and give a thorough analysis of the complexity of this attack.

Theorem 5. *Let $a \geq 0$ the smallest integer such that*

$$\mathcal{N}_{\mathbb{F}_q} \geq \tbinom{2n+n_1-a}{w_1+w_2} - M_a - \nu_{\mathbb{F}_q} - 1,$$

where $\mathcal{N}_{\mathbb{F}_q} = m\sum_{i=w_2}^{w_1+w_2}\binom{n_1-1}{i}\binom{n}{w_1+w_2-i}$, $\nu_{\mathbb{F}_q} = m\binom{n_1-1}{w_2-1}\binom{n-1}{w_1}$ and $M_a := \sum_{i=0}^{\omega_2-1}\binom{n_1}{i}\binom{2n-a}{\omega_1+\omega_2-i}$. The hybrid MaxMinors attack adapted to NHRSD costs

$$\mathcal{O}\left(q^{aw_1}\mathcal{N}_{\mathbb{F}_q}\left(\tbinom{2n+n_1-a}{w_1+w_2} - M_a - \nu_{\mathbb{F}_q}\right)^{\omega-1}\right)$$

operations in $\mathbb{F}_q$, where ω is a linear algebra constant.

MaxMinors linear system [12]. The MaxMinors system is a system of equations over $\mathbb{F}_{q^m}$ which vanish on the solutions to the RSD instance. Let $\mathbf{y} = \mathbf{c}+\mathbf{e} \in \mathbb{F}_{q^m}^{(2n+n_1)}$ be the noisy codeword to be decoded in a random $\mathbb{F}_{q^m}$-linear code $\mathcal{C}$ of length $2n+n_1$ and dimension n with generator matrix $\boldsymbol{G} \in \mathbb{F}_{q^m}^{n\times(2n+n_1)}$. The extended code $\mathcal{C}_{\mathbf{e}} = \mathcal{C}_{\mathbf{y}}$ is generated by the matrix $\boldsymbol{G}_{\mathbf{y}} := \begin{pmatrix}\boldsymbol{G}\\ \mathbf{y}\end{pmatrix}$, and we also

consider $\boldsymbol{H}_{\mathbf{y}} \in \mathbb{F}_{q^m}^{(n+n_1-1)\times(2n+n_1)}$ a full-rank parity-check matrix for this code. We clearly have

$$0 = \mathbf{e}\boldsymbol{H}_{\mathbf{y}}^{\mathsf{T}} = \boldsymbol{\beta}\,\mathrm{Mat}(\mathbf{e})\boldsymbol{H}_{\mathbf{y}}^{\mathsf{T}} = \boldsymbol{\beta}\boldsymbol{S}\boldsymbol{C}\boldsymbol{H}_{\mathbf{y}}^{\mathsf{T}},$$

so that the matrix $\boldsymbol{C}\boldsymbol{H}_{\mathbf{y}}^{\mathsf{T}}$ contains a non-zero vector $\boldsymbol{\beta}\boldsymbol{S}$ in its left kernel and cannot be full-rank. In particular, the MaxMinors system is the system of maximal minors $\mathcal{P} := \{P_J\}_J$ such that $P_J := \left|\boldsymbol{C}\boldsymbol{H}_{\mathbf{y}}^{\mathsf{T}}\right|_{*,J}$ for each subset $J \subset \{1..n+n_1-1\}$, $\#J = w_1 + w_2$. The crux is that these equations are actually *linear* in the minor variables $c_T := |\boldsymbol{C}|_{*,T} \in \mathbb{F}_q$ by using the Cauchy-Binet formula for the determinant of a product of rectangular matrices, see [11,12]. In this section, the c_T's will be sorted with respect to the following ordering on the T's: we consider that $T = \{t_1 < \cdots < t_r\} < T' = \{t'_1 < \cdots < t'_r\}$ if $t_j = t'_j$ for $j < j_0$ and $t_{j_0} < t'_{j_0}$ assuming that $1 < 2 < \cdots < n$. We will further assume that $\boldsymbol{H}_{\mathbf{y}} := \left(* \; \boldsymbol{I}_{n+n_1-1}\right)$ and from that assumption [12] derive the fundamental Lemma 1 on the shape of the MaxMinors equations:

Lemma 1 (Prop. 2, [12]).

$$P_J = c_{J+n+1} + \sum_{\substack{T^- \subset \{1..n+1\}, T^+ \subset (J+n+1) \\ T = T^- \cup T^+,\ \#T = w_1+w_2,\ T^- \neq \emptyset}} c_T |\boldsymbol{H}_{\mathbf{y}}|_{J,T}. \tag{10}$$

A direct consequence of Lemma 1 is that the equations of $\mathcal{P}$ are linearly independent over $\mathbb{F}_{q^m}$ as their leading terms are distinct.

Removing variables corresponding to zero minors. The very same MaxMinors system can be employed to attack NHRSD. A main difference in this case is that if one wants to decrease the number of minor variables by relying on the special structure of $\mathbf{e}$ as shown in [3,12], then linear relations between the equations after removing these variables also occur and must be taken into account in the analysis. Recall from [3, 6.2.2] that the row support of $\mathrm{Mat}(\mathbf{e}) \in \mathbb{F}_q^{m\times(n+n_1+n)}$ can be written as

$$\boldsymbol{C} = \begin{pmatrix} \boldsymbol{C}_1 & \boldsymbol{C}_2 & \boldsymbol{C}_3 \\ 0 & \boldsymbol{C}'_2 & 0 \end{pmatrix} \in \mathbb{F}_q^{(w_1+w_2)\times(n+n_1+n)}, \tag{11}$$

where $\boldsymbol{C}_1,\ \boldsymbol{C}_3 \in \mathbb{F}_q^{w_1\times n}$, $\boldsymbol{C}_2 \in \mathbb{F}_q^{w_1\times n_1}$ and $\boldsymbol{C}'_2 \in \mathbb{F}_q^{w_2\times n_1}$. From Equation (11), it has been noted that the minors $|\boldsymbol{C}|_{*,T}$ such that $T \cap \{n+1..n+n_1\} \leq w_2 - 1$ are always zero. This means that the

$$M := \sum_{i=0}^{w_2-1} \binom{n_1}{i} \binom{2n}{w_1+w_2-i} \tag{12}$$

variables from the set

$$\zeta := \{c_T,\ T \subset \{1..(2n+n_1)\},\ \#T = w_1+w_2,\ T \cap \{n+1..n+n_1\} \leq w_2 - 1\}$$

can be set to zero in the MaxMinors system. It is then relevant to separate the initial P_J equations into several subsets in function of the presence or the absence of these c_T variables. We consider the partition $\mathcal{P} := \mathcal{P}_{\text{lost}} \sqcup \mathcal{P}_{\text{rest}} \sqcup \mathcal{P}_{\text{indep}}$, where

$$\begin{aligned}
\mathcal{P}_{\text{lost}} &:= \{P_J : \ \#J = w_1 + w_2, \ \#(J \cap \{1..(n_1-1)\}) \leq w_2 - 2\} \\
\mathcal{P}_{\text{rest}} &:= \{P_J : \ \#J = w_1 + w_2, \ \#(J \cap \{1..(n_1-1)\}) = w_2 - 1\} \\
\mathcal{P}_{\text{indep}} &:= \{P_J : \ \#J = w_1 + w_2, \ \#(J \cap \{1..(n_1-1)\}) \geq w_2\}\,.
\end{aligned}$$

Using Lemma 1, it is easy to grasp the shape of the equations from $\mathcal{P}_{\text{lost}}$ and $\mathcal{P}_{\text{indep}}$ after removing the minor variables belonging to ζ:

Proposition 3. *After setting the minor variables from ζ to zero in the MaxMinors system $\mathcal{P}$, we have the following properties:*

1. *The equations in $\mathcal{P}_{lost}$ all become zero.*
2. *The equations in $\mathcal{P}_{indep}$ keep the same leading terms and therefore they are still linearly independent. We have*

$$\dim_{\mathbb{F}_{q^m}} \langle \mathcal{P}_{indep} \rangle = \#\mathcal{P}_{indep} = \sum_{i=w_2}^{w_1+w_2} \binom{n_1-1}{j}\binom{n}{w_1+w_2-j}.$$

Finally, the system $\mathcal{P}_{indep}$ contains at most $\binom{2n+n_1}{w_1+w_2} - M$ variables.

Proof. See Appendix A.3. □

Contrary to $\mathcal{P}_{\text{indep}}$, the equations in $\mathcal{P}_{\text{rest}}$ have their leading terms in ζ so that these monomials are destroyed after setting the M minor variables to zero. More precisely, by Lemma 1, an equation $P_J \in \mathcal{P}_{\text{rest}}$ becomes

$$\begin{aligned}
\widetilde{P_J} &= \sum_{\substack{T^- \subset \{1..n+1\},\ T^+ \subset (J+n+1) \\ T = T^- \cup T^+,\ n+1 \in T^-,\ \#(T^+ \cap \{n+2..n+n_1\}) = w_2 - 1}} c_T |\boldsymbol{H}_{\mathbf{y}}|_{J,T} \\
&= \sum_{\substack{T^- \subset \{1..n+1\},\ T^+ \subset (J+n+1) \\ T = T^- \cup T^+,\ n+1 \in T^-,\ T^+ \cap \{n+2..n+n_1\} = (J \cap \{1..(n_1-1)\}) + n + 1}} c_T |\boldsymbol{H}_{\mathbf{y}}|_{J,T} \qquad (13)
\end{aligned}$$

For clarity, we still denote the resulting system by $\mathcal{P}_{\text{rest}}$. We analyze it in the following Proposition 4:

Proposition 4. *After setting the minor variables from ζ to zero in $\mathcal{P}_{rest}$, one obtains a system of rank $\binom{n_1-1}{w_2-1}\binom{n-1}{w_1}$ and whose equations are also independent from $\mathcal{P}_{indep}$. Finally, these equations contain at most $\binom{n_1-1}{w_2-1}\binom{2n}{w_1}$ variables.*

The first part of Proposition 4 is obvious. Using Equation (13), the leading term of $\widetilde{P_J} \in \mathcal{P}_{\text{rest}}$ is a c_T variable such that $n+1 \in T$, whereas the leading term of any $P_{J'} \in \mathcal{P}_{\text{indep}}$ is $c_{J'+n+1}$ and $c_{J'+n+1} > c_T$ for any such T. Thus, what is left to prove in Proposition 4 is that $\dim_{\mathbb{F}_{q^m}} \langle \mathcal{P}_{\text{rest}} \rangle = \binom{n_1-1}{w_2-1}\binom{n-1}{w_1}$ and that the number of variables is $\binom{n_1-1}{w_2-1}\binom{2n}{w_1}$. For this we rely on the following lemma, whose proofs can be found in Appendix A.3:

Lemma 2. *For $A \subset \{n+2..n+n_1\}$, $\#A = w_2 - 1$, let*

$$\mathcal{P}_{rest,A} := \{P_J \in \mathcal{P}_{rest} : J \cap \{1..n_1 - 1\} = A - (n+1)\},$$

so that $\{\mathcal{P}_{rest,A}\}_A$ is a partition of $\mathcal{P}_{rest}$. We have $\langle \mathcal{P}_{rest} \rangle = \oplus_A \langle \mathcal{P}_{rest,A} \rangle$.

Lemma 3. *For $A \subset \{n+2..n+n_1\}$, $\#A = w_2 - 1$, let $\mathcal{P}_{rest,A}$ as defined in Lemma 2. With very high probability, we have $\dim_{\mathbb{F}_{q^m}} \langle \mathcal{P}_{rest,A} \rangle = \binom{n-1}{w_1}$.*

Finishing the attack by projecting over $\mathbb{F}_q$. The last step of the initial MaxMinors attack on RSD is by solving the "projected" linear system $\mathcal{P}_{\mathbb{F}_q} := \{P_{j,J}\}_{j,J}$ obtained by expressing the coefficients of the P_J's in a fixed basis of $\mathbb{F}_{q^m}$ over $\mathbb{F}_q$ and taking each component, yielding m times more equations. We proceed in a very similar way as in [12] and due to space constraints we do not recall all the details of this step. Our final complexity estimate relies on

Assumption 1 *Let $\mathcal{P}_{indep,\mathbb{F}_q}$ (resp. $\mathcal{P}_{rest,\mathbb{F}_q}$) be the system over $\mathbb{F}_q$ obtained by projecting $\mathcal{P}_{indep}$ (resp. $\mathcal{P}_{rest}$) where the variables in ζ had already been removed, let $\mathcal{N}_{\mathbb{F}_q} := \dim_{\mathbb{F}_q} \langle \mathcal{P}_{indep,\mathbb{F}_q} \rangle$, let $\nu_{\mathbb{F}_q} := \dim_{\mathbb{F}_q} \langle \mathcal{P}_{rest,\mathbb{F}_q} \rangle$ and let M as defined in Equation (12). We assume that*

$$\mathcal{N}_{\mathbb{F}_q} = m \dim_{\mathbb{F}_{q^m}} \langle \mathcal{P}_{indep} \rangle = m \sum_{i=w_2}^{w_1+w_2} \binom{n_1-1}{i} \binom{n}{w_1+w_2-i} \tag{14}$$

when this value is $\leq \binom{2n+n_1}{w_1+w_2} - M$ and $\mathcal{N}_{\mathbb{F}_q} = \binom{2n+n_1}{w_1+w_2} - M - 1$ otherwise, and

$$\nu_{\mathbb{F}_q} = m \dim_{\mathbb{F}_{q^m}} \langle \mathcal{P}_{rest} \rangle = m \binom{n_1-1}{w_2-1} \binom{n-1}{w_1}, \tag{15}$$

provided that this value is $\leq \binom{n_1-1}{w_2-1}\binom{2n}{w_1}$.

To solve the final system, one can start by performing linear algebra on $\mathcal{P}_{\mathrm{rest},\mathbb{F}_q}$ and then substitute $\nu_{\mathbb{F}_q}$ variables corresponding to an echelonized basis of $\langle \mathcal{P}_{\mathrm{rest},\mathbb{F}_q} \rangle$ in the system $\mathcal{P}_{\mathrm{indep},\mathbb{F}_q}$ to get a new system $\mathcal{P}'_{\mathrm{indep},\mathbb{F}_q}$. The final step is then to solve the linear system $\mathcal{P}'_{\mathrm{indep},\mathbb{F}_q}$ in $\binom{2n+n_1}{w_1+w_2} - M - \nu_{\mathbb{F}_q}$ variables.

Corollary 1 (Same notations as in Assumption 1) *Let $\mathcal{P}_{indep,\mathbb{F}_q}$ and let $\mathcal{P}_{rest,\mathbb{F}_q}$ denote the projected systems from Assumption 1. We consider $\mathcal{P}'_{indep,\mathbb{F}_q}$ the linear system obtained from $\mathcal{P}_{indep,\mathbb{F}_q}$ after plugging $\nu_{\mathbb{F}_q}$ equations from the echelon form of $\mathcal{P}_{rest,\mathbb{F}_q}$ to substitute variables. Assuming that the system $\mathcal{P}'_{indep,\mathbb{F}_q}$ can be solved, namely $\mathcal{N}_{\mathbb{F}_q} \geq \binom{2n+n_1}{w_1+w_2} - M - \nu_{\mathbb{F}_q} - 1$, the complexity of solving the system is*

$$\mathcal{O}\left(\mathcal{N}_{\mathbb{F}_q}\left(\binom{2n+n_1}{w_1+w_2} - M - \nu_{\mathbb{F}_q}\right)^{\omega-1}\right)$$

operations in $\mathbb{F}_q$, where ω is a linear algebra constant.

However, the projected linear system cannot be solved directly when there are not enough equations compared to the number of minor variables, *i.e.* $\mathcal{N}_{\mathbb{F}_q} <$

$\binom{2n+n_1}{w_1+w_2} - M - \nu_{\mathbb{F}_q} - 1$. In this case, a method suggested in [12] is an hybrid approach by adding linear constraints on these minor variables which are obtained by fixing the entries of $a \geq 0$ columns in the matrix $\boldsymbol{C}$. Here, like it was done in [3], it is possible to take advantage of the particular structure of $\boldsymbol{C}$ given in Equation (11) by fixing columns containing only w_1 non-zero coordinates, which leads to a smaller exponential factor of q^{aw_1} in the final cost instead of the naive $q^{a(w_1+w_2)}$. The cost claimed in Theorem 5 follows.

5.3 Attacks on the RSL problem

In this section, we consider an RSL (m, n, k, r, N)-instance, say N distinct RSD instances whose errors share the same support of dimension r. This number N is a crucial parameter to estimate the hardness of RSL and in particular to compare it to RSD. For instance, this problem can be solved in polynomial time when $N \geq nr$ due to [21]. A more powerful attack was later found in [16] and it suggests that secure RSL instances must satisfy a stronger condition: $N < kr$.

In what follows, we give a new combinatorial attack against RSL, it is more efficient than the previous combinatorial attacks, plus it enables us to decrease the threshold where the RSL problem starts to be solvable in polynomial time. In addition to this, we give more explicit formulas to clarify the recent algebraic attack of [10].

New combinatorial attack on RSL.

Theorem 6 (Combinatorial attack on RSL). *There exists a combinatorial attack on RSL (m, n, k, r, N) with complexity*

$$\widetilde{\mathcal{O}}\left(q^{r\left(m-\left\lfloor\frac{m(n-k)-N}{n-a}\right\rfloor\right)}\right)$$

operations in $\mathbb{F}_q$, where $a := \left\lfloor\frac{N}{r}\right\rfloor$.

Proof. Let $\mathbf{s}_i \in \mathbb{F}_{q^m}^{n-k}$, $1 \leq i \leq N$ denote the N syndromes from the RSL instance. By definition there exist $\mathbf{e}_i \in \mathbb{F}_{q^m}^n$, $\|\mathbf{e}_i\| = r$, $\boldsymbol{H}\mathbf{e}_i^\mathsf{T} = \mathbf{s}_i^\mathsf{T}$, where $\boldsymbol{H} \in \mathbb{F}_{q^m}^{(n-k)\times n}$ is a parity-check matrix and where $\mathrm{Supp}(\mathbf{e}_i)$ does not depend on i. Similarly to [24,10], this last property enables us to use the fact that there exists an $\mathbb{F}_q$-linear combination $(\ \mathbf{0}_a \mid \widetilde{\mathbf{e}}\ ) \in \mathbb{F}_{q^m}^n$ of the $\mathbf{e}_i$'s which is all-zero on its first $a := \left\lfloor\frac{N}{r}\right\rfloor$ coordinates. This error corresponds to a *secret* linear combination of the syndromes, more precisely

$$\exists \lambda_1, \lambda_2, \ldots, \lambda_N \in \mathbb{F}_q, \quad \boldsymbol{H}(\ \mathbf{0}_a \mid \widetilde{\mathbf{e}}\ )^\mathsf{T} = \textstyle\sum_{i=1}^N \lambda_i \mathbf{s}_i^\mathsf{T}.$$

By setting $\widetilde{\boldsymbol{H}} := \boldsymbol{H}_{*,\{a+1\ldots n\}}$, this is equivalent to

$$\widetilde{\boldsymbol{H}}\widetilde{\mathbf{e}}^\mathsf{T} = \sum_{i=1}^N \lambda_i \mathbf{s}_i^\mathsf{T}. \tag{16}$$

Equation (16) can be seen as $n-k$ parity-check equations which may be exploited by the classical combinatorial technique, see [23,8] or the discussion above Equation (6). The main difference here is that the right hand this equation also contains N unknowns $\lambda_i \in \mathbb{F}_q$. Still, we can pick a vector space V of dimension $r_1 \geq r$ and hoping that $\mathrm{Supp}(\widetilde{\mathbf{e}}) \subset V$. If this is the case, one can derive from (16) a linear system of $(n-k)m$ equations over $\mathbb{F}_q$ in $N + (n-a)r_1$ variables, where the first N variables merely correspond to the λ_i's. The final cost is then obtained by looking at the optimal value of r_1 which allows to solve this linear system, namely $r_1 := \left\lfloor \frac{m(n-k)-N}{n-a} \right\rfloor$. □

Thanks to Theorem 6, we are able to derive a value of N so that an RSL instance is solvable in polynomial time, this is the topic of Corollary 2.

Corollary 2 (New Bound for RSL). *An RSL instance with parameters (m, n, k, r, N) can be solved in polynomial time using the attack of Theorem 6 as long as*

$$N > kr\frac{m}{m-r}.$$

Proof. This is a straightforward application of the complexity given by Theorem 6 which states that the attack has a polynomial cost (hidden in the $\widetilde{\mathcal{O}}$) and an exponential cost of $q^{r(m-\delta)}$ where $\delta = \left\lfloor \frac{m(n-k)-N}{n-a} \right\rfloor$. Since $r \neq 0$ by definition, the only way for the exponential component to vanish is if $\delta = m$. Without loss of generality, we assume that $\frac{N}{r}$ and $\frac{m(n-k)-N}{n-N/r}$ are integers. By solving a simple equation, one gets that $\delta = m \iff N = kr\frac{m}{m-r}$, hence the result. □

Note that, as long as $\frac{m}{m-r}$, which is often the case for cryptographic parameters, our bound is lower than the previous one, given in [21], which was $N > nr$.

Algebraic attack of [10]. This attack consists in solving a bilinear system at some bi-degree $(b, 1)$ for $b \geq 1$ by using an XL approach similar to [12]. The two cases "$\delta = 0$" and "$\delta > 0$" presented below correspond to two different specializations of this bilinear system which lead to different costs. Here, we provide explicit formulas to compute these two complexities (for the binary field $\mathbb{F}_2$). In particular, we also include the values of α_R and α_λ which correspond to the hybrid approach mentioned in [10]. Finally, note that these formulas are valid only when $N > n-k-r$.

***First case:* $\delta = 0$.** Let a be the unique integer such that $ar < N \leq (a+1)r$, and let $N' := ar+1$. For $1 \leq b \leq r+1$, the number of variables for linearization is

$$\mathcal{M}^{\mathbb{F}_2}_{\leq b} := \sum_{i=1}^{b} \binom{n-a-\alpha_R}{r} \binom{N'-\alpha_\lambda}{i}, \tag{17}$$

where $0 \leq \alpha_R < n-a-r$, and $0 \leq \alpha_\lambda < N' - b$, and the number of linearly independent equations at hand is equal to $m\mathcal{N}_{\leq b}^{\mathbb{F}_2}$ where

$$\mathcal{N}_{\leq b}^{\mathbb{F}_2} := \sum_{i=1}^{b} \sum_{d=1}^{i} \sum_{j=1}^{n-k} \binom{j-1}{d-1} \binom{n-k-j}{r-d+1} \binom{N'-\alpha_\lambda-j}{i-d}. \tag{18}$$

The complexity is given by

$$\mathcal{O}\left(\min\left(2^{r\alpha_R+\alpha_\lambda} m\mathcal{N}_{\leq b}^{\mathbb{F}_2} (\mathcal{M}_{\leq b}^{\mathbb{F}_2})^{\omega-1}, \right.\right. \\ \left.\left. 2^{r\alpha_R+\alpha_\lambda} (N'-\alpha_\lambda) \binom{k-a+1+r}{r} (\mathcal{M}_{\leq b}^{\mathbb{F}_2})^2 \right) \right) \tag{19}$$

provided that $m\mathcal{N}_{\leq b}^{\mathbb{F}_2} \geq \mathcal{M}_{\leq b}^{\mathbb{F}_2} - 1$, and where the values of b, α_R, and α_λ are chosen to minimize the complexity.

$\delta > 0$ case. Let δ be a positive integer such that $N \geq \delta(n-r+\delta)$, let a be the greatest integer such that $N > \delta(n-r+\delta) + a(r-\delta)$ and let $N' := \delta(n-r+\delta) + a(r-\delta)$. To find the complexity of this attack, one replaces r by $r-\delta$ in the expressions of $\mathcal{M}_{\leq b}^{\mathbb{F}_2}$ and $\mathcal{N}_{\leq b}^{\mathbb{F}_2}$ from Equations (17) and (18). The complexity is finally obtained with Equation (19) and its minimal value now depends on $\delta > 0$ as well as b, α_R, and α_λ as above.

Visualization of the attacks against RSL. Last but not least, thanks to our analysis of the complexity to solve RSL with different attacks, we were able to draw a graph, see Figure 4, of the complexity to solve an RSD instance as a function of the number of given syndromes N.

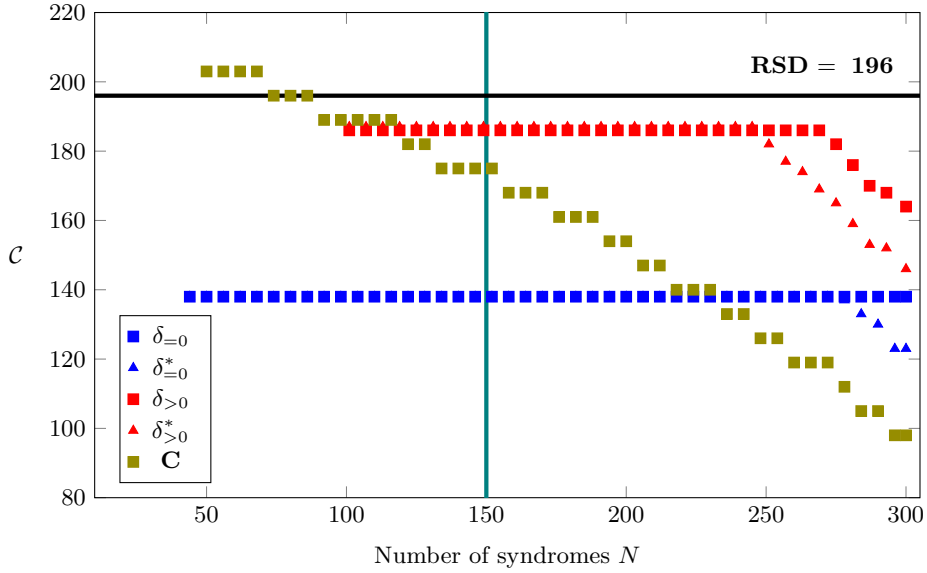
The instance parameters are $[m, n, k, r] = [61, 100, 50, 7]$, this is precisely the instance corresponding to attacking our scheme NH-Multi-RQC-AG-128 (see Table 1). The complexity to solve this RSD instance using the algebraic attack MaxMinors (see Section 5.1) is 196 bits; it corresponds to the horizontal black thick line. Starting with 44 syndromes; recall that it is the thresold for the algebraic attack against RSL (see Section 5.3), one sees that it beats the RSD attack. It is worth noticing that with approximately 225 syndromes, our new combinatorial attack against RSL (see Theorem 6), starts to beats the algebraic attack of [10]. And finally, one notices that, with a lot of syndromes, all the aforementioned RSL attacks complexities drops down, which is quite logical.

5.4 Combinatorial attack on NHRSL

In this section, we adapt the combinatorial attack against RSL, given in the proof of Theorem 6, to the case of non-homogeneous error, i.e. to the NHRSL problem (see Problem 4).

For the sake of simplicity, and since it is the case for all cryptographic parameters studied in this paper, we focus only on NHRSL instances where $n_1 < n$.

Fig. 4. Complexity $\mathcal{C}$ (in bits) of the best known attacks against an RSL instance with parameters $[m, n, k, r] = [61, 100, 50, 7]$ in terms of the number $N > n - k - r$ of syndromes. In the legend: **C** stands for our combinatorial attack (see Theorem 6), all the other symbols correspond to the 2 cases of the algebraic attack [10] where the "*" indicates the use of Wiedemann algorithm instead of Strassen's.



Theorem 7 (Combinatorial attack against NHRSL). *There exists a combinatorial attack against an NHRSL instance with parameters* (m, n, n_1, w_1, w_2) *whose complexity, in terms of elementary operations in* $\mathbb{F}_q$*, is given by*

$$\widetilde{\mathcal{O}}\left(q^{(w_1+w_2)(m-r)-w_2\rho}\right),$$

where r, ρ *are integers chosen to maximize the quantity* $(w_1 + w_2)r + w_2\rho$ *under the following constraints:* $N_1, N_2, r, \rho \in \mathbb{N}$, $N_1 + N_2 = N$, $w_1 \leq r$, $w_2 \leq \rho$, $r + \rho \leq m - 1$, $a := \left\lfloor \frac{N_1}{w_1} \right\rfloor \leq n_1$, $b := \left\lfloor \frac{N_2}{w_1+w_2} \right\rfloor \leq 2n$, $m(n + n_1) \geq (n_1 - b)(r + \rho) + (2n - a)r + N$.

Proof. Straightforward adaptation of the attack in the proof of Theorem 6, combined with the probability results given in Appendix A.1.

6 Security and parameters of our schemes

6.1 Security comparison for our schemes

According to Theorem 2, the security of Multi-RQC-AG relies on the Decisional Ideal Rank Syndrome Decoding problem (DIRSD) and on the Decisional Ideal

Non-Homogeneous Rank Support Learning problem (DNHIRSL). So far, there is no known attack to solve the decisional versions of these problems without solving the associated search instances. In addition to this, there is currently no attack that takes advantage of the ideal structure; thus, studying the security of Multi-RQC-AG comes down to evaluating the complexity of RSD and NHRSL. Unlike Multi-RQC-AG, our new scheme Multi-UR-AG does not use ideal structure. Despite the aforementioned absence of attack that exploits ideal structure, it might induce a weakness in a scheme. This is why Multi-RQC-AG, which does not use any structure like its name suggests it, is more secure. To study its complexity, according to Theorem 3, one has to study RSL and NHRSL. However, for an even better security, one could use Multi-UR-AG with homogeneous weight, making its security relying solely on RSL (see for instance the parameters sets Multi-UR-AG -128 and Multi-UR-AG -192 in Section 6).

6.2 Examples of parameters

In this section, we propose parameters for our different schemes, see Table 1; all parameters are chosen to resist to attacks described in Section 5. Among the different codes that can be attacked for each of our schemes (see proofs of Theorems 2 and 3), there is not a weaker one which enables us to fix all of our parameters. More precisely, sometimes attacking the public key, i.e. a code $[2n, n]$, gives the lowest complexity, but for another set of parameters, it will be the $[2n+n_1, n, w_1, w_2]$-code instead. However, there seems to be an invariant: no matter the length n of the code or the dimension m of the extension, it looks like the closer to GV bound the target rank r is, the better the combinatorial attacks are, and the worse are the algebraic attacks. In other words, for a given $[m, n, k]$-code, there seems to always be a value of r such that all the combinatorial attacks will beat the algebraic ones. This seems to be the case both for homogeneous and non-homogeneous versions of the aforementioned problems, and with or without multiple syndromes.

Instance	Struct.	m	n'	n	n_1	n_2	k	ε	w	w_1	w_2	DFR	Sizes in **KB**		
													pk	**ct**	Total
Loong-128 [37]	Random	191	182	35	13	14	6	0	8	11	0	0	10.9	16.0	**26.9**
Multi-RQC-AG-128	Ideal	83	82	-	5	38	2	74	7	11	0	-138	0.4	3.9	**4.4**
NH-Multi-RQC-AG-128	Ideal	61	60	-	3	50	3	51	7	7	5	-158	0.4	2.3	**2.7**
Multi-RQC-AG-192	Ideal	113	112	-	4	60	2	98	8	13	0	-215	0.9	6.8	**7.7**
NH-Multi-RQC-AG-192	Ideal	79	78	-	2	95	5	65	8	8	5	-238	0.9	3.8	**4.7**
Multi-UR-AG-128	Random	97	96	24	14	15	3	83	8	11	0	-190	4.1	6.9	**11.0**
NH-Multi-UR-AG-128	Random	73	72	22	13	14	2	66	8	8	4	-133	2.7	4.5	**7.1**
Multi-UR-AG-192	Random	127	126	35	15	16	3	93	9	12	0	-350	8.4	12.7	**21.1**
NH-Multi-UR-AG-192	Random	97	96	30	14	14	3	77	9	9	4	-214	5.1	7.5	**12.6**

Table 1. Parameters for our scheme

Similarly to [3], we use the fact that $1 \in \mathrm{Supp}(\mathbf{x}, \mathbf{y})$ to set $\delta := ww_1$ in the homogeneous case and $\delta := ww_1 + w_2$ in the non-homogeneous case. Recall

Instance	128 bits	192 bits
NH-Multi-UR-AG	**7,122**	**12,602**
LRPC-MS [1]	7,205	14,270
Multi-UR-AG	**11,026**	**21,075**
FrodoKEM [6]	19,336	31,376
Loong-128 [37]	26,948	-
Loidreau [35]	36,300	-
Classic McEliece [13]	261,248	524,348

Instance	128 bits	192 bits
NH-Multi-RQC-AG	**2,710**	**4,732**
ILRPC-MS [1]	2,439	4,851
BIKE [2]	3,113	6,197
Multi-RQC-AG	**4,378**	**7,668**
HQC [4]	6,730	13,548

Table 2. Comparison of sizes for unstructured (random) and structured (ideal) KEMs. The sizes represent the sum of the public key and the ciphertext, expressed in bytes.

that this quantity corresponds to the weight of the error decoded by the public Augmented Gabidulin code. For all our protocols (where "NH" denote non-homogeneous errors), both 128 and 192 bits security level are considered. As a comparison, we also updated the parameters of the code-based KEM Loong [37]. Note that this scheme does not use augmented Gabidulin codes nor non-homogeneous error but it does uses multiple syndromes.

The parameters sets given in Table 1 come with the sizes of the associated public key $\mathbf{pk}$ and ciphertext $\mathbf{ct}$ expressed in kilo-bytes (KB). For Multi-RQC-AG , $|\mathbf{pk}| = 40 + \left\lceil \frac{n_2 m}{8} \right\rceil$ and $|\mathbf{ct}| = \left\lceil \frac{2n_1 n_2 m}{8} \right\rceil$. For Multi-UR-AG , $|\mathbf{pk}| = 40 + \left\lceil \frac{nn_1 m}{8} \right\rceil$ and $|\mathbf{ct}| = \left\lceil \frac{m(nn_2 + n_1 n_2)}{8} \right\rceil$. The term 40 represents the length of a seed used to generate $(\mathbf{g}, \mathbf{h})$, recall that the public key consists in $(\mathbf{g}, \mathbf{h}, \mathbf{s})$ and the ciphertext in the couple $(\mathbf{u}, \mathbf{v})$. Note that the size of the secret key is not relevant since it is only a seed, thus it always has size 40 bytes.

To sum up, our most competitive set of parameters, in terms of sizes, is NH-Multi-RQC-AG-128 which uses ideal structure and non-homogeneous error; on the other side, the most secure set of parameters, whose security solely depends on RSL, and which does not use ideal structure, is Multi-UR-AG-128. Table 2 enables one to compare the sizes of our most competitive scheme to other KEMs using ideal or random (unstructured) matrices. Note that using non-homogeneous errors, our schemes are the shortest.

Last but not least, the vertical green line at $N = 150$ on Figure 4 shows the number of syndromes available for an adversary trying to attack a ciphertext of our scheme NH-MRQC-AG-128. It is worth noticing that, even though the blue squares are below the black line (complexity of the plain RSD attack), they are still way above the security level of 128 bits, and even given 150 syndromes, an attacker could not break our scheme. Note that it is far away from the area where the complexities of the different RSL attacks start to drop. More generally, we picked all our parameters that way, not only to resist to these attacks, but to be sure not to be targeted by any minor improvements.

7 Conclusion

In this paper, we introduce new variations on the RQC scheme, and more specifically, we introduce the Augmented Gabidulin codes which are very well suited to RQC. These new codes, together with the multiple syndrome and the non homogeneous approaches, lead to very small parameters which compare very well with other existing code-based schemes. In addition to this, we propose a meaningful scheme only relying on pure random instances, without any ideal structure and with small parameters, around 11KBytes.

We also study more deeply the security of the rank based problems used for our new schemes. Because of their properties, problems like NHRSD or RSL, are probably bound to be used in many future schemes based on rank metric.

Ackowledgements

The third author would like to thank Maxime Bombar for helpful discussion.

A Appendices

A.1 Computation of the success probability Π

To compute $\Pi := \Pr_{V,Z}[S_1 \subset V,\ S_2 \subset V \oplus Z]$ we use

Lemma 4. *Let $\Pi := \Pr_{V,Z}[S_1 \subset V,\ S_2 \subset V \oplus Z]$, where the randomness comes from the choice of a random subspace $V \subset \mathbb{F}_{q^m}$ and a random complementary subspace Z (hence isomorphic to a subspace of $\mathbb{F}_{q^m}/V$). We have*

$$\begin{aligned}\Pi &= \Pr[S_1 \subset V,\ S_2/S_1 \subset (V \oplus Z \oplus S_1)/S_1] \\ &= \Pr_V[S_1 \subset V] \Pr_{V,Z}[S_2/S_1 \subset (V \oplus Z \oplus S_1)/S_1 \mid S_1 \subset V] \\ &= \Pr_V[S_1 \subset V] \times \Pi_{cond},\end{aligned}$$

where $\Pi_{cond} := \Pr_{V,Z}[S_2/S_1 \subset (V \oplus Z \oplus S_1)/S_1 \mid S_1 \subset V]$.

Proof. The only non-trivial equality is the first one. For $\leq$, this is clear by taking the quotient by S_1. For $\geq$, let π_{S_1} denote the quotient map $\mathbb{F}_{q^m} \to \mathbb{F}_{q^m}/S_1$. The event at the right-hand side can be seen as $S_1 \subset V,\ \pi_{S_1}(S_2) \subset \pi_{S_1}(V \oplus Z \oplus S_1)$, and by considering the inverse image by π_{S_1}, this event is included in $S_1 \subset V,\ \pi_{S_1}^{-1}(\pi_{S_1}(S_2)) \subset \pi_{S_1}^{-1}(\pi_{S_1}(V \oplus Z \oplus S_1))$. This gives $S_1 \subset V$ and $S_2 + \ker(\pi_{S_1}) = S_1 + S_2 = S_2$. This space is included in $V \oplus Z \oplus S_1 + \ker(\pi_{S_1}) = V \oplus Z \oplus S_1 + S_1 = V \oplus Z \oplus S_1$, hence $S_1 \subset V$ and $S_2 \subset V \oplus Z$. □

We now focus on the Π_{cond} factor. Note that we have the decomposition

$$\begin{aligned}\{S_2/S_1 \subset (V \oplus Z \oplus S_1)/S_1 \mid S_1 \subset V\} &= \{S_2/S_1 \subset (V \oplus Z)/S_1\} \\ &= \coprod_{\ell=0}^{w_2} \left\{ \dim_{\mathbb{F}_q}(S_2/S_1 \cap V/S_1) = \ell,\ \frac{S_2/S_1}{S_2/S_1 \cap V/S_1} \subset \frac{(V \oplus Z)/S_1}{V/S_1} \right\} \\ &= \coprod_{\ell=0}^{w_2} \{A_\ell \cap B\},\end{aligned}$$

where A_ℓ : "$\dim_{\mathbb{F}_q}(S_2/S_1 \cap V/S_1) = \ell$" and B : "$\frac{S_2/S_1}{S_2/S_1 \cap V/S_1} \subset \frac{(V \oplus Z)/S_1}{V/S_1}$". For $0 \leq \ell \leq w_2$, let $p_\ell := \Pr[A_\ell \cap B]$, let $s_\ell := \Pr[A_\ell]$ and let $t_\ell := \Pr[B \mid A_\ell]$ so that $p_\ell = s_\ell t_\ell$ and $\Pi_{\text{cond}} = \sum_{\ell=0}^{w_2} p_\ell$. To compute s_ℓ, we rely on

Lemma 5 (§9.3.2 p. 269, [14]). *Let F be an $\mathbb{F}_q$-linear space of dimension n.*

1. *If X is a j-dimensional subspace of F, then there are $q^{ij}\binom{n-j}{i}_q$ i-dimensional subspaces Y such that $X \cap Y = 0$.*
2. *If X is a j-dimensional subspace of F, then there are $q^{(i-\ell)(j-\ell)}\binom{n-j}{i-\ell}_q\binom{j}{\ell}_q$ i-dimensional subspaces Y such that $X \cap Y$ has dimension ℓ.*

More precisely, we use Lemma 5, 2. with $F := \mathbb{F}_{q^m}/S_1$, fixed $X := S_2/S_1 \subset \mathbb{F}_{q^m}/S_1$ of dimension $j := w_2$ and random $Y := V/S_1 \subset \mathbb{F}_{q^m}/S_1$ of dimension $i := r - w_1$. We obtain

$$s_\ell = q^{(r-w_1-\ell)(w_2-\ell)} \frac{\binom{m-w_1-w_2}{r-w_1-\ell}_q \times \binom{w_2}{\ell}_q}{\binom{m-w_1}{r-w_1}_q}. \tag{20}$$

To compute t_ℓ, note that conditioned on $\dim_{\mathbb{F}_q}(S_2/S_1 \cap V/S_1) = \ell$ the probability that $\frac{S_2/S_1}{S_2/S_1 \cap V/S_1} \subset \frac{(V \oplus Z)/S_1}{V/S_1}$ is the probability that a random subspace of dimension ρ contains a fixed subspace of dimension $w_2 - \ell$ in the ambient space $\frac{\mathbb{F}_{q^m}/S_1}{V/S_1} \simeq \mathbb{F}_{q^m}/V$. From there we obtain $t_\ell = \frac{\binom{\rho}{w_2-\ell}_q}{\binom{m-r}{w_2-\ell}_q}$, and finally by combining this with Equation (20):

$$p_\ell = q^{(r-w_1-\ell)(w_2-\ell)} \frac{\binom{m-w_1-w_2}{r-w_1-\ell}_q \times \binom{w_2}{\ell}_q}{\binom{m-w_1}{r-w_1}_q} \times \frac{\binom{\rho}{w_2-\ell}_q}{\binom{m-r}{w_2-\ell}_q}.$$

A.2 Finishing the proof of Theorem 4

Obviously $p_0 < \Pi_{cond}$ and one can also easily show that $\Pi_{cond} = \Theta(p_0)$. By including the q^{-m} factor from [8], the number of $\mathbb{F}_q$-operations in the attack is

$$\mathcal{K} = \mathcal{O}\left(L \times \Pi^{-1} \times q^{-m}\right) = \widetilde{\mathcal{O}}\left(\Pr[C]^{-1} p_0^{-1} q^{-m}\right),$$

where $\Pr[C] := \Pr_V[S_1 \subset V]$ and where L is the polynomial factor coming from the linear algebra step whose exact formula is not relevant for the discussion. Using the classical $\binom{a}{b}_q = \Theta(q^{b(a-b)})$ when $\max(a,b) \to +\infty$ together with

$$p_0 = q^{(r-w_1)w_2} \frac{\binom{m-w_1-w_2}{r-w_1}_q}{\binom{m-w_1}{r-w_1}_q} \frac{\binom{\rho}{w_2}_q}{\binom{m-r}{w_2}_q},$$

we obtain $p_0 = \Theta\big(q^{(r-w_1)w_2} \times q^{-(r-w_1)w_2} \times q^{-w_2(m-r-\rho)}\big) = \Theta(q^{-w_2(m-r-\rho)})$, and similarly $\Pr[C] = \Theta\big(q^{-w_1(m-r)}\big)$. Therefore $\mathcal{K} = \widetilde{\mathcal{O}}\big(q^{(w_1+w_2)(m-r)-w_2\rho-m}\big)$, which is the statement of Theorem 4.

A.3 Proofs for the MaxMinors attack on NHRSD

Proof of Proposition 3. For item 1., let $J \subset \{1..n+n_1-1\}$, $\#J = w_1+w_2$ such that $P_J \in \mathcal{P}_{\text{lost}}$. By definition of $\mathcal{P}_{\text{lost}}$ the set $J+n+1$ has intersection $\leq w_2-2$ with $\{n+2..n+n_1\}$, hence any subset $T = T^- \cup T^+$, $T^- \subset \{1..n+1\}$, $T^+ \subset (J+n+1)$ satisfies $\#(T \cap \{n+1..n+n_1\}) \leq w_2 - 1$ since T^- might also contain $n+1$. This means that the corresponding minor variable c_T belongs to ζ and can be set to zero in P_J. Using the shape depicted in Equation (10), this implies that the whole P_J equation becomes zero. For item 2. , recall that the leading term of $P_J \in \mathcal{P}_{\text{indep}}$ is c_{J+n+1}. Moreover we have $\#(J \cap \{1..n_1-1\}) = \#(J+n+1 \cap \{n+2..n+n_1\}) \geq w_2$, which means $c_{J+n+1} \notin \zeta$. In particular, all the equations from $\mathcal{P}_{\text{indep}}$ keep the same leading terms after fixing the M minor variables to zero and therefore they remain linearly independent. The last statement on the number of variables in obvious.

Lemmata to prove Proposition 4.

Proof of Lemma 2. Using Equation (13), one has that the equations in $\mathcal{P}_{\text{rest},A}$ all have their monomials in $\mu_A := \big\{c_T,\ T \subset \{1..2n+n_1\},\ \#T = w_1+w_2,\ n+1 \in T,\ T \cap \{n+2..n+n_1\} = A\big\}$, and this set has size $\binom{2n}{w_1}$. Finally, note that μ_A and μ'_A are disjoint when $A \neq A'$, which concludes the proof.

Proof of Lemma 3, under assumptions. Using Equation (13), it is readily verified that the set of leading terms of all equations in $\mathcal{P}_{\text{rest},A}$ is

$$\tau_A := \big\{c_{\{n+1\}\cup A\cup U},\ U \subset \{(n+n_1+2)..(2n+n_1)\},\ \#U = w_1\big\},$$

and for instance note that the equation P_{J_U} with $J_U+n+1 = A \cup \{n+n_1+1\} \cup U$ has leading term $c_{\{n+1\}\cup A\cup U} \in \tau_A$. This already shows that $\dim_{\mathbb{F}_{q^m}} \langle \mathcal{P}_{\text{rest},A} \rangle \geq \#\tau_A = \binom{n-1}{w_1}$. For the converse inequality, we need to rely on some assumption on the randomness of the entries of the P_J's in $\mathbb{F}_{q^m}$ to argue that we cannot construct an element in $\langle \mathcal{P}_{\text{rest},A} \rangle$ whose leading term does not belong to τ_A with very high probability. First, note that the variables from $P_J \in \mathcal{P}_{\text{rest},A}$ with

$J+n+1 = A \cup V_J$ where $V_J = \left\{v_1^{(J)} < \cdots < v_{w_1+1}^{(J)}\right\}$ which belong to τ_A are the $c_{\{n+1\}\cup A\cup V_J\setminus\{v_j^{(J)}\}}$ for $1 \leq j \leq w_1+1$. To kill the leading term of P_J, one would then consider an equation with the same leading term, namely a P'_J with $J' \neq J$, $J'+n+1 = A \cup V_{J'}$ and such that $V_{J\setminus\{v_1^{(J)}\}} = V_{J'\setminus\{v_1^{(J')}\}} = B$ for some B. In this case, one can check that the only monomial from τ_A present in both P_J and $P_{J'}$ is $c_{\{n+1\}\cup A\cup B}$, so that $P_J + \lambda_{J'}P_{J'}$ contains at least $2w_1$ monomials from τ_A. Similarly, by using a third J'' one could kill at most one extra monomial in P_J and in the worst case one in $P_{J'}$ as well. This means that a linear combination of the form $P_J + \lambda_{J'}P_{J'} + \lambda_{J''}P_{J''}$ contains at least $2(w_1-1)+(w_1+1-2) = 3(w_1-1)$ monomials from τ_A, and the lower bound is reached if and only if those monomials in P_J and P'_J are killed at the same time by $\lambda_{J''}P_{J''}$. This is extremely unlikely if the coefficients of the MaxMinors equations are random elements in $\mathbb{F}_{q^m}$, so that we assume instead that $P_J+\lambda_{J'}P_{J'}+\lambda_{J''}P_{J''}$ contains at least $(w_1-1)+w_1+(w_1+1-1) = 3w_1-1$ monomials in τ_A. Relying on the same type of assumption, one can proceed by induction on the numbers of terms to show that a non-zero linear combination in $\langle \mathcal{P}_{\mathrm{rest},A} \rangle$ always has a monomial in τ_A.

References

1. Aguilar Melchor, C., Aragon, N., Dyseryn, V., Gaborit, P., Zémor, G.: LRPC codes with multiple syndromes: near ideal-size KEMs without ideals (2022), https://arxiv.org/abs/2206.11961
2. Aguilar Melchor, C., Aragon, N., Barreto, P., Bettaieb, S., Bidoux, L., Blazy, O., Deneuville, J.C., Gaborit, P., Gueron, S., Güneysu, T., Misoczki, R., Persichetti, E., Sendrier, N., Tillich, J.P., Zémor, G.: BIKE. Round 3 Submission to the NIST Post-Quantum Cryptography Call, v. 4.2 (Sep 2021)
3. Aguilar Melchor, C., Aragon, N., Bettaieb, S., Bidoux, L., Blazy, O., Bros, M., Couvreur, A., Deneuville, J.C., Gaborit, P., Zémor, G., Hauteville, A.: Rank quasi cyclic (RQC). Second Round submission to NIST Post-Quantum Cryptography call (Apr 2020)
4. Aguilar Melchor, C., Aragon, N., Bettaieb, S., Bidoux, L., Blazy, O., Deneuville, J.C., Gaborit, P., Persichetti, E., Zémor, G., Bos, J.: HQC. Round 3 Submission to the NIST Post-Quantum Cryptography Call (Jun 2021)
5. Alekhnovich, Michael: More on Average Case vs Approximation Complexity. In: 44th Symposium on Foundations of Computer Science (FOCS 2003), 11-14 October 2003, Cambridge, MA, USA, Proceedings. pp. 298–307. IEEE Computer Society (2003)
6. Alkim, E., Bos, J.W., Ducas, L., Longa, P., Mironov, I.: Frodokem. 3rd round submission to the nist (2021)
7. Aragon, N., Blazy, O., Gaborit, P., Hauteville, A., Zémor, G.: Durandal: a rank metric based signature scheme. In: Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19-23, 2019, Proceedings, Part III. vol. 11478, pp. 728–758. Springer (2019)
8. Aragon, N., Gaborit, P., Hauteville, A., Tillich, J.P.: A new algorithm for solving the rank syndrome decoding problem. In: 2018 IEEE International Symposium on

Information Theory, ISIT 2018, Vail, CO, USA, June 17-22, 2018. pp. 2421–2425. IEEE (2018)
9. Augot, D., Loidreau, P., Robert, G.: Generalized gabidulin codes over fields of any characteristic. Designs, Codes and Cryptography **86**(8), 1807–1848 (2018)
10. Bardet, M., Briaud, P.: An algebraic approach to the rank support learning problem. In: Cheon, J.H., Tillich, J.P. (eds.) PQCrypto. lncs, Springer International Publishing (2021)
11. Bardet, M., Briaud, P., Bros, M., Gaborit, P., Neiger, V., Ruatta, O., Tillich, J.: An algebraic attack on rank metric code-based cryptosystems. In: Advances in Cryptology - EUROCRYPT 2020 - 30th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May 10-14, 2020. Proceedings (2020)
12. Bardet, M., Bros, M., Cabarcas, D., Gaborit, P., Perlner, R., Smith-Tone, D., Tillich, J.P., Verbel, J.: Improvements of algebraic attacks for solving the rank decoding and minrank problems. In: ASIACRYPT 2020, International Conference on the Theory and Application of Cryptology and Information Security, 2020. Proceedings. pp. 507–536 (2020)
13. Bernstein, D.J., Chou, T., Lange, T., von Maurich, I., Misoczki, R., Niederhagen, R., Persichetti, E., Peters, C., Schwabe, P., Sendrier, N., et al.: Classic mceliece (2017)
14. Brouwer, A.E., Cohen, A.M., Neumaier, A.: Distance-Regular Graphs. No. 18 in Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge / A Series of Modern Surveys in Mathematics, Springer Verlag Berlin Heidelberg (1989)
15. Couvreur, A., Bombar, M.: Right-hand side decoding of gabidulin codes and applications. In: WCC 2022 (2022)
16. Debris-Alazard, T., Tillich, J.P.: Two attacks on rank metric code-based schemes: Ranksign and an identity-based-encryption scheme. In: 2018. vol. 11272, pp. 62–92. Springer, Brisbane, Australia (Dec 2018)
17. Gabidulin, E.M.: Theory of codes with maximum rank distance **21**(1), 3–16 (1985)
18. Gabidulin, E.M., Paramonov, A.V., Tretjakov, O.V.: Ideals over a non-commutative ring and their applications to cryptography. In: '91. pp. 482–489. No. 547, Brighton (Apr 1991)
19. Gabidulin, E.M., Pilipchuk, N.I.: Error and erasure correcting algorithms for rank codes. Designs, codes and Cryptography **49**(1), 105–122 (2008)
20. Gaborit, P., Galvez, L., Hauteville, A., Kim, J.L., Kim, M.J., Kim, Y.S.: Dual-ouroboros: an improvement of the mcnie scheme. Advances in Mathematics of Communications **14**(2), 301 (2020)
21. Gaborit, P., Hauteville, A., Phan, D.H., Tillich, J.: Identity-based encryption from rank metric. In: 2017. vol. 10403, pp. 194–226. Springer, Santa Barbara, CA, USA (Aug 2017)
22. Gaborit, P., Murat, G., Ruatta, O., Zémor, G.: Low rank parity check codes and their application to cryptography. In: Proceedings of the Workshop on Coding and Cryptography WCC'2013. Bergen, Norway (2013)
23. Gaborit, P., Ruatta, O., Schrek, J.: On the complexity of the rank syndrome decoding problem. IEEE Trans. Information Theory **62**(2), 1006–1019 (2016)
24. Gaborit, P., Ruatta, O., Schrek, J., Zémor, G.: New results for rank-based cryptography. In: 2014. vol. 8469, pp. 1–12 (2014)
25. Gaborit, P., Zémor, G.: On the hardness of the decoding and the minimum distance problems for rank codes **62(12)**, 7245–7252 (2016)

26. Hoffstein, J., Pipher, J., Silverman, J.H.: NTRU: A ring-based public key cryptosystem. In: Buhler, J. (ed.) Algorithmic Number Theory, Third International Symposium, ANTS-III, Portland, Oregon, USA, June 21-25, 1998, Proceedings. vol. 1423, pp. 267–288. Springer (1998)
27. Kabatianskii, G., Krouk, E., Smeets, B.J.M.: A digital signature scheme based on random error-correcting codes. In: IMA Int. Conf. vol. 1355, pp. 161–167. Springer (1997)
28. Loidreau, P.: Properties of codes in rank metric (2006)
29. Loidreau, P.: A new rank metric codes based encryption scheme. In: 2017. vol. 10346, pp. 3–17. Springer (2017)
30. Misoczki, R., Tillich, J.P., Sendrier, N., Barreto, P.S.L.M.: MDPC-McEliece: New McEliece variants from moderate density parity-check codes. pp. 2069–2073 (2013)
31. Ore, O.: On a special class of polynomials. Trans. Amer. Math. Soc. **35**(3), 559–584 (1933)
32. Otmani, A., Talé-Kalachi, H., Ndjeya, S.: Improved cryptanalysis of rank metric schemes based on Gabidulin codes **86**(9), 1983–1996 (2018)
33. Otmani, A., Tillich, J.P.: An efficient attack on all concrete KKS proposals. In: 2011. vol. 7071, pp. 98–116 (2011)
34. Overbeck, R.: A new structural attack for GPT and variants. In: Mycrypt. vol. 3715, pp. 50–63 (2005)
35. Pham, B.D.: Étude et conception de nouvelles primitives de chiffrement fondées sur les codes correcteurs d'erreurs en métrique rang. Ph.D. thesis, Rennes 1 (2021)
36. Reed, I.S., Solomon, G.: Polynomial codes over certain finite fields. Journal of the society for industrial and applied mathematics **8**(2), 300–304 (1960)
37. Wang, L.P.: Loong: a new ind-cca-secure code-based kem. In: 2019 IEEE International Symposium on Information Theory (ISIT). pp. 2584–2588. IEEE (2019)