The Hardness of Code Equivalence over F_q and its Application to Code-based Cryptography

Nicolas Sendrier, Dimitrios E. Simos

The Hardness of Code Equivalence over $\mathbb{F}_q$ and its Application to Code-based Cryptography

Nicolas Sendrier and Dimitris E. Simos

INRIA Paris-Rocquencourt
Project-Team SECRET
78153 Le Chesnay Cedex, France
{nicolas.sendrier,dimitrios.simos}@inria.fr

Abstract. The code equivalence problem is to decide whether two linear codes over $\mathbb{F}_q$ are identical up to a linear isometry of the Hamming space. In this paper, we review the hardness of code equivalence over $\mathbb{F}_q$ due to some recent negative results and argue on the possible implications in code-based cryptography. In particular, we present an improved version of the three-pass identification scheme of Girault and discuss on a connection between code equivalence and the hidden subgroup problem.

Keywords: Code Equivalence, Isometry, Hardness, Zero-Knowledge Protocols, Quantum Fourier Sampling, Linear Codes

1 Introduction

The purpose of this work is to examine the applications of the worst-case and average-case hardness of the Code Equivalence problem to the field of code-based cryptography. The latter problem is, given the generator matrices of two q-ary linear codes, how hard is it to decide whether or not these codes are identical up to an isometry of Hamming space? The support splitting algorithm ($\mathcal{SSA}$) [1] runs in polynomial time for all but a negligible proportion of all linear codes, and solves the latter problem by recovering the isometry when it is just a permutation of the code support.

The McEliece public-key cryptosystem [2] and Girault's zero-knowledge protocol [3], both candidates for post-quantum cryptography, are related to the hardness of permutationally equivalent linear codes. For the McEliece cryptosystem, the $\mathcal{SSA}$ is able to detect some weak keys but a polynomial attack is infeasible due to the large number of possible private keys. However, the security of Girault's zero-knowledge protocol is severely weakened and cannot longer be used with random codes but only with weakly self-dual codes (the hard instances of $\mathcal{SSA}$).

Recently in [4], the worst-case and average-case hardness of code equivalence over $\mathbb{F}_q$ was studied and it was shown that in practice, $\mathcal{SSA}$ could be extended for $q \in \{3, 4\}$, and similarly solve all but an exponentially small proportion of the instances in polynomial time, when isometries are under consideration. However, for any fixed $q \geq 5$, the problem seems to be intractable for almost all instances.

In light of these new results, we repair Girault's zero-knowledge protocol over $\mathbb{F}_q$, when $q \geq 5$, by showing that random codes are again a viable option. Moreover, the context of the framework built in [5] suggests that codes with large automorphism groups resist quantum Fourier sampling as long as permutation equivalence is considered. We examine whether it is possible to extend these results, when a more general notion of code equivalence over $\mathbb{F}_q$ is taken into account, in particular when the equivalence mapping is an isometry and not just a permutation of the code support.

The paper is structured as follows. In section 2 we define the different notions of equivalence of linear codes over $\mathbb{F}_q$ when isometries are considered, while in section 3 we formally define the Code Equivalence problem and present a thorough analysis of its hardness. In section 4 we review the protocol of Girault together with its weakness and repair its security using results based on the hardness of code equivalence, while in the last section we elaborate on the connection between code equivalence over $\mathbb{F}_q$ and the quantum Fourier sampling.

2 Equivalence of Linear Codes over $\mathbb{F}_q$

Code equivalence is a basic concept in coding theory with several applications in code-based cryptography; the McEliece public-key cryptosystem [2], Girault's identification scheme [3] and the CFS signature scheme [6], to name a few. The notion of equivalence of linear codes used in code-based cryptography usually involves only permutations as the code alphabet is the binary field. However, this is by far the case in coding theory where for a more general notion of equivalence all isometries of the Hamming space have to be included. In this section, we review the concept of what it means for codes to be "essentially different" by considering the metric Hamming space together with its isometries, which are the maps preserving the metric structure. This in turn will lead to a rigorous definition of equivalence of linear codes and as we shall see later on may provide additional applications in cryptography. In fact, we will call codes isometric if they are equivalent as subspaces of the Hamming space.

2.1 Three Notions of Equivalence

Let $\mathbb{F}_q$ be a finite field of cardinality $q = p^r$, where the prime number p is its characteristic, and r is a positive integer. As usual, a linear $[n, k]$ code C is a k-dimensional subspace of the finite vector space $\mathbb{F}_q^n$ and its elements are called codewords. We consider all vectors, as row vectors. Therefore, an element v of $\mathbb{F}_q^n$ is of the form $v := (v_1, \ldots, v_n)$. It can also be regarded as the mapping v from the set $\mathcal{I}_n = \{1, \ldots, n\}$ to $\mathbb{F}_q$ defined by $v(i) := v_i$. The Hamming distance (metric) on $\mathbb{F}_q^n$ is the following mapping,

$$d : \mathbb{F}_q^n \times \mathbb{F}_q^n \to \mathbb{N} : (x, y) \mapsto d(x, y) := \mid \{i \in \{1, 2, \ldots, n\} \mid x_i \neq y_i\} \mid .$$

The pair $(\mathbb{F}_q^n, d)$ is a metric space, called the Hamming space of dimension n over $\mathbb{F}_q$, denoted by $H(n, q)$. The Hamming weight $w(x)$ of a codeword $x \in C$ is simply the number of its non-zero coordinates, i.e. $w(x) := d(x, 0)$.

It is well-known due to a theorem of MacWilliams that any isometry between linear codes preserving the weight of the codewords induces an equivalence for codes [7]. Therefore, two codes C, C' are of the same quality if there exists a mapping $\iota : \mathbb{F}_q^n \mapsto \mathbb{F}_q^n$ with $\iota(C) = C'$ which preserves the Hamming distance, i.e. $d(\upsilon, \upsilon') = d(\iota(\upsilon), \iota(\upsilon'))$, for all $\upsilon, \upsilon' \in \mathbb{F}_q^n$. Mappings with the latter property are called the isometries of $H(n, q)$, and the two codes C and C' will be called isometric. Clearly, isometric codes have the same error-correction capabilities, and obvious permutations of the coordinates are isometries. We write $\mathcal{S}_n$ for the symmetric group acting on the set $\mathcal{I}_n$, equipped with the composition of permutations.

Definition 1. *Two linear codes $C, C' \subseteq \mathbb{F}_q^n$ will be called permutationally equivalent*[1]*, and will be denoted as $C \overset{\mathbf{PE}}{\sim} C'$, if there exists a permutation $\sigma \in \mathcal{S}_n$ that maps C onto C', i.e. $C' = \sigma(C) = \{\sigma(x) \mid x = (x_1, \ldots, x_n) \in C\}$ where $\sigma(x) = \sigma(x_1, \ldots, x_n) := (x_{\sigma^{-1}(1)}, \ldots, x_{\sigma^{-1}(n)})$.*

Note also that the use of σ^{-1} in the index is consisted as we have $\sigma(\pi(C)) = \sigma \circ \pi(C)$. This can easily be seen by considering $x \in C$, and $\sigma, \pi \in \mathcal{S}_n$ such that $\sigma(\pi(x)) = \sigma((x_{\pi^{-1}(i)})_{i \in \mathcal{I}_n})$. Let $y_i = x_{\pi^{-1}(i)}$, $i \in \mathcal{I}_n$. Then $\sigma(\pi(x)) = \sigma((y_i)_{i \in \mathcal{I}_n}) = (y_{\sigma^{-1}(i)})_{i \in \mathcal{I}_n} = (x_{\pi^{-1}\sigma^{-1}(i)})_{i \in \mathcal{I}_n} = (x_{(\sigma\pi)^{-1}(i)})_{i \in \mathcal{I}_n} = \sigma \circ \pi(x)$.

Moreover, there is a particular subgroup of $\mathcal{S}_n$ that maps C onto itself, the permutation group of C defined as $\mathrm{PAut}(C) := \{C = \sigma(C) \mid \sigma \in \mathcal{S}_n\}$. $\mathrm{PAut}(C)$ always contains the identity permutation. If it does not contain any other element, we will say that it is trivial.

Recall, that we defined two codes to be isometric if there exists an isometry that maps one into another. Isometries that are linear[2], are called linear isometries. Therefore, we can obtain a more general notion of equivalence for codes induced by linear isometries of $\mathbb{F}_q$. Moreover, it can be shown that any linear isometry between two linear codes $C, C' \subseteq \mathbb{F}_q^n$ can always be extended to an isometry of $\mathbb{F}_q^n$ [8].

The group of all linear isometries of $H(n, q)$ corresponds to the semidirect product of $\mathbb{F}_q^{*n}$ and $\mathcal{S}_n$, $\mathbb{F}_q^{*n} \rtimes \mathcal{S}_n = \{(\upsilon; \pi) \mid \upsilon : \mathcal{I}_n \mapsto \mathbb{F}_q^*, \pi \in \mathcal{S}_n\}$, called the monomial group of degree n over $\mathbb{F}_q^*$, where the multiplication within this group is defined by

$$(\upsilon; \pi)(\upsilon'; \pi') = (\upsilon\upsilon'_\pi, \pi\pi') \quad \text{and} \quad (\upsilon\upsilon'_\pi)_i := \upsilon_i \upsilon'_{\pi^{-1}(i)} \tag{1}$$

where $\mathbb{F}_q^*$ denotes the multiplicative group of $\mathbb{F}_q$. Hence, any linear isometry ι can be expressed as a pair of mappings $(\upsilon; \pi) \in \mathbb{F}_q^{*n} \rtimes \mathcal{S}_n$. Note that, some authors [8–10], describe this group as the wreath product $\mathbb{F}_q^* \wr_n \mathcal{S}_n$. The action of the latter group in an element of $\mathbb{F}_q^n$ is translated into an equivalence for linear codes.

[1] This definition can also met as permutationally isometric codes in the literature, see [8].

[2] For all $u, v \in \mathbb{F}_q^n$ we have $\iota(u + v) = \iota(u) + \iota(v)$ and $\iota(0) = 0$.

Definition 2. *Two linear codes $C, C' \subseteq \mathbb{F}_q^n$ will be called linearly or monomially equivalent, and will be denoted as $C \overset{\mathbf{LE}}{\sim} C'$, if there exists a linear isometry $\iota = (v;\sigma) \in \mathbb{F}_q^{*n} \rtimes \mathcal{S}_n$ that maps C onto C', i.e. $C' = (v;\sigma)(C) = \{(v;\sigma)(x) \mid (x_1,\ldots,x_n) \in C\}$ where $(v;\sigma)(x_1,\ldots,x_n) := (v_1 x_{\sigma^{-1}(1)},\ldots,v_n x_{\sigma^{-1}(n)})$.*

If $q = p^r$ is not a prime, then the Frobenius automorphism $\tau : \mathbb{F}_q \to \mathbb{F}_q, x \mapsto x^p$ applied on each coordinate of $\mathbb{F}_q^n$ preserves the Hamming distance, too. Moreover, for $n \geq 3$, the isometries of $\mathbb{F}_q^n$ which map subspaces onto subspaces are exactly the semilinear mappings[3] of the form $(v;(\alpha,\pi))$, where $(v;\pi)$ is a linear isometry and α is a field automorphism, i.e. $\alpha \in \mathrm{Aut}(\mathbb{F}_q)$ (c.f. [8, 11]). All these mappings form the group of semilinear isometries of $H(n,q)$ which is isomorphic to the semidirect product $\mathbb{F}_q^{*n} \rtimes (\mathrm{Aut}(\mathbb{F}_q) \times \mathcal{S}_n)$, where the multiplication of elements is given by

$$(v;(\alpha,\pi))(\varphi;(\beta,\sigma)) := (v \cdot \alpha(\varphi_\pi);(\alpha\beta,\pi\sigma)) \tag{2}$$

Moreover, there is a description of $\mathbb{F}_q^{*n} \rtimes (\mathrm{Aut}(\mathbb{F}_q) \times \mathcal{S}_n)$ as a generalized wreath product $\mathbb{F}_q^* \wr_n (\mathrm{Aut}(\mathbb{F}_q) \times \mathcal{S}_n)$, see [8, 12, 11]. Clearly, the notion of semilinear isometry which can be expressed as a group action on the set of linear subspaces gives rise to the most general notion of equivalence for linear codes.

Definition 3. *Two linear codes $C, C' \subseteq \mathbb{F}_q^n$ will be called semilinearly equivalent, and will be denoted as $C \overset{\mathbf{SLE}}{\sim} C'$, if there exists a semilinear isometry $(v;(\alpha,\sigma)) \in \mathbb{F}_q^{*n} \rtimes (\mathrm{Aut}(\mathbb{F}_q) \times \mathcal{S}_n)$ that maps C onto C', i.e. $C' = (v;(\alpha,\sigma))(C) = \{(v;(\alpha,\sigma))(x) \mid (x_i)_{i\in\mathcal{I}_n} \in C\}$ where $(v;(\alpha,\sigma))(x_1,\ldots,x_n) = (v_1\alpha(x_{\sigma^{-1}(1)}), \ldots, v_n\alpha(x_{\sigma^{-1}(n)}))$.*

Finally, we can define the monomial group of C as $\mathrm{MAut}(C) := \{C = (v;\sigma)(C) \mid (v;\sigma) \in \mathbb{F}_q^{*n} \rtimes \mathcal{S}_n\}$ and the automorphism group of C as $\mathrm{Aut}(C) := \{C = (v;(\alpha,\sigma))(C) \mid (v;(\alpha,\sigma)) \in \mathbb{F}_q^{*n} \rtimes (\mathrm{Aut}(\mathbb{F}_q) \times \mathcal{S}_n)\}$ where their elements map each codeword of C to another codeword of C, under the respective actions of the involved groups. For more details, on automorphism groups of linear codes we refer to [13]. In addition, we remark the following:

1. When $\mathbb{F}_q = \mathbb{F}_2$ the group of linear isometries of $H(n,2)$ is isomorphic to $\mathcal{S}_n$, therefore all notions of equivalence are the same.
2. The group of semilinear isometries of $H(n,q)$ is the same as the group of linear isometries if and only if q is a prime (since $\mathrm{Aut}(\mathbb{F}_q)$ is trivial if and only if q is a prime). Therefore, semilinear equivalence reduces to linear equivalence for prime fields, and is different for all other cases.

3 The Code Equivalence Problem

For efficient computation of codes we represent them with generator matrices. A $k \times n$ matrix G over $\mathbb{F}_q$, is called a generator matrix for the $[n,k]$ linear code

[3] $\sigma : \mathbb{F}_q^n \to \mathbb{F}_q^n$ is semilinear if there exists $\alpha \in \mathrm{Aut}(\mathbb{F}_q)$ such that for all $u, v \in \mathbb{F}_q^n$ and $k \in \mathbb{F}_q$ we have $\sigma(u+v) = \sigma(u) + \sigma(v)$ and $\sigma(ku) = \alpha(k)\sigma(u)$.

C if the rows of G form a basis for C, so that $C = \{xG \mid x \in \mathbb{F}_q^k\}$. In general, a linear code possess many different bases, and it is clear from linear algebra that the set of all generator matrices for C can be reached by $\{SG \mid S \in \mathrm{GL}_k(q)\}$, where $\mathrm{GL}_k(q)$ is the group of all $k \times k$ invertible matrices over $\mathbb{F}_q$.

For any $\sigma \in \mathcal{S}_n$ associate by $P_\sigma = [p_{i,j}]$ the $n \times n$ matrix such that $p_{i,j} = 1$ if $\sigma(i) = j$ and $p_{i,j} = 0$ otherwise, therefore P_σ is a permutation matrix. Note that, the action of $\sigma \in \mathcal{S}_n$ on $x \in \mathbb{F}_q^n$ agrees with the ordinary matrix multiplication. The permutation matrices form a subgroup of $M_n(q)$, the set of all $n \times n$ monomial matrices over $\mathbb{F}_q$, that is, matrices with exactly one nonzero entry per row and column from $\mathbb{F}_q$. If $M = [m_{i,j}] \in M_n(q)$, then $M = DP$, where P is a permutation matrix and $D = [d_{i,j}] = \mathrm{diag}(d_1, \ldots, d_n)$ is a diagonal matrix with $d_i = d_{i,i} = m_{i,j}$ if $m_{i,j} \neq 0$ and $d_{i,j} = 0$ if $i \neq j$. There is an isomorphism between diagonal matrices and $\mathbb{F}_q^{*n}$, therefore we associate $D_\upsilon = \mathrm{diag}(\upsilon_1, \ldots, \upsilon_n)$ for $\upsilon = (\upsilon_i)_{i \in \mathcal{I}_n} \in \mathbb{F}_q^{*n}$. Hence, we can map any linear isometry $(\upsilon; \sigma) \in \mathbb{F}_q^{*n} \rtimes \mathcal{S}_n$ to a monomial matrix $M_{(\upsilon;\sigma)} = D_\upsilon P_\sigma \in M_n(q)$, and this mapping is an isomorphism between $\mathbb{F}_q^{*n} \rtimes \mathcal{S}_n$ and $M_n(q)$. Therefore, we can express the equivalence between linear codes in terms of their generator matrices.

Problem 1. Given two $k \times n$ matrices G and G' over $\mathbb{F}_q$, whose rows span two $[n, k]$ linear codes C and C' over $\mathbb{F}_q$, does there exist $S \in \mathrm{GL}_k(q)$ and a monomial matrix $M_{(\upsilon;\sigma)} = D_\upsilon P_\sigma \in M_n(q)$ such that $G' = SGD_\upsilon P_\sigma$?

We will refer to the decidability of the previous problem, as the Linear Code Equivalence problem. The Semi-Linear Code Equivalence problem can be defined analogously by permitting the application of a field automorphism in the columns of the scrambled generator matrix. In particular, we define the following problem.

Problem 2. Given two $k \times n$ matrices G and G' over $\mathbb{F}_q$, whose rows span two $[n, k]$ linear codes C and C' over $\mathbb{F}_q$, does there exist $S \in \mathrm{GL}_k(q)$, a monomial matrix $M_{(\upsilon;\sigma)} = D_\upsilon P_\sigma \in M_n(q)$ and a field automorphism $\alpha \in \mathrm{Aut}(\mathbb{F}_q)$ such that $G' = S\alpha(GD_\upsilon P_\sigma)$?

3.1 The Hardness of Code Equivalence over $\mathbb{F}_q$

In this section, we review the hardness of the code equivalence problem, therefore we deem necessary to briefly mention the most significant results in terms of complexity, for deciding it, and algorithms, for solving it.

When the linear isometry $(\upsilon; \sigma)$ is just a permutation, i.e. D_υ is equal to I_n, we will call problem 1, as the Permutation Code Equivalence problem. The latter problem, was introduced in [14], who showed that if $\mathbb{F}_q = \mathbb{F}_2$ then it is harder than the Graph Isomorphism, there exists a polynomial time reduction, but not NP-complete unless P = NP. A different proof of this reduction is also given in [11]. Recently, the reduction of [14] was generalized in [15] over any field $\mathbb{F}_q$, hence Permutation Code Equivalence is harder than the Graph Isomorphism, for any field $\mathbb{F}_q$. The latter problem, has been extensively studied

for decades, but until now there is no polynomial-time algorithm for solving all of its instances. Clearly, (Semi)-Linear Code Equivalence for any $\mathbb{F}_q$ cannot be easier than the Graph Isomorphism, since it contains the Permutation Code Equivalence as a subproblem.

Last but not least, we would like to mention that the McEliece public-key cryptosystem [2] is related to the hardness of permutationally equivalent binary linear codes. Towards this direction, another important complexity result was shown in [5], that the Hidden Subgroup problem also reduces to Permutation Code Equivalence for any field $\mathbb{F}_q$.

The Support Splitting Algorithm can be used as an oracle to decide whether two binary codes are permutationally equivalent [1], as well as to retrieve the equivalence mapping. Other notable algorithms for code equivalence can be found in [16–18]. The main idea of $\mathcal{SSA}$ is to partition the support $\mathcal{I}_n$ of a code $C \subseteq \mathbb{F}_2^n$, into small sets that are fixed under operations of $\mathrm{PAut}(C)$. The algorithm employs the concept of invariants and signatures, defined in [1]. Invariants are mappings such that any two permutationally equivalent codes take the same value, while signatures depends on the code and one of its positions.

Definition 4. *A signature S over a set F maps a code $C \subseteq \mathbb{F}_q^n$ and an element $i \in \mathcal{I}_n$ into an element of F and is such that for all $\sigma \in \mathcal{S}_n$, $S(C,i) = S(\sigma(C), \sigma(i))$. Moreover, S is called discriminant for C if there exist $i, j \in \mathcal{I}_n$ such that $S(C,i) \neq S(C,j)$ and fully discriminant if this holds $\forall\ i, j \in \mathcal{I}_n$.*

The fundamental idea of $\mathcal{SSA}$ is to be able to find a distinct property for the code and one of its positions, and thus by labeling them accordingly it is possible to recover the permutation between equivalent codes.

The main difficulty of the algorithm, is to obtain a fully discriminant signature, for as many codes as possible. In [1] it was shown that such a signature, can be built from the weight enumerator of the hull of a code C, denoted by $\mathcal{H}(C)$, and defined as the intersection of the code with its dual, $\mathcal{H}(C) = C \cap C^\perp$ [19], because the hull commutes with permutations[4], $\mathcal{H}(\sigma(C)) = \sigma(\mathcal{H}(C))$, and therefore is an invariant for permutation equivalence. The (heuristic) complexity of $\mathcal{SSA}$ for an $[n,k]$ code C is $\mathcal{O}(n^3 + 2^h n^2 \log n)$ where h is the dimension of the hull $[20, 1]$. The first term is the cost of the Gaussian elimination needed to compute the hull. The second term is the (conjectured) number of refinements, $\log n$, multiplied by the cost one refinement (n weight enumerators of codes of dimension h and length n). Moreover, the cost of computing the weight enumerator of an $[n,h]$ code over $\mathbb{F}_q$ is proportional to nq^h operations in $\mathbb{F}_q$ [1].

In practice, for random codes, the hull has a small dimension with overwhelming probability [21] and the dominant cost for the average case is $\mathcal{O}(n^3)$. Note that, the worst case occurs when the hull dimension is maximal; weakly self-dual codes ($C \subset C^\perp$) are equal to their hulls. Then the algorithm becomes intractable with a complexity equal to $\mathcal{O}(2^k n^2 \log n)$.

[4] No such property exists in general for linear codes when (semi)-linear equivalence is considered, because the dual of equivalent codes do not remain equivalent with the same isometry as the original codes.

Reduction of Linear Code Equivalence to Permutation Code Equivalence was made possible via the introduction of the closure of a linear code in [4].

Definition 5. *Let $\mathbb{F}_q = \{a_0, a_1, \ldots, a_{q-1}\}$, with $a_0 = 0$, and a linear code $C \subseteq \mathbb{F}_q^n$. Define $\mathcal{I}_{q-1}^{(n)}$ as the cartesian product of $\mathcal{I}_{q-1} \times \mathcal{I}_n$. The closure $\widetilde{C}$ of the code C is a code of length $(q-1)n$ over $\mathbb{F}_q$ where,*

$$\widetilde{C} = \{(a_k x_i)_{(k,i)\in\mathcal{I}_{q-1}^{(n)}} \mid (x_i)_{i\in\mathcal{I}_n} \in C\}.$$

Clearly, we see that every coordinate of the closure $\widetilde{C}$, corresponds to a coordinate position of a codeword of C multiplied by a nonzero element of $\mathbb{F}_q$. Since, the index $(k, i) \in \mathcal{I}_{q-1}^{(n)}$ of a position of a codeword of the closure means that $k \in \mathcal{I}_{q-1}$ and $i \in \mathcal{I}_n$, we have taken into account every possible multiplication of x_i with nonzero elements of $\mathbb{F}_q$. The fundamental property of the closure is realised in the following theorem, first given in [4].

Theorem 1. *Let $C, C' \subseteq \mathbb{F}_q^n$. Then C and C' are linearly equivalent, i.e. $C \overset{\mathbf{LE}}{\sim} C'$, if and only if $\widetilde{C}$ and $\widetilde{C'}$ are permutationally equivalent, i.e. $\widetilde{C} \overset{\mathbf{PE}}{\sim} \widetilde{C'}$.*

Theorem 1 is of great importance, because it realizes a reduction from the Linear Code Equivalence problem to the Permutation Code Equivalence problem. Thus, we are able to decide if the codes C and C' are linearly equivalent by checking their closures for permutation equivalence. Moreover, if the closures are permutation equivalent then there exists an algorithmic procedure that allows the retrieval of the initial isometry between C and C' by considering that a signature for an extension of $\mathcal{SSA}$ can be built from the weight enumerator of the $\mathcal{H}(\widetilde{C})$.

Unfortunately, it turns out that the closure $\widetilde{C}$ is a weakly self-dual code for every $q \geq 5$, considering both Euclidean and Hermitian duals, which are exactly the hard instances of $\mathcal{SSA}$ [4]. Moreover, for $\mathbb{F}_3$ and $\mathbb{F}_4$ equipped with the Euclidean and Hermitian inner product, respectively, the distribution of the dimension of $\mathcal{H}(\widetilde{C})$ follows the distribution of the dimension $\mathcal{H}(C)$, since the closure has the same dimension as C, and will be on average a small constant, [21], except in the cases where C is also a weakly self-dual code. Therefore, the Linear Code Equivalence problem can be decided (and solved) in polynomial time using $\mathcal{SSA}$ only in $\mathbb{F}_3$ and $\mathbb{F}_4$, as long as the hull of the given code is small (the worst-case being a weakly self-dual code). However, for $q \geq 5$ its complexity growth becomes exponential for all instances. Moreover, it was conjectured in [4], that for $q \geq 5$, Code Equivalence is hard for almost all instances. This argument, was further supported by some impossibility results on the Tutte polynomial of a graph which corresponds to the weight enumerator of a code [22]. To conclude with, we would like to make clear that the hardness of the code equivalence arises from the absence of an easy computable invariant not the inexistence of an algorithm.

Table 1. Complexity estimates for $\mathcal{SSA}$ and its extension over $\mathbb{F}_q$

Algorithm	Field (alphabet)	Random codes (average-case)	Weakly self-dual codes (worst-case)
$\mathcal{SSA}$	$\mathbb{F}_2$	$\mathcal{O}(n^3)$	$\mathcal{O}(2^k n^2 \log n)$
$\mathcal{SSA}$ extension	$\mathbb{F}_3$	$\mathcal{O}(n^3)$	$\mathcal{O}(3^k n^2 \log n)$
$\mathcal{SSA}$ extension	$\mathbb{F}_4$	$\mathcal{O}(n^3)$	$\mathcal{O}(2^{2k} n^2 \log n)$
$\mathcal{SSA}$ extension	$\mathbb{F}_q$, $q \geq 5$	$\mathcal{O}(q^k n^2 \log n)$	$\mathcal{O}(q^k n^2 \log n)$

4 Zero-Knowledge Protocols

A central concept in cryptography is zero-knowledge protocols. These protocols allow a prover to convince a verifier that it knows a secret without the verifier learning any information about the secret. In practice, this is used to allow one party to prove its identity to another by proving it has a particular secret. For a protocol to be zero-knowledge, no information can be revealed no matter what strategy a so-called cheating verifier, simply cheater, follows when interacting with the prover. Therefore, an important question is what happens to these protocols when the cheater is a quantum computer. Are there any zero-knowledge protocols sufficient to withstand such a powerful cheater in a post-quantum era?

In this section, we deal with protocols based on a particular type of alternative cryptography originating from error-correcting codes, called code-based cryptography. In this emerging field of cryptography the underlying hard problems which pose as its security assumptions, decoding in a random linear code and recovering the code structure, does not seem so far to be susceptible to attacks mounted by quantum computers [20]. In addition, as we shall mention in the following section there is a negative result regarding the connection between coding theory and the Hidden Subgroup problem, which is the starting point for designing efficient quantum algorithms.

The idea of using error-correcting codes for identification schemes is due to Harari [23], followed by Stern (first protocol) [24] and Girault [3]. Harari's protocol was broken and the security of Girault's one was severely weakened (we shall explain this shortly after) while the protocol of Stern was five-pass and unpractical. At Crypto'93, Stern proposed a new scheme [25], which is one of the main references in this area. Recently, there has been an upsurge on designing identification schemes mainly due to the work of several researchers [26–28], where their efforts concentrated on both reducing the communication cost and the probabibility of someone impersonating an honest prover.

4.1 Girault's Three-Pass Identification Scheme

Girault's identification scheme is a three-pass one with a cheating probability of 1/2 (compared to the usual 2/3 of Stern's protocol), and has the additional advantage that all computations are performed on the standard model instead of the random oracle model since there is no involmement of a hash function in the

committments of the protocol. However, this advantage comes with a cost. At each round of the protocol a large number of bits has to be transmitted, which render the scheme unpractical. Its principle is as follows: Let H be an $(n-k)\times n$ matrix over the binary field $\mathbb{F}_2$ common to all users. Each prover $\mathcal{P}$ has an n-bit word e of small weight w randomly chosen by him and a public identifier $He = s$. Clearly, when H is a parity-check matrix of a linear code, computing e from H and s comes to finding a word of given small weight and given syndrome s, a well-known NP-hard problem. When $\mathcal{P}$ needs to authenticate to a verifier $\mathcal{V}$ as the owner of s, then $\mathcal{P}$ and $\mathcal{V}$ interact through the following scheme.

Step 1: $\mathcal{P}$ picks a random $n \times n$ permutation matrix P and a random $k \times k$ non-singular matrix S. P computes $H' = SHP$ and $s' = Ss$, and sends H' and s' to $\mathcal{V}$.
Step 2: $\mathcal{V}$ generates a random bit $c \in \{0, 1\}$ and sends it to $\mathcal{P}$.
Step 3a: If $c = 0$, $\mathcal{P}$ replies by delivering S, P to $\mathcal{V}$, who checks that $SHP = H'$ and $Ss = s'$.
Step 3b: If $c = 1$, $\mathcal{P}$ replies by delivering $e' = P^{-1}e$ to $\mathcal{V}$, who checks that the weight of e' is w and $H'e' = s'$.

The protocol is a multi-round one as it has to be repeated t times to reach a security level of $1 - (1/2)^t$ and was proved to be zero-knowledge on [3]. Its security is based on the hardness of two well-known problems in coding theory. The first one is the Binary Syndrome Decoding problem shown to be NP-complete in the worst case [29], but it is also widely believed that for the average case it still remains hard. The other assumption is related to the hardness of the Permutation Code Equivalence problem over the binary field since from the knowledge of H and H' someone must not be able to recover the scrambing matrix S and the permutation matrix P, as this would lead to information leakage about the secret key of $\mathcal{P}$. However, as we extensively discussed on section 3, $\mathcal{SSA}$ can recover the matrix P in (almost) polynomial time when the underlying code is chosen at random (see also the complexity figures in table 3), and then using elementary linear algebra the matrix S can also be found.

Still, the protocol can be used with weakly self-dual codes, the instances of Permutation Code Equivalence that the growth of $\mathcal{SSA}$ becomes exponential, however there is no significant advantage on decoding with self-dual codes and in addition this restrict too much the possibilities for the public key.

4.2 Improved Version of the Girault Protocol

We now consider, Girault's identification scheme in a q-ary setting. That is, the underlying finite field, will no longer be the binary field but the field $\mathbb{F}_q$ with q elements. For the security assumptions of the scheme we first have to consider syndrome decoding over $\mathbb{F}_q$. We define the decisional version of the q-ary Syndrome Decoding problem, below,

Problem 3. Given an $m \times n$ matrix H over $\mathbb{F}_q$, a target vector $s \in \mathbb{F}_q^m$ and an integer $w > 0$ does there exist a vector $x \in \mathbb{F}_q^n$ of weight $\leq w$ such that $Hx = s$?

which was also proven to be NP-complete in [30]. There are two main families of algorithms for solving the latter problem: Information Set Decoding (ISD) and (Generalized) Birthday algorithm (GBA). ISD has the lowest complexity of the two, and in a recent work [31] the complexity of a generalization of Stern's algorithm from [32] is analyzed which permits the decoding of linear codes over arbitrary finite fields $\mathbb{F}_q$. For a general treatment of the topic we refer to [20], while for the security of the scheme it is sufficient to consider that all known decoding attacks have an exponential cost on the code length.

Moreover, in an attempt to repair the security of the scheme we consider the (semi)-linear code equivalence instead of the permutation code equivalence, depending on whether $\mathbb{F}_q$ is a prime field or not. As one of the purposes of this paper, is to state the implications of the hardness of the (Semi)-Linear Code equivalence problem for designing cryptographic primitives, we choose the parameter q to be at least equal to 5, since we strongly believe that a random instance of the latter problem is hard for these cases (see also section 3).

The starting point of this improved version of Girault's scheme is the same as in the original one, with the expection that all operations now occur over $\mathbb{F}_q$, $q \geq 5$. Let H be an $(n-k) \times n$ matrix over $\mathbb{F}_q$ common to all users. Each prover $\mathcal{P}$ has an n-bit word e of small weight w randomly chosen by him and a public identifier $He = s$. As before, when a prover $\mathcal{P}$ needs to authenticate to a verifier $\mathcal{V}$ as the owner of s, then $\mathcal{P}$ and $\mathcal{V}$ interact through the following protocol.

Improved Version of Girault Identification Scheme

Key Generation: Random $[n, k]$ linear code with an $(n-k) \times n$ parity-check matrix H over $\mathbb{F}_q$
- **Private key:** A word $e \in \mathbb{F}_q^n$ of small weight w
- **Public key:** A public identifier $s \in \mathbb{F}_q^{n-k}$ such that $He = s$

Commitments:
- $\mathcal{P}$ picks a random $n \times n$ monomial matrix M, a random $k \times k$ non-singular matrix S and a field automorphism α of $\mathbb{F}_q$.
- $\mathcal{P}$ computes the commitments $s' = Ss$ and $H' = S\alpha(HM)$.
- $\mathcal{P}$ sends s' and H' to $\mathcal{V}$.

Challenge: $\mathcal{V}$ chooses randomly $c \in \{0, 1\}$ and sends it to $\mathcal{P}$.

Response:
- If $c = 0$ then $\mathcal{P}$ replies by delivering α, S, M to $\mathcal{V}$
- If $c = 1$ then $\mathcal{P}$ replies by delivering $e' = \alpha^{-1}(M^{-1}e)$ to $\mathcal{V}$

Verification:
- If $c = 0$ then $\mathcal{V}$ checks that $S\alpha(HM) = H'$ and $Ss = S'$.
- If $c = 1$ then $\mathcal{V}$ checks that the weight of e' is w and $H'e' = s'$.

The scheme is again a three-pass one and has to be repeated t times to reach a security level of $1-(1/2)^t$. The completeness, soundness and zero-knowledge of the scheme is a straight-forward verification of the proofs given by Girault in the original version [3], by replacing the permutation matrix P with the monomial matrix M and the field automorphism α (for non-prime fields) and therefore we avoid repeating them here to save space. Note that, the scheme is again usable in the standard model in contrast to the usual random oracle model.

We would like also to remark, that this q-ary version of Girault's scheme can be used again with the family of random linear codes for any field $\mathbb{F}_q$, $q \geq 5$. Moreover, we choose to commit the monomial matrix M instead of its (unique) factorization to a diagonal matrix D and a permutation matrix P (see also section 3) to reduce the (already) large cost of communication at each round (since we transmit matrices) as much as possible. A promising approach to circumvent this drawback could be to employ random structured codes as the public keys such as quasi-cyclic (QC) codes, similar to the work carried out in [28]. Although, there is no obvious advantage for an adversary mounting decoding attacks on QC codes, their rich structure may lead to structural attacks even when semi-linear code equivalence is considered (even though we are unaware of such kind of attacks) and a more careful analysis is required before proposing any specific parameters for the scheme.

5 A Note about Code Equivalence over $\mathbb{F}_q$ and Quantum Fourier Sampling

In [5], it was shown that permutation code equivalence over $\mathbb{F}_q$ has a direct reduction to a nonabelian Hidden Subgroup problem (HSP). It was further shown in the same paper that McEliece-type cryptosystems with certain conditions on the permutation automorphism groups of the underlying linear codes used as private keys, as is the case of rational Goppa codes, resist precisely the attacks to which the RSA and El Gamal cryptosystems are vulnerable, namely those based on generating and measuring coset states. This fact, eliminated the approach of strong Fourier sampling on which almost all known exponential speedups by quantum algorithms are based. In addition, these negative results have been extended in [33] for the case of Reed-Muller codes, which correspond to the particular case of the Sidelnikov cryptosystem.

There are two main questions arising from this framework: Whether there are any other families of codes suitable for cryptographic applications and what happens when we consider a more general notion of code equivalence over $\mathbb{F}_q$. We will investigate these matters, after briefly mentioning the conditions needed for the results of [5, 33].

Recall from [5] that a linear code C is HSP-hard if strong quantum Fourier sampling, reveals negligible information about the permutation $\sigma \in \mathcal{S}_n$ of permutationally equivalent codes, i.e. $C' = \sigma(C)$. Moreover, the support of a permutation $\sigma \in \mathcal{S}_n$ is the number of points that are not fixed by σ, and the minimal degree of a subgroup $H \leq \mathcal{S}_n$ is the smallest support of any non-identity $\pi \in H$.

Theorem 2 (Theorem 1, [33]). *Let C be a q-ary $[n, k]$ linear code such that $q^{k^2} \leq n^{0.2n}$. If $|\mathrm{PAut}(C)| \leq e^{o(n)}$ and the minimal degree of $|\mathrm{PAut}(C)|$ is $\Omega(n)$ then C is HSP-hard.*

We now consider $\mathbb{F}_q$ to be a prime field (hence $\mathrm{Aut}(\mathbb{F}_q)$ is trivial) and the monomial group $\mathrm{MAut}(C)$ of a code $C \subseteq \mathbb{F}_q^n$ for the notion of linear code equivalence.

Clearly, if the permutation part of $\mathrm{MAut}(C)$ satisfies the conditions of theorem 2, so does its closure $\widetilde{C}$ (see definition 5) which is a code of length $(q-1)n$ over the same field. Recall that two codes are linearly equivalent if and only if their closures are permutationally equivalent (c.f. theorem 1). In other words, the instances of codes that are HSP-hard for the Permutation Code Equivalence problem remain HSP-hard for the Linear Code Equivalence. This remark, would further imply that someone could design a McEliece-type cryptosystem by considering a monomial transformation of the private key instead of just a permutation without having to worry about attacks originating from the quantum Fourier sampling, based on rational Goppa codes over $\mathbb{F}_q$ for instance. However, we should note that these results apply only to high-rate $[n, k]$ codes over $\mathbb{F}_q$ (as $q^{k^2} \leq n^{0.2n}$ must be satisfied).

6 Conclusion

In this paper, we presented an analysis of the hardness of the Code Equivalence problem over $\mathbb{F}_q$ when the equivalence mapping is an isometry and not just a permutation of the code support. The hardness of the latter problem is of great importance when designing cryptographic primitives, such as public-key cryptosystems and identification schemes in the field of code-based cryptography. We stated the weaknesses of such an identification scheme (Girault's zero-knowledge protocol), and presented an improved version which relies on exactly these instances of the code equivalence that the problem is believed to be hard on average. Finally, we showed that some negative results regarding the possibility of attacking McEliece-type cryptosystems with quantum algorithms based on Fourier sampling apply also for other notions of code equivalence, besides the permutation equivalence, subject to certain conditions on the underlying family of codes used as private keys.

Acknowledgments. The work of the second author was carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme. This Programme is supported by the Marie Curie Co-funding of Regional, National and International Programmes (COFUND) of the European Commission.

References

1. Sendrier, N.: Finding the permutation between equivalent linear codes: The support splitting algorithm. IEEE Trans. Inform. Theory **26** (2000) 1193–1203
2. McEliece, R.J.: A public-key cryptosystem based on algebraic coding theory. Technical Report DSN Progress Report 42–44, California Institute of Technology, Jet Propulsion Laboratory, Pasadena, CA (1978)
3. Girault, M.: A (non-practical) three-pass identification protocol using coding theory. In Seberry, J., Pieprzyk, J., eds.: Advances in Cryptology AUSCRYPT '90. Volume 453 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (1990) 265–272

4. Sendrier, N., Simos, D.E.: How easy is code equivalence over $\mathbb{F}_q$? Preprint (2012)
5. Dinh, H., Moore, C., Russell, A.: Mceliece and niederreiter cryptosystems that resist quantum fourier sampling attacks. In: Proceedings of the 31st annual conference on Advances in cryptology. CRYPTO'11, Berlin, Heidelberg, Springer-Verlag (2011) 761–779
6. Courtois, N., Finiasz, M., Sendrier, N.: How to achieve a McEliece-based digital signature scheme. In Boyd, C., ed.: Advances in Cryptology - ASIACRYPT 2001. Volume 2248 of LNCS., Springer (2001) 157–174
7. MacWilliams, F.J.: Error-correcting codes for multiple-level transmission. Bell. Syst. Tech. J. **40** (1961) 281–308
8. Betten, A., Braun, M., Fripertinger, H., Kerber, A., Kohnert, A., Wassermann, A.: Error-Correcting Linear Codes: Classification by Isometry and Applications. Volume 18 of Algorithms and Computation in Mathematics. Springer, Berlin, Heidelberg (2006)
9. Fripertinger, H.: Enumeration of linear codes by applying methods from algebraic combinatorics. Grazer Math. Ber. **328** (1996) 31–42
10. Fripertinger, H., Kerber, A.: Isometry classes of indecomposable linear codes. In: Proceedings of the 11th International Symposium on Applied Algebra, Algebraic Algorithms and Error-Correcting Codes. AAECC-11, London, UK, Springer-Verlag (1995) 194–204
11. Kaski, P., Östergård, P.R.J.: Classification Algorithms for Codes and Designs. Volume 15 of Algorithms and Computation in Mathematics. Springer, Berlin, Heidelberg (2006)
12. Fripertinger, H.: Enumeration of the semilinear isometry classes of linear codes. Bayrether Mathematische Schriften **74** (2005) 100–122
13. Huffman, W.C.: Codes and groups. In Pless, V., Huffman, W.C., eds.: Handbook of Coding Theory. Elsevier, North-Holland, Amsterdam (1998) 1345–1440
14. Petrank, E., Roth, R.M.: Is code equivalence easy to decide? IEEE Trans. Inform. Theory **43** (1997) 1602–1604
15. Grochow, J.A.: Matrix lie algebra isomorphism. Technical Report TR11-168, Electronic Colloquium on Computational Complexity (2011) Also available as arXiv:1112.2012. To appear, IEEE Conference on Computational Complexity, 2012.
16. Babai, L., Codenotti, P., Grochow, J.A., Y.Qiao: Code equivalence and group isomorphism. In: Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms. SODA '11, SIAM (2011) 1395–1408
17. Bouyukliev, I.: About the code equivalence. Ser. Coding Theory Cryptol. **3** (2007) 126–151
18. Feulner, T.: The automorphism groups of linear codes and canonical representatives of their semilinear isometry classes. Adv. Math. Commun. **3** (2009) 363–383
19. Assmus, E.F.J., Key, J.D.: Designs and their Codes. Volume 103 of Cambridge Tracts in Mathematics. Cambridge University Press (1992) Second printing with corrections, 1993.
20. Overbeck, R., Sendrier, N.: Code-based cryptography. In Bernstein, D., Buchmann, J., Dahmen, E., eds.: Post-Quantum Cryptography. Springer (2009) 95–145
21. Sendrier, N.: On the dimension of the hull. SIAM J. Discrete Math. **10**(2) (1997) 282–293
22. Vertigan, D.: Bicycle dimension and special points of the Tutte polynomial. Journal of Combinatorial Theory, Series B **74** (1998) 378–396

23. Harari, S.: A new authentication algorithm. In Cohen, G., Wolfmann, J., eds.: Coding Theory and Applications. Volume 388 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (1989) 91–105
24. Stern, J.: An alternative to the fiat-shamir protocol. In Quisquater, J.J., Vandewalle, J., eds.: Advances in Cryptology EUROCRYPT 89. Volume 434 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (1990) 173–180
25. Stern, J.: A new identification scheme based on syndrome decoding. In: Proceedings of the 13th Annual International Cryptology Conference on Advances in Cryptology. CRYPTO '93, London, UK, UK, Springer-Verlag (1994) 13–21
26. Cayrel, P.L., Gaborit, P., Girault, M.: Identity-based identification and signature schemes using correcting codes. In Augot, D., Sendrier, N., Tillich, J.P., eds.: Workshop on Coding and Cryptography - WCC 2007, INRIA (2007) 69–78
27. Cayrel, P.L., Veron, P., Yousfi Alaoui, S.: A zero-knowledge identification scheme based on the q-ary syndrome decoding problem. In Biryukov, A., Gong, G., Stinson, D., eds.: Selected Areas in Cryptography. Volume 6544 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2011) 171–186
28. Aguilar, C., Gaborit, P., Schrek, J.: A new zero-knowledge code based identification scheme with reduced communication. In: Information Theory Workshop (ITW), 2011 IEEE. (2011) 648–652
29. Berlekamp, E., McEliece, R., van Tilborg, H.: On the inherent intractability of certain coding problems (corresp.). Information Theory, IEEE Transactions on **24** (1978) 384–386
30. Barg, S.: Some new NP-complete coding problems. Probl. Peredachi Inf. **30** (1994) 23–28
31. Peters, C.: Information-set decoding for linear codes over $\mathbb{F}_q$. In Sendrier, N., ed.: Post-Quantum Cryptography. Volume 6061 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2010) 81–94
32. Stern, J.: A method for finding codewords of small weight. In Cohen, G., Wolfmann, J., eds.: Coding Theory and Applications. Volume 388 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (1989) 106–113
33. Dinh, H., Moore, C., Russell, A.: Quantum fourier sampling, code equivalence, and the quantum security of the mceliece and sidelnikov cryptosystems. Technical report (2011) Also available as arXiv:1111.4382v1.