New Construction of Code-Based Signature Schemes

Yang Yang[1,2] and Fangguo Zhang[1,2(✉)]

[1] School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China
yangy789@mail2.sysu.edu.cn, isszhfg@mail.sysu.edu.cn
[2] Guangdong Key Laboratory of Information Security Technology, Guangzhou 510006, China

Abstract. In this paper, we present a novel approach to construct code-based digital signature schemes. Our focus is on addressing the efficiency issues stemming from the low proportion of decodable syndromes in classical CFS signature scheme. To overcome this challenge, we introduce a well-defined syndrome subspace wherein all syndromes can be efficiently decoded. Additionally, our scheme differs from traditional code-based approaches by employing a novel method for constructing public and private keys, replacing permutation matrices with invertible matrices. This transformation yields a public key matrix that is statistically indistinguishable from a random matrix. We demonstrate that our scheme achieves existential unforgeability under an adaptive chosen message attack (EUF-CMA) in the random oracle model. Comparative analysis against other code-based signature schemes, such as Wave and Enhanced pqsigRM, reveals significant advantages in both public key and signature sizes. Furthermore, compared to the three post-quantum signature schemes (Crystals-Dilithium, Falcon, and Sphincs+) selected as finalists by NIST PQC, our scheme still maintains a significant advantage in terms of signature size.

Keywords: Post-quantum cryptography · Code-based cryptography · Syndrome decoding · Digital signatures.

1 Introduction

Digital signature schemes, as one of the vital means to safeguard network information security, find applications across various domains such as commerce, politics, and the military. With the emergence of Shor's quantum algorithm [38,39], Grover's quantum algorithm [23], and the rapid development of quantum algorithms in recent years [19, 36, 37], classical digital signature schemes based on traditional problems, like RSA, DSA, and ECDSA face significant challenges. The development of an efficient, secure, and quantum-resistant digital signature scheme is a topic worthy of in-depth research.

In 1978, Berlekamp et al. [10] proved that the decoding problem for general linear codes is an NP-complete problem. In the same year, McEliece [30] leveraged this challenge, along with the fast decoding capabilities of Goppa codes, to

constructing a class of code-based public key cryptosystems. This cryptosystem is known for its simplicity, ease of encryption and decryption, and has undergone extensive analysis for over four decades. With appropriate parameter and code type selections, it is considered one of the most secure public key cryptosystems available today.

In 2001, Courtois, Finiasz, and Sendrier [16] (CFS) presented the first secure code-based digital signature scheme following the "hash-and-sign" paradigm, with security analyzed by Dallot [17]. The CFS signature scheme employs binary Goppa codes, with its security relying on the syndrome decoding problem and the distinguishing Goppa code problem. For Goppa codes capable of correcting t errors, the CFS signature scheme, on average, requires $t!$ hash computations to obtain a decodable syndrome. However, on the one hand, a large t value results in longer signature times, and on the other hand, a small t value leads to a higher code rate for Goppa codes, making them distinguishable from random linear codes [20]. Over the past two decades, many researchers have proposed various improved schemes to address the efficiency issues of the CFS scheme [15, 18, 22, 28]. Among them, the more effective ones are Wave [18] and Enhanced pqsigRM [15], which employ generalized $(U, U+V)$-codes and modified RM codes, respectively, as replacements for the Goppa codes in CFS. The Enhanced pqsigRM signature scheme proposed by Cho et al. [15] is an enhancement of the pqsigRM [28] scheme submitted to NIST during the first round of PQC standardization. It addresses the weaknesses of the early versions of pqsigRM by modifying the public code and is resistant to attacks by Minder-Shokrollahi [32] and Chizhov-Borodin [14]. In contrast to previous schemes, the Wave signature scheme employs ternary generalized $(U, U+V)$-codes and utilizes large-weight vectors as signatures. In the "hash-and-sign" paradigm, this scheme provides robust security against various attacks and is currently one of the best signature schemes available.

Additionally, other types of code-based signature schemes have been proposed in the literature, such as the KKS scheme [26] and its variant [5]. The KKS signature schemes achieve signature generation by constructing a specific subset of syndrome space, allowing the signer to perform efficient decoding operations on these syndromes and use the decoding results as signatures. However, the security of KKS-type signature schemes is relatively low and can at most be considered as one-time signature schemes [12, 33]. Another significant approach is the use of the Fiat-Shamir transformation to convert zero-knowledge identification schemes into signature schemes. This transformation is essentially an adaptation of interactive identification processes into non-interactive ones. In 1993, Stern [41] introduced the first code-based zero-knowledge identification scheme, which required three rounds of interaction and had a success probability of $2/3$ for deception. In other words, achieving λ-bit security necessitated approximately 1.7λ repetitions, resulting in relatively lengthy signatures. To reduce signature length, numerous scholars have proposed improvements, including reducing the communication overhead in a single round and lowering the deception probability [11, 13, 21, 31].

In comparison to schemes based on Hamming weight decoding problems, schemes based on rank metric decoding problems [3, 7–9] can achieve the same security level with smaller security parameters. Nevertheless, these schemes still face challenges related to signature length and zero-knowledge property [24, 27]. In 2019, Aragon et al. [1] introduced the Durandal signature scheme based on a variant of Lyubashevsky-type code-based identification protocols, which features small public key. Subsequently, Song et al. [40] made further enhancements, allowing the improved Durandal scheme's security to be reduced to the rank syndrome decoding problem and reducing the sizes of public key and signature. However, in 2023, Aragon et al. [2] presented a new attack against the product spaces subspaces indistinguishability problem, one of the three security foundations of the Durandal scheme, compromising all existing parameters of Durandal.

Our Contributions. In this work, we introduce a new code-based digital signature scheme following the "hash-and-sign" paradigm. Classical CFS digital signature scheme suffer from the impracticality of only a small fraction of the syndromes being efficiently decodable, with no discernible patterns among these efficiently decodable syndromes. One of the contributions of this paper is the design of a trapdoor that constructs a subspace where any syndrome is efficiently decodable, and this subspace can be randomly selected. In our scheme, although only a small fraction of the syndromes are efficiently decodable, as they constitute a subspace, the efficiency of the signatures can be improved by ensuring that the hash values of all messages lie within this subspace. Such a hash construction is straightforward, and in the construction of the signature scheme in Section 3 we give the concrete method.

The second significant contribution of this paper is the column transformation of the private key matrix, ensuring that the resulting public key matrix is statistically indistinguishable from a random matrix. Traditional signature schemes based on the syndrome decoding problem transform the private key matrix $\mathbf{H}_{\mathsf{sk}}$ into the public key matrix $\mathbf{H}_{\mathsf{pk}} = \mathbf{S}\mathbf{H}_{\mathsf{sk}}\mathbf{P}$ using a non-singular matrix $\mathbf{S}$ and a permutation matrix $\mathbf{P}$ for row and column transformations. The resulting matrix $\mathbf{H}_{\mathsf{pk}}$ obtained in this way has the same weight distribution as the codes generated by the matrix $\mathbf{H}_{\mathsf{sk}}$, which typically differs from that of a random code. In our scheme, we perform row and column transformations on the private key matrix $\mathbf{H}_{\mathsf{sk}}$ using two non-singular matrices $\mathbf{S}$ and $\mathbf{U}$ simultaneously to obtain the public key matrix $\mathbf{H}_{\mathsf{pk}} = \mathbf{S}\mathbf{H}_{\mathsf{sk}}\mathbf{U}$. Since $\mathbf{U}$ is a reversible matrix with uniformly distributed elements 0 and 1, the weight distribution of the code generated by $\mathbf{H}_{\mathsf{pk}}$ differs from that of $\mathbf{H}_{\mathsf{sk}}$. Additionally, since $\mathbf{H}_{\mathsf{sk}}$ is composed of three random matrices and a zero matrix, each column of $\mathbf{H}_{\mathsf{pk}}$ is also a random vector, making $\mathbf{H}_{\mathsf{pk}}$ a random matrix. Consequently, the public key matrix $\mathbf{H}_{\mathsf{pk}}$ is statistically indistinguishable from a random matrix.

Organization. The rest of the paper is structured as follows. In Section 2, we review some preliminaries that will be used in the following sections. In Section

3, we present the detailed construction of the signature scheme proposed in this paper. In Section 4, we provide a security analysis of our scheme, demonstrating its existential unforgeability under an adaptive chosen message attack. Section 5 provides two sets of parameters and offers a comparison with other signature schemes. Finally, some conclusions are given in Section 6.

2 Preliminaries

In this section, we review some fundamental definitions in coding theory and introduce some symbols that will be used in subsequent sections.

Notation. Throughout the paper, $\mathbb{F}_2$ represents the finite field with two elements. Bold lowercase letters (e.g., $\mathbf{x}$) are used to denote row vectors, while bold uppercase letters (e.g., $\mathbf{H}$) are used to denote matrices. Let n be a positive integer, and let $[n]$ denote the set $\{1, 2, \cdots, n\}$. For $\mathbf{x} \in \mathbb{F}_2^n$, the Hamming weight of $\mathbf{x}$ is denoted by $|\mathbf{x}|$, i.e.,

$$|\mathbf{x}| = |\{i \in [n] \mid x_i \neq 0\}|.$$

For arbitrary set $I \subseteq [n]$, we denote by $\mathbf{x}_I$ the vector whose coordinates are those of $\mathbf{x} = (x_i)_{i\in[n]}$ which are indexed by I (i.e., $\mathbf{x}_I = (x_i)_{i\in I}$). We also use $\mathbf{x}(i)$ to denote the i-th entry of a vector $\mathbf{x}$, and $\mathbf{H}(i, j)$ to denote the entry in row i and column j of a matrix $\mathbf{H}$. The concatenation of two vectors $\mathbf{x}$ and $\mathbf{y}$ is denoted by $(\mathbf{x}||\mathbf{y})$, and the transpose of a vector $\mathbf{x}$ is represented as $\mathbf{x}^\intercal$.

We use $s \leftarrow_\$ S$ if s is chosen uniformly at random from the finite set S. Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be two discrete probability distributions over the same probability space E. The statistical distance between them is defined as

$$\rho(\mathcal{D}_1, \mathcal{D}_2) = \frac{1}{2} \sum_{x\in E} |\mathcal{D}_1(x) - \mathcal{D}_2(x)|.$$

Coding Theory. An $[n, k]$ linear code $\mathcal{C}$ over the finite field $\mathbb{F}_2$ can be regarded as a k-dimensional subspace of the n-dimensional vector space $\mathbb{F}_2^n$. The vectors within the code $\mathcal{C}$ are referred to as codewords, with n representing the code's length, and k denoting its dimension. Similar to representing a linear space with a set of basis vectors, a linear code can also be represented by a set of codewords, and the matrix composed of the codewords is known as the generator matrix of the linear code $\mathcal{C}$.

Definition 1. *Let $\mathcal{C}$ be an $[n, k]$ linear code over the finite field $\mathbb{F}_2$. The generator matrix of the code $\mathcal{C}$ is a full-rank $k \times n$ matrix over $\mathbb{F}_2$. We refer to the matrix $\mathbf{G}$ as a generator matrix of the code $\mathcal{C}$ if there is*

$$\mathcal{C} = \{\mathbf{xG} \mid \mathbf{x} \in \mathbb{F}_2^k\}.$$

The code $\mathcal{C}$ can also be defined using a parity-check matrix, which is a full-rank $(n - k) \times n$ matrix over $\mathbb{F}_2$. We refer to the matrix $\mathbf{H}$ as a parity-check matrix of the code $\mathcal{C}$ if there is

$$\forall \mathbf{x} \in \mathcal{C} \Leftrightarrow \mathbf{xH}^\intercal = \mathbf{0}.$$

Certainly, the generator matrix and the parity-check matrix of code $\mathcal{C}$ are not unique. For any generator matrix $\mathbf{G}$ and parity-check matrix $\mathbf{H}$ of code $\mathcal{C}$, it holds that $\mathbf{G}\mathbf{H}^\intercal = \mathbf{0}$.

Definition 2. *Let $\mathcal{C}$ be an $[n,k]$ linear code over the finite field $\mathbb{F}_2$. Its dual code, denoted as $\mathcal{C}^\perp$, is defined as follows:*

$$\mathcal{C}^\perp = \{\mathbf{y} \in \mathbb{F}_2^n \mid \forall \mathbf{x} \in \mathcal{C}, \langle \mathbf{x}, \mathbf{y} \rangle = 0\}.$$

The code $\mathcal{C}^\perp$ is an $[n,r]$ linear code, with $r = n - k$. Its generator matrix is the parity-check matrix $\mathbf{H}$ of the code $\mathcal{C}$.

The information set of an $[n,k]$ linear code $\mathcal{C}$ is a set of k coordinate indices, denoted as $I \subseteq [n]$, which correspond to the indices of k linearly independent columns in the generator matrix $\mathbf{G}$. Its complement is the set of indices corresponding to the r $(r = n - k)$ linearly independent columns in the parity-check matrix $\mathbf{H}$ of the code $\mathcal{C}$. Given the parity-check matrix $\mathbf{H} \in \mathbb{F}_2^{r \times n}$ of code $\mathcal{C}$ and an information set I, for any vectors $\mathbf{s} \in \mathbb{F}_2^r$ and $\mathbf{x} \in \mathbb{F}_2^n$, there exists a unique vector $\mathbf{e} \in \mathbb{F}_2^n$ such that $\mathbf{e}\mathbf{H}^\intercal = \mathbf{s}$ and $\mathbf{x}_I = \mathbf{e}_I$.

Syndrome Decoding Problem. The syndrome decoding problem is one of the most commonly employed problems in code-based cryptographic schemes. Given a weight parameter w, the syndrome decoding problem requires finding a solution to a randomly generated system of linear equations over the finite field $\mathbb{F}_2$ with a Hamming weight of w. Definition 3 formalizes this as follows:

Definition 3. ***(Syndrome Decoding Problem)***

- ***Input:*** *A random matrix $\mathbf{H} \in \mathbb{F}_2^{r \times n}$, a syndrome $\mathbf{s} \in \mathbb{F}_2^r$, and a positive integer w*
- ***Output:*** *A vector $\mathbf{e} \in \mathbb{F}_2^n$ such that $\mathbf{e}\mathbf{H}^\intercal = \mathbf{s}$ and $|\mathbf{e}| = w$*

In this paper, we consider a variant of the syndrome decoding problem. We introduce the following:

Definition 4. ***(Decoding One Out of Many (DOOM))***

- ***Input:*** *A random matrix $\mathbf{H} \in \mathbb{F}_2^{r \times n}$, syndromes $\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N \in \mathbb{F}_2^r$, and a positive integer w*
- ***Output:*** *$(\mathbf{e}, i) \in \mathbb{F}_2^n \times [N]$ such that $\mathbf{e}\mathbf{H}^\intercal = \mathbf{s}_i$ and $|\mathbf{e}| = w$*

Prange Decoder. If $\frac{r}{2} \leq w \leq n - \frac{r}{2}$, then the above problem is relatively simple, as it can be effectively solved by calling the Prange algorithm [34], with the average number of calls being $2^r / \binom{r}{r/2}$. However, for values of w within the interval $[0, \frac{r}{2}) \cup (n - \frac{r}{2}, n]$, there is currently no known effective algorithm for solving the problem. In this context, we only consider small-weight vectors $\mathbf{e}$, as in the binary field, small-weight and large-weight are symmetric. Algorithm 1 represents the pseudocode of the Prange algorithm [34].

Algorithm 1 $\mathsf{Prange}(\mathbf{H}, \mathbf{s}, w)$

Parameters: (n, r, w)

Require: $\mathbf{H} \in \mathbb{F}_2^{r \times n}, \mathbf{s} \in \mathbb{F}_2^r$
Ensure: $\mathbf{e} \in \mathbb{F}_2^n$ with $\mathbf{eH}^\intercal = \mathbf{s}$
1: $t = \min\left(n - r, \max\left(0, w - \frac{r}{2}\right)\right)$
2: Randomly choose an information set $\mathcal{I} \subseteq [n]$ of size $n - r$
3: $\mathbf{x} \leftarrow_\$ \{\mathbf{x} \in \mathbb{F}_2^n \mid |\mathbf{x}| = |\mathbf{x}_\mathcal{I}| = t\}$
4: Randomly choose a permutation matrix $\mathbf{P} \in \mathbb{F}_2^{n \times n}$ such that $\mathcal{I}$ on the last $n - r$ coordinates
5: $(\mathbf{A}||\mathbf{B}) \leftarrow \mathbf{HP}$, $(\mathbf{0}||\mathbf{e}'') \leftarrow \mathbf{xP}$
6: $\mathbf{e} \leftarrow \left((\mathbf{s} - \mathbf{e}''\mathbf{B}^\intercal)(\mathbf{A}^{-1})^\intercal||\mathbf{e}''\right)\mathbf{P}^\intercal$
7: **return e**

Digital Signature. We now review the definitions of the signature scheme and the security model.

Definition 5. *A signature scheme $\mathcal{S}$ consists of three algorithms:* KeyGen*,* Sign*, and* Vrfy*, which are described as follows:*

- *Key Generation Algorithm* KeyGen*: A probabilistic algorithm that, given a security parameter λ, generates a pair of public and private keys $(\mathsf{pk}, \mathsf{sk})$;*
- *Signing Algorithm* Sign*: A probabilistic algorithm that takes as input a message $\mathbf{m} \in \{0,1\}^*$ and produces a signature $\sigma = \mathsf{Sign}_{\mathsf{sk}}(\mathbf{m})$;*
- *Verification Algorithm* Vrfy*: This algorithm takes a signature σ and a message $\mathbf{m}$ as input. It outputs 1 if the signature σ is valid for the message $\mathbf{m}$, and 0 otherwise. Specifically, it is required that $\mathsf{Vrfy}_{\mathsf{pk}}(\mathbf{m}, \mathsf{Sign}_{\mathsf{sk}}(\mathbf{m})) = 1$.*

The signature scheme proposed in this paper follows the "hash-and-sign" paradigm, where a message is first hashed and then signed. For proof purposes, hash function and signature algorithm are generally considered as random oracles. Existential unforgeability under an adaptive chosen message attack (EUF-CMA) is one of the common attack models against signature schemes. In this attack scenario, the adversary can query the hash oracle $q_\mathcal{H}$ times and the signature oracle $q_\Sigma < q_\mathcal{H}$ times, aiming to generate at least one message-signature pair that can pass verification. The formal definition of EUF-CMA security for a signature scheme is as follows.

Definition 6. (**EUF-CMA**) *Let $\mathcal{S}$ be a signature scheme. An adversary $\mathcal{A}$ is considered to be a $(\tau, \epsilon, q_\mathcal{H}, q_\Sigma)$-adversary against $\mathcal{S}$ in the* EUF-CMA *if, after at most $q_\mathcal{H}$ queries to the hash oracle $\mathcal{H}$, $q_\Sigma < q_\mathcal{H}$ queries to the signature oracle Σ, and running for at most time τ, it can output a valid forgery with probability at least ϵ. We define the* EUF-CMA *success probability against $\mathcal{S}$ as*

$$Succ_\mathcal{S}^{\text{EUF-CMA}}(\tau, q_\mathcal{H}, q_\Sigma) := \max_\mathcal{A}(\epsilon \mid \mathcal{A} \text{ is a } (\tau, \epsilon, q_\mathcal{H}, q_\Sigma)\text{-adversary}).$$

The signature scheme $\mathcal{S}$ is called $(\tau, q_\mathcal{H}, q_\Sigma)$-secure in EUF-CMA *if the above success probability is a negligible function of the security parameter λ.*

3 Code-Based Signature Scheme

In this section, we present the specific construction of the signature scheme proposed in this paper. A signature scheme $\mathcal{S}$ comprises three algorithms: KeyGen, Sign and Vrfy. The specific details are outlined as follows:

- $\mathsf{KeyGen}(1^\lambda)$: The algorithm is a key generation algorithm. Given the security parameter λ, it outputs the private key $\mathsf{sk} : (\mathbf{S}, \mathbf{H}_{\mathsf{sk}}, \mathbf{Q}, \mathbf{N})$ and the public key $\mathsf{pk} : (\mathbf{H}_{\mathsf{pk}}, \mathbf{M}, w, r_1, r_2, r, n_1, n_2, n, t, \lambda_0)$, where $n = n_1 + n_2$ and $r = r_1 + r_2$.
  1. Randomly choose three full-rank matrices $\mathbf{H}_1 \in \mathbb{F}_2^{r_1 \times n_1}$, $\mathbf{H}_2 \in \mathbb{F}_2^{r_1 \times n_2}$, and $\mathbf{H}_3 \in \mathbb{F}_2^{r_2 \times n_2}$;
  2. Randomly choose two matrices, denoted as $\mathbf{Q}_1 \in \mathbb{F}_2^{n \times n_1}$ and $\mathbf{Q}_2 \in \mathbb{F}_2^{n \times n_2}$, where $\mathbf{Q}_1$ is sparse and each column of $\mathbf{Q}_1$ has exactly t components of 1;
  3. Randomly choose two non-singular matrices $\mathbf{S} = (\mathbf{S}_1 || \mathbf{S}_2) \in \mathbb{F}_2^{r \times r}, \mathbf{N} \in \mathbb{F}_2^{r_1 \times r_1}$, where $\mathbf{S}_1 \in \mathbb{F}_2^{r \times r_1}, \mathbf{S}_2 \in \mathbb{F}_2^{r \times r_2}$;
  4. Let
$$\mathbf{H}_{\mathsf{sk}} = \begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{0} & \mathbf{H}_3 \end{pmatrix}, \ \mathbf{Q} = (\mathbf{Q}_1 || \mathbf{Q}_2), \ \mathbf{U} = \mathbf{Q}^{-1};$$
  5. Let
$$\mathsf{sk} \leftarrow (\mathbf{S}, \mathbf{H}_{\mathsf{sk}}, \mathbf{Q}, \mathbf{N}), \ \mathsf{pk} \leftarrow (\mathbf{H}_{\mathsf{pk}}, \mathbf{M}, w, r_1, r_2, r, n_1, n_2, n, t, \lambda_0),$$
  where $\mathbf{H}_{\mathsf{pk}} = \mathbf{S}\mathbf{H}_{\mathsf{sk}}\mathbf{U}, \mathbf{M} = \mathbf{S}_1\mathbf{N}$;
  6. Choose a cryptographic hash function $\mathsf{Hash} : \{0,1\}^* \to \mathbb{F}_2^{r_1}$.
- $\mathsf{Sign}_{\mathsf{sk}}(\mathbf{m})$: Given an arbitrary message $\mathbf{m}$ and private key $\mathsf{sk} = (\mathbf{S}, \mathbf{H}_{\mathsf{sk}}, \mathbf{Q}, \mathbf{N})$, the output is the signature σ corresponding to the message $\mathbf{m}$.
  1. $\mathbf{r} \leftarrow_{\$} \{0,1\}^{\lambda_0}$;
  2. $\mathbf{s}_1 \leftarrow \mathsf{Hash}(\mathbf{m}, \mathbf{r})$;
  3. $\mathbf{e}_1 \leftarrow \mathsf{Prange}(\mathbf{H}_1, \mathbf{s}_1\mathbf{N}^\intercal, \frac{r_1}{2})$, s.t., $\mathbf{e}_1\mathbf{H}_1^\intercal = \mathbf{s}_1\mathbf{N}^\intercal$;
  4. Compute $\mathbf{e} = \mathbf{e}_1\mathbf{Q}_1^\intercal$. If $|\mathbf{e}| = w$, continue; else return to the step 3;
  5. $\sigma \leftarrow (\mathbf{e}, \mathbf{r})$.
- $\mathsf{Vrfy}_{\mathsf{pk}}(\mathbf{m}, \sigma)$: Given the message $\mathbf{m}$, signature $\sigma = (\mathbf{e}', \mathbf{r})$, and public key pk, the output is a bit $b \in \{0,1\}$.
  1. Check whether $|\mathbf{e}'| = w$ and $\mathbf{r} \in \{0,1\}^{\lambda_0}$. If not, return $b = 0$;
  2. $\mathbf{s}_1' \leftarrow \mathsf{Hash}(\mathbf{m}, \mathbf{r})$;
  3. Check whether $\mathbf{H}_{\mathsf{pk}}\mathbf{e}'^\intercal = \mathbf{M}\mathbf{s}_1'^\intercal$. If true, return $b = 1$, else return $b = 0$.

Correctness. If the signature $\sigma = (\mathbf{e}, \mathbf{r})$ is generated strictly according to the signing process by the private key holder, then the corresponding message-signature pair $(\mathbf{m}, \sigma)$ will always pass the verification process.

$$\begin{aligned} \mathbf{H}_{\mathsf{pk}}\mathbf{e}^\intercal &= \mathbf{S}\mathbf{H}_{\mathsf{sk}}\mathbf{U}\mathbf{Q}_1\mathbf{e}_1^\intercal \\ &= \mathbf{S}\mathbf{H}_{\mathsf{sk}}\mathbf{U}(\mathbf{Q}_1||\mathbf{Q}_2)(\mathbf{e}_1||\mathbf{0})^\intercal \\ &= \mathbf{S}\mathbf{H}_{\mathsf{sk}}(\mathbf{e}_1||\mathbf{0})^\intercal \\ &= (\mathbf{S}_1||\mathbf{S}_2)(\mathbf{e}_1\mathbf{H}_1^\intercal||\mathbf{0})^\intercal \\ &= (\mathbf{S}_1||\mathbf{S}_2)(\mathbf{s}_1\mathbf{N}^\intercal||\mathbf{0})^\intercal \\ &= \mathbf{S}_1\mathbf{N}\mathbf{s}_1^\intercal = \mathbf{M}\mathbf{s}_1^\intercal. \end{aligned} \tag{1}$$

Theorem 1 guarantees that the private key holder can generate a valid and legitimate signature for each message in polynomial time.

Theorem 1. *For any given message* $\mathbf{m}$*, the probability of successfully generating a signature in strict accordance with scheme* $\mathcal{S}$ *is*

$$\frac{1}{2^{r_1}}\sum_{j=1}^{r_1}\binom{r_1}{j}\binom{n}{w}p_j^w(1-p_j)^{n-w},$$

where

$$p_j=\sum_{i=1}^{n_1/2}\binom{n_1}{2i-1}(\frac{t}{n}\cdot\frac{j}{n_1})^{2i-1}(1-\frac{t}{n}\cdot\frac{j}{n_1})^{n_1-2i+1}.$$

Here we omit the detailed proof. For the two sets of parameters given in Section 5, the probability of successfully generating a signature is 0.0157 and 0.0114, respectively. This means that on average, the Prange algorithm needs to be called 64 times and 88 times, respectively.

4 Security Analysis

In this section, we provide the security proof of the signature scheme $\mathcal{S}$. Under the conditions where the syndrome decoding problem and the distinguishing problem are hard, we demonstrate that scheme $\mathcal{S}$ achieves existential unforgeability under an adaptive chosen message attack (EUF-CMA).

4.1 Distinguishing Problem

The syndrome decoding problem is one of the most famous hard problems in coding theory and has been proven to be an NP-complete problem [10]. In this subsection, we prove that the public key matrix $\mathbf{H}_{\mathsf{pk}}$ constructed in this paper is statistically indistinguishable from a random matrix.

Let $\mathcal{H}_{\mathsf{pk}}$ denote the set of all public key matrices $\mathbf{H}_{\mathsf{pk}}$ constructed as described in Section 3. A distinguisher $\mathcal{D}$ for the $\mathcal{H}_{\mathsf{pk}}$ codes is a probabilistic polynomial-time algorithm. We define its advantage as follows:

$$\begin{aligned}Adv^{\mathcal{H}_{\mathsf{pk}}}(\mathcal{D})=\Big|&\mathbb{P}[\mathbf{H}\leftarrow_{\$}\mathcal{H}_{\mathsf{pk}}\,\mathrm{Codes}(n,k):\mathcal{D}(\mathbf{H})=1]\\&-\mathbb{P}[\mathbf{H}\leftarrow_{\$}\mathrm{Random\,Binary\,Codes}(n,k):\mathcal{D}(\mathbf{H})=1]\Big|.\end{aligned}$$

If the advantage is negligible for any distinguisher $\mathcal{D}$, then the $\mathcal{H}_{\mathsf{pk}}$ codes are indistinguishable from random linear codes.

Definition 7. *The* $\mathcal{H}_{\mathsf{pk}}$ *Codes Distinguishing Problem is considered to be* (τ,ϵ)*-hard if for any distinguisher* $\mathcal{D}$ *running in time at most* τ *we have*

$$Adv^{\mathcal{H}_{\mathsf{pk}}}(\mathcal{D})\leq\epsilon.$$

Our approach to constructing public and secret keys differs slightly from other cryptographic schemes based on the syndrome decoding problem. In traditional cryptographic schemes, the secret keys consist of randomly chosen non-singular matrix $\mathbf{S}$ and permutation matrix $\mathbf{P}$. Here, we disclose the subspace S_v formed by the columns of matrix $\mathbf{S}_1$ as part of the public key. Intuitively, this may seem to reduce the security of the scheme. However, we argue that this still presents a challenging task. Left-multiplying matrix $\mathbf{H}_{\mathsf{sk}} \in \mathbb{F}_2^{r\times n}$ by a randomly selected non-singular matrix $\mathbf{S} \in \mathbb{F}_2^{r\times r}$ is equivalent to randomly selecting r codewords from the code $\mathcal{C}$ with $\mathbf{H}_{\mathsf{sk}}$ as the generator matrix to form a new generator matrix $\mathbf{H}_{\mathsf{pk}}$, resulting in $\binom{2^r}{r}$ possibilities. In our scheme, even when the subspace S_v is disclosed, there are still $\binom{2^{r_1}}{r_1}\binom{2^r-2^{r_1}}{r_2}$ possibilities for generating the new matrix $\mathbf{H}_{\mathsf{pk}}$ by left-multiplying $\mathbf{H}_{\mathsf{sk}}$ by matrix $\mathbf{S}$, which is an exceedingly large number. Therefore, disclosing the subspace S_v will not significantly reduce security.

Furthermore, in traditional cryptographic schemes, the matrix $\mathbf{H}_{\mathsf{sk}}$ is also right-multiplied by a random permutation matrix $\mathbf{P}$ (to maintain the weight distribution of error vectors), which simply shuffles the columns of matrix $\mathbf{H}_{\mathsf{sk}}$ without altering the weight distribution of codewords in code $\mathcal{C}$. However, in our scheme, we right-multiply matrix $\mathbf{H}_{\mathsf{sk}}$ by an invertible matrix $\mathbf{U}$ with uniformly distributed elements 0 and 1, which makes each column of the final generated matrix $\mathbf{H}_{\mathsf{pk}}$ a random vector (as matrix $\mathbf{H}_{\mathsf{sk}}$ has n_2 columns of random vectors). Consequently, the weight distribution of the code $\mathcal{C}_{\mathsf{pk}}$ generated by matrix $\mathbf{H}_{\mathsf{pk}}$ is consistent with the weight distribution of a code generated by a random matrix of the same size, making the code distinguishing problem hard. In summary, the construction method of our scheme offers security no less than traditional methods and perhaps even greater security, as it alters the weight distribution of the code to approximate or equal the weight distribution of random codes.

Theorem 2 proves the indistinguishability between the public key matrix $\mathbf{H}_{\mathsf{pk}}$ and a random matrix. Prior to this, we introduce an important lemma, which can be found in [18].

Lemma 1. *[18] Let $\mathbf{H} \in \mathbb{F}_2^{r\times n}$ $(r \leq n)$ be a random full-rank matrix, and $\mathbf{s} \in \mathbb{F}_2^r$ be a random vector. Then, for any non-zero vector $\mathbf{x} \in \mathbb{F}_2^n$, it holds that*

$$\mathbb{P}(\mathbf{x}\mathbf{H}^\intercal = \mathbf{s}) = \frac{1}{2^r}. \tag{2}$$

Theorem 2. *Let $\mathbf{H}_1 \in \mathbb{F}_2^{r\times m}(r < m)$ be a random matrix and $\mathbf{U} \in \mathbb{F}_2^{m\times n}(m \leq n)$ be a full-rank matrix with uniformly distributed elements 0 and 1. Then $\mathbf{H}_2 = \mathbf{H}_1\mathbf{U}$ is statistically indistinguishable from a random matrix.*

Proof. Let $\mathbf{u}_i$ be the i-th column of matrix $\mathbf{U}$. According to Lemma 1, $\mathbf{H}_1\mathbf{u}_i$ is statistically indistinguishable from a vector randomly chosen from the vector space $\mathbb{F}_2^r$. Moreover, since matrix $\mathbf{U}$ is full-rank, matrix $\mathbf{H}_2$ is statistically indistinguishable from a random matrix.

According to Theorem 2, the public key matrix $\mathbf{H}_{\mathsf{pk}}$ is statistically indistinguishable from a random matrix. This property is highly desirable. To the best of our knowledge, this paper is the first to prove the statistical indistinguishability between the provided public key matrix $\mathbf{H}_{\mathsf{pk}}$ and a random matrix.

4.2 Domain Sampling with Uniform Output

Due to the construction of the signature scheme in this paper, the decodable syndromes collectively form a r_1-dimensional subspace of the r-dimensional vector space $\mathbb{F}_2^r$. Therefore, the syndrome distinguishing problem is confined to this subspace. Randomly select r_1 linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{r_1} \in \mathbb{F}_2^r$ to form a basis for the subspace S_v as well as the matrix $\mathbf{S}_1 \in \mathbb{F}_2^{r \times r_1}$. Following the method outlined in Section 3, the public-private key pair is constructed as follows:

$$\mathsf{sk} : (\mathbf{S}, \mathbf{H}_{\mathsf{sk}}, \mathbf{Q}, \mathbf{N}),\ \ \mathsf{pk} : (\mathbf{H}_{\mathsf{pk}} = \mathbf{S}\mathbf{H}_{\mathsf{sk}}\mathbf{U}, \mathbf{M} = \mathbf{S}_1\mathbf{N}),$$

where

$$\mathbf{S} = (\mathbf{S}_1||\mathbf{S}_2),\ \ \mathbf{H}_{\mathsf{sk}} = \begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{0} & \mathbf{H}_3 \end{pmatrix},\ \ \mathbf{Q} = (\mathbf{Q}_1||\mathbf{Q}_2),\ \ \mathbf{U} = \mathbf{Q}^{-1}.$$

We declare that the syndromes generated by the matrix $\mathbf{H}_{\mathsf{pk}}$ are indistinguishable from those randomly selected from the subspace S_v when the parameters are appropriately chosen. Theorem 3 provides the formal statement.

Theorem 3. *Let* $\mathbf{H}$ *be a* $r \times n$ *matrix over* $\mathbb{F}_2$*. Let* $\mathbf{H}^{n,w}$ *be the set of all vectors* $\mathbf{e} \in \mathbb{F}_2^n$ *such that* $\mathbf{e}\mathbf{H}^\intercal \in S_v$ *and* $\mathbf{e}$ *has weight* w*. Let* $\mathcal{D}_{\mathbf{H}^{n,w}}$ *be the distribution of the syndromes* $\mathbf{e}\mathbf{H}^\intercal$*, where* $\mathbf{e}$ *is chosen uniformly at random from the set* $\mathbf{H}^{n,w}$*. Let* $\mathcal{U}$ *be the uniform distribution on the syndrome subspace* S_v*, then*

$$\mathbb{E}_{\mathbf{H}_{\mathsf{pk}}}\Big(\rho(\mathcal{D}_{\mathbf{H}_{\mathsf{pk}}^{n,w}}, \mathcal{U})\Big) \leq \frac{1}{2}\sqrt{\theta},$$

where

$$\theta = \frac{2^r - 2^{r_2}}{\binom{n}{w}}.$$

Here we omit the detailed proof.

4.3 EUF-CMA Security

This subsection establishes the security of the signature scheme. We prove that the signature scheme $\mathcal{S}$ achieves EUF-CMA security in the random oracle model when the syndrome decoding problem and the distinguishing problem are hard.

Theorem 4. *Let* $\mathcal{H}_{\mathsf{pk}}$ Codes Distinguishing Problem *and* DOOM Problem *be* $(\tau_{CD}, \epsilon_{CD})$ *and* $(\tau_{DOOM}, \epsilon_{DOOM})$*-secure, respectively. Then the signature scheme* $\mathcal{S}$ *is* $(\tau, \epsilon, q_{\mathcal{H}}, q_\Sigma)$-EUF-CMA *secure in the random oracle model. Here,* τ *and* ϵ *are given as*

$$\epsilon \leq \frac{q_{\mathcal{H}} \cdot 2^{\lambda_0}}{r_1 \cdot (2^{\lambda_0} - q_{\mathcal{H}})}\epsilon_{DOOM} + \epsilon_{CD} + \frac{q_\Sigma}{2}\sqrt{\theta}, \tag{3}$$

and

$$\tau \geq \tau_{DOOM} - (q_{\mathcal{H}} + q_\Sigma)O\big(r_2^3\big),$$

where $O\big(r_2^3\big)$ *is the syndrome computation time of an* $[n, n-r]$ *linear code.*

Proof. Let $\mathcal{A}$ be a $(\tau, \epsilon, q_{\mathcal{H}}, q_{\Sigma})$-adversary in the EUF-CMA model against the signature scheme $\mathcal{S}$ and let $(\mathbf{H}_0, \mathbf{s}_1, \cdots, \mathbf{s}_{q_{\mathcal{H}}})$ be drawn uniformly at random among all instances of DOOM problem for parameters $n, r, q_{\mathcal{H}}$, and w. We explicitly state that the syndromes $\mathbf{s}_j$ are independent and random vectors in the vector space $\mathbb{F}_2^k$. Without loss of generality, let $\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_{r_1}$ be r_1 linearly independent vectors forming a basis for the vector space S_v, and let them form the matrix $\mathbf{S}_1 \in \mathbb{F}_2^{r \times r_1}$. $q_{\mathcal{H}} - r_1$ vectors $\mathbf{u}_{r_1+1}, \cdots, \mathbf{u}_{q_{\mathcal{H}}}$ are uniformly randomly selected from the vector space $\mathbb{F}_2^{r_1}$, and new syndromes $\mathbf{s}_j$ are computed as $\mathbf{s}_j = \mathbf{u}_j \mathbf{S}_1^{\intercal}$. We emphasize that these $q_{\mathcal{H}}$ vectors $\mathbf{s}_1, \cdots, \mathbf{s}_{q_{\mathcal{H}}}$ are independent and random vectors in the vector space S_v. Let **Game 0** (Fig. 1) be the standard EUF-CMA game adapted to this scheme. We denote $\mathbb{P}(S_i)$ as the probability of adversary $\mathcal{A}$ winning Game i. We have

$$\mathbb{P}(S_0) = Succ_{\mathcal{S}}^{\text{EUF-CMA}}(\mathcal{A}). \tag{4}$$

Input: An adversary $\mathcal{A}$
$(\mathbf{S}, \mathbf{H}_{\mathsf{sk}}, \mathbf{Q}, \mathbf{H}_{\mathsf{pk}}, S_v) \leftarrow \mathsf{KeyGen}(1^{\lambda})$
Set the random oracles $\mathcal{H}$ and Σ
$(\mathbf{m}', \mathbf{e}', \mathbf{r}') \leftarrow \mathcal{A}^{\mathcal{H}, \Sigma}(\mathbf{H}_{\mathsf{pk}}, S_v)$
If $\mathbf{e}'\mathbf{H}_{\mathsf{pk}}^{\intercal} = \mathcal{H}(\mathbf{m}', \mathbf{r}') \in S_v$, $|\mathbf{e}'| = w$ and Σ did not provide $\mathbf{e}'$ **then**
    $\mathcal{A}$ wins the game
else
    $\mathcal{A}$ loses the game

Fig. 1. Game 0: EUF-CMA game for $\mathcal{S}$

In order to prove that we use three lists, $\Lambda_{\mathcal{H}}, \Lambda_{\Sigma}$ and Λ, the oracles $\mathcal{H}$ and Σ maintain lists $\Lambda_{\mathcal{H}}$ and Λ_{Σ} respectively of queries with the corresponding output values. Each simulated oracle has access to these lists, since the oracles are controlled by the challenger. List $\Lambda_{\mathcal{H}}$ stores the syndrome $\mathbf{s}$ (or hash value $\mathbf{u}$ such that $\mathbf{s} = \mathbf{u}\mathbf{S}_1^{\intercal}$) and the corresponding error vector $\mathbf{e}$, i.e., $\Lambda_{\mathcal{H}}(\mathbf{m}, \mathbf{r}) = (\mathbf{s}, \mathbf{e})$ ($\Lambda_{\mathcal{H}}(\mathbf{m}, \mathbf{r}) = (\mathbf{u}, \mathbf{e})$) for any message $\mathbf{m}$ and any salt $\mathbf{r}$. List Λ_{Σ} stores the signature σ, that is, $\Lambda_{\Sigma}(\mathbf{m}) = (\mathbf{e}, \mathbf{r})$ for any message $\mathbf{m}$. List Λ stores the salt $\mathbf{r}$ corresponding to the message. If there is no value associated with an entry in a list, we denote the output by $\perp$.

Game 1. In this game, the challenger substitutes simulation $\mathcal{H}'$ (Fig. 2) for the hash oracle $\mathcal{H}$. For any query $(\mathbf{m}, \mathbf{r})$, we observe two scenarios: either $\mathbf{r} \neq \Lambda(\mathbf{m})$ or $\mathbf{r} = \Lambda(\mathbf{m})$. When $\mathbf{r} \neq \Lambda(\mathbf{m})$, the behavior of $\mathcal{H}'$ aligns with that of $\mathcal{H}$. However, when $\mathbf{r} = \Lambda(\mathbf{m})$, $\mathcal{H}'$ generates a decodable syndrome $\mathbf{s}$ and stores its decoding vector $\mathbf{e}$ in list $\Lambda_{\mathcal{H}}$: it chooses a vector $\mathbf{e}$ uniformly at random from the set $\mathbf{H}_{\mathsf{pk}}^{n,w}$, computes the syndrome $\mathbf{s} = \mathbf{e}\mathbf{H}_{\mathsf{pk}}^{\intercal}$ as the output, and finally stores $(\mathbf{s}, \mathbf{e})$ in list $\Lambda_{\mathcal{H}}$. At the end of the simulation, the oracle $\mathcal{H}'$ produces a total of $q_{\mathcal{H}} + q_{\Sigma}$ syndromes, some of which originate from the modified part of the oracle $\mathcal{H}$. Let X be a random variable representing the number of syndromes generated by the

modified part of the oracle $\mathcal{H}$. Hence, the probability of adversary $\mathcal{A}$ winning Game 1 is as follows:

$$\begin{aligned}\mathbb{P}(S_1) &= \mathbb{P}[(S_1 \cap (X=0)) \cup (S_1 \cap (X>0))] \\ &= \mathbb{P}[S_1 \cap (X=0)] + \mathbb{P}[S_1 \cap (X>0)].\end{aligned} \tag{5}$$

It is clear that $\mathbb{P}[S_1 \cap (X=0)]$ represents the probability that the adversary $\mathcal{A}$ wins Game 1 when all syndromes are generated by the random oracle $\mathcal{H}$, which is exactly the probability that the adversary $\mathcal{A}$ wins Game 0. Therefore, we have

$$\mathbb{P}(S_0) \leq \mathbb{P}(S_1). \tag{6}$$

Input: A pair $(\mathbf{m}, \mathbf{r})$
Output: A syndrome $\mathbf{s}$
If $\Lambda(\mathbf{m}) = \perp$ **then**
  $\Lambda(\mathbf{m}) \leftarrow_{\$} \{0,1\}^{\lambda_0}$
$(\mathbf{s}, \mathbf{e}) \leftarrow \Lambda_{\mathcal{H}}(\mathbf{m}, \mathbf{r})$
If $\mathbf{r} \neq \Lambda(\mathbf{m})$ **then**
  If $\mathbf{s} = \perp$ **then**
    $\mathbf{s} \leftarrow_{\$} S_v$, $\Lambda_{\mathcal{H}}(\mathbf{m}, \mathbf{r}) \leftarrow (\mathbf{s}, \perp)$
else
  If $\mathbf{s} = \perp$ **then**
    $\mathbf{e} \leftarrow_{\$} \mathbf{H}_{\mathsf{pk}}^{n,w}$, $\mathbf{s} \leftarrow \mathbf{e}\mathbf{H}_{\mathsf{pk}}^{\intercal}$, $\Lambda_{\mathcal{H}}(\mathbf{m}, \mathbf{r}) \leftarrow (\mathbf{s}, \mathbf{e})$
Return $\mathcal{H}(\mathbf{m}, \mathbf{r}) = \mathbf{s}$

Fig. 2. $\mathcal{H}'$: simulation of $\mathcal{H}$ (**Game 1**)

Game 2. In this game, the challenger substitutes simulation $\mathcal{H}''$ (Fig. 3) for simulation $\mathcal{H}'$. The only difference between $\mathcal{H}''$ and $\mathcal{H}'$ is the way the syndrome is generated in the Line 6. The oracle $\mathcal{H}''$ takes the elements of the set $\{\mathbf{s}_1, \cdots, \mathbf{s}_{q_{\mathcal{H}}}\}$ one by one as the output of the hash. As emphasized earlier, $\mathbf{s}_j$ are independent and random vectors in vector space S_v, and the simulation remains unchanged, hence

$$\mathbb{P}(S_2) = \mathbb{P}(S_1). \tag{7}$$

Game 3. In this game, the challenger substitutes simulation Σ' (Fig. 4) for the signature oracle Σ. Σ' does not need the private key to generate signatures since it queries $\mathcal{H}''$ on $(\mathbf{m}, \Lambda(\mathbf{m}))$, and the list $\Lambda_{\mathcal{H}}$ stores the corresponding decoding values and syndromes. The Line 5 in Σ' is in order to make the signature of the two different queries for the same message does not generate the same signature. The difference between Game 3 and Game 2 is the way the syndrome is generated in signature query, Game 2 chooses the syndrome from the set $\{\mathbf{s}_1, \cdots, \mathbf{s}_{q_{\mathcal{H}}}\}$, while Game 3 computes $\mathbf{e}\mathbf{H}_{\mathsf{pk}}^{\intercal}$ as output, where $\mathbf{e}$ is uniformly at random chosen from the set $\mathbf{H}_{\mathsf{pk}}^{n,w}$. Leveraging Proposition 4 and Lemma 4 from reference [18], we have

$$|\mathbb{P}(S_2) - \mathbb{P}(S_3)| \leq \frac{q_{\Sigma}}{2}\sqrt{\theta}, \tag{8}$$

where θ given in Theorem 3.

Input: A pair $(\mathbf{m}, \mathbf{r})$
Output: A syndrome $\mathbf{s}$
If $\Lambda(\mathbf{m}) = \perp$ **then**
  $\Lambda(\mathbf{m}) \leftarrow_{\$} \{0,1\}^{\lambda_0}$
$(\mathbf{s}, \mathbf{e}) \leftarrow \Lambda_{\mathcal{H}}(\mathbf{m}, \mathbf{r})$
If $\mathbf{r} \neq \Lambda(\mathbf{m})$ **then**
  If $\mathbf{s} = \perp$ **then**
    $\mathbf{s} \leftarrow \mathsf{Next}.[\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_{q_{\mathcal{H}}}]$, $\Lambda_{\mathcal{H}}(\mathbf{m}, \mathbf{r}) \leftarrow (\mathbf{s}, \perp)$
else
  If $\mathbf{s} = \perp$ then
    $\mathbf{e} \leftarrow_{\$} \mathbf{H}_{\mathsf{pk}}^{n,w}$, $\mathbf{s} \leftarrow \mathbf{e}\mathbf{H}_{\mathsf{pk}}^{\intercal}$, $\Lambda_{\mathcal{H}}(\mathbf{m}, \mathbf{r}) \leftarrow (\mathbf{s}, \mathbf{e})$
Return $\mathcal{H}(\mathbf{m}, \mathbf{r}) = \mathbf{s}$

Fig. 3. $\mathcal{H}''$: simulation of $\mathcal{H}$ (**Game 2**)

Input: A message $\mathbf{m}$
Output: A signature $(\mathbf{e}, \mathbf{r})$
if $\Lambda(\mathbf{m}) = \perp$ **then**
  $\Lambda(\mathbf{m}) \leftarrow_{\$} \{0,1\}^{\lambda_0}$
$\mathcal{H}''(\mathbf{m}, \Lambda(\mathbf{m}))$
$(\mathbf{s}, \mathbf{e}) \leftarrow \Lambda_{\mathcal{H}}(\mathbf{m}, \Lambda(\mathbf{m}))$
$\Lambda(\mathbf{m}) \leftarrow \perp$
return $\Lambda_{\Sigma}(\mathbf{m}) = (\mathbf{e}, \mathbf{r})$

Fig. 4. Σ': simulation of Σ (**Game 3**)

Game 4. In this game, the challenger replaces the public key $\mathbf{H}_{\mathsf{pk}}$ with matrix $\mathbf{H}_0$ of the instance $(\mathbf{H}_0, \mathbf{s}_1, \cdots, \mathbf{s}_{q_{\mathcal{H}}})$ of the DOOM problem. To establish a relationship between Game 4 and Game 3, we construct a distinguisher $\mathcal{D}$ (Fig. 5). If $\mathbf{H}$ is generated by the key generation algorithm KeyGen, $\mathcal{D}$ proceeds as Game 3 and therefore

$$\mathbb{P}[\mathbf{H} \leftarrow_{\$} \mathcal{H}_{\mathsf{pk}} \text{ Codes}(n,k) : \mathcal{D}(\mathbf{H}) = 1] = \mathbb{P}(S_3).$$

If $\mathbf{H}$ is matrix $\mathbf{H}_0$ of the instance $(\mathbf{H}_0, \mathbf{s}_1, \cdots, \mathbf{s}_{q_{\mathcal{H}}})$ of the DOOM problem, $\mathcal{D}$ proceeds as Game 4 and therefore

$$\mathbb{P}[\mathbf{H} \leftarrow \mathbf{H}_0 : \mathcal{D}(\mathbf{H}) = 1] = \mathbb{P}(S_4).$$

Then we have

$$Adv^{CD}(\mathcal{D}) = |\mathbb{P}(S_3) - \mathbb{P}(S_4)|,$$

and since the $\mathcal{H}_{\mathsf{pk}}$ codes distinguishing problem is $(\tau_{CD}, \epsilon_{CD})$-hard, we have

$$|\mathbb{P}(S_3) - \mathbb{P}(S_4)| \leq \epsilon_{CD}. \tag{9}$$

We assume the adversary $\mathcal{A}$ outputs a valid signature $(\mathbf{e}', \mathbf{r}')$ for the message $\mathbf{m}'$. Let p be the probability that $\mathbf{e}'\mathbf{H}_0^{\intercal} \in \{\mathbf{s}_1, \cdots, \mathbf{s}_{r_1}\}$, then we have

$$p \geq (1 - \frac{1}{2^{\lambda_0}})^{q_{\mathcal{H}}} \cdot \frac{r_1}{q_{\mathcal{H}}} \geq (1 - \frac{q_{\mathcal{H}}}{2^{\lambda_0}}) \cdot \frac{r_1}{q_{\mathcal{H}}}, \tag{10}$$

Input: A parity-check matrix $\mathbf{H}$
Output: A bit b
Set the oracles $\mathcal{H}''$ and Σ'
$(\mathbf{m}', \mathbf{e}', \mathbf{r}') \leftarrow \mathcal{A}^{\mathcal{H}'', \Sigma'}(\mathbf{H}, S_v)$
if $\mathbf{e}'\mathbf{H}^\intercal = \mathcal{H}''(\mathbf{m}', \mathbf{r}') \in S_v$, $|\mathbf{e}'| = w$ and Σ' did not provide $\mathbf{e}'$ **then**
    output 1
else
    output 0

Fig. 5. $\mathcal{D}(\mathbf{H})$ **(Game 4)**

that is,

$$Succ^{DOOM}(\mathcal{A}) = p \cdot \mathbb{P}(S_4) \geq (1 - \frac{q_{\mathcal{H}}}{2^{\lambda_0}}) \cdot \frac{r_1}{q_{\mathcal{H}}} \mathbb{P}(S_4). \tag{11}$$

Considering the DOOM problem is $(\tau_{DOOM}, \epsilon_{DOOM})$-hard, hence, we have

$$Succ^{DOOM}(\mathcal{A}) \leq \epsilon_{DOOM}. \tag{12}$$

Combining (4), (6), (7), (8), (9), (11) and (12), we have

$$Succ_{\mathcal{S}}^{\text{EUF}-\text{CMA}}(\mathcal{A}) \leq \frac{q_{\mathcal{H}} \cdot 2^{\lambda_0}}{r_1 \cdot (2^{\lambda_0} - q_{\mathcal{H}})} \epsilon_{DOOM} + \epsilon_{CD} + \frac{q_\Sigma}{2} \sqrt{\theta}. \tag{13}$$

The running time to successfully solve the instance $(\mathbf{H}_0, \mathbf{s}_1, \cdots, \mathbf{s}_{q_{\mathcal{H}}})$ of the DOOM problem is the running time of the adversary $\mathcal{A}$ and the time to compute $q_{\mathcal{H}} + q_\Sigma$ hash values, which is at most the time to compute $q_{\mathcal{H}} + q_\Sigma$ syndromes $\mathbf{e}\mathbf{H}_0^\intercal$, where $\mathbf{e}$ is uniformly at random chosen from the set $\mathbf{H}_0^{n,w}$. This gives the formula for τ.

5 Parameters and Comparison

In this section, we present two sets of parameters for our scheme and compare them with other schemes, including those that have been finalists in the NIST Post-Quantum Cryptography (PQC) competition, as well as recent submissions to NIST of several code-based schemes. Our scheme exhibits significant advantages in terms of signature size. Additionally, we validate through simulations our previous assertion that the inverse matrix $\mathbf{U}$ of the reversible matrix $\mathbf{Q}$ has uniformly distributed elements 0 and 1.

Parameter Selection. Based on the equation (3) in Theorem 4, it can be observed that the security of the scheme is related to the syndrome decoding problem and the distinguishing problem. From the previous analysis, the public key matrix $\mathbf{H}_{\mathsf{pk}}$ is indistinguishable from a random matrix. Regarding the syndrome decoding problem, we use the decoding complexity of the Prange algorithm [34] and the BJMM algorithm [6] as references. For a security level of λ, we consider $\lambda_0 = \lambda$ to be sufficient.

Table 1. Public key and signature sizes of our signature schemes $\mathcal{S}_1$ and $\mathcal{S}_2$

λ	Scheme	$(n_1, r_1, n_2, r_2, t, w)$	pk(MB)	σ(byte)
128	$\mathcal{S}_1$	(1000,590,1700,950,2,478)	0.322	161
	$\mathcal{S}_2$	(1000,900,2800,1900,3,967)	0.635	141

Table 1 illustrates the two sets of parameters we have chosen (schemes $\mathcal{S}_1$ and $\mathcal{S}_2$) along with their corresponding public key and signature sizes. Scheme $\mathcal{S}_1$ features a smaller public key size, while scheme $\mathcal{S}_2$ offers a smaller signature size. In our scheme, the public key comprises matrices $\mathbf{M} \in \mathbb{F}_2^{r \times r_1}$ and $\mathbf{R} \in \mathbb{F}_2^{r \times (n-r)}$, where $\mathbf{H}_{\mathsf{pk}} = (\mathbf{I}_r || \mathbf{R})$. Under the public key $(\mathbf{M}, \mathbf{R})$, a valid signature $\sigma = (\mathbf{e}, \mathbf{r}) \in \mathbb{F}_2^{n-r} \times \mathbb{F}_2^{\lambda_0}$ for a message $\mathbf{m}$ satisfies

$$|\mathbf{e}| + |\mathsf{Hash}(\mathbf{m}, \mathbf{r})\mathbf{M}^{\mathsf{T}} - \mathbf{e}\mathbf{R}^{\mathsf{T}}| = w.$$

Therefore, the size of the signature σ is given by: $|\sigma| = n - r + \lambda_0$.

Randomness Testing. The inverse transformation of a matrix is non-linear. Considering the construction of matrix $\mathbf{Q}$, we assert that the elements 0 and 1 in matrix $\mathbf{U}$ are uniformly distributed. We have reason to believe that as long as the ratio of the number of element 0 to the number of element 1 in matrix $\mathbf{U}$ approaches 1 : 1, our assertion is reasonable. Furthermore, if the elements 0 and 1 in matrix $\mathbf{U}$ are uniformly distributed, then the code generated by matrix $\mathbf{H}_{\mathsf{sk}}\mathbf{U}$ has the same weight distribution as a random code. That is, matrix $\mathbf{H}_{\mathsf{pk}} = \mathbf{S}\mathbf{H}_{\mathsf{sk}}\mathbf{U}$ is indistinguishable from a random matrix. Table 2 shows the ratio between the number of elements 0 and 1 in matrix $\mathbf{U}$ generated under two sets of parameters. The listed results represent the averages obtained after 10,000 simulation experiments. We can observe that the ratio between the number of elements 0 and 1 is close to 1 : 1. Therefore, matrix $\mathbf{H}_{\mathsf{pk}}$ is statistically indistinguishable from a random matrix.

Table 2. Randomness testing

Scheme	(1000,590,1700,950,2,478)	(1000,900,2800,1900,3,967)
0 : 1	0.999983936535054 : 1	1.000002220057866 : 1

Comparison with other code-based signature schemes. There are currently two mainstream approaches to constructing signatures. One approach is to use the Fiat-Shamir transformation to convert an identification scheme into a signature scheme. Such schemes have small public keys and high efficiency but a long signature. The another approach is to use trapdoors to construct signature schemes, which results in shorter signatures, but larger public keys and lower efficiency. Our scheme is also designed based on the "hash-and-sign" paradigm. Table 3 presents a comparison of our scheme with two code-based schemes

recently submitted to NIST. It is easy to see that our scheme has significant advantages in both public key and signature sizes. The public key size is only 0.161 times that of Enhanced pqsigRM [15] or 0.092 times that of Wave [4], and the signature size is only 0.137 times that of Enhanced pqsigRM or 0.172 times that of Wave.

Table 3. Public key and signature sizes of our signature schemes compared with other code-based signature schemes

λ	$\mathcal{S}_1$		$\mathcal{S}_2$		Enhanced pqsigRM [15]		Wave [4]	
	pk(MB)	σ(byte)	pk(MB)	σ(byte)	pk(MB)	σ(byte)	pk(MB)	σ(byte)
128	0.322	161	0.635	141	2.00	1032	3.51	822

Comparison with other post-quantum signature schemes. Finally, we compare our scheme with other post-quantum signature schemes, including the three signature schemes [25, 29, 35] finally selected by NIST. From table 4, it can be seen that although our scheme has slightly larger public key sizes compared to these three schemes, it has a significant advantage in terms of signature sizes. The signature sizes of schemes $\mathcal{S}_1$ and $\mathcal{S}_2$ are only 0.242 and 0.212 times that of the Falcon scheme, respectively, which has the shortest signature among the three NIST finalists.

Table 4. Public key and signature sizes of our signature schemes compared with the NIST PQC finalist signature schemes

λ	$\mathcal{S}_1/\mathcal{S}_2$		Crystals-Dilithium [29]		Falcon [35]		Sphincs+ [25]	
	pk(MB)	σ(byte)	pk(byte)	σ(byte)	pk(byte)	σ(byte)	pk(byte)	σ(byte)
128	0.322/0.635	161/141	1312	2420	897	666	32	7856

6 Conclusions

In this paper, we proposed a novel code-based digital signature scheme. Through a specific construction, all syndromes in a certain subspace can be effectively decoded, thus improving the decoding efficiency of the classical CFS signature scheme. Additionally, we replaced the permutation matrix with a reversible matrix as part of the private key, making the resulting public key matrix statistically indistinguishable from a random matrix. To the best of our knowledge, this is the first proof that the resulting public key matrix is statistically indistinguishable from a random matrix. We provided two set of parameters balancing public key and signature sizes and compared them with other post-quantum signature schemes. Compared to the three NIST PQC finalist signature schemes (Crystals-Dilithium, Falcon, and Sphincs+), our scheme has a significant advantage in signature size, less than a quarter of the Falcon scheme. Compared to

the other two code-based schemes (Wave and Enhanced pqsigRM), our scheme has significant advantages in both public key and signature sizes, both less than one-fifth of theirs.

Acknowledgements. This work is supported by the Guangdong Major Project of Basic and Applied Basic Research (2019B030302008) and the National Natural Science Foundation of China (No. 62272491) and the Project of Guangdong Provincial Key Laboratory of Information Security Technology (Grant No. 2023B1212060026).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aragon, N., Blazy, O., Gaborit, P., Hauteville, A., Zémor, G.: Durandal: A rank metric based signature scheme. In: Ishai, Y., Rijmen, V. (eds.) EUROCRYPT 2019. LNCS, vol. 11478, pp. 728–758. Springer (2019). https://doi.org/10.1007/978-3-030-17659-4_25
2. Aragon, N., Dyseryn, V., Gaborit, P.: Analysis of the security of the PSSI problem and cryptanalysis of the durandal signature scheme. In: Handschuh, H., Lysyanskaya, A. (eds.) CRYPTO 2023. LNCS, vol. 14083, pp. 127–149. Springer (2023). https://doi.org/10.1007/978-3-031-38548-3_5
3. Ayebie, E.B., Assidi, H., Souidi, E.M.: An efficient identification scheme based on rank metric. In: Benzekri, A., Barbeau, M., Gong, G., Laborde, R., García-Alfaro, J. (eds.) FPS 2019. LNCS, vol. 12056, pp. 273–289. Springer (2019). https://doi.org/10.1007/978-3-030-45371-8_17
4. Banegas, G., Carrier, K., Chailloux, A., Couvreur, A., Debris-Alazard, T., Gaborit, P., Karpman, P., Loyer, J., Niederhagen, R., Sendrier, N., Smith, B., Tillich, J.P.: Wave. NIST PQC Standardization of Additional Digital signature Schemes Round 1 Submission. (2023), https://wave-sign.org/wave_documentation.pdf
5. Barreto, P.S.L.M., Misoczki, R., Jr., M.A.S.: One-time signature scheme from syndrome decoding over generic error-correcting codes. J. Syst. Softw. **84**(2), 198–204 (2011). https://doi.org/10.1016/J.JSS.2010.09.016
6. Becker, A., Joux, A., May, A., Meurer, A.: Decoding random binary linear codes in $2^{n/20}$: How $1 + 1 = 0$ improves information set decoding. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012. LNCS, vol. 7237, pp. 520–536. Springer (2012). https://doi.org/10.1007/978-3-642-29011-4_31
7. Bellini, E., Caullery, F., Gaborit, P., Manzano, M., Mateu, V.: Improved veron identification and signature schemes in the rank metric. In: ISIT 2019. pp. 1872–1876. IEEE (2019). https://doi.org/10.1109/ISIT.2019.8849585
8. Bellini, E., Caullery, F., Hasikos, A., Manzano, M., Mateu, V.: Code-based signature schemes from identification protocols in the rank metric. In: Camenisch, J., Papadimitratos, P. (eds.) CANS 2018. LNCS, vol. 11124, pp. 277–298. Springer (2018). https://doi.org/10.1007/978-3-030-00434-7_14

9. Bellini, E., Gaborit, P., Hasikos, A., Mateu, V.: Enhancing code based zero-knowledge proofs using rank metric. In: Krenn, S., Shulman, H., Vaudenay, S. (eds.) CANS 2020. LNCS, vol. 12579, pp. 570–592. Springer (2020). https://doi.org/10.1007/978-3-030-65411-5_28
10. Berlekamp, E.R., McEliece, R.J., van Tilborg, H.C.A.: On the inherent intractability of certain coding problems (corresp.). IEEE Trans. Inf. Theory **24**(3), 384–386 (1978). https://doi.org/10.1109/TIT.1978.1055873
11. Bidoux, L., Gaborit, P., Kulkarni, M., Mateu, V.: Code-based signatures from new proofs of knowledge for the syndrome decoding problem. Des. Codes Cryptogr. **91**(2), 497–544 (2023). https://doi.org/10.1007/S10623-022-01114-3
12. Cayrel, P., Otmani, A., Vergnaud, D.: On kabatianskii-krouk-smeets signatures. In: Carlet, C., Sunar, B. (eds.) WAIFI 2007. LNCS, vol. 4547, pp. 237–251. Springer (2007). https://doi.org/10.1007/978-3-540-73074-3_18
13. Cayrel, P., Véron, P., Alaoui, S.M.E.Y.: A zero-knowledge identification scheme based on the q-ary syndrome decoding problem. In: Biryukov, A., Gong, G., Stinson, D.R. (eds.) SAC 2010. LNCS, vol. 6544, pp. 171–186. Springer (2010). https://doi.org/10.1007/978-3-642-19574-7_12
14. Chizhov, I.V., Borodin, M.A.: The failure of mceliece PKC based on reed-muller codes. IACR Cryptol. ePrint Arch. (2013), http://eprint.iacr.org/2013/287
15. Cho, J., No, J., Lee, Y., Kim, Y., Koo, Z.: Enhanced pqsigrm: Code-based digital signature scheme with short signature and fast verification for post-quantum cryptography. IACR Cryptol. ePrint Arch. p. 1493 (2022), https://eprint.iacr.org/2022/1493
16. Courtois, N.T., Finiasz, M., Sendrier, N.: How to achieve a mceliece-based digital signature scheme. In: Boyd, C. (ed.) ASIACRYPT 2001. LNCS, vol. 2248, pp. 157–174. Springer (2001). https://doi.org/10.1007/3-540-45682-1_10
17. Dallot, L.: Towards a concrete security proof of courtois, finiasz and sendrier signature scheme. In: Lucks, S., Sadeghi, A., Wolf, C. (eds.) WEWoRC 2007,. LNCS, vol. 4945, pp. 65–77. Springer (2007). https://doi.org/10.1007/978-3-540-88353-1_6
18. Debris-Alazard, T., Sendrier, N., Tillich, J.: Wave: A new family of trapdoor one-way preimage sampleable functions based on codes. In: Galbraith, S.D., Moriai, S. (eds.) ASIACRYPT 2019. LNCS, vol. 11921, pp. 21–51. Springer (2019). https://doi.org/10.1007/978-3-030-34578-5_2
19. Dong, X., Wang, X.: Quantum key-recovery attack on feistel structures. Sci. China Inf. Sci. **61**(10), 102501:1–102501:7 (2018). https://doi.org/10.1007/S11432-017-9468-Y
20. Faugère, J., Gauthier-Umaña, V., Otmani, A., Perret, L., Tillich, J.: A distinguisher for high-rate mceliece cryptosystems. IEEE Trans. Inf. Theory **59**(10), 6830–6844 (2013). https://doi.org/10.1109/TIT.2013.2272036
21. Feneuil, T., Joux, A., Rivain, M.: Shared permutation for syndrome decoding: new zero-knowledge protocol and code-based signature. Des. Codes Cryptogr. **91**(2), 563–608 (2023). https://doi.org/10.1007/S10623-022-01116-1
22. Finiasz, M.: Parallel-cfs - strengthening the CFS mceliece-based signature scheme. In: Biryukov, A., Gong, G., Stinson, D.R. (eds.) SAC 2010. LNCS, vol. 6544, pp. 159–170. Springer (2010). https://doi.org/10.1007/978-3-642-19574-7_11
23. Grover, L.K.: A fast quantum mechanical algorithm for database search. In: Miller, G.L. (ed.) Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, Philadelphia, Pennsylvania, USA, May 22-24, 1996. pp. 212–219. ACM (1996). https://doi.org/10.1145/237814.237866

24. Hauteville, A., Tillich, J.: New algorithms for decoding in the rank metric and an attack on the LRPC cryptosystem. In: ISIT 2015. pp. 2747–2751. IEEE (2015). https://doi.org/10.1109/ISIT.2015.7282956
25. Hulsing, A., Bernstein, D.J., Dobraunig, C., Eichlseder, M., Fluhrer, S., Gazdag, S.L., Kampanakis, P., Kolbl, S., Lange, T., Lauridsen, M.M., Mendel, F., Niederhagen, R., Rechberger, C., Rijneveld, J., Schwabe, P., Aumasson, J.P., Westerbaan, B., Beullens, W.: Sphincs+. Algorithm selected by NIST PQC. (2022), https://sphincs.org
26. Kabatianskii, G., Krouk, E.A., Smeets, B.J.M.: A digital signature scheme based on random error-correcting codes. In: Darnell, M. (ed.) IMA 1997. LNCS, vol. 1355, pp. 161–167. Springer (1997). https://doi.org/10.1007/BFB0024461
27. Lau, T.S.C., Tan, C.H., Prabowo, T.F.: Key recovery attacks on some rank metric code-based signatures. In: Albrecht, M. (ed.) Cryptography and Coding - 17th IMA International Conference, IMACC 2019. LNCS, vol. 11929, pp. 215–235. Springer (2019). https://doi.org/10.1007/978-3-030-35199-1_11
28. Lee, W., Kim, Y., Lee, Y., No, J.: Post quantum signature scheme based on modified reed-muller code pqsigrm. first round submission to the NIST post-quantum cryptography call, November 2017 (2017)
29. Lyubashevsky, V., Ducas, L., Kiltz, E., Lepoint, T., Schwabe, P., Seiler, G., Stehle, D., Bai, S.: Crystals-dilithium. Algorithm selected by NIST PQC. (2022), https://pq-crystals.org
30. McEliece, R.J.: A public key cryptosystem based on algebraic coding theory. DSN Progress Report **42**(44), 114–116 (1978)
31. Melchor, C.A., Gaborit, P., Schrek, J.: A new zero-knowledge code based identification scheme with reduced communication. In: ITW 2011. pp. 648–652. IEEE (2011). https://doi.org/10.1109/ITW.2011.6089577
32. Minder, L., Shokrollahi, A.: Cryptanalysis of the sidelnikov cryptosystem. In: Naor, M. (ed.) EUROCRYPT 2007. LNCS, vol. 4515, pp. 347–360. Springer (2007). https://doi.org/10.1007/978-3-540-72540-4_20
33. Otmani, A., Tillich, J.: An efficient attack on all concrete KKS proposals. In: Yang, B. (ed.) PQCrypto 2011. LNCS, vol. 7071, pp. 98–116. Springer (2011). https://doi.org/10.1007/978-3-642-25405-5_7
34. Prange, E.: The use of information sets in decoding cyclic codes. IRE Trans. Inf. Theory **8**(5), 5–9 (1962). https://doi.org/10.1109/TIT.1962.1057777
35. Prest, T., Fouque, P.A., Hoffstein, J., Kirchner, P., Lyubashevsky, V., Pornin, T., Ricosset, T., Seiler, G., Whyte, W., Zhang, Z.: Falcon. Algorithm selected by NIST PQC. (2022), https://falcon-sign.info
36. Regev, O.: An efficient quantum factoring algorithm. CoRR **abs/2308.06572** (2023), https://doi.org/10.48550/arXiv.2308.06572
37. Saxena, A., Shukla, A., Pathak, A.: A hybrid scheme for prime factorization and its experimental implementation using IBM quantum processor. Quantum Inf. Process. **20**(3), 1–15 (2021). https://doi.org/10.1007/S11128-021-03053-9
38. Shor, P.W.: Algorithms for quantum computation: Discrete logarithms and factoring. In: 35th Annual Symposium on Foundations of Computer Science, Santa Fe, New Mexico, USA, 20-22 November 1994. pp. 124–134. IEEE Computer Society (1994). https://doi.org/10.1109/SFCS.1994.365700
39. Shor, P.W.: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. SIAM Rev. **41**(2), 303–332 (1999). https://doi.org/10.1137/S0036144598347011
40. Song, Y., Huang, X., Mu, Y., Wu, W.: An improved durandal signature scheme. Sci. China Inf. Sci. **63**(3) (2020). https://doi.org/10.1007/S11432-019-2670-7

41. Stern, J.: A new identification scheme based on syndrome decoding. In: Stinson, D.R. (ed.) CRYPTO '93. LNCS, vol. 773, pp. 13–21. Springer (1993). https://doi.org/10.1007/3-540-48329-2_2