Reproducible Codes and Cryptographic Applications

Paolo Santini[1], Edoardo Persichetti[2], and Marco Baldi[1]

[1]Università Politecnica delle Marche, [2]Florida Atlantic University

Abstract. In this paper we study structured linear block codes, starting from well-known examples and generalizing them to a wide class of codes that we call *reproducible codes*. These codes have the property that they can be entirely generated from a small number of signature vectors, and consequently admit matrices that can be described in a very compact way. We then show some cryptographic applications of this class of codes and explain why the general framework we introduce may pave the way for future developments of code-based cryptography.

Keywords: Linear block codes, code-based cryptography, post-quantum cryptography, reproducible codes

1 Introduction

Defining linear block codes that possess a certain inner structure and verify some regularity properties is a natural process in coding theory. Arguably, the most relevant example is represented by the class of *cyclic codes*, which includes several families of codes that proved to be important throughout the history of communications, such as BCH and Hamming codes, as well as the binary Golay codes, Reed-Solomon codes and many others. This class is defined by the property of having codewords that are invariant under the action of a specific permutation, namely the cyclic (circular) shift, i.e. the rotation of the vector to the right (equivalently, to the left). Other examples which are well-known in literature include *constacyclic* codes, *negacyclic* codes, *quasi-cyclic* codes and many others.

In recent times, this research direction has been investigated further: Misoczki and Barreto in 2009 introduced the family of *quasi-dyadic* codes [29], which contain codewords that are invariant under a different type of permutation. The work was motivated by its applications in the framework of the McEliece cryptosystem [28], and in particular by the necessity of having a family of codes which possess generator and parity-check matrices that can be represented in a compact way. The reason behind this is that, in code-based cryptography, the public key of an encryption (or signature) scheme usually consists precisely of a generator or parity-check matrix of a linear block code. With the size of the codes used in code-based cryptography (typical code length values are in the order of 10^3 to 10^4), describing a whole matrix results in a public key of several kilobytes, and this size increases quadratically in the code length. This has historically prevented the use of the original McEliece cryptosystem, which exploits random-looking public codes, in many applications. On the other hand, structured codes admit a generator and parity-check matrix which can be entirely described by one or few rows; this allows for a very important reduction in public-key size, and it is arguably a fundamental step towards making code-based cryptography truly practical. Previous efforts to reduce key size were centered on quasi-cyclic algebraic codes [21] and have been since then extended to codes of a different nature, namely the low-density parity-check (LDPC) codes [3] and their recent generalization known as moderate-density parity-check (MDPC) codes [30]. These codes are characterized by sparse parity-check matrices and admit characteristic matrices in quasi-cyclic form, i.e. formed by circulant square blocks. Due to their efficient decoding algorithms and the lack of additional algebraic structure that could lead to structural attacks, schemes based on quasi-cyclic low-density parity-check (QC-LDPC) codes and quasi-cyclic moderate-density parity-check (QC-MDPC) codes are among the most promising solutions in the area.

The importance of code-based cryptography has risen dramatically in modern times due to the work of Peter Shor [35], which shows how it will be possible to effectively break cryptography based on “classical” number theory problems by introducing polynomial-time algorithms for factoring large integers and computing discrete logarithms on a quantum computer. This means that the cryptographic community has to devise primitives that rely on different hard problems, which will not be affected once quantum computers of an appropriate size will be available. Code-based cryptography is one of the most important areas in this scenario, and ever since McEliece’s seminal work in 1978, has shown no vulnerabilities against quantum attackers. Moreover, *generic* decoding attacks, which have exponential complexity, have improved only marginally over nearly 40 years of security history. Together with schemes based on lattice problems, code-based cryptography is at the basis of many candidates for the Post-Quantum Standardization call recently launched by NIST [2].

Our Contribution In this paper we analyze in detail the nature of structured codes. First, we introduce the notion of *reproducible* codes, which captures the generic idea of a code admitting characteristic matrices that can be entirely described by a subset of their rows. To the best of our knowledge, it is the first time such a broad concept is introduced and studied in its entirety. We then show that all the existing constructions of structured codes (cyclic, quasi-cyclic, dyadic etc.) are in fact but a special case of our general formulation – in particular, corresponding to the simplest case where the codes are “reproduced” via permutations applied to a single row. Finally, we propose a framework for constructing reproducible codes of any kind, and present concrete instantiations of non-trivial reproducible codes which have not appeared in literature before.

2 Preliminaries

We denote with $\mathbb{F}_q$ the finite field with q elements, where q is a prime power. For two sets X and Y we denote by X^Y the set of all functions from X to Y. For a set S we then denote by 2^S its power set, i.e. the set containing all possible subsets of S, exploiting the well-known bijection between the power set of S and the set of functions from S to $\{0,1\}$. We use bold letters to denote vectors and matrices. Given a vector $\mathbf{a}$, we refer to its element in position i as a_i. The size-k identity matrix is denoted as $\mathbf{I}_k$, while the $k \times n$ null matrix is denoted as $\mathbf{0}_{k\times n}$.

2.1 Coding theory background

A linear code $\mathcal{C}$ is a k-dimensional subspace of the n-dimensional vector space over the finite field $\mathbb{F}_q$. The parameters n (*length*) and k (*dimension*) are positive integers with $k \leq n$. The value $r = n - k$ is known as *codimension* of the code.

Definition 1 (Hamming metric). *The Hamming weight $\mathsf{wt}(\mathbf{x})$ of a vector $\mathbf{x} \in \mathbb{F}_q^n$ is the number of its non-zero entries. The Hamming distance $\mathsf{d}(\mathbf{x},\mathbf{y})$ between two vectors $\mathbf{x},\mathbf{y} \in \mathbb{F}_q^n$ is defined as the weight of their difference, i.e. $\mathsf{d}(\mathbf{x},\mathbf{y}) = \mathsf{wt}(\mathbf{x}-\mathbf{y})$. The minimum distance d of a code $\mathcal{C}$ is defined as the minimum distance between any two different codewords of $\mathcal{C}$, or equivalently as the minimum weight over all non-zero codewords.*

A linear code of length n, dimension k, and minimum distance d is called an $[n,k,d]$-code.

The error-correcting capability of a linear code is connected to its minimum distance, and in particular it corresponds to $\lfloor (d-1)/2 \rfloor$ under bounded distance decoding. When soft-decision decoding is used, a linear block code with distance d may correct up to $d-1$ symbol errors.

Definition 2 (Generator and parity-check matrices). *Let $\mathcal{C}$ be a linear code over $\mathbb{F}_q$. We call* generator matrix *of $\mathcal{C}$ a $k \times n$ matrix $\mathbf{G}$ whose rows form a basis for the vector space defined by $\mathcal{C}$, i.e.:*

$$\mathcal{C} = \{\mathbf{xG} : \mathbf{x} \in \mathbb{F}_q^k\}.$$

For any matrix $\mathbf{H}$ and any vector $\mathbf{x}$, the vector $\mathbf{Hx}^T$ is called syndrome *of $\mathbf{x}$. We then call* parity-check matrix *of $\mathcal{C}$ an $r \times n$ matrix $\mathbf{H}$ such that every codeword belonging to $\mathcal{C}$ has syndrome 0 with respect to $\mathbf{H}$, i.e.*

$$\mathcal{C} = \{\mathbf{x} \in \mathbb{F}_q^n : \mathbf{Hx}^T = 0\}.$$

Note that the parity-check matrix of a code $\mathcal{C}$ is also a generator matrix for the *dual* code $\mathcal{C}^\perp$, i.e. the linear code formed by all the words of $\mathbb{F}_q^n$ that are orthogonal to $\mathcal{C}$. It follows that for any generator matrix $\mathbf{G}$ and parity-check matrix $\mathbf{H}$ of a code, we have $\mathbf{HG}^T = \mathbf{0}_{r\times k}$.

Both matrices are required to have full rank. Moreover, notice that, clearly, neither matrix is unique: for instance, given a generator matrix $\mathbf{G}$ it is always possible to obtain another generator matrix for the same code by a linear transformation, that is, the multiplication on the left by an invertible $k \times k$ matrix $\mathbf{S}$, so that $\mathbf{G}' = \mathbf{S}\mathbf{G}$. This corresponds to a change of basis for the vector space. A similar property is verified by the parity-check matrix. Finally, two generator matrices generate *equivalent codes* if one is obtained from the other by a permutation of columns. These two facts are at the basis of the McEliece cryptosystem.

Joining these two properties, we can write any generator matrix $\mathbf{G}$ in *systematic form* as $\mathbf{G} = [\mathbf{I}_k|\mathbf{A}]$, where $\mathbf{I}_k$ is the identity matrix of size k and $|$ denotes concatenation. If $\mathcal{C}$ is generated by $\mathbf{G} = [\mathbf{I}_k|\mathbf{A}]$, then a (systematic) parity-check matrix for $\mathcal{C}$ is $\mathbf{H} = [-\mathbf{A}^T|\mathbf{I}_r]$.

2.2 The McEliece cryptosystem

The McEliece public-key encryption scheme [28] was introduced by R. J. McEliece in 1978. The original scheme uses binary Goppa codes, with which it remains unbroken (with a proper choice of parameters), but the scheme can be used with any class of codes for which an efficient decoding algorithm is known.

Key Generation Let $\mathbf{G}$ be a generator matrix for a linear $[n, k, d]$-code over $\mathbb{F}_q$ with an efficient decoding algorithm $\mathcal{D}$ which can correct up to $t = \lfloor (d-1)/2 \rfloor$ errors under bounded-distance decoding. Let $\mathbf{S}$ be an invertible $k \times k$ matrix and P be a random $n \times n$ permutation matrix over $\mathbb{F}_q$. The private key is $(\mathbf{S}, \mathbf{G}, \mathbf{P})$ and the public key is $\mathbf{G}' := \mathbf{S}\mathbf{G}\mathbf{P}$.

Encryption To be able to encrypt a plaintext, it has to be represented as a vector $\mathbf{m}$ of length k over $\mathbb{F}_q$. The encryption algorithm chooses a random error vector $\mathbf{e}$ of weight t in $\mathbb{F}_q^n$, and computes the ciphertext $\mathbf{c} = \mathbf{m}\mathbf{G}' + \mathbf{e}$.

Decryption The decryption algorithm first computes $\hat{\mathbf{c}} = \mathbf{c}\mathbf{P}^{-1} = \mathbf{m}\mathbf{S}\mathbf{G} + \mathbf{e}\mathbf{P}^{-1}$. As $\mathbf{P}$ is a permutation matrix, $\mathbf{e}\mathbf{P}^{-1}$ has the same weight as $\mathbf{e}$. Therefore, $\mathcal{D}$ can be used to decode the errors and obtain $\hat{\mathbf{m}} = \mathbf{m}\mathbf{S} = \mathcal{D}(\hat{\mathbf{c}})$. Finally, the plaintext is retrieved as $\mathbf{m} = \hat{\mathbf{m}}\mathbf{S}^{-1}$.

In successive papers, the original McEliece cryptosystem was refined and tweaked many times; for example it is now common practice to replace the scrambling method given by $\mathbf{S}$ and $\mathbf{P}$ with the computation of the systematic form, i.e. $\mathbf{G}'$ is the systematic form of $\mathbf{G}$. This is possible when the McEliece cryptosystem is embedded into a larger framework to convert it into an IND-CCA2 secure PKE or KEM, and has the additional advantage (beyond the obvious simpler formulation) of a smaller public key (since only the non-identity submatrix needs to be stored).

The (one-way) security of McEliece is largely based on the following hard problem.

Problem 1 (Syndrome Decoding Problem). Given an $r \times n$ full-rank matrix $\mathbf{H}$ and a vector $\mathbf{s}$, both with entries in $\mathbb{F}_q$, and a non-negative integer t; find a vector $\mathbf{e} \in \mathbb{F}_q^n$ of weight t such that $\mathbf{He}^T = \mathbf{s}^T$.

The Syndrome Decoding Problem (SDP) is a well-known problem in complexity theory, and it has been shown to be NP complete [13]. Note that, since the McEliece cryptosystem uses an $[n, k, d]$ code, the number of error vectors is $\binom{n}{t}(q-1)^t$, while the number of possible syndromes is q^r. Therefore

$$\binom{n}{t}(q-1)^t < q^r$$

is a necessary condition for the existence of at most one solution to the problem, i.e., for the decoding process to have a unique solution.

3 Sparse-matrix codes

One of the most delicate points about the McEliece cryptosystem is that, in order for the security to reduce to SDP, it is assumed that the matrix produced as the public key is indistinguishable from a uniformly random matrix of the same size. This is a plausible assumption, which however has been shown to be false in several cases. For many variants of McEliece (e.g. [36]), in fact, this opened up avenues of attack which simply ruled out the variant altogether. Even the long-standing binary Goppa codes have been shown to be distinguishable from random codes [20] when the code rate is chosen carelessly (too high). This is arguably one of the main reasons that pushed researchers away from algebraic codes, and towards codes of a different nature.

Low-Density Parity-Check (LDPC) codes are defined by parity-check matrices whose main requirement is to be sparse, with a very low row and column weight. These codes are easy to generate, and moreover admit a variety of choices for the decoding algorithm $\mathcal{D}$, like the Bit Flipping (BF) decoder of Gallager [22], which is very efficient in practice. For these reasons, this class of codes is a natural candidate for the McEliece cryptosystem. In such a framework, the private code $\mathcal{C}$ is represented through its sparse parity-check matrix $\mathbf{H}$, while the public key corresponds to a generator matrix $\mathbf{G}$ for $\mathcal{C}$. It is important to note that, from the knowledge of $\mathbf{G}$, the opponent can compute several parity-check matrices $\mathbf{H}'$ for $\mathcal{C}$, but they will not lead to an efficient decoding, unless they are sparse. As explained in section 2.2, typically having $\mathbf{G}$ in systematic form is enough to guarantee such a property. Indeed, we can always write $\mathbf{H} = [\mathbf{H}_0|\mathbf{H}_1]$, where $\mathbf{H}_0$ and $\mathbf{H}_1$ have size $r \times k$ and $r \times r$, respectively. Then, the corresponding generator matrix in systematic form is obtained as $\mathbf{G} = [\mathbf{I}_k|\mathbf{H}_0^T\mathbf{H}_1^{-T}]$. Typically (unless for specific choices of $\mathbf{H}$) the inverse of a sparse matrix is dense, and so $\mathbf{H}_1^{-T}$ is dense: in such a case, the multiplication of $\mathbf{H}_0^T$ by $\mathbf{H}_1^{-T}$ is enough to hide the structure of $\mathbf{H}$ into the one of $\mathbf{G}$.

It is important to note that, due to their probabilistic nature, decoding algorithms for LDPC codes are characterized by a non-trivial decoding failure rate (DFR). This means that, in the case of a decoding failure, Bob must ask Alice for a retransmission of the plaintext, encrypted with a different error vector. In order to avoid frequent retransmissions, which would obviously increase the latency of the system, the DFR must be kept sufficiently low; typically, values are in the range of 10^{-6} to 10^{-9}. As we will discuss later, this fact represents a crucial difference, with respect to the case of algebraic codes, since it leads to a new family of attacks, aimed at recovering the secret key by observing Bob's reactions. This also has implications on the security model against a Chosen Ciphertext Attack (CCA) for these systems.

3.1 Main attacks

We briefly recall the two main types of attacks that can be mounted against the McEliece cryptosystem and its variants when using sparse-matrix codes.

Decoding attacks Decoding attacks are aimed at recovering the plaintext from the ciphertext by performing decoding through the public code. In fact, being unable to retrieve the private code representation that enables efficient decoding, an attacker can still try to perform decoding through the public code, which looks like a general random code.

At the current state of the art, the best procedure for this task is the information set decoding (ISD) algorithm, which was first introduced by Prange in 1962 [32], and has received many improvements during the years [11, 26, 27, 37]. However, ISD and all its variants are characterized by an exponential complexity: the search for a weight-w codeword has asymptotic complexity equal to $2^{\alpha w}$, where the value of the constant α depends on the code parameters and on the particular algorithm we are analyzing. Even in a quantum setting, ISD algorithms are still characterized by exponential complexity: indeed, the only known application of a quantum algorithm to an ISD algorithm, which consists in using Grover's algorithm [23] to speed up the search, leads to a reduction in the complexity, with respect to the classical case, which cannot be larger than the square root of the exponent α [14].

Key-recovery attacks When LDPC codes are used, security against key recovery attacks depends on the difficulty of recovering low-weight codewords from the dual of the public code, which is again a decoding problem. In particular, let us denote by $\mathcal{C}^{\perp}$ the dual code of $\mathcal{C}$, which admits $\mathbf{H}$ as generator matrix. Since the rows of $\mathbf{H}$ are sparse, and have maximum weight $w \ll n$, with overwhelming probability they represent minimum-weight codewords in $\mathcal{C}^{\perp}$, and so can be searched with a generic algorithm for finding low-weight words. ISD algorithms can be exploited for this purpose as well.

Since the main threat to the use of LDPC codes is represented by a search for low-weight words, it makes sense to relax the notion of "low-density": the authors in [30] introduce the notion of "moderate-density" by increasing the allowed row weight in the parity-check matrix from $O(1)$ to $O(\sqrt{n})$. It is still possible to decode such codes (called MDPC by analogy) with the previously-mentioned algorithms; the error-correction capacity gets obviously worse, but the gain in security makes this tradeoff worth it.

In the end, the adoption of LDPC and MDPC codes does not reduce the security of the McEliece cryptosystem against key recovery attacks, since attacks deriving from the structure of the secret code can be easily avoided by fixing the minimum weight of the rows of $\mathbf{H}$.

3.2 Structured sparse-matrix codes

Using generic LDPC and MDPC codes without any structure is not a practical choice, since the resulting public key sizes are significantly larger than the ones we can obtain with other families of codes, like Goppa codes. In fact, even if the private sparse parity-check matrix can be compactly represented through the positions of its non-null entries (and so, a row with Hamming weight equal to w can be stored just with $w \log_2 n \log_2 q$ bits), applying this technique to the public key is not possible, since a sparse $\mathbf{G}$ might compromise the security of the system. One way to avoid this issue is to add some structure to the code family. This idea was first introduced by considering quasi-cyclic (QC) codes [21], and was then extended to LDPC codes [6] and algebraic codes [12]. In all cases, the authors propose to use QC codes to reduce key size. A QC code is simply a code which admits parity-check and generator matrices made of *circulant* blocks. A circulant matrix is a matrix in which every row is obtained as the cyclic shift of the previous one; an example of a circulant matrix is reported below.

$$\begin{bmatrix} a_0 & a_1 & \dots & a_{p-1} \\ a_{p-1} & a_0 & \dots & a_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \dots & a_0 \end{bmatrix}$$

This means that, in the McEliece cryptosystem, we can describe the public key completely using just the first row of each circulant block; it is clear that this results in a significant reduction in the public key size. However, this additional structure presents some drawbacks, since it exposes the system to structural weaknesses. In particular, the QC structure summed to the algebraic structure of the underlying codes provides a lot of information to the attacker, and opens up the possibility of structural attacks aimed at recovering the private code. The most famous structural attack of this type is known as FOPT [19], and works by solving a multivariate algebraic system with Gröbner bases techniques together with the QC property which greatly reduces the number of unknowns of the system. As a result, it seems very hard to provide secure schemes which involve QC algebraic codes (Goppa, GRS etc.), while still obtaining an effective key reduction: the recent NIST proposal BIG QUAKE [1] shows a reduction of about $1/4$ of the key size which would be obtained in a "classical" McEliece using unstructured binary Goppa codes.

Therefore, once again, it seems safer to deploy code-based schemes using sparse-matrix codes, since in this case there is no additional algebraic structure, and the QC property alone is not enough to provide a structural attack. However, some care is still necessary when using sparse-matrix codes. In particular, two main aspects have to be considered:

- ISD algorithms might obtain a speed up from the QC structure. This results in a complexity reduction for the relevant attacks. Such a speedup is achieved for both key recovering attacks and decoding attacks (following from the Decoding One Out of Many (DOOM) approach [34]). The attack complexity remains exponential in the key length, but the attack speedup leads to an increase in the row weight of $\mathbf{H}$ and in the number of errors to be used during encryption, which in turn results in an increase in the key length.

- It has been recently shown that the probability of a decoding failure depends on the number of overlapping ones between the error vector and rows of $\mathbf{H}$ [24]. In addition, in a circulant matrix, all the rows are characterized by the same set of cyclic distances between set symbols (given two ones at positions i and j, the corresponding cyclic distance is computed as $\min\{\pm(i-j) \mod p\}$, with p being the circulant size). Based on these considerations, it has been shown in [24] that an adversary can mount a key recovery attack by impersonating Alice, producing many ciphertexts and requesting Bob to decrypt them. The adversary can then exploit Bob's reactions concerning decoding failures, which are of public knowledge, in order to gather information about the secret key structure. The set of all distances of the rows of $\mathbf{H}$ is called *distance spectrum*, and can be used to reconstruct $\mathbf{H}$. This problem can be related to a graph problem, in which a row of $\mathbf{H}$ corresponds to a clique with maximum size. For a sparse QC matrix, such a graph is sparse as well, which gives a small number of cliques. This means that, once the distance spectrum is known, recovering the corresponding parity-check matrix is not a hard task in most cases.

Currently, the countermeasures that have been devised against the aforementioned reaction attacks exploit the use of ephemeral keys [4, 10], or of particular families of codes which make the reconstruction of the secret key unfeasible [33]. However, both these solutions come with some price to pay, since a new key-pair must be generated for each encryption (in the first case) or the size of the public key must be increased (in the second case). Another solution could be that of reducing the DFR to a negligible value, in order to increase the number of ciphertexts that the opponent must produce to recover the secret distance spectrum [38]. However, this requires the use of much longer codes, thus yielding a considerable increase in the public key size.

As we will see in the rest of this paper, the idea of using some structure to reduce the public key size can be strongly generalized. In particular, we will show that existing solutions are just very special cases of a wider framework, characterized by a large variety of options. This generalization comes with no increase in public key size, while on the other hand potentially allows to avoid DOOM and/or reaction attacks, or at the very least reduce their efficiency.

4 Reproducibility

We now introduce the main notions that we will use to provide a generalized approach to the design of structured codes.

Definition 3. *Consider a matrix $\mathbf{A} \in \mathbb{F}_q^{k \times n}$. Let $\mathcal{R}$ be the set of the rows of $\mathbf{A}$ and let $2^{\mathcal{R}}$ be its power set. We say that $\mathbf{A}$ is* reproducible *if $\mathbf{A}$ can be entirely described as $\mathcal{F}(\mathbf{a})$, where $\mathbf{a}$ is an element of $2^{\mathcal{R}}$, of cardinality $m < k$, called the* signature set*, and $\mathcal{F} = \{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \cdots, \boldsymbol{\sigma}_\ell\}$ is a family of linear transformations on elements of $\mathbb{F}_q^{m \times n}$.*

Definition 4. *Let $\mathcal{C}$ be a linear code over $\mathbb{F}_q$. If $\mathcal{C}$ can be described by a reproducible generator matrix $\mathbf{G} \in \mathbb{F}_q^{k \times n}$ and/or a reproducible parity-check matrix $\mathbf{H} \in \mathbb{F}_q^{r \times n}$, then we say that $\mathcal{C}$ is in* reproducible form*.*

Thus, a reproducible matrix is described just by its signature set and by its family of linear functions. Consequently, having the generator matrix (and/or the parity-check matrix) in reproducible form leads to a compact representation of the code. The condition on the reproducibility of a matrix can be relaxed, in order to take into account other structures that allow a compact representation.

Definition 5. *Let $\mathbf{A}_{i,j} \in \mathbb{F}_q^{k_{i,j} \times n_{i,j}}$ be reproducible matrices, each with its own dimensions, signature set $\mathbf{a}_{i,j} \in \mathbb{F}_q^{m_{i,j} \times n_{i,j}}$ and family of linear functions $\mathcal{F}_{i,j}$. Let $\mathbf{A}$ be a matrix obtained using as building blocks the matrices $\mathbf{A}_{i,j}$; then, we say that $\mathbf{A}$ is* quasi-reproducible*.*

Definition 6. *Let $\mathcal{C}$ be a linear code over $\mathbb{F}_q$. If $\mathcal{C}$ can be described by a quasi-reproducible generator matrix $\mathbf{G} \in \mathbb{F}_q^{k \times n}$ and/or a quasi-reproducible parity-check matrix $\mathbf{H} \in \mathbb{F}_q^{r \times n}$, then we say that $\mathcal{C}$ is in* quasi-reproducible form*.*

It is clear that, in order to describe a quasi-reproducible matrix, we just need the ensemble of the signature sets of its building blocks, together with the corresponding families of linear functions. Quasi-reproducibility generalizes the concept of reproducibility, since each reproducible code can be seen as a particular quasi-reproducible code, with a generator matrix described just by one signature. A particular type of quasi-reproducible codes is the one in which the blocks $\mathbf{A}_{i,j}$ are square matrices, defined by the same family $\mathcal{F}$.

At this point, we are ready to introduce a very important notion regarding the set of reproducible matrices obtained via a given family of transformations. Specifically, consider a family of linear functions $\mathcal{F} = \left\{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \cdots, \boldsymbol{\sigma}_{\frac{p}{m}-1}\right\}$, where each $\boldsymbol{\sigma}_i$ is a $p \times p$ matrix over $\mathbb{F}_q$. We denote by $\mathcal{M}_q^{\mathcal{F},m}$ the set of all reproducible matrices over $\mathbb{F}_q$ obtained via signatures of size $m \times p$ and $\mathcal{F}$, equipped with the usual operations of matrix sum and multiplication. Then the following results[1] hold.

Theorem 1. *The set $\mathcal{M}_q^{\mathcal{F},m}$ is an abelian group with respect to the sum.*

Proof. Showing that $\mathcal{M}_q^{\mathcal{F},m}$ is an additive abelian group is quite straightforward. In fact, the signature of the sum of two matrices corresponds to the sum of the original signatures. Commutativity and associativity follow from the element-wise sum between two matrices. The identity is given by the null signature (i.e., the signature made of all zeros), while the inverse of a matrix with signature $\mathbf{a}$ is the matrix with signature $-\mathbf{a}$. □

On the other hand, it is possible to show that the set, with respect to the multiplication, is a semigroup; in this case, the only requirements are closure and associativity. While associativity easily follows from the properties of the multiplication between two matrices, in order to guarantee closure, we must make an additional assumption.

Theorem 2. *$\mathcal{M}_q^{\mathcal{F},m}$ is a semigroup with respect to the multiplication if and only if for every matrix $\mathbf{M} \in \mathcal{M}_q^{\mathcal{F},m}$, we have*

$$\boldsymbol{\sigma}_i \mathbf{M} = \mathbf{M} \boldsymbol{\sigma}_i, \quad \forall i \in \mathbb{N}, 0 \leq i \leq \frac{p}{m} - 1.$$

Proof. We show that commutativity is necessary first. For what we discussed above, we only need to prove closure. Let $\mathbf{A}$ and $\mathbf{B}$ be two matrices of $\mathcal{M}_q^{\mathcal{F},m}$, with respective signatures $\mathbf{a}_0$, $\mathbf{b}_0$, that is

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_0\boldsymbol{\sigma}_1 \\ \vdots \\ \mathbf{a}_0\boldsymbol{\sigma}_{\frac{p}{m}-1} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_{\frac{p}{m}-1} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_0\boldsymbol{\sigma}_1 \\ \vdots \\ \mathbf{b}_0\boldsymbol{\sigma}_{\frac{p}{m}-1} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_0 \\ \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_{\frac{p}{m}-1} \end{bmatrix}.$$

Multiplying these two matrices we get

$$\mathbf{C} = \mathbf{A}\mathbf{B} = \begin{bmatrix} \mathbf{a}_0\mathbf{B} \\ \mathbf{a}_1\mathbf{B} \\ \vdots \\ \mathbf{a}_{\frac{p}{m}-1}\mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_0\mathbf{B} \\ \mathbf{a}_0\boldsymbol{\sigma}_1\mathbf{B} \\ \vdots \\ \mathbf{a}_0\boldsymbol{\sigma}_{\frac{p}{m}-1}\mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_0 \\ \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_{\frac{p}{m}-1} \end{bmatrix}. \tag{1}$$

[1] For simplicity we assume $\boldsymbol{\sigma}_0 = \mathbf{I}_p$, but this is not necessary and the results holds even if $\mathcal{F}$ does not contain the identity function.

Now by hypothesis

$$\mathbf{c}_i = \mathbf{a}_0\boldsymbol{\sigma}_i\mathbf{B} = \mathbf{a}_0\mathbf{B}\boldsymbol{\sigma}_i = \mathbf{c}_0\boldsymbol{\sigma}_i, \tag{2}$$

for all $i \leq \frac{p}{m} - 1$. It follows that $\mathbf{C}$ is reproducible and defined by $\mathcal{F}$.

Conversely, suppose $\mathcal{M}_q^{\mathcal{F},m}$ is a semigroup, and in particular that it is closed with respect to multiplication. Consider again two matrices $\mathbf{A}$ and $\mathbf{B}$ and their product, defined as in Equation 1. Since by hypothesis $\mathbf{C} \in \mathcal{M}_q^{\mathcal{F},m}$, and therefore reproducible, we have that $\mathbf{c}_i = \mathbf{c}_0\boldsymbol{\sigma}_i$ for all $i \leq \frac{p}{m} - 1$. It follows that

$$\mathbf{a}_0\boldsymbol{\sigma}_i\mathbf{B} = \mathbf{c}_i = \mathbf{c}_0\boldsymbol{\sigma}_i = \mathbf{a}_0\mathbf{B}\boldsymbol{\sigma}_i. \tag{3}$$

Now, since equation (3) holds in general for every signature $\mathbf{a}_0$, it must be that $\boldsymbol{\sigma}_i\mathbf{B} = \mathbf{B}\boldsymbol{\sigma}_i$, which concludes the proof. □

Finally, note that multiplication distributes over addition, as usual. This means that, if the conditions of Theorem 2 are satisfied, $\mathcal{M}_q^{\mathcal{F},m}$ verifies all the requisites of a mathematical *pseudo-ring*, i.e. a ring without multiplicative identity (also known as *rng*). We call this the *reproducible pseudo-ring* induced by $\mathcal{F}$ over $\mathbb{F}_q$.

4.1 Pseudo-rings induced by families of permutations

In the particular case of signatures made of just one row (i.e., $m = 1$) and the functions $\boldsymbol{\sigma}_i$ being permutations, we have a further result, which is described in Theorem 3. We point out that all the results we present in this section can be generalized, in order to consider the case $m > 1$, but we will not go into further details here. Since a $p \times p$ permutation corresponds to a matrix in which every row and column has weight equal to 1, it can equivalently be described as a bijection over $[0, p-1] \subset \mathbb{N}$. Given a permutation matrix $\boldsymbol{\sigma}_i$, we denote the corresponding bijection as $f_{\boldsymbol{\sigma}_i}$. If the element of $\boldsymbol{\sigma}_i$ in position (v, z) is equal to 1, then $f_{\boldsymbol{\sigma}_i}(v) = z$. The inverse of $f_{\boldsymbol{\sigma}_i}$ is denoted as $f_{\boldsymbol{\sigma}_i}^{-1}$, which is the bijection associated to the permutation matrix $\boldsymbol{\sigma}_i^{-1} = \boldsymbol{\sigma}_i^T$; if $f_{\boldsymbol{\sigma}_i}(v) = j$, then $f_{\boldsymbol{\sigma}_i}^{-1}(j) = v$. Let $\mathbf{a}$ and $\mathbf{a}'$ be two row vectors with entries $\{a_0, a_1, a_2, \ldots\}$ and $\{a'_0, a'_1, a'_2, \ldots\}$ respectively, such that $\mathbf{a}' = \mathbf{a}\boldsymbol{\sigma}_i$. Then, $a'_j = a_{f_{\boldsymbol{\sigma}_i}^{-1}(j)}$. If instead $\mathbf{a}'^T = \boldsymbol{\sigma}_i\mathbf{a}^T$, then $a'_j = a_{f_{\boldsymbol{\sigma}_i}(j)}$. We use $f_{\boldsymbol{\sigma}_i} \circ f_{\boldsymbol{\sigma}_j}$ to denote the bijection defined by the application of $f_{\boldsymbol{\sigma}_i}$ after $f_{\boldsymbol{\sigma}_j}$. In other words, $f_{\boldsymbol{\sigma}_i} \circ f_{\boldsymbol{\sigma}_j}$ corresponds to the permutation matrix $\boldsymbol{\sigma}_i\boldsymbol{\sigma}_j$, and $f_{\boldsymbol{\sigma}_i} \circ f_{\boldsymbol{\sigma}_j}(v) = f_{\boldsymbol{\sigma}_i}\left(f_{\boldsymbol{\sigma}_j}(v)\right)$. The identity $\mathbf{I}_p$ can be seen as the particular permutation that does not change the order of the elements; the corresponding bijection, which will be denoted as $f_{\mathbf{I}_p}$, is such that each element is mapped into itself (in other words, $f_{\mathbf{I}_p}(v) = v$).

Theorem 3. *Let $\mathcal{F} = \{\boldsymbol{\sigma}_0 = \mathbf{I}_p, \boldsymbol{\sigma}_1, \cdots, \boldsymbol{\sigma}_{p-1}\}$ be a family of linear transformations, with each $\boldsymbol{\sigma}_i$ being a permutation, and suppose that $\mathcal{F}$ induces the reproducible pseudo-ring $\mathcal{M}_q^{\mathcal{F},1}$ over $\mathbb{F}_q$. Then, the following relation must be satisfied*

$$\boldsymbol{\sigma}_j\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_{f_{\boldsymbol{\sigma}_i}(j)}, \quad \forall i, j \in \mathbb{N}, \;\; 0 \leq i \leq p-1, \;\; 0 \leq j \leq p-1.$$

Proof. Since $\mathcal{M}_q^{\mathcal{F},1}$ is a pseudo-ring, we know from Theorem 2 that, for every matrix $\mathbf{B} \in \mathcal{M}_q^{\mathcal{F},1}$ and every function $\boldsymbol{\sigma}_i \in \mathcal{F}$, it must be $\boldsymbol{\sigma}_i\mathbf{B} = \mathbf{B}\boldsymbol{\sigma}_i$. In particular, the left-hand term multiplication of $\boldsymbol{\sigma}_i$ by $\mathbf{B}$ corresponds to a row permutation, such that

$$\boldsymbol{\sigma}_i\mathbf{B} = \begin{bmatrix} \mathbf{b}_{f_{\boldsymbol{\sigma}_i}(0)} \\ \mathbf{b}_{f_{\boldsymbol{\sigma}_i}(1)} \\ \vdots \\ \mathbf{b}_{f_{\boldsymbol{\sigma}_i}(p-1)} = \end{bmatrix} = \begin{bmatrix} \mathbf{b}_0\boldsymbol{\sigma}_{f_{\boldsymbol{\sigma}_i}(0)} \\ \mathbf{b}_0\boldsymbol{\sigma}_{f_{\boldsymbol{\sigma}_i}(1)} \\ \vdots \\ \mathbf{b}_0\boldsymbol{\sigma}_{f_{\boldsymbol{\sigma}_i}(p-1)} \end{bmatrix}, \tag{4}$$

where $\mathbf{b}_i$ denotes the i-th row of $\mathbf{B}$. The product $\mathbf{B}\boldsymbol{\sigma}_i$ instead defines a column permutation of $\mathbf{B}$, and can be expressed as

$$\mathbf{B}\sigma_i = \begin{bmatrix} \mathbf{b}_0\boldsymbol{\sigma}_0 \\ \mathbf{b}_0\boldsymbol{\sigma}_1 \\ \vdots \\ \mathbf{b}_0\boldsymbol{\sigma}_{p-1} \end{bmatrix} \boldsymbol{\sigma}_i = \begin{bmatrix} \mathbf{b}_0\boldsymbol{\sigma}_0\boldsymbol{\sigma}_i \\ \mathbf{b}_0\boldsymbol{\sigma}_1\boldsymbol{\sigma}_i \\ \vdots \\ \mathbf{b}_0\boldsymbol{\sigma}_{p-1}\boldsymbol{\sigma}_i \end{bmatrix}. \tag{5}$$

Putting together equations (4) and (5), we obtain

$$\boldsymbol{\sigma}_j\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_{f_{\boldsymbol{\sigma}_i}(j)}, \tag{6}$$

which must be satisfied for every pair of indexes (i, j). □

Starting from the result of Theorem 3, we can easily derive some other properties that $\mathcal{F}$ must satisfy.

Corollary 1. *Let $\mathcal{F}$ be a family of permutations satisfying Theorem 3. Then, $\mathcal{F}$ has the following properties*

(a) $f_{\boldsymbol{\sigma}_i}(0) = i, \ \forall i;$
(b) $\forall i \ \exists j$ *s.t.* $f_{\boldsymbol{\sigma}_i} \circ f_{\boldsymbol{\sigma}_j} = f_{\mathbf{I}_p}.$

Proof. According to Theorem 3, we have

$$\boldsymbol{\sigma}_{f_{\boldsymbol{\sigma}_i}(0)} = \boldsymbol{\sigma}_0\boldsymbol{\sigma}_i = \mathbf{I}_p\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_i, \tag{7}$$

which can be satisfied only if $f_{\boldsymbol{\sigma}_i}(0) = i$, and this proves property (a).
Since each $f_{\boldsymbol{\sigma}_i}$ is a bijection of the integers in $[0, p-1]$, we know that, for a fixed value of i, there is a value $j \in [0, p-1]$ such that $f_{\boldsymbol{\sigma}_i}(j) = 0$. Then, we have

$$\boldsymbol{\sigma}_j\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_{f_{\sigma_i}(j)} = \boldsymbol{\sigma}_0 = \mathbf{I}_p. \tag{8}$$

In other words, the bijections corresponding to $f_{\boldsymbol{\sigma}_i}$ and $f_{\boldsymbol{\sigma}_j}$ are one the inverse of the other, and this proves property (b). □

Corollary 2. *Let $\mathcal{F}$ be a family of permutations satisfying Theorem 3. Then, $\mathcal{M}_q^{\mathcal{F},1}$ is a ring, which we call, by analogy,* reproducible ring *induced by $\mathcal{F}$.*

Proof. Let us show that $\mathcal{M}_q^{\mathcal{F},1}$ contains the multiplicative identity, i.e., the $p \times p$ identity matrix. Because of Corollary 1, $\mathcal{F}$ is formed by $p \times p$ permutations such that $f_{\boldsymbol{\sigma}_i}(0) = i, \forall i$. If we generate the element of $\mathcal{M}_q^{\mathcal{F},1}$ corresponding to the signature $\mathbf{u} = [1, 0, \cdots, 0]$, we easily obtain the $p \times p$ identity matrix $\mathbf{I}_p$. □

Theorem 4. *Let $\mathcal{F}$ be a family of permutations satisfying Theorem 3. Then, $\mathcal{M}_q^{\mathcal{F},1}$ is a reproducible ring and the invertible elements of $\mathcal{M}_q^{\mathcal{F},1}$ constitute a multiplicative group over $\mathcal{M}_q^{\mathcal{F},1}$.*

Proof. Based on Corollary 2, $\mathcal{M}_q^{\mathcal{F},1}$ is a reproducible ring provided with multiplicative identity. Now, we need to prove that any non-singular matrix in $\mathcal{M}_q^{\mathcal{F},1}$ admits inverse in $\mathcal{M}_q^{\mathcal{F},1}$. Let us consider a matrix $\mathbf{A} \in \mathcal{M}_q^{\mathcal{F},m}$, with signature $\mathbf{a}$, and let $\mathbf{B}$ be its inverse. Since $\mathbf{AB} = \mathbf{I}_p$, we have

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a} \\ \mathbf{a}\boldsymbol{\sigma}_1 \\ \vdots \\ \mathbf{a}\boldsymbol{\sigma}_{p-1} \end{bmatrix} \mathbf{B} = \mathbf{I}_p = \begin{bmatrix} \mathbf{u} \\ \mathbf{u}\boldsymbol{\sigma}_1 \\ \vdots \\ \mathbf{u}\boldsymbol{\sigma}_{p-1} \end{bmatrix},$$

with $\mathbf{u} = [1, 0, \cdots, 0]$ as in Corollary 2. Then we have $\mathbf{a}\boldsymbol{\sigma}_i\mathbf{B} = \mathbf{u}\boldsymbol{\sigma}_i$. For $i = 0$, we have $\mathbf{u} = \mathbf{aB}$. Hence, for whichever value i, we get

$$\mathbf{a}\boldsymbol{\sigma}_i\mathbf{B} = \mathbf{u}\boldsymbol{\sigma}_i = \mathbf{aB}\boldsymbol{\sigma}_i,$$

which can be satisfied for whichever $\mathbf{a}$ only if $\boldsymbol{\sigma}_i$ and $\mathbf{B}$ commute. Because of Theorem 2, this means that $\mathbf{B} \in \mathcal{M}_q^{\mathcal{F},1}$. □

Sum and multiplication are not the only matrix operations we consider. In Theorem 5 we analyze how transposition acts on the matrices belonging to a reproducible pseudo-ring $\mathcal{M}_q^{\mathcal{F},1}$.

Theorem 5. *Let $\mathcal{M}_q^{\mathcal{F},1}$ be a reproducible pseudo-ring; if*

$$f_{\boldsymbol{\sigma}_j}^{-1}(i) = f_{\boldsymbol{\sigma}_v}^{-1}(0), \;\; v = f_{\boldsymbol{\sigma}_i}^{-1}(j), \;\; \forall i, j \;\; s.t. \;\; 0 \le i \le p-1, 0 \le j \le p-1$$

then $\mathcal{M}_q^{\mathcal{F},1}$ is closed under the transposition operation.

Proof. Let $\mathbf{A} \in \mathcal{M}_q^{\mathcal{F},1}$, with signature $\mathbf{a}$, and denote as $\mathbf{B} = \mathbf{A}^T$ its transpose. The i-th row of $\mathbf{B}$ corresponds to the i-th column of $\mathbf{A}$. In particular, the i-th column of $\mathbf{A}$ is defined as

$$\begin{bmatrix} a_i \\ a_{f_{\boldsymbol{\sigma}_1}^{-1}(i)} \\ a_{f_{\boldsymbol{\sigma}_2}^{-1}(i)} \\ \vdots \\ a_{f_{\boldsymbol{\sigma}_{p-1}}^{-1}(i)} \end{bmatrix}.$$

Because $\mathbf{B}$ is the transpose of $\mathbf{A}$, the i-th row of $\mathbf{B}$ corresponds to the i-th column of $\mathbf{A}$. Let us denote as $\mathbf{b}_0$ the first row of $\mathbf{B}$, that is

$$\mathbf{b}_0 = [a_0, a_{f_{\boldsymbol{\sigma}_1}^{-1}(0)}, \cdots, a_{f_{\boldsymbol{\sigma}_{p-1}}^{-1}(0)}] = [a_{f_{\boldsymbol{\sigma}_0}^{-1}(0)}, a_{f_{\boldsymbol{\sigma}_1}^{-1}(0)}, \cdots, a_{f_{\boldsymbol{\sigma}_{p-1}}^{-1}(0)}]. \tag{9}$$

Let us consider the i-th row of $\mathbf{B}$, and denote it as $\mathbf{b}_i$; if transposition has closure in $\mathcal{M}_q^{\mathcal{F},1}$, then it must be

$$\mathbf{b}_i = [a_i, a_{f_{\boldsymbol{\sigma}_1}^{-1}(i)}, \cdots, a_{f_{\boldsymbol{\sigma}_{p-1}}^{-1}(i)}] = [a_{f_{\boldsymbol{\sigma}_0}^{-1}(i)}, a_{f_{\boldsymbol{\sigma}_1}^{-1}(i)}, \cdots, a_{f_{\boldsymbol{\sigma}_{p-1}}^{-1}(i)}] = \mathbf{b}_0\boldsymbol{\sigma}_i. \tag{10}$$

Now suppose that $f_{\boldsymbol{\sigma}_i}(v) = j$; then, the j-th entry of $\mathbf{b}_i$ corresponds to the v-th entry of $\mathbf{b}_0$, that is $a_{f_{\boldsymbol{\sigma}_v}^{-1}(0)}$. In other words, we have $b_{i,j} = a_z$, with

$$z = f_{\boldsymbol{\sigma}_v}^{-1}(0), \;\; v = f_{\boldsymbol{\sigma}_i}^{-1}(j). \tag{11}$$

In order to satisfy eq. (10), a_z must be equal to the j-th entry of the i-th column of $\mathbf{A}$, that is $a_{f_{\boldsymbol{\sigma}_j}^{-1}(i)}$. Then, it must be $f_{\boldsymbol{\sigma}_j}^{-1}(i) = z$, that is

$$f_{\boldsymbol{\sigma}_j}^{-1}(i) = f_{\boldsymbol{\sigma}_v}^{-1}(0), \;\; v = f_{\boldsymbol{\sigma}_i}^{-1}(j), \tag{12}$$

which concludes the proof. □

Depending on the properties stated in the previous theorems, the family $\mathcal{F}$ might induce different algebraic structures over $\mathbb{F}_q^{p\times p}$. In particular, let us consider the case of $\mathcal{F}$ corresponding to $\mathcal{M}_q^{\mathcal{F},1}$ satisfying both Theorems 4 and 5. Let $\mathbf{A}$ be a square matrix whose elements are picked from $\mathcal{M}_q^{\mathcal{F},1}$. By definition, we have $\mathbf{A}^{-1} = det(\mathbf{A})^{-1}adj(\mathbf{A})$, where $det(\mathbf{A})$ is the determinant of $\mathbf{A}$ and $adj(\mathbf{A})$ is the adjugate of $\mathbf{A}$. Computing $det(\mathbf{A})$ involves only sums and multiplications: this means that $det(\mathbf{A}) \in \mathcal{M}_q^{\mathcal{F},1}$; because of Theorem 4, $det(\mathbf{A})^{-1} \in \mathcal{M}_q^{\mathcal{F},1}$. Computing $adj(\mathbf{A})$ involves sums, multiplications and transpositions: because of Theorem 5, we have that the entries of $adj(\mathbf{A})$ are again elements of $\mathcal{M}_q^{\mathcal{F},1}$. This means that $\mathbf{A}^{-1}$ is a matrix whose elements belong to $\mathcal{M}_q^{\mathcal{F},1}$, and so has the same quasi-reproducible structure of $\mathbf{A}$.

4.2 Known examples of reproducible pseudo-rings

In Section 4.1 we have described some properties that a family of permutations $\mathcal{F}$ must have to guarantee that it induces algebraic structures on $\mathbb{F}_q^{p\times p}$. Well-known cases of such objects, with common use in cryptography, are circulant matrices and dyadic matrices.

Circulant Matrices As we have seen before, a circulant matrix is a $p\times p$ matrix for which each row is obtained as the cyclic shift of the previous one. In particular, a circulant matrix can be seen as a square reproducible matrix, whose signature corresponds to the first row and the functions $\boldsymbol{\sigma}_i$ defining $\mathcal{F}$ correspond to $\boldsymbol{\pi}^i$, where $\boldsymbol{\pi}$ is the unitary circulant permutation matrix with entries

$$\pi_{l,j} = \begin{cases} 1 & \text{if } l+1 \equiv j \mod p \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Basically, the bijection representing $\boldsymbol{\pi}$ is defined as

$$f_{\boldsymbol{\pi}}(v) = v + 1 \mod p. \tag{14}$$

It can be easily shown that

$$f_{\boldsymbol{\sigma}_i}(v) = f_{\boldsymbol{\pi}^i}(v) = \underbrace{f_{\boldsymbol{\pi}} \circ f_{\boldsymbol{\pi}} \cdots \circ f_{\boldsymbol{\pi}}}_{i \text{ times}}(v) = v + i \mod p, \tag{15}$$

which leads to $\boldsymbol{\pi}^p = \mathbf{I}_p$ and $\boldsymbol{\pi}^i\boldsymbol{\pi}^j = \boldsymbol{\pi}^{i+j \mod p}$. Since permutation matrices are orthogonal, their inverses correspond to their transposes, and thus $(\boldsymbol{\pi}^i)^T = \boldsymbol{\pi}^{p-i}$. With these properties, we have

$$\boldsymbol{\sigma}_i\boldsymbol{\sigma}_j = \boldsymbol{\pi}^{i+j \mod p} = \boldsymbol{\sigma}_{i+j \mod p}, \tag{16}$$

which is compliant with Theorem 3, since $f_{\boldsymbol{\sigma}_i}(j) = i+j \mod p$. With some simple computations, it can be easily shown that circulant matrices satisfy Theorem 5 and that the multiplication between two circulant matrices is commutative.

Dyadic Matrices A *dyadic* matrix is a $p \times p$ matrix, with p being a power of 2, whose signature is again its first row. The rows of a dyadic matrix are obtained by permuting the elements of the signature, such that the element at position (i, j) is the one in the signature at position $i \oplus j$, where $\oplus$ denotes the bitwise XOR between i and j. Then, a dyadic matrix can be described in terms of reproducible matrices, for which each function $\boldsymbol{\sigma}_i$ is the dyadic matrix whose signature has all-zero entries, except that at position i. This means that $\boldsymbol{\sigma}_i$ can be described by the following bijection

$$f_{\boldsymbol{\sigma}_i}(v) = v \oplus i \mod p. \tag{17}$$

If we combine two transformations, we obtain

$$f_{\boldsymbol{\sigma}_i} \circ f_{\boldsymbol{\sigma}_j}(v) = (v \oplus j) \oplus i = v \oplus (i \oplus j) = f_{\boldsymbol{\sigma}_{i\oplus j}}(v). \tag{18}$$

Since $f_{\boldsymbol{\sigma}_i}(j) = i \oplus j$, this proves that the family of dyadic matrices is compliant with Theorem 3. It can be straightforwardly proven that dyadic matrices are symmetric (and so, satisfy Theorem 5), and that the multiplication between two dyadic matrices is commutative.

Circulant and dyadic matrices are just two particular cases of reproducible pseudo-rings, and can obviously be further generalized by considering signatures that are composed by more than one row. In addition, several more constructions can be obtained. For instance, for every permutation matrix $\boldsymbol{\psi}$ and every reproducible pseudo-ring $\mathcal{M}_q^{\mathcal{F},m}$, induced by $\mathcal{F} = \left\{\sigma_0 = \mathbf{I}_p, \sigma_1, \cdots, \sigma_{\frac{p}{m}-1}\right\}$, we can obtain a new pseudo-ring as

$$\mathcal{M}_q^{\mathcal{F}',m} = \left\{\mathbf{M}' \middle| \, \mathbf{M}' = \boldsymbol{\psi}\mathbf{M}\boldsymbol{\psi}^T, \;\; \mathbf{M} \in \mathcal{M}_q^{\mathcal{F},m}\right\}. \tag{19}$$

The corresponding family of transformations is $\mathcal{F}' = \left\{\sigma'_0, \sigma'_1, \cdots, \sigma'_{\frac{p}{m}-1}\right\}$, with $\boldsymbol{\sigma}'_i = \boldsymbol{\sigma}_{f_{\boldsymbol{\psi}}(i)}\boldsymbol{\psi}^T$. Proving that $\mathcal{F}'$ actually induces a pseudo-ring is quite simple; indeed, for any two matrices $\mathbf{A} = \boldsymbol{\psi}\mathbf{M}_A\boldsymbol{\psi}^T$ and $\mathbf{B} = \boldsymbol{\psi}\mathbf{M}_B\boldsymbol{\psi}^T$, with $\mathbf{M}_A, \mathbf{M}_B \in \mathcal{M}_{\mathcal{F},m}$, we have

$$\mathbf{A} + \mathbf{B} = \boldsymbol{\psi}\mathbf{M}_A\boldsymbol{\psi}^T + \boldsymbol{\psi}M_B\boldsymbol{\psi}^T = \boldsymbol{\psi}(\mathbf{M}_A + \mathbf{M}_B)\boldsymbol{\psi}^T, \tag{20}$$

$$\mathbf{AB} = \boldsymbol{\psi}\mathbf{M}_A\boldsymbol{\psi}^T\boldsymbol{\psi}\mathbf{M}_B\boldsymbol{\psi}^T = \boldsymbol{\psi}\mathbf{M}_A\mathbf{M}_B\boldsymbol{\psi}^T, \tag{21}$$

which return matrices belonging to $\mathcal{M}_q^{\mathcal{F}',m}$, since $\mathbf{M}_A + \mathbf{M}_B \in \mathcal{M}_q^{\mathcal{F},m}$ and $\mathbf{M}_A\mathbf{M}_B \in \mathcal{M}_q^{\mathcal{F},m}$. In addition, if multiplication is commutative in $\mathcal{M}_q^{\mathcal{F},m}$, then it will be commutative in $\mathcal{M}_q^{\mathcal{F}',m}$ too. To prove this fact, let us consider two matrices $\mathbf{M}_A, \mathbf{M}_B \in \mathcal{M}_q^{\mathcal{F},m}$, such that $\mathbf{M}_A\mathbf{M}_B = \mathbf{M}_B\mathbf{M}_A$. Then, for $\mathbf{A} = \boldsymbol{\psi}\mathbf{M}_A\boldsymbol{\psi}^T$ and $\mathbf{B} = \boldsymbol{\psi}\mathbf{M}_B\boldsymbol{\psi}^T$, we have

$$\begin{aligned}\mathbf{AB} &= \boldsymbol{\psi}\mathbf{M}_A\boldsymbol{\psi}^T\boldsymbol{\psi}\mathbf{M}_B\boldsymbol{\psi}^T = \boldsymbol{\psi}\mathbf{M}_A\mathbf{M}_B\boldsymbol{\psi}^T = \\ &= \boldsymbol{\psi}\mathbf{M}_B\mathbf{M}_A\boldsymbol{\psi}^T = \boldsymbol{\psi}\mathbf{M}_B\boldsymbol{\psi}^T\boldsymbol{\psi}\mathbf{M}_A\boldsymbol{\psi}^T = \mathbf{BA}.\end{aligned}$$

It is easy to prove that, if $\mathcal{M}_q^{\mathcal{F},m}$ is closed under transposition, $\mathcal{M}_q^{\mathcal{F}',m}$ is too.

5 Codes in reproducible form

In the previous section we have described the properties that a family of functions $\mathcal{F}$ must have in order to generate reproducible matrices. This opens a wide range of possibilities for obtaining codes with compact representations. In fact, reproducible pseudo-rings allow to design codes that can be described in a very compact manner. Codes of this type are of interest in code-based cryptography, where small public keys are important.

In this section we describe how to design codes with reproducible representations, and the properties that characterize them. In particular we study how to achieve a reproducible representation for the parity-check matrix $\mathbf{H}$ starting from a generator matrix $\mathbf{G}$ in reproducible form. In addition, we provide intuitive methods to obtain random-looking codes in reproducible form, starting from their parity-check matrix. The following theorem states some properties about the parity-check matrix that are sufficient (but not necessary) conditions for having a code with $\mathbf{G}$ and $\mathbf{H}$ in reproducible form.

Theorem 6. *Let $\mathbf{G} \in \mathbb{F}_q^{k\times n}$ be a reproducible matrix, with signature $\mathbf{g}_0 \in \mathbb{F}_q^{m\times n}$ (hence, m is among the factors of k) and family $\mathcal{F} = \left\{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \cdots, \boldsymbol{\sigma}_{\frac{k}{m}-1}\right\}$. For simplicity, we suppose that $\boldsymbol{\sigma}_0 = \mathbf{I}_n$. Let $\mathbf{H} \in \mathbb{F}_q^{r\times n}$, such that $\mathbf{g}_0\mathbf{H}^t = \mathbf{0}_{m\times r}$. We denote as $\mathbf{h}_j$ the subset of rows of $\mathbf{H}$ at positions $\{js, js+1, \cdots, (j+1)s-1\}$. If we can define a function $f(x_0, x_1) : [0, \frac{k}{m}-1] \times [0, \frac{r}{s}-1] \subset \mathbb{N}^2 \to [0, \frac{r}{s}-1] \subset \mathbb{N}$, such that*

$$\mathbf{h}_j\boldsymbol{\sigma}_i^T = \mathbf{h}_{f(i,j)}, \ \ \forall i, j \in \mathbb{N}, \ \ 0 \le i \le \frac{k}{m} - 1, \ \ 0 \le j \le \frac{r}{s} - 1, \tag{22}$$

then $\mathbf{G}$ and $\mathbf{H}^T$ are orthogonal, i.e. $\mathbf{GH}^T = \mathbf{0}_{k\times r}$.

Proof. Since the generator matrix $\mathbf{G}$ is reproducible, with signature $\mathbf{g}_0$, we have

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_{\frac{k}{m}-1} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_0\boldsymbol{\sigma}_1 \\ \vdots \\ \mathbf{g}_0\boldsymbol{\sigma}_{\frac{k}{m}-1} \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{h}_0 \\ \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_{\frac{r}{s}-1} \end{bmatrix}. \tag{23}$$

In order for $\mathbf{G}$ to be a valid generator matrix, it must be $\mathbf{G}\mathbf{H}^T = \mathbf{0}_{k\times r}$, that is

$$\mathbf{g}_i\mathbf{h}_j^T = \mathbf{g}_0\boldsymbol{\sigma}_i\mathbf{h}_j^T = \mathbf{0}_{m\times s}, \;\; \forall i,j \in \mathbb{N} \text{ s.t. } 0 \leq i \leq \frac{k}{m}-1, \;\; 0 \leq j \leq \frac{r}{s}-1. \tag{24}$$

By hypothesis, $\mathbf{g}_0$ is an $m \times n$ matrix such that $\mathbf{g}_0\mathbf{H}^T = \mathbf{0}_{m\times r}$, which means

$$\mathbf{g}_0\mathbf{h}_j^T = \mathbf{0}_{m\times s}, \;\; \forall j \in \mathbb{N} \text{ s.t. } 0 \leq j \leq \frac{r}{s}-1. \tag{25}$$

Consider now the product $\mathbf{g}_i\mathbf{h}_j^T = \mathbf{g}_0\boldsymbol{\sigma}_i\mathbf{h}_j^T$, for $i \geq 1$. If we can define a function $f(x_0, x_1) : [0, \frac{k}{m}-1] \times [0, \frac{r}{s}-1] \subset \mathbb{N}^2 \rightarrow [0, \frac{r}{s}-1] \subset \mathbb{N}$ with the aforementioned property described by (22), then for all couples of indexes i, j we have

$$\boldsymbol{\sigma}_i\mathbf{h}_j^T = \mathbf{h}_{f(i,j)}^T, \tag{26}$$

and (24) is surely satisfied, since

$$\mathbf{g}_i\mathbf{h}_j^T = \mathbf{g}_0\boldsymbol{\sigma}_i\mathbf{h}_j^T = \mathbf{g}_0\mathbf{h}_{f(i,j)}^T = \mathbf{0}_{m\times s}, \tag{27}$$

where $\mathbf{g}_0\mathbf{h}_{f(i,j)}^T = \mathbf{0}_{m\times s}$ because of (25). □

For $\mathbf{G}$ and $\mathbf{H}$ to be, respectively, generator and parity-check matrix of a code $\mathcal{C}$, some conditions have to be verified, given in Corollary 3 below.

Corollary 3. *Let $\mathbf{G} \in \mathbb{F}_q^{k\times n}$ be a reproducible matrix, with signature $\mathbf{g}_0 \in \mathbb{F}_q^{m\times n}$ (hence, m is among the factors of k) and family $\mathcal{F} = \left\{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \cdots, \boldsymbol{\sigma}_{\frac{k}{m}-1}\right\}$. Let $\mathbf{H} \in \mathbb{F}_q^{r\times n}$ be a matrix such that $\mathbf{G}\mathbf{H}^T = \mathbf{0}_{k\times n}$, and suppose that it satisfies the hypothesis of Theorem 6. For $\mathbf{H}$ and $\mathbf{G}$ to be, respectively, the parity-check and generator matrices of a code $\mathcal{C}$ with length n, dimension k and redundancy r, the following conditions are necessary*

(a) $\mathcal{F}$ contains $\frac{k}{m}$ distinct linear transformations;
(b) $\frac{k}{m} \leq \frac{r}{s}$;
(c) for any three integers $i \in [0, \frac{k}{m}-1]$ and $j', j'' \in [0, \frac{r}{s}-1]$, with $j' \neq j''$, it must be $f(i, j') \neq f(i, j'')$.

Proof. We want the reproducible $k \times n$ matrix $\mathbf{G}$ to be the generator matrix of a code with dimension k: then, $\mathbf{G}$ must have rank equal to k. If $\mathcal{F}$ contains two transformations $\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_j$, with $i \neq j$, then the rows of $\mathbf{G}$ obtained as $\mathbf{g}_0\boldsymbol{\sigma}_i$ are identical to the ones obtained as $\mathbf{g}_0\boldsymbol{\sigma}_j$. If $\mathbf{G}$ has some identical rows, then its rank cannot be maximum, and this proves condition (a). It is straightforward to show that this condition can also be expressed as follows: there cannot exists three integers $i', i'' \in [0, \frac{k}{m} - 1]$, with $i' \neq i''$, and $j \in [0, \frac{r}{s} - 1]$, such that $f(i', j) = f(i'', j)$. Indeed, if we can determine such integers, then

$$\mathbf{h}_j \boldsymbol{\sigma}_{i'}^T = \mathbf{h}_{f(i',j)} = \mathbf{h}_{f(i'',j)} = \mathbf{h}_j \boldsymbol{\sigma}_{i''}^T,$$

which results in $\boldsymbol{\sigma}_{i'} = \boldsymbol{\sigma}_{i''}$.

We can then easily prove condition (b). Indeed, fix an integer $j \in [0, \frac{r}{s} - 1]$ and consider, for all $i \in [0, \frac{k}{m} - 1]$, all the images $f(i, j)$: because of condition (a), these images must be distinct. However, the dimension of the codomain of $f(i, j)$ is equal to $\frac{r}{s}$: if $\frac{k}{m} > \frac{r}{s}$, then (a) cannot be satisfied. This proves (b).
If $\mathbf{H}$ is the parity-check matrix a code with redundancy r, then it must have rank equal to r. If we suppose that there exists three integers $i \in [0, \frac{k}{m} - 1]$, $j', j'' \in [0, \frac{r}{s} - 1]$, with $j' \neq j''$, such that $f(i, j') = f(i, j'')$ then, because of Theorem 6, we also have $\mathbf{h}_{j'} \boldsymbol{\sigma}_i^T = \mathbf{h}_{j''} \boldsymbol{\sigma}_i^T$, which implies $\mathbf{h}_{j'} = \mathbf{h}_{j''}$. If $\mathbf{H}$ has some identical rows, then its rank must be $< r$, and this proves condition (c). □

Theorem 6 and Corollary 3 allow obtaining a code in reproducible form in a very simple way. Given a family of transformations $\mathcal{F}$, a matrix $\mathbf{H}$ with the characteristics required by the theorem can be found. Then, for the code $\mathcal{C}$ having $\mathbf{H}$ as parity-check matrix, a variety of reproducible generator matrices can be found. Indeed, let $\mathbf{G}$ be a generator matrix for $\mathcal{C}$: by definition, since $\mathbf{G}\mathbf{H}^T = \mathbf{0}_{k \times r}$, we know that whichever subset $\mathbf{g}_0$ formed by m rows of $\mathbf{G}$ is such that $\mathbf{g}_0 \mathbf{H}^T = \mathbf{0}_{m \times r}$. Then, $\mathbf{g}_0$ is a valid signature for a reproducible generator matrix, defined by the family $\mathcal{F}$. On condition that both $\mathbf{H}$ and $\mathbf{G}$ have full rank, then they can be used to represent the code $\mathcal{C}$, with length n, dimension k and redundancy r.

In some cases, a quasi-reproducible code can be seen as a particular case of a reproducible code (and viceversa). Let us consider a code $\mathcal{C}$ with length $n = n_0 p$, dimension $k = p$ and codimension $r = (n_0 - 1)p$, for some integer $n_0 \in \mathbb{N}$. Let us suppose that $\mathbf{G}$ is obtained as a row of n_0 blocks with dimensions $p \times p$, that is

$$\mathbf{G} = [\mathbf{G}_0 | \mathbf{G}_1 | \cdots | \mathbf{G}_{n_0 - 1}]. \tag{28}$$

This form of the generator matrix is commonly used in sparse-matrix code-based cryptosystems [5,30]. Suppose that $\mathbf{G}$ in (28) is in quasi-reproducible form, i.e., each $\mathbf{G}_i$ is an element of the reproducible pseudo-ring $\mathcal{M}_q^{\mathcal{F}_i, m_i}$, and has signature V_i. If the signatures have all the same number of rows (that is, $m_i = m$), then such a $\mathbf{G}$ can be characterized as a particular reproducible matrix. Let us write the i-th family of transformations as $\mathcal{F}_i = \left\{ \boldsymbol{\sigma}_0^{(i)}, \boldsymbol{\sigma}_1^{(i)}, \cdots, \boldsymbol{\sigma}_{\frac{p}{m}-1}^{(i)} \right\}$ and define an overall family of transformations $\mathcal{F} = \left\{ \boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \cdots, \boldsymbol{\sigma}_{\frac{p}{m}-1} \right\}$, such that

$$\boldsymbol{\sigma}_i = \begin{bmatrix} \boldsymbol{\sigma}_i^{(0)} & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \cdots & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \boldsymbol{\sigma}_i^{(1)} & \mathbf{0}_{p\times p} & \cdots & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \boldsymbol{\sigma}_i^{(2)} & \cdots & \mathbf{0}_{p\times p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \cdots & \boldsymbol{\sigma}_i^{(n_0-1)} \end{bmatrix}. \tag{29}$$

Then, it is easy to see that a matrix in the form (28) can be described as a reproducible matrix obtained through $\mathcal{F}$ in (29), with signature

$$\mathbf{g}_0 = \left[\mathbf{g}_0^{(0)}\middle|\, \mathbf{g}_0^{(1)}\middle| \cdots \middle|\mathbf{g}_0^{(n_0-1)}\right]. \tag{30}$$

5.1 Reproducible Codes from Householder Matrices

A *Householder matrix* [25] is a matrix that is at the same time orthogonal and symmetric. Let us consider a set of distinct Householder matrices $\boldsymbol{\psi}_0, \cdots, \boldsymbol{\psi}_{v-1}$. We have that, for all $j = 0, \ldots, v-1$, it must be $\boldsymbol{\psi}_j^{-1} = \boldsymbol{\psi}_j^T = \boldsymbol{\psi}_j$. In order to fulfill the conditions of Theorem 6, these matrices must form a commutative group, that is

$$\boldsymbol{\psi}_i\boldsymbol{\psi}_j = \boldsymbol{\psi}_j\boldsymbol{\psi}_i,\ 0 \le i, j \le v-1. \tag{31}$$

Let us consider two sets containing all the 2^v distinct binary v-tuples, i.e.

$$\begin{gathered} \left\{\mathbf{a}^{(i)}\middle|\ 0 \le i \le 2^v-1,\ \mathbf{a}^{(i)} \in \mathbb{F}_2^v,\ \text{ s.t. } \mathbf{a}^{(i)} \ne \mathbf{a}^{(j)},\ \ \forall i \ne j\right\}, \\ \left\{\mathbf{b}^{(i)}\middle|\ 0 \le i \le 2^v-1,\ \mathbf{b}^{(i)} \in \mathbb{F}_2^v,\ \text{ s.t. } \mathbf{b}^{(i)} \ne \mathbf{b}^{(j)},\ \ \forall i \ne j\right\}. \end{gathered} \tag{32}$$

For the sake of simplicity, let us fix $\mathbf{a}^{(0)} = \mathbf{0}_{1\times v}$. It is clear that these two sets are identical, except for the order of their elements. We can now define a family of transformations $\mathcal{F}$, containing 2^v linear functions $\boldsymbol{\sigma}_i$, defined as

$$\boldsymbol{\sigma}_i = \prod_{l=0}^{v-1} \boldsymbol{\psi}_l^{a_l^{(i)}}, \tag{33}$$

where $a_l^{(i)}$ is the l-th entry of $\mathbf{a}^{(i)}$. Since we are considering Householder matrices with the property (31), it is easy to verify that $\boldsymbol{\sigma}_i^2 = \mathbf{I}_n$, and it follows that each function is an involution.

The family $\mathcal{F}$ can be used to define a reproducible generator matrix $\mathbf{G}$ of a code $\mathcal{C}$; a parity-check matrix for $\mathcal{C}$ can be the reproducible matrix $\mathbf{H}$, with signature $\mathbf{h}_0\mathbb{F}_q^{s\times n}$, whose rows are obtained as

$$\mathbf{h}_j = \mathbf{h}_0 \left(\prod_{l=0}^{v-1} \boldsymbol{\psi}_l^{b_l^{(j)}}\right)^T. \tag{34}$$

If $\mathbf{H}$ has full rank, the corresponding code has redundancy $r = s2^v$, and

$$\begin{aligned} \mathbf{h}_j\boldsymbol{\sigma}_i^T = \mathbf{h}_j\left(\prod_{l=0}^{v-1}\boldsymbol{\psi}_l^{a^{(i)}}\right)^T &= \mathbf{h}_0\left(\prod_{l=0}^{v-1}\boldsymbol{\psi}_l^{b^{(j)}}\right)^T\left(\prod_{l=0}^{v-1}\boldsymbol{\psi}_l^{a_l^{(i)}}\right)^T = \\ &= \mathbf{h}_0\left(\prod_{l=0}^{v-1}\boldsymbol{\psi}_l^{a_l^{(j)}\oplus b_l^{(i)}}\right)^T = \mathbf{h}_{f(i,j)}, \end{aligned}$$

where $\oplus$ denotes the modulo 2 sum and

$$f(i,j) = u, \;\text{ s.t. }\; \mathbf{b}^{(u)} = \mathbf{a}^{(i)} \oplus \mathbf{b}^{(j)}. \tag{35}$$

It is straightforward to show that such a function satisfies the properties required by Theorem 6 and Corollary 3. The corresponding code has length n, dimension $k = m2^v$ and redundancy $r = s2^v$, thus the code rate corresponds to $\frac{m}{m+s}$. In addition, we point out that it might be possible to tune the code parameters, by selecting only proper subsets of all the binary v-tuples, in order to form the rows of both $\mathbf{G}$ and $\mathbf{H}$.

5.2 Reproducible codes from powers of a single function

In this section we present another construction of reproducible codes satisfying Theorem 6. Let us consider an $n \times n$ matrix $\boldsymbol{\pi}$ such that $\boldsymbol{\pi}^b = \mathbf{I}_n$, for some integer b. Let v be a divisor of b; obviously, if b is a prime, then $v = 1$. Then, we can use $\boldsymbol{\pi}$ to build a family $\mathcal{F}$ of $\frac{k}{m} \leq \frac{b}{v}$ linear transformations, where k is the desired code dimension and m is the number of rows in a signature. Indeed, the functions in $\mathcal{F}$ can be defined as $\boldsymbol{\sigma}_i = \boldsymbol{\pi}^{vz_i}$, where the values z_i are distinct integers $\leq \frac{b}{v}$. For simplicity, we assume $z_0 = 0$, i.e. $\boldsymbol{\sigma}_0 = \mathbf{I}_n$. Then, given a $m \times n$ signature $\mathbf{g}_0$, we can use the family $\mathcal{F}$ to obtain a generator matrix $\mathbf{G}$ for a code $\mathcal{C}$ as

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_1 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{g}_{\frac{k}{m}-1} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_0\boldsymbol{\pi}^{vz_1} \\ \mathbf{g}_0\boldsymbol{\pi}^{vz_2} \\ \vdots \\ \mathbf{g}_0\boldsymbol{\pi}^{vz_{\frac{k}{m}-1}} \end{bmatrix}. \tag{36}$$

A parity-check matrix for $\mathcal{C}$ can be chosen in reproducible form, by taking an $s \times n$ matrix $\mathbf{h}_0$, and use it to generate the parity-check matrix $\mathbf{H}$ as

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_0 \\ \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_{\frac{b}{v}-1} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_0 \\ \mathbf{h}_0(\boldsymbol{\pi}^{b-v})^T \\ \mathbf{h}_0(\boldsymbol{\pi}^{b-2v})^T \\ \vdots \\ \mathbf{h}_0(\boldsymbol{\pi}^{v})^T \end{bmatrix}. \tag{37}$$

When $\mathbf{H}$ has full rank, then $\mathcal{C}$ has redundancy $r = s\frac{b}{v}$; the code dimension and redundancy must be linked to the code length according to $k + s\frac{b}{v} = n$.

It is quite easy to show that such a parity check matrix is compliant with Theorem 6. In fact, we have

$$\mathbf{h}_j\boldsymbol{\sigma}_i^T = \mathbf{h}_0(\boldsymbol{\pi}^{b-jv})^T(\boldsymbol{\pi}^{vz_i})^T = \mathbf{h}_0\left[\boldsymbol{\pi}^{b+(z_i-j)v}\right]^T. \tag{38}$$

If $z_i \geq j$, we have

$$\begin{aligned}\left[\boldsymbol{\pi}^{b+(z_i-j)v}\right]^T &= \left[\boldsymbol{\pi}^{2b-b+(z_i-j)v}\right]^T = \left[\boldsymbol{\pi}^{b-(\frac{b}{v}+j-z_i)v}\right]^T\left[\boldsymbol{\pi}^{b}\right]^T = \\ &= \left[\boldsymbol{\pi}^{b-(\frac{b}{v}+j-z_i)v}\right]^T = \left[\boldsymbol{\pi}^{b-(j-z_i \mod \frac{b}{v})v}\right]^T.\end{aligned}$$

In the case of $z_i < j$, we can write

$$\left[\boldsymbol{\pi}^{b+(z_i-j)v}\right]^T = \left[\boldsymbol{\pi}^{b-(j-z_i)v}\right]^T\left[\boldsymbol{\pi}^{b-(j-z_i \mod \frac{b}{v})v}\right]^T. \tag{39}$$

Thus, we have proven that

$$\mathbf{h}_j\boldsymbol{\sigma}_i^T = \mathbf{h}_0\left[\boldsymbol{\pi}^{b-(j-z_i \mod \frac{b}{v})v}\right]^T = \mathbf{h}_{(j-z_i \mod \frac{b}{v})}, \tag{40}$$

such that the function $f(x_0, x_1)$ required by Theorem 6 is defined as

$$f(x_0, x_1) = x_1 - z_{x_0} \mod \frac{b}{v}. \tag{41}$$

For instance, a simple construction can be obtained by choosing $m = s = 1$ and $k = r = n/2$: the matrices $\mathbf{G}$ and $\mathbf{H}$ are two reproducible matrices, with signatures that are row vectors of length n, and are characterized by the same number of rows (thus, $\mathcal{C}$ has rate 1/2).

5.3 Code-based schemes from quasi-reproducible codes

The algebraic structures we have introduced in the previous sections can be used to generate key pairs in code-based cryptosystems. For instance, let us consider a parity-check matrix $\mathbf{H}$ made of $r_0 \times n_0$ matrices belonging to a pseudo-ring $\mathcal{M}_q^{\mathcal{F},m}$. In order to use $\mathbf{H}$ as the private key of a sparse-matrix code-based instance of the Niederreiter cryptosystem, we must guarantee that $\mathbf{H}$ is sufficiently sparse: this property can be easily achieved by choosing a family $\mathcal{F}$ of sparse matrices $\boldsymbol{\sigma}_i$, which guarantee that a matrix defined by a sparse signature will be sparse as well. In such a case, we can obtain the public key as $\mathbf{H}' = \mathbf{SH}$, where $\mathbf{S}$ is a random dense matrix, whose elements are picked over $\mathcal{M}_q^{\mathcal{F},m}$. Because of Theorem 2, the entries of $\mathbf{H}'$ belong to $\mathcal{M}_q^{\mathcal{F},m}$, thus they maintain the same reproducible structure defined by $\mathcal{F}$.

If $m = 1$ and $\mathcal{F}$ is a family of permutations satisfying Theorem 3, then $\mathcal{F}_q^{\mathcal{F},1}$ is actually a ring (see Corollary 2). Then, the secret key can be chosen as $\mathbf{H} = [\mathbf{H}_0, \mathbf{H}_1, \cdots, \mathbf{H}_{n_0-1}]$, with $\mathbf{H}_i \in \mathcal{M}_q^{\mathcal{F},1}$, while the public key can corresponds to the systematic form of $\mathbf{H}$, that is $\mathbf{H}' = \mathbf{H}_0^{-1}\mathbf{H}$. Indeed, because of Theorem 4, we have $\mathbf{H}_0^{-1} \in \mathcal{M}_q^{\mathcal{F},1}$, and so $\mathbf{H}'$ is a matrix constituted of blocks over $\mathcal{M}_q^{\mathcal{F},1}$. This is the approach followed in previous LDPC and MDPC code-based instances of the McEliece and Niederreiter cryptosystems [5, 30], in which however only the special case of $\mathbf{H}_i$ in the form of a circulant matrix was considered.

Suppose we have a family $\mathcal{F}$ satisfying Theorem 5, for which multiplication in $\mathcal{M}_q^{\mathcal{F},1}$ is commutative (see Section 4.2 for some examples). Then, we can use the reproducible pseudo-ring induced by $\mathcal{F}$ to obtain key pairs for a McEliece cryptosystem. For instance, we can choose $\mathbf{H} = [\mathbf{H}_0, \mathbf{H}_1]$, with $\mathbf{H}_i \in \mathcal{M}_q^{\mathcal{F},1}$, as the secret parity-check matrix, and obtain a generator matrix as $\mathbf{G} = \mathbf{S}[\mathbf{H}_1^T, -\mathbf{H}_0^T]$, with $\mathbf{S} \in \mathcal{M}_q^{\mathcal{F},1}$. The matrices $\mathbf{H}$ and $\mathbf{G}$ can be used as the private and public key, respectively, for a McEliece cryptosystem. Even if this case might seem quite specific, it is of significant interest since it is exactly the structure appearing in the first of the three variants (BIKE-1) of the BIKE proposal to the NIST competition.

When both Theorems 4 and 5 are satisfied, we can obtain a generator matrix in systematic form, which maintains a quasi-reproducible structure. In fact, starting from a $r \times n$ parity-check matrix $\mathbf{H}$, where the elements are picked randomly from $\mathcal{M}_q^{\mathcal{F},1}$, we can use the corresponding parity-check matrix in systematic form as the public key for a Niederreiter cryptosystem instance. In the same way, we can compute the systematic generator matrix, and use it as the public key in a McEliece cryptosystem instance.

The idea of using codes that are completely reproducible, and not formed by reproducible pseudo-rings, opens up for the possibility of a whole new way of generating key pairs in the McEliece cryptosystem. Indeed, once we have generated a sparse parity-check matrix $\mathbf{H}$, we can use it as the secret key. Then, a possible public key can be obtained by taking a bunch of linearly independent codewords, and using them as the signature of the public generator matrix. If such codewords correspond to rows of the generator matrix in systematic form, then we obviously obtain another significant reduction in the public key size, since there is no need for publishing the first k bits of each one of the selected codewords.

It is clear that having the public code described by a matrix with some reproducible structure leads to a very significant reduction in the public-key size. Indeed, once the structure of the matrix is fixed by the protocol (i.e. dimensions, family $\mathcal{F}$), the whole public-key can be efficiently represented using just the signatures of each building block.

6 Cryptographic Properties and Attacks

In the previous sections we have introduced the notion of reproducibility and have described some properties of codes that are built with such a feature. Our analysis has shown that there can be a wide variety of methods which allow obtaining codes with some reproducible structure. As we have seen in Section

5.3, these codes can be used to generate key-pairs in code-based cryptosystems. The main advantage is the possibility of reducing the information needed to represent the matrix used as the public key. In particular, following the considerations in Section 3, this framework is well suited for sparse-matrix code-based cryptosystems. Let $\mathcal{C}$ be a secret code with parity-check matrix $\mathbf{H}$, and suppose that the public key is constituted by a general generator matrix (for the McEliece case) or parity-check matrix (for the Niederreiter case) of $\mathcal{C}$. Then, the following properties must be satisfied

(a) $\mathbf{H}$ is sufficiently sparse to perform efficient decoding;
(b) the knowledge of the public key does not admit efficient techniques for obtaining $\mathbf{H}$ or another valid sparse parity-check matrix $\mathbf{H}'$.

When property (a) is satisfied, $\mathcal{C}$ is an LDPC code and so admits an efficient decoding algorithm $\mathcal{D}$. We point out that this property can be easily satisfied if we choose $\mathcal{F}$ as a family of sparse matrices: this way, choosing a sparse signature for $\mathbf{H}$ guarantees that $\mathbf{H}$ will be sparse as well. Satisfying property (b) might result in being the most delicate part, since it depends on the particular reproducible structure we consider. However, as the case of circulant matrices clearly shows, this property might not be hard to satisfy. For instance, let us consider the systematic form of $\mathbf{H} = [\mathbf{H}_0|\mathbf{H}_1]$ obtained as $\mathbf{H}' = \mathbf{H}_1^{-1}\mathbf{H}$. For a generic sparse matrix, there is no constraint regarding the density of its inverse. This means that, unless for particular structures (like orthogonal matrices), $\mathbf{H}_1^{-1}$ is dense with overwhelming probability, and this is enough to hide the structure of $\mathbf{H}$ into that of $\mathbf{H}'$. For the systematic generator matrix, we have $\mathbf{G}' = [\mathbf{I}_k|(\mathbf{H}_1^{-1}\mathbf{H}_0)^T]$, and so we can make analogous considerations.

Regardless of the particular choice of $\mathcal{F}$, it is important to note that this additional structure does not expose the secret key to the risk of enumeration. For instance, let us consider the construction described in Section 5.2, in which the signature $\mathbf{H}$ is defined by a signature of size $m \times n$, with all the rows having weight w. If we assume that the rows are picked in such a way as to be linearly independent, the cardinality of the secret key is then approximately equal to $\binom{n}{w}^m$. It is easy to see that, for practical choices of the parameters, this number is sufficiently high to make attacks based on the enumeration of the secret key unfeasible.

6.1 Reaction Attacks

Reaction attacks [16–18, 24] are a recent kind of attacks aimed at recovering the private key by exploiting events of decoding failure. In this section we briefly describe the attack proposed in [24], and then we make some considerations about reproducible codes. In particular, we consider a QC code with parity-check matrix $\mathbf{H} = [\mathbf{H}_0|\mathbf{H}_1]$, where each $\mathbf{H}_i$ is a sparse $p \times p$ circulant with row and column weight equal to w. Then, the resulting code has length $n = 2p$, dimension and redundancy equal to p.

In a reaction attack, the opponent impersonates Alice, producing ciphertexts and sending them to Bob. Events of decoding failure can be detected since, in the case of a decoding failure, Bob must ask for a retransmission. A crucial player in a reaction attack is the distance spectrum, that is the set of all distances produced by the elements of value 1 in a vector [24]. If a distance d appears μ times in the spectrum, we say that it has multiplicity equal to μ; if a distance is not in the spectrum, we say that it has zero multiplicity. In the case of QC codes, these distances are computed cyclically: given two ones at positions x_0 and x_1, the corresponding distance is obtained as $d = \min\left\{\pm(x0 - x_1) \mod p\right\}$. In a circulant matrix, all the rows are characterized by the same distance spectrum; in particular, an opponent performing a reaction attack aims to obtain the distance spectrum of the rows of $\mathbf{H}_0$. For this purpose, he collects the produced ciphertexts into subsets Σ_d, such that each error vector used for the encryption of a ciphertext in Σ_d has d in the distance spectrum of its first circulant block. Then he observes a sufficiently large number of Bob's reactions and assigns a decoding failure probability to each set. As observed in [24], the decoding failure probability of Σ_d depends on the presence of couples of ones in the rows of $\mathbf{H}_0$, at the same distance d. Indeed, suppose that the first length-p block of $\mathbf{e}$ has a couple of ones forming the distance d; then, the following properties hold

- if the distance spectrum of $\mathbf{H}_0$ contains d with multiplicity μ, then the couple of ones overlaps with μ rows of $\mathbf{H}$;
- if the distance spectrum of $\mathbf{H}_0$ does not contain d, then the couple of ones does not overlap with any row of $\mathbf{H}$.

These justify the fact that the average syndrome weight of the ciphertexts belonging to the same set Σ_d depends on the multiplicity of d in the spectrum of $\mathbf{H}_0$, as observed in [16]. In particular, the syndrome weight slightly decreases as μ increases, and this causes the difference in the corresponding decoding failure probabilities [16]. This allows an opponent to obtain the distance spectrum of $\mathbf{H}_0$, since he can guess the multiplicity of each distance d by looking at the decoding failure probability of the corresponding set Σ_d. Since $\mathbf{H}_0$ is sparse, its distance spectrum is sparse as well, which means that it contains a small number of distances. It is then possible to recover $\mathbf{H}_0$ from the knowledge of its distance spectrum, with a procedure that can be related to the solution of a graph problem, in which solutions are represented as cliques of size w. In principle, finding cliques of maximal size in a graph is a hard problem; however, the fact that the graph is sparse allows for an efficient solution of the problem [17, 24].

Consider now the case of reproducible codes in general. For simplicity, we will focus on a code with $k = r = p$ and $n = 2p$, with a signature made of just one row, and a family $\mathcal{F}$ of functions $\boldsymbol{\sigma}_i$ that are obtained as consecutive powers of a permutation $\boldsymbol{\psi}$. In addition, let us suppose that $\boldsymbol{\psi}$ is obtained as the product of two disjoint p-cycles. In other words, ψ is such that that we can find two disjoint sets $\left\{a_0^{(0)}, a_1^{(0)}, \cdots, a_{p-1}^{(0)}\right\}$ and $\left\{a_0^{(1)}, a_1^{(1)}, \cdots, a_{p-1}^{(1)}\right\}$, such that

$$f_{\boldsymbol{\psi}}\left(a_j^{(b)}\right) = a_{j+1 \mod p}^{(b)}, \quad b \in \{0, 1\}. \tag{42}$$

It is clear that

$$f_{\boldsymbol{\sigma}_i}\left(a_j^{(b)}\right) = a_{j+i \mod p}^{(b)}, \quad b \in \{0, 1\}, \forall i. \tag{43}$$

Suppose that the signature of $\mathbf{H}$ has two ones at positions $a_v^{(0)}$ and $a_l^{(0)}$, with $a_l^{(0)} - a_v^{(0)} = d$. Then, in the i-th row of $\mathbf{H}$ these ones correspond to the positions $a_{v+i \mod p}^{(0)}$ and $a_{l+i \mod p}^{(0)}$. The corresponding distance is $d' = a_{l+i \mod p}^{(0)} - a_{v+i \mod p}^{(0)}$. It is clear that d' and d are, in general, different.

As a toy example, set $p = 7$ and $\boldsymbol{\psi}$ formed by the cycles $\{1, 8, 5, 3, 16, 0, 13\}$ and $\{4, 12, 10, 6, 15, 11, 16\}$. For simplicity, suppose that in the signature there are two ones at positions 0 and 1. These correspond to the ones at positions 13 and 8 in the second row of $\mathbf{H}$, at positions 1 and 8 in the third row, etc. The distances between these ones are all different, depending on the considered row.

This removes the basis on which reaction attacks are built, and proves that families of reproducible codes exist that can make reaction attacks infeasible. Asserting the resistance of general families of transformations requires deeper investigations, although some conclusions can already be drawn. First of all, in the case of reproducible codes it is no longer guaranteed that the opponent can find a way to produce subsets Σ_d that are associated to different decoding failure probabilities. Secondly, as opposed to the case of QC codes, it may be impossible for the opponent to obtain the distance spectrum of a single row of $\mathbf{H}$.

6.2 Decoding One Out of Many

In [34], Sendrier introduced a technique, called *Decoding One Out of Many* (DOOM), which speeds up the execution of ISD algorithms for certain families of codes, including QC codes. In general, this technique can be applied whenever there are multiple instances of SDP with just one solution. When ISD is used to perform a decoding attack, the gain obtained from DOOM can be explained as follows. Consider the public parity-check matrix $\mathbf{H}'$ and a set of N different syndromes $\mathcal{S} = \left\{\mathbf{s}^{(0)}, \mathbf{s}^{(1)}, \cdots, \mathbf{s}^{(N-1)}\right\}$ to be decoded. Suppose that, $\forall \mathbf{e}^{(i)}$ such that $\mathbf{H}'{\mathbf{e}^{(i)}}^T = \mathbf{s}^{(i)}$, there exists a bijective function that allows to obtain $\mathbf{e}^{(i)}$ from $\mathbf{e}^{(0)}$ and vice-versa. We denote such a function by $\mathcal{B}$, so that $\mathbf{e}^{(i)} = \mathcal{B}(\mathbf{e}^{(0)})$ and $\mathbf{e}^{(0)} = \mathcal{B}^{-1}(\mathbf{e}^{(i)})$. Then each pair $\left\{\mathbf{s}^{(i)}, \mathbf{H}'\right\}$ can be considered as the input of an ISD instance aimed at finding $\mathbf{e}^{(0)}$ with weight $\leq w$ such that $\mathbf{H}'\mathcal{B}(\mathbf{e}^{(0)})^T = \mathbf{H}'{\mathbf{e}^{(i)}}^T = \mathbf{s}^{(i)}$. According to DOOM, we consider N_i independent calls to an ISD algorithm: as soon as one of these runs successfully ends, the whole algorithm ends since $\mathbf{e}^{(0)}$ has been found. The corresponding gain is equal to $|\mathcal{S}| / \sqrt{N_i} = N/\sqrt{N_i}$, which becomes $\sqrt{N}$ when $N_i = N$. Obviously, exploiting DOOM is beneficial when the N_i decoding instances have comparable complexity. This occurs on condition that $\mathbf{e}^{(i)} = \mathcal{B}(\mathbf{e}^{(0)})$ has the same Hamming weight as $\mathbf{e}^{(0)}$, or almost the same.

So, the rationale of exploiting DOOM for a decoding attack is to intercept one ciphertext and then try to obtain other valid ciphertexts from it, corresponding to transformed versions of the same error vector. Let us consider the case in which the opponent intercepts a ciphertext corresponding to an initial syndrome $\mathbf{s}^{(0)}$, and wants to recover the vector $\mathbf{e}^{(0)}$ used during encryption. Then, in order to apply DOOM, the opponent must produce other syndromes corresponding to as many error vectors being deterministic functions of $\mathbf{e}^{(0)}$. In other words, suppose that ISD returns the solution $\mathbf{e}^{(i)}$ for $\mathbf{s}^{(i)}$, then it must be $\mathbf{e}^{(i)} = \mathbf{A}\mathbf{e}^{(0)}$, with $\mathbf{A}$ being a full-rank matrix. For instance, in the QC case, the opponent can obtain a set of p syndromes $\mathcal{S}$ just by cyclically shifting the initial syndrome $\mathbf{s}^{(0)}$ and the corresponding error vector $\mathbf{e}^{(0)}$.

In general terms, the applicability of DOOM can be modeled as follows. Starting from a syndrome $\mathbf{s}^{(0)} = \mathbf{H}'\mathbf{e}^{(0)T}$, we want to determine a transformation $\boldsymbol{\Phi}$ of the syndrome that corresponds to a transformation $\boldsymbol{\Psi}$ of the error vector, that is

$$\boldsymbol{\Phi}\mathbf{s}^{(0)} = \boldsymbol{\Phi}\mathbf{H}'\mathbf{e}^{(0)T} = \mathbf{H}'\left(\mathbf{e}^{(0)}\boldsymbol{\Psi}\right)^{T} = \mathbf{H}'\boldsymbol{\Psi}^{T}\mathbf{e}^{(0)T}, \tag{44}$$

where $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are two matrices over $\mathbb{F}_q$, with size $r \times r$ and $n \times n$, respectively. The previous equation must be satisfied for every vector $\mathbf{e}^{(0)}$; this can happen only if

$$\exists\ \boldsymbol{\Phi} \in \mathbb{F}_q^{r\times r},\ \boldsymbol{\Psi}\mathbb{F}_q^{n\times n}\ \text{ s.t. }\ \boldsymbol{\Phi}\mathbf{H}' = \mathbf{H}'\boldsymbol{\Psi}^{T}. \tag{45}$$

For the general class of reproducible codes, the applicability of DOOM must be carefully analyzed. For instance, consider a code obtained with the procedure described in Section 5.2, using a family of functions $\mathcal{F}$ consisting of powers of a single function. If this is a permutation, ue to Theorem 6, we have that $\mathbf{H}\boldsymbol{\sigma}_i$ with $\boldsymbol{\sigma}_i \in \mathcal{F}$ always results in a permutation of the rows of $\mathbf{H}$. So, the opponent can build the set $\mathcal{S}$, which is used as input for the DOOM algorithm, by multiplying the initial syndrome by the matrices $\boldsymbol{\sigma}_i$. However, the case in which $\boldsymbol{\pi}$ is not a permutation matrix is of interest. For instance, take three commuting permutation matrices $\boldsymbol{\pi}^{(a)}$, $\boldsymbol{\pi}^{(b)}$ and $\boldsymbol{\pi}^{(c)}$, all having order equal to $n = 2p$, for some integer p. We focus here on the binary case, and suppose $\boldsymbol{\pi} = \boldsymbol{\pi}_a + \boldsymbol{\pi}_b + \boldsymbol{\pi}_c$. It can be easily verified that $\boldsymbol{\pi}$ has order n, since

$$\boldsymbol{\pi}^{n} = \boldsymbol{\pi}_a^{2p} + \boldsymbol{\pi}_b^{2p} + \boldsymbol{\pi}_c^{2p} = \mathbf{I}_n + \mathbf{I}_n + \mathbf{I}_n = \mathbf{I}_n.$$

We then choose $\mathcal{F} = \{\mathbf{I}_n, \boldsymbol{\sigma}_0, \cdots, \boldsymbol{\sigma}_{p-1}\}$, with $\boldsymbol{\sigma}_i = \boldsymbol{\pi}^{2i}$. Since $\boldsymbol{\pi}_a$, $\boldsymbol{\pi}_b$ and $\boldsymbol{\pi}_c$ commute, we have $\boldsymbol{\sigma}_i = \boldsymbol{\pi}_a^{2i} + \boldsymbol{\pi}_b^{2i} + \boldsymbol{\pi}_c^{2i}$. This means that each $\boldsymbol{\sigma}_i$ is a sparse matrix, with row and column weight ≤ 3. Then, starting from a sparse signature $\mathbf{h}_0$, we can obtain a sparse parity-check matrix.

In this case, the opponent can still produce a set $\mathcal{S}$, since equation (44) can be satisfied by choosing $\boldsymbol{\Psi} = \boldsymbol{\sigma}_i$; the corresponding reordering of the rows of $\mathbf{H}$ is a cyclic shift by i positions. However, it results that $\mathbf{e}^{(i)} = \mathbf{e}^{(0)}\boldsymbol{\sigma}_i$ and $\mathbf{e}^{(i)}$ does not have the same Hamming weight of $\mathbf{e}^{(0)}$. In fact, since both $\mathbf{e}^{(0)}$ and $\boldsymbol{\sigma}_i$ are sparse matrices, the weight of $\mathbf{e}^{(i)}$ is close to 3 times the weight of $\mathbf{e}^{(0)}$. According to [15], we can approximate the time complexity for solving ISD on $\mathbf{s}^{(0)}$ as 2^{ct}, where $c = -\log_2(1 - \frac{k}{n})$ and t is the Hamming weight of $\mathbf{s}^{(0)}$. Since the syndromes $\mathbf{s}^{(i)}$, $i \geq 1$, are associated to error vectors that have weight approximately equal to $3t$, applying ISD on them requires a time-complexity that can be estimated as 2^{3ct}. Then, there is no gain in considering this set of multiple instances, since the additional instances produced by the opponent are associated to an ISD complexity that is significantly larger than that of the original instance.

7 Explicit Constructions

In this section we introduce some explicit constructions of reproducible codes that can be advantageous for the use in code-based cryptographic schemes, with the aim of illustrating the potential of this theoretical framework and paving the way for further developments and research on the topic.

7.1 Quasi-Dyadic MDPC Codes

Dyadic matrices, defined in Section 4.2, have been used with some success in cryptography, but always in the context of algebraic codes. The first proposal [29] using quasi-dyadic (QD) Goppa codes was cryptanalyzed [19] almost in its entirety. A later proposal based on Generalized Srivastava (GS) codes [31] proved to be more robust, and led to one of the current NIST submissions, DAGS [8]. Nevertheless, the threat of structural attacks is always present, as shown by the recent results of Barelli and Couvreur [9]. On the other hand, using dyadic matrices has undeniable advantages in terms of key reduction, and leads to fast and efficient arithmetic (as shown in [7]) while at the same time featuring a reproducible structure which is less "obvious" than that provided by circulant matrices.

This is why we believe that designing MDPC codes with QD structure, i.e. QD-MDPC codes, has a great potential in cryptography. Constructing a code-based scheme from QD-MDPC codes is actually rather intuitive. Since dyadic matrices have many good properties (e.g. they are symmetric) and satisfy Theorems 2-4, we can follow the guidelines detailed in Section 5.3. Indeed, in the simplest instantiation, one can form a parity-check matrix by selecting just two blocks, i.e. $\mathbf{H} = [\mathbf{H}_0, \mathbf{H}_1]$, with $\mathbf{H}_i \in \mathcal{M}_q^{\mathcal{F},1}$ of size $r \times r$. In this case $\mathcal{F} = \{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_{r-1}\}$ consists of r distinct dyadic permutations (as described in Section 4.2), for r a power of 2. The resulting code has rate $1/2$. Then we can obtain a generator matrix as $\mathbf{G} = \mathbf{S}[\mathbf{H}_1^T, -\mathbf{H}_0^T]$, where $\mathbf{S} \in \mathcal{M}_q^{\mathcal{F},1}$ is dense. This is essentially a dyadic version of BIKE-1, where $\mathbf{H}$ and $\mathbf{G}$ are used as private and public key in a framework based on McEliece.

On the other hand, using a matrix in systematic form requires some caution. In fact, as explained in Section 6, since dyadic matrices are orthogonal, the density of the inverse matrix is not guaranteed, which could create problems in terms of security. Therefore, if one wants to use a matrix in systematic form, we recommend to choose $\mathbf{H}_i$ to be 2×2 block matrices, with blocks in $\mathcal{M}_q^{\mathcal{F},1}$ of size $r/2 \times r/2$. Then it is possible, for instance, to form the matrix $\mathbf{H}' = \mathbf{H}_1^{-1}\mathbf{H}$ and use it as a public key in a framework based on Niederreiter.

Note that in both cases the size of the public matrix is identical, since in the first case we have two fully-dyadic blocks of size r, for a total of $2r = n$ bits, while in the second case we have one quasi-dyadic block for which we need four signatures, totaling again $4 \cdot r/2 = 2r = n$ bits.

7.2 Block-wise circulant matrices

As shown in Section 4.2, circulant matrices are a classic special case of reproducible matrices and have already been used in cryptography for some time. For a circulant matrix, the signature corresponds to its first row and $\mathcal{F} = \{\boldsymbol{\sigma}_0 = \mathbf{I}_p, \boldsymbol{\sigma}_1 = \boldsymbol{\pi}, \boldsymbol{\sigma}_2 = \boldsymbol{\pi}^2, \ldots, \boldsymbol{\sigma}_{p-1} = \boldsymbol{\pi}^{p-1}\}$, where $\boldsymbol{\pi}$ is the unitary circulant permutation matrix (13).

The concept of circulant matrix can be easily generalized into that of a block-wise circulant matrix as follows. Let us consider $m > 1$, such that $m|p$, and an $m \times p$ signature $\mathbf{z}$ formed by m independent rows of p elements each, with entries over $\mathbb{F}_q$. Then, let us consider the set of permutations

$$\mathcal{F} = \left\{\boldsymbol{\sigma}_0 = \mathbf{I}_p, \boldsymbol{\sigma}_1 = \boldsymbol{\pi}^m, \boldsymbol{\sigma}_2 = \boldsymbol{\pi}^{2m}, \cdots, \boldsymbol{\sigma}_{\frac{p}{m}-1} = \boldsymbol{\pi}^{p-m}\right\}, \tag{46}$$

which induces $\mathcal{M}_q^{\mathcal{F},m}$ as the set of all reproducible matrices of the type

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z} \\ \mathbf{z}\boldsymbol{\pi}^m \\ \mathbf{z}\boldsymbol{\pi}^{2m} \\ \vdots \\ \mathbf{z}\boldsymbol{\pi}^{p-m} \end{bmatrix}. \tag{47}$$

These matrices are indeed block-wise circulant matrices, in the sense that any block of m rows is originated by the previous block of m rows through a cyclic shift by m positions. It is easy to verify that, for every matrix $\mathbf{Z} \in \mathcal{M}_q^{\mathcal{F},m}$, we have

$$\boldsymbol{\sigma}_i\mathbf{Z} = \boldsymbol{\pi}^{im}\mathbf{Z} = \mathbf{Z}\boldsymbol{\pi}^{im} = \mathbf{Z}\boldsymbol{\sigma}_i, \ \ \forall i \in \mathbb{N}, 0 \leq i \leq \frac{p}{m} - 1.$$

Based on Theorem 2, $\mathcal{M}_q^{\mathcal{F},m}$ is a semigroup with respect to the multiplication.

Circulant matrices have the property that any distance between a pair of ones in their first row can be found in any other position in one of the other rows, due to the unitary cyclic shift between any row and the subsequent one. In this more general formulation, shifts by m positions replace unitary shifts, therefore the aforementioned property no longer holds. This hinders reactions attacks of the type introduced in [24], which rely on such a property of circulant matrices.

To the best of our knowledge, it is the first time that this type of reproducible structure is proposed in literature, and this construction therefore represents even more of a novelty, compared to the one presented in the previous section.

References

1. https://bigquake.inria.fr/.
2. https://csrc.nist.gov/projects/post-quantum-cryptography.
3. M. Baldi. *LDPC codes in the McEliece cryptosystem: attacks and countermeasures*, volume 23 of *NATO Science for Peace and Security Series - D: Information and Communication Security*, pages 160–174. IOS Press, 2009.
4. M. Baldi, A. Barenghi, F. Chiaraluce, G. Pelosi, and P. Santini. Ledakem: A post-quantum key encapsulation mechanism based on QC-LDPC codes. In *Post-Quantum Cryptography - 9th International Conference, PQCrypto 2018, Fort Lauderdale, FL, USA, April 9-11, 2018, Proceedings*, pages 3–24, 2018.
5. M. Baldi, M. Bianchi, and F. Chiaraluce. Optimization of the parity-check matrix density in QC-LDPC code-based McEliece cryptosystems. In *Proc. IEEE ICC 2013 - Workshop on Information Security over Noisy and Lossy Communication Systems*, Budapest, Hungary, June 2013.
6. M. Baldi, M. Bodrato, and F. Chiaraluce. A new analysis of the McEliece cryptosystem based on QC-LDPC codes. In *Security and Cryptography for Networks*, volume 5229 of *Lecture Notes in Computer Science*, pages 246–262. Springer Verlag, 2008.
7. G. Banegas, P. S. Barreto, E. Persichetti, and P. Santini. Designing efficient dyadic operations for cryptographic applications. *IACR Cryptology ePrint Archive*, 2018(650), 2018.
8. G. Banegas, P. S. L. M. Barreto, B. O. Boidje, P. Cayrel, G. N. Dione, K. Gaj, C. T. Gueye, R. Haeussler, J. B. Klamti, O. Ndiaye, D. T. Nguyen, E. Persichetti, and J. E. Ricardini. DAGS: key encapsulation using dyadic GS codes. *IACR Cryptology ePrint Archive*, 2017:1037, 2017.
9. E. Barelli and A. Couvreur. An efficient structural attack on nist submission dags. *arXiv preprint arXiv:1805.05429*, 2018.
10. P. S. Barreto, S. Gueron, T. Gueneysu, R. Misoczki, E. Persichetti, N. Sendrier, and J.-P. Tillich. Cake: code-based algorithm for key encapsulation. In *IMA International Conference on Cryptography and Coding*, pages 207–226. Springer, 2017.
11. A. Becker, A. Joux, A. May, and A. Meurer. Decoding random binary linear codes in $2^{n/20}$: How $1 + 1 = 0$ improves information set decoding. In D. Pointcheval and T. Johansson, editors, *Advances in Cryptology - EUROCRYPT 2012*, volume 7237 of *Lecture Notes in Computer Science*, pages 520–536. Springer Verlag, 2012.
12. T. P. Berger, P.-L. Cayrel, P. Gaborit, and A. Otmani. Reducing key length of the McEliece cryptosystem. In *Progress in Cryptology - AFRICACRYPT 2009*, volume 5580 of *Lecture Notes in Computer Science*, pages 77–97. Springer Verlag, 2009.
13. E. Berlekamp, R. McEliece, and H. van Tilborg. On the inherent intractability of certain coding problems. *IEEE Trans. Inform. Theory*, 24(3):384–386, May 1978.
14. D. J. Bernstein. Grover vs. mceliece. In *PQCrypto*, 2010.
15. R. Canto Torres and N. Sendrier. *Analysis of Information Set Decoding for a Sub-linear Error Weight*, pages 144–161. Springer International Publishing, 2016.
16. E. Eaton, M. Lequesne, A. Parent, and N. Sendrier. QC-MDPC: A Timing Attack and a CCA2 KEM. In *PQCrypto*, 2018.
17. T. Fabšič, V. Hromada, P. Stankovski, P. Zajac, Q. Guo, and T. Johansson. A reaction attack on the QC-LDPC McEliece cryptosystem. In *Post-Quantum Cryptography*, volume 10346 of *LNCS*, pages 51–68. Springer, 2017.

18. T. Fabsic, V. Hromada, and P. Zajac. A reaction attack on LEDApkc. *IACR Cryptology ePrint Archive*, 2018:140, 2018.
19. J.-C. Faugère, A. Otmani, L. Perret, and J.-P. Tillich. Algebraic cryptanalysis of McEliece variants with compact keys. In *EUROCRYPT 2010*, volume 6110 of *Lecture Notes in Computer Science*, pages 279–298. Springer Verlag, 2010.
20. J.-C. Faugère, A. Otmani, L. Perret, and J.-P. Tillich. A distinguisher for high rate McEliece cryptosystems. In *Proc. IEEE Information Theory Workshop (ITW)*, pages 282–286, Paraty, Brazil, Oct. 2011.
21. P. Gaborit. Shorter keys for code based cryptography. In *Proc. Int. Workshop on Coding and Cryptography (WCC 2005)*, pages 81–90, Bergen, Norway, Mar. 2005.
22. R. G. Gallager. *Low-density parity-check codes.* M.I.T. Press, 1963.
23. L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proc. 28th Annual ACM Symposium on the Theory of Computing*, pages 212–219, Philadephia, PA, May 1996.
24. Q. Guo, T. Johansson, and P. Stankovski. A key recovery attack on MDPC with CCA security using decoding errors. In *ASIACRYPT*, volume 10031 of *LNCS*, pages 789–815. Springer, 2016.
25. A. S. Householder. Unitary triangularization of a nonsymmetric matrix. *J. ACM*, 5:339–342, 1958.
26. J. Leon. A probabilistic algorithm for computing minimum weights of large error-correcting codes. *IEEE Trans. Inform. Theory*, 34(5):1354–1359, Sept. 1988.
27. A. May, A. Meurer, and E. Thomae. Decoding random linear codes in $O(2^{0.054n})$. In *ASIACRYPT*, volume 7073 of *LNCS*, pages 107–124. Springer, 2011.
28. R. J. McEliece. A public-key cryptosystem based on algebraic coding theory. *DSN Progress Report*, pages 114–116, 1978.
29. R. Misoczki and P. S. L. M. Barreto. Compact McEliece keys from Goppa codes. In *Selected Areas in Cryptography*, volume 5867 of *Lecture Notes in Computer Science*, pages 376–392. Springer Verlag, 2009.
30. R. Misoczki, J. P. Tillich, N. Sendrier, and P. S. L. M. Barreto. MDPC-McEliece: New McEliece variants from moderate density parity-check codes. In *2013 IEEE International Symposium on Information Theory*, pages 2069–2073, Jul. 2013.
31. E. Persichetti. Compact mceliece keys based on quasi-dyadic srivastava codes. *Journal of Mathematical Cryptology*, 6(2):149–169, 2012.
32. E. Prange. The use of information sets in decoding cyclic codes. *IRE Transactions on Information Theory*, 8(5):5–9, Sept. 1962.
33. P. Santini, M. Baldi, G. Cancellieri, and F. Chiaraluce. Hindering reaction attacks by using monomial codes in the mceliece cryptosystem. *CoRR*, abs/1805.04722, 2018.
34. N. Sendrier. Decoding one out of many. In B.-Y. Yang, editor, *Post-Quantum Cryptography*, volume 7071 of *Lecture Notes in Computer Science*, pages 51–67. Springer Verlag, 2011.
35. P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, Oct. 1997.
36. V. M. Sidelnikov and S. O. Shestakov. On insecurity of cryptosystems based on generalized reed-solomon codes. *Discrete Mathematics and Applications*, 2(4):439–444, 1992.
37. J. Stern. A method for finding codewords of small weight. In G. Cohen and J. Wolfmann, editors, *Coding Theory and Applications*, volume 388 of *Lecture Notes in Computer Science*, pages 106–113. Springer Verlag, 1989.
38. J.-P. Tillich. The decoding failure probability of mdpc codes. *CoRR*, abs/1801.04668, 2018.