Durandal: a rank metric based signature scheme

Nicolas Aragon*† Olivier Blazy* Philippe Gaborit* Adrien Hauteville*
Gilles Zémor‡

December 9, 2018

Abstract

We describe a variation of the Schnorr-Lyubashevsky approach to devising signature schemes that is adapted to rank based cryptography. This new approach enables us to obtain a randomization of the signature, which previously seemed difficult to derive for code-based cryptography. We provide a detailed analysis of attacks and an EUF-CMA proof for our scheme. Our scheme relies on the security of the Ideal Rank Support Learning and the Ideal Rank Syndrome problems and a newly introduced problem: Product Spaces Subspaces Indistinguishability, for which we give a detailed analysis. Overall the parameters we propose are efficient and comparable in terms of signature size to the Dilithium lattice-based scheme, with a signature size of less than 4kB for a public key of size less than 20kB.

1 Introduction

During the last few years and especially since the 2017 call for proposals of the NIST for post-quantum cryptosystems, there has been a burst of activity in post-quantum cryptography and notably in code-based cryptography.

As far as encryption schemes are concerned, code-based cryptography has satisfactory solutions, in the form of cryptosystems whose security is reduced to well known problems: decoding random structured matrices like ideal or quasi-cyclic matrices ([2, 4, 3]). However, the situation is very different for signature schemes.

Essentially there exist two types of signature schemes: hash-and-sign schemes and proof of knowledge based signatures.

For hash-and-sign signature schemes, signing consists in finding a small weight preimage of a random syndrome, with a non-negligible probability. For instance: CFS in code based cryptography ([7]), GPV for lattices [21], Ranksign for rank metric [17], NTRUSign for lattices [25], pqsigRM [26]... The main drawback of this approach is that

*Université de Limoges, XLIM-DMI, 123, Av. Albert Thomas, 87060 Limoges Cedex, France.
†This work was partially funded by French DGA
‡Université de Bordeaux, Institut de Mathématiques, UMR 5251, 351 cours de la Libération, 33400 Talence, France.

the system relies on hiding a trapdoor within the public key: typically the secret is a decoding (or approximate decoding) algorithm which is hidden in the public matrix that describes the code. Whereas for lattices this type of masking can be efficiently randomized because of properties of the Euclidean distance [21], it has proved much more difficult for coding theory. In practice there exist two published code-based signature schemes: CFS [7] and RankSign [18] (see also SURF [8]), but for these schemes the public key can be distinguished from a random matrix [11], [9]. Overall, for signature, this approach is similar to classical McEliece Encryption for which there is always a sword of Damocles lying over its head, namely the possible existence of a structural attack which recovers the hidden structure of and hence breaks the scheme. Relating the distinguishing problem to another well known problem seems a difficult feature to obtain. For the case of the RankSign scheme, a structural attack was recently found in [9]; it is always possible to repair and counter such attacks, like it was the case for all the sequels of NTRUSign [25], but this illustrates the difficulty of relying on this approach, when the secret trapdoor is not randomized.

The second approach for devising a signature scheme consists in proving that one knows a small weight vector associated to a given syndrome. It can done in two ways.

A first way consists in considering a zero-knowledge authentication algorithm and turning it into a signature scheme through the Fiat-Shamir transform. If the probability of cheating (associated to soundness) is very small, this approach can be efficient, but when the cheating probability is of order $1/2$, it leads to very large signature sizes, since the number of necessary rounds is very large. It is typically the case for the Stern authentication protocol [32] for which the cheating probability is $2/3$ (it was decreased to $1/2$ in [1] and adapted to the rank metrix in [19]). Overall this approach is very interesting in terms of security reductions since one is reduced to generic problems *without* any masking, but rather inefficient in terms of signature size which can easily reach several hundred thousand bits, which is questionable in practice.

A second approach was proposed in a sequence of papers initiated by V. Lyubashevsky [28] in 2009. This approach is in the spirit of the Schnorr signature scheme [31] but adapted to the lattice context. The idea works as follows: for a public random matrix H, the secret is a matrix S of small weight vectors, to which one associates a matrix of syndromes HS^T. The signature consists in a proof of knowledge of the small weight matrix S from a sparse challenge c. The signature has the form $z = y + cS$, for y a random vector of moderate weight, typically of several orders of magnitude higher than the weights of cS. The idea of the proof of knowledge is that through z the verifier is convinced that the prover knows the secret matrix S because of the use of cS in the signature. At the same time, the vector y guarantees the randomization of the signature since its more noisy distribution enables it to absorb the less noisy distribution of cS. The main appeal of this approach is that it enables one to avoid the repetition related to zero-knowledge protocols with high probability of cheating, for instance the Dilithium [30] signature of NIST has a length of only 4kB.

This previous approach can be straightforwardly adapted globally to code-based cryptography, but there is a problem is the randomization part: for the Hamming metric the

randomization has to be considered on the whole length of the word, and not only on independent coordinates as when dealing with the Euclidean metric. In practice it means that it seems difficult to randomize the signature [12]: consequently, whenever a signature is produced, information leaks from the secret, so that after only a few signatures it becomes possible to recover the whole secret.

Overall, this second approach seems very promising but finding a good randomization strategy is a challenge.

Our contribution.

We build upon the Schnorr-Lyubashevsky approach in a rank metric context and propose a way to efficiently randomize the signature. The main idea consists in extending the number of small weight secret vectors and adding another secret matrix S', so that the signature has the form $z = y + cS + pS'$ where p serves the purpose of providing extra randomization. In this way, the prover benefits from relaxed conditions that he uses to derive a randomization of the signature. We give a proof in the EUF-CMA security model, reducing the security of the scheme to the Rank Support Learning (RSL), the (ideal) Rank Syndrome Decoding (RSD) problem and a newly introduced problem, the Product Spaces Subspaces Indistinguishability (PSSI) problem for which we give a detailed analysis of a distinguisher. Our approach is developed for the rank metric and does not have an obvious Hamming metric counterpart. Overall our scheme is efficient in terms of signature size (a few kB) and of key sizes of (of order 20kB), with a security reduction to the ideal-RSD problem (a generalization of QCRSD problem).

Roadmap.

The paper is organized as follows: Section 2 recalls the required material from rank based cryptography, Section 3 gives a general overview and a precise description of the scheme, Section 4 is concerned with the security of the scheme. Finally, Sections 5 and 6 describe main practical attacks and examples of parameters for our scheme.

2 Presentation of rank metric codes

Notations

In what follows, q denotes a power of a prime p. The finite field with q elements is denoted by $\mathbb{F}_q$ and for any positive integer m the finite field with q^m elements is denoted by $\mathbb{F}_{q^m}$. We will frequently view $\mathbb{F}_{q^m}$ as an m-dimensional vector space over $\mathbb{F}_q$. The Grassmannian $\mathbf{Gr}(k, \mathbb{F}_{q^m})$ represents the set of all subspaces of $\mathbb{F}_{q^m}$ of dimension k.

We use bold lowercase and capital letters to denote vectors and matrices respectively.

2.1 General definitions

Let us begin by defining the rank metric.

Definition 1 (Rank metric over $\mathbb{F}_{q^m}^n$)**.** *Let $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{F}_{q^m}^n$ and $(\beta_1, \ldots, \beta_m) \in \mathbb{F}_{q^m}^m$ be a basis of $\mathbb{F}_{q^m}$ viewed as an m-dimensional vector space over $\mathbb{F}_q$. Each coordinate*

x_j is associated to a vector of $\mathbb{F}_q^m$ in this basis: $x_j = \sum_{i=1}^m m_{ij}\beta_i$. The $m \times n$ matrix associated to $\boldsymbol{x}$ is given by $\boldsymbol{M}(\boldsymbol{x}) = (m_{ij})_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}}$.

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

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

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

This metric is adapted to a particular family of codes: the $\mathbb{F}_{q^m}$-linear codes.

Definition 2 ($\mathbb{F}_{q^m}$-linear code)**.** *An $\mathbb{F}_{q^m}$-linear code $\mathcal{C}$ of dimension k and length n is a subspace of dimension k of $\mathbb{F}_{q^m}^n$ embedded with the rank metric. In this case we speak of an $[n, k]_{q^m}$ code.*

$\mathcal{C}$ can be represented in two equivalent ways:

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

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

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

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

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

As for the Hamming metric, a crucial notion in rank metric is the notion of the support of a word. This notion appears very often in rank metric code-based cryptography, especially to compute the complexity of some algorithms.

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

$$E = \langle x_1, \ldots, x_n \rangle_{\mathbb{F}_q}$$

This definition is coherent with the definition of the rank weight since we have $\dim E = \|\boldsymbol{x}\|$.

The number of supports of dimension w of $\mathbb{F}_{q^m}$ is denoted by the Gaussian coefficient

$$\begin{bmatrix} m \\ w \end{bmatrix}_q = \prod_{i=0}^{w-1} \frac{q^m - q^i}{q^w - q^i} = \Theta(q^{w(m-w)})$$

We also need to define homogeneous matrices.

Definition 4 (Homogeneous matrices)**.** *Let $\boldsymbol{M} \in \mathbb{F}_{q^m}^{k \times n}$ be a matrix over $\mathbb{F}_{q^m}$. The matrix $\boldsymbol{M} = (m_{ij})$ is said to be homogeneous of support E if the $\mathbb{F}_q$-subspace of $\mathbb{F}_{q^m}$ spanned by its coefficients m_{ij} is equal to E. If $d = \dim E$, then $\boldsymbol{M}$ is also said to be homogeneous of weight d.*

2.2 Double circulant and ideal codes

To describe an $[n,k]_{q^m}$ linear code, we can give its systematic generator matrix or its systematic parity-check matrix. In both cases, the number of bits needed to represent such a matrix is $k(n-k)m\lceil \log_2 q \rceil$. To reduce the size of the representation of a code, we introduce double circulant codes.

First we need to define circulant matrices.

Definition 5 (Circulant matrix)**.** *A square $n \times n$ matrix $\boldsymbol{M}$ is said to be circulant if it is of the form*

$$\boldsymbol{M} = \begin{pmatrix} m_0 & m_1 & \dots & m_{n-1} \\ m_{n-1} & m_0 & \ddots & m_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ m_1 & m_2 & \dots & m_0 \end{pmatrix}$$

We denote $\mathcal{M}_n(\mathbb{F}_{q^m})$ the set of circulant matrices of size $n \times n$ over $\mathbb{F}_{q^m}$.

The following proposition states an important property of circulant matrices.

Proposition 6. *$\mathcal{M}_n(\mathbb{F}_{q^m})$ is an $\mathbb{F}_{q^m}$-algebra isomorphic to $\mathbb{F}_{q^m}[X]/(X^n-1)$. The canonical isomorphism is given by*

$$\begin{aligned} \varphi : \mathbb{F}_{q^m}[X]/(X^n-1) &\longrightarrow \mathcal{M}_n(\mathbb{F}_{q^m}) \\ \sum_{i=0}^{n-1} m_i X^i &\longmapsto \begin{pmatrix} m_0 & m_1 & \dots & m_{n-1} \\ m_{n-1} & m_0 & \ddots & m_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ m_1 & m_2 & \dots & m_0 \end{pmatrix} \end{aligned}$$

In the following, in order to simplify notation, we will identify the polynomial $G(X) = \sum_{i=0}^{n-1} g_i X^i \in \mathbb{F}_{q^m}[X]$ with the vector $\boldsymbol{g} = (g_0, \dots, g_{n-1}) \in \mathbb{F}_{q^m}^n$. We will denote $\boldsymbol{ug} \mod P$ the vector of the coefficients of the polynomial

$$\left(\sum_{j=0}^{n-1} u_j X^j\right)\left(\sum_{i=0}^{n-1} g_i X^i\right) \mod P$$

or simply $\boldsymbol{ug}$ if there is no ambiguity in the choice of the polynomial P.

Definition 7 (Double circulant codes)**.** *A $[2n,n]_{q^m}$ linear code $\mathcal{C}$ is said to be double circulant if it has a generator matrix $\boldsymbol{G}$ of the form $\boldsymbol{G} = (\boldsymbol{A}|\boldsymbol{B})$ where $\boldsymbol{A}$ and $\boldsymbol{B}$ are two circulant matrices of size n.*

With the previous notation, we have $\mathcal{C} = \{(\boldsymbol{xa}, \boldsymbol{xb}), \boldsymbol{x} \in \mathbb{F}_{q^m}^n\}$. If $\boldsymbol{a}$ is invertible in $\mathbb{F}_{q^m}[X]/(X^n-1)$, then $\mathcal{C} = \{(\boldsymbol{x}, \boldsymbol{xg}), \boldsymbol{x} \in \mathbb{F}_{q^m}^n\}$ where $\boldsymbol{g} = \boldsymbol{a}^{-1}\boldsymbol{b}$. In this case we say that $\mathcal{C}$ is generated by $\boldsymbol{g} \pmod{X^n-1}$. Thus we only need $nm\lceil \log_2 q \rceil$ bits to describe a $[2n,n]_{q^m}$ double circulant code.

We can generalize double circulant codes by choosing another polynomial P, rather than X^n-1, to define the quotient-ring $\mathbb{F}_{q^m}[X]/(P)$. These codes are called ideal codes.

Definition 8 (Ideal codes)**.** *Let $P(X) \in \mathbb{F}_q[X]$ be a polynomial of degree n and $\boldsymbol{g}_1, \boldsymbol{g}_2 \in \mathbb{F}_{q^m}^n$. Let $G_1(X) = \sum_{i=0}^{n-1} g_{1i} X^i$ and $G_2(X) = \sum_{j=0}^{n-1} g_{1j} X^j$ be the polynomials associated respectively to $\boldsymbol{g}_1$ and $\boldsymbol{g}_2$.*

The $[2n, n]_{q^m}$ ideal code $\mathcal{C}$ with generator $(\boldsymbol{g}_1, \boldsymbol{g}_2)$ is the code with generator matrix

$$\boldsymbol{G} = \left(\begin{array}{cc|cc} G_1(X) & \bmod P & G_2(X) & \bmod P \\ XG_1(X) & \bmod P & XG_2(X) & \bmod P \\ \multicolumn{2}{c|}{\vdots} & \multicolumn{2}{c}{\vdots} \\ X^{n-1}G_1(X) & \bmod P & X^{n-1}G_2(X) & \bmod P \end{array}\right)$$

More concisely, we have $\mathcal{C} = \{(\boldsymbol{x}\boldsymbol{g}_1 \mod P, \boldsymbol{x}\boldsymbol{g}_2 \mod P), \boldsymbol{x} \in \mathbb{F}_{q^m}^n\}$. We will often omit mentioning the polynomial P if there is no ambiguity.

If $\boldsymbol{g}_1$ is invertible, we may express the code in systematic form, $\mathcal{C} = \{(\boldsymbol{x}, \boldsymbol{x}\boldsymbol{g}), \boldsymbol{x} \in \mathbb{F}_{q^m}^n\}$ with $\boldsymbol{g} = \boldsymbol{g}_1^{-1}\boldsymbol{g}_2 \mod P$.

The advantage of ideal codes over double circulant codes is that they are resistant to the folding attack of [24]. Such codes have been used for NIST propositions LAKE and LOCKER.

2.3 Difficult problems in rank metric

In order to design rank metric code-based cryptosystems, we need to define difficult problems in rank metric. The first problem corresponds to the classical problem of syndrome decoding, adapted to the rank metric.

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

This problem is probabilistically reduced to the well-known NP-complete Syndrome Decoding problem in the Hamming metric [20].

The following problem was introduced in [13]. It is similar to the RSD problem, the difference is that instead of having one syndrome, we are given several syndromes of errors of same support and the goal is to find this support.

Problem 2. ***Rank Support Learning (RSL)*** *[14] Let $\boldsymbol{H}$ be a random full-rank $(n-k) \times n$ matrix over $\mathbb{F}_{q^m}$.*

Let $\mathcal{O}$ be an oracle which, given $\boldsymbol{H}$, gives samples of the form $\boldsymbol{H}\boldsymbol{s}_1^T, \boldsymbol{H}\boldsymbol{s}_2^T, \ldots, \boldsymbol{H}\boldsymbol{s}_N^T$, with the vectors $\boldsymbol{s}_i$ randomly chosen from a space E^n, where E is a random subspace of $\mathbb{F}_{q^m}$ of dimension r. The $RSL_{q,m,n,k,r}$ problem is to recover E given only access to the oracle.

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

$$(\boldsymbol{H}, \boldsymbol{H}\boldsymbol{s}_1^T, \boldsymbol{H}\boldsymbol{s}_2^T, \ldots, \boldsymbol{H}\boldsymbol{s}_N^T)$$

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

The last problem we need before introducing our scheme is a variant of the RSD problem. Instead of searching for the error associated to a syndrome, this problem consists in finding an error associated to a syndrome which belongs to a given $\mathbb{F}_q$-affine subspace of $\mathbb{F}_{q^m}^{n-k}$. Formally:

Problem 3. ***Affine Rank Syndrome Decoding (***ARSD***)*** *Let* $\boldsymbol{H}$ *be an* $(n-k) \times n$ *parity-check matrix of an* $[n,k]$ $\mathbb{F}_{q^m}$*-linear code,* $\boldsymbol{H}'$ *an* $(n-k) \times n'$ *random matrix over* $\mathbb{F}_{q^m}$*,* F *an* $\mathbb{F}_q$*-subspace of* $\mathbb{F}_{q^m}$ *of dimension* r'*,* $\boldsymbol{s} \in \mathbb{F}_{q^m}^{n-k}$ *and* r *an integer. The* $\mathrm{ARSD}_{q,m,n,k,r,n',F}$ *problem is to find* $\boldsymbol{e} \in \mathbb{F}_{q^m}^n$ *and* $\boldsymbol{e}' \in \mathbb{F}_{q^m}^{n'}$ *such that*

$$\begin{cases} \boldsymbol{H}\boldsymbol{e}^T + \boldsymbol{H}'\boldsymbol{e}'^T = \boldsymbol{s} \\ \|\boldsymbol{e}\| = r \\ \mathrm{Supp}(\boldsymbol{e}') \subseteq F \end{cases}$$

Remark: This problem can seen as that of finding a vector $\boldsymbol{x}$ of weight r such that $\boldsymbol{H}\boldsymbol{x}^T = \boldsymbol{s}'$ with $\boldsymbol{s}' \in \{\boldsymbol{s} - \boldsymbol{H}'\boldsymbol{x}'^T : \mathrm{Supp}(\boldsymbol{x}') \subseteq F\}$. This set is an $\mathbb{F}_q$-affine subspace of $\mathbb{F}_{q^m}^{n-k}$, which explains the name of the problem.

The following proposition shows that the ARSD problem in the worst case is as hard as the RSD problem for large values of m.

Proposition 9. *Let* $\mathcal{A}$ *be an algorithm which can solve the* $\mathrm{ARSD}_{q,m,n,k,r,n',F}$ *problem with* $m \geqslant \frac{r(n-r)+n' \dim F}{n-k-r}$*. Then* $\mathcal{A}$ *can be used to solve the* $\mathrm{RSD}_{q,m,n,k,r}$ *problem with non negligible probability.*

Proof. Let $\boldsymbol{H} \in \mathbb{F}_{q^m}^{(n-k)\times n}, \boldsymbol{s} \in \mathbb{F}_{q^m}^{n-k}$ such that $\boldsymbol{s} = \boldsymbol{H}\boldsymbol{e}^T$ with $\|\boldsymbol{e}\| = r$ be an instance of the $\mathrm{RSD}_{q,m,n,k,r}$ problem. First we need to transform this instance into an instance of the ARSD problem. Let $\boldsymbol{H}' \in \mathbb{F}_{q^m}^{(n-k)\times n'}$ and let F be a subspace of $\mathbb{F}_{q^m}$ of dimension r' such that $m \geqslant \frac{r(n-r)+n' \dim F}{n-k-r}$. Let $\boldsymbol{s}' = \boldsymbol{s} + \boldsymbol{H}'\boldsymbol{e}'^T$ with $\mathrm{Supp}(\boldsymbol{e}') = F$.

$(\boldsymbol{H}, \boldsymbol{s}', r, \boldsymbol{H}', F)$ is an instance of the $\mathrm{ARSD}_{q,m,n,k,r,n',F}$ problem. Let $(\boldsymbol{x}, \boldsymbol{x}')$ be a solution of this instance given by algorithm $\mathcal{A}$. Now we will prove that this solution is unique with a non negligible probability.

Let us consider the application f defined by

$$\begin{aligned} f : \mathcal{S}_{\mathbb{F}_{q^m}^n}(r) \times F^{n'} &\to \quad \mathbb{F}_{q^m}^{n-k} \\ (\boldsymbol{x}, \boldsymbol{x}') &\mapsto \boldsymbol{H}\boldsymbol{x}^T + \boldsymbol{H}'\boldsymbol{x}'^T \end{aligned}$$

where $\mathcal{S}_{\mathbb{F}_{q^m}^n}(r)$ is the set of words of $\mathbb{F}_{q^m}^n$ of rank r. By definition of the ARSD problem, we have $(\boldsymbol{x}, \boldsymbol{x}') \in f^{-1}(\{\boldsymbol{s}'\})$.

Let $S(\mathbb{F}_{q^m}^n, r)$ denote the cardinality of $\mathcal{S}_{\mathbb{F}_{q^m}^n}(r)$. By definition of the rank metric, $S(\mathbb{F}_{q^m}^n, r)$ is equal to the number of matrices of $\mathbb{F}_q^{m\times n}$ of rank r and we have

$$S(\mathbb{F}_{q^m}^n, r) = \prod_{i=0}^{r-1} \frac{(q^m - q^i)(q^n - q^i)}{q^r - q^i} = \Theta\left(q^{r(m+n-r)}\right).$$

Thus the cardinality of the codomain of f is in $\Theta(q^{r(n+n-r)+n'r'})$ and the cardinality of its domain is equal to $q^{m(n-k)}$. We have $m \geqslant \frac{r(n-r)+n'r'}{n-k-r}$ which implies $m(n-k) \geqslant r(m+n-r)+n'r'$, hence $\boldsymbol{s}'$ has only one preimage with a non negligible probability. Thus $\boldsymbol{H}\boldsymbol{x}^T + \boldsymbol{H}'\boldsymbol{x}'^T = \boldsymbol{s}' = \boldsymbol{H}\boldsymbol{e}^T + \boldsymbol{H}\boldsymbol{e}'^T$ implies $(\boldsymbol{x}, \boldsymbol{x}') = (\boldsymbol{e}, \boldsymbol{e}')$ so $\boldsymbol{x}$ is a solution of the instance of the $\mathrm{RSD}_{q,m,n,k,r}$ problem. □

Remark: All these problems are defined for random codes but can straightforwardly be specialized to the families of double circulant codes or of ideal random codes. In this case, these problems are denoted $\mathrm{I}-\mathrm{RSD}$, $\mathrm{I}-\mathrm{ARSD}$ and $\mathrm{I}-\mathrm{RSL}$ respectively. The reductions are unchanged, the only difference being that the $\mathrm{I}-\mathrm{RSD}$ problem is reduced to the Syndrome Decoding problem for ideal codes, which has not been proven NP-complete. However this problem is considered hard by the community since the best attacks stay exponential.

2.4 Bounds on rank metric codes

One can define bounds on the size or the minimum distance of rank metric codes that are similar to well-known bounds for Hamming metric codes. The rank Gilbert-Varshamov bound (or rank Gilbert-Varshamov distance, denoted d_{RGV}) corresponds to the case when the RSD problem has typically a unique solution.

Definition 10 (Rank Gilbert-Varshamov (RGV) bound)**.** *Let $B(\mathbb{F}_{q^m}^n, t)$ be the size of the ball of radius t in rank rank metric. The quantity d_{RGV} is defined as the smallest t such that $B(\mathbb{F}_{q^m}^n, t) \geqslant q^{m(n-k)}$.*

Asymptotically we have

$$d_{RGV}(m,n,k) \sim \frac{m+n-\sqrt{(m-n)^2+4km}}{2}$$
$$d_{RGV}(n,n,k) \sim n\left(1-\sqrt{\frac{k}{n}}\right) \text{ when } m=n. \quad (1)$$

The quantity $q^{m(n-k)}$ corresponds to the number of syndromes $\boldsymbol{s} \in \mathbb{F}_{q^m}^{n-k}$ and by definition,

$$B(\mathbb{F}_{q^m}^n, t) = \sum_{i=0}^{t} S(\mathbb{F}_{q^m}^n, i)$$

where $S(\mathbb{F}_{q^m}^n, i)$ is the size of the sphere of radius i, which correspond to the number of matrices of size $m \times n$ and of rank i over $\mathbb{F}_q$. This quantity is equal to $\prod_{j=0}^{i-1} \frac{(q^m-q^j)(q^n-q^j)}{q^i-q^j} = \Theta\left(q^{i(m+n-i)}\right)$. The asymptotic expressions are obtained by solving for t the equation $t(m+n-t) = m(n-k)$ [27].

The rank Singleton bound corresponds to the case when the RSD problem becomes polynomial.

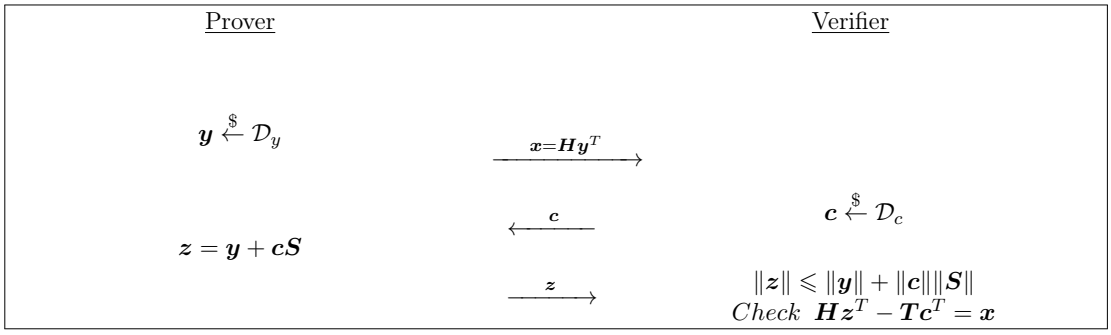

Figure 1: Overview of the authentication framework from [29]

Definition 11 (Rank Singleton bound)**.** *The rank Singleton bound $d_{RS}(m,n,k)$ for an $\mathbb{F}_{q^m}$-linear $[n,k]$ code is defined as*

$$d_{RS}(m,n,k) = \frac{m(n-k)}{\max(m,n)}.$$

We can obtain this equality by counting the number of equations and unknowns over $\mathbb{F}_q$ of the RSD problem. Indeed, given a random support E of dimension r, we can express the error $\boldsymbol{e}$ in a basis of E with nr unknowns over $\mathbb{F}_q$ (r unknowns per coordinate). The parity-check equations gives us $(n-k)$ equations over $\mathbb{F}_{q^m}$, meaning $m(n-k)$ equations over $\mathbb{F}_q$. If $nr \geqslant m(n-k)$ then this instance of the RSD problem has a solution $\boldsymbol{e}$ of support E with a non-negligible probability. Such a solution can easily be found by solving a linear system.Therefore, the RSD problem becomes polynomial if $r \geqslant \left\lceil \frac{m(n-k)}{n} \right\rceil$.

3 A new signature scheme based on the RSL problem

3.1 General overview

Our scheme consists of adapting to the rank metric the idea proposed in [29]. This idea can be viewed as a framework for an authentication scheme and can be loosely described as follows. Two matrices, over some fixed finite field, $\boldsymbol{H}$ and $\boldsymbol{T}$, are public, and a Prover wishes to prove that she is in possession of secret matrix $\boldsymbol{S}$ with "small" entries such that $\boldsymbol{T} = \boldsymbol{H}\boldsymbol{S}^T$. She chooses a random vector $\boldsymbol{y}$ of small norm (to be defined appropriately) according to some distribution $\mathcal{D}_y$. She computes the syndrome $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{y}^T$ of $\boldsymbol{y}$ and sends it to the verifier. The verifier chooses a random vector $\boldsymbol{c}$ of the appropriate length and of small norm according to some distribution $\mathcal{D}_c$ and sends it as a challenge to the Prover. The Prover computes $\boldsymbol{z} = \boldsymbol{y} + \boldsymbol{c}\boldsymbol{S}$ and sends it to the Verifier. The verifier checks that $\boldsymbol{z}$ is of small norm, and that

$$\boldsymbol{H}\boldsymbol{z}^T - \boldsymbol{T}\boldsymbol{c}^T = \boldsymbol{x}.$$

This scheme is described on Figure 1.

The general idea is that cheating is difficult for the Prover because it requires finding a vector $\boldsymbol{z}$ of small norm such that $\boldsymbol{H}\boldsymbol{z}^T$ equals a prescribed value, and this is an instance of the decoding problem for a random code. Also, the vector $\boldsymbol{z}$ sent by the legitimate Prover should yield no useful information on the secret $\boldsymbol{S}$, because the noisy random vector $\boldsymbol{y}$ drowns out the sensitive quantity $\boldsymbol{cS}$.

If we instantiate this scheme in the rank metric, $\boldsymbol{H}$ would be a random matrix over $\mathbb{F}_{q^m}$, and for $\boldsymbol{S}$ to be a matrix of small norm would mean it to be homogeneous matrix of some small rank r. Requiring that the vectors $\boldsymbol{y}$ and $\boldsymbol{c}$ are also small will mean that they are taken in random subspaces of $\mathbb{F}_{q^m}$ of fixed dimensions respectively w and d.

The problem with this approach in the rank metric is that adding $\boldsymbol{y}$ to $\boldsymbol{cS}$ does not hide $\boldsymbol{cS}$ properly. Indeed, the verifier, or any witness to the protocol of Figure 1, can recover the support of the secret matrix $\boldsymbol{S}$ even after a single instance of the protocol, using techniques from the decoding of LRPC codes [15]: since the verifier has $\boldsymbol{c}$ he can choose a basis $f_1, \ldots, f_d$ of $\mathrm{Supp}(\boldsymbol{c})$ and then with high probability it will occur that:

$$\bigcap_{i=1}^{d} f_i^{-1} \mathrm{Supp}(\boldsymbol{z}) = \mathrm{Supp}(\boldsymbol{S})$$

and with the support of $\boldsymbol{S}$ the verifier can compute $\boldsymbol{S}$ explicitely from the linear equations $\boldsymbol{H}\boldsymbol{S}^T = T$.

To tackle this problem, we will modify the protocol of Figure 1 by adding an other term to $\boldsymbol{z}$.

3.2 An authentication scheme

We will first describe our scheme as an authentication scheme. It calls upon the notion of product of $\mathbb{F}_q$-linear subspaces of $\mathbb{F}_{q^m}$.

Definition 12. *If E and F are two $\mathbb{F}_q$-linear subspaces of $\mathbb{F}_{q^m}$, their* product *is defined as the* $\mathbb{F}_q$-subspace *consisting of the $\mathbb{F}_q$-linear span of the set of vectors*

$$\{ef, e \in E, f \in F\}$$

where ef stands for the product of e by f in the field $\mathbb{F}_{q^m}$. The product of E with itself will be denoted $E^{\langle 2 \rangle}$, so as not to confuse it with the cartesian product.

The public key consists of a random $(n-k) \times n$ matrix $\boldsymbol{H}$ over $\mathbb{F}_{q^m}$ and two matrices $\boldsymbol{T}$ and $\boldsymbol{T}'$, of size $(n-k) \times lk$ and $(n-k) \times l'k$ respectively, and such that $(\boldsymbol{H}, \boldsymbol{T} | \boldsymbol{T}')$ is an instance of the RSL problem, where $|$ denotes matrix concatenation. The private key consist of two homogeneous matrices $\boldsymbol{S}$ and $\boldsymbol{S}'$ of weight r such that $\boldsymbol{H}\boldsymbol{S}^T = \boldsymbol{T}$ and $\boldsymbol{H}\boldsymbol{S}'^T = \boldsymbol{T}'$. Accordingly, $\boldsymbol{S}$ and $\boldsymbol{S}'$ are $lk \times n$ and $l'k \times n$ matrices respectively. We denote by E the vector space spanned by the coordinates of $\boldsymbol{S}$ and $\boldsymbol{S}'$.

In the commitment step, we sample uniformly at random two vector spaces : $W \in \mathbf{Gr}(w, \mathbb{F}_{q^m})$ and $F \in \mathbf{Gr}(d, \mathbb{F}_{q^m})$. We then randomly choose $\boldsymbol{y} \in (W + EF)^n$. This vector will be used to mask the secret information in answer to the challenge. The commitment consists of $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{y}^T$ together with the subspace F.

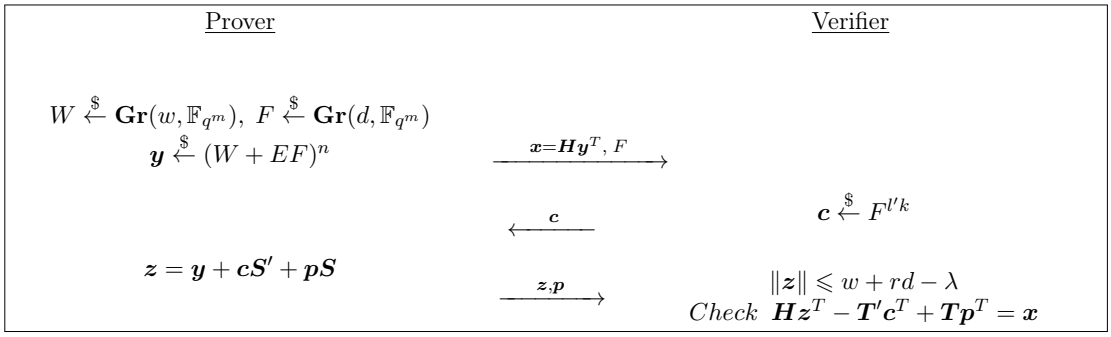

Figure 2: Overview as an authentication scheme

The verifier then chooses a challenge $\boldsymbol{c} \in F^{l'k}$.

To answer the challenge, the prover first computes $\boldsymbol{y} + \boldsymbol{c}\boldsymbol{S}'$. Since the entries of the vector $\boldsymbol{c}$ are in F and the entries of the matrix $\boldsymbol{S}'$ are in E, we have that $\boldsymbol{c}\boldsymbol{S}'$ has its entries in the product space EF, and the vector $\boldsymbol{y} + \boldsymbol{c}\boldsymbol{S}'$ has its entries in the space $W + EF$, like the vector $\boldsymbol{y}$. The linear span of the coordinates of $\boldsymbol{y} + \boldsymbol{c}\boldsymbol{S}'$ is typically equal, or very close to $W + EF$, and this yields too much information on the secret space E to be given to the verifier. To counter this, we add a vector $\boldsymbol{p}\boldsymbol{S}$. Coordinates of $\boldsymbol{p}$ are chosen in F, so that the coordinates of $\boldsymbol{p}\boldsymbol{S}$ fall in the product space EF, and through linear algebra the prover chooses $\boldsymbol{p}$ such that the the linear span of the coordinates of the sum $\boldsymbol{z} = \boldsymbol{y} + \boldsymbol{c}\boldsymbol{S}' + \boldsymbol{p}\boldsymbol{S}$ is restricted to a smaller subspace: namely a subspace $W + U$ for U some subspace of EF of codimension λ inside EF. In other words, $\boldsymbol{z} = \boldsymbol{y} + \boldsymbol{c}\boldsymbol{S}' + \boldsymbol{p}\boldsymbol{S}$ is computed so as to be of rank at most $w + rd - \lambda$. The vector $\boldsymbol{z}$ is then sent to the verifier, together with the vector $\boldsymbol{p}$. This operation is at the heart of the present protocol and the derived signature scheme. More details are given about this in the following section and in Section 4.1.

The verifier accepts if $\|\boldsymbol{z}\| \leqslant rd + w - \lambda$ and $\boldsymbol{H}\boldsymbol{z}^T - \boldsymbol{T}'\boldsymbol{c}^T + \boldsymbol{T}\boldsymbol{p}^T = \boldsymbol{x}$. An overview of this protocol is given in Figure 2.

Using the Fiat-Shamir heuristic, we turn this authentication scheme into a signature scheme.

3.3 Signature scheme

Key generation

- Randomly choose an $(n - k) \times n$ ideal double circulant matrix $\boldsymbol{H}$ as in definition 8 for an irreducible polynomial P, in practice we consider $k = \frac{n}{2}$
- Choose a random subspace E of dimension r of $\mathbb{F}_{q^m}$ and sample l vectors $\boldsymbol{s}_i$ and l' vectors $\boldsymbol{s}'_i$ of length n from the same support E of dimension r
- Set $\boldsymbol{t}_i = \boldsymbol{H}\boldsymbol{s}_i^T$ and $\boldsymbol{t'}_i = \boldsymbol{H}\boldsymbol{s'}_i^T$

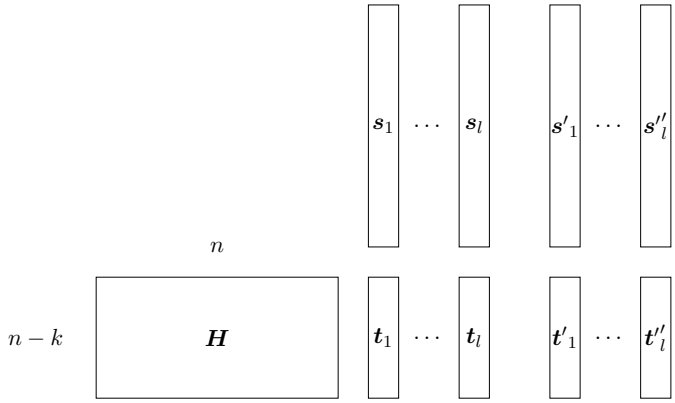

Figure 3: Overview of public and secret key

- Output $(\boldsymbol{H}, \boldsymbol{t}_1, \ldots, \boldsymbol{t}_l, \boldsymbol{t'}_1, \ldots, \boldsymbol{t'}_{l'})$ as public key, and $(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_l, \boldsymbol{s'}_1, \ldots, \boldsymbol{s'}_{l'})$ as secret key

Note that, since $\boldsymbol{H}$ has an ideal structure, each relation of the form $\boldsymbol{H}\boldsymbol{s}_i^T = \boldsymbol{t}_i$ can be shifted $\mod P$ to generate k syndrome relations. We denote $\boldsymbol{S}$ (respectively $\boldsymbol{S'}$) the matrix consisting of all $\boldsymbol{s}_i$ (respectively $\boldsymbol{s'}_i$) and their ideal shifts. Let $\boldsymbol{T} = \boldsymbol{H}\boldsymbol{S}^T$ and $\boldsymbol{T'} = \boldsymbol{H}\boldsymbol{S'}^T$: the public key consists of $(\boldsymbol{H}, \boldsymbol{T}, \boldsymbol{T'})$. $\boldsymbol{T}$ and $\boldsymbol{T'}$ are $\frac{n}{2} \times lk$ and $\frac{n}{2} \times l'k$ matrices respectively, but can be described using only the vectors $(\boldsymbol{t}_1, \ldots, \boldsymbol{t}_l)$ and $(\boldsymbol{t'}_1, \ldots, \boldsymbol{t'}_{l'})$. The secret key consists of the homogeneous matrices $\boldsymbol{S}$ and $\boldsymbol{S'}$ of rank r such that $\boldsymbol{H}\boldsymbol{S}^T = \boldsymbol{T}$ and $\boldsymbol{H}\boldsymbol{S'}^T = \boldsymbol{T'}$.

Figure 3 describes the key pair.

Signature of a message μ

- Randomly choose W, a subspace of $\mathbb{F}_{q^m}$ of dimension w.
- Randomly choose F, a subspace of $\mathbb{F}_{q^m}$ of dimension d.
- Sample $\boldsymbol{y} \in (W + EF)^n$ and set $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{y}^T$.
- For some hash function $\mathcal{H}$, set $\boldsymbol{c} = \mathcal{H}(\boldsymbol{x}, F, \mu)$, $\boldsymbol{c} \in F^{l'k}$. This is done by using the output of $\mathcal{H}$ as the coordinates of $\boldsymbol{c}$ in a basis of F.
- Choose U, a subspace of the product space EF, of dimension $rd - \lambda$, and such that U contains no non-zero elements of the form ef, for $e \in E$ and $f \in F$. More details on this process are given in subsection 3.4.
- Solve $\boldsymbol{z} = \boldsymbol{y} + \boldsymbol{c}\boldsymbol{S'} + \boldsymbol{p}\boldsymbol{S}$ with $\boldsymbol{p} \in F^{lk}$ as unknown, such that $\mathrm{Supp}(\boldsymbol{z}) \subset W + U$: as mentioned in the previous section, $\boldsymbol{p}$ is computed through linear algebra. Specifically, we write $\boldsymbol{p} = (p_1, \ldots, p_{lk})$, and each coordinate $p_i \in F$ of $\boldsymbol{p}$ is decomposed

as

$$p_i = \sum_{\ell=1}^{d} p_{i\ell} f_\ell$$

where $f_1, \ldots, f_d$ is a basis of F that will be used to describe the space F. The j-th coordinate of the vector $\boldsymbol{cS}$ is therefore equal to

$$(\boldsymbol{cS})_j = \sum_{i=1}^{lk} \sum_{\ell=1}^{d} p_{i\ell} f_\ell \boldsymbol{S}_{ij}. \tag{2}$$

Recall that $f_\ell \boldsymbol{S}_{ij}$ is in FE because $\boldsymbol{S}$ has support E. Choose a basis of EF of the form $u_1, \ldots, u_{rd-\lambda}, v_1, \ldots, v_\lambda$, where $u_1, \ldots, u_{rd-\lambda}$ is a basis of U (the typical dimension of EF is rd). Let $\pi_1, \ldots, \pi_\lambda$ be the projections of elements of EF on the last λ coordinates of the above basis. For $h = 1 \ldots \lambda$, applying π_h to all n coordinates of the vector $\boldsymbol{y} + \boldsymbol{cS'} + \boldsymbol{pS}$ and declaring the result to equal 0, we will obtain a linear system of λn equations in the variables p_{ij} by using linearity in (2) to express $\pi_h[(\boldsymbol{cS})_j]$ as

$$\sum_{i=1}^{lk} \sum_{\ell=1}^{d} p_{i\ell} \pi_h(f_\ell \boldsymbol{S}_{ij}). \tag{3}$$

Parameters are chosen so that this system has more variables than equations and typically has a solution. If it doesn't, another space U is sampled.

Figure 4 gives a schematic view of how $\boldsymbol{z} \in W + U$ is obtained.

- Output $(\boldsymbol{z}, F, \boldsymbol{c}, \boldsymbol{p})$ as signature.

The signature consists therefore of the challenge $\boldsymbol{c}$, computed through a hash function, together with the answer to this challenge.

Verification of a signature $(\mu, \boldsymbol{z}, F, \boldsymbol{c}, \boldsymbol{p})$

- Check that $\|zv\| \leqslant rd + w - \lambda$,
- Verify that $\mathcal{H}(\boldsymbol{H}\boldsymbol{z}^T - \boldsymbol{T'}\boldsymbol{c}^T + \boldsymbol{T}\boldsymbol{p}^T, F, \mu) = \boldsymbol{c}$.

To verify the signature, we have to check the rank weight of $\boldsymbol{z}$ and that $\mathcal{H}(\boldsymbol{x}, F, \mu) = \boldsymbol{c}$. The vector $\boldsymbol{x}$ is recomputed using the answer to the challenge. Our signature scheme is summarized on Figure 5.

3.4 Filtering vector spaces

The goal of filtering U during the signature step is to ensure to there is no non-zero element of the form ef in the support of $\boldsymbol{z}$, for $e \in E$ and $f \in F$. This is to prevent an attack that would recover E through techniques for decoding LRPC codes [15]. Indeed, if there is an element of the form ef in $\mathrm{Supp}(\boldsymbol{z})$, then $e \in E \cap f^{-1}\,\mathrm{Supp}(\boldsymbol{z})$ which allows an

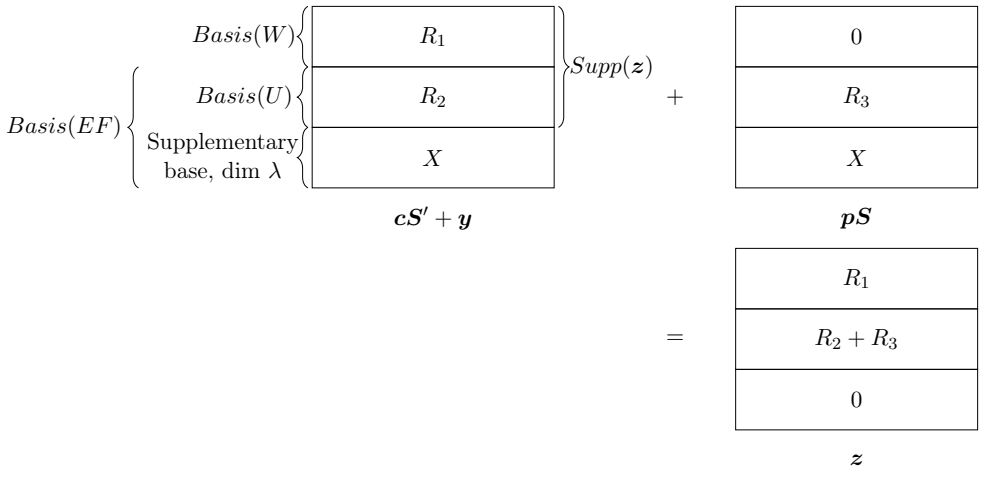

Figure 4: Construction of $\boldsymbol{z}$. R_1, R_2, R_3 are random.

Key generation: $E \xleftarrow{\$} \mathbf{Gr}(r, \mathbb{F}_{q^m})$
Signing key: $\boldsymbol{S} \xleftarrow{\$} E^{n \times lk}$, $\boldsymbol{S'} \xleftarrow{\$} E^{n \times l'k}$
Verification key: $\boldsymbol{H} \xleftarrow{\$}$ ideal $\mathcal{M}^{\frac{n}{2} \times n}$, $\boldsymbol{T} = \boldsymbol{H}\boldsymbol{S}^T, \boldsymbol{T'} = \boldsymbol{H}\boldsymbol{S'}^T$

Sign$(\mu, \boldsymbol{S}, \boldsymbol{S'})$:

1. $W \xleftarrow{\$} \mathbf{Gr}(w, \mathbb{F}_{q^m})$, $F \xleftarrow{\$} \mathbf{Gr}(d, \mathbb{F}_{q^m})$
2. $\boldsymbol{y} \xleftarrow{\$} (W + EF)^n$, $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{y}^T$
3. $\boldsymbol{c} = \mathcal{H}(\boldsymbol{x}, F, \mu)$, $\boldsymbol{c} \in F^{l'k}$
4. $U \xleftarrow{\$}$ filtered subspace of EF of dimension $rd - \lambda$
5. $\boldsymbol{z} = \boldsymbol{y} + \boldsymbol{c}\boldsymbol{S'} + \boldsymbol{p}\boldsymbol{S}$, $\boldsymbol{z} \in W + U$
6. Output $(\boldsymbol{z}, F, \boldsymbol{c}, \boldsymbol{p})$

Verify$(\mu, \boldsymbol{z}, F, \boldsymbol{c}, \boldsymbol{p}, \boldsymbol{H}, \boldsymbol{T}, \boldsymbol{T'})$:

1. Accept if and only if : $\|\boldsymbol{z}\| \leqslant w + rd - \lambda$ and $\mathcal{H}(\boldsymbol{H}\boldsymbol{z}^T - \boldsymbol{T'}\boldsymbol{c}^T + \boldsymbol{T}\boldsymbol{p}^T, F, \mu) = \boldsymbol{c}$

Figure 5: The Durandal Signature scheme

attack against the secret key (moreover elements of this form can be used to distinguish between signatures and randomly generated vectors, as explained in 4.1). To achieve that, we need to find a pair (U, F) such that:

- U is a subspace of EF of dimension $rd - \lambda$
- For every non-zero $x = ef$ with $e \in E$ and $f \in F$, we have that $x \notin U$.

We argue that, for a given F, finding the required space U is quite manageable. We use the following obvious proposition to check the second condition:

Proposition 13. *Let U be a subspace of EF of dimension $rd - \lambda$. Let $E/\mathbb{F}_q$ be a set of representatives of the equivalence relation $x \equiv y \iff \exists \alpha \in \mathbb{F}_q^*$ such that $x = \alpha y$. We have the following equivalence:*

$$\left\{ef : e \in E, f \in F\right\} \cap U = \{0\} \iff \forall e \in E/\mathbb{F}_q, eF \cap U = \{0\}$$

Hence, the cost of this verification is $(q^r - 1)/(q - 1)$ intersections of subspaces of dimension d and $rd - \lambda$, that is to say $\frac{q^r-1}{q-1} \times (d + rd - \lambda)^2 m$ operations in $\mathbb{F}_q$.

We now briefly estimate the probability that a random U contains no element $x = ef$. For simplicity, we only deal with a typical practical case, namely $q = 2$ and $d = r$.

The subspace U is chosen randomly and uniformly of codimension λ inside the vector space EF: we study the probability that U contains no non-zero product ef. Let $x = ef$ be such a non-zero product. Let $\mathcal{U}_x$ be the event $\{x \in U\}$. We are interested in $1 - \mathrm{Prob}\,\mathcal{U}$ where

$$\mathcal{U} = \bigcup_{x=ef,\, x\neq 0} \mathcal{U}_x.$$

Clearly,

$$\mathrm{Prob}\,\mathcal{U}_x = 2^{-\lambda}.$$

Our goal is to evaluate $\mathrm{Prob}\,\mathcal{U}$ through inclusion-exclusion, i.e.

$$\mathrm{Prob}\,\mathcal{U} = \sum_x \mathrm{Prob}\,\mathcal{U}_x - \sum_{x,y} \mathrm{Prob}\,\mathcal{U}_x \cap \mathcal{U}_y + \cdots + (-1)^i \sum_{X\in\Pi, |X|=i} \mathrm{Prob} \bigcap_{x\in X} \mathcal{U}_x + \cdots \quad (4)$$

where Π denotes the set of non-zero elements of EF of the form $x = ef$. We have $|\Pi| = (2^r - 1)^2$. Note that whenever X is made up of linearly independent elements, then the events $\mathcal{U}_x, x \in X$ are independent in the sense of probability, so that

$$\mathrm{Prob} \bigcap_{x\in X} \mathcal{U}_x = 2^{-\lambda|X|}.$$

More generally, since any linear combination of vectors that are in U is also in U, we have

$$\mathrm{Prob} \bigcap_{x\in X} \mathcal{U}_x = 2^{-\lambda \mathrm{rk}(X)}$$

where $\mathrm{rk}(X)$ denotes the rank of X.

For $\lambda = 2r-1$, tedious computations show that the contribution of the non full-rank subsets X for a growing (with r) set of first terms of (4) is negligible, so that we have

$$\mathrm{Prob}\,\mathcal{U} \approx 2 - \frac{4}{2!} + \frac{8}{3!} + \cdots \approx 1 - e^{-2}$$

Giving $1 - \mathrm{Prob}\,\mathcal{U} \approx e^{-2}$.

3.5 Value of λ

In order to find U that contains no element $x = ef$, we need to take the highest value possible for λ. We denote $\boldsymbol{z}_1 = \boldsymbol{y} + \boldsymbol{cS}'$. When $\boldsymbol{z}_1$ is written as a $rd \times n$ matrix over $\mathbb{F}_q$ by rewriting each of its coordinates in a basis of EF of the form $\{u_1, \ldots, u_{rd-\lambda}, v_1, \ldots v_\lambda\}$ such that $U = \{u_1, \ldots, u_{rd-\lambda}\}$, we want $\boldsymbol{pS}$ to be equal to z_1 on the last λ lines, corresponding to $\{v_1, \ldots v_\lambda\}$. This gives λn equations in the base field, and the system has dlk unknowns (the coordinates of $\boldsymbol{p}$ in a basis of F). This gives the following condition on λ :

$$\lambda n < dlk \Leftrightarrow \lambda < \frac{dlk}{n}$$

Since we want to maximize the value of λ, we take $\lambda = \lfloor \frac{dlk}{n} \rfloor$.

3.6 Computational cost

Key generation

The most costly operation of the key generation step is the multiplication of $\boldsymbol{H}$ and the syndromes $\boldsymbol{s}_i$. Each matrix-vector multiplication costs n^2 multiplications in $\mathbb{F}_{q^m}$, hence a total cost of $(l + l')n^2$ multiplications.

Signature of a message μ

The signature step splits naturally in two phases: an offline phase during which the signature support is computed (this is the most costly part) and an online phase to compute the actual signature. The two phases are as follows:

1. Offline phase
   - Choose the vector spaces W and F.
   - Sample $\boldsymbol{y} \in (W + EF)^n$ and set $\boldsymbol{x} = \boldsymbol{H}\boldsymbol{y}^T$.
   - Choose U, a random subspace of EF of dimension $rd - \lambda$. If U contains non-zero elements of the form ef, $e \in E$ and $f \in F$, choose another U.
   - Write the $\mathbb{F}_{q^m}$-coordinates of the vector $\boldsymbol{pS}$ in a basis of EF of the form $\{u_1, \ldots, u_{rd-\lambda}, v_1, \ldots, v_\lambda\}$ where $U = \langle u_1, \ldots, u_{rd-\lambda} \rangle$ to obtain linear expressions in the variables p_{ij} of the form (3). Compute a $\lambda n \times \lambda n$ matrix $\boldsymbol{D}$ that

inverts this linear mapping of the p_{ij}. This will allow to compute $\boldsymbol{p}$ such that $\boldsymbol{z} \in U$ in the online phase with a matrix multiplication instead of an inversion. If the linear map cannot be inverted to produce the matrix $\boldsymbol{D}$, choose another random subspace U of EF.

2. Online phase
   - Set $\boldsymbol{c} = \mathcal{H}(\boldsymbol{x}, F, \mu)$, $\boldsymbol{c} \in F^{l'k}$
   - Solve $\boldsymbol{z} = \boldsymbol{y} + \boldsymbol{c}\boldsymbol{S}' + \boldsymbol{p}\boldsymbol{S}$ with $\boldsymbol{p} \in F^{lk}$, using the matrix $\boldsymbol{D}$ computed during the offline phase
   - Output $(\boldsymbol{z}, F, \boldsymbol{c}, \boldsymbol{p})$ as signature.

The most costly step in the offline phase is the computation of the matrix $\boldsymbol{D}$, which requires inverting a linear system over $\mathbb{F}_q$ with λn equations, hence the cost is $(\lambda n)^3$ multiplications in $\mathbb{F}_q$.

The online phase consists in the computation of $\boldsymbol{p}$ which costs $(\lambda n)^2$ multiplications in $\mathbb{F}_q$ as well as the computation of $\boldsymbol{z} = \boldsymbol{y} + \boldsymbol{c}\boldsymbol{S}' + \boldsymbol{p}\boldsymbol{S}$ which costs $(l'k)^2 + (lk)^2$ multiplications in $\mathbb{F}_{q^m}$ for computing the matrix/vector products.

Verification of a signature

The most costly step during the verification phase is the computation of $\boldsymbol{H}\boldsymbol{z}^T - \boldsymbol{T}'\boldsymbol{c}^T + \boldsymbol{T}\boldsymbol{p}^T$, which costs $n^2 + (l'k)^2 + (lk)^2$ multiplications in $\mathbb{F}_{q^m}$.

4 Security of the scheme

4.1 Product Spaces Subspaces Indistinguishability (PSSI)

The PSSI problem is a new problem which appears naturally when we try to prove the indistinguishability of the signatures.

Problem 4. *Product Spaces Subspaces Indistinguishability. Let E be a fixed $\mathbb{F}_q$-subspace of $\mathbb{F}_{q^m}$ of dimension r. Let F_i, U_i and W_i be subspaces defined as follow:*

- $F_i \xleftarrow{\$} \mathbf{Gr}(d, \mathbb{F}_{q^m})$
- $U_i \xleftarrow{\$} \mathbf{Gr}(rd - \lambda, EF_i)$ *such that* $\{ef, e \in E, f \in F\} \cap U_i = \{0\}$
- $W_i \xleftarrow{\$} \mathbf{Gr}(w, \mathbb{F}_{q^m})$

The $\mathrm{PSSI}_{r,d,\lambda,w,\mathbb{F}_{q^m}}$ *problem consists in distinguishing samples of the form $(\boldsymbol{z}_i, F_i)$ where $\boldsymbol{z}_i$ is a random vector of $\mathbb{F}_{q^m}^n$ of support $W_i + U_i$ from samples of the form $(\boldsymbol{z}'_i, F_i)$ where $\boldsymbol{z}'_i$ is a random vector of $\mathbb{F}_{q^m}^n$ of weight $w + rd - \lambda$.*

In order to study the complexity of this problem, we first reduce it to the case where the samples are of the form (Z_i, F_i) with $Z_i = \mathrm{Supp}(\boldsymbol{z}_i)$. Let us suppose we have a distinguisher D for this last case. Then given N samples of the PSSI problem, it is easy to compute the supports Z_i of the vectors $\boldsymbol{z}_i$ and to use D to distinguish if Z_i is a random subspace of dimension $w + rd - \lambda$ or if it is of the form $W_i + U_i$ with U_i a subspace of the product space EF_i.

Conversely, let us suppose we have a distinguisher D' for the PSSI problem. We are given N samples of the form (Z_i, F_i). For each sample, we can compute a random vector $\boldsymbol{z}_i$ of support Z_i and use D' to distinguish whether $\boldsymbol{z}_i$ is a random vector of weight $w + rd - \lambda$ or if its support is of the form $W_i + U_i$.

Thus we can consider the case when the samples are only couples of subspaces of $\mathbb{F}_{q^m}$.

This problem is related to the decoding of LRPC codes [15]. Indeed we can consider a subspace $Z = U + W$ as the noisy support of a syndrome for an LRPC code, the noise corresponding to W. Consequently, it is natural to try and apply techniques used for decoding LRPC codes in order to solve the PSSI problem. The first idea is to use the basic decoding algorithm (see [15]). It consists in computing intersections I of the form $f^{-1}Z \cap f'^{-1}Z$ with $(f, f') \in F^2$. If Z is of the form $U + W$ then the probability that $\dim I \neq 0$ is much higher than if Z were truly random. However, this technique cannot be used because we filter the subspace U.

The decoding algorithm for LRPC codes has been improved in [6]. The idea is to consider product spaces of the form ZF_i where F_i is a subspace of F of dimension 2. The probability that $\dim ZF_i = 2(w + rd - \lambda)$ depends on whether Z is random or not. We study in detail the advantage of this distinguisher in the following paragraphs.

Consider the product subspace EF inside $\mathbb{F}_{q^m}$ with $\dim F = \dim E = r$. Suppose E is unknown, the typical dimension of the product EF is then r^2, if we assume $r^2 \ll m$. We now suppose we are given a subspace Q of $\mathbb{F}_{q^m}$ of dimension r^2 that is either a product space EF or a randomly chosen one, and we wish to distinguish between the two events. One easy way to do so, if the dimension m of the ambient space $\mathbb{F}_{q^m}$ is large enough, is to multiply Q by F. If Q is random, we will get the typical product dimension $\dim FQ = r \dim Q = r^3$. While if $Q = EF$, we will get $\dim FQ \leq \binom{r+1}{2} r < r^3$. In fact, to distinguish the two cases it is enough to multiply Q by any subspace A of F of dimension 2, since we will have $\dim AQ \leq 2r^2 - r$ when $Q = EF$ and $\dim AQ = 2r^2$ in the typical random Q case.

To make our two cases difficult to distinguish, our query space Q is actually chosen to be constructed in one of two ways, making up a *distinguishing problem*:

Distinguishing problem. Distinguish whether Q is of the form (i) or (ii) below:

(i) $Q = U + W$ where U is a subspace of EF of codimension λ. The space E is chosen randomly of dimension r as before. The subspace U is chosen in such a way so that, for any subspace A of F dimension 2, we have $\dim AU = 2 \dim U$. The space W is chosen randomly of dimension w, so that $\dim Q = r^2 - \lambda + w$.

(ii) Q is a random subspace of dimension $r^2 - \lambda + w$. Equivalently we may think of Q of the form $Q = V + W$ where both V and W are random (and independent) of dimensions $r^2 - \lambda$ and w respectively.

The purpose of choosing such a subspace U of EF is to make the dimension of AU equal to that of AV for a random V. Adding the random space W to U should keep the probability distributions of $\dim AQ$ equal for both ways of construction Q. The purpose of W is to make the dimension of Q sufficiently large with respect to the dimension m of the ambient space, so that multiplying Q by a space of dimension more than 2 will typically fill up the whole space $\mathbb{F}_{q^m}$ anyway. In this manner, the two ways of constructing Q will be indistinguishable by measuring dimensions of the product of Q by a subspace.

First, we give a criterion for a subspace U of EF to have the property that $\dim AU = 2\dim U$ for any subspace A of dimension 2 of F.

Let E, F be two subspaces of $\mathbb{F}_{q^m}$, both of dimension r over $\mathbb{F}_q$. Let us make the remark that the maximum possible dimension of $F^{\langle 2\rangle}$ is $\binom{r+1}{2}$, and the maximum possible dimension of $F^{\langle 2\rangle}E$ is therefore $r\binom{r+1}{2}$.

Let $f_1, f_2, \ldots, f_r$ be a basis of F. Denote by F_2 the subspace of F generated by f_1, f_2, by F_3 the subspace of F generated by f_1, f_2, f_3, and so on.

Lemma 14. *Suppose* $\dim F^{\langle 2\rangle}E = r\binom{r+1}{2}$. *Then* $f_1FE \cap f_2FE = f_1f_2E$.

Proof. We have $F_2FE = f_1FE + f_2FE$, and $f_1FE \cap f_2FE \supset f_1f_2E$, therefore

$$\dim F_2FE \geq 2\dim EF - \dim F \tag{5}$$

by using the formula $\dim(A + B) = \dim A + \dim B - \dim A \cap B$. Similarly, $F_{i+1}FE = F_iFE + f_{i+1}FE$ and $F_iFE \cap f_{i+1}FE \supset f_jf_{i+1}E$ for all $j = 1, 2, \ldots i$. From which we have

$$\dim F_{i+1}FE \geq \dim F_iFE + \dim FE - i\dim E.$$

Now $\dim F^{\langle 2\rangle}E = r\binom{r+1}{2}$ only occurs when we have equality in all the above inequalities, in particular we have equality in (5) which implies that the inclusion $f_1f_2E \subset f_1FE \cap f_2FE$ is also an equality. □

Lemma 15. *Let U be a subspace of EF. Under the hypothesis* $\dim F^{\langle 2\rangle}E = r\binom{r+1}{2}$, *we have that there exists a subspace $A \subset F$ of dimension 2 such that,*

$$\dim AU < 2\dim U$$

if and only if U contains two non-zero elements of the form fe and $f'e$ $f, f' \in F$, $e \in E$ where f and f' are two linearly independent elements of F.

Proof. Let A be a subspace of F of dimension 2 generated by f_1, f_2. We have $AS = f_1U + f_2U$ so that $\dim AU < 2\dim U$ if and only if $f_1U \cap f_2U \neq \{0\}$. But we have

$$f_1U \cap f_2U \subset f_1FE \cap f_2FE$$

and under the hypothesis $\dim F^{\langle 2\rangle}E = r\binom{r+1}{2}$ we have that $\dim AFE = 2r^2 - r$ and $f_1FE \cap f_2FE = f_1f_2E$. Therefore $f_1U \cap f_2U$ contains a non-zero element if and only if U contains an element of the form f_2e and an element of the form f_1e, for $e \in E$, $e \neq 0$. □

Corollary 16. *Suppose* $\dim F^{\langle 2\rangle}E = r\binom{r+1}{2}$*, and that* U *is a subspace of* FE *such that for any two non-zero elements* $f \in F$ *and* $e \in E$, $ef \notin U$*. Then we have, for any subspace* $A \subset F$ *of dimension* 2*,*

$$\dim AU = 2 \dim U.$$

Next, we study the probability distribution of the dimension of the product space $A(U + W)$, where W is random of dimension w, and U is either constructed as above or uniform random. We only focus on the binary extension field case $q = 2$, and from the previous discussion we only keep the property that $\dim AU$ is maximal. In other words, for the purpose of the following analysis, U is a fixed subspace of $\mathbb{F}_{2^m}$ with $\dim U = u$, A is a fixed subspace of $\mathbb{F}_{2^m}$ of dimension $\dim A = 2$ and we suppose that we have $\dim AU = 2u$. Let W be a *random* subspace of dimension $\dim W = w$ of $\mathbb{F}_{2^m}$. The space W is chosen by choosing $x_1, x_2, \ldots, x_w$ random independent (in the sense of probability) elements of $\mathbb{F}_{2^m}$ and W is taken to be the subspace generated by the x_i. Strictly speaking, $x_1, \ldots, x_w$ may turn out not to be linearly independent and not generate a space of dimension w. However, w will be taken to be much smaller than m, so that this event happens with negligible probability.

Our goal is to study the probability that $A(U+W)$ does not have dimension $2(u+w)$ and see how it may vary for two different spaces U_1 and U_2.

Consider the mapping

$$\begin{array}{rcl} A^w & \xrightarrow{\Phi} & \mathbb{F}_{2^m} \\ (a_1, a_2, \ldots, a_w) & \mapsto & a_1x_1 + a_2x_2 + \cdots + a_wx_w \end{array}$$

The product space $A(U + W)$ does not have maximal dimension, namely $2(u + w)$, if and only if there is a non-zero

$$\mathbf{a} = (a_1, a_2, \ldots, a_w)$$

in A^w such that $\Phi(\mathbf{a}) \in AU$. This event $\mathcal{E}$, over all choices of $\mathbf{x} = (x_1, \ldots, x_w)$, can therefore be written as:

$$\mathcal{E} = \bigcup_{\substack{\mathbf{a} \in A^w \\ \mathbf{a} \neq 0}} \mathcal{E}_{\mathbf{a}}$$

where $\mathcal{E}_{\mathbf{a}}$ denotes the event $\Phi(\mathbf{a}) \in AU$. Since

$$\operatorname{Prob} \mathcal{E}_{\mathbf{a}} = \frac{4^u}{2^m}$$

the union bound gives us

$$\operatorname{Prob} \mathcal{E} \leq (4^w - 1)\frac{4^u}{2^m}. \tag{6}$$

We now study the lower bound

$$\operatorname{Prob}\mathcal{E} \geq \sum_{\substack{\mathbf{a}\in A^w \\ \mathbf{a}\neq 0}} \operatorname{Prob}\mathcal{E}_{\mathbf{a}} - \sum_{\mathbf{a},\mathbf{b}} \operatorname{Prob}\mathcal{E}_{\mathbf{a}} \cap \mathcal{E}_{\mathbf{b}} \tag{7}$$

where the second sum runs over all unordered pairs of distinct w-tuples $\mathbf{a}$ and $\mathbf{b}$. To evaluate this second sum we split the pairs $\mathbf{a}, \mathbf{b}$ into two disjoint sets:

1. linearly independent pairs $\mathbf{a}, \mathbf{b}$. In which case the two random variables $\mathbf{ax}$ and $\mathbf{bx}$ are independent, and we have

$$\operatorname{Prob}\mathcal{E}_{\mathbf{a}} \cap \mathcal{E}_{\mathbf{b}} = \operatorname{Prob}\mathcal{E}_{\mathbf{a}} \operatorname{Prob}\mathcal{E}_{\mathbf{b}} = \left(\frac{4^u}{2^m}\right)^2.$$

2. linearly dependent pairs $\mathbf{a}, \lambda\mathbf{a}$, for some $\lambda \in \mathbb{F}_{2^m}$, $\lambda \neq 1$, such that $\lambda\mathbf{a} \in A^w$. In this case, we have

$$\operatorname{Prob}\mathcal{E}_{\mathbf{a}} \cap \mathcal{E}_{\mathbf{b}} = |AU \cap \lambda AU| \frac{1}{2^m} \leq \frac{4^u}{2^m}.$$

   We now estimate the number of such pairs $\mathbf{a}, \lambda\mathbf{a}$.

   Denote the non-zero elements of A by $a_1, a_2, a_3 = a_1 + a_2$ (recall that A is a vector space). Suppose we have $\lambda a_1 = a_2$ and $\lambda a_2 = a_3 = a_1 + a_2$ ($\lambda a_2 = a_2$ would imply $\lambda = 1$ and $\lambda a_2 = a_1$ would imply $\lambda^2 = 1$ hence $\lambda = 1$ in a field of characteristic 2). Then $a_2 a_1^{-1} = \lambda$ satisfies $\lambda^2 + \lambda + 1$ which is not possible for odd m and happens with negligible probability for even m. Assuming this does not happen we have therefore that any $\mathbf{a} \in A^w$ such that $\lambda\mathbf{a} \in A^w$ must have all non-zero coefficients equal. Hence the number of such pairs $\mathbf{a}, \lambda\mathbf{a}$ is at most 3.2^w. Inequality (7) gives us therefore:

$$\operatorname{Prob}\mathcal{E} \geq (4^w - 1)\frac{4^u}{2^m} - \binom{4^w - 1}{2}\frac{4^{2u}}{2^{2m}} - 3.2^w \frac{4^u}{2^m}. \tag{8}$$

From which we get:

Proposition 17. *If U and V are two spaces of dimension u such that $\dim AU = \dim AV = 2u$ then*

$$\left|\operatorname{Prob}\left[\dim A(U+W) < 2(u+w)\right] - \operatorname{Prob}\left[\dim A(V+W) < 2(u+w)\right]\right| \leq \binom{4^w - 1}{2}\frac{4^{2u}}{2^{2m}} + 3.2^w \frac{4^u}{2^m}.$$

Product space distinguisher. We go back to our distinguishing problem defined above. As mentioned in the discussion leading up to the problem, it is natural to try and distinguish between (i) and (ii) by computing the dimension of some AQ for many instances of Q and basing the decision on the number of times an abnormal (less than

$2 \dim Q$) turns up. The consequence of Proposition 17 is that to distinguish confidently with this method requires a very large number of queries. Specifically, if the two probabilities of producing an abnormal dimension are p and $p(1+\varepsilon)$, then the number of products AQ that one must produce is of the order $1/p\varepsilon^2$. Proposition 17 gives $p \approx 2^{2u+2w-m}$ and $\varepsilon = 2^{\log_2 3 - w}$.

Proposition 18. *By applying proposition 17 to the* $\mathrm{PSSI}_{r,d,\lambda,w,\mathbb{F}_{q^m}}$ *problem, the advantage with which one may distinguish the two distributions is of the order of* $2^{m-2(rd-\lambda)}$.

Remark: One might also consider computing product spaces of the form ZE' where E' is a subspace of E of dimension larger than 2. However, we have chosen our parameters such that $3(w + rd - \lambda) > m$ so this idea cannot work.

New Problem: Advanced Product Subspaces Indistinguishability (PSSI^+) The PSSI^+ problem is a generalization of the previous problem, with some extra side information. We need to consider this problem for our security proof.

Problem 5. *Advanced Product Spaces Subspaces Indistinguishability. Let* E *be a fixed* $\mathbb{F}_q$*-subspace of* $\mathbb{F}_{q^m}$ *of dimension* r*. Let* F_i, U_i *and* W_i *be subspaces defined as before:*

- $F_i \xleftarrow{\$} \mathbf{Gr}(d, \mathbb{F}_{q^m})$
- $U_i \xleftarrow{\$} \mathbf{Gr}(rd - \lambda, EF_i)$ *such that* $\{ef, e \in E, f \in F\} \cap U_i = \{0\}$
- $W_i \xleftarrow{\$} \mathbf{Gr}(w, \mathbb{F}_{q^m})$

Let $\boldsymbol{H}$ *be a Randomly chosen* $(n-k) \times n$ *ideal double circulant matrix as in definition 8 for an irreducible polynomial* P*.*

- *Sample* l *vectors* $\boldsymbol{s}_i$ *and* l' *vectors* $\boldsymbol{s}'_i$ *of length* n *from the same support* E *of dimension* r
- *Set* $\boldsymbol{S}$ *(respectively* $\boldsymbol{S}'$*) the matrix consisting of all* $\boldsymbol{s}_i$ *(respectively* $\boldsymbol{s}'_i$*) and their ideal shifts. Let* $\boldsymbol{T} = \boldsymbol{H}\boldsymbol{S}^T$ *and* $\boldsymbol{T}' = \boldsymbol{H}\boldsymbol{S}'^T$*.*

The $\mathrm{PSSI}^+(N)_{r,d,\lambda,w,\mathbb{F}_{q^m}}$ *problem consists in distinguishing* N *samples of the form* $(\boldsymbol{z}_i, F_i)$ *where* $\boldsymbol{z}_i$ *is a random vector of* $\mathbb{F}_{q^m}^n$ *of support* $W_i + U_i$ *from samples of the form* $(\boldsymbol{z}'_i, F_i)$ *where* $\boldsymbol{z}'_i$ *is a random vector of* $\mathbb{F}_{q^m}^n$ *of weight* $w + rd - \lambda$ *when additionally given* $\boldsymbol{H}, \boldsymbol{T}, \boldsymbol{T}'$*.*

The PSSI^+ consists of an instance of the PSSI problem and an instance of the RSL problem that share the same secret support E. The question is to determine whether or not the instance of RSL can be used in order to reduce the difficulty of PSSI.

In general, combining two difficult problem together does not necessarily lead to another hard problem. For example, combining two difficult instances of the factorization

of large integers n, n' with $n = pq$ and $n' = pq'$ where p, q and q' are prime is a an easy problem.

In our case, the knowledge of an instance of RSL could be useful if it gives us some information on the support E. But, for our parameters, the best attacks on the RSL problem are based on the GRS$^+$ algorithm [5, 13, 10] and this algorithm recovers the whole support or nothing. Moreover, the main idea behind the GRS$^+$ algorithm (which consists of looking for a subspace E' which contains E) cannot be applied to the PSSI$^+$ problem since E is "multiplied" by an F_i at each sample. Thus it appears that the knowledge of an instance of RSL that shares the same secret support E does not help to solve the PSSI problem and we will consider that the PSSI$^+$ problem is as hard to attack as the PSSI problem.

4.2 Security model

One of the security models for signature schemes is existential unforgeability under an adaptive chosen message attack (EUF-CMA). Basically, it means that even if an adversary has access to a signature oracle, it cannot produce a valid signature for a new message with a non negligible probability.

Existential Unforgeability under Chosen Message Attacks [22] ($\mathsf{EUF-CMA}$). Even after querying N valid signatures on chosen messages (μ_i), an adversary should not be able to output a valid signature on a fresh message μ. To formalize this notion, we define a signing oracle OSign:

- $\mathsf{OSign}(\mathsf{vk}, \mu)$: This oracle outputs a signature on μ valid under the verification key vk. The requested message is added to the signed messages set $\mathcal{SM}$.

$\mathbf{Exp}^{\mathsf{euf}}_{\mathcal{S},\mathcal{A}}(\lambda)$
1. $\mathsf{param} \leftarrow \mathsf{Setup}(1^\lambda)$
2. $(\mathsf{vk}, \mathsf{sk}) \leftarrow \mathsf{KeyGen}(\mathsf{param})$
3. $(\mu^*, \sigma^*) \leftarrow \mathcal{A}(\mathsf{vk}, \mathsf{OSign}(\mathsf{vk}, \cdot))$
4. $b \leftarrow \mathsf{Verify}(\mathsf{vk}, \mu^*, \sigma^*)$
5. `IF` $\mu^* \in \mathcal{SM}$ `RETURN` 0
6. `ELSE RETURN` b

The probability of success against this game is denoted by

$$\mathsf{Succ}^{\mathsf{euf}}_{\mathcal{S},\mathcal{A}}(\lambda) = \Pr[\mathbf{Exp}^{\mathsf{euf}}_{\mathcal{S},\mathcal{A}}(\lambda) = 1], \quad \mathsf{Succ}^{\mathsf{euf}}_{\mathcal{S}}(\lambda, t) = \max_{\mathcal{A} \leq t} \mathsf{Succ}^{\mathsf{euf}}_{\mathcal{S},\mathcal{A}}(\lambda).$$

4.3 $\mathsf{EUF-CMA}$ proof

To prove the $\mathsf{EUF-CMA}$ security of our scheme, we proceed in two steps. In the first step, we show that an adversary with access to N valid signatures has a negligible advantage on the same adversary with only access to the public keys. In other words, we prove that signatures do not leak information of the secret keys. In the second step, we show that if we only have access to the public keys, a valid signature allows us to solve an instance of the I − ARSD problem.

A Technical Lemma

Lemma 19. *Let $\mathcal{F}$ be a family of functions defined by*

$$\mathcal{F} = \left\{ \begin{array}{rccc} f_{\boldsymbol{H}} : & (W+EF)^n & \rightarrow & \mathbb{F}_{q^m}^{n-k} \\ & \boldsymbol{y} & \mapsto & \boldsymbol{x} = \boldsymbol{y}\boldsymbol{H}^T \end{array} \right\}$$

Since $\boldsymbol{H}$ is chosen uniformly at random amongst the $(n-k)\times n$ ideal double circulant matrices, $\mathcal{F}$ is a pairwise independent family of function.
The number of choices for $\boldsymbol{y}$ depends on W and F and on the choice of the coordinates of $\boldsymbol{y}$. Overall, the entropy of $\boldsymbol{y}$ is equal to

$$\Theta\left(\begin{bmatrix} m \\ w \end{bmatrix}_q \begin{bmatrix} m \\ d \end{bmatrix}_q q^{(w+rd)n}\right) = 2^{(w(m-w)+d(m-d)+(w+rd)n)\log q + O(1)}$$

Since $\|\boldsymbol{y}\| > d_{RGV}$, any vector of $\mathbb{F}_{q^m}^{n-k}$ can be reached, thus the entropy of $\boldsymbol{x}$ is equal to $2^{(n-k)m\log q}$. According to the Leftover Hash Lemma [23], we have

$$\Delta(\mathcal{D}_{\mathbf{G_0}}, \mathcal{U}) < \frac{\varepsilon}{2}$$

where $\Delta(X,Y)$ denotes the statistical distance between X and Y, $\mathcal{D}_{\mathbf{G_0}}$ denotes the distribution of $\boldsymbol{x}$ in game $\mathbf{G_0}$, $\mathcal{U}$ denotes the uniform distribution over $\mathbb{F}_{q^m}^{n-k}$ (so the distribution of $\boldsymbol{x}'$ in game $\mathbf{G_1}$ and

$$\varepsilon = 2^{\frac{((n-k)m - w(m-w)+d(m-d)+(w+rd)n)\log q}{2} + O(1)}$$ □

Proofs For the first step, we proceed in a sequence of games. We denote $\mathbb{P}_{\mathbf{G_i}}$ the probability that the adversary returns 1 in the end of the game $\mathbf{G_i}$ and $\mathsf{Adv}(\mathbf{G_i}) = |\mathbb{P}_{\mathbf{G_i}} - \frac{1}{2}|$ the advantage of the adversary for the game $\mathbf{G_i}$.

- $\mathbf{G_0}$: this is the real $\mathsf{EUF-CMA}$ game for $\mathcal{S}$. The adversary has access to the signature oracle OSign to obtain valid signatures.

  $$\mathbb{P}_{\mathbf{G_0}} = \mathsf{Succ}^{\mathsf{euf}}_{\mathcal{S},\mathcal{A}}(\lambda)$$

- $\mathbf{G_1}$: we replace $\boldsymbol{z}$ by a vector $\boldsymbol{z}'$ of the same weight chosen uniformly at random in the correct subspace U of $W+EF$, and sample $\boldsymbol{c}', \boldsymbol{p}'$ uniformly with support F.

  Now set $\boldsymbol{x}' = \boldsymbol{H}\boldsymbol{z}' - \boldsymbol{c}'\boldsymbol{T}' - \boldsymbol{p}'\boldsymbol{T}$ and use the Random Oracle to set $\boldsymbol{c} = \mathcal{H}(\boldsymbol{x}', F, \mu)$.

  In $\mathbf{G_0}$, $\boldsymbol{x}$ is the syndrome of the vector $\boldsymbol{y}$ of support of the form $EF+W$, while here $\boldsymbol{x}'$ is not necessarily. Under Lemma 19 we conclude.

  $$\mathsf{Adv}(\mathbf{G_1}) \leq \mathsf{Adv}(\mathbf{G_0}) + \varepsilon$$

  The parameters of the signature are chosen such that ε is lower than the security parameter.

- $\mathbf{G_2}$: We now sample $\boldsymbol{z}$ at random in $\mathbb{F}_{q^m}^n$ with the same weight, and proceeds as in $\mathbf{G_2}$.

  This corresponds to an instance of the PSSI$^+(N)$ problem 5. Since the adversary can have access to at most N signatures, we have

$$|\mathsf{Adv}(\mathbf{G_2}) - \mathsf{Adv}(\mathbf{G_1})| \leqslant \mathsf{Adv}(\text{PSSI}^+(N))$$

- $\mathbf{G_3}$: We now pick $\boldsymbol{T}, \boldsymbol{T}'$ at random and proceed as before. The difference between $\mathbf{G_3}$ and $\mathbf{G_2}$ resides in the public key, on whether it was sampled using vectors in a given subspace or not.

$$|\mathsf{Adv}(\mathbf{G_3}) - \mathsf{Adv}(\mathbf{G_2})| \leqslant \mathsf{Adv}(\text{DRSL})$$

At this step, everything we send to the adversary is random, and independent from any secret keys. Hence the security of our scheme is reduced to the case where no signature is given to the attacker.

If he can compute a valid signature after the game $\mathbf{G_3}$, then the challenger can compute a solution of the $\mathrm{I}-\mathrm{ARSD}$ problem. Indeed, the couple $(\boldsymbol{z}, \boldsymbol{p})$ is a solution of the instance $(\boldsymbol{H}, -\boldsymbol{T}, F, \boldsymbol{x} + \boldsymbol{T}'\boldsymbol{c}^T, \omega + rd - \lambda)$ of the $\mathrm{I}-\mathrm{ARSD}$ problem. According to Proposition 9, the $\mathrm{I}-\mathrm{ARSD}$ problem is reduced to the $\mathrm{I}-\mathrm{RSD}$ problem.

Finally, we can give the main theorem of our article:

Theorem 20 (EUF-CMA security)**.** *Under the hypothesis of the hardness of the* PSSI$^+$ *problem 4.1 and of the* $\mathrm{DRSL}, \mathrm{I}-\mathrm{RSD}$ *problems 1, our signature scheme is secure under the EUF-CMA model in the Random Oracle Model.*

5 Attacks

5.1 Attacks on the RSL problem

In this section we will study the hardness of recovering the secret matrices $\boldsymbol{S}$ and $\boldsymbol{S}'$ from $\boldsymbol{H}, \boldsymbol{T}, \boldsymbol{T}'$. This is exactly an instance of the $\mathrm{RSL}_{q,m,n,k,w,N}$ problem.

We will use the setting proposed in [13]. First, we recall how the problem is reduced to searching for a codeword of weight w in a code containing q^N codewords of this form. We introduce the following $\mathbb{F}_q$-linear code :

$$C = \{\boldsymbol{x} \in \mathbb{F}_{q^m}^n : \boldsymbol{A}\boldsymbol{x} \in W_T\}$$

Where W_T is the $\mathbb{F}_q$-linear subspace of $\mathbb{F}_{q^m}^{n-k}$ generated by the linear combinations of the elements of the public matrices $\boldsymbol{T}$ and $\boldsymbol{T}'$. As in the lemma 1 in [13], we define :

$$C' = \{\sum_i \alpha_i \boldsymbol{s}_i, \alpha_i \in \mathbb{F}_q\}$$

We have :

- $dim_{\mathbb{F}_q} C \leqslant km + N$
- $C' \subset C$
- the elements of C' are of weight $\leqslant w$.

Combinatorial attack

In [13], the authors search for codewords of rank w in C by using information-set decoding techniques, using the fact that C' contains q^N words of weight w. As this codeword will very likely be a linear combinations of the vectors $\boldsymbol{s}_i$, it will reveal the secret support E with high probability. Theorem 2 in [13] gives a complexity of $q^{min(e_-,e_+)}$, where :

$$e_- = \left(w - \left\lfloor \frac{N}{n} \right\rfloor\right)\left(\left\lfloor \frac{K}{n} \right\rfloor - \left\lfloor \frac{N}{n} \right\rfloor\right)$$
$$e_+ = \left(w - \left\lfloor \frac{N}{n} \right\rfloor - 1\right)\left(\left\lfloor \frac{K}{n} \right\rfloor - \left\lfloor \frac{N}{n} \right\rfloor - 1\right) + n\left(\left\lfloor \frac{K}{n} \right\rfloor - \left\lfloor \frac{N}{n} \right\rfloor - 1\right)$$

Where $K = km + N$. See [13] for more details about this.

Algebraic attack

We will now study how algebraic attacks can be used to find codewords of weight w in C.

We are looking for $X \in C$ such that $X \in E^n$. We can write X as :

$$X = \begin{matrix} \sum_{i=1}^{w} x_1^{(i)} y_1^{(i)} & \dots & \sum_{i=1}^{w} x_1^{(i)} y_n^{(i)} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{w} x_m^{(i)} y_1^{(i)} & \dots & \sum_{i=1}^{w} x_m^{(i)} y_n^{(i)} \end{matrix}$$

Where $(x^{(1)}, \dots, x^{(w)})$ represent a basis of E, and the $(y_i^j), 1 \leqslant i \leqslant w, 1 \leqslant j \leqslant n$ represent the coordinates of X written in this basis.

C has length nm and dimension $N + km$ in $\mathbb{F}_q$, which gives $(n-k)m - N$ parity check equations, and $(n+m)w$ unknowns (the $x_i^{(j)}$ and the $(y_i^{(j)})$).

To decrease the number of unknowns, we will first write the basis of E in an echeloned form, which removes w^2 unknowns :

$$\forall (i,j) \in [1,w]^2, i \neq j, x_i^{(j)} = 0$$

$$x_i^{(i)} = 1$$

Then we will use the fact that for a fixed basis of E, the solution space has dimension N, which allows us to set N of the $(y_i^{(j)})$ to specialize one solution, as in [9] : for a random subset $I \subseteq [1,n] \times [1,w]$ of size $N-1$:

$$\forall (i,j) \in I, y_i^{(j)} = 0$$

$$y_{i_0}^{(j_0)} = 1, (i_0, j_0) \notin I$$

Which removes N unknowns.

Proposition 21. *Using this setting we obtain :*

- $(n-k)m - N$ *equations*
- $(n+m)w - w^2 - N$ *unknowns.*

We implemented this approach in Magma to try it on small examples, and the combinatorial attacks become much more efficient than the algebraic approach when the number of samples is around kr, while this attack is faster when the number of samples is higher. Another drawback of this attack is the high memory cost, making parameters as small as $n = m = 30, k = 15, r = 3$ with kr samples too big for a computer using $16GB$ of RAM.

For concrete parameters (section 6), we chose N, the number of samples for the RSL problem, equal to either $k(r-1)$ or $k(r-2)$. Our experiments on smaller parameters showed that combinatorial attacks should be way faster for this number of samples. This also defeats the setting proposed in [9] since it needs at least kr samples.

The parameter set I gives 2117 unknowns for 23836 equations and the parameter set II gives 2809 unknowns for 29154 equations. Based on our experiments on smaller parameters this seems really hard to reach.

5.2 Attack on the ARSD problem

As explained in the security proof in Section 4.3, a forgery attack consists in solving an instance of the ARSD problem 3. In order to choose the parameters of our signature, we need to deal with the complexity of the attacks on this problem. Since it is very similar to the RSD problem, these attacks are adaptations of the attacks against the RSD problem [16, 5].

The following proposition gives a bound from which the problem becomes polynomial.

Proposition 22. *Let* $(\boldsymbol{H}, \boldsymbol{H'}, \boldsymbol{s}, F)$ *be an instance of the* $ARSD_{q,m,n,k,r,n',F}$ *problem. If* $\max(m,n)r + n'r' \geqslant m(n-k)$ *then the* ARSD *problem can be solved in polynomial time with a probabilistic algorithm.*

Proof. To prove this proposition, we will use the method used to compute the Singleton bound.

Let us begin with the case $n \geqslant m$. Let E be a subspace of $\mathbb{F}_{q^m}$ of dimension r and suppose that there exists a solution $(\boldsymbol{x}, \boldsymbol{x}')$ of the ARSD problem such that $\mathrm{Supp}(\boldsymbol{x}) = E$. Then, we can express the coordinate x_i of $\boldsymbol{x}$ in a basis E_j of E:

$$\forall i \in \{1 \dots n\}, x_j = \sum_{i=1}^{r} \lambda_{ij} E_i$$

Likewise, we can express the coordinates of $\boldsymbol{x}'$ in a basis of F:

$$\forall i \in \{1 \dots n'\}, x'_t = \sum_{s=1}^{r'} \lambda'_{st} F_s$$

Let us write the linear system satisfied by the unknown $(\lambda_{ij}, \lambda'_{st})$:

$$\begin{aligned}
& \boldsymbol{H}\boldsymbol{x}^T \quad + \quad \boldsymbol{H}'\boldsymbol{x}'^T \quad = \boldsymbol{s} \\
\iff \forall i \in \{1..n-k\}, & \sum_{j=1}^{n} H_{ij} x_j \quad + \quad \sum_{t=1}^{n'} H'_{it} x'_t \quad = s_i \\
\iff \forall i \in \{1..n-k\}, & \sum_{j=1}^{n} H_{ij} \sum_{i=1}^{r} \lambda_{ij} E_i + \sum_{t=1}^{n'} H'_{it} \sum_{s=1}^{r'} \lambda'_{st} F_s = s_i \qquad (9)
\end{aligned}$$

The $(n-k)$ linear equations (9) over $\mathbb{F}_{q^m}$ can be projected on $\mathbb{F}_q$ to obtain $m(n-k)$ linear equations over $\mathbb{F}_q$. Since we have $nr + n'r \geqslant m(n-k)$, there are more unknowns than equations so the system admits at least a solution with a non negligible probability.

In the case $m > n$, we need to consider the matrix $\boldsymbol{M}(\boldsymbol{x})$ associated to $\boldsymbol{x}$ (cf definition 1) and express its rows in a basis of a subspace E of dimension r of $\mathbb{F}_q^n$. Since the support of $\boldsymbol{x}'$ is fixed, its coordinates still give us $n'r'$ unknowns over $\mathbb{F}_q$. This gives us $mr + n'r'$ unknowns over $\mathbb{F}_q$ in total. Then we transform the equation $\boldsymbol{H}\boldsymbol{x}^T + \boldsymbol{H}'\boldsymbol{x}'^T = \boldsymbol{s}$ into a linear system over $\mathbb{F}_q$ as previously. This operation is not difficult but technically complicated and we do not give the details of the equations. Finally we obtain a linear system of $m(n-k)$ equations and $mr+n'r'$ unknowns over $\mathbb{F}_q$. This system has a solution with a non negligible probability since $mr + n'r' \geqslant m(n-k)$.

In both case, the solution of the system gives us a solution to the instance of the ARSD problem. □

In the case where $\max(m,n)r + n'r' < m(n-k)$, we need to adapt the best attacks against the RSD problem in the case of the ARSD problem. The attack is detailed in [16]. The general idea is to find a subspace E of dimension δ such that $\mathrm{Supp}(\boldsymbol{x}) \subset E$ (in the case $n \geqslant m$). Then we can express the coordinates of $\boldsymbol{x}$ if $n \geqslant m$ or the rows of the matrix $\boldsymbol{M}(\boldsymbol{x})$ if $m > n$ in a basis of E exactly as in the previous proposition. We want δ as large as possible to increase the probability that $\mathrm{Supp}(\boldsymbol{x}) \subset E$ but we have to take δ such that $\max(m,n)\delta + n'r' < m(n-k)$ in order to obtain an over-constrained linear system. Hence $\delta = \left\lfloor \frac{m(n-k)-n'r'}{\max(m,n)} \right\rfloor$. The probability that $E \supset \mathrm{Supp}(\boldsymbol{x})$ depends on m and n:

- If $n \geqslant m$, then E is a subspace of $\mathbb{F}_{q^m}$: $\mathbb{P}(E \supset \mathrm{Supp}(\boldsymbol{x})) = \dfrac{\begin{bmatrix}\delta\\ r\end{bmatrix}_q}{\begin{bmatrix}m\\ r\end{bmatrix}_q} = \Theta(q^{-r(m-\delta)})$.
- If $n < m$ then E is a subspace of $\mathbb{F}_q^n$: $\mathbb{P}(E \supset \mathrm{Supp}(\boldsymbol{x})) = \dfrac{\begin{bmatrix}\delta\\ r\end{bmatrix}_q}{\begin{bmatrix}n\\ r\end{bmatrix}_q} = \Theta(q^{-r(n-\delta)})$.

In order to respect the constraints of our signature, we have to take parameters such that the instance of the ARSD problem has several solutions. Thus, the average complexity of this attack is equal to the inverse of the probability $\mathbb{P}(E \supset \mathrm{Supp}(\boldsymbol{x}))$ divided by the number of solutions times the cost of the linear algebra. The number of solutions is in $\Theta\left(q^{r(m+n-r)+n'r'-m(n-k)}\right)$ (see Proposition 9 for the details).

Proposition 23. *In the case* $\max(m,n)r + n'r' < m(n-k)$*, the complexity of the best attack against the* $\mathrm{ARSD}_{q,m,n,k,r,n',F}$ *problem is in*

$$\mathcal{O}\left(m^3(n-k)^3 q^{r\left\lceil \frac{km+n'r'}{\max(m,n)}\right\rceil - r(m+n-r)-n'r'+m(n-k)}\right).$$

Remark: We did not consider the improvement of the attack of the RSD problem in [5] because this attack does not fit the case where there are several solutions to the RSD problem.

6 Parameters

6.1 Constraints

In this section we recap the different constraints on our parameters.

Choice of l, l', r and d

First we need to choose l' such that the entropy of $\boldsymbol{c}$ is high enough. For our parameters, $l' = 1$ is always enough since $\boldsymbol{c} \in F^{l'dk}$ and $dk > 512$. In practice using less than dk coordinates for $\boldsymbol{c}$ is a possibility to make the parameters a little smaller.

We then need to choose r high enough such that both the attack on both the RSD and RSL problems are hard. d and l must be chosen such that $\lambda \geqslant r + d$: $d = r$ and $l = 4$ is a way to meet this condition. In the sets of parameters given below, this value of l leads to $N = k(r-1)$ and $N = k(r-2)$ respectively, which allows us to be pretty conservative with respect to the attacks on the RSL problem.

Choice of m

In order to avoid the distinguisher attack for a security parameter of 128, the relation $m - 2u \geqslant 128 + 64$ (we use Proposition 18, where $u = rd - \lambda$, must be verified to fit

the security proof : we consider that the adversary has access to 2^{64} signatures, so the probability of distinguishing signatures and random vectors must be lower than 2^{-192}. We choose a prime m (so there is no intermediate field between $\mathbb{F}_q$ and $\mathbb{F}_{q^m}$) such that $m \geqslant 192 + 2u$.

Choice of n, k and w

n, k and w must be chosen such that $3(u+w) > m$ to avoid the distinguisher attack using subspaces of dimension 3, and $(u+w) < (n-k) - \lambda$ in order to keep the weight of the signature below the Singleton bound $-\lambda$ (due to ARSD). k is taken prime for having access to really sparse polynomials to define the ideal codes.

6.2 Example of parameters

The public key consists of :

- $\boldsymbol{H}$ which can be recovered from a seed (256 bits)
- $l(n-k)m\log(q)$ bits to describe the syndromes

The signature consists of :

- $(rd + \omega - \lambda)(n + m - rd - \omega + \lambda)\log(q)$ bits to describe $\boldsymbol{z}$. We give $\mathrm{Supp}(\boldsymbol{z})$ in echelon form as well as the coordinates in this basis
- A seed to describe F (256 bits)
- 512 bits to describe $\boldsymbol{c}$
- $dlk\log(q)$ bits to describe $\boldsymbol{p}$

The complexity of the key recovery attack is computed using the complexity of the combinatorial attack given in Section 5.1.

For our parameters, the complexity of the forgery attack using the algorithm against ARSD described in Section 5.2 is disproportionately large compared to the key recovery attack.

As stated in 6.1, to target a security level of 2^{128}, we choose our parameters such that the probability of distinguishing signatures from random vectors is smaller than 2^{-192}, considering the maximum number of calls to the signature oracle an attacker can make is 2^{64}.

	m	n	k	l	l'	d	r	ω	λ	q	Public key size	Signature size	Key recovery attack	Distinguisher	Security
I	241	202	101	4	1	6	6	57	12	2	121 961	32 514	461	193	128
II	263	226	113	4	1	7	7	56	14	2	148 851	40 150	660	193	128

The implementation of our scheme on an Intel(R) Core(TM) i5-7440HQ CPU running at 2.80GHz gives the following computation times :

Parameter	Keygen	Online signature phase	Verification
I	4ms	4ms	5ms
II	5ms	5ms	6ms

For the offline phase, the most costly step, the computation of the matrix $\boldsymbol{D}$, takes 350ms for parameter I and 700ms for parameter II.

7 Conclusion

We have described a variation on the Schnorr-Lyubashevsky approach for rank based cryptography. This new approach enables us to obtain a randomization of the signature, which seemed difficult to derive before this work for code-based cryptography. We provide a detailed analysis of attacks and an EUF-CMA proof for our scheme. Overall the parameters we propose are efficient and comparable in terms of signature size to the Dilithium lattice-based scheme.

References

[1] Carlos Aguilar, Philippe Gaborit, and Julien Schrek. A new zero-knowledge code based identification scheme with reduced communication. In *Proc. IEEE Inf. Theory Workshop- ITW 2011*, pages 648–652. IEEE, October 2011. *2*

[2] Carlos Aguilar Melchor, Nicolas Aragon, Paulo Barreto, Slim Bettaieb, Loïc Bidoux, Olivier Blazy, Jean-Christophe Deneuville, Philippe Gaborit, Shay Gueron, Tim Güneysu, Rafael Misoczki, Edoardo Persichetti, Nicolas Sendrier, Jean-Pierre Tillich, and Gilles Zémor. BIKE. first round submission to the NIST post-quantum cryptography call, November 2017. *1*

[3] Carlos Aguilar Melchor, Nicolas Aragon, Slim Bettaieb, Loïc Bidoux, Olivier Blazy, Jean-Christophe Deneuville, Philippe Gaborit, and Gilles Zémor. Rank quasi cyclic (RQC). first round submission to the NIST post-quantum cryptography call, November 2017. *1*

[4] Nicolas Aragon, Slim Bettaieb, Loic Bidoux, Olivier Blazy, Jean-Christophe Deneuville, Phillipe Gaborit, Carlos Aguilar Melchor, Edoardo Persichetti, and Gilles Zémor. HQC, December 2017. NIST Round 1 submission for Post-Quantum Cryptography. *1*

[5] Nicolas Aragon, Philippe Gaborit, Adrien Hauteville, and Jean-Pierre Tillich. A new algorithm for solving the rank syndrome decoding problem. In *Proc. IEEE Int. Symposium Inf. Theory - ISIT*, Vail, USA, 2018. IEEE. *23, 27, 29*

[6] Nicolas Aragon Aragon, Philippe Gaborit, Olivier Ruatta, and Gilles Zemor. More on lrpc codes and their cryptographic applications. `https://www.unilim.fr/pages_perso/philippe.gaborit/newLRPC.pdf`, 2017. *18*

[7] Nicolas Courtois, Matthieu Finiasz, and Nicolas Sendrier. How to achieve a McEliece-based digital signature scheme. In *Advances in Cryptology - ASIACRYPT 2001*, volume 2248 of *LNCS*, pages 157–174, Gold Coast, Australia, 2001. Springer. *1, 2*

[8] Thomas Debris-Alazard, Nicolas Sendrier, and Jean-Pierre Tillich. The problem with the surf scheme. preprint, November 2017. arXiv:1706.08065. *2*

[9] Thomas Debris-Alazard and Jean-Pierre Tillich. A polynomial attack on a NIST proposal: Ranksign, a code-based signature in rank metric. preprint, April 2018. IACR Cryptology ePrint Archive. *2, 27*

[10] Thomas Debris-Alazard and Jean-Pierre Tillich. Two attacks on rank metric code-based schemes: Ranksign and an identity-based-encryption scheme. In *Advances in Cryptology - ASIACRYPT 2018*, LNCS, Brisbane, Australia, December 2018. Springer. *23*

[11] Jean-Charles Faugère, Valérie Gauthier, Ayoub Otmani, Ludovic Perret, and Jean-Pierre Tillich. A distinguisher for high rate McEliece cryptosystems. IACR Cryptology ePrint Archive, Report2010/331, 2010. http://eprint.iacr.org/. *2*

[12] Kazuhide Fukushima, Partha Sarathi Roy, Rui Xu, Shinsaku Kiyomoto, Kirill Morozov, and Tsuyoshi Takagi. RaCoSS, December 2017. NIST Round 1 submission for Post-Quantum Cryptography. *3*

[13] Philippe Gaborit, Adrien Hauteville, Duong Hieu Phan, and Jean-Pierre Tillich. Identity-based encryption from rank metric. IACR Cryptology ePrint Archive, Report2017/623, May 2016. `http://eprint.iacr.org/`. *6, 23, 25, 26*

[14] Philippe Gaborit, Adrien Hauteville, Duong Hieu Phan, and Jean-Pierre Tillich. Identity-based encryption from rank metric. In *Advances in Cryptology - CRYPTO2017*, volume 10403 of *LNCS*, pages 194–226, Santa Barbara, CA, USA, August 2017. Springer. *6*

[15] Philippe Gaborit, Gaétan Murat, Olivier Ruatta, and Gilles Zémor. Low rank parity check codes and their application to cryptography. In *Proceedings of the Workshop on Coding and Cryptography WCC'2013*, Bergen, Norway, 2013. Available on `www.selmer.uib.no/WCC2013/pdfs/Gaborit.pdf`. *10, 13, 18*

[16] Philippe Gaborit, Olivier Ruatta, and Julien Schrek. On the complexity of the rank syndrome decoding problem. *IEEE Trans. Information Theory*, 62(2):1006–1019, 2016. *27, 28*

[17] Philippe Gaborit, Olivier Ruatta, Julien Schrek, and Gilles Zémor. New results for rank-based cryptography. In *Progress in Cryptology - AFRICACRYPT 2014*, volume 8469 of *LNCS*, pages 1–12, 2014. *1*

[18] Philippe Gaborit, Olivier Ruatta, Julien Schrek, and Gilles Zémor. Ranksign: An efficient signature algorithm based on the rank metric (extended version on arxiv). In *Post-Quantum Cryptography 2014*, volume 8772 of *LNCS*, pages 88–107. Springer, 2014. *2*

[19] Philippe Gaborit, Julien Schrek, and Gilles Zémor. Full cryptanalysis of the chen identification protocol. In *Post-Quantum Cryptography - 4th International Workshop, PQCrypto 2011, Taipei, Taiwan, November 29 - December 2, 2011. Proceedings*, pages 35–50, 2011. *2*

[20] Philippe Gaborit and Gilles Zémor. On the hardness of the decoding and the minimum distance problems for rank codes. *IEEE Trans. Information Theory*, 62(12):7245–7252, 2016. *6*

[21] Craig Gentry, Chris Peikert, and Vinod Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 197–206. ACM, 2008. *1, 2*

[22] Shafi Goldwasser, Silvio Micali, and Ronald L Rivest. A digital signature scheme secure against adaptive chosen-message attacks. *SIAM Journal on Computing*, 17(2):281–308, 1988. *23*

[23] Johan Håstad, Russell Impagliazzo, Leonid A Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999. *24*

[24] Adrien Hauteville and Jean-Pierre Tillich. New algorithms for decoding in the rank metric and an attack on the LRPC cryptosystem, 2015. abs/1504.05431. *6*

[25] Jeffrey Hoffstein, Nick Howgrave-Graham, Jill Pipher, Joseph H Silverman, and William Whyte. Ntrusign: Digital signatures using the ntru lattice. In *Cryptographers' Track at the RSA Conference*, pages 122–140. Springer, 2003. *1, 2*

[26] Wijik Lee, Young-Sik Kim, Yong-Woo Lee, and Jong-Seon No. pqsigRM, December 2017. NIST Round 1 submission for Post-Quantum Cryptography. *1*

[27] Pierre Loidreau. On cellular code and their cryptographic applications. In I. Landjev G. Kabatiansky, editor, *Proceedings of ACCT14 (algebraic and combinatorial coding theory)*, pages 234–239, Svetlogorsk, Russia, September 2014. *8*

[28] Vadim Lyubashevsky. Fiat-shamir with aborts: Applications to lattice and factoring-based signatures. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 598–616. Springer, 2009. *2*

[29] Vadim Lyubashevsky. Lattice signatures without trapdoors. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 738–755. Springer, 2012. *9*

[30] Vadim Lyubashevsky, Leo Ducas, Eike Kiltz, Tancrede Lepoint, Peter Schwabe, Gregor Seiler, and Damien Stehle. CRYSTALS-DILITHIUM, December 2017. NIST Round 1 submission for Post-Quantum Cryptography. *2*

[31] Claus-Peter Schnorr. Efficient signature generation by smart cards. *Journal of cryptology*, 4(3):161–174, 1991. *2*

[32] Jacques Stern. A new identification scheme based on syndrome decoding. In D.R. Stinson, editor, *Advances in Cryptology - CRYPTO'93*, volume 773 of *LNCS*, pages 13–21. Springer, 1993. *2*