Suppressing Hidden Extension-Field Linearity in Rank-Metric Cryptography via Structural Incompatibility

Dengchuan Liao[1], Xiangxue Li[1], and Yu Yu[2]

[1]School of Software Engineering, School of Cryptology, ECNU
[2]Shanghai Qizhi Institute
xxli@cs.ecnu.edu.cn; yyuu@sjtu.edu.cn

Abstract. A prominent line of rank-metric code-based cryptography has long relied on highly structured algebraic code families, such as Gabidulin codes, for their optimal rank-distance properties and efficient decoding. However, this structure exposes algebraic invariants (most notably extension-field linearity and Frobenius invariance) that enable powerful polynomial-time distinguishers and effective key-recovery attacks. In this work, we revisit this structural tension from a new perspective. Rather than relying solely on masking, we identify a simple yet fundamental structural incompatibility that rules out the direct extension-field linear representation on which these attacks rely.

Building on this insight, we introduce *Enhanced Gabidulin Matrix Subcodes* (EnGMS), a family of masked matrix codes obtained from K'-dimensional $\mathbb{F}_q$-subcodes of expanded Gabidulin codes. For $m \nmid K'$, where m is the extension degree, this dimension mismatch is not merely a randomization heuristic. It deterministically rules out hidden $\mathbb{F}_{q^m}$-linear expansion structure, a key algebraic prerequisite for the relevant attacks in [5, 43]. Using a generic transform, EnGMS-based constructions yield IND-CCA2-secure public-key encryption schemes and key encapsulation mechanisms, while retaining deterministic decoding and zero decryption failure. At standard security levels, our schemes achieve very compact ciphertexts with moderate public-key sizes, demonstrating that provable structural guarantees can coexist with competitive size efficiency.

1 Introduction

Rank-Metric Code-Based Cryptography and Structural Tension. Rank-metric code-based cryptography has a long and intricate history, with early milestones such as the introduction of Gabidulin codes [20], which can be viewed as rank-metric analogues of Reed–Solomon codes [46]. Owing to their ability to efficiently correct rank errors, Gabidulin codes have played a central role in early rank-metric code-based cryptosystems [10, 19, 21, 22, 41], most notably the GPT scheme [21]. Indeed, the very algebraic regularity that enables efficient decoding has repeatedly led to powerful structural attacks, beginning with Overbeck's attack [42] and subsequent refinements (e.g., [40]).

This phenomenon closely mirrors the classical experience in the Hamming metric, where highly structured codes such as Reed–Solomon codes have long been considered unsuitable for McEliece-like masking [13,48], whereas less structured families eventually led to long-standing constructions such as the original McEliece scheme with binary Goppa codes [35]. In the rank metric, this tension between *decodability* and *structural security* has shaped much of the subsequent research [1–3, 5, 7–9, 18, 23, 24, 36–38, 45].

Moving Away from Gabidulin Structure. Motivated by the vulnerability of Gabidulin-based constructions, a growing body of work has explored alternatives inspired by NTRU-like [29] and MDPC-type designs [39], leading to rank-metric schemes based on Low-Rank Parity-Check (LRPC) codes [23] and their variants. These ideas culminated in several NIST PQC submissions, including LAKE, LOCKER, OUROBOROS-R, and RQC [1,2,36,37]; several of these proposals were later consolidated into the ROLLO scheme [3]. While these constructions substantially reduce algebraic structure, they introduce new challenges related to parameter selection, decoding complexity, and algebraic modeling.

A fundamental computational problem underlying many rank-metric code-based cryptosystems is the *MinRank problem*, which naturally arises in the analysis of decoding and structural attacks. Recent advances in algebraic modeling, notably the Support Minors framework [9], have led to improved strategies for solving structured MinRank instances and have played a role in the cryptanalytic evaluation of ROLLO and RQC [7]. These developments highlight that simply shifting structure does not necessarily mitigate attack surfaces; rather, it changes their nature.

Vector Codes, Matrix Codes, and Structural Attacks. Rank-metric code-based cryptosystems can be formulated either in terms of vector codes over extension fields or as matrix codes over the base field, a viewpoint with early precursors [31]. Such matrix codes are $\mathbb{F}_q$-linear subspaces of $\mathbb{F}_q^{m\times n}$ endowed with the rank metric. In this setting, the generic decoding problem can be formulated as a MinRank instance, whose decision version is NP-complete [12]. A rank-metric code of length n over $\mathbb{F}_{q^m}$ is an $\mathbb{F}_{q^m}$-linear vector subspace of $\mathbb{F}_{q^m}^n$, also endowed with the rank metric; we refer to such objects as (rank-metric) vector codes. Fixing an $\mathbb{F}_q$-basis of $\mathbb{F}_{q^m}$, the expansion map transforms a vector code into a matrix code by expanding each coordinate into its m-dimensional representation over $\mathbb{F}_q$.

While these representations are equivalent via the usual expansion map, they expose different forms of structure. In particular, Gabidulin vector codes over $\mathbb{F}_{q^m}$ allow for efficient distinguishers constructed on the basis of Frobenius closures [24], while their matrix expansions give rise to non-trivial stabilizer algebras [5]. Both of these phenomena stand in stark contrast to the characteristics of random rank-metric codes and thus enable the design of efficient polynomial-time distinguishers. Aragon *et al.* [5] introduced a masking method for matrix Gabidulin codes, but the resulting public codes still admit a combinatorial distinguisher for residual $\mathbb{F}_{q^m}$-linearity and an Overbeck-like attack exploiting surviving Frobenius structure. Building on this line of work, Porwal *et al.* [43] gave

a key-recovery attack against the EGMC scheme of Aragon *et al.* Their attack is asymptotically cheaper than the previously best distinguisher and shows that some original EGMC parameter sets do not achieve their claimed security levels.

A concise review of these distinguishers, as well as of the remaining attacks against enhanced Gabidulin constructions, is deferred to Appendix A.

Key Insight and Approach. Our starting point is the observation that prominent known attacks often rely on the existence of an underlying $\mathbb{F}_{q^m}$-linear structure, regardless of whether it is explicitly visible. We show that this dependence can be suppressed at a structural level, not by masking the full expansion itself, but instead by working with carefully chosen *matrix subcodes* whose dimensions are incompatible with extension-field linearity.

Concretely, we select an $\mathbb{F}_q$-subcode of dimension K' inside the matrix expansion of a Gabidulin code, with $m \nmid K'$, and then apply masking. The unmasked subcode already rules out direct $\mathbb{F}_{q^m}$-linear expansion structure; the resulting public code still retains efficient Gabidulin decoding while excluding the key algebraic preconditions underlying the relevant known attacks in [5, 43].

Beyond the Subcode Paradigm. At a conceptual level, the use of subcodes is not new in code-based cryptography. For example, Couvreur *et al.* [14] use matrix subcodes to hide Gabidulin structure in a hash-and-sign signature setting, rather than for public-key encryption or key encapsulation. In that setting and in much of the related prior work, subcodes are employed primarily as a *heuristic masking tool*, without a precise characterization of which algebraic invariants are preserved or ruled out.

In contrast, our work departs from this viewpoint and treats subcode selection as a *structural-obstruction mechanism* with provable guarantees. Rather than relying on the generic intuition that "subcodes look more random", we use subcode selection to enforce a specific algebraic obstruction. In this way, subcodes are transformed from a heuristic technique into a *provable structural tool*.

Construction Overview. Enhanced Gabidulin Matrix Code (EGMC) schemes [5] demonstrate that Gabidulin-based constructions can achieve extremely compact ciphertexts. However, the public code in EGMC still derives from a full expansion of an $\mathbb{F}_{q^m}$-linear Gabidulin code, and therefore retains sufficient extension-field structure to enable the combinatorial and Overbeck-like distinguishers identified in [5]. Motivated by these observations, our goal is to preserve the efficiency benefits of EGMC while suppressing the algebraic invariants on which such distinguishers rely. To this end, we depart from full expansion and instead construct public codes from carefully chosen matrix subcodes of reduced dimension.

Specifically, we begin with a Gabidulin code $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ of dimension k over $\mathbb{F}_{q^m}$ and fix an $\mathbb{F}_q$-basis of $\mathbb{F}_{q^m}$. This yields an associated matrix code $\mathcal{C}_{\mathrm{mat}} \subseteq \mathbb{F}_q^{m \times n}$ of dimension mk over $\mathbb{F}_q$. For a target dimension $K' \leq mk$, we sample a random full column-rank matrix $\mathbf{T} \in \mathbb{F}_q^{mk \times K'}$ and use its columns as coefficient vectors to form K' random linear combinations of a fixed $\mathbb{F}_q$-basis of $\mathcal{C}_{\mathrm{mat}}$. The resulting code is an $\mathbb{F}_q$-linear matrix subcode of dimension K'.

A crucial property of this construction is that choosing K' such that $m \nmid K'$ provably excludes the possibility that the resulting subcode arises as the expan-

sion of any $\mathbb{F}_{q^m}$-linear code (see Lemma 1). This dimension mismatch removes the extension-field linearity that underlies known structural attacks discussed above. Finally, we apply a masking procedure of Definition 11, based on row and column padding together with left and right scramblers, yielding the public matrix code. While the public code no longer exhibits $\mathbb{F}_{q^m}$-linearity, its security remains grounded in the MinRank problem, and the reduction in code dimension can translate into smaller public key and ciphertext sizes.

The novelty of our approach does not lie in the use of subcodes per se, but in using subcode-dimension selection to enforce a structural obstruction. Our contributions can be summarized as follows:

- We identify a simple but fundamental structural obstruction to extension-field linearity in matrix codes: any $\mathbb{F}_q$-linear matrix code of dimension not divisible by m cannot admit a direct $\mathbb{F}_{q^m}$-linear expansion representation.
- We define the Enhanced Gabidulin Matrix Subcode (EnGMS) family by applying a subcode-based matrix code transformation to expanded Gabidulin codes, where the subcode selected before masking has $\mathbb{F}_q$-dimension K'.
- We show that, for EnGMS parameters satisfying $m \nmid K'$, the relevant structural attacks in [5, 43] do not apply as stated, while the considered variants do not yield effective attacks against our proposed parameter sets.
- We instantiate the EnGMS framework in public-key encryption and KEMs with deterministic decoding and zero decryption failure, achieving compact ciphertexts and moderate public-key sizes.

Structural Perspective and Efficiency. From a structural standpoint, EnGMS follows a McEliece-like paradigm: a structured code is transformed, via base-field subcodes and masking, into a public code whose security rests on the apparent absence of algebraic invariants exploitable by known structural attacks. Specifically, our construction starts from a Gabidulin code over $\mathbb{F}_{q^m}$, passes to an $\mathbb{F}_q$-subcode of its associated matrix code, and finally applies a masking step. This design is conceptually analogous to Classic McEliece, where binary Goppa codes can be viewed as base-field subcodes of suitable extension-field codes before masking is applied [35].

The original McEliece cryptosystem, introduced in 1978, and its core design principles have withstood more than four decades of cryptanalysis and are widely regarded as among the most conservative approaches in code-based cryptography [35]. The Classic McEliece proposal builds explicitly on this long-standing line of work by instantiating these principles within a modern parameter and standardization framework [11]. By contrast, schemes such as HQC [25] rely on quasi-cyclic structure and the QCSD problem, which has been the subject of extensive cryptanalytic investigation and parameter refinements over time [28, 34, 47]. By avoiding quasi-cyclic structure and obscuring extension-field linearity through base-field subcodes, EnGMS can therefore be viewed as closer in spirit to the Classic McEliece design philosophy than to HQC.

Relative to EGMC, the dimension-incompatible subcode step removes the direct precondition for the relevant structural attacks in [5, 43]. Rank-metric NIST

submissions such as RQC [38] and ROLLO [3] rely on ideal or quasi-cyclic rank structures, which have been the focus of recent algebraic attacks and parameter revisions [8, 9]. Compared with these two schemes, EnGMS inherits deterministic Gabidulin decoding and therefore exhibits a zero decryption failure rate, matching RQC in this regard and contrasting with ROLLO.

Beyond structural considerations, EnGMS offers a favorable efficiency profile. At the 128-bit security level, we achieve a balanced parameter set with a 60 B ciphertext and a 90 kB public key. By comparison, the original EGMC-Niederreiter parameters in [5] use a 65 B ciphertext with a 98 kB public key, while Classic McEliece requires a 96 B ciphertext and a 261 kB public key [11]. Additional parameter sets and comparisons at higher security levels are reported in Section 6.

Organization of This Work. The remainder of this paper is organized as follows. Section 2 fixes notation and recalls the necessary background on public-key encryption and key encapsulation mechanisms, Gabidulin codes, matrix codes, and the MinRank problem. Section 3 details the subcode-dimension selection step for matrix codes and applies it to Gabidulin matrix codes, leading to the EnGMS construction and the induced public-code distributions. Section 4 presents the EnGMS-Niederreiter encryption scheme, establishes correctness, and proves OW-CPA security. Section 5 analyzes known attacks relevant to our constructions, including those of [5, 43], together with related variants considered in this work. Section 6 provides concrete parameter choices for EnGMS-Niederreiter and compares its efficiency with existing code-based and lattice-based constructions. Section 7 concludes the paper.

Standard security definitions, the application of the FO transformation yielding IND-CCA2-secure KEM and PKE instantiations, the EnGMS-McEliece construction, additional discussions of structural attacks and dual-code attacks, and a comparison of size-related metrics for CPA-secure PKE schemes are deferred to the Appendix.

2 Notation and Preliminaries

The security parameter is denoted by $\lambda \in \mathbb{N}$. Matrices are written in boldface (e.g., $\mathbf{A}$). For a matrix $\mathbf{A}$, $\mathrm{rank}(\mathbf{A})$ denotes its rank, and $\mathbf{A}^{-1}$ its inverse whenever $\mathbf{A}$ is invertible. Unless stated otherwise, all vectors are row vectors. For a row vector $\boldsymbol{x}$ or a matrix $\mathbf{A}$, their transposes are denoted by $\boldsymbol{x}^\top$ and $\mathbf{A}^\top$, respectively. In particular, for a parity-check matrix $\mathbf{H} \in \mathbb{F}_q^{\ell \times n}$, the syndrome of a vector $\boldsymbol{x} \in \mathbb{F}_q^n$ is defined as $\boldsymbol{s} = \boldsymbol{x}\mathbf{H}^\top \in \mathbb{F}_q^\ell$.

Let q be a prime power. We denote by $\mathbb{F}_{q^m}$ the degree-m extension of $\mathbb{F}_q$ and by $\mathbb{F}_q^{m \times n}$ the set of all $m \times n$ matrices over $\mathbb{F}_q$. The set of all $\mathbb{F}_q$-bases of $\mathbb{F}_{q^m}$ is denoted by $\mathcal{B}(\mathbb{F}_{q^m})$, and $\mathrm{GL}_m(\mathbb{F}_q)$ denotes the group of invertible $m \times m$ matrices over $\mathbb{F}_q$.

For a square matrix $\mathbf{A} = [a_{ij}] \in \mathbb{F}_q^{m \times m}$, the matrix trace is defined as $\mathrm{tr}(\mathbf{A}) := \sum_{i=1}^m a_{ii}$. For a finite set $\mathcal{S}$, the notation $x \xleftarrow{\$} \mathcal{S}$ denotes uniform

sampling from $\mathcal{S}$, and $\#\mathcal{S}$ its cardinality. For a subset $\mathcal{S}$ of a vector space, $\mathrm{Span}_{\mathbb{F}_q}(\mathcal{S})$ denotes its $\mathbb{F}_q$-linear span. For an $\mathbb{F}_q$-linear subspace $\mathcal{C}$, $\dim_{\mathbb{F}_q}(\mathcal{C})$ denotes its dimension over $\mathbb{F}_q$. We write PPT for probabilistic polynomial time.

The terms EGMC and EGMC-Niederreiter refer to the Enhanced Gabidulin Matrix Code (EGMC) schemes and their Niederreiter variant introduced in [5].

2.1 Public-Key Encryption and Key Encapsulation Mechanisms

We consider two standard primitives in public-key cryptography: public-key encryption (PKE) and key encapsulation mechanisms (KEMs).

A public-key encryption scheme is a triple of probabilistic polynomial-time (PPT) algorithms $\mathsf{PKE} = (\mathsf{KeyGen}, \mathsf{Encrypt}, \mathsf{Decrypt})$. The key generation algorithm $\mathsf{KeyGen}(1^\lambda)$ outputs a public/secret key pair $(\mathsf{pk}, \mathsf{sk})$. Encryption $\mathsf{Encrypt}(\mathsf{pk}, m)$ takes a public key pk and a message m, and outputs a ciphertext c. Decryption $\mathsf{Decrypt}(\mathsf{sk}, c)$ outputs either the corresponding message or the failure symbol $\perp$. We require the standard notion of correctness, namely that decryption of an honestly generated ciphertext succeeds with overwhelming probability.

A key encapsulation mechanism consists of a triple of PPT algorithms $\mathsf{KEM} = (\mathsf{KeyGen}, \mathsf{Encaps}, \mathsf{Decaps})$. The key generation algorithm is identical to that of PKE. The encapsulation algorithm $\mathsf{Encaps}(\mathsf{pk})$ outputs a shared secret K together with a ciphertext c, while the decapsulation algorithm $\mathsf{Decaps}(\mathsf{sk}, c)$ outputs either the same shared secret or $\perp$. Correctness requires that decapsulation of an honestly generated encapsulation recovers the shared secret with overwhelming probability.

We adopt the standard security notions for both primitives. For PKE, we focus on one-wayness under chosen-plaintext attack (OW-CPA) and indistinguishability under adaptive chosen-ciphertext attack (IND-CCA2). For KEM, we adopt the corresponding notion of IND-CCA2 security.

For completeness, we provide the formal game-based definitions of the above correctness and security notions in Appendix B.

2.2 Gabidulin Codes

A q-polynomial over $\mathbb{F}_{q^m}$ is a polynomial of the form

$$f(X) = \sum_{i=0}^{s} a_i X^{q^i}, \quad a_i \in \mathbb{F}_{q^m},$$

and its q-degree is defined as $\deg_q f := \max\{\, i \mid a_i \neq 0 \,\}$. We write $\mathcal{L}_q[X]$ for the set of all such polynomials.

Definition 1 (Gabidulin Code). *Let $k, m, n \in \mathbb{N}$ satisfy $k \leq n \leq m$. Let $\boldsymbol{g} = (g_1, \ldots, g_n) \in \mathbb{F}_{q^m}^n$ be a vector whose entries are linearly independent over $\mathbb{F}_q$. Define*

$$\mathcal{G}_{\boldsymbol{g}}(n, k, m) := \big\{\, (f(g_1), \ldots, f(g_n)) \;\big|\; f \in \mathcal{L}_q[X],\ \deg_q f < k \big\}.$$

The code $\mathcal{G}_{\boldsymbol{g}}(n,k,m)$ is called a Gabidulin code with parameters $[n,k]_{q^m}$ and evaluation vector $\boldsymbol{g}$.

Gabidulin codes were proposed in [20] as a family of maximum rank distance (MRD) codes. They are particularly appealing in cryptography because efficient decoding algorithms can correct errors of rank weight up to $\left\lfloor \frac{n-k}{2} \right\rfloor$. However, their strong algebraic structure also gives rise to efficient distinguishers from random codes [24].

2.3 Matrix Codes

Definition 2 (Ψ_γ-Expansion). *Let $\gamma = (\gamma_1, \ldots, \gamma_m) \in \mathcal{B}(\mathbb{F}_{q^m})$. The Ψ_γ-expansion is defined by*

$$\Psi_\gamma : x \in \mathbb{F}_{q^m} \mapsto (x_1, \ldots, x_m)^\top \in \mathbb{F}_q^{m\times 1},$$

where $x = \sum_{i=1}^m x_i \gamma_i$.

The expansion Ψ_γ extends component-wise to vectors $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathbb{F}_{q^m}^n$, yielding a matrix $\Psi_\gamma(\boldsymbol{y}) \in \mathbb{F}_q^{m\times n}$ with j-th column $\Psi_\gamma(y_j)$.

Definition 3 (Rank-Metric Vector Codes). *A vector code $\mathcal{C}_{\mathrm{vec}}$ is an $\mathbb{F}_{q^m}$-linear subspace of $\mathbb{F}_{q^m}^n$, endowed with the rank metric. The rank weight $\mathrm{wt}_R(\boldsymbol{x})$ of a vector $\boldsymbol{x} \in \mathbb{F}_{q^m}^n$ is defined as the rank of the matrix $\Psi_\gamma(\boldsymbol{x})$ for any choice of $\gamma \in \mathcal{B}(\mathbb{F}_{q^m})$. The weight is invariant under the choice of γ [49].*

Definition 4 (Matrix Codes). *A matrix code $\mathcal{C}_{\mathrm{mat}}$ is an $\mathbb{F}_q$-subspace of $\mathbb{F}_q^{m\times n}$ equipped with the rank metric. Its minimum rank distance is given by*

$$d(\mathcal{C}_{\mathrm{mat}}) := \min_{\substack{\mathbf{A}\in\mathcal{C}_{\mathrm{mat}} \\ \mathbf{A}\neq\mathbf{0}}} \mathrm{rank}(\mathbf{A}).$$

Let $\gamma \in \mathcal{B}(\mathbb{F}_{q^m})$. An $\mathbb{F}_{q^m}$-linear $[n,k]_{q^m}$ vector code $\mathcal{C}_{\mathrm{vec}}$ induces an mk-dimensional $\mathbb{F}_q$-matrix code:

$$\Psi_\gamma(\mathcal{C}_{\mathrm{vec}}) := \{\Psi_\gamma(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathcal{C}_{\mathrm{vec}}\}.$$

Definition 5 (Singleton Bound [16]). *Let $\mathcal{C}_{\mathrm{mat}} \subseteq \mathbb{F}_q^{m\times n}$ be an $\mathbb{F}_q$-linear matrix code of dimension K and minimum rank distance $d = d(\mathcal{C}_{\mathrm{mat}})$. Then $\mathcal{C}_{\mathrm{mat}}$ satisfies the Singleton bound*

$$d \leq \min(m,n) - \frac{K}{\max(m,n)} + 1.$$

A matrix code achieving equality in this bound is called an MRD code.

Definition 6 (Matrix Code Equivalence). *Two matrix codes $\mathcal{C}_{\mathrm{mat}}$ and $\mathcal{D}_{\mathrm{mat}}$ in $\mathbb{F}_q^{m\times n}$ are equivalent if one can be transformed into the other by multiplying with invertible matrices from the left and right:*

$$\mathcal{C}_{\mathrm{mat}} = \mathbf{P}\mathcal{D}_{\mathrm{mat}}\mathbf{Q} := \{\mathbf{PDQ} \mid \mathbf{D} \in \mathcal{D}_{\mathrm{mat}}\},$$

where $\mathbf{P} \in \mathrm{GL}_m(\mathbb{F}_q)$ and $\mathbf{Q} \in \mathrm{GL}_n(\mathbb{F}_q)$.

Definition 7 (Folding Map). *Let* $\boldsymbol{v} = (\boldsymbol{v}_1 \| \cdots \| \boldsymbol{v}_n) \in \mathbb{F}_q^{mn}$ *be the concatenation of* $\boldsymbol{v}_j \in \mathbb{F}_q^m$*. Define* $\mathrm{Fold}(\boldsymbol{v}) \in \mathbb{F}_q^{m\times n}$ *to be the matrix whose* j*-th column is* $\boldsymbol{v}_j^\top$*, i.e.,* $\mathrm{Fold}(\boldsymbol{v}) := (\boldsymbol{v}_1^\top \mid \cdots \mid \boldsymbol{v}_n^\top)$*. Its inverse* $\mathrm{Unfold} : \mathbb{F}_q^{m\times n} \to \mathbb{F}_q^{mn}$ *denotes the column-wise vectorization of a matrix.*

Definition 8 (Left Stabilizer Algebra of a Matrix Code). *Let* $\mathcal{C}_{\mathrm{mat}} \subseteq \mathbb{F}_q^{m\times n}$ *be a matrix code over* $\mathbb{F}_q$*. The left stabilizer algebra of* $\mathcal{C}_{\mathrm{mat}}$ *is defined as the set of all* $m\times m$ *matrices over* $\mathbb{F}_q$ *that preserve* $\mathcal{C}_{\mathrm{mat}}$ *under left multiplication:*

$$\mathrm{Stab}_L(\mathcal{C}_{\mathrm{mat}}) := \left\{\mathbf{S} \in \mathbb{F}_q^{m\times m} \mid \mathbf{SC} \in \mathcal{C}_{\mathrm{mat}},\ \forall\, \mathbf{C} \in \mathcal{C}_{\mathrm{mat}}\right\}.$$

2.4 MinRank Problem

Definition 9 (MinRank Problem). *Let* $\mathbf{Y}, \mathbf{M}_1, \ldots, \mathbf{M}_K \in \mathbb{F}_q^{m\times n}$ *and let* $r \in \mathbb{N}$*. The MinRank problem with parameters* (q, m, n, K, r) *is to find coefficients* $c_1, \ldots, c_K \in \mathbb{F}_q$ *together with a matrix* $\mathbf{E} \in \mathbb{F}_q^{m\times n}$ *with* $\mathrm{rank}(\mathbf{E}) \le r$ *such that* $\mathbf{Y} = \sum_{i=1}^K c_i\mathbf{M}_i + \mathbf{E}$*.*

Definition 10 (MinRank-Syndrome Problem [5]). *Given vectors* $\boldsymbol{y}, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{mn} \in \mathbb{F}_q^{mn-K}$*, the MinRank-Syndrome problem with parameters* (q, m, n, K, r) *is the task of finding a vector* $\boldsymbol{e} = (e_1, \ldots, e_{mn}) \in \mathbb{F}_q^{mn}$ *such that* $\boldsymbol{y} = \sum_{i=1}^{mn} e_i\boldsymbol{\mu}_i$ *and* $\mathrm{rank}\big(\mathrm{Fold}(\boldsymbol{e})\big) \le r$*.*

For an $\mathbb{F}_q$-linear matrix code $\mathcal{C}_{\mathrm{mat}}$ of size $m \times n$ and dimension K, the decoding task is: given a received matrix $\mathbf{Y} \in \mathbb{F}_q^{m\times n}$, recover a codeword $\mathbf{X} \in \mathcal{C}_{\mathrm{mat}}$ and an error matrix $\mathbf{E} \in \mathbb{F}_q^{m\times n}$ such that $\mathbf{Y} = \mathbf{X} + \mathbf{E}$ with $\mathrm{rank}(\mathbf{E}) \le r$. Equivalently, fixing an $\mathbb{F}_q$-basis $\{\mathbf{M}_i\}_{i=1}^K$ of $\mathcal{C}_{\mathrm{mat}}$, decoding reduces to an instance of $\mathrm{MinRank}(q, m, n, K, r)$: given $\mathbf{Y}$ and $\mathbf{M}_i$, find $\boldsymbol{c} \in \mathbb{F}_q^K$ and $\mathbf{E} \in \mathbb{F}_q^{m\times n}$ with $\mathrm{rank}(\mathbf{E}) \le r$ such that $\mathbf{Y} = \sum_{i=1}^K c_i\mathbf{M}_i + \mathbf{E}$. The MinRank and MinRank-Syndrome problems are equivalent, for a similar reason as in vector codes, where the decoding problem can be reformulated as the syndrome decoding problem. Moreover, the decision MinRank problem is known to be NP-complete [12].

3 Subcode-Dimension Selection in Matrix Codes

3.1 Structural Obstruction to Extension-Field Linearity

We describe a subcode-dimension selection step for matrix codes, whose goal is to conceal the algebraic structure of an underlying secret code. The starting point of our approach is a simple, yet fundamental, structural obstruction to extension-field linearity in matrix codes.

Let $\mathcal{C}_{\mathrm{mat}} \subseteq \mathbb{F}_q^{m\times n}$ be a matrix code obtained as the Ψ_γ-expansion of an $\mathbb{F}_{q^m}$-linear vector code. Then $\mathcal{C}_{\mathrm{mat}}$ is necessarily an $\mathbb{F}_{q^m}$-linear space when viewed inside $\mathbb{F}_q^{m\times n}$, and in particular its $\mathbb{F}_q$-dimension must be a multiple of m.

This observation yields a robust criterion: any matrix code whose $\mathbb{F}_q$-dimension is not divisible by m cannot itself be $\mathbb{F}_{q^m}$-linear, regardless of any dimension-preserving transformations. In particular, such a code cannot arise as the expansion of a Gabidulin code or any $\mathbb{F}_{q^m}$-linear rank-metric code. We exploit this obstruction as a central design principle in our construction.

The following elementary observation will be used repeatedly in the sequel. It formalizes the subcode-selection step and, at the same time, records the dimension-based obstruction to hidden $\mathbb{F}_{q^m}$-linearity.

Lemma 1. *Fix $\gamma \in \mathcal{B}(\mathbb{F}_{q^m})$, and let $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ be a k-dimensional $\mathbb{F}_{q^m}$-linear code. Let $\Psi_\gamma(\mathcal{C}) \subseteq \mathbb{F}_q^{m\times n}$ denote its matrix expansion, and let $\mathbf{A}_1, \ldots, \mathbf{A}_{mk}$ be an $\mathbb{F}_q$-basis of $\Psi_\gamma(\mathcal{C})$. Given a full column-rank matrix $\mathbf{T} \in \mathbb{F}_q^{mk\times K'}$ with $1 \le K' \le mk$, define $\mathbf{A}'_j := \sum_{i=1}^{mk} T_{i,j}\mathbf{A}_i$ for $j = 1, \ldots, K'$, and let $\mathcal{S} := \mathrm{Span}_{\mathbb{F}_q}\{\mathbf{A}'_1, \ldots, \mathbf{A}'_{K'}\}$. Then*

1. *the subspace $\mathcal{S}$ has dimension K' over $\mathbb{F}_q$ and is contained in $\Psi_\gamma(\mathcal{C})$;*
2. *if $m \nmid K'$, then $\mathcal{S}$ is not $\mathbb{F}_{q^m}$-linear. In particular, $\mathcal{S}$ cannot be the Ψ_γ-expansion of any $\mathbb{F}_{q^m}$-linear subcode of $\mathbb{F}_{q^m}^n$.*

From Subspace Sampling to Structural Obstruction. At first glance, the above construction may appear as a standard subcode sampling procedure. However, the role of the transformation differs from that in many related subcode-based approaches. The key point is that the dimension condition $m \nmid K'$ induces a *structural obstruction*: it excludes the existence of a direct $\mathbb{F}_{q^m}$-linear expansion structure on the selected subspace, independently of how the subspace is represented through dimension-preserving $\mathbb{F}_q$-linear transformations. In particular, this is not a probabilistic statement about typical subcodes, but a deterministic constraint that holds for all subspaces of dimension K'. Therefore, the subcode selection step should be viewed not merely as a means of introducing randomness, but as a mechanism for enforcing this algebraic obstruction on the selected subcode. The randomness in the choice of $\mathbf{T}$ is used to place the subcode in a generic position and to avoid obvious additional structure.

For attacks relying on such a hidden expansion structure, the condition $m \nmid K'$ removes the direct prerequisite; the corresponding algebraic framework is no longer immediately available from the selected subcode itself. Importantly, this is not merely a degradation of the attack efficiency, but a qualitative change in the direct attack surface: the adversary can no longer directly follow the known route of recovering an extension-field structure from the selected subcode or its masked image, and would typically need to operate through an intermediate code with sufficient inherited structure, or resort to generic $\mathbb{F}_q$-linear matrix-code attacks.

3.2 Random Matrix Subcodes and Subcode-Based Masking

Guided by the above obstruction, we consider matrix subcodes of expanded Gabidulin codes whose dimensions deliberately violate the divisibility condition by m. Among all such subcodes, random $\mathbb{F}_q$-subcodes constitute a natural choice:

they preserve the ambient rank-metric geometry while avoiding unnecessary algebraic regularity that could be exploited by structural attacks.

Formally, let $\mathcal{C}_{\mathrm{mat}} \subseteq \mathbb{F}_q^{m\times n}$ be a matrix code of dimension mk. A random matrix subcode of dimension K' is obtained by sampling a full column-rank linear map $\mathbf{T} \in \mathbb{F}_q^{mk\times K'}$ and taking the image of $\mathcal{C}_{\mathrm{mat}}$ under $\mathbf{T}$ (with respect to a fixed $\mathbb{F}_q$-basis of $\mathcal{C}_{\mathrm{mat}}$). With overwhelming probability (by a standard subspace-counting argument via Gaussian binomial coefficients [44]), the resulting subcode admits no nontrivial direct $\mathbb{F}_{q^m}$-linear structure.

Definition 11 (Subcode-Based Matrix Code Transformation). *Fix integers $m, n, K, \ell_1, \ell_2 \in \mathbb{N}$. Let $\mathcal{C}_{\mathrm{mat}} \subseteq \mathbb{F}_q^{m\times n}$ be a matrix code of dimension K over $\mathbb{F}_q$ with basis $\mathcal{B} = \{\mathbf{A}_1, \ldots, \mathbf{A}_K\}$. Choose $K' \in \{1, \ldots, K\}$ and a full column-rank matrix $\mathbf{T} \in \mathbb{F}_q^{K\times K'}$. For $j = 1, \ldots, K'$, define*

$$\mathbf{A}'_j := \sum_{i=1}^{K} T_{i,j}\, \mathbf{A}_i \in \mathbb{F}_q^{m\times n},$$

and set $\mathcal{B}' = \{\mathbf{A}'_1, \ldots, \mathbf{A}'_{K'}\}$. This gives a K'-dimensional subcode $\mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}') \subseteq \mathcal{C}_{\mathrm{mat}}$. We then draw the following matrices uniformly at random:

- $\mathbf{P} \xleftarrow{\$} \mathrm{GL}_{m+\ell_1}(\mathbb{F}_q)$,
- $\mathbf{Q} \xleftarrow{\$} \mathrm{GL}_{n+\ell_2}(\mathbb{F}_q)$,
- *for $i = 1, \ldots, K'$, sample independently*

$$\mathbf{R}_i \xleftarrow{\$} \mathbb{F}_q^{m\times \ell_2}, \quad \mathbf{R}'_i \xleftarrow{\$} \mathbb{F}_q^{\ell_1\times n}, \quad \mathbf{R}''_i \xleftarrow{\$} \mathbb{F}_q^{\ell_1\times \ell_2}.$$

The transformed matrix code is defined by the basis

$$\mathcal{RB}' = \left\{ \mathbf{P} \begin{pmatrix} \mathbf{A}'_1 & \mathbf{R}_1 \\ \mathbf{R}'_1 & \mathbf{R}''_1 \end{pmatrix} \mathbf{Q},\ \ldots,\ \mathbf{P} \begin{pmatrix} \mathbf{A}'_{K'} & \mathbf{R}_{K'} \\ \mathbf{R}'_{K'} & \mathbf{R}''_{K'} \end{pmatrix} \mathbf{Q} \right\}.$$

Implementing Structure-Obstructing Subcodes. Having established the dimension obstruction induced by $m \nmid K'$, realizing such subspaces in a cryptographic construction requires additional care. A naive restriction to a subset of basis elements would typically preserve significant portions of the original algebraic structure, in particular dependencies induced by the underlying extension-field representation. Such constructions may thus preserve attack-relevant invariants.

Instead, we construct subcodes via full-rank linear combinations over $\mathbb{F}_q$, parameterized by a matrix $\mathbf{T} \in \mathbb{F}_q^{K\times K'}$. Uniform sampling of a full column-rank matrix $\mathbf{T}$ makes the resulting subcode uniformly distributed among the K'-dimensional $\mathbb{F}_q$-subspaces of the fixed ambient code $\mathcal{C}_{\mathrm{mat}}$, thereby avoiding the deterministic bias that would arise from selecting a prescribed subset of basis elements. Thus, $\mathbf{T}$ does not create the dimension obstruction itself; rather, it realizes this obstruction without imposing additional obvious structure through a prescribed basis choice.

From now on, unless stated otherwise, we assume that $k, m, n, \ell_1, \ell_2, K' \in \mathbb{N}$ satisfy $k \le n \le m$ and $1 \le K' \le mk$.

Definition 12 (Enhanced Gabidulin Matrix Subcode). *Let $\gamma \in \mathcal{B}(\mathbb{F}_{q^m})$, and let $\mathcal{G} \subseteq \mathbb{F}_{q^m}^n$ be a k-dimensional Gabidulin code over $\mathbb{F}_{q^m}$. Set $\mathcal{C}_{\mathrm{mat}} := \Psi_\gamma(\mathcal{G}) \subseteq \mathbb{F}_q^{m\times n}$. The Enhanced Gabidulin Matrix Subcode $\mathcal{EG}_{\mathrm{sub}}(k, m, n, K', \ell_1, \ell_2)$ is defined as the matrix code obtained by applying the transformation of Definition 11 to $\mathcal{C}_{\mathrm{mat}}$ with $K = mk$.*

Masking Matrix Subcodes. The role of masking in our construction is not to suppress extension-field linearity (this is already ensured by the choice of the subcode dimension), but to hide the specific representation of the resulting matrix code. Concretely, we apply standard row and column scramblers together with padding operations to the selected matrix subcode. These transformations preserve the decoding capability inherited from the underlying Gabidulin code given the private information, while preventing the public code from directly exposing the particular basis or embedding used in its construction.

3.3 The EnGMS Distributions

We now formalize the public-code distributions induced by the above subcode selection and masking procedure, which we collectively refer to as the EnGMS distributions. These distributions abstract away from the choice of decoding algorithm and serve as the basis for our security analysis.

Intuitively, they aim to suppress extension-field linear structure in the public code. Security is then formulated via the indistinguishability of these distributions from uniformly random matrix codes with matching parameters, following the standard indistinguishability-based paradigm in code-based cryptography.

Definition 13 (EnGMS Distribution). *Sample $\boldsymbol{g} \xleftarrow{\$} \mathbb{F}_{q^m}^n$ conditioned on the entries of $\boldsymbol{g}$ being $\mathbb{F}_q$-linearly independent, and $\gamma \xleftarrow{\$} \mathcal{B}(\mathbb{F}_{q^m})$. Let $\mathcal{G}$ be the k-dimensional Gabidulin code of length n over $\mathbb{F}_{q^m}$ with evaluation vector $\boldsymbol{g}$, and let $\mathcal{C}_{\mathrm{mat}} := \Psi_\gamma(\mathcal{G}) \subseteq \mathbb{F}_q^{m\times n}$. We then apply the transformation of Definition 11 to $\mathcal{C}_{\mathrm{mat}}$ with $K = mk$. The output of the distribution EnGMS$(k, m, n, K', \ell_1, \ell_2)$ is the resulting transformed matrix code with basis $\mathcal{RB}'$.*

Definition 14 (EnGMS-IND Problem). *Let $\mathcal{C}'_{\mathrm{mat}} \subseteq \mathbb{F}_q^{(m+\ell_1)\times(n+\ell_2)}$ be a matrix code of dimension K' over $\mathbb{F}_q$. The EnGMS-IND problem is to decide, given $\mathcal{C}'_{\mathrm{mat}}$, whether $\mathcal{C}'_{\mathrm{mat}}$ is sampled according to the EnGMS$(k, m, n, K', \ell_1, \ell_2)$ distribution, or whether it is drawn uniformly at random from the set of all K'-dimensional $\mathbb{F}_q$-subspaces of $\mathbb{F}_q^{(m+\ell_1)\times(n+\ell_2)}$.*

Analysis of the EnGMS-IND Assumption. The EnGMS-IND assumption formalizes the hardness of distinguishing EnGMS public codes from random matrix codes with matching parameters after the subcode-selection and masking steps. Our OW-CPA security proof relies on this public-code indistinguishability assumption together with the corresponding MinRank assumption.

This assumption is motivated by the structural-attack analysis in Section 5. In the recommended regime $m \nmid K'$, the relevant structural attacks in [5, 43] do

not apply as stated by Lemma 1. The natural variants examined in our analysis do not lead to effective attacks against the proposed parameter sets.

For these attack mechanisms, no PPT distinguisher with non-negligible advantage against EnGMS-IND is known in this setting. Thus, EnGMS-IND captures the remaining public-code indistinguishability assumption after the known attack preconditions no longer apply directly. Further discussion of the assumption and its supporting observations is deferred to Appendices C.1 and C.2.

Definition 15 (EnGMS-Search Problem). *The EnGMS-Search problem is to recover, given a matrix code $\mathcal{C}'_{\mathrm{mat}}$ sampled from the EnGMS$(k, m, n, K', \ell_1, \ell_2)$ distribution, the underlying $\mathbb{F}_q$-basis $\gamma \in \mathcal{B}(\mathbb{F}_{q^m})$ and the evaluation vector $\boldsymbol{g} \in \mathbb{F}_{q^m}^n$ used to generate $\mathcal{C}'_{\mathrm{mat}}$.*

We next characterize the dual of the EnGMS matrix subcode, which will be useful in the analysis of potential dual-based attacks. Let $\mathcal{G}$ be an $[n, k]$ Gabidulin code over $\mathbb{F}_{q^m}$, and let $\gamma, \gamma' \in \mathcal{B}(\mathbb{F}_{q^m})$ be orthogonal bases in the sense of [45, Definition 19]. By [45, Theorem 21] we have $\Psi_\gamma(\mathcal{G})^\perp = \Psi_{\gamma'}(\mathcal{G}^\perp)$, where $\mathcal{G}^\perp$ denotes the Euclidean dual of $\mathcal{G}$, which is an $[n, n-k]$ Gabidulin code over $\mathbb{F}_{q^m}$. Together with Definition 21 (Appendix F.1), this implies the following theorem.

Theorem 1 (Dual of EnGMS). *We adopt the notation of Definitions 11 and 12. Fix an integer K' with $1 \le K' \le mk$, and define the $\mathbb{F}_q$-subcode*

$$\mathcal{C}_{\mathrm{sub}} := \mathrm{Span}_{\mathbb{F}_q}\{\mathbf{A}'_1, \ldots, \mathbf{A}'_{K'}\} \subseteq \Psi_\gamma(\mathcal{G}).$$

Let

$$\mathcal{V} := \left\{ \begin{pmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \middle| \ \mathbf{R} \in \mathbb{F}_q^{m\times n} \right\} \subseteq \mathbb{F}_q^{(m+\ell_1)\times(n+\ell_2)}.$$

Then there exists a subspace $\mathcal{W}$ of $\mathbb{F}_q^{(m+\ell_1)\times(n+\ell_2)}$ that complements $\mathcal{V}$, such that the dual of the EnGMS matrix subcode satisfies

$$\mathcal{E}\mathcal{G}_{\mathrm{sub}}^\perp = (\mathbf{P}^\top)^{-1}\, \mathcal{D}'_{\mathrm{mat}}\, (\mathbf{Q}^\top)^{-1},$$

where

$$\mathcal{D}'_{\mathrm{mat}} = \big(\mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}_1) \oplus \mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}_2)\big) \oplus \mathcal{W},$$

with

$$\mathcal{B}_1 := \left\{ \begin{pmatrix} \mathbf{B}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \middle| \ i = 1, \ldots, m(n-k) \right\},\ \mathcal{B}_2 := \left\{ \begin{pmatrix} \widetilde{\mathbf{B}}_j & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \middle| \ j = 1, \ldots, mk - K' \right\}.$$

Here, $\{\mathbf{B}_i\}_{i=1}^{m(n-k)}$ denotes an $\mathbb{F}_q$-basis of $\Psi_\gamma(\mathcal{G})^\perp$, while $\{\widetilde{\mathbf{B}}_j\}_{j=1}^{mk-K'}$ denotes an $\mathbb{F}_q$-basis of a complement of $\Psi_\gamma(\mathcal{G})^\perp$ inside $\mathcal{C}_{\mathrm{sub}}^\perp$.

Proof Sketch. The proof proceeds by explicitly characterizing the structure of the dual of the EnGMS matrix subcode. First, by using the contragredient transformation rule for the matrix trace inner product under left and right

multiplication by invertible matrices, the effect of the scramblers $\mathbf{P}$ and $\mathbf{Q}$ is factored out, reducing the analysis to an unscrambled subcode, denoted by $\widehat{\mathcal{D}}$. A blockwise examination then shows that $\widehat{\mathcal{D}}^{\perp} \cap \mathcal{V}$ is precisely the zero-padded embedding of $\mathcal{C}_{\mathrm{sub}}^{\perp}$ into the top-left $m \times n$ block, while the remaining part of $\widehat{\mathcal{D}}^{\perp}$ forms a complementary subspace. Finally, the inclusion $\mathcal{C}_{\mathrm{sub}} \subseteq \Psi_{\gamma}(\mathcal{G})$ allows a basis of $\Psi_{\gamma}(\mathcal{G})^{\perp}$ to be extended to a basis of $\mathcal{C}_{\mathrm{sub}}^{\perp}$, giving the stated top-left decomposition. The full technical proof, including the precise construction of the involved subspaces, is deferred to Appendix C.3.

4 A Niederreiter-Type PKE from Matrix Subcodes

We now present a Niederreiter-type formulation of the EnGMS construction, which is the main focus of our presentation. The McEliece-type scheme, which can be regarded as its dual counterpart, is provided in Appendix D for reference. In this formulation, encryption computes a syndrome with respect to the public parity-check matrix rather than adding an error to a codeword.

The Niederreiter variant does not introduce additional algebraic structure beyond that already present in the public code. In particular, the public-key distribution coincides with that of the McEliece-type variant, and the security analysis is based on the same EnGMS-IND assumption. This equivalence ensures that both variants inherit identical resistance to known structural attacks.

We first describe the key generation, encryption, and decryption algorithms of the Niederreiter-type scheme, and then establish correctness and OW-CPA security. Applying the FO transformation yields an IND-CCA2-secure KEM in the random oracle model and, via the common KEM-DEM paradigm, an IND-CCA2-secure PKE scheme. We emphasize that the Niederreiter variant is suited to compact ciphertext instantiations: encryption outputs only the syndrome of the encoded low-rank message, and decryption remains deterministic.

4.1 EnGMS-Niederreiter Encryption Scheme

In this section, we restrict to the case $n = m$. We specialize the framework to a Niederreiter-style variant (the dual form of McEliece). This requires adapting the standard syndrome decoding approach to matrix codes. The resulting scheme is given in Fig. 1.

Convention. Let $N := (m + \ell_1)(m + \ell_2)$. Given a matrix code $\mathcal{C}'_{\mathrm{mat}}$ of $\mathbb{F}_q$-dimension K', viewed as a subspace of $\mathbb{F}_q^{(m+\ell_1)\times(m+\ell_2)}$, we associate with it a linear code of length N

$$\mathcal{C}' := \{\mathrm{Unfold}(\mathbf{U}) \mid \mathbf{U} \in \mathcal{C}'_{\mathrm{mat}}\}.$$

From this point onward, we will work with a parity-check matrix $\bar{\mathbf{H}}' \in \mathbb{F}_q^{(N-K')\times N}$ of $\mathcal{C}'$. By abuse of language, we also refer to $\bar{\mathbf{H}}'$ as a parity-check matrix of the

KeyGen(1^λ):
– Sample $\boldsymbol{g} \in \mathbb{F}_{q^m}^m$ whose entries are $\mathbb{F}_q$-linearly independent, and define $\mathcal{G} := \mathcal{G}_{\boldsymbol{g}}(m,k,m)$, a $[m,k]_{q^m}$ Gabidulin code that corrects rank errors of weight at most $t = \lfloor \frac{m-k}{2} \rfloor$.
– Sample uniformly at random a basis $\gamma \xleftarrow{\$} \mathcal{B}(\mathbb{F}_{q^m})$.
– Compute an $\mathbb{F}_q$-basis $\{\mathbf{A}_1, \ldots, \mathbf{A}_{mk}\}$ of the code $\Psi_\gamma(\mathcal{G})$.
– Sample a full column-rank matrix $\mathbf{T} \xleftarrow{\$} \mathbb{F}_q^{mk \times K'}$ with $1 \le K' \le mk$.
– Compute $\mathbf{A}'_j = \sum_{i=1}^{mk} T_{i,j}\mathbf{A}_i$ for $j = 1, \ldots, K'$.
– For each $j = 1, \ldots, K'$ sample independently $\mathbf{R}_j \xleftarrow{\$} \mathbb{F}_q^{m \times \ell_2}$, $\mathbf{R}'_j \xleftarrow{\$} \mathbb{F}_q^{\ell_1 \times m}$ and $\mathbf{R}''_j \xleftarrow{\$} \mathbb{F}_q^{\ell_1 \times \ell_2}$.
– Sample $\mathbf{P} \xleftarrow{\$} \mathrm{GL}_{m+\ell_1}(\mathbb{F}_q)$ and repeat until the top-left $m \times m$ block $\mathbf{S}_1$ of $\mathbf{P}^{-1}$ is invertible. Next, sample $\mathbf{Q} \xleftarrow{\$} \mathrm{GL}_{m+\ell_2}(\mathbb{F}_q)$ and repeat until $\mathbf{Q}^{-1}$ has an invertible top-left $m \times m$ block $\mathbf{S}_2$.
– For $j = 1, \ldots, K'$, define $\mathbf{M}'_j = \mathbf{P}\begin{pmatrix} \mathbf{A}'_j & \mathbf{R}_j \\ \mathbf{R}'_j & \mathbf{R}''_j \end{pmatrix}\mathbf{Q}$.
– Define the matrix code $\mathcal{C}'_{\mathrm{mat}} = \mathrm{Span}_{\mathbb{F}_q}\{\mathbf{M}'_1, \ldots, \mathbf{M}'_{K'}\}$.
– Compute a parity-check matrix $\bar{\mathbf{H}}' \in \mathbb{F}_q^{(N-K') \times N}$ for $\mathcal{C}'_{\mathrm{mat}}$ (by convention).
– Return $\mathsf{pk} = \bar{\mathbf{H}}'$ and $\mathsf{sk} = (\boldsymbol{g}, \gamma, \mathbf{P}, \mathbf{Q})$.

Encrypt$(\mathsf{pk}, \boldsymbol{\mu})$:
Input: $\mathsf{pk} = \bar{\mathbf{H}}'$, and a message $\boldsymbol{\mu} \in \mathcal{M}_{\le r}$, where $\mathcal{M}_{\le r} := \{\boldsymbol{m} \in \mathbb{F}_q^{m^2} \mid \mathrm{rank}(\mathrm{Fold}(\boldsymbol{m})) \le r\}$.
– Let $\widetilde{\mathbf{U}} \in \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_2)}$ be the matrix defined by

$$\widetilde{\mathbf{U}} = \begin{pmatrix} \mathrm{Fold}(\boldsymbol{\mu}) & \mathbf{0}_{m\times\ell_2} \\ \mathbf{0}_{\ell_1\times m} & \mathbf{0}_{\ell_1\times\ell_2} \end{pmatrix}.$$

– Define $\boldsymbol{\mu}' = \mathrm{Unfold}(\widetilde{\mathbf{U}}) \in \mathbb{F}_q^N$.
– For every integer i from 1 to N, let $\boldsymbol{h}'_i$ be the i-th column of $\bar{\mathbf{H}}'$.
– Return $\boldsymbol{c} = \sum_{i=1}^N \mu'_i \boldsymbol{h}'^{\top}_i \in \mathbb{F}_q^{N-K'}$.

Decrypt$(\mathsf{sk}, \boldsymbol{c})$:
Input: $\mathsf{sk} = (\boldsymbol{g}, \gamma, \mathbf{P}, \mathbf{Q})$ and a ciphertext $\boldsymbol{c} \in \mathbb{F}_q^{N-K'}$, where $\mathbf{S}_1, \mathbf{S}_2 \in \mathbb{F}_q^{m\times m}$ are the top-left blocks of $\mathbf{P}^{-1}$ and $\mathbf{Q}^{-1}$, respectively.
– Find any $\bar{\boldsymbol{y}} \in \mathbb{F}_q^N$ such that $\boldsymbol{c} = \sum_{i=1}^N \bar{y}_i \boldsymbol{h}'^{\top}_i$.
– Let $\mathbf{Y} = \mathrm{Fold}(\bar{\boldsymbol{y}})$. Compute $\mathbf{P}^{-1}\mathbf{Y}\mathbf{Q}^{-1}$ and delete its last ℓ_1 rows to obtain $\mathbf{M} = (\mathbf{M}_0 \mid \mathbf{M}_1) \in \mathbb{F}_q^{m\times(m+\ell_2)}$ with $\mathbf{M}_0 \in \mathbb{F}_q^{m\times m}$.
– Compute $\boldsymbol{y} = \Psi_\gamma^{-1}(\mathbf{M}_0) \in \mathbb{F}_{q^m}^m$.
– Apply the decoder of $\mathcal{G}$ to $\boldsymbol{y}$; the decoder outputs an error vector $\boldsymbol{e} \in \mathbb{F}_{q^m}^m$ of rank weight at most r (as $r \le t$).
– Let $\mathbf{E} = \Psi_\gamma(\boldsymbol{e})$ and compute $\hat{\mathbf{E}} = \mathbf{S}_1^{-1}\mathbf{E}\mathbf{S}_2^{-1}$.
– Return $\hat{\boldsymbol{\mu}} = \mathrm{Unfold}(\hat{\mathbf{E}})$.

Fig. 1: EnGMS-Niederreiter encryption scheme.

matrix code $\mathcal{C}'_{\mathrm{mat}}$ under the above identification induced by Unfold.

Key Generation Checks (Sketch). To ensure zero decryption failures for the EnGMS-Niederreiter variant, the key generation procedure enforces a simple invertibility condition on certain subblocks of the secret matrices. Concretely, the matrices $\mathbf{P}$ and $\mathbf{Q}$ are sampled from the corresponding general linear groups and are accepted only if the top-left $m \times m$ blocks of $\mathbf{P}^{-1}$ and $\mathbf{Q}^{-1}$ are invertible. Once this condition is satisfied, decryption becomes fully deterministic and cannot fail.

The validity of this approach is justified by the fact that these subblock invertibility conditions hold with non-negligible probability for uniformly random invertible matrices. As a result, the rejection-sampling step during key generation incurs only a constant expected overhead, with the expected number of resamplings remaining below 2.4 in the worst case among the considered parameter sets. The underlying analysis relies on elementary linear-algebraic counting arguments over finite fields, based on a block-matrix decomposition and the Schur complement, and is therefore deferred to Appendix E.1.

4.1.1 Message Space: Encoding and Size Estimation

We define the message space and describe its encoding and size estimation to ensure correct encryption and decryption.

(1) Message space and encoding. We define the message space as

$$\mathcal{M}_{\leq r} := \{\boldsymbol{m} \in \mathbb{F}_q^{m^2} \mid \mathrm{rank}(\mathrm{Fold}(\boldsymbol{m})) \leq r\}.$$

Since encryption takes $\boldsymbol{\mu} \in \mathcal{M}_{\leq r}$ as input (see Fig. 1), we view messages as vectors whose Fold images are low-rank matrices.

For bit-string plaintexts, we use a reversible message-encoding layer: an injective and efficiently decodable map

$$\begin{aligned} &\mathrm{Enco}_{\mathcal{M}} : \{0,1\}^{\kappa} \to \mathcal{M}_{\leq r}, \\ &\mathrm{Deco}_{\mathcal{M}} : \mathrm{Im}(\mathrm{Enco}_{\mathcal{M}}) \to \{0,1\}^{\kappa},\ \mathrm{Deco}_{\mathcal{M}}(\mathrm{Enco}_{\mathcal{M}}(x)) = x. \end{aligned}$$

Here, $\mathrm{Im}(\mathrm{Enco}_{\mathcal{M}}) := \{\mathrm{Enco}_{\mathcal{M}}(x) \mid x \in \{0,1\}^{\kappa}\}$. This requires $2^{\kappa} \leq |\mathcal{M}_{\leq r}|$, i.e., $\kappa \leq \lfloor \log_2 |\mathcal{M}_{\leq r}| \rfloor$. We stress that $\mathrm{Enco}_{\mathcal{M}}$ need not be surjective onto $\mathcal{M}_{\leq r}$; $\mathrm{Deco}_{\mathcal{M}}$ is only defined on $\mathrm{Im}(\mathrm{Enco}_{\mathcal{M}})$.

A concrete realization is as follows. Fix $s \leq r$. Parse $x \in \{0,1\}^{\kappa}$ into matrices $\mathbf{A} \in \mathbb{F}_q^{(m-s)\times s}$ and $\mathbf{B} \in \mathbb{F}_q^{s\times(m-s)}$ (by interpreting blocks of bits as elements of $\mathbb{F}_q$), so that $\kappa \leq 2s(m-s)\log_2 q$, and set

$$\mathbf{M}_s := \begin{pmatrix} \mathbf{I}_s & \mathbf{B} \\ \mathbf{A} & \mathbf{AB} \end{pmatrix}, \qquad \mathrm{Enco}_{\mathcal{M}}(x) := \mathrm{Unfold}(\mathbf{M}_s).$$

By the factorization

$$\mathbf{M}_s = \begin{pmatrix} \mathbf{I}_s \\ \mathbf{A} \end{pmatrix} (\mathbf{I}_s \ \mathbf{B}),$$

we have $\mathrm{rank}(\mathbf{M}_s) = s \le r$, hence $\mathrm{Enco}_{\mathcal{M}}(x) \in \mathcal{M}_{\le r}$.

Conversely, for $\boldsymbol{\mu} \in \mathrm{Im}(\mathrm{Enco}_{\mathcal{M}})$, let $\mathbf{M}_s = \mathrm{Fold}(\boldsymbol{\mu})$ and recover

$$\mathbf{A} = \mathbf{M}_s[(s+1) : m,\ 1 : s], \qquad \mathbf{B} = \mathbf{M}_s[1 : s,\ (s+1) : m],$$

where $\mathbf{M}_s[i_1 : i_2,\ j_1 : j_2]$ denotes the submatrix of $\mathbf{M}_s$ formed by rows $i_1, \ldots, i_2$ and columns $j_1, \ldots, j_2$. Finally, output $\mathrm{Deco}_{\mathcal{M}}(\boldsymbol{\mu})$ by applying the inverse of the fixed binary encoding used in the parsing step entrywise to $\mathbf{A}$ and $\mathbf{B}$, and concatenating the resulting bit strings in the reverse order of the parsing.

(2) Estimating the message space size. Since $\mathrm{Fold} : \mathbb{F}_q^{m^2} \to \mathbb{F}_q^{m \times m}$ is a bijection, we have

$$|\mathcal{M}_{\le r}| = \left|\{\, \mathbf{M} \in \mathbb{F}_q^{m \times m} \mid \mathrm{rank}(\mathbf{M}) \le r \,\}\right| = \sum_{k=0}^{r} N_m(k),$$

where the number of $m \times m$ matrices of rank exactly k is

$$N_m(k) = \frac{\prod_{i=0}^{k-1}(q^m - q^i)^2}{\prod_{i=0}^{k-1}(q^k - q^i)} = q^{k(2m-k)} \cdot \frac{\prod_{i=0}^{k-1}(1 - q^{i-m})^2}{\prod_{i=0}^{k-1}(1 - q^{i-k})}.$$

In particular, the largest term corresponds to $k = r$, so $|\mathcal{M}_{\le r}|$ has leading exponent $r(2m - r)$, i.e., $|\mathcal{M}_{\le r}| \approx q^{r(2m-r)}$ up to lower order factors. For the parameter sets in Section 6, the above estimate gives a very large message space $\mathcal{M}_{\le r}$, with the smallest leading-term estimate about 2^{309}.

4.1.2 Decryption Correctness

We now prove the correctness of decryption for the EnGMS-Niederreiter scheme, showing that the original message can be fully recovered.

Theorem 2 (Decryption Correctness). *In the EnGMS-Niederreiter scheme, let $\boldsymbol{\mu} \in \mathcal{M}_{\le r}$ be a message and suppose that $r \le \lfloor (m-k)/2 \rfloor$. Then the decryption algorithm outputs the correct message $\boldsymbol{\mu}$.*

Proof. Recall that $N = (m + \ell_1)(m + \ell_2)$. Let $\mathsf{pk} = \bar{\mathbf{H}}' \in \mathbb{F}_q^{(N-K') \times N}$. For $i = 1, \ldots, N$, let $\boldsymbol{h}'_i$ denote the i-th column of $\bar{\mathbf{H}}'$. Let $\boldsymbol{\mu} \in \mathcal{M}_{\le r}$ be the message such that $\mathrm{rank}(\mathrm{Fold}(\boldsymbol{\mu})) \le r$ and define

$$\widetilde{\mathbf{U}} := \begin{pmatrix} \mathrm{Fold}(\boldsymbol{\mu}) & \mathbf{0}_{m \times \ell_2} \\ \mathbf{0}_{\ell_1 \times m} & \mathbf{0}_{\ell_1 \times \ell_2} \end{pmatrix} \in \mathbb{F}_q^{(m+\ell_1) \times (m+\ell_2)}.$$

Thus $\mathrm{rank}(\widetilde{\mathbf{U}}) = \mathrm{rank}(\mathrm{Fold}(\boldsymbol{\mu})) \le r$.

Step 1: Choose $\bar{\boldsymbol{y}}$ with the given syndrome. Let $\boldsymbol{\mu}' = \mathrm{Unfold}(\widetilde{\mathbf{U}}) \in \mathbb{F}_q^N$ and define the ciphertext $\boldsymbol{c} = \sum_{i=1}^{N} \mu'_i \boldsymbol{h}_i'^{\top}$. Pick any $\bar{\boldsymbol{y}} \in \mathbb{F}_q^N$ such that $\boldsymbol{c} = \sum_{i=1}^{N} \bar{y}_i \boldsymbol{h}_i'^{\top}$. Subtracting the two identities gives

$$\sum_{i=1}^{N} (\bar{y}_i - \mu'_i)\, \boldsymbol{h}_i'^{\top} = \mathbf{0}, \quad \text{i.e.,} \quad (\bar{\boldsymbol{y}} - \boldsymbol{\mu}')\, \bar{\mathbf{H}}'^{\top} = \mathbf{0}.$$

Since $\bar{\mathbf{H}}'$ is a parity-check matrix of the linear code $\mathcal{C}' := \{\mathrm{Unfold}(\mathbf{U}) \mid \mathbf{U} \in \mathcal{C}'_{\mathrm{mat}}\}$ (by the above convention), $(\bar{\boldsymbol{y}} - \boldsymbol{\mu}')\,\bar{\mathbf{H}}'^{\top} = \mathbf{0}$ implies $\bar{\boldsymbol{y}} - \boldsymbol{\mu}' \in \mathcal{C}'$. Under the identification induced by Unfold, this is equivalent to $\mathrm{Fold}(\bar{\boldsymbol{y}} - \boldsymbol{\mu}') \in \mathcal{C}'_{\mathrm{mat}}$. Let $\mathbf{Y} = \mathrm{Fold}(\bar{\boldsymbol{y}}) \in \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_2)}$. Hence $\mathbf{Y} = \widetilde{\mathbf{U}} + \mathbf{W}$ with $\mathbf{W} \in \mathcal{C}'_{\mathrm{mat}}$.

Step 2: Recover the top-left $m \times m$ error block. Compute $\mathbf{P}^{-1}\mathbf{Y}\mathbf{Q}^{-1}$ and delete its last ℓ_1 rows, obtaining $\mathbf{M} = (\mathbf{M}_0 \mid \mathbf{M}_1) \in \mathbb{F}_q^{m\times(m+\ell_2)}$ with $\mathbf{M}_0 \in \mathbb{F}_q^{m\times m}$.

Since $\mathbf{W} \in \mathcal{C}'_{\mathrm{mat}} = \mathrm{Span}_{\mathbb{F}_q}\{\mathbf{M}'_1, \ldots, \mathbf{M}'_{K'}\}$, let $\mathbf{C}_0$ denote the top-left $m \times m$ block of the matrix obtained from $\mathbf{P}^{-1}\mathbf{W}\mathbf{Q}^{-1}$ after deleting its last ℓ_1 rows. Then $\mathbf{C}_0 \in \mathrm{Span}_{\mathbb{F}_q}\{\mathbf{A}'_1, \ldots, \mathbf{A}'_{K'}\} \subseteq \Psi_\gamma(\mathcal{G})$. Hence there exists $\boldsymbol{g}_0 \in \mathcal{G}$ such that $\Psi_\gamma(\boldsymbol{g}_0) = \mathbf{C}_0$.

Let $\boldsymbol{y} = \Psi_\gamma^{-1}(\mathbf{M}_0) \in \mathbb{F}_{q^m}^m$. Then $\boldsymbol{y} = \boldsymbol{g}_0 + \boldsymbol{e}$ for some $\boldsymbol{e} \in \mathbb{F}_{q^m}^m$. Moreover, since $\mathbf{Y} = \widetilde{\mathbf{U}} + \mathbf{W}$ and applying $\mathbf{P}^{-1}$, $\mathbf{Q}^{-1}$, and truncation does not increase rank, we have $\mathrm{wt}_R(\boldsymbol{e}) = \mathrm{rank}(\Psi_\gamma(\boldsymbol{e})) \le \mathrm{rank}(\widetilde{\mathbf{U}}) \le r \le t$. Hence the decoder of $\mathcal{G}$ succeeds and returns $\boldsymbol{e}$. Let $\mathbf{E} = \Psi_\gamma(\boldsymbol{e}) \in \mathbb{F}_q^{m\times m}$, thus $\mathbf{E}$ is known. Let

$$\Phi : \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_2)} \to \mathbb{F}_q^{m\times(m+\ell_2)}$$

be the linear truncation map that keeps the first m rows of a matrix. Let $\mathbf{V} := \Phi(\mathbf{P}^{-1}\widetilde{\mathbf{U}}\mathbf{Q}^{-1}) \in \mathbb{F}_q^{m\times(m+\ell_2)}$, where $\widetilde{\mathbf{U}}$ is the matrix obtained by applying the Fold operation to the message $\boldsymbol{\mu}$ and appending ℓ_1 zero rows and ℓ_2 zero columns.

Let $\mathbf{V} = (\mathbf{V}_0 \mid \mathbf{V}_1)$ with $\mathbf{V}_0 \in \mathbb{F}_q^{m\times m}$. Since $\Psi_\gamma(\boldsymbol{y}) = \mathbf{M}_0$, we have $\Psi_\gamma(\boldsymbol{y} - \boldsymbol{g}_0) = \mathbf{M}_0 - \mathbf{C}_0 = \mathbf{V}_0$. Therefore, $\mathbf{E} = \Psi_\gamma(\boldsymbol{e}) = \Psi_\gamma(\boldsymbol{y} - \boldsymbol{g}_0) = \mathbf{V}_0$.

Step 3: Recover message $\boldsymbol{\mu}$. By Step 2, $\mathbf{E}$ is precisely the left $m \times m$ block of $\Phi(\mathbf{P}^{-1}\widetilde{\mathbf{U}}\mathbf{Q}^{-1})$. Let $\mathbf{S}_1 \in \mathbb{F}_q^{m\times m}$ (resp., $\mathbf{S}_2 \in \mathbb{F}_q^{m\times m}$) be the top-left $m \times m$ submatrix of $\mathbf{P}^{-1}$ (resp., of $\mathbf{Q}^{-1}$). Write the block decompositions

$$\mathbf{P}^{-1} = \begin{pmatrix} \mathbf{S}_1 & * \\ * & * \end{pmatrix}, \qquad \mathbf{Q}^{-1} = \begin{pmatrix} \mathbf{S}_2 & * \\ * & * \end{pmatrix}.$$

Then

$$\mathbf{P}^{-1}\widetilde{\mathbf{U}}\mathbf{Q}^{-1} = \begin{pmatrix} \mathbf{S}_1\,\mathrm{Fold}(\boldsymbol{\mu})\mathbf{S}_2 & * \\ * & * \end{pmatrix},$$

so the top-left $m \times m$ block of $\Phi\big(\mathbf{P}^{-1}\widetilde{\mathbf{U}}\mathbf{Q}^{-1}\big)$ is $\mathbf{S}_1\,\mathrm{Fold}(\boldsymbol{\mu})\mathbf{S}_2$. Consequently, we obtain $\mathbf{E} = \mathbf{S}_1\,\mathrm{Fold}(\boldsymbol{\mu})\mathbf{S}_2$. Since $\mathbf{S}_1$ and $\mathbf{S}_2$ are invertible, it follows that $\mathrm{Fold}(\boldsymbol{\mu}) = \mathbf{S}_1^{-1}\mathbf{E}\,\mathbf{S}_2^{-1}$. Finally, decryption outputs $\mathrm{Unfold}\big(\mathbf{S}_1^{-1}\mathbf{E}\,\mathbf{S}_2^{-1}\big)$, which equals the original message $\boldsymbol{\mu}$.

4.2 Security

We now establish the OW-CPA and IND-CCA2 security guarantees of the EnGMS-Niederreiter scheme.

Theorem 3 (OW-CPA security). *Assume that no PPT algorithm solves MinRank-Syndrome$(q, m+\ell_1, m+\ell_2, K', r)$ with non-negligible probability, and that every PPT distinguisher for the EnGMS-IND problem has only negligible advantage. Then the EnGMS-Niederreiter encryption scheme is OW-CPA secure.*

Proof Sketch. The proof follows a reduction-based argument. Suppose that there exists a PPT adversary that inverts the EnGMS-Niederreiter encryption with non-negligible probability. We use this adversary as a black box and show that its success can be explained in only two ways. On the one hand, when the public key corresponds to a uniformly random matrix code, a successful inversion directly yields a low-rank solution to a MinRank-Syndrome instance. On the other hand, if the public key is sampled from the EnGMS distribution, then a non-negligible deviation in the adversary's behavior induces a distinguisher for the EnGMS-IND problem. Since both breaking the MinRank-Syndrome assumption and distinguishing EnGMS codes from random are assumed to be infeasible for PPT algorithms, the adversary's success probability must be negligible. This establishes OW-CPA security of the EnGMS-Niederreiter scheme. The full proof can be found in Appendix E.2.

Building on Theorems 2 and 3, we apply the FO transformation [30] to the EnGMS-Niederreiter encryption scheme. The resulting construction achieves IND-CCA2 security in the random oracle model.

Theorem 4 (IND-CCA2 Security). *Work in the random oracle model and assume that:*

1. *no PPT algorithm solves the MinRank-Syndrome*$(q, m+\ell_1, m+\ell_2, K', r)$ *problem with non-negligible probability; and*
2. *no PPT distinguisher for the EnGMS-IND problem exists with non-negligible advantage.*

Let $\mathsf{PKE}_{\mathrm{N}}$ *denote the EnGMS-Niederreiter public-key encryption scheme, and let* G *and* H *be independent random oracle hash functions instantiated as required by the FO transformation. Applying the transformation* $\mathrm{FO}_m^{\not\perp}[\mathsf{PKE}_{\mathrm{N}}, \mathsf{G}, \mathsf{H}]$ *yields a key encapsulation mechanism* $\mathsf{KEM}_{\mathrm{N}}^{\mathrm{FO}}$*. Let* $\mathsf{PKE}_{\mathrm{N}}^{\mathrm{FO}}$ *denote the public-key encryption scheme obtained from* $\mathsf{KEM}_{\mathrm{N}}^{\mathrm{FO}}$ *via the common KEM-DEM hybrid encryption framework. Under the above assumptions, both* $\mathsf{KEM}_{\mathrm{N}}^{\mathrm{FO}}$ *and* $\mathsf{PKE}_{\mathrm{N}}^{\mathrm{FO}}$ *achieve IND-CCA2 security.*

The full descriptions of $\mathsf{KEM}_{\mathrm{N}}^{\mathrm{FO}}$ and $\mathsf{PKE}_{\mathrm{N}}^{\mathrm{FO}}$ are provided in Appendix E.3 (Figs. 5 and 6).

4.3 Algorithmic Complexity

We turn to the complexity analysis of the Niederreiter variant. To streamline notation, set $N = (m+\ell_1)(m+\ell_2)$.

- *Key generation cost.* The dominant cost comes from: (i) constructing K' intermediate matrices $\mathbf{A}'_j$ from linear combinations of $m \times m$ matrices; (ii) computing the K' masked matrices $\mathbf{M}'_j$, where $\mathbf{M}'_j \in \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_2)}$; and (iii) deriving the public parity-check matrix $\bar{\mathbf{H}}' \in \mathbb{F}_q^{(N-K')\times N}$ via Gaussian elimination. With classical matrix multiplication, the overall computational cost is $O\Big(K'\big(km^3 + (2m+\ell_1+\ell_2)N\big) + (N-K')^2 N\Big)$ operations in $\mathbb{F}_q$.

– *Encryption cost.* The encryption algorithm outputs the syndrome $\boldsymbol{c} = \boldsymbol{\mu}' \bar{\mathbf{H}}'^{\top}$, with $\bar{\mathbf{H}}' \in \mathbb{F}_q^{(N-K')\times N}$. Computing $\boldsymbol{c}$ costs $O(N(N-K'))$ operations in $\mathbb{F}_q$.
– *Decryption cost.* The decryption process requires solving a linear system over $\mathbb{F}_q$ with N unknowns, i.e., finding any $\bar{\boldsymbol{y}} \in \mathbb{F}_q^N$ such that $\bar{\boldsymbol{y}} \bar{\mathbf{H}}'^{\top} = \boldsymbol{c}$. Using Gaussian elimination, this costs $O(N^3)$ operations in $\mathbb{F}_q$.

5 Structural Attacks and Message-Recovery Attacks

Where Does the Hardness Come From? A natural question is whether the use of subcodes merely shifts the structure without changing the attack surface. Our analysis shows that this is not the case. In previous constructions, extension-field linearity provides the algebraic pathway exploited by the structural attacks considered here, reducing the relevant distinguishing or recovery tasks to algebraic problems of comparatively low complexity. By contrast, once the dimension constraint $m \nmid K'$ is enforced, this known direct pathway is no longer available from the public code itself. An adversary would typically need to recover an intermediate code with sufficient inherited structure, or else resort to generic $\mathbb{F}_q$-linear matrix-code attacks; for message recovery, the latter are captured by the corresponding MinRank problem. Thus, the resulting security gain does not arise merely from hiding structure, but from blocking the direct algebraic pathway that connects the public code to these attacks.

We next analyze the impact of known structural attacks, together with closely related attack variants, on the EnGMS-Niederreiter construction presented in Section 4 (and on the EnGMS-McEliece scheme described in Appendix D). Our focus is on attacks that exploit algebraic regularities in the public code or in derived projected/completed codes, rather than on generic message-recovery attacks.

The known structural attacks developed for Gabidulin-based and matrix-Gabidulin-based schemes typically rely on directly exploitable $\mathbb{F}_{q^m}$-linear or Frobenius-related structure. For parameter choices with $m \nmid K'$, the EnGMS public code is not $\mathbb{F}_{q^m}$-linear. Thus, the algebraic preconditions required by these structural attacks [5, 43] do not hold directly for the public code.

The purpose of this section is not to claim unconditional resistance to all conceivable attacks, but rather to clarify which known structural mechanisms no longer apply directly, and which ones would require auxiliary steps to expose sufficient inherited structure.

We organize the discussion according to two broad cryptanalytic objectives relevant to our construction:

1) *Structural Distinguishers and Key-Recovery Attacks:* In this first class, the adversary exploits residual algebraic structure in the public code or in suitably derived codes, either to distinguish the public distribution from random or to reconstruct an equivalent hidden description of the secret Gabidulin structure. In our setting, full key recovery should be understood as recovering enough hidden structure to derive an equivalent decryption procedure (closely related to the EnGMS-Search problem), and the attacks considered below also include distinguishers that stop short of full secret-key recovery.

2) *Message Recovery Attacks:* In this second class, the adversary does not seek an equivalent secret key, but only the plaintext corresponding to a given ciphertext and public key. For our construction, this reduces to solving the corresponding MinRank instance induced by the public code, equivalently viewed in syndrome form for the Niederreiter variant.

5.1 Attacks on the Key

Recovering the secret key can be viewed as solving the EnGMS-Search problem: given a public matrix code $\mathcal{C}'_{\mathrm{mat}}$ sampled from the EnGMS$(k, m, n, K', \ell_1, \ell_2)$ distribution, recover a secret $\mathbb{F}_q$-basis $\gamma \in \mathcal{B}(\mathbb{F}_{q^m})$ and the Gabidulin evaluation vector $\boldsymbol{g} \in \mathbb{F}_{q^m}^n$. Since efficient structural attacks are known for Gabidulin matrix codes, EGMC applies a masking transformation combining random row and column padding with left and right multiplication by secret invertible matrices $\mathbf{P}$ and $\mathbf{Q}$. Despite this masking, Aragon et al. [5] consider two distinguishers against EGMC. The first is a combinatorial attack that detects the underlying $\mathbb{F}_{q^m}$-linear structure, whereas the second is an algebraic attack in the spirit of Overbeck. We also consider the key-recovery attack of Porwal *et al.* [43], which attacks the EGMC scheme and relies on the same $\mathbb{F}_{q^m}$-linear expansion structure.

Motivated by these attacks, our scheme is designed to suppress the underlying $\mathbb{F}_{q^m}$-linear structure of the public code by publishing only a K'-dimensional subcode. We then analyze the impact of these attacks on our construction.

We keep the notation of Definitions 12 and 13. Let $\mathbf{P} \in \mathrm{GL}_{m+\ell_1}(\mathbb{F}_q)$ and $\mathbf{Q} \in \mathrm{GL}_{n+\ell_2}(\mathbb{F}_q)$ be random matrices. Our public matrix code, denoted by $\mathcal{C}'_{\mathrm{pub}}$ (and by $\mathcal{C}_{\mathrm{pub}}$ in the EGMC scheme), is given by a basis of K' matrices:

$$\mathcal{RB}'_{\mathrm{pub}} \;=\; \left\{ \mathbf{P} \begin{pmatrix} \mathbf{A}'_1 & \mathbf{R}_1 \\ \mathbf{R}'_1 & \mathbf{R}''_1 \end{pmatrix} \mathbf{Q} \;,\; \dots \;,\; \mathbf{P} \begin{pmatrix} \mathbf{A}'_{K'} & \mathbf{R}_{K'} \\ \mathbf{R}'_{K'} & \mathbf{R}''_{K'} \end{pmatrix} \mathbf{Q} \right\},$$

where the matrices $\mathbf{A}'_i$ form an $\mathbb{F}_q$-basis of a subcode of $\Psi_\gamma(\mathcal{G})$, and $\mathbf{R}_i$, $\mathbf{R}'_i$, $\mathbf{R}''_i$ are random matrices of respective sizes $m \times \ell_2$, $\ell_1 \times n$ and $\ell_1 \times \ell_2$.

5.1.1 A Combinatorial Distinguisher Targeting Hidden $\mathbb{F}_{q^m}$-Linearity

We first recall the combinatorial distinguisher used in [5], together with its underlying idea and complexity estimate. Given a public matrix code $\mathcal{C}_{\mathrm{pub}}$, the distinguisher operates as follows.

- *Projections via* $\mathbf{U}'$ *and* $\mathbf{V}'$*:* One selects projection matrices $\mathbf{U}' \in \mathbb{F}_q^{m \times (m+\ell_1)}$ and $\mathbf{V}' \in \mathbb{F}_q^{(n+\ell_2)\times(k+1)}$ of appropriate rank, and applies them to $\mathcal{C}_{\mathrm{pub}}$ from the left and the right. The resulting projected code $\mathbf{U}'\mathcal{C}_{\mathrm{pub}}\mathbf{V}'$ is designed to suppress the contribution of the random padding components while retaining sufficient information about the underlying $\mathbb{F}_{q^m}$-linear structure, possibly at the cost of partially puncturing the original Gabidulin code.

– *Stabilizer algebra test:* For each choice of $(\mathbf{U}', \mathbf{V}')$, one computes the left stabilizer algebra of $\mathbf{U}'\mathcal{C}_{\text{pub}}\mathbf{V}'$. If the dimension of this stabilizer is at least m, the code is identified as exhibiting a Gabidulin-type structure; otherwise, it is deemed indistinguishable from a random matrix code.

A suitable $(\mathbf{U}', \mathbf{V}')$ is found with probability approximately $q^{-(m\ell_1+(k+1)\ell_2)}$, yielding an expected running time of $\widetilde{O}\big(q^{m\ell_1+(k+1)\ell_2}\big)$.

This procedure thus provides a statistical test for distinguishing public codes derived from Gabidulin constructions from uniformly random matrix codes of the same parameters.

1) Case $m \nmid K'$: Absence of $\mathbb{F}_{q^m}$-Linearity

We now analyze the behavior of the above distinguisher against our construction in the case $m \nmid K'$.

– *Absence of $\mathbb{F}_{q^m}$-linearity:* By Lemma 1, the $\mathbb{F}_q$-linear space

$$\mathcal{S} = \text{Span}_{\mathbb{F}_q}\{\mathbf{A}'_1, \ldots, \mathbf{A}'_{K'}\}$$

is not $\mathbb{F}_{q^m}$-linear.

– *Impact on the stabilizer-based test:* An adversary may nevertheless attempt to emulate the strategy of [5] by selecting projection matrices $\mathbf{U}'$ and $\mathbf{V}'$ and forming $\mathbf{U}'\mathcal{C}'_{\text{pub}}\mathbf{V}'$. However, without such $\mathbb{F}_{q^m}$-linearity in $\mathcal{S}$, the dimension of the left stabilizer algebra of the projected code is expected to be comparable to that observed for a uniformly random matrix code. Consequently, the stabilizer dimension no longer provides a reliable distinguishing criterion.

Hence, the combinatorial distinguisher of [5] does not yield a non-negligible advantage when applied directly to our public code in this parameter regime.

2) Case $m \mid K'$: Conservative Scenario

We next consider the case where $K' = mk'$ with $k' < k$. While the condition $m \mid K'$ is necessary for $\mathcal{S} = \text{Span}_{\mathbb{F}_q}\{\mathbf{A}'_1, \ldots, \mathbf{A}'_{K'}\}$ to be $\mathbb{F}_{q^m}$-linear, it is by no means sufficient. Nevertheless, in order to assess the most conservative scenario, we assume that $\mathcal{S}$ does admit an $\mathbb{F}_{q^m}$-linear structure. Under this assumption, the combinatorial distinguisher of [5] applies essentially as in the EGMC setting. This is formalized below.

Theorem 5 (Combinatorial Distinguisher for $K' = mk'$). *Let $\ell'_1, \ell'_2 \geq 0$, and consider the public matrix code of size $(m + \ell'_1) \times (n + \ell'_2)$. The distinguisher of [5] requires retaining $n_\star > k'$ columns after projection. A single trial succeeds with probability approximately $q^{-(m\ell'_1+n_\star\ell'_2)}$, yielding an expected complexity $\widetilde{O}\big(q^{m\ell'_1+n_\star\ell'_2}\big)$. Setting $n_\star = k' + 1$ yields complexity $\widetilde{O}(q^{m\ell'_1+(k'+1)\ell'_2})$.*

Let $E_{\text{EGMC}} = m\ell_1 + (k+1)\ell_2$ denote the exponent associated with the original EGMC parameters (ℓ_1, ℓ_2, k), $E_{\text{EnGMS}} = m\ell'_1 + (k'+1)\ell'_2$ be the corresponding exponent for our parameters (ℓ'_1, ℓ'_2, k'). Ensuring $E_{\text{EnGMS}} \geq E_{\text{EGMC}}$ would require increasing at least one of ℓ'_1 or ℓ'_2 due to the inequality $k' < k$.

In practice, this entails the following trade-offs:

1. When $m = n$, it is common to choose $\ell'_1 = \ell'_2$, as near-square matrices are empirically observed to mitigate algebraic attacks against MinRank. Increasing these parameters increases the matrix size $(m + \ell'_1) \times (n + \ell'_2)$, thereby raising the cost of the scheme's linear-algebraic operations.
2. Larger values of ℓ'_1 or ℓ'_2 directly increase public-key and ciphertext sizes.

As a result, when $K' = mk'$ with $k' < k$, the complexity $\widetilde{O}(q^{m\ell'_1+(k'+1)\ell'_2})$ is typically smaller than in the original EGMC setting, rendering the combinatorial distinguisher comparatively more effective. For this reason, we favor parameter choices with $m \nmid K'$, for which the selected subcode is provably not $\mathbb{F}_{q^m}$-linear and the above distinguisher no longer applies directly.

5.1.2 An Overbeck-Like Distinguisher

We recall the Overbeck-like distinguisher described in [5], which aims at detecting a hidden $\mathbb{F}_{q^m}$-linear Gabidulin structure by exploiting q-polynomial representations and Frobenius-related algebraic properties. The distinguisher is analyzed under the dimensional condition

$$2\dim_{\mathbb{F}_q} \mathcal{C}_{\mathrm{pub}} \;\geq\; (m+\ell_1)(n+\ell_2).$$

The approach constructs an auxiliary matrix code $\mathcal{D}' \subseteq \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_1)}$ obtained from the code $\mathcal{D}$ of [5, Proposition 7] via the similarity transformation induced by the left scrambler $\mathbf{P}$, namely $\mathcal{D}' := \mathbf{P}\mathcal{D}\mathbf{P}^{-1}$. It is shown in [5, Corollary 1] that left multiplication by elements of $\mathcal{D}'$, that is, the map $\mathbf{C} \in \mathcal{C}_{\mathrm{pub}} \mapsto \mathbf{DC}$ for $\mathbf{D} \in \mathcal{D}'$, does not span the full ambient space $\mathbb{F}_q^{(m+\ell_1)\times(n+\ell_2)}$ when $\mathcal{C}_{\mathrm{pub}}$ arises from an $\mathbb{F}_{q^m}$-linear Gabidulin construction. Consequently, under the above dimensional condition, there exists a nonzero matrix $\mathbf{M} \in \mathcal{C}_{\mathrm{pub}}^{\perp}$ such that

$$\mathrm{tr}(\mathbf{DCM}^{\top}) = 0 \quad \text{for all } \mathbf{C} \in \mathcal{C}_{\mathrm{pub}},$$

where $\mathbf{D}$ ranges over a structured subspace of $\mathcal{D}'$. In contrast, a uniformly random $\mathcal{C}_{\mathrm{pub}}$ would typically not admit such a nonzero $\mathbf{M}$.

Instantiating these relations for a basis of $\mathcal{C}_{\mathrm{pub}}$ yields a bilinear system in the unknowns describing $\mathbf{D}$ and $\mathbf{M}$ (quadratic when expanded entry-wise). For parameter choices with $k = \Theta(m)$ and $m \approx n$, this system typically involves $\Theta(m^2)$ equations in $\Theta(m^2)$ algebraic unknowns; a naive linearization introduces $\Theta(m^4)$ variables. For the moderate parameter sizes considered in [5] (e.g., $m \approx 40$), this places the direct solution of the resulting system beyond the reach of elementary algebraic techniques. Unless a substantially more efficient modeling or solving strategy is identified, this Overbeck-like approach does not appear to be practical in this regime.

1) Case $m \nmid K'$: Absence of $\mathbb{F}_{q^m}$-Linearity

Assume $m \nmid K'$ and $2K' \geq (m+\ell_1)(n+\ell_2)$. Then $\mathcal{S} = \mathrm{Span}_{\mathbb{F}_q}\{\mathbf{A}'_1, \ldots, \mathbf{A}'_{K'}\}$ is not $\mathbb{F}_{q^m}$-linear and, in particular, cannot be the Ψ_γ-expansion of any $\mathbb{F}_{q^m}$-linear code. Consequently, the public code does not exhibit the relevant algebraic invariants exploited by Overbeck-like distinguishers.

Using the notation of [5, Proposition 7], let $\mathcal{C}'_0 = \mathbf{P}^{-1}\mathcal{C}'_{\text{pub}}\mathbf{Q}^{-1}$. For $\mathbf{D} \in \mathcal{D} \subseteq \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_1)}$ and $\mathbf{C} \in \mathcal{C}'_0 \subseteq \mathbb{F}_q^{(m+\ell_1)\times(n+\ell_2)}$, consider the block decompositions

$$\mathbf{D} = \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{T}_1 & \mathbf{T}_2 \end{pmatrix}, \qquad \mathbf{C} = \begin{pmatrix} \mathbf{A} & \mathbf{R} \\ \mathbf{R}' & \mathbf{R}'' \end{pmatrix}.$$

Here, $\mathbf{B} \in \Psi_\gamma(\mathcal{G}(m, n-k, m))$ and $\mathbf{A} \in \mathcal{S}$, while the remaining blocks are arbitrary matrices of compatible dimensions, namely $\mathbf{T}_1 \in \mathbb{F}_q^{\ell_1\times m}$, $\mathbf{T}_2 \in \mathbb{F}_q^{\ell_1\times\ell_1}$, $\mathbf{R} \in \mathbb{F}_q^{m\times\ell_2}$, $\mathbf{R}' \in \mathbb{F}_q^{\ell_1\times n}$ and $\mathbf{R}'' \in \mathbb{F}_q^{\ell_1\times\ell_2}$.

Theorem 6. *Let* $\mathcal{U}_{BA} = \text{Span}_{\mathbb{F}_q}\{\mathbf{BA} \mid \mathbf{B} \in \Psi_\gamma(\mathcal{G}(m, n-k, m)),\ \mathbf{A} \in \mathcal{S}\}$. *Then*

$$\dim_{\mathbb{F}_q}(\mathcal{U}_{BA}) \leq \min\{m(n-k)K',\, mn\}.$$

Proof. Since $\Psi_\gamma(\mathcal{G}(m, n-k, m))$ has dimension $m(n-k)$, and $\mathcal{S}$ has dimension K', we immediately have: $\dim_{\mathbb{F}_q}(\mathcal{U}_{BA}) \leq m(n-k)\cdot K'$.

Simultaneously, since $\mathbf{B} \in \mathbb{F}_q^{m\times m}$ and $\mathbf{A} \in \mathbb{F}_q^{m\times n}$, the product $\mathbf{BA}$ resides in the ambient space $\mathbb{F}_q^{m\times n}$, which has dimension mn. The dimension of $\mathcal{U}_{BA}$ is therefore bounded by: $\dim_{\mathbb{F}_q}(\mathcal{U}_{BA}) \leq \min\{m(n-k)K',\, mn\}$.

Impact on Overbeck-Like Attacks. The Overbeck-like argument of [5] crucially relies on the existence of a constrained $\mathbb{F}_{q^m}$-expansion, leading to a dimension drop relative to the ambient space $\mathbb{F}_q^{(m+\ell_1)\times(n+\ell_2)}$ as formalized in [5, Corollary 1]. In our setting with $m \nmid K'$ and $2K' \geq (m+\ell_1)(n+\ell_2)$, Theorem 6 yields only the generic bound $\dim_{\mathbb{F}_q}(\mathcal{U}_{BA}) \leq mn$, matching the dimension of the full matrix space $\mathbb{F}_q^{m\times n}$. For a random choice of the subcode $\mathcal{S}$, one therefore expects $\mathcal{U}_{BA}$ to span the ambient space $\mathbb{F}_q^{m\times n}$ with high probability.

As a result, the dimensional collapse exploited in [5] does not occur, and the resulting bilinear system no longer exhibits the structural imbalance required for a practical Overbeck-like distinguisher. In this sense, the Overbeck-like technique of [5] does not directly extend to our parameter regime.

2) Case $m \mid K'$: Conservative Scenario

We consider the case $K' = mk'$ with $k' < k$. The Overbeck-like distinguisher of [5] requires not only $\mathbb{F}_{q^m}$-linearity but also that $\mathcal{S} = \text{Span}_{\mathbb{F}_q}\{\mathbf{A}'_1, \ldots, \mathbf{A}'_{K'}\}$ arises as the Ψ_γ-expansion of an $\mathbb{F}_{q^m}$-linear k'-dimensional Gabidulin code. The divisibility condition $m \mid K'$ is therefore necessary but not sufficient.

To assess the most conservative scenario, we assume that $\mathcal{S}$ indeed has such a Ψ_γ-expanded Gabidulin structure and that $2mk' \geq (m+\ell_1)(n+\ell_2)$. Under these assumptions, the Overbeck-like distinguisher can be instantiated analogously to the EGMC case, with k replaced by k'.

The resulting bilinear system contains mk' equations. The number of unknowns, arising from both the $\mathbf{D}$-part and the $\mathbf{M}$-part, remains of the same order as in the original EGMC analysis. In particular, for $k' = \Theta(m)$ and $m \approx n$, linearization again leads to $\Theta(m^4)$ variables for only $\Theta(m^2)$ equations.

Replacing k by k' thus decreases the number of equations without reducing the number of unknowns, and does not yield an asymptotic advantage for the

attacker. Consequently, even under this pessimistic structural assumption, the Overbeck-like approach does not appear easier than in the original EGMC setting for the parameter sizes considered in this work.

5.1.3 Applicability of the Key-Recovery Attack in [43]

Porwal *et al.* [43] recently introduced an exponential-time key-recovery attack targeting the EGMC scheme. Unlike previous distinguisher-based approaches [5], this attack aims to recover an equivalent secret key. The asymptotic complexity of the attack, up to polynomial factors, improves upon the previously best-known distinguisher. Moreover, concrete estimates provided in [43] suggest that several original EGMC parameter sets no longer meet their claimed security levels.

From a structural perspective, however, the efficacy of the new attack is predicated on the existence of hidden $\mathbb{F}_{q^m}$-linear expansions. More precisely, the public code is modeled as a masked matrix code of the form $\Psi_\gamma(\mathcal{C}_{\mathrm{vec}})$ for an $[n,k]_{\mathbb{F}_{q^m}}$ code $\mathcal{C}_{\mathrm{vec}}$. The analysis explicitly assumes that $\dim_{\mathbb{F}_q} \mathrm{Stab}_L(\Psi_\gamma(\mathcal{C}_{\mathrm{vec}})) = m$, a property invoked in the proof of Theorem 4 of [43]. Indeed, the success of the recovery procedure (Algorithm 1 in [43]) is strictly contingent on the input code being an $\mathbb{F}_{q^m}$-linear expansion with stabilizer dimension m.

In contrast, under our parameter choice, the public code is derived from an $\mathbb{F}_q$-subcode that does not admit the direct $\mathbb{F}_{q^m}$-linear expansion structure required by this attack. Since this structural requirement is not met, the key-recovery attack of [43] cannot be applied to our construction as stated.

Remark 1. In [43], the naive key-recovery algorithm and the ACDGV distinguisher are formulated for public codes arising from hidden $\mathbb{F}_{q^m}$-linear expansions, for which the stabilizer algebra must be non-trivial (see Algorithms 2 and 3 in [43]). When $m \nmid K'$, the stabilizer-based distinguishing strategy is no longer effective. Hence, these attacks do not apply as stated in this setting.

5.1.4 A Combinatorial Distinguisher via Subcode Completion

Inspired by the combinatorial distinguisher of [5], we consider the following iterative procedure:

- Guess matrices $\mathbf{U}' \in \mathbb{F}_q^{m\times(m+\ell_1)}$ and $\mathbf{V}' \in \mathbb{F}_q^{(n+\ell_2)\times n^\star}$, where $k < n^\star \le n$, and compute the projected code $\mathcal{S}^* := \mathbf{U}'\mathcal{C}'_{\mathrm{pub}}\mathbf{V}' \subseteq \mathbb{F}_q^{m\times n^\star}$. Denote $\ell := \dim_{\mathbb{F}_q}(\mathcal{S}^*) \le K'$.
- Complete $\mathcal{S}^*$ to an mk-dimensional candidate code, viewed as an $\mathbb{F}_q$-subspace of $\mathbb{F}_q^{m\times n^\star}$ containing $\mathcal{S}^*$.
- Compute the left stabilizer algebra of the resulting code.

This procedure is repeated until a code whose left stabilizer algebra has dimension at least m is obtained.

In the favorable case, one may take projection matrices of the form

$$\mathbf{U}' = (\mathbf{U_0} \mid \mathbf{0}_{m\times\ell_1})\mathbf{P}^{-1}, \qquad \mathbf{V}' = \mathbf{Q}^{-1}\begin{pmatrix}\mathbf{V_0}\\ \mathbf{0}_{\ell_2\times n^\star}\end{pmatrix},$$

with $\mathbf{U_0} \in \mathrm{GL}_m(\mathbb{F}_q)$ and $\mathbf{V_0} \in \mathbb{F}_q^{n\times n^\star}$ of full rank. For such a choice of $(\mathbf{U}', \mathbf{V}')$, the analysis in [5] shows that the projected code $\mathcal{S}^*$ is contained in the matrix expansion of an $\mathbb{F}_{q^m}$-linear Gabidulin code of length $n^\star$ and dimension k. Our goal in Step 2 is then to complete $\mathcal{S}^*$ to the corresponding mk-dimensional expanded code inside $\mathbb{F}_q^{m\times n^\star}$. However, the projected code does not necessarily have dimension K'. We therefore write $\ell := \dim_{\mathbb{F}_q}(\mathcal{S}^*)$, with $\ell \le K'$.

Complexity of step 1. From Section 5.1.1, the expected cost of reaching the target projected situation is $\widetilde{O}\big(q^{\,m\ell_1+n^\star\ell_2}\big)$.

Complexity of step 2. The ambient quotient space $\mathbb{F}_q^{m\times n^\star}/\mathcal{S}^*$ has dimension $mn^\star - \ell$. Therefore, the number of possible completions, counted at the level of $\mathbb{F}_q$-subspaces, is the Gaussian binomial

$$\binom{mn^\star-\ell}{mk-\ell}_q \approx q^{(mk-\ell)\left((mn^\star-\ell)-(mk-\ell)\right)} = q^{m(mk-\ell)(n^\star-k)}.$$

Thus, the expected cost of Step 2 is $\widetilde{O}\big(q^{m(mk-\ell)(n^\star-k)}\big)$.

Total expected complexity. Each iteration of Step 1 invokes Step 2 to complete the candidate to mk dimensions. Step 3 computes the left stabilizer algebra, a polynomial-time operation neglected in the asymptotic analysis.

Combining the expected costs of Steps 1 and 2, the expected runtime is $\widetilde{O}\Big(q^{\,m\ell_1+n^\star\ell_2+m(mk-\ell)(n^\star-k)}\Big)$. For the optimal parameters $(n^\star = k+1, \ell = K')$, the expected cost simplifies to

$$\widetilde{O}\Big(q^{\,m\ell_1+(k+1)\ell_2+m(mk-K')}\Big). \tag{1}$$

Remark 2. For the proposed parameters, this distinguisher has complexity well above the targeted security level. Allowing completions to any equivalent mk-dimensional expansion reduces the search space by a multiplicity factor ν. In hidden-code search settings, this effect is modeled as a speed-up from T to T/ν, where T denotes the work factor of the search attack. Given the conjecture in [43] that $\nu = 1$ with high probability, such completions are not expected to compromise the security of our construction.

Remark 3. The naive key-recovery algorithm and the ACDGV distinguisher from [43] incorporating subcode completion necessarily face a complexity lower bounded by $\widetilde{O}(q^{\,m(mk-K')})$ to restore the prerequisite algebraic structure. For our recommended parameters, this completion cost alone exceeds the targeted security level by a margin of at least 273 bits.

Furthermore, the key-recovery attack of [43] satisfies the following conservative complexity bound when augmented with subcode completion:

$$\widetilde{O}\Big(k(m-k)m^6 \cdot q^{\,b_1\ell_1+b_2\ell_2+(mk-K')(b_1b_2-mk)}\Big), \tag{2}$$

where the pair (b_1, b_2) is optimized according to [43, Theorem 4]. For our proposed parameter sets, this estimate strictly exceeds the targeted security level with a minimum margin of 41 bits. Consequently, these hybridizations do not compromise the security of our scheme.

5.1.5 A Structural Distinguisher via Subcode Projection

We evaluate the structural distinguisher from [14] based on the observation that, in the successful projection case of Section 5.1.4, the code $\mathcal{S}^* \subseteq \mathbb{F}_q^{m\times n^\star}$ is inherited from the matrix expansion of an $\mathbb{F}_{q^m}$-linear Gabidulin code. This specific structure leads to the following rank weight characteristics.

- *Rank weight distribution.* The subcode $\mathcal{S}^*$ satisfies $d(\mathcal{S}^*) \geq n^\star - k + 1$, which precludes rank $n^\star - k$ codewords. Conversely, random codes of equivalent dimension possess such low rank elements.
- *Parameter regime.* For $\ell := \dim_{\mathbb{F}_q}(\mathcal{S}^*) = am + b \leq K'$ with $0 < b \leq m$, our parameters fall into the critical regime identified in [14], where $s = n^\star - k < n^\star - a$. In this setting, the absence of such rank-s codewords distinguishes $\mathcal{S}^*$ from the corresponding random-code distribution.
- *Attack complexity.* The attack requires a projection step followed by a rank search. The combined expected complexity is

$$\widetilde{O}\Big((km^w + mnk^{w-1})q^{\,m\ell_1+n^\star\ell_2+a(n^\star-m-k)+k(m+k-n^\star)+\min(0,\,n^\star-k-b+1)}\Big).$$

  Setting $a = k - 1$ to minimize the work factor yields

$$\widetilde{O}\Big((km^w + mnk^{w-1})q^{\,m(\ell_1+1)+n^\star(\ell_2-1)+k+\min(0,\,n^\star-k-b+1)}\Big), \tag{3}$$

  where w denotes the matrix multiplication exponent.

Remark 4. In the projected setting considered above, $\mathcal{S}^\star$ is contained in the matrix expansion of a Gabidulin code, and our analysis only uses the resulting low-rank distinguisher from [14]. The subsequent key-recovery attack in [14] relies on a stronger random-subcode model inside a hidden Gabidulin matrix code, together with the recovery of the hidden $\mathbb{F}_{q^m}$-linear structure. Since $\mathcal{S}^\star$ arises from our specific projection procedure and is not modeled as such a random subcode, the key-recovery attack of [14] does not directly apply to our setting as stated, while the low-rank distinguisher is retained as a conservative estimate.

5.1.6 Attacks on the Dual Code

When the dimension of the EGMC public code satisfies

$$2\dim_{\mathbb{F}_q}\mathcal{C}_{\mathrm{pub}} \;<\; (m+\ell_1)(n+\ell_2),$$

the rate of its dual code exceeds $1/2$. In this regime, the Overbeck-like distinguisher of [5] no longer applies directly to the public code, which naturally motivates considering analogous attacks on the dual code.

We therefore examine the dual of the EnGMS public code under the same parameter conditions and compare its behavior with the EGMC dual setting. For fixed $(k, m, n, \ell_1, \ell_2)$, the Overbeck-like approach yields a strictly more restrictive distinguishing condition than in EGMC, both when $m \nmid K'$ and, under

a conservative structural assumption, when $m \mid K'$. As a consequence, the admissible parameter range in which the dual distinguisher may apply is reduced, making the attack harder to trigger.

Regarding combinatorial distinguishers, when $m \mid K'$ (and under a pessimistic structural scenario) the dual-side search problem is essentially comparable to the public-code case, while for $m \nmid K'$ we are not aware of a dual-code approach achieving strictly lower complexity than existing public-code attacks.

Due to space constraints, the detailed technical analysis of attacks on the dual code is deferred to Appendix F.2.

5.2 Attacks on the Message

Recovering a message without knowledge of the secret key in our scheme amounts to solving a MinRank$(q,\, m+\ell_1,\, m+\ell_2,\, K',\, r)$ instance. We briefly review the main classes of attacks against this problem and discuss their relevance at the parameter sizes considered in this work. Throughout this section, ω denotes the exponent of matrix multiplication, i.e., the smallest constant such that two $s \times s$ matrices over $\mathbb{F}_q$ can be multiplied in $O(s^\omega)$ field operations. We defer the numerical instantiation of these attacks to Section 6, where we verify that all considered parameters meet the claimed security levels.

Kernel Attack. The kernel attack, introduced in [27], exploits the fact that a rank-r matrix $\mathbf{E}$ admits a kernel of codimension at most r. The attacker repeatedly samples vectors in the hope that they belong to this kernel and uses the resulting relations to derive a linear system whose solution reveals $\mathbf{E}$. For MinRank(q, m, n, K, r) instances, the expected complexity of this approach is

$$O\Big(q^{\,r\left\lceil \frac{K}{\max\{m,n\}} \right\rceil} \cdot K^\omega\Big)\,, \tag{4}$$

where the factor $q^{\,r\left\lceil \frac{K}{\max\{m,n\}} \right\rceil}$ corresponds to the expected number of trials needed to sample sufficiently many kernel vectors, and the factor K^ω accounts for the cost of solving the induced linear system. For the parameter sets considered in this work, this complexity is well above the targeted security levels.

Minors Attack. The minors attack is an algebraic approach that models MinRank as a system of polynomial equations by setting all $(r{+}1)$-minors of the unknown matrix $\mathbf{E}$ to zero [18, 26]. The resulting determinantal system is typically solved using Gröbner-basis algorithms.

In the generic well-defined regime $K = (m-r)(n-r)$, the complexity of computing a grevlex Gröbner basis is commonly estimated as $O(M(d_{\mathrm{reg}})^\omega)$ [18], where d_{reg} denotes the degree of regularity and $M(d)$ the number of monomials of total degree at most d. Several refinements have been proposed, notably in [26], exploiting syzygies of determinantal ideals to cut down redundant eliminations. They remain within the Gröbner-basis framework, with cost bounded by linear algebra on the intermediate Macaulay matrices. At the parameter sizes relevant to our construction, the resulting systems are far beyond the reach of current algebraic solvers.

Support Minors Attack. The support minors attack, introduced in [9], combines a hybridization step on a low-dimensional support representation of the matrix $\mathbf{E}$ with a minors-based algebraic resolution. Its complexity is commonly estimated using hybrid techniques that balance the cost of this hybridization against the cost of solving the resulting reduced system.

In our analysis, we evaluate this attack using the puncturing–hybrid variant implemented in the public estimator `https://estimators.crypto.tii.ae` [17], which optimizes the attack parameters and outputs a concrete work factor combining the hybridization cost with linear-algebra solving steps. For all parameter sets considered, the estimated complexity of the support minors attack remains above the claimed security levels.

Concrete complexity estimates for these attacks, instantiated at our chosen parameter sets, are summarized in Tables 1 and 2 (see Section 6).

Remark 5. The fast linearization regimes in [9,32] do not appear to apply to our scheme. Thus, they are not expected to outperform the attacks considered.

6 Parameter Selection and Size Comparison

6.1 Parameters and Sizes for EnGMS-Niederreiter

We instantiate the EnGMS-Niederreiter encryption framework using a Gabidulin code $\mathcal{G}[m,k]_{q^m}$, which admits an efficient decoding algorithm correcting errors of rank weight up to $\lfloor \frac{m-k}{2} \rfloor$. As established above, decrypting a ciphertext $\boldsymbol{c}$ without knowledge of the secret key sk can be formulated as solving a MinRank instance with parameters $(q,\ m+\ell_1,\ m+\ell_2,\ K',\ r)$.

Parameter Selection. For a fixed target rank weight r, we first select m and k according to the decoding-radius relation $r = \lfloor (m-k)/2 \rfloor$, so that the resulting $\mathrm{MinRank}(q, m+\ell_1, m+\ell_2, K', r)$ instance remains secure against the currently best-known attacks relevant to the considered parameter regime.

As discussed above, when $m \nmid K'$, the public-code attacks considered in [5,43] do not apply directly to our construction. In the same regime, the corresponding dual-code strategies do not seem to offer an advantage over those analyzed in [5]. Accordingly, we focus on the case $m \nmid K'$.

Throughout all tables, the abbreviations "SL", "Ker.", "Minor.", and "Struc." denote the targeted classical security level (in bits), the kernel attack (Eq. (4)), the support minors attack [17], and the relevant structural distinguisher, respectively. Here, "Struc." stands for the combinatorial distinguisher [5] for the EGMC scheme, and for the structural distinguisher (Eq. (3)) for our scheme. Moreover, K_{pub} denotes the dimension of the public code (equal to K' for our scheme and to mk for the scheme in [5]).

We now fix q, k, and m, and set $r = \lfloor (m-k)/2 \rfloor$. We consider parameters $0 \le \ell_1, \ell_2 \le 10$ (not both zero), $1 \le K' < mk$, and $m \nmid K'$. Fixing the matrix multiplication exponent to $\omega = 2$, Tables 1 and 2 report, for $q = 2$ and $q = 16$, respectively, the smallest achievable public key size (pk) and ciphertext size (ct) subject to the following constraints:

1. the security level against the kernel attack is essentially preserved, within ± 0.5 bits;
2. the target security level is met against the kernel attack, the relevant structural distinguisher, and the support minors attack (and the EGMC reference parameters meet their claimed security levels against the attacks in [43]);
3. our selected parameters ensure that the complexities in Eqs. (1) and (2) exceed the targeted security levels by a substantial margin.

Let $N := (m+\ell_1)(m+\ell_2)$. The public key pk consists of a parity-check matrix $\bar{\mathbf{H}}' \in \mathbb{F}_q^{(N-K')\times N}$ stored in systematic form, hence requires $K'(N-K')\log_2 q$ bits; a ciphertext $\boldsymbol{c} \in \mathbb{F}_q^{N-K'}$ requires $(N-K')\log_2 q$ bits.

Table 1: Cost comparison at matched kernel attack complexity for $q = 2$.

SL	k	m	ℓ_1	ℓ_2	K_{pub}	r	Ker.	Scheme	Struc.	Minor.	pk	ct
128	27	37	4	0	999	5	145	[5]	148	147	65 kB	65 B
			3	0	981			Ours	146	151	61 kB	62 B
	35	43	4	0	1505	4	153	[5]	172	147	97 kB	65 B
			3	0	1495			Ours	157	156	90 kB	60 B
	47	53	3	0	2491	3	158	[5]	159	157	149 kB	60 B
			2	0	2472			Ours	144	163	137 kB	55 B
192	29	43	4	2	1247	7	209	[5]	232	220	135 kB	109 B
			3	1	1213			Ours	211	214	123 kB	101 B
	35	47	4	1	1645	6	219	[5]	224	217	165 kB	100 B
			4	0	1634			Ours	217	213	156 kB	95 B
	43	53	3	2	2279	5	227	[5]	247	236	228 kB	100 B
			3	0	2248			Ours	209	222	202 kB	90 B
	51	59	3	1	3009	4	219	[5]	229	221	267 kB	89 B
			3	0	2995			Ours	210	217	248 kB	83 B
256	35	51	5	1	1785	8	278	[5]	291	298	251 kB	141 B
			5	0	1773			Ours	285	292	240 kB	135 B
	39	53	5	1	2067	7	274	[5]	305	286	275 kB	133 B
			5	0	2057			Ours	294	281	261 kB	127 B
	71	79	3	1	5609	4	301	[5]	309	301	667 kB	119 B
			3	0	5594			Ours	273	297	618 kB	110 B

Performance Comparison with EGMC. We consider NIST-compliant parameter sets in Tables 1 and 2, while maintaining a security margin of 15 bits above the targeted security level. Under these comparable security constraints, our construction achieves a noticeable reduction in both public key size and ciphertext size when compared to the EGMC-Niederreiter encryption scheme.

For $q = 2$, as reported in Table 1, the public key size is reduced by 3.81%–11.40%, while the ciphertext size is reduced by 2.60%–10.00%. Table 2 shows a

Table 2: Cost comparison at matched kernel attack complexity for $q = 16$.

SL	k	m	ℓ_1	ℓ_2	K_{pub}	r	Ker.	Scheme	Struc.	Minor.	pk	ct
128	17	23	2	0	391	3	209	[5]	184	199	36 kB	92 B
			1	0	382			Ours	146	207	32 kB	85 B
	5	23	2	1	115	9	194	[5]	208	151	28 kB	243 B
			2	0	107			Ours	216	148	25 kB	234 B
192	19	25	2	1	475	3	234	[5]	280	214	54 kB	114 B
			2	0	467			Ours	251	214	49 kB	104 B
	15	23	2	1	345	4	241	[5]	248	223	44 kB	128 B
			2	0	337			Ours	234	221	40 kB	119 B
256	19	27	2	1	513	4	306	[5]	296	276	77 kB	150 B
			2	0	503			Ours	275	286	70 kB	140 B
	15	25	1	3	375	5	297	[5]	292	273	66 kB	177 B
			1	2	367			Ours	278	272	61 kB	168 B

similar trend for $q = 16$, with reductions ranging from 7.58%–11.11% in public key size and 3.70%–8.77% in ciphertext size.

Quantum Security Estimates. The "Ker." attack [27] is a combinatorial attack. In the quantum setting, applying the square-root (Grover-type) speedup to the dominant exponential search term results in halving the corresponding exponent of the underlying q-ary search space, while the polynomial factors are preserved as multiplicative overheads [38].

For a MinRank instance with parameters $(q,\, m+\ell_1,\, m+\ell_2,\, K',\, r)$ and fixing $\omega = 2$, Eq. (4) yields a classical complexity of $O\left(q^{r\left\lceil \frac{K'}{\max\{m+\ell_1,\, m+\ell_2\}} \right\rceil} \cdot K'^2\right)$. Thus, the quantum complexity of the Kernel attack is estimated as

$$O\left(q^{\frac{r}{2}\left\lceil \frac{K'}{\max\{m+\ell_1,\, m+\ell_2\}} \right\rceil} \cdot K'^2\right). \tag{5}$$

For representative parameter sets targeting the 128-bit, 192-bit, and 256-bit classical security levels, the estimated classical costs of the kernel attack are $O(2^{145})$, $O(2^{209})$, and $O(2^{274})$, respectively. The corresponding quantum cost estimates obtained from Eq. (5) are $O(2^{82})$, $O(2^{115})$, and $O(2^{148})$, respectively.

The "Minor." attack [17] is an algebraic attack, as it reduces the problem to solving a system of multivariate polynomial equations, typically via Gröbner basis techniques and, in certain parameter regimes, large linear systems [9]. As observed in [38], no generic quantum speedup with clear asymptotic advantage is currently known for this class of algebraic attacks.

The structural distinguisher in Eq. (3) comprises a combinatorial projection step followed by a rank search phase. In the absence of a dedicated quantum evaluation for this specific combination, we do not assign it a separate quantum complexity estimate, relying instead on classical security margins.

6.2 Comparative Analysis of Existing Schemes

Table 3 reports a comparison of public key and ciphertext sizes between the FO-based IND-CCA2 KEM variant of our EnGMS-Niederreiter construction in the random oracle model and a selection of representative code-based and lattice-based KEM schemes[1], including several submissions from the NIST post-quantum cryptography standardization process. Moreover, the FO-based KEM variant in our construction preserves the public-key and ciphertext sizes of the underlying EnGMS-Niederreiter scheme.

Across all three security levels, EnGMS-Niederreiter consistently achieves very compact ciphertexts, often among the smallest reported in the comparison, while keeping public key sizes within a practical range. At the 128-bit security level, for instance, a balanced parameter set yields a 60 B ciphertext, compared to 96 B for Classic McEliece and 65 B for the EGMC-Niederreiter scheme of [5]. A similar behavior is observed at the 192- and 256-bit security levels, where EnGMS-Niederreiter attains ciphertext sizes ranging from 83 B to 168 B.

Rather than optimizing solely for minimal public key size, the proposed construction explores a different region of the design space, in which ciphertext size is treated as a primary optimization objective. This choice is motivated by settings where ciphertext expansion constitutes a dominant cost, while moderately larger public keys can be accommodated. To reflect this flexibility, we provide multiple parameter sets that allow one to navigate the trade-off between public key size and ciphertext compactness, depending on application requirements.

Overall, the results in Table 3 illustrate that EnGMS-Niederreiter occupies a favorable point in the ciphertext/public-key trade-off among existing post-quantum proposals, particularly in regimes where compact ciphertexts are desirable. We emphasize that our goal is not to minimize public key size per se, but to demonstrate that, within code-based cryptography, substantially smaller ciphertexts can be achieved without relying on quasi-cyclic or idealized structures, and without sacrificing standard security assumptions. This in turn leaves the overall design philosophy closer in spirit to that of Classic McEliece than to schemes such as HQC, RQC, or ROLLO.

7 Conclusion and Future Work

We isolated a structural obstruction to hidden extension-field linearity and presented a subcode-dimension selection step within a masking procedure for matrix codes. Applied to matrix expansions of Gabidulin codes over $\mathbb{F}_{q^m}$, this framework leads to a new family of public-key encryption schemes, collectively denoted as EnGMS, whose security is based on the hardness of the MinRank problem. From a structural perspective, the condition $m \nmid K'$ removes the direct prerequisites of

[1] A detailed performance comparison of CPA-secure public-key encryption schemes, highlighting the flexible trade-offs of our EnGMS-Niederreiter construction (in particular its support for significantly larger plaintexts and tunable size parameters) is provided in Appendix G.

Table 3: Comparison of CCA2-secure KEM schemes.

Scheme	SL=128		SL=192		SL=256	
	pk	ct	pk	ct	pk	ct
EnGMS-Niederreiter	137 kB	55 B	248 kB	83 B	618 kB	110 B
EnGMS-Niederreiter	90 kB	60 B	202 kB	90 B	261 kB	127 B
EnGMS-Niederreiter	61 kB	62 B	156 kB	95 B	240 kB	135 B
EnGMS-Niederreiter	32 kB	85 B	49 kB	104 B	70 kB	140 B
EnGMS-Niederreiter	25 kB	234 B	40 kB	119 B	61 kB	168 B
Classic McEliece [11]	261 kB	96 B	524 kB	156 B	1044 kB	208 B
ROLLO-I [3]	696 B	696 B	958 B	958 B	1371 B	1371 B
KYBER [6]	800 B	768 B	1184 B	1088 B	1568 B	1568 B
BIKE [4]	1541 B	1573 B	3083 B	3115 B	5122 B	5154 B
RQC [38]	1834 B	3652 B	2853 B	5690 B	4090 B	8164 B
HQC [25]	2241 B	4433 B	4514 B	8978 B	7237 B	14421 B

the relevant structural attacks in [5, 43]; the corresponding variants considered in this work do not yield more effective attacks for the proposed parameter sets.

Under the MinRank and public-code indistinguishability assumptions, both EnGMS-McEliece and EnGMS-Niederreiter achieve OW-CPA security. In the random oracle model, their FO-based KEM and hybrid PKE variants are IND-CCA2 secure under the same assumptions. Beyond security, the proposed constructions occupy a favorable point in the efficiency trade-off. Concrete balanced parameter sets yield ciphertext sizes of 60 B, 83 B, and 110 B at the 128-bit, 192-bit, and 256-bit NIST security levels, respectively, resulting in significantly more compact ciphertexts than prior Gabidulin-based constructions, including the EGMC-Niederreiter scheme of [5].

Several directions for future work naturally arise. One is a finer-grained study of the subcode selection and masking steps, with the goal of identifying which design choices are most influential in resisting known structural attacks and their possible variants. Another is to further investigate the algebraic relationship between the public matrix code and its dual, as well as between the matrix representation of a Gabidulin code and its selected matrix subcode, in order to clarify whether such relations may lead either to new distinguishing techniques or to sharper security criteria for this class of constructions.

References

1. Aragon, N., Blazy, O., Deneuville, J.C., Gaborit, P., Hauteville, A., Ruatta, O., Tillich, J.P., Zémor, G.: LAKE (2017), Round 1 submission to the NIST Post-Quantum Cryptography Standardization Project
2. Aragon, N., Blazy, O., Deneuville, J.C., Gaborit, P., Hauteville, A., Ruatta, O., Tillich, J.P., Zémor, G.: LOCKER (2017), Round 1 submission to the NIST Post-Quantum Cryptography Standardization Project

3. Aragon, N., Blazy, O., Deneuville, J.C., Gaborit, P., Hauteville, A., Ruatta, O., Tillich, J.P., Zémor, G., Aguilar Melchor, C., Bettaieb, S., Bidoux, L., Bardet, M., Otmani, A.: ROLLO (merger of Rank-Ouroboros, LAKE, and LOCKER). `https://pqc-rollo.org` (Apr 2020), Round 2 submission to the NIST Post-Quantum Cryptography Standardization Project
4. Aragon, N., Barreto, P.S.L.M., Bettaieb, S., Bidoux, L., Blazy, O., Deneuville, J.C., Gaborit, P., Ghosh, S., Gueron, S., Güneysu, T., Aguilar Melchor, C., Misoczki, R., Persichetti, E., Richter-Brockmann, J., Sendrier, N., Tillich, J.P., Vasseur, V., Zémor, G.: BIKE. `https://bikesuite.org` (Oct 2024), Round 4 submission to the NIST Post-Quantum Cryptography Standardization Project
5. Aragon, N., Couvreur, A., Dyseryn, V., Gaborit, P., Vinçotte, A.: MinRank Gabidulin encryption scheme on matrix codes. In: Advances in Cryptology – ASIACRYPT 2024. pp. 68–100. Springer Nature Singapore (2025)
6. Avanzi, R., Bos, J., Ducas, L., Kiltz, E., Lepoint, T., Lyubashevsky, V., Schanck, J.M., Schwabe, P., Seiler, G., Stehlé, D.: CRYSTALS-KYBER. `https://pq-crystals.org` (Aug 2021), Round 3 submission to the NIST Post-Quantum Cryptography Standardization Project
7. Bardet, M., Briaud, P., Bros, M., Gaborit, P., Tillich, J.P.: Revisiting algebraic attacks on MinRank and on the rank decoding problem. Designs, Codes and Cryptography **91**(11), 3671–3707 (2023)
8. Bardet, M., Briaud, P., Bros, M., Gaborit, P., Neiger, V., Ruatta, O., Tillich, J.P.: An algebraic attack on rank metric code-based cryptosystems. In: Advances in Cryptology – EUROCRYPT 2020. pp. 64–93. Springer International Publishing (2020)
9. Bardet, M., Bros, M., Cabarcas, D., Gaborit, P., Perlner, R., Smith-Tone, D., Tillich, J.P., Verbel, J.: Improvements of algebraic attacks for solving the rank decoding and MinRank problems. In: Advances in Cryptology – ASIACRYPT 2020. pp. 507–536. Springer International Publishing (2020)
10. Berger, T.P., Loidreau, P.: How to mask the structure of codes for a cryptographic use. Designs, Codes and Cryptography **35**(1), 63–79 (2005)
11. Bernstein, D.J., Chou, T., Cid, C., Gilcher, J., Lange, T., Maram, V., von Maurich, I., Misoczki, R., Persichetti, E., Peters, C., Sendrier, N., Szefer, J., Tjhai, C.J., Tomlinson, M., Wang, W.: Classic McEliece. `https://classic.mceliece.org` (2025), Round 4 submission to the NIST Post-Quantum Cryptography Standardization Process
12. Buss, J.F., Frandsen, G.S., Shallit, J.O.: The computational complexity of some problems of linear algebra. Journal of Computer and System Sciences **58**(3), 572–596 (1999)
13. Couvreur, A., Gaborit, P., Gauthier-Umaña, V., Otmani, A., Tillich, J.P.: Distinguisher-based attacks on public-key cryptosystems using Reed–Solomon codes. Designs, Codes and Cryptography **73**(2), 641–666 (2014)
14. Couvreur, A., Debris-Alazard, T., Gaborit, P., Vinçotte, A.: MIRANDA: short signatures from a leakage-free full-domain-hash scheme. arXiv preprint arXiv:2510.07479 (2025)
15. Cramer, R., Shoup, V.: Design and analysis of practical public-key encryption schemes secure against adaptive chosen ciphertext attack. SIAM Journal on Computing **33**(1), 167–226 (2003)
16. Delsarte, P.: Bilinear forms over a finite field, with applications to coding theory. Journal of combinatorial theory, Series A **25**(3), 226–241 (1978)

17. Esser, A., Verbel, J., Zweydinger, F., Bellini, E.: SoK: CryptographicEstimators – a software library for cryptographic hardness estimation. In: Proceedings of the 19th ACM Asia Conference on Computer and Communications Security. pp. 560–574 (2024)
18. Faugère, J.C., El Din, M.S., Spaenlehauer, P.J.: Computing loci of rank defects of linear matrices using Gröbner bases and applications to cryptology. In: Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation. pp. 257–264 (2010)
19. Faure, C., Loidreau, P.: A new public-key cryptosystem based on the problem of reconstructing p–polynomials. In: Coding and Cryptography. pp. 304–315. Springer Berlin Heidelberg (2006)
20. Gabidulin, E.M.: Theory of codes with maximum rank distance. Problemy Peredachi Informatsii **21**(1), 3–16 (1985)
21. Gabidulin, E.M., Paramonov, A.V., Tretjakov, O.V.: Ideals over a non-commutative ring and their application in cryptology. In: Advances in Cryptology — EUROCRYPT 1991. pp. 482–489. Springer Berlin Heidelberg (1991)
22. Gabidulin, E., Ourivski, A.: Modified GPT PKC with right scrambler. Electronic Notes in Discrete Mathematics **6**, 168–177 (2001)
23. Gaborit, P., Murat, G., Ruatta, O., Zémor, G.: Low rank parity check codes and their application to cryptography. In: Proceedings of the Workshop on Coding and Cryptography WCC. vol. 2013 (2013)
24. Gaborit, P., Otmani, A., Kalachi, H.T.: Polynomial-time key recovery attack on the Faure–Loidreau scheme based on Gabidulin codes. Designs, Codes and Cryptography **86**(7), 1391–1403 (2018)
25. Gaborit, P., Aguilar-Melchor, C., Aragon, N., Bettaieb, S., Bidoux, L., Blazy, O., Deneuville, J.C., Persichetti, E., Zémor, G., Bos, J., Dion, A., Lacan, J., Robert, J.M., Véron, P., Barreto, P.L., Ghosh, S., Gueron, S., Güneysu, T., Misoczki, R., Richter-Brokmann, J., Sendrier, N., Tillich, J.P., Vasseur, V.: HQC. `https://pqc-hqc.org` (Aug 2025), Round 4 submission to the NIST Post-Quantum Cryptography Standardization Process
26. Gopalakrishnan, S., Neiger, V., Safey El Din, M.: Refined F5 algorithms for ideals of minors of square matrices. In: Proceedings of the 2023 International Symposium on Symbolic and Algebraic Computation. pp. 270–279 (2023)
27. Goubin, L., Courtois, N.T.: Cryptanalysis of the TTM cryptosystem. In: Advances in Cryptology — ASIACRYPT 2000. pp. 44–57. Springer Berlin Heidelberg (2000)
28. Guo, Q., Johansson, T., Löndahl, C.: A new algorithm for solving Ring-LPN with a reducible polynomial. IEEE Transactions on Information Theory **61**(11), 6204–6212 (2015)
29. Hoffstein, J., Pipher, J., Silverman, J.H.: NTRU: A ring-based public key cryptosystem. In: Algorithmic Number Theory. pp. 267–288. Springer Berlin Heidelberg (1998)
30. Hofheinz, D., Hövelmanns, K., Kiltz, E.: A modular analysis of the Fujisaki-Okamoto transformation. In: Theory of Cryptography. pp. 341–371. Springer International Publishing (2017)
31. Hua, L.K.: A theorem on matrices over a sfield and its applications. Acta Mathematica Sinica, Chinese Series **1**(2), 109–163 (1951)
32. Kipnis, A., Shamir, A.: Cryptanalysis of the HFE public key cryptosystem by relinearization. In: Wiener, M. (ed.) Advances in Cryptology — CRYPTO' 99. pp. 19–30. Springer Berlin Heidelberg, Berlin, Heidelberg (1999)
33. Lidl, R., Niederreiter, H.: Finite fields. No. 20, Cambridge University Press (1997)

34. Löndahl, C., Johansson, T., Shooshtari, M.K., Ahmadian-Attari, M., Aref, M.R.: Squaring attacks on McEliece public-key cryptosystems using quasi-cyclic codes of even dimension. Designs, Codes and Cryptography **80**(2), 359–377 (2016)
35. McEliece, R.J.: A public-key cryptosystem based on algebraic coding theory. The Deep Space Network Progress Report **42-44**, 114–116 (1978)
36. Melchor, C.A., Aragon, N., Bettaieb, S., Bidoux, L., Blazy, O., Deneuville, J.C., Gaborit, P., Hauteville, A., Zémor, G.: Ouroboros-R (2017), Round 1 submission to the NIST Post-Quantum Cryptography Standardization Project
37. Melchor, C.A., Aragon, N., Bettaieb, S., Bidoux, L., Blazy, O., Deneuville, J.C., Gaborit, P., Zémor, G.: Rank Quasi-Cyclic (RQC) (2017), Round 1 submission to the NIST Post-Quantum Cryptography Standardization Project
38. Melchor, C.A., Aragon, N., Bettaieb, S., Bidoux, L., Blazy, O., Deneuville, J.C., Gaborit, P., Zémor, G., Couvreur, A., Hauteville, A.: Rank Quasi-Cyclic (RQC). `https://pqc-rqc.org` (Apr 2020), Round 2 submission to the NIST Post-Quantum Cryptography Standardization Project
39. Misoczki, R., Tillich, J.P., Sendrier, N., Barreto, P.S.L.M.: MDPC-McEliece: New McEliece variants from moderate density parity-check codes. In: 2013 IEEE International Symposium on Information Theory. pp. 2069–2073 (2013)
40. Otmani, A., Kalachi, H.T., Ndjeya, S.: Improved cryptanalysis of rank metric schemes based on Gabidulin codes. Designs, Codes and Cryptography **86**(9), 1983–1996 (2018)
41. Ourivski, A., Gabidulin, E.: Column scrambler for the GPT cryptosystem. Discrete Applied Mathematics **128**(1), 207–221 (2003)
42. Overbeck, R.: A new structural attack for GPT and variants. In: Progress in Cryptology – Mycrypt 2005. pp. 50–63. Springer Berlin Heidelberg (2005)
43. Porwal, A., Wachter-Zeh, A., Loidreau, P.: Key attack on the ACDGV matrix encryption scheme. In: Advances in Cryptology – EUROCRYPT 2026. pp. 420–448. Springer Nature Switzerland (2026)
44. Prasad, A.: Counting subspaces of a finite vector space – 1. Resonance **15**(11), 977–987 (2010)
45. Ravagnani, A.: Rank-metric codes and their duality theory. Designs, Codes and Cryptography **80**(1), 197–216 (2016)
46. Reed, I.S., Solomon, G.: Polynomial codes over certain finite fields. Journal of the Society for Industrial and Applied Mathematics **8**(2), 300–304 (1960)
47. Sendrier, N.: Decoding one out of many. In: Post-Quantum Cryptography. pp. 51–67. Springer Berlin Heidelberg (2011)
48. Sidelnikov, V.M., Shestakov, S.O.: On insecurity of cryptosystems based on generalized Reed–Solomon codes. Discrete Mathematics and Applications **2**(4), 439–444 (1992)
49. Weger, V., Gassner, N., Rosenthal, J.: A survey on code-based cryptography. CoRR **abs/2201.07119** (2022), `https://arxiv.org/abs/2201.07119`
50. Zhang, F.: The Schur complement and its applications, vol. 4. Springer Science & Business Media (2006)

A Structural Distinguishers and Known Attacks on Gabidulin-Based Constructions

Structural Distinguishers. The vulnerability of Gabidulin-based constructions can be formalized through efficient polynomial-time distinguishers that separate them from random rank-metric codes.

Let m, n be positive integers, $1 < k < n-1$. For a vector $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{F}_{q^m}^n$, we denote by $\boldsymbol{x}^{[1]}$ the vector obtained by applying the Frobenius map $a \mapsto a^q$ to all its coordinates. At the vector-code level, let $\mathcal{C} \subseteq \mathbb{F}_{q^m}^n$ be an $[n, k]$ linear code and denote by $\mathcal{C}^{[1]}$ the code obtained by applying the Frobenius map $a \mapsto a^q$ coordinatewise, i.e., $\mathcal{C}^{[1]} := \{\boldsymbol{c}^{[1]} \mid \boldsymbol{c} \in \mathcal{C}\}$. As shown in [24], with high probability a random linear code satisfies

$$\dim(\mathcal{C} \cap \mathcal{C}^{[1]}) = \max\{2k - n, 0\},$$

whereas a Gabidulin $[n, k]$ code (with $1 < k < n-1$) satisfies $\dim(\mathcal{C} \cap \mathcal{C}^{[1]}) = k-1$. Since this invariant can be computed efficiently by linear algebra over $\mathbb{F}_{q^m}$, it yields a practical distinguisher between Gabidulin codes and random codes.

A related phenomenon appears after expanding Gabidulin codes into matrix codes over $\mathbb{F}_q$. Such matrix Gabidulin codes exhibit non-trivial stabilizer algebras that are typically absent in random matrix codes. In particular, when $n = m$, the right stabilizer algebra has dimension at least m [5]. These stabilizers can again be computed in polynomial time, providing efficient distinguishers at the matrix-code level.

Enhanced Gabidulin Matrix Codes and Remaining Attacks. To mitigate these structural weaknesses, Aragon *et al.* [5] introduced Enhanced Gabidulin Matrix Codes (EGMC), which apply a masking procedure designed to obscure extension-field artifacts in matrix Gabidulin codes. While this masking weakens direct Frobenius-based distinguishers, it still leaves enough hidden structure for further attacks against EGMC. More precisely, Aragon *et al.* [5] identified a combinatorial approach that searches, via projections and stabilizer tests, for traces of hidden $\mathbb{F}_{q^m}$-linearity, as well as an Overbeck-like approach based on the surviving Frobenius-related structure. Porwal *et al.* [43] subsequently turned this hidden-structure perspective into an exponential-time key-recovery attack against the same masked-expansion framework, leading to concrete security losses for several original EGMC parameter sets. Taken together, these attacks rely on hidden $\mathbb{F}_{q^m}$-linear expansion structure, including traces of $\mathbb{F}_{q^m}$-linearity, Frobenius-related invariants, and associated stabilizer structure.

B Game-Based Security Definitions for PKE and KEM

Definition 16 (PKE). *A public-key encryption scheme is defined by a triple of PPT algorithms:*

- $\mathsf{KeyGen}(1^\lambda)$*: Given a security parameter* λ*, outputs a public key* pk *and a secret key* sk*.*

- $\mathsf{Encrypt}(\mathsf{pk}, m)$*: Outputs a ciphertext* c *for message* m *under public key* pk*.*
- $\mathsf{Decrypt}(\mathsf{sk}, c)$*: Outputs the message* m *corresponding to ciphertext* c *under secret key* sk*, or an error symbol* $\perp$ *if decryption fails.*

We say that a PKE scheme is correct if, for every security parameter λ*, every key pair* $(\mathsf{pk}, \mathsf{sk})$ *output by* $\mathsf{KeyGen}(1^\lambda)$*, and every message* m*, it holds that*

$$\Pr\big[\mathsf{Decrypt}(\mathsf{sk}, \mathsf{Encrypt}(\mathsf{pk}, m)) = m\big] \geq 1 - \mathrm{negl}(\lambda),$$

where $\mathrm{negl}(\lambda)$ *denotes a negligible function in* λ*.*

Definition 17 (KEM). *A key encapsulation mechanism is specified by a triple of PPT algorithms:*

- $\mathsf{KeyGen}(1^\lambda)$*: On input the security parameter* λ*, generates a public key* pk *and a secret key* sk*.*
- $\mathsf{Encaps}(\mathsf{pk})$*: Outputs a pair* (K, c) *consisting of a shared secret key* K *and a ciphertext* c *under public key* pk*.*
- $\mathsf{Decaps}(\mathsf{sk}, c)$*: Given* sk *and a ciphertext* c*, returns the associated shared secret key* K*, or an error symbol* $\perp$ *if decapsulation rejects.*

A KEM is correct if, for every security parameter λ *and every key pair* $(\mathsf{pk}, \mathsf{sk})$ *generated by* $\mathsf{KeyGen}(1^\lambda)$*, we have*

$$\Pr\big[\mathsf{Decaps}(\mathsf{sk}, c) = K \;\big|\; (K, c) \leftarrow \mathsf{Encaps}(\mathsf{pk})\big] \geq 1 - \mathrm{negl}(\lambda),$$

where $\mathrm{negl}(\lambda)$ *denotes a negligible function in* λ*.*

Definition 18 (OW-CPA Security for PKE). *Let* $\mathcal{A}$ *be an adversary against a PKE scheme* $\mathsf{PKE} = (\mathsf{KeyGen}, \mathsf{Encrypt}, \mathsf{Decrypt})$*, and* λ *the security parameter. The OW-CPA game is defined as follows:*

- Challenge. *The challenger generates a key pair* $(\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{KeyGen}(1^\lambda)$*, samples a message* m *uniformly from the message space, and computes the ciphertext* $c \leftarrow \mathsf{Encrypt}(\mathsf{pk}, m)$*. The challenger then sends* (pk, c) *to the adversary* $\mathcal{A}$*.*
- Guess. *The adversary* $\mathcal{A}$ *outputs a guess* m' *for the message* m*. The adversary wins if* $m' = m$*.*

PKE *is said to be OW-CPA secure if, for any PPT adversary* $\mathcal{A}$*, the probability that* $\mathcal{A}$ *wins the game is negligible in* λ*.*

Definition 19 (IND-CCA2 Security for PKE). *Let* $\mathcal{A}$ *be an adversary against a PKE scheme* $\mathsf{PKE} = (\mathsf{KeyGen}, \mathsf{Encrypt}, \mathsf{Decrypt})$*, and let* λ *be the security parameter. The IND-CCA2 game is defined as follows:*

- Setup. *The challenger runs* $(\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{KeyGen}(1^\lambda)$ *and gives* pk *to the adversary* $\mathcal{A}$*.*
- Query Phase 1. *The adversary* $\mathcal{A}$ *is given adaptive access to a decryption oracle* $\mathcal{O}_{\mathsf{Dec}}(\cdot)$ *that on input any ciphertext* c *returns* $\mathsf{Decrypt}(\mathsf{sk}, c)$*.*

- Challenge. $\mathcal{A}$ *outputs two distinct messages* m_0, m_1 *from the same message space. The challenger samples a random bit* $b \xleftarrow{\$} \{0,1\}$*, computes the challenge ciphertext* $c^\star \leftarrow \mathsf{Encrypt}(\mathsf{pk}, m_b)$*, and sends* $c^\star$ *to* $\mathcal{A}$*.*
- Query Phase 2. *The adversary continues to have access to* $\mathcal{O}_{\mathsf{Dec}}(\cdot)$*, with the restriction that it may not query* $\mathcal{O}_{\mathsf{Dec}}(c^\star)$*.*
- Guess. $\mathcal{A}$ *outputs a guess* b'*. The adversary wins the game if* $b' = b$*.*

The advantage of $\mathcal{A}$ *in the above game is defined as*

$$\mathrm{Adv}^{\mathsf{IND\text{-}CCA2}}_{\mathsf{PKE}}(\mathcal{A}) \;:=\; \big|\Pr\big[b' = b\big] - \tfrac{1}{2}\big|.$$

We say that PKE *is IND-CCA2 secure if, for every PPT adversary* $\mathcal{A}$*, the advantage* $\mathrm{Adv}^{\mathsf{IND\text{-}CCA2}}_{\mathsf{PKE}}(\mathcal{A})$ *is negligible in* λ*.*

Definition 20 (IND-CCA2 Security for KEM). *Let* $\mathcal{A}$ *be an adversary against a KEM* $\mathsf{KEM} = (\mathsf{KeyGen}, \mathsf{Encaps}, \mathsf{Decaps})$*, and let* λ *be the security parameter. The IND-CCA2 game is specified as follows:*

- Setup. *The challenger samples* $(\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{KeyGen}(1^\lambda)$ *and provides* pk *to* $\mathcal{A}$*.*
- Query Phase 1. $\mathcal{A}$ *may adaptively query a decapsulation oracle* $\mathcal{O}_{\mathsf{Decap}}(\cdot)$*, which on input a ciphertext* c *returns* $\mathsf{Decaps}(\mathsf{sk}, c)$*.*
- Challenge. *The challenger computes* $(K_0^\star, c^\star) \leftarrow \mathsf{Encaps}(\mathsf{pk})$*, samples* $K_1^\star$ *uniformly at random from the same key space as* $K_0^\star$*, draws* $b \xleftarrow{\$} \{0,1\}$*, and returns* $(K_b^\star, c^\star)$ *to* $\mathcal{A}$*.*
- Query Phase 2. $\mathcal{A}$ *continues to access* $\mathcal{O}_{\mathsf{Decap}}(\cdot)$*, under the condition that it does not submit the challenge ciphertext* $c^\star$*.*
- Guess. $\mathcal{A}$ *outputs a bit* b'*. The adversary wins the game if* $b' = b$*.*

The advantage of $\mathcal{A}$ *is defined as*

$$\mathrm{Adv}^{\mathsf{IND\text{-}CCA2}}_{\mathsf{KEM}}(\mathcal{A}) \;:=\; \big|\Pr\big[b' = b\big] - \tfrac{1}{2}\big|.$$

We call KEM *IND-CCA2 secure if, for every PPT adversary* $\mathcal{A}$*,* $\mathrm{Adv}^{\mathsf{IND\text{-}CCA2}}_{\mathsf{KEM}}(\mathcal{A})$ *is negligible in* λ*.*

C Additional Details of EnGMS Distributions

C.1 Further Discussion of the EnGMS-IND Assumption

The EnGMS-IND assumption should not be read as a claim that the EnGMS distribution is statistically close to the uniform distribution over all K'-dimensional matrix subspaces. Before masking, the selected code remains a subcode of an expanded Gabidulin code. The assumption is that this origin is not efficiently detectable from the public code after random subcode selection and masking.

For our parameter sets, where $m \nmid K'$, the selected subcode cannot be the expansion of an $\mathbb{F}_{q^m}$-linear code. Random padding, together with secret row and column scramblers, further hides its embedding in the ambient matrix space.

A distinguisher would therefore have to expose an invariant that survives these transformations. The projection and completion variants, as well as the dual-code considerations analyzed in Section 5, are treated from this perspective, and their costs are reflected in the parameter analysis.

In particular, EnGMS-IND concerns the distributional hiding of the public code, including efficiently derived representations such as the dual code. Thus, dual-code distinguishers fall within the scope of the EnGMS-IND assumption. This is separate from the MinRank assumption, which captures the hardness of recovering the low-rank message or error from the associated MinRank instance.

C.2 On the Role of the EnGMS-IND Assumption

Indistinguishability assumptions between masked structured codes and uniformly random codes are a common component of code-based cryptography. In the case of EnGMS, the EnGMS-IND assumption arises naturally as a formal expression of the residual uncertainty that remains once the known algebraic structures and attack-enabling artifacts have been suppressed from the public code.

In our security analysis, the role of EnGMS-IND is best understood by comparing it with the MinRank assumption. The MinRank assumption concerns message recovery: given the public key and ciphertext, recover the low-rank message or error. EnGMS-IND instead concerns public-code recognition: decide whether the published code comes from the EnGMS distribution or from the corresponding random-code distribution. Hence, attacks exploiting recognizable structure in the public code are captured by EnGMS-IND, whereas generic low-rank recovery attacks are captured by MinRank.

This separation places EnGMS-IND within a standard paradigm in code-based cryptography: a structured code is masked, and the resulting public code is treated as random for the purposes of security analysis. Such public-code hiding assumptions are typically paired with a decoding-type hardness assumption; in our setting, this role is played by MinRank. Analogous assumptions appear in classical McEliece-type constructions and in many rank-metric schemes based on LRPC and related code families, although their precise formulations depend on the underlying code family.

C.3 Proof of Theorem 1

We decouple the argument into three separate steps.

(1) Removing the $\mathbf{P}, \mathbf{Q}$ *scramblers.* Let

$$\widehat{\mathcal{D}} := \mathrm{Span}_{\mathbb{F}_q} \left\{ \begin{pmatrix} \mathbf{A}'_j & \mathbf{R}_j \\ \mathbf{R}'_j & \mathbf{R}''_j \end{pmatrix} \middle| \; j = 1, \ldots, K' \right\}.$$

Let $\mathbf{P}\widehat{\mathcal{D}}\mathbf{Q} := \{\mathbf{PXQ} \mid \mathbf{X} \in \widehat{\mathcal{D}}\}$. By definition of the matrix trace inner product $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{tr}(\mathbf{X}\mathbf{Y}^\top)$, we have

$$(\mathbf{P}\widehat{\mathcal{D}}\mathbf{Q})^\perp = (\mathbf{P}^\top)^{-1}\, \widehat{\mathcal{D}}^\perp\, (\mathbf{Q}^\top)^{-1}. \tag{6}$$

Hence, it suffices to determine $\widehat{\mathcal{D}}^{\perp}$ and then apply (6).

(2) Structure of $\widehat{\mathcal{D}}^{\perp}$. Write

$$\mathbf{M}_j = \begin{pmatrix} \mathbf{A}'_j & \mathbf{R}_j \\ \mathbf{R}'_j & \mathbf{R}''_j \end{pmatrix}, \qquad \mathbf{Y} = \begin{pmatrix} \mathbf{B} & \mathbf{U} \\ \mathbf{V} & \mathbf{W} \end{pmatrix},$$

where the blocks have sizes $m \times n$, $m \times \ell_2$, $\ell_1 \times n$, and $\ell_1 \times \ell_2$, respectively. Using the trace inner product $\langle \mathbf{X}, \mathbf{Y} \rangle = \mathrm{tr}(\mathbf{X}\mathbf{Y}^{\top})$, for every j we have

$$\langle \mathbf{M}_j, \mathbf{Y} \rangle = \mathrm{tr}(\mathbf{A}'_j\mathbf{B}^{\top}) + \mathrm{tr}(\mathbf{R}_j\mathbf{U}^{\top}) + \mathrm{tr}(\mathbf{R}'_j\mathbf{V}^{\top}) + \mathrm{tr}(\mathbf{R}''_j\mathbf{W}^{\top}).$$

Consider the subspace

$$\mathcal{D}^0_{\mathrm{mat}} := \left\{ \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \;\middle|\; \mathbf{B} \in \mathcal{C}^{\perp}_{\mathrm{sub}} \right\} \subseteq \mathcal{V}.$$

Indeed, if $\mathbf{B} \in \mathcal{C}^{\perp}_{\mathrm{sub}}$ and let $\mathbf{U} = \mathbf{V} = \mathbf{W} = \mathbf{0}$, then $\langle \mathbf{M}_j, \mathbf{Y} \rangle = \mathrm{tr}(\mathbf{A}'_j\mathbf{B}^{\top}) = 0$ for all j, so $\mathcal{D}^0_{\mathrm{mat}} \subseteq \widehat{\mathcal{D}}^{\perp} \cap \mathcal{V}$. Conversely, if $\mathbf{Y} = \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \in \widehat{\mathcal{D}}^{\perp} \cap \mathcal{V}$, then we obtain $\mathrm{tr}(\mathbf{A}'_j\mathbf{B}^{\top}) = 0$ for all j, i.e., $\mathbf{B} \in \mathcal{C}^{\perp}_{\mathrm{sub}}$. So we have

$$\widehat{\mathcal{D}}^{\perp} \cap \mathcal{V} = \mathcal{D}^0_{\mathrm{mat}}.$$

Since $\widehat{\mathcal{D}}$ has dimension K' and the ambient space $\mathbb{F}_q^{(m+\ell_1)\times(n+\ell_2)}$ has dimension $(m+\ell_1)(n+\ell_2)$, we have $\dim_{\mathbb{F}_q} \widehat{\mathcal{D}}^{\perp} = (m+\ell_1)(n+\ell_2) - K'$. On the other hand, $\dim_{\mathbb{F}_q} \mathcal{C}_{\mathrm{sub}} = K'$, so $\dim_{\mathbb{F}_q} \mathcal{D}^0_{\mathrm{mat}} = \dim_{\mathbb{F}_q} \mathcal{C}^{\perp}_{\mathrm{sub}} = mn - K'$.

Choose a subspace $\mathcal{W} \subseteq \mathbb{F}_q^{(m+\ell_1)\times(n+\ell_2)}$ such that

$$\widehat{\mathcal{D}}^{\perp} = \mathcal{D}^0_{\mathrm{mat}} \oplus \mathcal{W}.$$

Then $\dim_{\mathbb{F}_q} \mathcal{W} = \dim_{\mathbb{F}_q} \widehat{\mathcal{D}}^{\perp} - \dim_{\mathbb{F}_q} \mathcal{D}^0_{\mathrm{mat}} = (m+\ell_1)(n+\ell_2) - mn$. Since $\widehat{\mathcal{D}}^{\perp} \cap \mathcal{V} = \mathcal{D}^0_{\mathrm{mat}}$, we have $\mathcal{W} \cap \mathcal{V} = \{0\}$. Thus we obtain $\mathbb{F}_q^{(m+\ell_1)\times(n+\ell_2)} = \mathcal{V} \oplus \mathcal{W}$, i.e., $\mathcal{W}$ indeed complements $\mathcal{V}$ in the ambient space.

(3) Decomposing $\mathcal{C}^{\perp}_{\mathrm{sub}}$ inside the top-left block. Because $\mathcal{C}_{\mathrm{sub}} \subseteq \Psi_\gamma(\mathcal{G})$, we have $\Psi_\gamma(\mathcal{G})^{\perp} \subseteq \mathcal{C}^{\perp}_{\mathrm{sub}}$. Choose a basis $\{\mathbf{B}_i\}_{i=1}^{m(n-k)}$ of $\Psi_\gamma(\mathcal{G})^{\perp}$ and extend it to a basis of $\mathcal{C}^{\perp}_{\mathrm{sub}}$ by adding $\{\widetilde{\mathbf{B}}_j\}_{j=1}^{mk-K'}$. Equivalently,

$$\mathcal{C}^{\perp}_{\mathrm{sub}} = \mathrm{Span}_{\mathbb{F}_q}\{\mathbf{B}_i\}_{i=1}^{m(n-k)} \oplus \mathrm{Span}_{\mathbb{F}_q}\{\widetilde{\mathbf{B}}_j\}_{j=1}^{mk-K'}. \tag{7}$$

Extending these basis elements by zeros in the lower and right blocks gives the two subspaces $\mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}_1)$ and $\mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}_2)$. Then $\mathcal{D}^0_{\mathrm{mat}} = \mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}_1) \oplus \mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}_2)$ by (7). Adjoining the complement $\mathcal{W}$ from Step (2) yields

$$\widehat{\mathcal{D}}^{\perp} = \big(\mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}_1) \oplus \mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}_2)\big) \oplus \mathcal{W}.$$

Then from (6), we obtain $\mathcal{EG}^{\perp}_{\mathrm{sub}} = (\mathbf{P}^{\top})^{-1}\mathcal{D}'_{\mathrm{mat}}(\mathbf{Q}^{\top})^{-1}$, as claimed.

D A McEliece-Type PKE from Matrix Subcodes

We show how the matrix subcode distributions introduced in Section 3 can be lifted to a full-fledged public-key encryption scheme in the McEliece paradigm. The resulting construction, which we call EnGMS-McEliece, follows the classical design principle of masking a structured code to obtain a public instance that is computationally indistinguishable from a random one, while retaining efficient decoding via the private structure.

Concretely, the public key consists of a masked matrix subcode sampled from the EnGMS distributions, whereas the secret key encodes the information required to recover the underlying Gabidulin code and to perform deterministic decoding. Encryption proceeds by adding a low-rank error sampled from an appropriate distribution, while decryption relies on inverting the scramblers, performing a truncation step, and applying the Gabidulin decoding algorithm.

The security of EnGMS-McEliece is established in two steps. First, we prove that the scheme achieves one-way security under chosen-plaintext attack (OW-CPA) assuming the hardness of the underlying MinRank problem and the hardness of distinguishing EnGMS public codes from random matrix codes of the same parameters. Second, we apply the FO transformation in the random oracle model to obtain an IND-CCA2-secure KEM, and then derive an IND-CCA2-secure PKE via the common KEM-DEM paradigm. Throughout this section, we emphasize that the role of masking is primarily to hide the representation of the matrix subcode, while the absence of extension-field linearity (ensured by the subcode construction itself) rules out some known structural attacks.

D.1 EnGMS-McEliece Encryption Scheme

We instantiate our construction by applying the subcode-based matrix code transformation (Definition 11) to the Gabidulin matrix code $\Psi_\gamma(\mathcal{G})$ induced by a Gabidulin code $\mathcal{G}$ of length n over $\mathbb{F}_{q^m}$. Here $\gamma \in \mathcal{B}(\mathbb{F}_{q^m})$ is a randomly selected $\mathbb{F}_q$-basis used for the expansion during key generation, and we focus on the case $n = m$. The task of decrypting a ciphertext without recovering the secret key amounts to solving a MinRank instance over $\mathbb{F}_q$. The resulting EnGMS-McEliece encryption scheme is shown in Fig. 2.

After applying the inverse scramblers and performing the final truncation, only the first m columns correspond to a received word over $\mathbb{F}_{q^m}$ with the hidden Gabidulin structure and can be recovered via Gabidulin decoding; the remaining ℓ_2 columns can then be computed by solving an $\mathbb{F}_q$-linear system.

Theorem 7 (Decryption Correctness). *In the EnGMS-McEliece scheme, assume the error matrix* $\mathbf{E}$ *used in encryption satisfies* $\mathrm{rank}(\mathbf{E}) \le r \le \left\lfloor \frac{m-k}{2} \right\rfloor$. *Then the decryption algorithm outputs the correct message* $\boldsymbol{\mu}$.

Proof. Let the ciphertext be $\mathbf{Y} = \sum_{j=1}^{K'} \mu_j \mathbf{M}'_j + \mathbf{E}$, where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{K'}) \in \mathbb{F}_q^{K'}$ is the original message, $\{\mathbf{M}'_j\}_{j=1}^{K'}$ is the public basis of the EnGMS code and $\mathrm{rank}(\mathbf{E}) \le r$.

KeyGen(1^λ):
– Sample $\boldsymbol{g} \in \mathbb{F}_{q^m}^m$ with $\mathbb{F}_q$-linearly independent entries, and let $\mathcal{G} = \mathcal{G}_{\boldsymbol{g}}(m, k, m)$ be the corresponding $[m, k]_{q^m}$ Gabidulin code with an efficient decoder correcting up to $t = \left\lfloor \frac{m-k}{2} \right\rfloor$ errors.
– Sample uniformly at random a basis $\gamma \xleftarrow{\$} \mathcal{B}(\mathbb{F}_{q^m})$.
– Compute an $\mathbb{F}_q$-basis $\{\mathbf{A}_1, \ldots, \mathbf{A}_{mk}\}$ of the code $\Psi_\gamma(\mathcal{G})$.
– Sample a full column-rank matrix $\mathbf{T} \xleftarrow{\$} \mathbb{F}_q^{mk \times K'}$ with $1 \le K' \le mk$.
– Compute $\mathbf{A}'_j = \sum_{i=1}^{mk} T_{i,j} \mathbf{A}_i$ for $j = 1, \ldots, K'$.
– For each $j = 1, \ldots, K'$ sample independently $\mathbf{R}_j \xleftarrow{\$} \mathbb{F}_q^{m \times \ell_2}$, $\mathbf{R}'_j \xleftarrow{\$} \mathbb{F}_q^{\ell_1 \times m}$ and $\mathbf{R}''_j \xleftarrow{\$} \mathbb{F}_q^{\ell_1 \times \ell_2}$.
– Sample $\mathbf{P} \xleftarrow{\$} \mathrm{GL}_{m+\ell_1}(\mathbb{F}_q)$ and $\mathbf{Q} \xleftarrow{\$} \mathrm{GL}_{m+\ell_2}(\mathbb{F}_q)$.
– For $j = 1, \ldots, K'$, define $\mathbf{M}'_j = \mathbf{P} \begin{pmatrix} \mathbf{A}'_j & \mathbf{R}_j \\ \mathbf{R}'_j & \mathbf{R}''_j \end{pmatrix} \mathbf{Q}$.
– Return $\mathsf{pk} = \{\mathbf{M}'_1, \ldots, \mathbf{M}'_{K'}\}$ and $\mathsf{sk} = (\boldsymbol{g}, \gamma, \mathbf{P}, \mathbf{Q})$.

Encrypt$(\mathsf{pk}, \boldsymbol{\mu})$:
Input: pk and a message $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{K'}) \in \mathbb{F}_q^{K'}$.
– Sample $\mathbf{U} \xleftarrow{\$} \mathbb{F}_q^{(m+\ell_1) \times r}$ and $\mathbf{V} \xleftarrow{\$} \mathbb{F}_q^{r \times (m+\ell_2)}$, and compute $\mathbf{E} = \mathbf{UV}$. This produces a random matrix $\mathbf{E}$ of rank at most r.
– Return the ciphertext $\mathbf{Y} = \sum_{j=1}^{K'} \mu_j \mathbf{M}'_j + \mathbf{E}$.

Decrypt$(\mathsf{sk}, \mathbf{Y})$:
Input: $\mathsf{sk} = (\boldsymbol{g}, \gamma, \mathbf{P}, \mathbf{Q})$ and a ciphertext $\mathbf{Y}$.
– Compute $\mathbf{P}^{-1}\mathbf{Y}\mathbf{Q}^{-1}$ and delete its last ℓ_1 rows to obtain $\mathbf{M} \in \mathbb{F}_q^{m \times (m+\ell_2)}$, written as $\mathbf{M} = (\mathbf{M}_0 \mid \mathbf{M}_1)$ with $\mathbf{M}_0 \in \mathbb{F}_q^{m \times m}$.
– Compute $\boldsymbol{y} = \Psi_\gamma^{-1}(\mathbf{M}_0) \in \mathbb{F}_{q^m}^m$.
– Apply the decoder of $\mathcal{G}$ to $\boldsymbol{y}$; the decoder outputs an error vector $\boldsymbol{e} \in \mathbb{F}_{q^m}^m$ of rank weight at most r (as $r \le t$).
– Compute $\mathbf{E}' = \Psi_\gamma(\boldsymbol{e}) \in \mathbb{F}_q^{m \times m}$.
– Solve the linear system $\mathbf{M} - (\mathbf{E}' \mid \mathbf{N}) = \sum_{j=1}^{K'} \mu_j (\mathbf{A}'_j \mid \mathbf{R}_j)$, where $\mathbf{N} \in \mathbb{F}_q^{m \times \ell_2}$ is the unknown part of the error. This yields $m(m + \ell_2)$ linear equations in the $K' + m\ell_2$ unknowns μ_j and the entries of $\mathbf{N}$.
– Return $\hat{\boldsymbol{\mu}}$, the solution of the above system.

Fig. 2: EnGMS-McEliece encryption scheme.

Step 1: Scrambling removal and truncation. Let

$$\Phi : \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_2)} \to \mathbb{F}_q^{m\times(m+\ell_2)}$$

be the linear truncation map that keeps the first m rows of a matrix. By construction of $\mathbf{M}'_j$ we have $\Phi(\mathbf{P}^{-1}\mathbf{M}'_j\mathbf{Q}^{-1}) = (\mathbf{A}'_j \mid \mathbf{R}_j)$ for all j, hence

$$\mathbf{M} = \Phi\big(\mathbf{P}^{-1}\mathbf{Y}\mathbf{Q}^{-1}\big) = \sum_{j=1}^{K'} \mu_j(\,\mathbf{A}'_j \mid \mathbf{R}_j\,) + \Phi\big(\mathbf{P}^{-1}\mathbf{E}\mathbf{Q}^{-1}\big) \in \mathbb{F}_q^{m\times(m+\ell_2)}.$$

We write $\mathbf{M} = (\mathbf{M}_0 \mid \mathbf{M}_1)$ with $\mathbf{M}_0 \in \mathbb{F}_q^{m\times m}$. Since multiplying by invertible matrices does not change the rank and Φ only discards rows, we have $\mathrm{rank}(\Phi(\mathbf{P}^{-1}\mathbf{E}\mathbf{Q}^{-1})) \le \mathrm{rank}(\mathbf{E}) \le r$.

Step 2: Gabidulin decoding recovers the error block $\mathbf{E}'$. Let $\mathbf{E}'$ denote the top-left $m \times m$ block of $\Phi(\mathbf{P}^{-1}\mathbf{E}\mathbf{Q}^{-1})$, so that $\mathbf{M}_0 = \sum_{j=1}^{K'} \mu_j \mathbf{A}'_j + \mathbf{E}'$ and $\mathrm{rank}(\mathbf{E}') \le \mathrm{rank}(\mathbf{E}) \le r$. Let $\boldsymbol{y} = \Psi_\gamma^{-1}(\mathbf{M}_0) \in \mathbb{F}_{q^m}^m$. By construction, each $\mathbf{A}'_j$ lies in $\Psi_\gamma(\mathcal{G})$, so there exist codewords $\boldsymbol{g}'_j \in \mathcal{G}$ with $\Psi_\gamma(\boldsymbol{g}'_j) = \mathbf{A}'_j$. Then $\boldsymbol{y} = \sum_{j=1}^{K'} \mu_j \boldsymbol{g}'_j + \boldsymbol{e}$, where $\boldsymbol{e} = \Psi_\gamma^{-1}(\mathbf{E}') \in \mathbb{F}_{q^m}^m$. Moreover, $\mathrm{wt}_R(\boldsymbol{e}) = \mathrm{rank}(\mathbf{E}') \le r$. By the choice of parameters, $r \le \lfloor (m-k)/2 \rfloor$, hence the decoding algorithm of $\mathcal{G}$ succeeds and returns the error vector $\boldsymbol{e} \in \mathbb{F}_{q^m}^m$. Then $\mathbf{E}' = \Psi_\gamma(\boldsymbol{e}) \in \mathbb{F}_q^{m\times m}$ is now known.

Step 3: Solving the linear system recovers $\boldsymbol{\mu}$. Therefore, $\mathbf{M} = \sum_{j=1}^{K'} \mu_j \left(\mathbf{A}'_j \mid \mathbf{R}_j\right) + \left(\mathbf{E}' \mid \mathbf{N}\right)$ for some unknown matrix $\mathbf{N} \in \mathbb{F}_q^{m\times\ell_2}$. The next step is to solve over $\mathbb{F}_q$ the linear system $\mathbf{M} - (\mathbf{E}' \mid \mathbf{N}) = \sum_{j=1}^{K'} \mu_j\,(\mathbf{A}'_j \mid \mathbf{R}_j)$, whose unknowns are the μ_j and the entries of $\mathbf{N} \in \mathbb{F}_q^{m\times\ell_2}$. This system consists of $m(m+\ell_2)$ equations in $K' + m\ell_2$ unknowns. By construction we have $K' \le mk$ and $k \le m$, hence $K' + m\ell_2 \le m(m+\ell_2)$, i.e., the system has at least as many equations as unknowns. We now argue existence and uniqueness of the solution.

Existence. Let $\boldsymbol{\mu}^{\mathrm{enc}} \in \mathbb{F}_q^{K'}$ be the plaintext message, and let $\mathbf{E}^{\mathrm{enc}}$ be the random matrix sampled in encryption. Then $\mathbf{Y} = \sum_{j=1}^{K'} \mu_j^{\mathrm{enc}}\,\mathbf{M}'_j + \mathbf{E}^{\mathrm{enc}}$. Using the truncation map Φ defined above, define $\mathbf{M} := \Phi(\mathbf{P}^{-1}\mathbf{Y}\mathbf{Q}^{-1})$ and $(\mathbf{E}' \mid \mathbf{N}^{\mathrm{enc}}) := \Phi(\mathbf{P}^{-1}\mathbf{E}^{\mathrm{enc}}\mathbf{Q}^{-1})$, where $\mathbf{E}' \in \mathbb{F}_q^{m\times m}$ and $\mathbf{N}^{\mathrm{enc}} \in \mathbb{F}_q^{m\times\ell_2}$ are the left and right blocks. Moreover, $\Phi(\mathbf{P}^{-1}\mathbf{M}'_j\mathbf{Q}^{-1}) = (\mathbf{A}'_j \mid \mathbf{R}_j)$ for all j. Hence,

$$\mathbf{M} - (\mathbf{E}' \mid \mathbf{N}^{\mathrm{enc}}) = \sum_{j=1}^{K'} \mu_j^{\mathrm{enc}}\,(\mathbf{A}'_j \mid \mathbf{R}_j),$$

so $(\boldsymbol{\mu}^{\mathrm{enc}}, \mathbf{N}^{\mathrm{enc}})$ is a solution to the linear system.

Uniqueness. Suppose that $(\boldsymbol{\mu}, \mathbf{N})$ and $(\boldsymbol{\mu}^{\mathrm{enc}}, \mathbf{N}^{\mathrm{enc}})$ are two solutions over $\mathbb{F}_q$ of the above linear system, i.e., both satisfy

$$\mathbf{M} - (\mathbf{E}' \mid \mathbf{N}) = \sum_{j=1}^{K'} \mu_j\,(\mathbf{A}'_j \mid \mathbf{R}_j) \tag{8}$$

and

$$\mathbf{M} - (\mathbf{E}' \mid \mathbf{N}^{\mathrm{enc}}) = \sum_{j=1}^{K'} \mu_j^{\mathrm{enc}} \, (\mathbf{A}'_j \mid \mathbf{R}_j). \tag{9}$$

Subtracting (9) from (8) yields

$$(\mathbf{0} \mid \mathbf{N}^{\mathrm{enc}} - \mathbf{N}) = \sum_{j=1}^{K'} (\mu_j - \mu_j^{\mathrm{enc}}) \, (\mathbf{A}'_j \mid \mathbf{R}_j). \tag{10}$$

Taking the left $m \times m$ block (i.e., the first m columns) in (10) gives $\mathbf{0} = \sum_{j=1}^{K'} (\mu_j - \mu_j^{\mathrm{enc}}) \, \mathbf{A}'_j$. Since $\{\mathbf{A}'_j\}_{j=1}^{K'}$ is $\mathbb{F}_q$-linearly independent by Lemma 1, we obtain $\mu_j = \mu_j^{\mathrm{enc}}$ for all j, hence $\boldsymbol{\mu} = \boldsymbol{\mu}^{\mathrm{enc}}$. Substituting $\boldsymbol{\mu} = \boldsymbol{\mu}^{\mathrm{enc}}$ into (10) implies $\mathbf{N} = \mathbf{N}^{\mathrm{enc}}$. Therefore the solution is unique, and solving the above linear system over $\mathbb{F}_q$ recovers the original message $\boldsymbol{\mu}$.

D.2 Security Proofs

Theorem 8 (OW-CPA Security). *If no PPT algorithm can solve the MinRank problem with non-negligible probability, and every PPT distinguisher for the EnGMS-IND problem has only negligible advantage, then the* EnGMS-McEliece *scheme is OW-CPA secure.*

Proof. Assume that there exists an efficient PPT algorithm $\mathcal{A}_{\mathrm{ME}}$ for the EnGMS-McEliece scheme, which takes as input a public key and a ciphertext and outputs the underlying message with non-negligible probability $\varepsilon(\lambda)$. We denote this success probability by $\mathrm{Succ}(\mathcal{A}_{\mathrm{ME}})$. Thus, $\mathrm{Succ}(\mathcal{A}_{\mathrm{ME}}) \geq \varepsilon(\lambda)$. Using $\mathcal{A}_{\mathrm{ME}}$, we construct the following distinguisher $\mathcal{D}$ for the EnGMS-IND problem.

The challenger samples a K'-dimensional matrix code $\mathcal{C}'_{\mathrm{mat}} \subseteq \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_2)}$ either from the EnGMS distribution (Definition 13) or uniformly at random as in Definition 14, and provides $\mathcal{D}$ with a basis $\mathcal{RB}' = \{\mathbf{M}'_1, \ldots, \mathbf{M}'_{K'}\}$ of $\mathcal{C}'_{\mathrm{mat}}$.

Distinguisher $\mathcal{D}(\mathcal{RB}')$:

- Choose $\boldsymbol{\mu} \xleftarrow{\$} \mathbb{F}_q^{K'}$.
- Choose $\mathbf{U} \xleftarrow{\$} \mathbb{F}_q^{(m+\ell_1)\times r}$, $\mathbf{V} \xleftarrow{\$} \mathbb{F}_q^{r\times(m+\ell_2)}$, and compute $\mathbf{E} = \mathbf{UV}$ with $\mathrm{rank}(\mathbf{E}) \leq r$.
- Compute $\mathbf{Y} \leftarrow \sum_{j=1}^{K'} \mu_j \mathbf{M}'_j + \mathbf{E}$.
- Compute $\hat{\boldsymbol{\mu}} \leftarrow \mathcal{A}_{\mathrm{ME}}(\mathcal{RB}', \mathbf{Y})$.
- Output 1 if $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}$, and output 0 otherwise.

Define an algorithm $\mathcal{A}_{\mathrm{MR}}$ for $\mathrm{MinRank}(q, m+\ell_1, m+\ell_2, K', r)$ as follows:

- Input $(\mathcal{RB}', \mathbf{Y})$.
- Compute $\hat{\boldsymbol{\mu}} \leftarrow \mathcal{A}_{\mathrm{ME}}(\mathcal{RB}', \mathbf{Y})$.
- Output $\hat{\mathbf{E}} = \mathbf{Y} - \sum_{j=1}^{K'} \hat{\mu}_j \mathbf{M}'_j$.

Then $\mathrm{Succ}(\mathcal{A}_{\mathrm{MR}}) = \Pr\left[\mathrm{rank}(\hat{\mathbf{E}}) \leq r\right]$. When $\mathcal{RB}'$ is a basis of a uniformly random K'-dimensional matrix code, producing $\hat{\mathbf{E}} = \mathbf{Y} - \sum_{j=1}^{K'} \hat{\mu}_j \mathbf{M}'_j$ of rank at most r amounts to solving an instance of the MinRank problem.

We denote by Unif the event that $\mathcal{RB}'$ is the basis of a uniformly random K'-dimensional matrix code, and by Sub the event that $\mathcal{RB}'$ is the basis of an EnGMS matrix code. By definition of the advantage of $\mathcal{D}$ in the EnGMS-IND problem, we have

$$\begin{aligned}\mathrm{Adv}(\mathcal{D}) &= \left|\Pr\left[\mathcal{D}(\mathcal{RB}') = 1 \mid \mathsf{Unif}\right] - \Pr\left[\mathcal{D}(\mathcal{RB}') = 1 \mid \mathsf{Sub}\right]\right| \\ &= \left|\mathrm{Succ}_{\mathsf{Unif}}(\mathcal{A}_{\mathrm{ME}}) - \mathrm{Succ}_{\mathsf{Sub}}(\mathcal{A}_{\mathrm{ME}})\right|.\end{aligned}$$

Moreover, $\mathrm{Succ}_{\mathsf{Unif}}(\mathcal{A}_{\mathrm{ME}}) \leq \mathrm{Succ}(\mathcal{A}_{\mathrm{MR}})$, since $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}$ implies $\hat{\mathbf{E}} = \mathbf{E}$. Therefore,

$$\mathrm{Succ}_{\mathsf{Sub}}(\mathcal{A}_{\mathrm{ME}}) \leq \mathrm{Succ}(\mathcal{A}_{\mathrm{MR}}) + \mathrm{Adv}(\mathcal{D}).$$

Since $\mathrm{Succ}_{\mathsf{Sub}}(\mathcal{A}_{\mathrm{ME}}) = \mathrm{Succ}(\mathcal{A}_{\mathrm{ME}}) \geq \varepsilon(\lambda)$, it follows that either $\mathrm{Succ}(\mathcal{A}_{\mathrm{MR}}) \geq \frac{1}{2}\varepsilon(\lambda)$ or $\mathrm{Adv}(\mathcal{D}) \geq \frac{1}{2}\varepsilon(\lambda)$, both of which are non-negligible. Hence there exists either an efficient PPT algorithm to solve the MinRank problem, or an efficient PPT distinguisher for the EnGMS-IND problem, contradicting our assumptions. This proves that the $\mathsf{EnGMS\text{-}McEliece}$ scheme is OW-CPA secure.

The OW-CPA security of the $\mathsf{EnGMS\text{-}McEliece}$ encryption scheme relies on the hardness of the MinRank problem and on the EnGMS-IND assumption. To obtain an IND-CCA2 secure variant in the random oracle model, we apply the FO transformation [30] to our $\mathsf{EnGMS\text{-}McEliece}$ PKE scheme, yielding an FO-based KEM. The resulting construction $\mathsf{KEM}^{\mathrm{FO}}_{\mathrm{ME}}$ is given in Fig. 3 and is IND-CCA2 secure under the two assumptions of Theorem 9.

We derive a PKE scheme $\mathsf{PKE}^{\mathrm{FO}}_{\mathrm{ME}}$ from $\mathsf{KEM}^{\mathrm{FO}}_{\mathrm{ME}}$ via the generic KEM-DEM (hybrid-encryption) paradigm: the KEM encapsulates a symmetric session key, and a DEM encrypts the payload under this key. In our setting, the DEM is instantiated by a one-time CCA-secure symmetric encryption scheme $\mathsf{SKE} = (\mathsf{Enc}_{\mathrm{SKE}}, \mathsf{Dec}_{\mathrm{SKE}})$, and the resulting $\mathsf{PKE}^{\mathrm{FO}}_{\mathrm{ME}}$ scheme is given in Fig. 4. By the KEM-DEM composition theorem (see [15, Theorem 7.2]), assuming such a one-time CCA-secure DEM, $\mathsf{PKE}^{\mathrm{FO}}_{\mathrm{ME}}$ is IND-CCA2 secure.

Theorem 9 (IND-CCA2 Security). *In the random oracle model, assume that*

1. *no PPT algorithm can solve the MinRank$(q, m+\ell_1, m+\ell_2, K', r)$ problem with non-negligible probability, and*
2. *no PPT distinguisher for the EnGMS-IND problem has a non-negligible advantage.*

Let $\mathsf{PKE}_{\mathrm{ME}}$ *be the* $\mathsf{EnGMS\text{-}McEliece}$ *PKE scheme, and let* G, H *be hash functions modeled as random oracles with output lengths as in the FO construction [30]. Let* $\mathsf{KEM}^{\mathrm{FO}}_{\mathrm{ME}}$ *be the KEM in Fig. 3 obtained from* $\mathsf{PKE}_{\mathrm{ME}}$ *via the transformation* $\mathrm{FO}^{\not\perp}_m[\mathsf{PKE}_{\mathrm{ME}}, \mathsf{G}, \mathsf{H}]$*, and let* $\mathsf{PKE}^{\mathrm{FO}}_{\mathrm{ME}}$ *be the PKE scheme in Fig. 4 obtained from* $\mathsf{KEM}^{\mathrm{FO}}_{\mathrm{ME}}$ *via the KEM-DEM hybrid paradigm recalled above. Then both* $\mathsf{KEM}^{\mathrm{FO}}_{\mathrm{ME}}$ *and* $\mathsf{PKE}^{\mathrm{FO}}_{\mathrm{ME}}$ *are IND-CCA2 secure.*

Table 4: Notation for $\mathsf{KEM}^{\mathrm{FO}}_{\mathrm{ME}}$.

Symbol	Meaning
$\mathcal{M}$	Message space, $\mathcal{M} = \mathbb{F}_q^{K'}$.
$\mathcal{E}_r$	Error set $\{\mathbf{E} \in \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_2)} \mid \mathrm{rank}(\mathbf{E}) \le r\}$.
κ	Session-key length (bits).
$\mathrm{enc}(\cdot)$	Fixed injective encoding into $\{0,1\}^*$ (for vectors or matrices).
$\boldsymbol{s}$	Secret value sampled in $\mathsf{KeyGen}^{\mathrm{FO}}$, $\boldsymbol{s} \xleftarrow{\$} \mathcal{M}$.
G	Hash function modeled as a random oracle, $\mathsf{G} : \mathcal{M} \to \{0,1\}^*$ (with sufficiently long output).
$\mathcal{R}$	Randomness space of the underlying encryption algorithm $\mathsf{Enc}_{\mathrm{ME}}$: $\mathcal{R} = \mathbb{F}_q^{(m+\ell_1)\times r} \times \mathbb{F}_q^{r\times(m+\ell_2)}$.
Parse_E	Deterministic parser mapping $\{0,1\}^*$ to $\mathcal{R}$: on input ω it outputs $(\mathbf{U},\mathbf{V}) \in \mathcal{R}$ and sets $\mathbf{E} := \mathbf{UV} \in \mathcal{E}_r$.
H	Hash function modeled as a random oracle, $\mathsf{H} : \{0,1\}^* \to \{0,1\}^\kappa$.
K	Session key (with encapsulated message $\boldsymbol{\mu} \in \mathcal{M}$, ciphertext $\mathbf{Y}$, and secret $\boldsymbol{s}$ from $\mathsf{KeyGen}^{\mathrm{FO}}$): $\mathsf{K} := \mathsf{H}(\mathrm{enc}(\boldsymbol{\mu}))$ if accepted; otherwise $\mathsf{K} := \mathsf{H}(\mathrm{enc}(\mathbf{Y})\|\mathrm{enc}(\boldsymbol{s}))$ (implicit rejection).

Notation: See Table 4 and Fig. 2.

$\mathbf{KeyGen}^{\mathrm{FO}}(1^\lambda)$:
– Run $(\mathsf{pk}_{\mathrm{ME}}, \mathsf{sk}_{\mathrm{ME}}) \leftarrow \mathsf{KeyGen}(1^\lambda)$ of $\mathsf{PKE}_{\mathrm{ME}}$, where $\mathsf{pk}_{\mathrm{ME}} = \{\mathbf{M}'_1, \ldots, \mathbf{M}'_{K'}\}$ and $\mathsf{sk}_{\mathrm{ME}} = (\boldsymbol{g}, \gamma, \mathbf{P}, \mathbf{Q})$.
– Sample $\boldsymbol{s} \xleftarrow{\$} \mathcal{M}$.
– Return $\mathsf{pk} = \mathsf{pk}_{\mathrm{ME}}$ and $\mathsf{sk} = (\mathsf{sk}_{\mathrm{ME}}, \boldsymbol{s})$.

$\mathbf{Encaps}^{\mathrm{FO}}(\mathsf{pk})$:
Input: $\mathsf{pk} = \{\mathbf{M}'_1, \ldots, \mathbf{M}'_{K'}\}$.
– Sample $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{K'}) \xleftarrow{\$} \mathcal{M}$.
– Compute $(\mathbf{U},\mathbf{V}) \leftarrow \mathsf{Parse}_E(\mathsf{G}(\boldsymbol{\mu}))$ and set $\mathbf{E} := \mathbf{UV}$.
– Compute $\mathbf{Y} := \sum_{j=1}^{K'} \mu_j \mathbf{M}'_j + \mathbf{E}$.
– Set $\mathsf{K} := \mathsf{H}(\mathrm{enc}(\boldsymbol{\mu}))$.
– Return $(\mathsf{K}, \mathbf{Y})$.

$\mathbf{Decaps}^{\mathrm{FO}}(\mathsf{sk}, \mathbf{Y})$:
Input: $\mathsf{sk} = (\mathsf{sk}_{\mathrm{ME}}, \boldsymbol{s})$ and a ciphertext $\mathbf{Y}$.
– Compute $\hat{\boldsymbol{\mu}} \leftarrow \mathsf{Dec}(\mathsf{sk}_{\mathrm{ME}}, \mathbf{Y})$ of $\mathsf{PKE}_{\mathrm{ME}}$.
– If $\hat{\boldsymbol{\mu}} = \perp$, set $\mathsf{K} := \mathsf{H}(\mathrm{enc}(\mathbf{Y})\|\mathrm{enc}(\boldsymbol{s}))$ and return K.
– Parse $(\hat{\mathbf{U}}, \hat{\mathbf{V}}) \leftarrow \mathsf{Parse}_E(\mathsf{G}(\hat{\boldsymbol{\mu}}))$, set $\hat{\mathbf{E}} := \hat{\mathbf{U}}\hat{\mathbf{V}}$, and re-encrypt $\hat{\mathbf{Y}} := \sum_{j=1}^{K'} \hat{\mu}_j \mathbf{M}'_j + \hat{\mathbf{E}}$.
– If $\hat{\mathbf{Y}} \neq \mathbf{Y}$, set $\mathsf{K} := \mathsf{H}(\mathrm{enc}(\mathbf{Y})\|\mathrm{enc}(\boldsymbol{s}))$; else set $\mathsf{K} := \mathsf{H}(\mathrm{enc}(\hat{\boldsymbol{\mu}}))$.
– Return K.

Fig. 3: $\mathsf{KEM}^{\mathrm{FO}}_{\mathrm{ME}}$ derived from $\mathsf{PKE}_{\mathrm{ME}}$ (Fig. 2) via $\mathrm{FO}^{\not\perp}_m$.

Table 5: Notation for $\mathsf{PKE}^{\mathrm{FO}}_{\mathrm{ME}}$.

Symbol	Meaning
K	Session key output by $\mathsf{Encaps}^{\mathrm{FO}}$ or $\mathsf{Decaps}^{\mathrm{FO}}$ (implicit rejection as in Table 4).
κ	Session-key length (bits).
$\mathbf{Y}$	KEM ciphertext output by $\mathsf{Encaps}^{\mathrm{FO}}$.
SKE	One-time chosen-ciphertext secure symmetric encryption scheme $\mathsf{SKE} = (\mathsf{Enc}_{\mathrm{SKE}}, \mathsf{Dec}_{\mathrm{SKE}})$ with key space $\{0,1\}^{\kappa}$ and message space $\mathcal{M}_{\mathrm{SKE}}$.
m	Plaintext message, $m \in \mathcal{M}_{\mathrm{SKE}}$.
χ	SKE ciphertext of message m under session key K.
$\mathrm{enc}(\cdot,\cdot)$	Canonical, uniquely decodable encoding of a pair into $\{0,1\}^*$ (used for $(\mathbf{Y},\chi)$).
$\hat{c}$	Hybrid ciphertext (canonical encoding of $(\mathbf{Y},\chi)$).
Parse_c	Deterministic parser for hybrid ciphertexts: on input $\hat{c}$ it outputs the unique $(\mathbf{Y},\chi)$ such that $\hat{c} = \mathrm{enc}(\mathbf{Y},\chi)$, or $\perp$.

Notation: See Table 5 and Fig. 3.

KeyGen(1^λ):
– Run $(\mathsf{pk},\mathsf{sk}) \leftarrow \mathsf{KeyGen}^{\mathrm{FO}}(1^\lambda)$ of $\mathsf{KEM}^{\mathrm{FO}}_{\mathrm{ME}}$.
– Return $\mathsf{pk}' = \mathsf{pk}$ and $\mathsf{sk}' = \mathsf{sk}$.

Encrypt(pk', m):
Input: pk' and a message $m \in \mathcal{M}_{\mathrm{SKE}}$.
– Run $(\mathsf{K},\mathbf{Y}) \leftarrow \mathsf{Encaps}^{\mathrm{FO}}(\mathsf{pk}')$ of $\mathsf{KEM}^{\mathrm{FO}}_{\mathrm{ME}}$.
– Compute $\chi \leftarrow \mathsf{Enc}_{\mathrm{SKE}}(\mathsf{K}, m)$.
– Compute the hybrid ciphertext $\hat{c} := \mathrm{enc}(\mathbf{Y},\chi)$, i.e., a canonical bit-string encoding of the pair $(\mathbf{Y},\chi)$.
– Return $\hat{c}$.

Decrypt$(\mathsf{sk}', \hat{c})$:
Input: sk' and a ciphertext $\hat{c}$.
– $(\mathbf{Y},\chi) \leftarrow \mathsf{Parse}_c(\hat{c})$; if $(\mathbf{Y},\chi) = \perp$, return $\perp$.
– Compute $\mathsf{K} \leftarrow \mathsf{Decaps}^{\mathrm{FO}}(\mathsf{sk}', \mathbf{Y})$ of $\mathsf{KEM}^{\mathrm{FO}}_{\mathrm{ME}}$.
– Compute $\hat{m} \leftarrow \mathsf{Dec}_{\mathrm{SKE}}(\mathsf{K},\chi)$.
– Return $\hat{m}$ (possibly $\perp$).

Fig. 4: $\mathsf{PKE}^{\mathrm{FO}}_{\mathrm{ME}}$ derived from $\mathsf{KEM}^{\mathrm{FO}}_{\mathrm{ME}}$ (Fig. 3) via KEM-DEM paradigm.

Proof. We proceed in three steps.

Step 1: OW-CPA security of the underlying PKE. By Theorem 8, the EnGMS-McEliece scheme $\mathsf{PKE}_{\mathrm{ME}}$ is OW-CPA secure under the stated MinRank hardness and EnGMS-IND assumptions. Since $\mathsf{PKE}_{\mathrm{ME}}$ does not make use of any hash oracle, the OW-CPA security statement also holds in the random oracle model.

Step 2: Applying the $\mathrm{FO}_m^{\not\perp}$ *transformation.* Theorem 7 shows that the EnGMS-McEliece scheme $\mathsf{PKE}_{\mathrm{ME}}$ is perfectly correct, i.e. it is δ_1-correct with $\delta_1 = 0$ in the sense of [30]. Together with Step 1, this means that $\mathsf{PKE}_{\mathrm{ME}}$ satisfies the hypotheses of Theorem 6 in [30].

By [30, Theorem 6], for the transformation $\mathrm{FO}_m^{\not\perp}$, if a PKE scheme PKE is OW-CPA secure and δ_1-correct, then the KEM obtained by applying $\mathrm{FO}_m^{\not\perp}$ to PKE is IND-CCA2 secure in the random oracle model. Applying this with $\mathsf{PKE} = \mathsf{PKE}_{\mathrm{ME}}$ yields that $\mathsf{KEM}_{\mathrm{ME}}^{\mathrm{FO}}$ is IND-CCA2 secure.

Step 3: From IND-CCA2 KEM to IND-CCA2 PKE. It is well known that, in the KEM–DEM paradigm, any IND-CCA2 secure KEM combined with a one-time chosen-ciphertext secure symmetric encryption scheme yields an IND-CCA2 secure PKE scheme (see [15, Theorem 7.2]). Since Step 2 shows that $\mathsf{KEM}_{\mathrm{ME}}^{\mathrm{FO}}$ is IND-CCA2 secure in the random oracle model, the associated hybrid scheme $\mathsf{PKE}_{\mathrm{ME}}^{\mathrm{FO}}$ is therefore also IND-CCA2 secure.

Combining Steps 1–3, we conclude that both $\mathsf{KEM}_{\mathrm{ME}}^{\mathrm{FO}}$ and $\mathsf{PKE}_{\mathrm{ME}}^{\mathrm{FO}}$ are IND-CCA2 secure in the random oracle model.

D.3 Algorithmic Complexity

We next estimate the algorithmic complexity of the McEliece variant. Writing $N := (m+\ell_1)(m+\ell_2)$, we measure the cost of the key generation, encryption, and decryption algorithms by counting arithmetic operations in $\mathbb{F}_q$.

- *Key generation cost.* The dominant costs are: (i) forming the K' matrices $\mathbf{A}'_j = \sum_{i=1}^{mk} T_{i,j}\mathbf{A}_i$ with $\mathbf{A}_i \in \mathbb{F}_q^{m\times m}$, and (ii) computing the K' masked matrices $\mathbf{M}'_j$, where $\mathbf{M}'_j \in \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_2)}$. The overall cost (using classical matrix multiplication) is $O\Big(K'\big(km^3 + (2m+\ell_1+\ell_2)N\big)\Big)$ operations in $\mathbb{F}_q$.
- *Encryption cost.* The encryption algorithm outputs $\mathbf{Y} = \sum_{j=1}^{K'} \mu_j \mathbf{M}'_j + \mathbf{E}$, where $\mathbf{M}'_j, \mathbf{E}, \mathbf{Y} \in \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_2)}$ and $\mathbf{E} = \mathbf{UV}$ with $\mathbf{U} \in \mathbb{F}_q^{(m+\ell_1)\times r}$ and $\mathbf{V} \in \mathbb{F}_q^{r\times(m+\ell_2)}$. Computing the linear combination $\sum_{j=1}^{K'} \mu_j \mathbf{M}'_j$ costs $O(K'N)$ operations in $\mathbb{F}_q$. Using classical matrix multiplication, computing $\mathbf{E} = \mathbf{UV}$ costs $O(rN)$ operations in $\mathbb{F}_q$. Hence, the arithmetic cost of forming $\mathbf{Y}$ is $O((K'+r)N)$ operations in $\mathbb{F}_q$.
- *Decryption cost.* The decryption algorithm requires solving one linear system over $\mathbb{F}_q$ with $K' + m\ell_2$ unknowns, costing $O\big((K'+m\ell_2)^3\big)$ operations in $\mathbb{F}_q$.

E Security Proof and FO-Based Constructions for EnGMS-Niederreiter

E.1 Key Generation Checks: Invertibility Conditions and Resampling Cost

This subsection discusses the invertibility checks during key generation and the associated probabilities to ensure zero decryption failures.

Since $\mathbf{P}$ (resp., $\mathbf{Q}$) is sampled uniformly from $\mathrm{GL}_{m+\ell_1}(\mathbb{F}_q)$ (resp., $\mathrm{GL}_{m+\ell_2}(\mathbb{F}_q)$), its inverse $\mathbf{P}^{-1}$ (resp., $\mathbf{Q}^{-1}$) is also uniform over the same group. More generally, fix $\ell \geq 0$ and let $\mathbf{F}$ be uniform in $\mathrm{GL}_{m+\ell}(\mathbb{F}_q)$. Decompose $\mathbf{F}$ as

$$\mathbf{F} = \begin{pmatrix} \mathbf{S} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix},$$

where $\mathbf{S} \in \mathbb{F}_q^{m\times m}$, $\mathbf{B} \in \mathbb{F}_q^{m\times \ell}$, $\mathbf{C} \in \mathbb{F}_q^{\ell\times m}$, and $\mathbf{D} \in \mathbb{F}_q^{\ell\times \ell}$. We will compute the probability that $\mathbf{S}$ is invertible.

Let $P_1 := \#\{(\mathbf{F},\mathbf{S}) \mid \mathbf{F} \in \mathrm{GL}_{m+\ell}(\mathbb{F}_q)\ ,\ \mathbf{S} \in \mathrm{GL}_m(\mathbb{F}_q)\}$. We are interested in the probability

$$\Pr\left[\mathbf{S} \in \mathrm{GL}_m(\mathbb{F}_q) \mid \mathbf{F} \in \mathrm{GL}_{m+\ell}(\mathbb{F}_q)\right] = \frac{P_1}{|\mathrm{GL}_{m+\ell}(\mathbb{F}_q)|}.$$

To count the numerator, choose $\mathbf{S} \in \mathrm{GL}_m(\mathbb{F}_q)$, and choose $\mathbf{B} \in \mathbb{F}_q^{m\times\ell}$ and $\mathbf{C} \in \mathbb{F}_q^{\ell\times m}$ arbitrarily. For such fixed $(\mathbf{S},\mathbf{B},\mathbf{C})$, the Schur-complement criterion (see [50, Theorem 1.1]) implies that the matrix $\mathbf{F}$ is invertible if and only if

$$\mathbf{J} := \mathbf{D} - \mathbf{C}\mathbf{S}^{-1}\mathbf{B} \in \mathrm{GL}_\ell(\mathbb{F}_q).$$

Since $\mathbf{D} \mapsto \mathbf{J}$ is a bijection on $\mathbb{F}_q^{\ell\times\ell}$, there are precisely $|\mathrm{GL}_\ell(\mathbb{F}_q)|$ admissible choices of $\mathbf{D}$. Since $\mathbf{B}$ and $\mathbf{C}$ are arbitrary, each admits $q^{m\ell}$ choices. Therefore, $P_1 = |\mathrm{GL}_m(\mathbb{F}_q)| \cdot q^{m\ell} \cdot q^{m\ell} \cdot |\mathrm{GL}_\ell(\mathbb{F}_q)| = |\mathrm{GL}_m(\mathbb{F}_q)|\,|\mathrm{GL}_\ell(\mathbb{F}_q)|\,q^{2m\ell}$. Dividing by $|\mathrm{GL}_{m+\ell}(\mathbb{F}_q)|$ yields the exact probability $\Pr\left[\mathbf{S} \in \mathrm{GL}_m(\mathbb{F}_q) \mid \mathbf{F} \in \mathrm{GL}_{m+\ell}(\mathbb{F}_q)\right] = \frac{|\mathrm{GL}_m(\mathbb{F}_q)|\,|\mathrm{GL}_\ell(\mathbb{F}_q)|\,q^{2m\ell}}{|\mathrm{GL}_{m+\ell}(\mathbb{F}_q)|}$. Using the well-known formula $|\mathrm{GL}_t(\mathbb{F}_q)| = \prod_{i=0}^{t-1}(q^t - q^i)$ [33], we can rewrite the above conditional probability as

$$\Pr\big[\mathbf{S} \in \mathrm{GL}_m(\mathbb{F}_q) \ \big|\ \mathbf{F} \in \mathrm{GL}_{m+\ell}(\mathbb{F}_q)\big] = \frac{\prod_{k=1}^{m}(1-q^{-k})}{\prod_{k=\ell+1}^{m+\ell}(1-q^{-k})}.$$

(1) Invertibility of the top-left block of $\mathbf{P}^{-1}$. In this case, we take $\mathbf{F} = \mathbf{P}^{-1}$, $\ell = \ell_1$, hence the same expression applies to the top-left $m \times m$ block $\mathbf{S}_1$ of $\mathbf{P}^{-1}$. For convenience, denote this success probability by

$$p_{\mathrm{succ}_1} := \Pr\big[\mathbf{S}_1 \in \mathrm{GL}_m(\mathbb{F}_q) \ \big|\ \mathbf{P}^{-1} \in \mathrm{GL}_{m+\ell_1}(\mathbb{F}_q)\big].$$

Among the 16 parameter instances reported in Section 6, the worst case for the $\mathbf{P}$-sampling step is attained at $(q, m, \ell_1, \ell_2) = (2, 53, 5, 0)$. Accordingly,

$$p_{\mathrm{succ}_1} = \prod_{i=1}^{5} \frac{1-2^{-i}}{1-2^{-(53+i)}} \approx 0.2980.$$

Hence, the expected number of draws is $1/p_{\mathrm{succ}_1} \approx 3.356$, which means that the expected number of resamplings is about 2.356.

(2) Invertibility of the top-left block of $\mathbf{Q}^{-1}$. Specializing to $\mathbf{F} = \mathbf{Q}^{-1}$ with $\ell = \ell_2$, let $\mathbf{S}_2$ denote the top-left $m \times m$ block of $\mathbf{Q}^{-1}$ and set

$$p_{\mathrm{succ}_2} := \Pr\big[\mathbf{S}_2 \in \mathrm{GL}_m(\mathbb{F}_q) \;\big|\; \mathbf{Q}^{-1} \in \mathrm{GL}_{m+\ell_2}(\mathbb{F}_q)\big].$$

For the $\mathbf{Q}$-sampling step, the smallest success probability is attained at $(q, m, \ell_1, \ell_2) = (2, 43, 3, 1)$. Thus,

$$p_{\mathrm{succ}_2} = \frac{1 - 2^{-1}}{1 - 2^{-44}} \approx 0.5000.$$

Therefore, the expected number of draws is $1/p_{\mathrm{succ}_2} \approx 2.000$, and the corresponding expected number of resamplings is about 1.000.

E.2 Proof of Theorem 3

Proof. Recall that $N = (m+\ell_1)(m+\ell_2)$. Let $\mathcal{A}_{\mathrm{N}}$ be an efficient PPT algorithm for the EnGMS-Niederreiter scheme, which, given a public key and a ciphertext, outputs the underlying message with non-negligible probability $\varepsilon(\lambda)$. We denote this success probability by $\mathrm{Succ}(\mathcal{A}_{\mathrm{N}})$. Thus, $\mathrm{Succ}(\mathcal{A}_{\mathrm{N}}) \geq \varepsilon(\lambda)$.

We build a distinguisher $\mathcal{D}'$ for the EnGMS-IND problem. In the EnGMS-IND experiment, the challenger samples a K'-dimensional matrix code $\mathcal{C}'_{\mathrm{mat}} \subseteq \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_2)}$ either from the EnGMS distribution or uniformly at random. Let $\bar{\mathbf{H}}' \in \mathbb{F}_q^{(N-K')\times N}$ be a parity-check matrix of $\mathcal{C}'_{\mathrm{mat}}$ (according to the convention in Section 4.1). For each $i \in \{1, \ldots, N\}$, let $\boldsymbol{h}'_i$ be the i-th column of $\bar{\mathbf{H}}'$. The challenger gives $\bar{\mathbf{H}}'$ to $\mathcal{D}'$.

Distinguisher $\mathcal{D}'(\bar{\mathbf{H}}')$:

- Sample $\boldsymbol{\mu} \xleftarrow{\$} \Big\{\boldsymbol{m} \in \mathbb{F}_q^{m^2} \;\Big|\; \mathrm{rank}\big(\mathrm{Fold}(\boldsymbol{m})\big) \leq r\Big\}$.
- Let $\widetilde{\mathbf{U}} := \begin{pmatrix} \mathrm{Fold}(\boldsymbol{\mu}) & \mathbf{0}_{m\times\ell_2} \\ \mathbf{0}_{\ell_1\times m} & \mathbf{0}_{\ell_1\times\ell_2} \end{pmatrix} \in \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_2)}$, and set $\boldsymbol{\mu}' := \mathrm{Unfold}(\widetilde{\mathbf{U}})$.
- Compute the ciphertext $\boldsymbol{c} = \sum_{i=1}^{N} \mu'_i \boldsymbol{h}_i'^{\top}$.
- Compute $\hat{\boldsymbol{\mu}} \leftarrow \mathcal{A}_{\mathrm{N}}(\bar{\mathbf{H}}', \boldsymbol{c})$.
- Output 1 iff $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}$, otherwise output 0.

Now we define an algorithm $\mathcal{A}_{\mathrm{MRS}}$ for MinRank-Syndrome$(q, m+\ell_1, m+\ell_2, K', r)$. Given an instance $(\boldsymbol{y}, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_N)$ with $\boldsymbol{y}, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_N \in \mathbb{F}_q^{N-K'}$, where we set $\boldsymbol{y} := \boldsymbol{c}$ and $\boldsymbol{\mu}_i := \boldsymbol{h}_i'^{\top}$ for $i = 1, \ldots, N$. $\mathcal{A}_{\mathrm{MRS}}$ proceeds as follows:

- Input $(\bar{\mathbf{H}}', \boldsymbol{c})$.
- Compute $\hat{\boldsymbol{\mu}} \leftarrow \mathcal{A}_{\mathrm{N}}(\bar{\mathbf{H}}', \boldsymbol{c})$.
- Let $\widetilde{\mathbf{U}}' := \begin{pmatrix} \mathrm{Fold}(\hat{\boldsymbol{\mu}}) & \mathbf{0}_{m\times\ell_2} \\ \mathbf{0}_{\ell_1\times m} & \mathbf{0}_{\ell_1\times\ell_2} \end{pmatrix} \in \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_2)}$ and output $\boldsymbol{e} := \mathrm{Unfold}(\widetilde{\mathbf{U}}')$.

We say that $\mathcal{A}_{\mathrm{MRS}}$ succeeds if it outputs $\boldsymbol{e}$ such that $\boldsymbol{c} = \sum_{i=1}^{N} e_i \, \boldsymbol{h}_i'^{\top}$ and $\mathrm{rank}\big(\mathrm{Fold}(\boldsymbol{e})\big) \leq r$. We denote this success probability by $\mathrm{Succ}(\mathcal{A}_{\mathrm{MRS}})$.

Let Sub denote the event that $\bar{\mathbf{H}}'$ is sampled from the EnGMS distribution, and let Unif denote the event that $\bar{\mathbf{H}}'$ is sampled from a uniformly random K'-dimensional matrix code. By the definition of the advantage of the distinguisher $\mathcal{D}'$, we have

$$\begin{aligned}\mathrm{Adv}(\mathcal{D}') &= \big|\Pr\big[\mathcal{D}'(\bar{\mathbf{H}}') = 1 \mid \mathsf{Unif}\big] - \Pr\big[\mathcal{D}'(\bar{\mathbf{H}}') = 1 \mid \mathsf{Sub}\big]\big| \\ &= \big|\,\mathrm{Succ}_{\mathsf{Unif}}(\mathcal{A}_{\mathrm{N}}) - \mathrm{Succ}_{\mathsf{Sub}}(\mathcal{A}_{\mathrm{N}})\big|.\end{aligned}$$

Furthermore, if $\mathcal{A}_{\mathrm{N}}$ outputs $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}$, then the above construction yields $\boldsymbol{e} = \boldsymbol{\mu}'$. In particular,

$$\mathrm{rank}\big(\mathrm{Fold}(\boldsymbol{e})\big) = \mathrm{rank}\big(\mathrm{Fold}(\boldsymbol{\mu}')\big) = \mathrm{rank}\big(\mathrm{Fold}(\boldsymbol{\mu})\big) \leq r.$$

Therefore, $\mathrm{Succ}_{\mathsf{Unif}}(\mathcal{A}_{\mathrm{N}}) \leq \mathrm{Succ}(\mathcal{A}_{\mathrm{MRS}})$. Then we have

$$\mathrm{Succ}_{\mathsf{Sub}}(\mathcal{A}_{\mathrm{N}}) \leq \mathrm{Succ}(\mathcal{A}_{\mathrm{MRS}}) + \mathrm{Adv}(\mathcal{D}').$$

If $\mathrm{Succ}_{\mathsf{Sub}}(\mathcal{A}_{\mathrm{N}}) = \mathrm{Succ}(\mathcal{A}_{\mathrm{N}}) \geq \varepsilon(\lambda)$ is non-negligible, then either $\mathrm{Succ}(\mathcal{A}_{\mathrm{MRS}}) \geq \varepsilon(\lambda)/2$ or $\mathrm{Adv}(\mathcal{D}') \geq \varepsilon(\lambda)/2$, contradicting the assumptions. Therefore, the $\mathsf{EnGMS\text{-}Niederreiter}$ scheme is OW-CPA secure.

E.3 FO-based KEM and Hybrid PKE for EnGMS-Niederreiter

Table 6: Notation for $\mathsf{KEM}_{\mathrm{N}}^{\mathrm{FO}}$.

Symbol	Meaning
$\mathcal{M}_{\leq r}$	Message space in Fig. 1, $\mathcal{M}_{\leq r} := \{\boldsymbol{m} \in \mathbb{F}_q^{m^2} \mid \mathrm{rank}(\mathrm{Fold}(\boldsymbol{m})) \leq r\}$.
$\mathrm{enc}(\cdot)$	Fixed injective encoding into $\{0,1\}^*$ (for vectors or matrices).
$\boldsymbol{s}$	Secret value sampled in $\mathsf{KeyGen}^{\mathrm{FO}}$, $\boldsymbol{s} \xleftarrow{\$} \mathcal{M}_{\leq r}$.
G	Hash function modeled as a random oracle, $\mathsf{G} : \mathcal{M}_{\leq r} \to \{0,1\}^*$ (with sufficiently long output).
κ	Session-key length (bits).
H	Hash function modeled as a random oracle, $\mathsf{H} : \{0,1\}^* \to \{0,1\}^\kappa$.
K	Session key (with encapsulated message $\boldsymbol{\mu} \in \mathcal{M}_{\leq r}$, ciphertext $\boldsymbol{c}$, and secret $\boldsymbol{s}$): $\mathsf{K} := \mathsf{H}(\mathrm{enc}(\boldsymbol{\mu}))$ if accepted; otherwise $\mathsf{K} := \mathsf{H}(\mathrm{enc}(\boldsymbol{c}) \| \mathrm{enc}(\boldsymbol{s}))$ (implicit rejection).

Notation: See Table 6 and Fig. 1.

KeyGen$^{\mathrm{FO}}(1^\lambda)$:
– Run $(\mathsf{pk}_\mathrm{N}, \mathsf{sk}_\mathrm{N}) \leftarrow \mathsf{KeyGen}(1^\lambda)$ of PKE_N, where $\mathsf{pk}_\mathrm{N} = \bar{\mathbf{H}}'$ and $\mathsf{sk}_\mathrm{N} = (\boldsymbol{g}, \gamma, \mathbf{P}, \mathbf{Q})$.
– Sample $\boldsymbol{s} \xleftarrow{\$} \mathcal{M}_{\leq r}$.
– Return $\mathsf{pk} = \mathsf{pk}_\mathrm{N}$ and $\mathsf{sk} = (\mathsf{sk}_\mathrm{N}, \boldsymbol{s})$.

Encaps$^{\mathrm{FO}}(\mathsf{pk})$:
Input: $\mathsf{pk} = \bar{\mathbf{H}}'$.
– Sample $\boldsymbol{\mu} \xleftarrow{\$} \mathcal{M}_{\leq r}$.
– Compute $\omega \leftarrow \mathsf{G}(\boldsymbol{\mu})$.
– Compute $\boldsymbol{c} \leftarrow \mathsf{Enc}(\mathsf{pk}, \boldsymbol{\mu})$ of PKE_N (ignoring ω if Enc is deterministic).
– Set $\mathsf{K} \leftarrow \mathsf{H}(\mathrm{enc}(\boldsymbol{\mu}))$.
– Return $(\mathsf{K}, \boldsymbol{c})$.

Decaps$^{\mathrm{FO}}(\mathsf{sk}, \boldsymbol{c})$:
Input: $\mathsf{sk} = (\mathsf{sk}_\mathrm{N}, \boldsymbol{s})$ and ciphertext $\boldsymbol{c}$.
– Compute $\hat{\boldsymbol{\mu}} \leftarrow \mathsf{Dec}(\mathsf{sk}_\mathrm{N}, \boldsymbol{c})$ of PKE_N.
– If $\hat{\boldsymbol{\mu}} = \perp$, set $\mathsf{K} := \mathsf{H}(\mathrm{enc}(\boldsymbol{c}) \| \mathrm{enc}(\boldsymbol{s}))$ and return K.
– Compute $\hat{\omega} \leftarrow \mathsf{G}(\hat{\boldsymbol{\mu}})$ and re-encrypt $\hat{\boldsymbol{c}} \leftarrow \mathsf{Enc}(\mathsf{pk}_\mathrm{N}, \hat{\boldsymbol{\mu}})$ of PKE_N (ignoring $\hat{\omega}$ if Enc is deterministic).
– If $\hat{\boldsymbol{c}} \neq \boldsymbol{c}$, set $\mathsf{K} := \mathsf{H}(\mathrm{enc}(\boldsymbol{c}) \| \mathrm{enc}(\boldsymbol{s}))$; else set $\mathsf{K} := \mathsf{H}(\mathrm{enc}(\hat{\boldsymbol{\mu}}))$.
– Return K.

Fig. 5: $\mathsf{KEM}_\mathrm{N}^{\mathrm{FO}}$ derived from PKE_N (Fig. 1) via $\mathrm{FO}_m^{\not\perp}$.

Notation: See Table 7 and Fig. 5.

KeyGen(1^λ):
– Run $(\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{KeyGen}^{\mathrm{FO}}(1^\lambda)$ of $\mathsf{KEM}_\mathrm{N}^{\mathrm{FO}}$.
– Return $\mathsf{pk}' = \mathsf{pk}$ and $\mathsf{sk}' = \mathsf{sk}$.

Encrypt(pk', m):
Input: pk' and a message $m \in \mathcal{M}_{\mathrm{SKE}}$.
– Run $(\mathsf{K}, \boldsymbol{c}) \leftarrow \mathsf{Encaps}^{\mathrm{FO}}(\mathsf{pk}')$ of $\mathsf{KEM}_\mathrm{N}^{\mathrm{FO}}$.
– Compute $\chi \leftarrow \mathsf{Enc}_{\mathrm{SKE}}(\mathsf{K}, m)$.
– Compute the hybrid ciphertext $\hat{c} := \mathrm{enc}(\boldsymbol{c}, \chi)$, i.e., a canonical bit-string encoding of the pair $(\boldsymbol{c}, \chi)$.
– Return $\hat{c}$.

Decrypt$(\mathsf{sk}', \hat{c})$:
Input: sk' and a ciphertext $\hat{c}$.
– $(\boldsymbol{c}, \chi) \leftarrow \mathsf{Parse}_c(\hat{c})$; if $(\boldsymbol{c}, \chi) = \perp$, return $\perp$.
– Compute $\mathsf{K} \leftarrow \mathsf{Decaps}^{\mathrm{FO}}(\mathsf{sk}', \boldsymbol{c})$ of $\mathsf{KEM}_\mathrm{N}^{\mathrm{FO}}$.
– Compute $\hat{m} \leftarrow \mathsf{Dec}_{\mathrm{SKE}}(\mathsf{K}, \chi)$.
– Return $\hat{m}$ (possibly $\perp$).

Fig. 6: $\mathsf{PKE}_\mathrm{N}^{\mathrm{FO}}$ derived from $\mathsf{KEM}_\mathrm{N}^{\mathrm{FO}}$ (Fig. 5) via KEM-DEM paradigm.

Table 7: Notation for $\mathsf{PKE}_{\mathrm{N}}^{\mathrm{FO}}$.

Symbol	Meaning
K	Session key output by $\mathsf{Encaps}^{\mathrm{FO}}$ or $\mathsf{Decaps}^{\mathrm{FO}}$ (implicit rejection as in Table 6).
κ	Session-key length (bits).
$\boldsymbol{c}$	KEM ciphertext output by $\mathsf{Encaps}^{\mathrm{FO}}$ of $\mathsf{KEM}_{\mathrm{N}}^{\mathrm{FO}}$.
SKE	One-time chosen-ciphertext secure symmetric encryption scheme $\mathsf{SKE} = (\mathsf{Enc}_{\mathrm{SKE}}, \mathsf{Dec}_{\mathrm{SKE}})$ with key space $\{0,1\}^{\kappa}$ and message space $\mathcal{M}_{\mathrm{SKE}}$.
m	Plaintext message, $m \in \mathcal{M}_{\mathrm{SKE}}$.
χ	SKE ciphertext of message m under session key K.
$\mathrm{enc}(\cdot,\cdot)$	Canonical, uniquely decodable encoding of a pair into $\{0,1\}^*$ (used for $(\boldsymbol{c}, \chi)$).
$\hat{c}$	Hybrid ciphertext (canonical encoding of $(\boldsymbol{c}, \chi)$).
Parse_c	Deterministic parser: on input $\hat{c}$ it outputs the unique $(\boldsymbol{c}, \chi)$ such that $\hat{c} = \mathrm{enc}(\boldsymbol{c}, \chi)$, or $\perp$.

F Matrix-Code Duality and Dual-Code Attack Analysis

F.1 Matrix-Code Duality: Definitions and Representations

Under the identification induced by Unfold, a matrix code $\mathcal{C}_{\mathrm{mat}} \subseteq \mathbb{F}_q^{m\times n}$ of dimension K can be specified by a generator matrix $\mathbf{G} \in \mathbb{F}_q^{K\times mn}$. Here, each row of $\mathbf{G}$ is the Unfold representation of a basis matrix of $\mathcal{C}_{\mathrm{mat}}$. Let $\mathrm{Unfold}(\mathcal{C}_{\mathrm{mat}}) := \{\mathrm{Unfold}(\mathbf{A}) \mid \mathbf{A} \in \mathcal{C}_{\mathrm{mat}}\}$ denote the associated linear code of length mn. Accordingly, one can choose a parity-check matrix $\mathbf{H} \in \mathbb{F}_q^{(mn-K)\times mn}$ for $\mathrm{Unfold}(\mathcal{C}_{\mathrm{mat}})$. Moreover, this representation provides a convenient way to describe the dual matrix code: by folding the rows of $\mathbf{H}$ back into $m \times n$ matrices, one obtains a set of generators for $\mathcal{C}_{\mathrm{mat}}^{\perp}$. We provide the formal definition next.

Definition 21 (Matrix Code Duality). *Let $\mathcal{C}_{\mathrm{mat}}$ be a matrix code of size $m \times n$ and dimension K. Its dual code, denoted by $\mathcal{C}_{\mathrm{mat}}^{\perp}$, is the matrix code of size $m \times n$ and dimension $mn - K$, defined by:*

$$\mathcal{C}_{\mathrm{mat}}^{\perp} := \left\{\mathbf{B} \in \mathbb{F}_q^{m\times n} \mid \mathrm{tr}(\mathbf{A}\mathbf{B}^{\top}) = 0,\ \forall\, \mathbf{A} \in \mathcal{C}_{\mathrm{mat}}\right\}.$$

F.2 Dual-Code Attack Analysis

For comparison, we briefly recall the distinguisher on the dual of the public code $\mathcal{C}_{\mathrm{pub}}$ in the original EGMC setting.

After left-multiplication by $\mathbf{P}^{\top}$ and right-multiplication by $\mathbf{Q}^{\top}$, the dual of the EGMC public code can be written as

$$\mathcal{D}_{\mathrm{mat}} = \mathcal{U}_{n-k} \oplus \mathcal{W},$$

where $\mathcal{U}_{n-k}$ denotes a matrix representation of a Gabidulin code of length n and $\mathbb{F}_{q^m}$-dimension $n-k$ *extended by zero*, i.e., embedded into $\mathbb{F}_q^{(m+\ell_1)\times(n+\ell_2)}$ by padding ℓ_1 all-zero rows and ℓ_2 all-zero columns, and $\mathcal{W}$ is a complementary subspace of dimension $n\ell_1 + m\ell_2 + \ell_1\ell_2$.

Let $\mathbf{D} \in \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_1)}$ be a matrix corresponding to a Gabidulin code of length m and $\mathbb{F}_{q^m}$-dimension b, extended by zero. Using the inclusion

$$\mathcal{U}_{n-k} + \mathbf{D}\mathcal{U}_{n-k} \subseteq \mathcal{U}_{n-k+b-1},$$

one obtains the dimension bound

$$\dim_{\mathbb{F}_q}(\mathcal{D}_{\mathrm{mat}} + \mathbf{D}\mathcal{D}_{\mathrm{mat}}) \le m(n-k+b-1) + 2(n\ell_1 + m\ell_2 + \ell_1\ell_2).$$

Since the rate of $\mathcal{D}_{\mathrm{mat}}$ is greater than $1/2$, for a uniformly random $\mathcal{D}_{\mathrm{mat}}$, $\mathcal{D}_{\mathrm{mat}} + \mathbf{D}\mathcal{D}_{\mathrm{mat}}$ is expected to span the entire ambient space $\mathbb{F}_q^{(m+\ell_1)\times(n+\ell_2)}$ with high probability. In contrast, for the structured EGMC dual code, this fails whenever

$$b < k + 1 - \frac{n}{m}\ell_1 - \ell_2 - \frac{\ell_1\ell_2}{m}. \tag{11}$$

Exploiting this condition leads to a bilinear system in the unknowns $\mathbf{D}$ and $\mathbf{M} \in \mathcal{C}_{\mathrm{pub}}$, defined by

$$\mathrm{tr}(\mathbf{D}\mathbf{C}\mathbf{M}^\top) = 0 \quad \text{for all } \mathbf{C} \in \mathcal{C}_{\mathrm{pub}}^\perp.$$

As in the public-code setting, the number of unknowns significantly exceeds the number of equations, rendering the system computationally intractable for the parameter sizes considered in [5].

We now turn to attacks on the dual of our public code. In this section, we assume that $2K' < (m+\ell_1)(n+\ell_2)$ and follow the notation of Theorem 1.

1) Case $m \nmid K'$: Absence of $\mathbb{F}_{q^m}$-Linearity

After left- and right-multiplication by $\mathbf{P}^\top$ and $\mathbf{Q}^\top$, respectively, the dual of our public code takes the form

$$\mathcal{D}'_{\mathrm{mat}} = \left(\mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}_1) \oplus \mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}_2)\right) \oplus \mathcal{W},$$

where

$$\mathcal{B}_1 := \left\{ \begin{pmatrix} \mathbf{B}_i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \middle| \; i = 1, \ldots, m(n-k) \right\}, \; \mathcal{B}_2 := \left\{ \begin{pmatrix} \widetilde{\mathbf{B}}_j & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \middle| \; j = 1, \ldots, mk - K' \right\}.$$

Here, $\{\mathbf{B}_i\}_{i=1}^{m(n-k)}$ is an $\mathbb{F}_q$-basis of $\Psi_\gamma(\mathcal{G})^\perp$, $\mathcal{W}$ is a complementary subspace of dimension $n\ell_1 + m\ell_2 + \ell_1\ell_2$, and $\{\widetilde{\mathbf{B}}_j\}_{j=1}^{mk-K'}$ is an $\mathbb{F}_q$-basis of a complement of $\Psi_\gamma(\mathcal{G})^\perp$ inside $\mathcal{C}_{\mathrm{sub}}^\perp$. Clearly, $\mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}_1)$ corresponds to $\mathcal{U}_{n-k}$ in the original EGMC scheme. Thus, our scheme introduces an additional term $\mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}_2)$.

When $m \nmid K'$, the dimension of $\mathrm{Span}_{\mathbb{F}_q}(\mathcal{B}_2)$ equals $mk - K'$ and is not a multiple of m. By the same dimension argument as in Lemma 1, this space is not $\mathbb{F}_{q^m}$-linear. We therefore denote it by $\mathcal{V}_{\text{non-}\mathbb{F}_{q^m}}$.

Next we show that, for the same parameters $(k, m, n, \ell_1, \ell_2)$ as in [5], this dual-side Overbeck-like attack is more difficult to carry out against our scheme.

1. The space $\mathcal{V}_{\text{non-}\mathbb{F}_{q^m}}$ is not the Ψ_γ-expansion of any $\mathbb{F}_{q^m}$-linear subcode of length n extended by zero. Let $\mathbf{D} \in \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_1)}$ be a matrix corresponding to a Gabidulin code of length m and $\mathbb{F}_{q^m}$-dimension b, extended by zero. Typically, $\mathbf{D}\,\mathcal{V}_{\text{non-}\mathbb{F}_{q^m}}$ is not contained in some $\mathcal{U}_\star$ (a matrix form of some Gabidulin code over $\mathbb{F}_{q^m}$ extended by zero).
2. The original bound $\dim_{\mathbb{F}_q}\big(\mathcal{D}_{\text{mat}} + \mathbf{D}\,\mathcal{D}_{\text{mat}}\big) \le m(n-k+b-1) + 2(n\ell_1 + m\ell_2 + \ell_1\ell_2)$ is valid only for the pure $\mathcal{U}_{n-k}$ block. With $\mathcal{V}_{\text{non-}\mathbb{F}_{q^m}}$ present, the dimension bound must be augmented by an additional contribution $\delta := \dim_{\mathbb{F}_q}\big(\mathcal{V}_{\text{non-}\mathbb{F}_{q^m}} + \mathbf{D}\mathcal{V}_{\text{non-}\mathbb{F}_{q^m}}\big)$, with $0 < mk - K' \le \delta \le 2(mk - K')$. Consequently, the analogue of (11) becomes

$$b < k + 1 - \frac{n}{m}\ell_1 - \ell_2 - \frac{\ell_1\ell_2}{m} - \frac{\delta}{m}. \tag{12}$$

For fixed parameters $(k, m, n, \ell_1, \ell_2)$, this condition is strictly more restrictive than (11). As a result, the admissible range of b for which the dual distinguisher may apply is reduced, making the attack correspondingly harder to mount using the same methodology as in [5].

2) Case $m \mid K'$: Conservative Structural Assumption

We consider the case $K' = mk'$ with $k' < k$. The condition $m \mid K'$ alone does not guarantee that $\text{Span}_{\mathbb{F}_q}(\mathcal{B}_2)$ is the matrix form of a Gabidulin code of $\mathbb{F}_{q^m}$-dimension $k - k'$ extended by zero; in particular, it may even fail to be $\mathbb{F}_{q^m}$-linear. For a conservative analysis, we therefore assume that $\text{Span}_{\mathbb{F}_q}(\mathcal{B}_2)$ is the matrix form of such a Gabidulin code extended by zero. Under this stronger hypothesis, the dual Overbeck-like distinguisher used in [5] remains applicable.

Let $\text{Span}_{\mathbb{F}_q}(\mathcal{B}_2)$ be abbreviated as $\mathcal{V}_{k-k'}$, which is the matrix form of a Gabidulin code of $\mathbb{F}_{q^m}$-dimension $k - k'$ extended by zero. Since the dual of our public code has the structure

$$\mathcal{D}'_{\text{mat}} = \mathcal{U}_{n-k} \oplus \mathcal{V}_{k-k'} \oplus \mathcal{W},$$

let $\mathbf{D} \in \mathbb{F}_q^{(m+\ell_1)\times(m+\ell_1)}$ be a matrix associated with a Gabidulin code of length m and $\mathbb{F}_{q^m}$-dimension b, extended by zero. Then $\mathcal{V}_{k-k'} + \mathbf{D}\,\mathcal{V}_{k-k'} \subseteq \mathcal{U}_{k-k'+b-1}$, where $\mathcal{U}_{k-k'+b-1}$ denotes the matrix form of a Gabidulin code of length n and $\mathbb{F}_{q^m}$-dimension $k - k' + b - 1$ extended by zero. Consequently,

$$\mathcal{U}_{n-k} + \mathbf{D}\mathcal{U}_{n-k} + \mathcal{V}_{k-k'} + \mathbf{D}\mathcal{V}_{k-k'} \subseteq \mathcal{U}_{\max\{n-k+b-1,\, k-k'+b-1\}}.$$

Since $k' < k \le n$, it follows that $n - k < n - k'$ and $k - k' \le n - k'$. Hence, $\max\{n - k + b - 1,\ k - k' + b - 1\} \le n - k' + b - 1$. Thus we may write $(\mathcal{U}_{n-k} + \mathbf{D}\mathcal{U}_{n-k}) + (\mathcal{V}_{k-k'} + \mathbf{D}\mathcal{V}_{k-k'}) \subseteq \mathcal{U}_{n-k'+b-1}$.

This yields the condition $m(n - k' + b - 1) + 2(n\ell_1 + m\ell_2 + \ell_1\ell_2) < (m + \ell_1)(n + \ell_2)$, which is equivalent to

$$b < k' + 1 - \frac{n}{m}\ell_1 - \ell_2 - \frac{\ell_1\ell_2}{m}. \tag{13}$$

For the same system parameters $(k, m, n, \ell_1, \ell_2)$, the bound in (13) is strictly smaller than that in (11) because $k' < k$. Hence, under identical parameters, the distinguisher can be triggered only for smaller values of b, making the attack correspondingly more difficult.

Remark 6 (Case $b = 0$). Assume that $K' = mk'$ with $k' < k$ and the code $\mathcal{D}'_{\text{mat}} = (\mathcal{U}_{n-k} \oplus \mathcal{V}_{k-k'}) \oplus \mathcal{W}$, where $\mathcal{U}_{n-k}$ and $\mathcal{V}_{k-k'}$ are both $\mathbb{F}_{q^m}$-linear codes extended by zero. In this case, the dual-side Overbeck-like approach with parameter $b = 0$ reduces to detecting hidden $\mathbb{F}_{q^m}$-linearity of the code $\mathcal{U}_{n-k} \oplus \mathcal{V}_{k-k'}$.

When $m \nmid K'$, we have $\mathcal{D}'_{\text{mat}} = (\mathcal{U}_{n-k} \oplus \mathcal{V}_{\text{non-}\mathbb{F}_{q^m}}) \oplus \mathcal{W}$, where $\mathcal{U}_{n-k}$ is the known $\mathbb{F}_{q^m}$-linear part, while $\mathcal{V}_{\text{non-}\mathbb{F}_{q^m}}$ is not $\mathbb{F}_{q^m}$-linear, and $\mathcal{W}$ is not necessarily $\mathbb{F}_{q^m}$-linear. Thus, in the case $b = 0$, the $\mathbb{F}_{q^m}$-linear part $\mathcal{U}_{n-k}$ is mixed with these additional structures, and we currently do not know how to exploit this structure to derive a more effective attack.

Remark 7 (Combinatorial Approach on the Dual). Consider the case where $K' = mk'$ with $k' < k$ and the code $\mathcal{D}'_{\text{mat}} = (\mathcal{U}_{n-k} \oplus \mathcal{V}_{k-k'}) \oplus \mathcal{W}$ has an $\mathbb{F}_{q^m}$-linear part $\mathcal{U}_{n-k} \oplus \mathcal{V}_{k-k'}$ and a complementary subspace $\mathcal{W}$. A combinatorial approach on the dual that tries to exploit this $\mathbb{F}_{q^m}$-linear part would typically need to guess a shortening of the row and column spaces in order to suppress the contribution of $\mathcal{W}$. This leads to a search problem that is very similar in nature to the combinatorial distinguisher for the public code discussed in Section 5.1.1.

When $m \nmid K'$, we have $\mathcal{D}'_{\text{mat}} = (\mathcal{U}_{n-k} \oplus \mathcal{V}_{\text{non-}\mathbb{F}_{q^m}}) \oplus \mathcal{W}$, where $\mathcal{U}_{n-k}$ is the $\mathbb{F}_{q^m}$-linear part, while $\mathcal{V}_{\text{non-}\mathbb{F}_{q^m}}$ is not $\mathbb{F}_{q^m}$-linear, and $\mathcal{W}$ is not necessarily $\mathbb{F}_{q^m}$-linear. In this situation, a combinatorial reasoning on the dual would first need to get rid of the contributions of $\mathcal{V}_{\text{non-}\mathbb{F}_{q^m}}$ and $\mathcal{W}$ before exposing a Gabidulin-type structure, and we currently do not know how to achieve this with lower complexity than in the combinatorial attack on the public code described above.

G Size Trade-offs of CPA-Secure PKE Schemes

This section provides a comparative evaluation of public-key encryption (PKE) schemes achieving CPA security, with an emphasis on size-related performance metrics. We compare public key sizes (pk), ciphertext sizes (ct), and supported plaintext lengths (pt) across three standardized security levels (SL=128, 192, 256). For our EnGMS-Niederreiter scheme, the supported plaintext length follows from the encoding in Section 4.1.1 and equals $2r(m-r)\log_2 q$ bits for parameters (q, m, r). The results are summarized in Table 8.

A key feature of the EnGMS-Niederreiter family is the flexibility it offers in parameter selection. Across all security levels, EnGMS admits multiple instantiations that expose a wide spectrum of trade-offs between public key size, ciphertext expansion, and plaintext length. At SL=128, public key sizes range from 25 kB to 137 kB, while ciphertext sizes vary from as low as 55 B up to 234 B. At the highest security level SL=256, this flexibility persists, with public keys spanning 61 kB to 618 kB and ciphertexts between 110 B and 168 B. Notably,

several parameter sets allow plaintext sizes exceeding 100 B, which is uncommon among code-based CPA-secure PKE schemes.

In contrast, many existing schemes follow a more rigid design philosophy. Lattice-based constructions such as KYBER enforce a fixed plaintext size of 32 B across all security levels, while achieving relatively compact public keys and ciphertexts. Code-based schemes optimized for IND-CPA security, including BIKE, ROLLO-II, RQC, and HQC, similarly favor fixed and short plaintexts, typically between 16 B and 64 B. This design choice simplifies parameterization but limits applicability in scenarios requiring the encryption of larger payloads.

Several schemes incur noticeable ciphertext expansion as security increases. For instance, RQC and HQC incur ciphertext sizes exceeding 8 kB and 14 kB respectively at SL=256, while BIKE ciphertexts grow beyond 5 kB. By comparison, EnGMS-Niederreiter maintains ciphertext sizes below 235 B across all security levels, at the cost of larger public keys. This asymmetric trade-off is particularly attractive in settings where public keys can be amortized or distributed infrequently, while bandwidth-efficient ciphertexts are critical.

Overall, the EnGMS-Niederreiter family offers a scalable and versatile performance profile for CPA-secure PKE. Rather than optimizing a single metric, it exposes a broad design space that allows practitioners to balance public key size, ciphertext overhead, and plaintext capacity according to application-specific constraints. This flexibility distinguishes EnGMS from existing lattice-based and code-based alternatives, which typically commit to a narrow operating point in the design space.

Table 8: Comparison of CPA-secure PKE schemes.

Scheme	SL=128			SL=192			SL=256		
	pk	ct	pt	pk	ct	pt	pk	ct	pt
EnGMS-Niederreiter	137 kB	55 B	37 B	248 kB	83 B	55 B	618 kB	110 B	75 B
EnGMS-Niederreiter	90 kB	60 B	39 B	202 kB	90 B	60 B	261 kB	127 B	80 B
EnGMS-Niederreiter	61 kB	62 B	40 B	156 kB	95 B	61 B	240 kB	135 B	86 B
EnGMS-Niederreiter	32 kB	85 B	60 B	49 kB	104 B	66 B	70 kB	140 B	92 B
EnGMS-Niederreiter	25 kB	234 B	126 B	40 kB	119 B	76 B	61 kB	168 B	100 B
KYBER [6] (IND-CPA)	800 B	768 B	32 B	1184 B	1088 B	32 B	1568 B	1568 B	32 B
BIKE [4] (OW-CPA)	1541 B	1541 B	252 B	3083 B	3083 B	398 B	5122 B	5122 B	561 B
BIKE [4] (IND-CPA)	1541 B	1573 B	32 B	3083 B	3115 B	32 B	5122 B	5154 B	32 B
ROLLO-II [3] (IND-CPA)	1961 B	2025 B	64 B	2341 B	2405 B	64 B	2559 B	2623 B	64 B
RQC [38] (IND-CPA)	1834 B	3588 B	48 B	2853 B	5626 B	95 B	4090 B	8100 B	68 B
HQC [25] (IND-CPA)	2241 B	4417 B	16 B	4514 B	8962 B	24 B	7237 B	14405 B	32 B