Research Article

Paolo Santini, Edoardo Persichetti*, and Marco Baldi

Reproducible families of codes and cryptographic applications

https://doi.org/10.1515/jmc-2020-0003
received January 24, 2020; accepted August 10, 2021

Abstract: Structured linear block codes such as cyclic, quasi-cyclic and quasi-dyadic codes have gained an increasing role in recent years both in the context of error control and in that of code-based cryptography. Some well known families of structured linear block codes have been separately and intensively studied, without searching for possible bridges between them. In this article, we start from well known examples of this type and generalize them into a wider class of codes that we call $\mathcal{F}$-reproducible codes. Some families of $\mathcal{F}$-reproducible 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 denote these codes as compactly reproducible codes and show that they encompass known families of compactly describable codes such as quasi-cyclic and quasi-dyadic codes. We then consider some cryptographic applications of codes of this type and show that their use can be advantageous for hindering some current attacks against cryptosystems relying on structured codes. This suggests that the general framework we introduce may enable future developments of code-based cryptography.

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

MSC 2020: 11T71, 94A60

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, which consists of cyclically rotating a vector by one position to the right (equivalently, to the left). Other examples which are well known in the literature include *constacyclic* codes, *negacyclic* codes, *quasi-cyclic* codes, and many others.

Recently, this research direction has been investigated further: Misoczki and Barreto in 2009 introduced *quasi-dyadic* codes [1], which contain codewords invariant under a different type of permutation. The work was motivated by its implications for the McEliece cryptosystem [2], and in particular by the necessity of having a family of codes whose generator and parity-check matrices can be represented in a compact way. This is because, in code-based cryptography, the public key of an encryption (or signature) scheme

*** Corresponding author: Edoardo Persichetti,** Department of Mathematical Sciences, Florida Atlantic University, Boca Raton, Florida 33431, United States, e-mail: epersichetti@fau.edu
Paolo Santini: Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy, e-mail: p.santini@pm.univpm.it
Marco Baldi: Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy, e-mail: m.baldi@staff.univpm.it

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 lengths 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 toward making code-based cryptography truly practical. Previous efforts to reduce key size were centered on quasi-cyclic algebraic codes [3] and have been since then extended to codes of a different nature, namely the Low-Density Parity-Check (LDPC) codes [4] and their recent generalization known as Moderate-Density Parity-Check (MDPC) codes [5]. These codes are characterized by sparse parity-check matrices and admit matrices in quasi-cyclic form, 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 [6] and Quasi-Cyclic Moderate-Density Parity-Check (QC-MDPC) codes [5] are among the most promising solution in this area.

The importance of code-based cryptography has risen dramatically in modern times due to the work of Shor [7], who showed 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 calls for cryptographic 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 [2], it has shown no vulnerabilities against quantum attackers. Moreover, *generic* decoding attacks, which have exponential complexity, have improved only marginally over nearly 40 years of cryptanalysis. Together with lattice-based schemes, code-based cryptography is at the basis of many candidates for the post-quantum standardization call recently launched by NIST [8].

In this article, we provide a general framework for the definition of structured codes, which are of increasing interest in several McEliece and Niederreiter cryptosystem variants. First, we introduce the notion of $\mathcal{F}$-reproducible codes as a general framework for describing both structured and unstructured codes. Then, we introduce some special families of $\mathcal{F}$-reproducible codes, that we denote as compactly reproducible (CR) codes, which require a smaller-than-maximum number of degrees of freedom for the representation of each code belonging to the same family. This generalizes existing families of structured codes used in code-based cryptosystems. We also propose a framework for constructing $\mathcal{F}$-reproducible codes of any kind and present concrete families of non-trivial CR codes which have not appeared in literature before. Our goal is to provide a generic framework to serve as a basis for future constructions, as indeed was the case in ref. [9], which references a preprint version of this work.

To highlight the importance of these codes in cryptography, we mention that among the 26 candidates that were admitted to the second round of the NIST's standardization effort [10], 5 are based on structured random and pseudo-random codes, which are the focus of this article. In particular, BIKE and LEDAcrypt are two public-key encryption schemes based on, respectively, QC-MDPC and QC-LDPC codes, which naturally fit into the general framework we describe in this article. The same occurs for the system named HQC, in which part of the public key consists in a random QC code. Although we focus on the Hamming metric case, the framework we describe could also be applied to the generation of structured codes in the rank metric (with the proper modifications). ROLLO and RQC are other two candidates that could be encompassed by such a framework in the rank metric domain.

The article is organized as follows. In Section 2, we recall some basic concepts and introduce the notation we use throughout the article. In Section 3, we introduce $\mathcal{F}$-reproducible matrices, and we use them to define the new class of codes in Section 4. Section 5 is devoted to the study of their possible use in code-based cryptosystems and provides some practical constructions for this purpose. In Section 6, we draw some conclusions.

2 Preliminaries and notation

We denote with $\mathbb{F}_q$ the finite field with q elements, where q is a prime power. For two sets X and Y, X^Y denotes the set of all maps from Y to X. 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 with 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 $\mathbf{0}_{k\times n}$. Finally, we use the term *pseudo-ring* to denote a structure that satisfies all the ring axioms, apart from the existence of the multiplicative identity. Such a structure is also typically known as *rng*.

2.1 Coding theory background

A linear code 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 2.1. (Hamming metric) The Hamming weight $\mathrm{wt}(\mathbf{x})$ of a vector $\mathbf{x} \in \mathbb{F}_q^n$ is the number of its non-zero entries. The Hamming distance $\mathrm{d}(\mathbf{x}, \mathbf{y})$ between two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{F}_q^n$ is defined as the weight of their difference, i.e., $\mathrm{d}(\mathbf{x}, \mathbf{y}) = \mathrm{wt}(\mathbf{x} - \mathbf{y})$. The minimum distance d of a code C is defined as the minimum distance between any two different codewords of 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 2.2. (Generator and parity-check matrices) Let C be a linear code over $\mathbb{F}_q$. We call *generator matrix* of C a $k \times n$ matrix $\mathbf{G}$ whose rows form a basis for the vector space defined by C, i.e.:

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

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 C a full rank $r \times n$ matrix $\mathbf{H}$ such that every codeword belonging to C has syndrome 0 with respect to $\mathbf{H}$, i.e.,

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

Note that the parity-check matrix of a code C is also a generator matrix of the *dual* code $C^\perp$, i.e., the linear code formed by all the words of $\mathbb{F}_q^n$ that are orthogonal to 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, note 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 left multiplication by an invertible $k \times k$ matrix $\mathbf{S}$, so that $\mathbf{G}' = \mathbf{SG}$. 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 the two properties above, we can write any generator matrix $\mathbf{G}$ in *systematic form* as $\mathbf{G} = [\mathbf{I}_k|\mathbf{A}]$, where $|$ denotes concatenation. If C is generated by $\mathbf{G} = [\mathbf{I}_k|\mathbf{A}]$, then a (systematic) parity-check matrix for C is $\mathbf{H} = [-\mathbf{A}^T|\mathbf{I}_r]$.

2.2 The McEliece cryptosystem

The McEliece public-key encryption scheme [2] 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.

2.2.1 Key generation

Let $\mathbf{G}$ be a generator matrix of 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 $\mathbf{G}' \coloneqq \mathbf{SGP}$.

2.2.2 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 $\mathbf{c} = \mathbf{mG}' + \mathbf{e}$.

2.2.3 Decryption

The decryption algorithm first computes $\hat{\mathbf{c}} = \mathbf{cP}^{-1} = \mathbf{mSG} + \mathbf{eP}^{-1}$. As $\mathbf{P}$ is a permutation matrix, $\mathbf{eP}^{-1}$ has the same weight as $\mathbf{e}$. Therefore, $\mathcal{D}$ can be used to decode the errors and obtain $\hat{\mathbf{m}} = \mathbf{mS} = \mathcal{D}(\hat{\mathbf{c}})$. 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[1] secure Public Key Encryption (PKE) scheme or Key Encapsulation Mechanism (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 based on the following hard problem.

Problem 2.3
(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 [11]. Note that, since the McEliece cryptosystem uses an $[n, k, d]$ code, the number of error vectors of weight t 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.

1 The term IND-CCA2 stands for Indistinguishability under Adaptively Chosen Ciphertext Attack, which is the highest security notion for a PKE and KEM since it considers the strongest adversarial model.

2.3 Sparse-matrix codes

One of the most delicate points about the McEliece cryptosystem is that, in order for the security to reduce to the SDP, it is assumed that the matrix used 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., ref. [12]), 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 [13] 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 toward codes of a different nature.

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 $\mathcal{D}$, inspired by the Bit Flipping (BF) decoder of Gallager [14], which is very efficient in practice. For these reasons, this class of codes is a natural candidate for the McEliece cryptosystem. A first instantiation was studied in ref. [4], where a private LDPC matrix was considered, along with a linearly transformed version of the same matrix used as the public key. As highlighted in ref. [4], security of the private LDPC code is not preserved unless the public matrix is dense. Thus, in such a framework, the private LDPC code C is represented through its sparse parity-check matrix $\mathbf{H}$, while the public key corresponds to a dense generator matrix $\mathbf{G}$ for C. It is important to note that, from the knowledge of $\mathbf{G}$, the opponent can compute several parity-check matrices $\mathbf{H}'$ for 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, and $\mathbf{H}_1$ is full rank. 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 particular 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 [15]. Therefore, finding reliable models for their DFR is necessary to ensure that its value is negligible for those instances designed to achieve indistinguishability under chosen ciphertext attack (IND-CCA) [16].

2.4 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.

2.4.1 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 [17] and has received many improvements during

the years [18–21]. 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 [22] to speed up the search, leads to a reduction in the complexity, with respect to the classical case, which cannot be larger than half the exponent α [23].

2.4.2 Key-recovery attacks

When LDPC codes are used, key recovery attacks boil down to recovering low-weight codewords from the dual of the public code, which is again a decoding problem. Let us denote by $C^{\perp}$ the dual code of C, having generator matrix $\mathbf{H}$. Since the rows of $\mathbf{H}$ are sparse, and of maximum weight $w \ll n$, they are minimum-weight codewords in $C^{\perp}$ with overwhelming probability, and so can be searched with a generic algorithm for finding low-weight words, for which ISD algorithms can be used as well.

Since the difficulty of such a task increases with the weight of the searched codewords, it makes sense to relax the notion of "low-density": the authors in ref. [5] introduce the notion of "moderate-density" by increasing the allowed row weight in the parity-check matrix from $O(\log(n))$ to $O(\sqrt{n})$, thus defining moderate-density parity-check (MDPC) codes. It is still possible to decode MDPC codes 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 in modern variants of the McEliece cryptosystem does not reduce the security 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 $\mathbf{H}$.

2.5 Structured sparse-matrix codes

Using generic LDPC and MDPC codes without any structure in the McEliece cryptosystem is not a practical choice, as pointed out in ref. [4]. This is because the need to avoid sparse public matrices makes the resulting public key sizes 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 [3] and was then extended to LDPC codes [24] and algebraic codes [25]. In all cases, the authors propose to use QC codes to reduce the public key size. A QC code can be simply seen as 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

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

Any circulant matrix is fully described by one of its rows, conventionally the first one. This means that, in the McEliece cryptosystem, we can describe the public key completely using just the first row of each one of its circulant blocks; it is clear that this results in a significant reduction in the public key size with respect to instances using non-structured public matrices. 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 [26] 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 [27] shows a reduction of about 1/4 in the key size compared to what 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 [28]). 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 $\mathbf{H}$ [29]. 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 $\min\{\pm(i-j) \bmod p\}$, with p being the circulant size). Based on these considerations, it has been shown in ref. [29] 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 [30,31], of special iterative decoders that allow theoretical modeling of their failure rate [32,33], or of particular families of codes that make the reconstruction of the secret key unfeasible [34]. However, all these solutions come with some price to 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 and third cases).

As we will see in the rest of this article, 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 us to avoid DOOM and/or reaction attacks, or at least to reduce their efficiency.

3 Reproducibility

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

Definition 3.1. Let $n, k \in \mathbb{N}$, with $k = \ell m$ where also $\ell, m \in \mathbb{N}$. Let $\mathcal{F} = \{\boldsymbol{\sigma}_0, \dots, \boldsymbol{\sigma}_{\ell-1}\}$ be a family of ℓ linear maps, with $\boldsymbol{\sigma}_i : \mathbb{F}_q^n \mapsto \mathbb{F}_q^n$ (thus, we can think of each σ_i as a square matrix of size n and values in $\mathbb{F}_q$). We say that a $k \times n$ matrix A is an $\mathcal{F}$-reproducible matrix if there exists an $m \times n$ matrix $\mathbf{a}$ such that

$$A = \begin{bmatrix} a \cdot \sigma_0 \\ a \cdot \sigma_1 \\ \vdots \\ a \cdot \sigma_{\ell-1}. \end{bmatrix} \tag{3.1}$$

We call m the *reproducible order* and $\mathbf{a}$ the *signature set* and write $\mathbf{A} = \mathcal{F}(\mathbf{a})$. We say that a code $C \subseteq \mathbb{F}_q^n$ is an $\mathcal{F}$-reproducible code if it admits a generator matrix and/or a parity-check matrix which are $\mathcal{F}$-reproducible.

Let us consider an $\mathcal{F}$-reproducible code described by an $\mathcal{F}$-reproducible generator matrix $\mathbf{G} \in \mathbb{F}_q^{k\times n}$ such that, for $\mathcal{F} = \{\boldsymbol{\sigma}_0, \ldots, \boldsymbol{\sigma}_{\ell-1}\}$, we have

$$\mathbf{G} = \begin{bmatrix} \mathbf{g} \cdot \boldsymbol{\sigma}_0 \\ \mathbf{g} \cdot \boldsymbol{\sigma}_1 \\ \vdots \\ \mathbf{g} \cdot \boldsymbol{\sigma}_{\ell-1} \end{bmatrix}, \tag{3.2}$$

where $\mathbf{g}$ is the $m \times n$ signature set of $\mathbf{G}$. Then, for the fixed family $\mathcal{F}$ of linear maps, the code is completely represented through $\mathbf{g}$. The same reasoning applies to an $\mathcal{F}$-reproducible code described by an $\mathcal{F}$-reproducible parity-check matrix $\mathbf{H} \in \mathbb{F}_q^{r\times n}$ with signature set $\mathbf{h}$.

Proposition 3.2. *Any $[n, k, d]$-code over $\mathbb{F}_q$ is an $\mathcal{F}$-reproducible code for at least one choice of $\mathcal{F}$ and the corresponding signature set. Such a choice corresponds to $\ell = 1$, $m = k$, $\mathbf{g} = \mathbf{G}$, and $\mathcal{F} = \{\mathbf{I}_n\}$, where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Equivalently, the code can be described through the parity-check matrix $\mathbf{H}$ considering $\ell = 1$, $m = r$, $\mathbf{h} = \mathbf{H}$, and $\mathcal{F} = \{\mathbf{I}_n\}$.*

Once the family $\mathcal{F}$ is defined, an $\mathcal{F}$-reproducible matrix can be described just by its signature set. Consequently, when the family of maps $\mathcal{F}$ is fixed and universally known, having an $\mathcal{F}$-reproducible generator matrix (or equivalently parity-check matrix) with $\ell > 1$ leads to a more compact representation of the code with respect to storing its full generator or parity-check matrix. This happens because $\mathcal{F}$ is universally known, and it does not need to be included in the code representation, thus the signature set alone is sufficient for representing the code.

If we consider a single code, then it is always possible to find some family $\mathcal{F}$ according to which such a code has an $\mathcal{F}$-reproducible generator matrix (or equivalently parity-check matrix) with $\ell > 1$. This is detailed in the following two propositions.

Proposition 3.3. *Any single $[n, k, d]$-code over $\mathbb{F}_q$ admits multiple generator and parity-check matrices, thus it can be an $\mathcal{F}$-reproducible code for several choices of $\mathcal{F}$ and the corresponding signature set.*

Proof. The proof is straightforward and omitted for saving space. □

Proposition 3.4. *For any single $[n, k, d]$-code C over $\mathbb{F}_q$, a family $\mathcal{F}$ with $\ell = k$ entries can be defined according to which such a code admits an $\mathcal{F}$-reproducible generator matrix with reproducible order $m = 1$. Similarly, a family $\mathcal{F}$ with $\ell = r$ entries can be defined according to which C admits an $\mathcal{F}$-reproducible parity-check matrix with reproducible order $m = 1$.*

Proof. Let $\mathbf{G} \in \mathbb{F}_q^{k\times n}$ be a valid generator matrix for the code C. Let us consider the ith row $\mathbf{g}_i$ of $\mathbf{G}$ and define $\boldsymbol{\sigma}_i$, $i \in [1;k]$, as the $n \times n$ matrix $\in \mathbb{F}_q^{n\times n}$ having its first row equal to $\mathbf{g}_i$, and all the other rows filled with arbitrary entries. Then, $\mathbf{G}$ is easily obtained as $\mathbf{G} = F(\mathbf{a})$, with $\mathbf{a} = [1, 0, 0, \ldots, 0]$. The fact that C admits an $\mathcal{F}$-reproducible parity-check matrix with reproducible order $m = 1$ can be proved with a similar reasoning. □

From Proposition 3.4, we know that any single code is $\mathcal{F}$-reproducible for some family $\mathcal{F}$ yielding $\ell > 1$ and $m < k$ (considering the generator matrix) or $m < r$ (considering the parity-check matrix). However, if

instead of a single code we consider a group of codes and aim at representing all of them as $\mathcal{F}$-reproducible codes for the same, universally known family of maps $\mathcal{F}$, then it is not always possible to find a solution with $\ell > 1$ and $m < k$ (considering the generator matrix) or $m < r$ (considering the parity-check matrix). The only trivial solutions that always exist are those of the type considered in Proposition 3.2, yielding $\ell = 1$ and $m = k$ (considering the generator matrix) or $m = r$ (considering the parity-check matrix), and thus not enabling more compact code representations than those corresponding to storing the full generator or parity-check matrix. We are instead interested in group of codes that, besides these trivial solutions, also admit $\mathcal{F}$-reproducible generator and parity-check matrices for a fixed $\mathcal{F}$ with $\ell > 1$ and $m < k$ or $m < r$, as detailed in the next definition.

Definition 3.5. We say that a group of $[n, k, d]$-codes over $\mathbb{F}_q$ are Compactly Reproducible (CR) codes if, for a fixed $\mathcal{F}$ with $\ell > 1$, each of them admits at least one $\mathcal{F}$-reproducible generator matrix with $m < k$, or at least one $\mathcal{F}$-reproducible parity-check matrix with $m < r$, thus enabling a more compact code representation with respect to storing the full generator or parity-check matrix.

The condition for a code to be CR can be generalized, in order to take into account other structures that enable a compact representation.

Definition 3.6. Let $\mathbf{A}_{i,j} \in \mathbb{F}_q^{k_{i,j} \times n_{i,j}}$ be $\mathcal{F}$-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 *$\mathcal{F}$-quasi-reproducible*.

Definition 3.7. Let us consider a group of linear codes over $\mathbb{F}_q$. If, for a fixed $\mathcal{F}$ with $\ell > 1$, any code C in such a group can be described by an $\mathcal{F}$-quasi-reproducible generator matrix $\mathbf{G} \in \mathbb{F}_q^{k \times n}$ such that $m < k$, and/or an $\mathcal{F}$-quasi-reproducible parity-check matrix $\mathbf{H} \in \mathbb{F}_q^{r \times n}$ such that $m < r$, then we say that C is a quasi-compactly reproducible (QCR) code.

It is clear that, in order to describe an $\mathcal{F}$-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 set. A particular type of quasi-reproducible code is the one in which the blocks $\mathbf{A}_{i,j}$ are square matrices, defined by the same family $\mathcal{F}$.

We are now ready to introduce a very important notion regarding the set of $\mathcal{F}$-reproducible matrices obtained via a given family of transformations. Specifically, consider a family of linear functions $\mathcal{F} = \left\{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_{\frac{p}{m}-1}\right\}$, 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 $\mathcal{F}$-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[2] hold.

Theorem 3.8. *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 $-\mathbf{a}$. □

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

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 3.9. *$\mathcal{M}_q^{\mathcal{F},m}$ 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*

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

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

$$\mathbf{C} = \mathbf{AB} = \begin{bmatrix} \mathbf{a}_0\mathbf{B} \\ \mathbf{a}_1\mathbf{B} \\ \vdots \\ \mathbf{a}_{\frac{p}{m}-1}\mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_0\mathbf{B} \\ \mathbf{a}_0\boldsymbol{\sigma}_1\mathbf{B} \\ \vdots \\ \mathbf{a}_0\boldsymbol{\sigma}_{\frac{p}{m}-1}\mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_0 \\ \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_{\frac{p}{m}-1} \end{bmatrix}. \tag{3.3}$$

Now by hypothesis

$$\mathbf{c}_i = \mathbf{a}_0\boldsymbol{\sigma}_i\mathbf{B} = \mathbf{a}_0\mathbf{B}\boldsymbol{\sigma}_i = \mathbf{c}_0\boldsymbol{\sigma}_i, \tag{3.4}$$

for all $i \le \frac{p}{m} - 1$. It follows that $\mathbf{C}$ is $\mathcal{F}$-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 (3.3). Since by hypothesis $\mathbf{C} \in \mathcal{M}_q^{\mathcal{F},m}$, and therefore is $\mathcal{F}$-reproducible, we have that $\mathbf{c}_i = \mathbf{c}_0\boldsymbol{\sigma}_i$ for all $i \le \frac{p}{m} - 1$. It follows that

$$\mathbf{a}_0\boldsymbol{\sigma}_i\mathbf{B} = \mathbf{c}_i = \mathbf{c}_0\boldsymbol{\sigma}_i = \mathbf{a}_0\mathbf{B}\boldsymbol{\sigma}_i. \tag{3.5}$$

Now, since equation (3.5) 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 Theorem 3.9 holds, $\mathcal{M}_q^{\mathcal{F},m}$ verifies all the requisites of a mathematical *pseudo-ring*, i.e., a ring without multiplicative identity, as defined in Section 2. We call this the *$\mathcal{F}$-reproducible pseudo-ring* induced by $\mathcal{F}$ over $\mathbb{F}_q$.

3.1 Pseudo-rings induced by families of permutations

In the particular case of signatures made of just one row (i.e., reproducible order $m = 1$) and the functions $\boldsymbol{\sigma}_i$ being permutations, we have a further result, which is described in Theorem 3.10. 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 $f^{-1}_{\boldsymbol{\sigma}_i}$, which is the bijection associated with the permutation matrix $\boldsymbol{\sigma}_i^{-1} = \boldsymbol{\sigma}_i^T$; if $f_{\boldsymbol{\sigma}_i}(v) = j$, then $f^{-1}_{\boldsymbol{\sigma}_i}(j) = v$. 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, $a'_j = a_{f^{-1}_{\boldsymbol{\sigma}_i}(j)}$. 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 $f_{\boldsymbol{\sigma}_i} \circ f_{\boldsymbol{\sigma}_j}(v) = f_{\boldsymbol{\sigma}_i}(f_{\boldsymbol{\sigma}_j}(v))$. 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 3.10. *Let $\mathcal{F} = \{\boldsymbol{\sigma}_0 = \mathbf{I}_p, \boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_{p-1}\}$ be a family of linear transformations, with each $\boldsymbol{\sigma}_i$ being a permutation, and suppose that $\mathcal{F}$ induces the $\mathcal{F}$-reproducible pseudo-ring $\mathcal{M}_q^{\mathcal{F},1}$ over $\mathbb{F}_q$. Then, the following relation must be satisfied*

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

Proof. Since $\mathcal{M}_q^{\mathcal{F},1}$ is a pseudo-ring, we know from Theorem 3.9 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

$$\boldsymbol{\sigma}_i\mathbf{B} = \begin{bmatrix} \mathbf{b}_{f_{\boldsymbol{\sigma}_i}(0)} \\ \mathbf{b}_{f_{\boldsymbol{\sigma}_i}(1)} \\ \vdots \\ \mathbf{b}_{f_{\boldsymbol{\sigma}_i}(p-1)=} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_0\boldsymbol{\sigma}_{f_{\boldsymbol{\sigma}_i}(0)} \\ \mathbf{b}_0\boldsymbol{\sigma}_{f_{\boldsymbol{\sigma}_i}(1)} \\ \vdots \\ \mathbf{b}_0\boldsymbol{\sigma}_{f_{\boldsymbol{\sigma}_i}(p-1)} \end{bmatrix}, \tag{3.6}$$

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

$$\mathbf{B}\sigma_i = \begin{bmatrix} \mathbf{b}_0\boldsymbol{\sigma}_0 \\ \mathbf{b}_0\boldsymbol{\sigma}_1 \\ \vdots \\ \mathbf{b}_0\boldsymbol{\sigma}_{p-1} \end{bmatrix}\boldsymbol{\sigma}_i = \begin{bmatrix} \mathbf{b}_0\boldsymbol{\sigma}_0\boldsymbol{\sigma}_i \\ \mathbf{b}_0\boldsymbol{\sigma}_1\boldsymbol{\sigma}_i \\ \vdots \\ \mathbf{b}_0\boldsymbol{\sigma}_{p-1}\boldsymbol{\sigma}_i \end{bmatrix}. \tag{3.7}$$

Putting together equations (3.6) and (3.7), we obtain

$$\boldsymbol{\sigma}_j\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_{f_{\boldsymbol{\sigma}_i}(j)}, \tag{3.8}$$

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

Starting from the result of Theorem 3.10, we can easily derive some other properties that $\mathcal{F}$ must satisfy.

Corollary 3.11. *Let $\mathcal{F}$ be a family of permutations such that the induced $\mathcal{M}_q^{\mathcal{F},m}$ is a pseudo-ring. Then, $\mathcal{F}$ has the following properties*:
(*a*) $f_{\boldsymbol{\sigma}_i}(0) = i, \quad \forall i$;
(*b*) $\forall i\ \exists j$ s.t. $f_{\boldsymbol{\sigma}_i} \circ f_{\boldsymbol{\sigma}_j} = f_{\mathbf{I}_p}$.

Proof. Since $\mathcal{F}$ satisfies the hypothesis of Theorem 3.10, we have

$$\boldsymbol{\sigma}_{f_{\boldsymbol{\sigma}_i}(0)} = \boldsymbol{\sigma}_0\boldsymbol{\sigma}_i = \mathbf{I}_p\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_i, \tag{3.9}$$

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

$$\boldsymbol{\sigma}_j\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_{f_{\sigma_i}(j)} = \boldsymbol{\sigma}_0 = \mathbf{I}_p. \tag{3.10}$$

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 3.12. *Let $\mathcal{F}$ be a family of permutations such that the induced $\mathcal{M}_q^{\mathcal{F},m}$ is a pseudo-ring. Then, $\mathcal{M}_q^{\mathcal{F},1}$ is a ring, which we call, by analogy, $\mathcal{F}$-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 3.11, $\mathcal{F}$ 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 $\mathbf{u} = [1, 0, \ldots, 0]$, we easily obtain the $p \times p$ identity matrix $\mathbf{I}_p$. □

Theorem 3.13. *Let $\mathcal{F}$ be a family of permutations such that the induced $\mathcal{M}_q^{\mathcal{F},m}$ is a pseudo-ring. Then, $\mathcal{M}_q^{\mathcal{F},1}$ is an $\mathcal{F}$-reproducible ring and the invertible elements of $\mathcal{M}_q^{\mathcal{F},1}$ form a multiplicative group.*

Proof. Based on Corollary 3.12, $\mathcal{M}_q^{\mathcal{F},1}$ is an $\mathcal{F}$-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 $\mathbf{u} = [1, 0, \ldots, 0]$ as in Corollary 3.12. 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 3.9, 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 3.14, we analyze how transposition acts on the matrices belonging to an $\mathcal{F}$-reproducible pseudo-ring $\mathcal{M}_q^{\mathcal{F},1}$.

Theorem 3.14. *Let $\mathcal{M}_q^{\mathcal{F},1}$ be an $\mathcal{F}$-reproducible pseudo-ring; if*

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

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 ith row of $\mathbf{B}$ corresponds to the ith column of $\mathbf{A}$. In particular, the ith 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 ith row of $\mathbf{B}$ corresponds to the ith column of $\mathbf{A}$. Let us denote as $\mathbf{b}_0$ the first row of $\mathbf{B}$, that is,

$$\mathbf{b}_0 = \left[a_0, a_{f_{\boldsymbol{\sigma}_1}^{-1}(0)}, \ldots, a_{f_{\boldsymbol{\sigma}_{p-1}}^{-1}(0)}\right] = \left[a_{f_{\boldsymbol{\sigma}_0}^{-1}(0)}, a_{f_{\boldsymbol{\sigma}_1}^{-1}(0)}, \ldots, a_{f_{\boldsymbol{\sigma}_{p-1}}^{-1}(0)}\right]. \tag{3.11}$$

Let us consider the ith 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

$$\mathbf{b}_i = \left[a_i, a_{f_{\boldsymbol{\sigma}_1}^{-1}(i)}, \ldots, a_{f_{\boldsymbol{\sigma}_{p-1}}^{-1}(i)}\right] = \left[a_{f_{\boldsymbol{\sigma}_0}^{-1}(i)}, a_{f_{\boldsymbol{\sigma}_1}^{-1}(i)}, \ldots, a_{f_{\boldsymbol{\sigma}_{p-1}}^{-1}(i)}\right] = \mathbf{b}_0\boldsymbol{\sigma}_i. \tag{3.12}$$

Now suppose that $f_{\boldsymbol{\sigma}_i}(v) = j$; then, the jth entry of $\mathbf{b}_i$ corresponds to the vth 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

$$z = f_{\boldsymbol{\sigma}_v}^{-1}(0), \quad v = f_{\boldsymbol{\sigma}_i}^{-1}(j). \tag{3.13}$$

In order to satisfy eq. (3.12), a_z must be equal to the jth entry of the ith 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,

$$f_{\boldsymbol{\sigma}_j}^{-1}(i) = f_{\boldsymbol{\sigma}_v}^{-1}(0), \quad v = f_{\boldsymbol{\sigma}_i}^{-1}(j), \tag{3.14}$$

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 3.13 and 3.14. Let $\mathbf{A}$ be a square matrix whose elements are picked from $\mathcal{M}_q^{\mathcal{F},1}$. By definition, we have $\mathbf{A}^{-1} = \det(\mathbf{A})^{-1}\mathrm{adj}(\mathbf{A})$, where $\det(\mathbf{A})$ is the determinant of $\mathbf{A}$ and $\mathrm{adj}(\mathbf{A})$ 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 3.13, $\det(\mathbf{A})^{-1} \in \mathcal{M}_q^{\mathcal{F},1}$. Computing $\mathrm{adj}(\mathbf{A})$ involves sums, multiplications and transpositions: because of Theorem 3.14, we have that the entries of $\mathrm{adj}(\mathbf{A})$ 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 $\mathcal{F}$-quasi-reproducible structure of $\mathbf{A}$.

3.2 Known examples of $\mathcal{F}$-reproducible pseudo-rings

In Section 3.1, 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.

3.2.1 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 $\mathcal{F}$-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

$$\pi_{l,j} = \begin{cases} 1 & \text{if } l + 1 \equiv j \mod p \\ 0 & \text{otherwise}. \end{cases} \tag{3.15}$$

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

$$f_{\boldsymbol{\pi}}(v) = v + 1 \bmod p. \tag{3.16}$$

It can be easily shown that

$$f_{\boldsymbol{\sigma}_i}(v) = f_{\boldsymbol{\pi}^i}(v) = \underbrace{f_{\boldsymbol{\pi}} \circ f_{\boldsymbol{\pi}} \cdots \circ f_{\boldsymbol{\pi}}}_{i \text{ times}}(v) = v + i \mod p, \tag{3.17}$$

which leads to $\boldsymbol{\pi}^p = \mathbf{I}_p$ and $\boldsymbol{\pi}^i\boldsymbol{\pi}^j = \boldsymbol{\pi}^{i+j \bmod p}$. 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

$$\boldsymbol{\sigma}_i\boldsymbol{\sigma}_j = \boldsymbol{\pi}^{i+j \bmod p} = \boldsymbol{\sigma}_{i+j \bmod p}, \tag{3.18}$$

which is compliant with Theorem 3.10, since $f_{\boldsymbol{\sigma}_i}(j) = i + j \bmod p$. With some simple computations, it can be easily shown that circulant matrices satisfy Theorem 3.14 and that the multiplication between two circulant matrices is commutative.

3.2.2 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 written as an $\mathcal{F}$-reproducible matrix, 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:

$$f_{\boldsymbol{\sigma}_i}(v) = v \oplus i \bmod p. \tag{3.19}$$

If we combine two transformations, we obtain

$$f_{\boldsymbol{\sigma}_i} \circ f_{\boldsymbol{\sigma}_j}(v) = (v \oplus j) \oplus i = v \oplus (i \oplus j) = f_{\boldsymbol{\sigma}_{i\oplus j}}(v). \tag{3.20}$$

Since $f_{\boldsymbol{\sigma}_i}(j) = i \oplus j$, this proves that the family of dyadic matrices is compliant with Theorem 3.10. It can be straightforwardly proven that dyadic matrices are symmetric (and so satisfy Theorem 3.14), and that the multiplication between two dyadic matrices is commutative.

Circulant and dyadic matrices are just two particular cases of $\mathcal{F}$-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 $\mathcal{F}$-reproducible pseudo-ring $\mathcal{M}_q^{\mathcal{F},m}$, induced by $\mathcal{F} = \left\{\sigma_0 = \mathbf{I}_p, \sigma_1, \ldots, \sigma_{\frac{p}{m}-1}\right\}$, we can obtain a new $\mathcal{F}$-reproducible pseudo-ring as

$$\mathcal{M}_q^{\mathcal{F}',m} = \{\mathbf{M}' | \mathbf{M}' = \boldsymbol{\psi}\mathbf{M}\boldsymbol{\psi}^T, \quad \mathbf{M} \in \mathcal{M}_q^{\mathcal{F},m}\}. \tag{3.21}$$

The corresponding family of transformations is $\mathcal{F}' = \left\{\sigma'_0, \sigma'_1, \ldots, \sigma'_{\frac{p}{m}-1}\right\}$, 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

$$\mathbf{A} + \mathbf{B} = \boldsymbol{\psi}\mathbf{M}_A\boldsymbol{\psi}^T + \boldsymbol{\psi}M_B\boldsymbol{\psi}^T = \boldsymbol{\psi}(\mathbf{M}_A + \mathbf{M}_B)\boldsymbol{\psi}^T, \tag{3.22}$$

$$\mathbf{AB} = \boldsymbol{\psi}\mathbf{M}_A\boldsymbol{\psi}^T\boldsymbol{\psi}\mathbf{M}_B\boldsymbol{\psi}^T = \boldsymbol{\psi}\mathbf{M}_A\mathbf{M}_B\boldsymbol{\psi}^T, \tag{3.23}$$

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

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

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.

4 Compactly reproducible codes

In the previous section, we have described the properties that a family of functions $\mathcal{F}$ must have in order to generate $\mathcal{F}$-reproducible matrices. This opens a wide range of possibilities for obtaining codes with compact representations, that is, CR codes according to Definition 3.5. In fact, $\mathcal{F}$-reproducible pseudo-rings

allow us 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 CR codes, and the properties that characterize them. In particular, we study how to achieve an $\mathcal{F}$-reproducible representation for the parity-check matrix $\mathbf{H}$ starting from an $\mathcal{F}$-reproducible generator matrix $\mathbf{G}$. In addition, we provide intuitive methods to obtain random-looking CR codes, starting from their parity-check matrix.

Let C be a CR code over $\mathbb{F}_q$, with length n, dimension k, and codimension $r = n - k$, with an $\mathcal{F}$-reproducible generator matrix $\mathbf{G} \in \mathbb{F}_q^{k\times n}$ defined by the signature $\mathbf{g}_0 \in \mathbb{F}_q^{m\times n}$ and the fixed and universally known family of transformations $\mathcal{F}$. In particular, according to Definition 3.5 we have $\ell = \frac{k}{m} > 1$ and we write $\mathcal{F} = \{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_{l-1}\}$. Without loss of generality, we can suppose that $\boldsymbol{\sigma}_0 = \mathbf{id} = \mathbf{I}_n$. The matrix $\mathbf{G}$ can thus be expressed as

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_{\ell-1} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_0\boldsymbol{\sigma}_1 \\ \vdots \\ \mathbf{g}_0\boldsymbol{\sigma}_{\ell-1} \end{bmatrix}. \tag{4.1}$$

Let $\mathbf{H} \in \mathbb{F}_q^{r\times n}$ be a parity-check matrix for C and s be one of the factors of r; if r is a prime, necessarily $s = 1$. Then, $\mathbf{H}$ can be expressed as

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_0 \\ \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_{\frac{r}{s}-1} \end{bmatrix}, \tag{4.2}$$

where each $\mathbf{h}_i$ is a matrix with dimensions $s \times n$. Since by definition $\mathbf{G}\mathbf{H}^T = \mathbf{0}_{k\times r}$, it must be

$$\mathbf{g}_i\mathbf{h}_j^T = \mathbf{g}_0\boldsymbol{\sigma}_i\mathbf{h}_j^T = 0_{m\times s}, \quad \forall i, j \in \mathbb{N} \quad \text{s.t. } 0 \le i \le l-1,\ 0 \le j \le \frac{r}{s} - 1. \tag{4.3}$$

Let us assume that $\mathbf{g}_0\mathbf{H}^T = \mathbf{0}_{m\times n}$: as we explain later, in the practical case of a cryptographic scheme, this condition can be easily satisfied. The following theorem considers a particular construction for a CR code and states some properties that its parity-check matrix must satisfy.

Theorem 4.1. *Let $\mathbf{G} \in \mathbb{F}_q^{k\times n}$ be an $\mathcal{F}$-reproducible matrix, with signature $\mathbf{g}_0 \in \mathbb{F}_q^{m\times n}$ (hence, m divides k) and family $\mathcal{F} = \left\{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_{\frac{k}{m}-1}\right\}$. For simplicity, we suppose $\boldsymbol{\sigma}_0 = \mathbf{I}_n$. Let $r = n - k$, and $\mathbf{H} \in \mathbb{F}_q^{r\times n}$ such that $\mathbf{g}_0\mathbf{H}^T = \mathbf{0}_{m\times r}$. Let s be a factor of r, and denote by $\mathbf{h}_j$ the subset of rows of $\mathbf{H}$ at positions $\{js, js+1, \ldots, (j+1)s-1\}$. If we can define a function $f(x_0, x_1) : \left[0, \frac{k}{m} - 1\right] \times \left[0, \frac{r}{s} - 1\right] \subset \mathbb{N}^2 \to \left[0, \frac{r}{s} - 1\right] \subset \mathbb{N}$, such that*

$$\mathbf{h}_j\boldsymbol{\sigma}_i^T = \mathbf{h}_{f(i,j)}, \quad \forall i, j \in \mathbb{N}, \quad 0 \le i \le \frac{k}{m} - 1,\ 0 \le j \le \frac{r}{s} - 1, \tag{4.4}$$

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

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

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_{\frac{k}{m}-1} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_0\boldsymbol{\sigma}_1 \\ \vdots \\ \mathbf{g}_0\boldsymbol{\sigma}_{\frac{k}{m}-1} \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} \mathbf{h}_0 \\ \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_{\frac{r}{s}-1} \end{bmatrix}. \tag{4.5}$$

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,

$$\mathbf{g}_i\mathbf{h}_j^T = \mathbf{g}_0\boldsymbol{\sigma}_i\mathbf{h}_j^T = \mathbf{0}_{m\times s}, \quad \forall i,j \in \mathbb{N} \text{ s.t. } 0 \le i \le \frac{k}{m} - 1,\ 0 \le j \le \frac{r}{s} - 1. \tag{4.6}$$

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

$$\mathbf{g}_0\mathbf{h}_j^T = \mathbf{0}_{m\times s}, \quad \forall j \in \mathbb{N} \text{ s.t. } 0 \le j \le \frac{r}{s} - 1. \tag{4.7}$$

Consider now the product $\mathbf{g}_i\mathbf{h}_j^T = \mathbf{g}_0\boldsymbol{\sigma}_i\mathbf{h}_j^T$, for $i \ge 1$. If we can define a function $f(x_0, x_1) : \left[0, \frac{k}{m} - 1\right] \times \left[0, \frac{r}{s} - 1\right] \subset \mathbb{N}^2 \to \left[0, \frac{r}{s} - 1\right] \subset \mathbb{N}$ with the aforementioned property described by (4.4), then for all couples of indexes i, j we have

$$\boldsymbol{\sigma}_i\mathbf{h}_j^T = \mathbf{h}_{f(i,j)}^T, \tag{4.8}$$

and (4.6) is surely satisfied, since

$$\mathbf{g}_i\mathbf{h}_j^T = \mathbf{g}_0\boldsymbol{\sigma}_i\mathbf{h}_j^T = \mathbf{g}_0\mathbf{h}_{f(i,j)}^T = \mathbf{0}_{m\times s}, \tag{4.9}$$

where $\mathbf{g}_0\mathbf{h}_{f(i,j)}^T = \mathbf{0}_{m\times s}$ because of (4.7). □

Remark 4.2. Note that if r is a prime, then we either have $s = r$ or $s = 1$. The first case may lead to somehow trivial constructions: we have that the function f is constant, since it maps any pair $(x_0, 0)$ $\left(\text{with } x_0 \in \left[0, \frac{k}{m} - 1\right]\right)$ to 0. This implies that the matrix $\mathbf{H}$ is such that $\mathbf{H}\boldsymbol{\sigma}_i^T = \mathbf{H}$, for any $\boldsymbol{\sigma}_i \in \mathcal{F}$: if the functions $\boldsymbol{\sigma}_i$ have all full rank (for instance, they are permutations), then $\mathbf{H}$ cannot have maximum rank r. Hence, when r is a prime, the only case with practical interest is that of $s = 1$ (i.e., the one in which each $\mathbf{h}_j$ is actually a row vector).

For $\mathbf{G}$ and $\mathbf{H}$ to be, respectively, the generator and parity-check matrix of a code C, some conditions have to be verified, given in Corollary 4.3.

Corollary 4.3. *Let $\mathbf{G} \in \mathbb{F}_q^{k\times n}$ be an $\mathcal{F}$-reproducible matrix, with signature $\mathbf{g}_0 \in \mathbb{F}_q^{m\times n}$ (hence, m is among the factors of k) and family $\mathcal{F} = \left\{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_{\frac{k}{m}-1}\right\}$. 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 4.1. *For $\mathbf{H}$ and $\mathbf{G}$ to be, respectively, the parity-check and generator matrices of a code 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*) $\frac{k}{m} \le \frac{r}{s}$;

(*c*) *For any three integers $i \in \left[0, \frac{k}{m} - 1\right]$ and $j', j'' \in \left[0, \frac{r}{s} - 1\right]$, with $j' \ne j''$, it must be $f(i, j') \ne f(i, j'')$.*

Proof. We want the $\mathcal{F}$-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 $i \ne j$, 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 exist three integers $i', i'' \in \left[0, \frac{k}{m} - 1\right]$, with $i' \ne i''$ and $j \in \left[0, \frac{r}{s} - 1\right]$, 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 $j \in \left[0, \frac{r}{s} - 1\right]$ and consider, for all $i \in \left[0, \frac{k}{m} - 1\right]$, 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 of a code with redundancy r, then it must have rank equal to r. If we suppose that there exists three integers $i \in \left[0, \frac{k}{m} - 1\right]$, $j', j'' \in \left[0, \frac{r}{s} - 1\right]$, with $j' \neq j''$, such that $f(i, j') = f(i, j'')$ then, because of Theorem 4.1, 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 4.1 and Corollary 4.3 allow us to generate a CR code in a very simple way. Given a family of transformations $\mathcal{F}$, first obtain a matrix $\mathbf{H}$ with the characteristics required by the theorem. Then, for the code C having $\mathbf{H}$ as parity-check matrix, a variety of $\mathcal{F}$-reproducible generator matrices can be found. Indeed, let $\mathbf{G}$ be a generator matrix for 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 an $\mathcal{F}$-reproducible generator matrix, defined by the family $\mathcal{F}$. On condition that both $\mathbf{H}$ and $\mathbf{G}$ have full rank, and $m < k \Rightarrow l > 1$, then they can be used to represent the CR code C with length n, dimension k, and redundancy r.

We point out that the properties defined by Theorem 4.1 can be described in a graphical way, considering the fact that the linear functions $\boldsymbol{\sigma}_i$ define a mapping acting on the ensemble of matrices $\mathbf{h}_j$. We can consider a directed graph $\mathcal{G}$, with $\frac{r}{s}$ nodes, labeled from 0 to $\frac{r}{s} - 1$. In such a graph, we have an edge from a node j_0 to a node j_1 if there exists an integer i such that $\mathbf{h}_{j_0}\boldsymbol{\sigma}_i^T = \mathbf{h}_{j_1}$. In addition, every edge is labeled with the corresponding function $\boldsymbol{\sigma}_i^T$. With this construction, the graph $\mathcal{G}$ contains all the information about the mapping defined by $\mathcal{F}$. The meaning of the graph is the following: if there exists a length-l path from a node j_0 to a node j_1, whose edges have labels $\mathfrak{I} = \{i_0, i_1, \ldots, i_{l-1}\}$, then it must be

$$\mathbf{h}_{j_1} = \mathbf{h}_{j_0}\prod_{i\in\mathfrak{I}}\boldsymbol{\sigma}_i^T. \tag{4.10}$$

We can now consider two different paths having the same starting and final nodes, with the corresponding sets of edges labeled as $\mathfrak{I}^a$ and $\mathfrak{I}^b$. Then, it must be

$$\prod_{i\in\mathfrak{I}^a}\boldsymbol{\sigma}_i^T = \prod_{i\in\mathfrak{I}^b}\boldsymbol{\sigma}_i^T. \tag{4.11}$$

The definitions we have introduced in the previous section describe codes whose generator matrices can be efficiently described just by a subset of their entries; for this reason, they are natural candidates for being used in a McEliece cryptosystem. Actually, some variants of this type have already been proposed during the years, with the aim of reducing the public-key size by exploiting such a property. We show that these already existing variants are encompassed by our general framework and that the possibilities for obtaining such features are actually many more than those already exploited.

In some cases, a QCR code can be seen as a particular case of a CR code (and viceversa). Let us consider a code 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 size $p \times p$, that is,

$$\mathbf{G} = \left[\mathbf{G}_0|\mathbf{G}_1|\cdots|\mathbf{G}_{n_0-1}\right]. \tag{4.12}$$

This form of the generator matrix is commonly used in sparse-matrix code-based cryptosystems [5,35]. Suppose that $\mathbf{G}$ in (4.12) is an $\mathcal{F}$-quasi-reproducible matrix, i.e., each $\mathbf{G}_i$ is an element of the 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 seen as a particular $\mathcal{F}$-reproducible matrix. Let us write the ith family of transformations as $\mathcal{F}_i = \left\{\boldsymbol{\sigma}_0^{(i)}, \boldsymbol{\sigma}_1^{(i)}, \ldots, \boldsymbol{\sigma}_{\frac{p}{m}-1}^{(i)}\right\}$ and define an overall family of transformations $\mathcal{F} = \left\{\boldsymbol{\sigma}_0, \boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_{\frac{p}{m}-1}\right\}$, such that

$$\boldsymbol{\sigma}_i = \begin{bmatrix} \boldsymbol{\sigma}_i^{(0)} & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \cdots & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \boldsymbol{\sigma}_i^{(1)} & \mathbf{0}_{p\times p} & \cdots & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \boldsymbol{\sigma}_i^{(2)} & \cdots & \mathbf{0}_{p\times p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times p} & \cdots & \boldsymbol{\sigma}_i^{(n_0-1)} \end{bmatrix}. \tag{4.13}$$

Then, it is easy to see that a matrix in the form (4.12) is also an $\mathcal{F}$-reproducible matrix obtained through $\mathcal{F}$ in (4.13), with signature

$$\mathbf{g}_0 = [\mathbf{g}_0^{(0)}|\mathbf{g}_0^{(1)}|\cdots|\mathbf{g}_0^{(n_0-1)}]. \tag{4.14}$$

4.1 CR codes from Householder matrices

A *Householder matrix* [36] is a matrix that is at the same time orthogonal and symmetric. Let us consider a set of distinct Householder matrices $\boldsymbol{\psi}_0, \ldots, \boldsymbol{\psi}_{v-1}$. 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 4.1, these matrices must form a commutative group, that is,

$$\boldsymbol{\psi}_i\boldsymbol{\psi}_j = \boldsymbol{\psi}_j\boldsymbol{\psi}_i, \quad 0 \le i,\ j \le v-1. \tag{4.15}$$

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

$$\begin{aligned} &\{\mathbf{a}^{(i)}|0 \le i \le 2^v-1,\ \mathbf{a}^{(i)} \in \mathbb{F}_2^v,\ \text{s.t. } \mathbf{a}^{(i)} \ne \mathbf{a}^{(j)},\ \forall i \ne j\}, \\ &\{\mathbf{b}^{(i)}|\ 0 \le i \le 2^v-1,\ \mathbf{b}^{(i)} \in \mathbb{F}_2^v,\ \text{s.t. } \mathbf{b}^{(i)} \ne \mathbf{b}^{(j)},\ \forall i \ne j\}. \end{aligned} \tag{4.16}$$

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

$$\boldsymbol{\sigma}_i = \prod_{l=0}^{v-1} \boldsymbol{\psi}_l^{a_l^{(i)}}, \tag{4.17}$$

where $a_l^{(i)}$ is the lth entry of $\mathbf{a}^{(i)}$. Since we are considering Householder matrices with the property (4.15), 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 an $\mathcal{F}$-reproducible generator matrix $\mathbf{G}$ for a code C; a parity-check matrix for C can then be the $\mathcal{F}$-reproducible matrix $\mathbf{H}$, with signature $\mathbf{h}_0 \in \mathbb{F}_q^{s\times n}$, whose rows are obtained as

$$\mathbf{h}_j = \mathbf{h}_0\left(\prod_{l=0}^{v-1} \boldsymbol{\psi}_l^{b_l^{(j)}}\right)^T. \tag{4.18}$$

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

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

where $\oplus$ denotes the modulo 2 sum and

$$f(i,j) = u,\ \text{s.t. } \mathbf{b}^{(u)} = \mathbf{a}^{(i)} \oplus \mathbf{b}^{(j)}. \tag{4.19}$$

It is straightforward to show that such a function satisfies the properties required by Theorem 4.1 and Corollary 4.3. 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 $\mathbf{H}$.

4.2 CR codes from powers of a single function

In this section, we present another construction of reproducible codes satisfying Theorem 4.1. 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 $v = 1$. 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 an $m \times n$ signature $\mathbf{g}_0$, we can use the family $\mathcal{F}$ to obtain a generator matrix $\mathbf{G}$ for a code C as

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_1 \\ \mathbf{g}_2 \\ \vdots \\ \mathbf{g}_{\frac{k}{m}-1} \end{bmatrix} = \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_0\boldsymbol{\pi}^{vz_1} \\ \mathbf{g}_0\boldsymbol{\pi}^{vz_2} \\ \vdots \\ \mathbf{g}_0\boldsymbol{\pi}^{vz_{\frac{k}{m}-1}} \end{bmatrix}. \tag{4.20}$$

An $\mathcal{F}$-reproducible parity-check matrix for C can be obtained by taking an $s \times n$ matrix $\mathbf{h}_0$, and using it to generate the parity-check matrix $\mathbf{H}$ as

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_0 \\ \mathbf{h}_1 \\ \mathbf{h}_2 \\ \vdots \\ \mathbf{h}_{\frac{b}{v}-1} \end{bmatrix} = \begin{bmatrix} \mathbf{h}_0 \\ \mathbf{h}_0(\boldsymbol{\pi}^{b-v})^T \\ \mathbf{h}_0(\boldsymbol{\pi}^{b-2v})^T \\ \vdots \\ \mathbf{h}_0(\boldsymbol{\pi}^{v})^T \end{bmatrix}. \tag{4.21}$$

If $\mathbf{H}$ is full rank, then 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 4.1. In fact, we have

$$\mathbf{h}_j\boldsymbol{\sigma}_i^T = \mathbf{h}_0(\boldsymbol{\pi}^{b-jv})^T\left(\boldsymbol{\pi}^{vz_i}\right)^T = \mathbf{h}_0[\boldsymbol{\pi}^{b+(z_i-j)v}]^T. \tag{4.22}$$

If $z_i \geq j$, we have

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

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

$$[\boldsymbol{\pi}^{b+(z_i-j)v}]^T = [\boldsymbol{\pi}^{b-(j-z_i)v}]^T\left[\boldsymbol{\pi}^{b-\left(j-z_i \bmod \frac{b}{v}\right)v}\right]^T. \tag{4.23}$$

Thus, we have proven that

$$\mathbf{h}_j\boldsymbol{\sigma}_i^T = \mathbf{h}_0\left[\boldsymbol{\pi}^{b-\left(j-z_i \bmod \frac{b}{v}\right)v}\right]^T = \mathbf{h}_{\left(j-z_i \bmod \frac{b}{v}\right)}, \tag{4.24}$$

such that the function $f(x_0, x_1)$ required by Theorem 4.1 is defined as

$$f(x_0, x_1) = x_1 - z_{x_0} \bmod \frac{b}{v}. \tag{4.25}$$

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 $\mathcal{F}$-reproducible matrices, with signatures that are row vectors of length n and are characterized by the same number of rows (thus, C has rate 1/2).

For what concerns property (b), we can consider the following equivalence:

$$x_0 - x_1' \equiv x_0 - x_1'' \bmod \frac{r}{s}, \tag{4.26}$$

which turns into

$$x_1'' - x_1' \equiv 0 \bmod \frac{r}{s}. \tag{4.27}$$

Then, it is clear that it must be $x', x'' < \frac{r}{s}$: however, this condition is quite straightforward, since j denotes the row index of the matrix blocks in $\mathbf{H}$. In the same way, when considering the index of the transformation $\boldsymbol{\sigma}_i$, we have

$$x'_0 - x_1 \equiv x''_0 - x_1 \bmod \frac{r}{s}, \tag{4.28}$$

which turns into

$$x'_0 - x''_0 \equiv 0 \bmod \frac{r}{s}. \tag{4.29}$$

Again, in order to guarantee that the previous equivalence has no solution, it must be $x'_0, x''_0 < \frac{r}{s}$. This basically means that we must have $k \leq m\frac{r}{s}$.

Remark. There is a clear analogy between the concept of reproducibility and that of automorphism group of a code. Remember that, by automorphism group, we refer to the set of functions that map a code into itself. For instance, consider codes obtained from generator matrices as in (4.20) and assume that $\boldsymbol{\pi}$ is a permutation. Let us further assume, for simplicity, that $v = 1$ and choose $k = b$, i.e., suppose the code has dimension equal to the order of the considered permutation $\boldsymbol{\pi}$. We then have $\mathcal{F} = \{\mathbf{I}_n, \boldsymbol{\pi}, \boldsymbol{\pi}^2, \ldots, \boldsymbol{\pi}^{k-1}\}$, and for each each $\mathbf{g}_0 \in \mathbb{F}_q^n$ we obtain an $\mathcal{F}$-reproducible generator matrix as

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_0\boldsymbol{\pi} \\ \mathbf{g}_0\boldsymbol{\pi}^2 \\ \vdots \\ \mathbf{g}_0\boldsymbol{\pi}^{k-1} \end{bmatrix}.$$

It is trivial to show that $\mathcal{F}$ is in the automorphism group of the code C having $\mathbf{G}$ as a generator matrix. Indeed, each codeword is obtained as

$$\mathbf{c} = \mathbf{u}\mathbf{G} = \sum_{j=0}^{k-1} u_j \mathbf{g}_0 \boldsymbol{\pi}^j, \quad u_j \in \mathbb{F}_q.$$

If we permute $\mathbf{c}$ according to a permutation $\boldsymbol{\pi}^i$, we obtain

$$\mathbf{c}\boldsymbol{\pi}^i = \sum_{j=0}^{k-1} u_i \mathbf{g}_0 \boldsymbol{\pi}^{i+j} = \sum_{j=0}^{k-1} u'_j \mathbf{g}_0 \boldsymbol{\pi}^j = \mathbf{u}'\mathbf{G}, \quad \text{with } u'_j = u_{j-i \bmod k}.$$

Thus, $\mathbf{u}'$ is a cyclic permutation of $\mathbf{u}$: this proves that $\mathbf{c}\boldsymbol{\pi}^i \in C$. Hence, the automorphism group of C contains all permutations of the form $\boldsymbol{\pi}^i$, for $i \in [1; k-1]$. With similar arguments, one can prove that analogous results hold for other families of transformations that we consider in this article.

4.3 Code-based schemes from QCR 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 an $\mathcal{F}$-reproducible matrix defined by a sparse signature will be sparse as well. In such a case, we can obtain the public key as $\mathbf{H}' = \mathbf{S}\mathbf{H}$, where $\mathbf{S}$ is a random dense matrix, whose elements are picked over $\mathcal{M}_q^{\mathcal{F},m}$. Because of Theorem 3.9, the entries of $\mathbf{H}'$ belong to $\mathcal{M}_q^{\mathcal{F},m}$, thus they maintain the same structure defined by $\mathcal{F}$.

If $m = 1$ and $\mathcal{F}$ is a family of permutations satisfying Theorem 3.10, then $\mathcal{M}_q^{\mathcal{F},1}$ is actually a ring (see Corollary 3.12). Then, the secret key can be chosen as $\mathbf{H} = [\mathbf{H}_0, \mathbf{H}_1, \ldots, \mathbf{H}_{n_0-1}]$, with $\mathbf{H}_i \in \mathcal{M}_q^{\mathcal{F},1}$, while the public key can correspond to the systematic form of $\mathbf{H}$, that is, $\mathbf{H}' = \mathbf{H}_0^{-1}\mathbf{H}$. Indeed, because of Theorem 3.13, 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 instances of the McEliece and Niederreiter cryptosystems based on QC-LDPC and QC-MDPC codes [5,35], which, however, only considered the special case of circulant matrices as $\mathbf{H}_i$.

Suppose we have a family $\mathcal{F}$ satisfying Theorem 3.14, for which multiplication in $\mathcal{M}_q^{\mathcal{F},1}$ is commutative (see Section 3.2 for some examples). Then, we can use the $\mathcal{F}$-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 $\mathbf{H}_i \in \mathcal{M}_q^{\mathcal{F},1}$, 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 [37].

When both Theorems 3.13 and 3.14 are satisfied, we can obtain a generator matrix in systematic form, which is still an $\mathcal{F}$-reproducible matrix. In fact, starting from an $r \times n$ 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 a CR public code may lead to a 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.

5 Cryptographic properties and attacks

In the previous sections, we have introduced the notion of reproducibility and have described some properties of reproducible codes. Our analysis has shown that there can be a wide variety of methods which allow us obtaining reproducible codes. As we have seen in Section 4.3, 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 2.3, this framework is well suited for sparse-matrix code-based cryptosystems. Let 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 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, 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 4.2, 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 large to make attacks based on the enumeration of the secret key unfeasible. In the next sections, we provide some considerations on attacks that work for QC codes and that may be hindered by proper families of reproducible codes. We only provide some qualitative arguments and leave detailed and thorough considerations about these attacks for future works.

5.1 Reaction attacks

Reaction attacks [29,38–40] 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 ref. [29], and then we make some considerations about reproducible codes. In particular, we consider a binary 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 [29]. 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 $d = \min\{\pm(x_0 - x_1) \bmod p\}$. 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 ref. [29], 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 ref. [40]. In particular, the syndrome weight slightly decreases as μ increases, and this causes the difference in the corresponding decoding failure probabilities [40]. 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 not dense, which means that it contains a small number of

distances, with multiplicities that generically are rather low. It is then possible to recover $\mathbf{H}_0$ from the knowledge of its distance spectrum, with a procedure that can be related to that of finding cliques of prefixed size in a given graph. In principle, cliques finding algorithms run with a time complexity that grows exponentially with the clique size; however, for sparse graphs (i.e., graphs that contain a small number of edges), the problem becomes significantly easier [29,38].

In summary, reaction attacks against QC codes are possible because of two factors:

(i) A sufficiently high DFR;

(ii) The invariance of the set of distances between pairs of ones in a row of the secret key with respect to the row index. This guarantees feasibility of the key reconstruction phase, since the resulting graph (in which rows of the secret key are represented by cliques of fixed size) is sparse.

In particular, one can try to counter reaction attacks by choosing codes for which condition (ii) is not met. For instance, in ref. [34] authors propose to use a specific family of QC monomial codes with the property that the distances between pairs of ones in the secret key fill the distance spectrum. In this way, the density in the obtained graph becomes maximal and, as a consequence, reconstructing the secret key becomes unfeasible. We argue that families of reproducible codes may, in general, be characterized by analogous properties.

For simplicity, consider the example of a reproducible 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, 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 $\{a_0^{(0)}, a_1^{(0)}, \dots, a_{p-1}^{(0)}\}$ and $\{a_0^{(1)}, a_1^{(1)}, \dots, a_{p-1}^{(1)}\}$, for which

$$f_{\boldsymbol{\psi}}(a_j^{(b)}) = a_{j+1 \bmod p}^{(b)}, \quad b \in \{0, 1\}. \tag{5.1}$$

It is clear that

$$f_{\boldsymbol{\sigma}_i}(a_j^{(b)}) = a_{j+i \bmod p}^{(b)}, \quad b \in \{0, 1\}, \quad \forall i. \tag{5.2}$$

Suppose now 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 ith row of $\mathbf{H}$ these ones correspond to the positions $a_{v+i \bmod p}^{(0)}$ and $a_{l+i \bmod p}^{(0)}$. The corresponding distance is $d' = a_{l+i \bmod p}^{(0)} - a_{v+i \bmod p}^{(0)}$ which, in general, is different from d.

As a toy example, set $p = 7$ and suppose $\boldsymbol{\psi}$ is formed by the cycles $\{1, 8, 5, 3, 7, 0, 13\}$ and $\{4, 12, 10, 6, 15, 11, 2\}$. For simplicity, suppose that in the secret signature there are two ones in 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 and, furthermore, are not an invariant of the row index. Thus, differently from the case of QC codes, the distances that are produced between ones in the first row of the secret key are not maintained in the other rows.

With this simple example we have shown that, differently from the QC case, the distance spectrum of generic reproducible codes becomes richer and, as a consequence, the graph which is used to discover the secret key becomes denser. Thus, the secret key reconstruction phase, which is the final step of a reaction attack, may be hindered, and this may be enough to remove the basis upon which reaction attacks are built. Asserting the resistance of general families of transformations requires a deeper investigation, although some conclusions can already be drawn.

5.2 DOOM

In ref. [28], Sendrier introduced a technique, called DOOM, which is able to speed 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 $\mathcal{S} = \{\mathbf{s}^{(0)}, \mathbf{s}^{(1)}, \dots, \mathbf{s}^{(N-1)}\}$ 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 us 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 $\{\mathbf{s}^{(i)}, \mathbf{H}'\}$ can be considered as the input of an ISD algorithm 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 comes to an end, the whole algorithm ends as well, 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 independent decoding instances have comparable complexity. This only occurs on the condition that $\mathbf{e}^{(i)} = \mathcal{B}(\mathbf{e}^{(0)})$ has the same Hamming weight as $\mathbf{e}^{(0)}$, or almost the same.

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 $\mathbf{\Phi}$ of the syndrome that corresponds to a transformation $\mathbf{\Psi}$ of the error vector, that is,

$$\mathbf{\Phi}\mathbf{s}^{(0)} = \mathbf{\Phi}\mathbf{H}'\mathbf{e}^{(0)T} = \mathbf{H}'(\mathbf{e}^{(0)}\mathbf{\Psi})^T = \mathbf{H}'\mathbf{\Psi}^T\mathbf{e}^{(0)T}, \tag{5.3}$$

where $\mathbf{\Phi}$ and $\mathbf{\Psi}$ 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

$$\exists \mathbf{\Phi} \in \mathbb{F}_q^{r\times r}, \quad \mathbf{\Psi}\mathbb{F}_q^{n\times n} \quad \text{s.t. } \mathbf{\Phi}\mathbf{H}' = \mathbf{H}'\mathbf{\Psi}^T. \tag{5.4}$$

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 4.2, using a family of functions $\mathcal{F}$ consisting of powers of a single function. If this is a permutation, due to Theorem 4.1, 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, as we have described in the previous sections, reproducible families of codes can be obtained in many different ways. For instance, we can use functions $\boldsymbol{\sigma}_i$ that are powers of a matrix $\boldsymbol{\theta}$ that is not a permutation. In this case, the opponent can still produce a set $\mathcal{S}$, since equation (5.3) can be satisfied by choosing $\mathbf{\Psi} = \boldsymbol{\sigma}_i$; the corresponding reordering of the rows of $\mathbf{H}$ is a cyclic shift by i positions. However, it results that $\mathbf{e}^{(i)} = \mathbf{e}^{(0)}\boldsymbol{\sigma}_i$. Unless $\boldsymbol{\theta}$ is a permutation, powers of this matrix would contain a rather large number of non-null entries: for instance, if $\boldsymbol{\theta}$ is selected at random, then we expect that for any $\boldsymbol{\sigma}_i$ the portion of non-null components is close to $\frac{q-1}{q}$. In such a case, any $\mathbf{e}^{(i)}$ would have a rather large Hamming weight $\left(\text{say, close to } \frac{q-1}{q}\right)$, way larger than that of $\mathbf{e}^{(0)}$. According to ref. [41], we can approximate the time complexity of an ISD algorithm searching for a vector with weight t as 2^{ct}, where $c = -\log_2\left(1 - \frac{k}{n}\right)$. If t is the weight of $\mathbf{e}^{(0)}$, then we have that the ISD algorithm taking $\mathbf{s}^{(0)}$ as input is expected to run in time 2^{ct}. Since all the other syndromes $\mathbf{s}^{(i)}$, with $i \geq 1$, are associated with error vectors with weights significantly larger than t, applying ISD on them requires a time complexity that is significantly larger than 2^{ct}. Then, there is no gain in considering this set of multiple instances, since the additional instances (which are produced by the opponent) are associated with an ISD complexity that is significantly larger than that of the original one.

We note that codes of this type may be employed in cryptosystems where codes in compact form are not required to admit efficient decoding. This is the case, for instance, of the HQC KEM [42] and the AGS identification scheme [43]. In both schemes, a code in compact form is needed to obtain a syndrome decoding instance: while in HQC decoding is done with a public and fixed code, in AGS decoding is not involved at all.

Hence, in this type of applications, the adoption of reproducible families of codes may be convenient: defeating DOOM would obviously result in the possibility of choosing better parameters for a scheme.

5.3 Construction examples

We provide 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 the introduced theoretical framework.

5.3.1 Quasi-dyadic MDPC codes

Dyadic matrices, which we have already mentioned in Section 3.2, have been used with some measure of success in cryptography, but always in the context of algebraic codes. The first proposal using quasi-dyadic (QD) Goppa codes [1] was cryptanalyzed [26] almost in its entirety. A later proposal based on generalized Srivastava (GS) codes [44] was designed to be more robust against the previous attack and led to one of the NIST submissions for the key exchange functionality, DAGS [45,46]. Nevertheless, the threat of structural attacks is always present, as shown by the recent results of Barelli and Couvreur [47]. On the other hand, using dyadic matrices has undeniable advantages, not only in terms of key reduction but also because it leads to fast and efficient arithmetic (as shown in ref. [48]) while at the same time featuring a reproducible structure which is less "obvious" than that provided by circulant matrices.

The reasons mentioned above are why we believe that designing MDPC codes with a QD structure, i.e., QD-MDPC codes, has potential in cryptography. Dyadic matrices have many good properties (e.g., they are symmetric and orthogonal) and satisfy Theorems 3.9–3.13, which means the ensemble $\mathcal{M}_q^{\mathcal{F},1}$ of dyadic matrices forms a fully-fledged ring (which is also commutative). A formal definition of reproducible codes having such a structure is given below.

Definition 5.1. (QD-MDPC codes) Let $\mathcal{M}_q^{\mathcal{F},1}$ be the ring of dyadic matrices. We call *Quasi-Dyadic MDPC* (*QD-MDPC*) *code of type* (r_0, n_0) a linear code of length $n = n_0 p$ and redundancy $r \leq r_0 p$ that admits a parity-check matrix in the form $\mathbf{H} = \{\mathbf{Z}_{ij}\}$, where $\mathbf{Z}_{ij} \in \mathcal{M}_q^{\mathcal{F},1}$ for all $0 \leq i \leq r_0 - 1, 0 \leq j \leq n_0 - 1$, such that $\mathbf{H}$ has row weight $O(\sqrt{n})$.

Constructing a code-based cryptosystem from QD-MDPC codes is actually rather intuitive, since we can follow the guidelines detailed in Section 4.3. However, due to the very same properties we just mentioned, building QD-MDPC codes for cryptographic purposes requires some caution. For example, in the simplest instantiation, one could 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 $p \times p$. However, this would not be secure. In fact, since dyadic matrices are orthogonal, the density of the inverse matrix is not guaranteed. This means that a Niederreiter instantiation would not be secure, since the non-systematic block is obtained as $\mathbf{H}_0^{-1}\mathbf{H}_1$. Similarly, to use the McEliece framework, one could compute a generator matrix as $\mathbf{G} = [\mathbf{G}_0, \mathbf{G}_1] = \mathbf{S}[\mathbf{H}_1^T, -\mathbf{H}_0^T]$, where $\mathbf{S} \in \mathcal{M}_q^{\mathcal{F},1}$ is dense, but then the product $\mathbf{G}_0\mathbf{G}_1^{-1}$ may still reveal the private key, due to the sparsity of the inverse of a dyadic matrix.

As a consequence, to construct code-based schemes using this particular family of reproducible codes, it is recommended to choose $r_0 \geq 2$ and employ "true" block matrices, with blocks in $\mathcal{M}_q^{\mathcal{F},1}$.

5.3.2 Block-wise circulant matrices

As shown in Section 3.2, circulant matrices are a classic special case of reproducible matrices and have already been used in cryptography for quite some time. For a traditional circulant matrix, the signature

corresponds to its first row and the set of transformations is $\mathcal{F} = \{\boldsymbol{\sigma}_0 = \mathbf{I}_p, \boldsymbol{\sigma}_1 = \boldsymbol{\pi}, \boldsymbol{\sigma}_2 = \boldsymbol{\pi}^2, \ldots, \boldsymbol{\sigma}_{p-1} = \boldsymbol{\pi}^{p-1}\}$, where $\boldsymbol{\pi}$ is the unitary circulant permutation matrix (3.15).

The concept of circulant matrix can be easily generalized into that of a block-wise circulant matrix, or a periodically circulant matrix as defined in ref. [49]. Such a generalization of circulant matrices can be described in the form of $\mathcal{F}$-reproducible matrices 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 a fixed family of linear maps $\mathcal{F}$ formed by the set of permutations

$$\mathcal{F} = \left\{\boldsymbol{\sigma}_0 = \mathbf{I}_p, \boldsymbol{\sigma}_1 = \boldsymbol{\pi}^m, \boldsymbol{\sigma}_2 = \boldsymbol{\pi}^{2m}, \ldots, \boldsymbol{\sigma}_{\frac{p}{m}-1} = \boldsymbol{\pi}^{p-m}\right\}, \tag{5.5}$$

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

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z} \\ \mathbf{z}\boldsymbol{\pi}^m \\ \mathbf{z}\boldsymbol{\pi}^{2m} \\ \vdots \\ \mathbf{z}\boldsymbol{\pi}^{p-m} \end{bmatrix}. \tag{5.6}$$

These matrices are indeed block-wise circulant, 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

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

Based on Theorem 3.9, $\mathcal{M}_q^{\mathcal{F},m}$ is a semigroup with respect to the multiplication, and therefore a pseudo-ring. With this in mind, we can define the following object.

Definition 5.2. (BC-MDPC codes) Let $\mathcal{M}_q^{\mathcal{F},m}$ be the pseudo-ring formed by block-wise circulant matrices of the form (5.6). We call *Block-wise Cyclic MDPC* (*BC-MDPC*) *code of type* (r_0, n_0) a linear code of length $n = n_0 p$ and redundancy $r \leq r_0 p$ that admits a parity-check matrix in the form $\mathbf{H} = \{\mathbf{Z}_{ij}\}$, where $\mathbf{Z}_{ij} \in \mathcal{M}_q^{\mathcal{F},m}$ for all $0 \leq i \leq r_0 - 1$, $0 \leq j \leq n_0 - 1$, such that $\mathbf{H}$ has row weight $O(\sqrt{n})$.

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. Therefore, we expect that using BC-MDPC codes could hinder reaction attacks of the type introduced in ref. [29], which rely on such a property of circulant matrices.

Remark 5.3. Note that the above formulation of BC-MDPC codes could be made even more general. In fact, in Definition 5.2, these codes are described as made of blocks all coming from the same pseudo-ring $\mathcal{M}_q^{\mathcal{F},m}$. However, this is not strictly necessary to preserve a reproducible structure. One could in fact select block-wise circulant components with different reproducible orders, which would lead to a BC-MDPC code of reproducible order $m = \mathrm{lcm}(m_i)$. We believe that such a formulation could be an interesting avenue to investigate in future works.

6 Conclusion

We have introduced the notions of reproducibility and quasi-reproducibility. They capture the idea of matrices that can be compactly represented through a signature, i.e., a subset of rows, and a family of functions which generate all remaining rows. We have provided theoretical results about the existence and properties of these families of matrices, which only depend on the chosen family of transformations. Alongside, we have extended these notions to coding theory and have introduced the concept of

reproducible and quasi-reproducible codes, which are codes described by a generator or a parity-check matrix yielding a compact representation. We have shown that existing and well known families of structured codes are encompassed within this framework, and have provided some concrete constructions of other families of reproducible codes.

A direct application of this work is in code-based cryptography, where the representation of a code is commonly used as the public key. As the recent NIST call for the standardization of post-quantum cryptosystems clearly emphasizes, random and pseudo-random codes are of interest for many code-based cryptosystems. In particular, at the current state of the art, many systems rely on the quasi-cyclic structure of codes in order to reduce the public key size. Essentially, all the schemes employing such structured codes can be generalized to the use of reproducible codes, via some of the constructions we have shown in this article. While the compactness of the public key is preserved, advantages come from the fact that attacks targeting the specific quasi-cyclic structure can be avoided when more general code constructions are considered. Although a complete cryptanalysis of these new families of codes requires a deeper investigation, and is out of the scope of this article, these potential benefits motivate the study of reproducible codes as a generalization of quasi-cyclic and other known structured codes.

Acknowledgements: Edoardo Persichetti and Paolo Santini were supported by National Science Foundation (NSF) grant CNS-1906360.

Conflict of interest: Authors state no conflict of interest.

References

[1] Misoczki R, Barreto PSLM. Compact McEliece keys from Goppa codes. In: Selected Areas in Cryptography, Lecture Notes in Computer Science 5867. Springer Verlag; 2009. p. 376–92.
[2] McEliece RJ. A public-key cryptosystem based on algebraic coding theory. DSN Progress Report. 1978;4244:114–6.
[3] Gaborit P. Shorter keys for code based cryptography. In: Proceedings of the International Workshop on Coding and Cryptography (WCC 2005). Bergen, Norway; March 2005. p. 81–90.
[4] Monico C, Rosenthal J, Shokrollahi A. Using low density parity check codes in the McEliece cryptosystem. In: Proceedingsof the IEEE International Symposium on Information Theory (ISIT 2000) Sorrento, Italy; June 2000. p. 215.
[5] Misoczki R, Tillich JP, Sendrier N, Barreto PSLM. MDPC-McEliece: New McEliece variants from moderate density parity-check codes. In: 2013 IEEE International Symposium on Information Theory; July 2013. p. 2069–73.
[6] Baldi M. LDPC codes in the McEliece cryptosystem: attacks and countermeasures. NATO Science for Peace and Security Series - D: Information and Communication Security 23. IOS Press; 2009. p. 160–74.
[7] Shor PW. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM J Comput. 1997;26:1484–509.
[8] https://csrc.nist.gov/Projects/Post-Quantum-Cryptography.
[9] Bootland C, Castryck W, Szepieniec A, Vercauteren F. A framework for cryptographic problems from linear algebra. J Math Cryptol. 2019;14:202–17.
[10] Alagic C, Alperin-Sheriff J, Apon D, Cooper D, Dang Q, Liu YK, et al. Status report on the first round of the NIST post-quantum cryptography standardization process. Washington, DC: US Department of Commerce, National Institute of Standards and Technology; 2019.
[11] Berlekamp E, McEliece R, van Tilborg H. On the inherent intractability of certain coding problems. IEEE Trans Inform Theory 1978;24:384–86.
[12] Sidelnikov VM, Shestakov SO. On insecurity of cryptosystems based on generalized Reed–Solomon codes. Discr Math Appl. 1992;2:439–44.
[13] Faugère J-C, Otmani A, Perret L, Tillich J-P. A distinguisher for high rate McEliece cryptosystems. In: Proceedings of IEEE Information Theory Workshop (ITW). Paraty, Brazil; October 2011. p. 282–6.
[14] Gallager RG. Low-density parity-check codes. IRE Transactions on Information Theory. IEEE; 1963;8(1).
[15] Hofheinz D, Hövelmanns K, Kiltz E. A modular analysis of the Fujisaki-Okamoto transformation. In:Kalai Y, Reyzin L, editors. Theory of Cryptography. Cham: Springer International Publishing; 2017. p. 341–71.

[16] Baldi M, Barenghi A, Chiaraluce F, Pelosi G, Santini P. Failure rate model of bit-flipping decoders for QC-LDPC and QC-MDPC code-based cryptosystems. In: Proceedings of the 17th International Joint Conference on e-Business and Telecommunications - Volume 3: SECRYPT, INSTICC. SciTePress; 2020. p. 238–49.
[17] Prange E. The use of information sets in decoding cyclic codes. IRE Trans Inform Theory. 1962;8:5–9.
[18] Leon JS. A probabilistic algorithm for computing minimum weights of large error-correcting codes. IEEE Trans Inform Theory. 1988;34:1354–9.
[19] Stern J. A method for finding codewords of small weight. In: Coding Theory and Applications. Cohen G, Wolfmann J, editors. Lecture Notes in Computer Science 388. Springer Verlag; 1989. p. 106–13.
[20] May A, Meurer A, Thomae E. Decoding random linear codes in $O(2^{0.054n})$. ASIACRYPT, LNCS 7073. Springer; 2011. p. 107–24.
[21] Becker A, Joux A, May A, Meurer A. Decoding random binary linear codes in $2^{n/20}$: How $1 + 1 = 0$ improves information set decoding. In: Pointcheval D, Johansson T, editors. Advances in Cryptology – EUROCRYPT 2012, Lecture Notes in Computer Science 7237. Springer Verlag; 2012. p. 520–36.
[22] Grover LK. A fast quantum mechanical algorithm for database search. In: Proceedings of the 28th Annual ACM Symposium on the Theory of Computing. Philadephia, PA; May 1996. p. 212–9.
[23] Bernstein DJ. Grover vs. McEliece. In: PQCrypto. 2010.
[24] Baldi M, Bodrato M, Chiaraluce F. A new analysis of the McEliece Cryptosystem based on QC-LDPC codes. In: Security and Cryptography for Networks, Lecture Notes in Computer Science 5229. Springer Verlag; 2008. p. 246–62.
[25] Berger TP, Cayrel P-L, Gaborit P, Otmani A. Reducing key length of the McEliece cryptosystem. In: Progress in Cryptology - AFRICACRYPT 2009, Lecture Notes in Computer Science 5580. Springer Verlag; 2009. p. 77–97.
[26] Faugère J-C, Otmani A, Perret L, Tillich J-P. Algebraic cryptanalysis of McEliece variants with compact keys. In: EUROCRYPT 2010, Lecture Notes in Computer Science 6110. Springer Verlag; 2010. p. 279–98.
[27] https://bigquake.inria.fr/.
[28] Sendrier N. Decoding one out of many. In: Post-quantum cryptography. Yang BY, editor. Lecture Notes in Computer Science 7071. Springer Verlag; 2011. p. 51–67.
[29] Guo Q, Johansson T, Stankovski P. A key recovery attack on MDPC with CCA security using decoding errors. In: ASIACRYPT, LNCS 10031. Springer; 2016. p. 789–815.
[30] Baldi M, Barenghi A, Chiaraluce F, Pelosi G, Santini P. LEDAkem: A Post-quantum Key Encapsulation Mechanism Based on QC-LDPC Codes. In: 9th International Conference on Post-Quantum Cryptography. Fort Lauderdale, FL, USA: PQCrypto; April 9–11 2018. p. 3–24.
[31] Barreto PSLM, Gueron S, Gueneysu T, Misoczki R, Persichetti E, Sendrier N, et al. CAKE: code-based algorithm for key encapsulation. In: IMA International Conference on Cryptography and Coding. Springer; 2017. p. 207–26.
[32] Tillich J-P. The decoding failure probability of MDPC codes. In:2018 IEEE International Symposium on Information Theory (ISIT), IEEE; 2018. p. 941–5.
[33] Santini P, Battaglioni M, Baldi M, Chiaraluce F. Analysis of the error correction capability of LDPC and MDPC codes under parallel bit-flipping decoding and application to cryptography. IEEE Trans Commun. 2020;68:4648–60.
[34] Santini P, Baldi M, Cancellieri G, Chiaraluce F. Hindering reaction attacks by using monomial codes in the McEliece cryptosystem. In:2018 IEEE International Symposium on Information Theory (ISIT), IEEE; 2018. p. 951–5.
[35] Baldi M, Bianchi M, Chiaraluce F. 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.
[36] Householder AS. Unitary triangularization of a nonsymmetric matrix. J ACM. 1958;5:339–42.
[37] Aragon N, Barreto PSLM, Bettaieb S, Bidoux L, Blazy O, Deneuville Jc. BIKE: Bit flipping key encapsulation; 2017.
[38] Fabšič T, Hromada V, Stankovski P, Zajac P, Guo Q, Johansson T. A reaction attack on the QC-LDPC McEliece cryptosystem. In:Post-Quantum Cryptography, LNCS 10346. Cham: Springer; 2017. p. 51–68.
[39] Fabsic T, Hromada V, Zajac P. A reaction attack on LEDApkc. IACR Cryptol ePrint Archive. 2018;2018:140.
[40] Eaton E, Lequesne M, Parent A, Sendrier N. QC-MDPC: A timing attack and a CCA2 KEM. In: PQCrypto. Cham: Springer; 2018. p. 47–76.
[41] CantoTorres R, Sendrier N. Analysis of information set decoding for a sub-linear error weight. Cham: Springer International Publishing; 2016. p. 144–61.
[42] Melchor CA, Aragon N, Bettaieb S, Bidoux L, Blazy O, Deneuville Jc, et al. HQC: Hamming Quasi Cyclic. 2017.
[43] Aguilar C, Gaborit P, Schrek J. A new zero-knowledge code based identification scheme with reduced communication. In:2011 IEEE Information Theory Workshop (ITW). Paraty, Brazil; Oct 2011. p. 648–52.
[44] Persichetti E. Compact McEliece keys based on quasi-dyadic Srivastava codes. J Math Cryptol 2012;6:149–69.
[45] Banegas G, Barreto PSLM, Boidje BO, Cayrel P-L, Dione K, Gaj GN, et al. DAGS: Key encapsulation using dyadic GS codes. J Math Cryptol. 2018;12:221–39.
[46] Banegas G, Barreto PSLM, Boidje BO, Cayrel P-L, Dione K, Gaj GN, et al. Dags: Reloaded revisiting dyadic key encapsulation. In: Code-Based Cryptography Workshop. Springer; 2019. p. 69–85.

[47] Barelli E, Couvreur A. An efficient structural attack on NIST submission DAGS. In: International Conference on the Theory and Application of Cryptology and Information Security. Springer; 2018. p. 93–118.
[48] Banegas G, Barreto PSLM, Persichetti E, Santini P. Designing efficient dyadic operations for cryptographic applications. J Math Cryptol. 2020;14:95–109.
[49] Battaglioni M, Chiaraluce F, Baldi M, Lentmaier M. Girth analysis and design of periodically time-varying SC-LDPC codes. IEEE Trans Inform Theor. 2021;67(4):2217–35.