A Simple Noncommutative UOV Scheme

Lih-Chung Wang*, Po-En Tseng†, Yen-Liang Kuan‡, and Chun-Yen Chou§

Department of Applied Mathematics, National Dong Hwa University, Taiwan, R.O.C.

Abstract. UOV (Unbalanced Oil and Vinegar signature scheme), initially introduced in 1997, has undergone extensive research and is widely recognized as a robust signature scheme with excellent efficiency. Nonetheless, UOV is hindered by its substantial public key sizes. Specifically, when targeting NIST security level I, UOV public keys typically span from 40KB to 60KB in size. We propose a new multivariate signature scheme SNOVA (Simple Noncommutative unbalanced Oil and Vinegar scheme with randomness Alignment), which is a UOV-variant scheme over noncommutative rings. In order to enhance the comprehension of SNOVA, we introduce an intermediary phase called ring UOV, which generalizes UOV to any noncommutative ring. However, a ring UOV may be viewed as a big UOV system with sparse matrix representations. We further modified ring UOV to SNOVA, which resolves the sparsity problem. In comparison to UOV, SNOVA achieves a remarkable reduction in the public key size, making it to a mere 1KB, while maintaining commendable performance levels.

Keywords: PQC, MQ, digital signature, UOV, ring UOV, SNOVA

*Corresponding author: Lih-Chung Wang, Email: lcwang@gms.ndhu.edu.tw
†briantseng0320@gmail.com
‡ylkuan@gms.ndhu.edu.tw
§choucy@gms.ndhu.edu.tw

1 Introduction

The Unbalanced Oil and Vinegar (UOV) signature scheme [26] is a slight modification of the Oil and Vinegar (OV) [35] signature scheme, proposed by Patarin in 1997.

The UOV signature scheme has been studied and analyzed for a long time. To this day, it is still believed to be a secure scheme. However, as a multivariate signature scheme, it still suffers from the problem of having excessively large public keys. In the literature, many related variants have been proposed, which try to address the issue of large public keys while retaining the advantages of UOV [42, 14, 5].

On the other hand, fundamental public key compression methods have been proposed. A. Petzoldt [36, 37] and Rainbow [13] of the third-round of NIST proposal showed that part of the randomness of the private key can be transferred to the public key and then a large part of public key can be generated by a PRNG (Pseudorandom Number Generator) which we called "randomness alignment" technique here. This reduces the public key size of UOV to the order $O(m^3 \cdot \log q)$ where m is number of equations and q is the order of finite field in UOV scheme. For the modern parameters of UOV which aiming at NIST security level I [32], the public key sizes are about 40KB to 60KB. However, these sizes of the public key of UOV scheme are still too large.

To alleviate the problem, new possibilities have come into our view. By generalizing the UOV scheme to noncommutative rings, we can further reduce the size of the public key. Through some appropriate modifications, the public key compression techniques of UOV remain applicable to our new signature scheme on noncommutative rings.

Our contribution. In this paper, we propose a new UOV variant over noncommutative rings called SNOVA.

In SNOVA, we see several advantages:

- By building on noncommutative rings, we can reduce the size of the public key while still maintaining the advantage of short signatures.
- The randomness alignment key-compression technique of Petzoldt [36] can be successfully adapted to SNOVA without being affected by noncommutativity.

- There is an intuitive connection between SNOVA and UOV. In the case that $l = 1$ of the underlying matrix ring, SNOVA reduces to UOV scheme.

We propose parameter sets aiming for NIST security levels I, III, and V. For security level I, one of our parameter sets results in a public key size of 1000 bytes and a signature size of 232 bytes. With these performance, we believe that the SNOVA scheme has strong competitiveness compared to other post-quantum signature schemes. Additionally, through the generalization of UOV to noncommutative rings, we hope to open up new possibilities for designing signature schemes.

2 Preliminaries

The following Tables 1 and 2 are tables that list symbols fixed with specific meaning and conventions on notations, respectively.

Table 1: The table of conventions on notations in this paper.

Description	**The font denoted with**	**Example**
Integers and elements in finite field $\mathbb{F}_q$	lower case letters	n, m and l
Elements in ring $\mathcal{R}$	upper case letters	A, S and Q
Variables over $\mathcal{R}$	upper case letters	$X_1, \cdots, X_n$
Variables over $\mathbb{F}_q$	lower case letters	$x_1, \cdots, x_n$
Vectors of any dimension	boldface letters with an arrow on top	$\vec{\mathbf{X}}$ and $\vec{\mathbf{x}}$
Vector spaces and rings	calligraphic font	$\mathcal{O}$ and $\mathcal{R}$
The (j, k)-th entry of the matrix $[F_i]$, $[T]$ and $[P_i]$, respectively	subscript j, k	$F_{i,jk}$, T_{jk} and $P_{i,jk}$

Table 2: The table of symbols fixed with specific meaning in this paper.

Symbol	Description
$\mathbb{F}_q$	finite field of order q
$\mathcal{R}$	$\mathrm{Mat}_{l\times l}(\mathbb{F}_q)$, matrix ring consisting of $l \times l$ matrices over $\mathbb{F}_q$
v, o	numbers of vinegar and oil variables, respectively
S	symmetric matrix in $\mathcal{R}$ with its characteristic polynomial irreducible over $\mathbb{F}_q$
$n = v + o, m = o$	numbers of variables and equations, respectively
$F = [F_1, \ldots, F_m]$	central map of the ring UOV scheme
$[F_i]$	matrix corresponding to F_i in F
$\tilde{F} = \left[\tilde{F}_1, \ldots, \tilde{F}_m\right]$	central map of the SNOVA scheme
T	invertible linear map in signature scheme
$[T]$	matrix corresponding to T
$P = [P_1, \ldots, P_m]$	public map of the ring UOV scheme
$[P_i]$	matrix corresponding to P_i in P
$\tilde{P} = \left[\tilde{P}_1, \ldots, \tilde{P}_m\right]$	public map of the SNOVA scheme
$\mathcal{O}$	oil space
$MQ(N, M, q)$	complexity of an MQ system of M equations in N variables over $\mathbb{F}_q$

2.1 Basic Notions

MQ problem. Let $\mathbb{F}_q$ be a finite field of order q. Given a multivariate quadratic map $P(\vec{\mathbf{x}}) = [P_1(\vec{\mathbf{x}}), \ldots, P_M(\vec{\mathbf{x}})]$ of M components in N variables $\vec{\mathbf{x}} = (x_1, \ldots, x_N)$ and a vector $\vec{\mathbf{y}} \in \mathbb{F}_q^M$, to find a vector $\vec{\mathbf{u}} \in \mathbb{F}_q^N$ such that $P(\vec{\mathbf{u}}) = [P_1(\vec{\mathbf{u}}), \ldots, P_M(\vec{\mathbf{u}})] = \vec{\mathbf{y}}$. This problem is known to be NP-hard [21].

In this paper, we use $MQ(N, M, q)$ to denote the complexity of solving such an MQ problem. There are several algorithms to solve a multivariate quadratic system of M equations in N variables over $\mathbb{F}_q$ such as F_4 [17], F_5 [18] and XL variants [12, 43].

Polar forms. The polar form of a homogeneous multivariate quadratic map $P(\vec{\mathbf{x}}) = [P_1(\vec{\mathbf{x}}), \ldots, P_M(\vec{\mathbf{x}})]$ is defined to be the map

$$P'(\vec{\mathbf{x}}, \vec{\mathbf{y}}) = [P'_1(\vec{\mathbf{x}}, \vec{\mathbf{y}}), \ldots, P'_M(\vec{\mathbf{x}}, \vec{\mathbf{y}})] \tag{2.1}$$

where for each $i \in \{1, \ldots, M\}$ the polar form of $P_i(\vec{\mathbf{x}})$ is defined by

$$P'_i(\vec{\mathbf{x}}, \vec{\mathbf{y}}) = P_i(\vec{\mathbf{x}} + \vec{\mathbf{y}}) - P_i(\vec{\mathbf{x}}) - P_i(\vec{\mathbf{y}}). \tag{2.2}$$

If we write $P_i(\vec{\mathbf{x}}) = \vec{\mathbf{x}}^t [P_i] \vec{\mathbf{x}}$ where $[P_i]$ is the matrix representation of P_i then the matrix representation of P'_i is symmetric and

$$[P'_i] = [P_i] + [P_i]^t . \tag{2.3}$$

2.2 Unbalanced Oil and Vinegar Signature Scheme

A (v, o, q) UOV signature scheme with $v > o$ is defined with a triple of positive integers so that the number of variables $n = v + o$, the number of equations $m = o$, and the scheme is over $\mathbb{F}_q$.

Central map. The central map of UOV scheme is $F = [F_1, \ldots, F_m] : \mathbb{F}_q^n \to \mathbb{F}_q^m$ where each F_i is of the form

$$F_i(x_1, \ldots, x_n) = \sum_{j=1}^{v} \sum_{k=j}^{n} f_{i,jk} x_j x_k. \tag{2.4}$$

The coefficients $f_{i,jk}$'s are chosen randomly from $\mathbb{F}_q$. Note that each F_i is a homogeneous quadratic polynomials in n variables which has no terms $x_j x_k$ for $j, k \in \{v + 1, \ldots, n\}$ over $\mathbb{F}_q$.

Private key and public key. The private key of UOV is the pair (F, T) where $T : \mathbb{F}_q^n \to \mathbb{F}_q^n$ is an invertible linear map which is randomly chosen. The public key is the map $P = [P_1, \ldots, P_m] = F \circ T : \mathbb{F}_q^n \to \mathbb{F}_q^m$ where $P_i = F_i \circ T$.

Oil space, $\mathcal{O}$. The central map F of UOV scheme vanishes on the linear space $\mathcal{O} = \{\vec{\mathbf{x}} \in \mathbb{F}_q^n : x_1 = \cdots = x_v = 0\}$ called the oil space. Then the public map P vanishes on the space $T^{-1}(\mathcal{O})$. For key recovery attacks against UOV, the most important task is to find a nonzero vector in $T^{-1}(\mathcal{O})$. It is because once such a vector is found, we can use this vector and the differential of the public map to successively get more vectors in $T^{-1}(\mathcal{O})$, and finally to obtain a basis of $T^{-1}(\mathcal{O})$. And then such a basis can be used to induce an equivalent key [4].

3 Ring UOV

In order to enhance the comprehension of SNOVA, we now introduce an intermediary phase called ring UOV, which generalizes UOV to any noncommutative ring $\mathcal{R}$. There are other schemes involving noncommutative rings but with different techniques been proposed [19, 47].

Similar to UOV, Let $n = v + o$ and $m = o$. However, due to the noncommutativity of $\mathcal{R}$ we need to explicitly denote the following index set which will be used below by

$$\Omega = \{(j,k) : 1 \leq j,k \leq n\} \setminus \{(j,k) : v+1 \leq j,k \leq n\}. \tag{3.1}$$

The basic structure of ring UOV. The central map of ring UOV is the map $F = [F_1, \ldots, F_m] : \mathcal{R}^n \to \mathcal{R}^m$ with each F_i defined by

$$F_i(X_1, \ldots, X_n) = \sum_{(j,k)\in\Omega} \phi(X_j) F_{i,jk} X_k \tag{3.2}$$

where the coefficients $F_{i,jk}$ are randomly chosen from $\mathcal{R}$. The map ϕ is a ring map with "factor order reversed" property, i.e., $\phi\left(\sum_j C_j X_j\right) = \sum_j \phi(X_j)\,\phi(C_j)$ where $C_j \in \mathcal{R}$. The (ring) variables $X_1, \ldots, X_v$ are called the vinegar variables and $X_{v+1}, \ldots, X_n$ are called the oil variables.

A concrete example of ring UOV. For the purpose of explaining SNOVA, we now fix the noncommutative ring to be $\mathcal{R} = \mathrm{Mat}_{l\times l}(\mathbb{F}_q)$ and the ring map ϕ to be the matrix transpose. Then, we have a (v, o, q, l)-type ring UOV scheme. And, for brevity, we will call it a (v, o, q, l) ring UOV or simply a ring UOV. Due to these specification, the i-th component, for $i \in \{1, \ldots, m\}$, of the central map $F = [F_1, \ldots, F_m] : \mathcal{R}^n \to \mathcal{R}^m$ becomes

$$F_i(X_1, \ldots, X_n) = \sum_{(j,k)\in\Omega} X_j^t F_{i,jk} X_k. \tag{3.3}$$

Note that we can write F_i into quadratic form over $\mathcal{R}$. That is,

$$F_i(\vec{\mathbf{X}}) = \vec{\mathbf{X}}^t\,[F_i]\,\vec{\mathbf{X}} \tag{3.4}$$

where $\vec{\mathbf{X}} = (X_1, \ldots, X_n)^t$ and the matrix representation $[F_i]$ over $\mathcal{R}$ corresponding to F_i is of the form

$$[F_i] = \left[F_{i,jk}\right] = \begin{bmatrix} F_i^{11} & F_i^{12} \\ F_i^{21} & 0 \end{bmatrix}, \tag{3.5}$$

F_i^{11}, F_i^{12} and F_i^{21} are matrices over $\mathcal{R}$ of size $v \times v$, $v \times o$ and $o \times v$, respectively.

Similar to UOV scheme, the public map $P = [P_1, \cdots, P_m]$ is the composition of central map F and an invertible ring linear map $T : \mathcal{R}^n \to \mathcal{R}^n$, i.e., $P(\vec{\mathbf{U}}) = (F \circ T)(\vec{\mathbf{U}})$ where $P_i(\vec{\mathbf{U}}) = (F_i \circ T)(\vec{\mathbf{U}})$ for each $i \in \{1, 2, \ldots, m\}$. The map T is defined by its matrix representation

$$[T] = \begin{bmatrix} I^{11} & T^{12} \\ 0 & I^{22} \end{bmatrix} \tag{3.6}$$

where T^{12} is a $v \times o$ random matrix over $\mathcal{R}$ and I^{11}, I^{22} are identity matrices over $\mathcal{R}$ of size $v \times v$ and $o \times o$, respectively.

Public key and private key. For each $i \in \{1, \ldots, m\}$, we have

$$P_i(\vec{\mathbf{U}}) = (F_i \circ T)(\vec{\mathbf{U}}) = \vec{\mathbf{U}}^t \cdot \Big([T]^t\,[F_i]\,[T] \Big) \cdot \vec{\mathbf{U}}. \tag{3.7}$$

Therefore, the public key consists of the corresponding matrices generated by the following congruence relation, for $i \in \{1, \cdots, m\}$,

$$[P_i] = \left[P_{i,d_j d_k}\right] = [T]^t\,[F_i]\,[T] \tag{3.8}$$

and the private key is (F, T), i.e., the matrix $[T]$ and the matrices $[F_i]$.

4 SNOVA: A Simple Noncommutative UOV Scheme

In this section, we introduce SNOVA signature scheme whose central map is a modified ring UOV map. In order to eliminate the sparsity of ring UOV map (when we regard it as a UOV map over field), some specific matrices will be introduced into the ring UOV map. And we will see that, through appropriate design, these introduced matrices will not affect the process of SNOVA public key generation. The key generation will be almost identical to the case of the ring UOV scheme, which will be explained below.

4.1 Description

Let v, o be positive integers with $v > o$ and $\mathbb{F}_q$ be of characteristic 2. For example, we choose $\mathbb{F}_q = \mathrm{GF}(16)$ for our implementation. Let $n = v + o$ and $m = o$. Next, we will proceed to introduce the subring of the matrix ring $\mathcal{R}$, $\mathbb{F}_q[S]$. Then, we will define a (v, o, q, l) SNOVA scheme.

The subring $\mathbb{F}_q[S]$. Let S be an $l \times l$ symmetric matrix with its characteristic polynomial irreducible over $\mathbb{F}_q$. The subring $\mathbb{F}_q[S]$ of $\mathcal{R}$ is defined to be

$$\mathbb{F}_q[S] = \{a_0 + a_1 S + \cdots + a_{l-1} S^{l-1} : a_0, a_1, \cdots, a_{l-1} \in \mathbb{F}_q\} \tag{4.1}$$

and note that the elements in $\mathbb{F}_q[S]$ are also symmetric and they all commute.

Central map and its core part. let $\Omega = \{(j,k) : 1 \leq j,k \leq n\} \setminus \{(j,k) : v+1 \leq j,k \leq n\}$. The central map of SNOVA scheme is $\tilde{F} = \left[\tilde{F}_1, \cdots, \tilde{F}_m\right] : \mathcal{R}^n \to \mathcal{R}^m$ and, for $i \in \{1, \cdots, m\}$, F_i is defined to be

$$\tilde{F}_i(X_1, \ldots, X_n) = \sum_{\alpha=1}^{l^2} A_\alpha \cdot \left(\sum_{(j,k)\in\Omega} X_j^t \left(Q_{\alpha 1} F_{i,jk} Q_{\alpha 2}\right) X_k \right) \cdot B_\alpha \tag{4.2}$$

where $F_{i,jk}$'s are randomly chosen from $\mathcal{R}$, A_α and B_α are invertible elements randomly chosen from $\mathcal{R}$, and $Q_{\alpha 1}$, $Q_{\alpha 2}$ are invertible matrices randomly chosen from $\mathbb{F}_q[S]$.

For the central map $\tilde{F}$ of SNOVA, we define its core part to be the corresponding ring UOV map. That is, for $i \in \{1, \ldots, m\}$, we define

$$core(\tilde{F}_i) := F_i = \sum_{(j,k)\in\Omega} X_j^t F_{i,jk} X_k. \tag{4.3}$$

From the above definition, we can observe that for a central map of SNOVA, there always exists a corresponding ring UOV map. Through the core part, even if the central map of SNOVA can not be represented as a quadratic form over ring (due to matrices A_α, B_α, $Q_{\alpha 1}$ and $Q_{\alpha 2}$), its ring coefficients can still be recorded by the matrix representation of its core part, i.e., the matrices

$$\left[core(\tilde{F}_i)\right] := [F_i] = \left[F_{i,jk}\right] = \begin{bmatrix} F_i^{11} & F_i^{12} \\ F_i^{21} & 0 \end{bmatrix} \tag{4.4}$$

where $[F_i]$ is the matrix representation of the ring UOV map corresponding to $core(\tilde{F}_i)$.

Invertible linear map. The invertible linear map in SNOVA scheme is the map $T : \mathcal{R}^n \to \mathcal{R}^n$ corresponding to the matrix

$$[T] = [T_{ij}] = \begin{bmatrix} I^{11} & T^{12} \\ 0 & I^{22} \end{bmatrix}, \tag{4.5}$$

where T^{12} is a $v \times o$ matrix consisting of nonzero entries T_{ij} chosen randomly in $\mathbb{F}_q[S]$. Note that T_{ij} is symmetric and commutes with other elements in $\mathbb{F}_q[S]$. In particular, T_{ij} commutes with $Q_{\alpha 1}$ and $Q_{\alpha 2}$. The matrices I^{11} and I^{22} are identity matrices over $\mathcal{R}$. Therefore, $[T]$ is invertible and hence T. Note that since $\mathbb{F}_q$ is of characteristic 2, the matrix $[T^{-1}] = [T]$.

Public map. Let $\tilde{P} = \tilde{F} \circ T$ be the public map of SNOVA scheme. For $i \in \{1, 2, \ldots, m\}$, $\tilde{P}_i = \tilde{F}_i \circ T$. The relation $\overrightarrow{\mathbf{X}} = [T] \cdot \overrightarrow{\mathbf{U}}$ where $\overrightarrow{\mathbf{U}} = (U_1, \cdots, U_n) \in \mathcal{R}^n$ implies that

$$\tilde{P}_i(\overrightarrow{\mathbf{U}}) = \tilde{F}_i(T(\overrightarrow{\mathbf{U}})) = \sum_{\alpha=1}^{l^2} \sum_{d_j=1}^{n} \sum_{d_k=1}^{n} A_\alpha \cdot U_{d_j}^t (Q_{\alpha 1} P_{i,d_j d_k} Q_{\alpha 2}) U_{d_k} \cdot B_\alpha \tag{4.6}$$

where $P_{i,d_j d_k} = \sum_{\Omega} T_{j,d_j} \cdot F_{i,jk} \cdot T_{k,d_k}$ by the commutativity of $\mathbb{F}_q[S]$ and that all elements in $\mathbb{F}_q[S]$ are symmetric. Similarly, we define the core part of the public map $\tilde{P}$ by

$$core(\tilde{P}_i) := P_i = core(\tilde{F}_i) \circ T = F_i \circ T. \tag{4.7}$$

Therefore, the matrix representation of the map $core(\tilde{P}_i)$ consists of the corresponding matrices

$$\left[core(\tilde{P}_i)\right] := [P_i] = \left[P_{i,d_j d_k}\right] = [T]^t \,[F_i]\,[T] \tag{4.8}$$

for $i \in \{1, \ldots, m\}$. By introducing the matrices $A_\alpha, B_\alpha, Q_{\alpha 1}, Q_{\alpha 2}$, the public map $\tilde{P}$ is not a sparse UOV map when we regard it as over $\mathbb{F}_q$.

Public key and private key. The public key is the matrices $\left[core(\tilde{P}_i)\right]$ that records the ring coefficients of $core(\tilde{P}_i)$ and the matrices A_α, B_α, $Q_{\alpha 1}$ and $Q_{\alpha 2}$ for $\alpha = 1, 2, \ldots, l^2$, or simply the seed $\mathbf{s_{public}}$ which generates them. By utilizing matrices $\left[core(\tilde{P}_i)\right]$ and the seed $\mathbf{s_{public}}$, the verifier is capable to obtain the public map $\tilde{P}$ and subsequently verify the received signature.

The private key of SNOVA is (F, T), i.e., the matrix $[T]$ and the matrices $[F_i]$ for $i = 1, 2, \ldots, m$. Note that we can use the private seed $\mathbf{s_{private}}$ to generate T.

Signature. Let D be the document to be signed and $Hash(D) = \overrightarrow{\mathbf{Y}} = (Y_1, \cdots, Y_m) \in \mathcal{R}^m$ be its hash value. We compute the signature $\overrightarrow{\mathbf{U}}$ step by step. First, We assign values to vinegar variables $X_1, \cdots, X_v$ randomly and the resulting system can be seen as a linear system over the $\mathbb{F}_q$-entries of oil variables $X_{v+1}, \cdots, X_n$. The remaining is the same as in UOV scheme by regarding SNOVA as a UOV over $\mathbb{F}_q$. Secondly, the signature is $\overrightarrow{\mathbf{U}} = T^{-1}(\overrightarrow{\mathbf{X}}) \in \mathcal{R}^n$.

Verification. Let $\overrightarrow{\mathbf{U}} = (U_1, \cdots, U_n) \in \mathcal{R}^n$ be the signature to be verified. If $Hash(D) = \tilde{P}(\overrightarrow{\mathbf{U}})$, then the signature is accepted, otherwise rejected.

4.2 Key generation process of SNOVA

In this section, we give the standard key generation process of SNOVA and the key generation process with randomness alignment key-compression technique [36]. Note that, in SNOVA scheme, $\mathbb{F}_q$ is of the characteristic 2.

Standard key generation process. For $i \in \{1, \ldots, m\}$, the matrix $[P_i]$ is obtained by relation

$$[P_i] = [T]^t [F_i] [T]. \tag{4.9}$$

Then, we have the following

$$P_i^{11} = F_i^{11} \tag{4.10}$$
$$P_i^{12} = F_i^{11}T^{12} + F_i^{12} \tag{4.11}$$
$$P_i^{21} = (T^{12})^t F_i^{11} + F_i^{21} \tag{4.12}$$
$$P_i^{22} = (T^{12})^t \cdot \left(F_i^{11}T^{12} + F_i^{12}\right) + F_i^{21}T^{12}. \tag{4.13}$$

Therefore, to get $[P_i]$, we generate the matrices $[F_i]$, $[T]$ from a seed $\mathbf{s_{private}}$ at first and then compute $[P_i]$ for $i \in \{1, \ldots, m\}$ with the formulas above.

Key generation with randomness alignment. The following are steps of key generation process of SNOVA with key randomness alignment.

First Step: Generate S, P_i^{11}, P_i^{12} and P_i^{21} for $i \in \{1, \ldots, m\}$, and $[T]$ from two seeds $\mathbf{s_{public}}$ and $\mathbf{s_{private}}$ respectively. We also generate the matrices A_α, B_α, $Q_{\alpha 1}$ and $Q_{\alpha 2}$ for $\alpha = 1, 2, \ldots, l^2$ from $\mathbf{s_{public}}$.

Second Step: Compute the matrix $F_i^{11}, F_i^{12}, F_i^{21}, P_i^{22}$ for $i \in \{1, \ldots, m\}$ as below.

For $i \in \{1, \ldots, m\}$, we have

$$[F_i] = \left[T^{-1}\right]^t [P_i] \left[T^{-1}\right]. \tag{4.14}$$

Therefore, the following equations hold

$$F_i^{11} = P_i^{11} \tag{4.15}$$
$$F_i^{12} = P_i^{11} T^{12} + P_i^{12} \tag{4.16}$$
$$F_i^{21} = (T^{12})^t P_i^{11} + P_i^{21} \tag{4.17}$$
$$0 = F_i^{22} = (T^{12})^t \cdot \left(P_i^{11} T^{12} + P_i^{12}\right) + P_i^{21} T^{12} + P_i^{22}. \tag{4.18}$$

In other words, we then have

$$P_i^{22} = \left(T^{12}\right)^t \cdot \left(P_i^{11} T^{12} + P_i^{12}\right) + P_i^{21} T^{12}. \tag{4.19}$$

Public key size. The reduced size of the public key of SNOVA using alignment is

$$\text{Size}_{\text{SNOVA}} = m \cdot m^2 \cdot l^2 \tag{4.20}$$

field elements of $\mathbb{F}_q$. Note that the key size here does not include the size of the public seed $\mathbf{s_{public}}$ which is negligible in comparison to P_i^{22}'s.

5 Security Analysis

The SNOVA scheme can be considered as both a UOV-like signature scheme over the matrix ring $\mathcal{R}$ and a UOV over $\mathbb{F}_q$. The security analysis are presented from two different aspects: over the ring $\mathcal{R}$ and over the finite field $\mathbb{F}_q$.The target of this section is to explore various methods of attacking the SNOVA and to assess their feasibility.

For forgery attacks, the security analysis of SNOVA mainly based on the public map of SNOVA. Since the public keys of SNOVA and the ring UOV corresponding to its core part both are generated by the congruence relation $[P_i] = [T]^t [F_i] [T]$, they share the same private key $[T]$. For key recovery attacks, the security of SNOVA will be evaluated by analyzing the complexity of such attacks against the ring UOV scheme corresponding to $core(\tilde{F}_i)$ which has a much simpler structure.

5.1 Solving MQ systems and Complexity Estimation

There are several algorithms to solve a quadratic system of M equations in N variables over finite fields such as F_4 [17], F_5 [18] and XL variants [9, 12, 43].

Solving MQ problem. The complexity of solving M homogeneous quadratic equations in N variables [9] can be estimated by

$$MQ(N, M, q) = 3 \cdot \binom{N - 1 + d_{reg}}{d_{reg}}^2 \cdot \binom{N + 1}{2} \tag{5.1}$$

field multiplications. The term d_{reg}, degree of regularity of a semi-regular polynomial system [1], equals to the smallest positive integer d such that the coefficient of t^d term in the series generated by

$$\frac{(1 - t^2)^M}{(1 - t)^N} \tag{5.2}$$

is non-positive.

Hybrid approach. The hybrid approach [2] randomly guesses k variables before solving the MQ system and the corresponding complexity is

$$HMQ(N, M, q) = \min_k q^k \cdot MQ(N - k + 1, M, q) \tag{5.3}$$

field multiplications for the classical case and

$$\min_k q^{k/2} \cdot MQ(N - k + 1, M, q) \tag{5.4}$$

field multiplications when applying Grover's algorithm [22] for the quantum case.

Methods solving underdetermined MQ. On the other hand, several methods [20, 23, 41] have been proposed to solve underdetermined MQ more efficiently. These methods can transform an underdetermined $MQ(N, M, q)$ problem to an $MQ(M - k - \alpha_k, M - \alpha_k, q)$ problem where the value of α_k depends on the approach utilized in each method. (Generally, the attack in [23] would be the sharpest among [20, 23, 41].) Hence, the main term of complexity of solving MQ system under this technique is given by

$$\min_k q^k \cdot MQ(M - k - \alpha_k + 1, M - \alpha_k, q) \tag{5.5}$$

field multiplications in the classical case and

$$\min_k q^{k/2} \cdot MQ(M - k - \alpha_k + 1, M - \alpha_k, q) \tag{5.6}$$

in the quantum case with different optimal values α_k corresponding to different methods.

Recently, the algorithm in [23] has been revised. The updated algorithm has become more efficient. It reduces the complexity of direct attack on the MAYO scheme with the latest parameters in the submission of the additional NIST PQC standardization [6], making it unable to meet NIST security levels. When solving an underdetermined MQ system, our complexity estimations consider the method with the lowest complexity.

Algorithms for super-underdetermined MQ. Note that, [10, 11, 27, 30] indicate that when the number of variables N is sufficiently larger than the number of equations M in an MQ problem then we can solve this MQ in polynomial time. Please refer to the table in [23] for more information. Note that these four algorithms are not applicable to the parameter sets of SNOVA.

5.2 To Attain EUF-CMA Security

For practical considerations, we use a random binary vector, called salt in order to achieve Existential Unforgeability under Chosen Message Attack (EUF-CMA) Security [33].

Signature. Let D be the document to be signed, we randomly choose **salt** and then generate a signature for the hash value $\overrightarrow{\mathbf{Y}} = Hash(Hash(D)||\mathbf{salt})$. Therefore, the corresponding signature is of the form $\vec{\sigma} = (\overrightarrow{\mathbf{U}}||\mathbf{salt})$ where $\overrightarrow{\mathbf{U}}$ is the signature of $\overrightarrow{\mathbf{Y}}$ generated by the SNOVA signer. Note that we want almost no **salt** is used for more than one signature. Therefore, the length of **salt** is chosen to be 16 Bytes under the assumption of up to 2^{64} signatures being generated with the system.

Verification. If $P(\overrightarrow{\mathbf{U}}) = Hash(Hash(D)||\mathbf{salt})$, the signature is accepted, otherwise rejected.

5.3 Forgery attacks

In this section, we will give the security analysis of two main types of forgery attacks: direct attack and collision attack. The ideas behind these two attacks are straightforward. They directly ignore the structure possessed by the central map and attack the scheme by generating fake signatures.

Finding the preimage of the public map for the hash value of a message is what constitutes signature forgery. However, the public maps of SNOVA and ring UOV are only weakly connected as a result of the use of l^2 copies with different A_α, $Q_{\alpha 1}$, $Q_{\alpha 2}$, and B_α in $\tilde{F}_i$ of SNOVA. Consequently, solving the equations derived from the public map of ring UOV corresponding to the core part does not aid in solving the equations produced by the public map of SNOVA for the purpose of forgery attacks.

Besides, one may try to directly forge valid fake signature of SNOVA over $\mathcal{R}$ not returning to field level. This approach will suffer from the fact that there is no efficient algorithm like F_4, F_5 and XL to solve multivariate quadratic system over the noncommutative ring $\mathcal{R}$. Therefore, the security of forgery attacks will be analyzed with respect to the public map of the SNOVA scheme in the sense that regarding the public map of SNOVA as a UOV public map over $\mathbb{F}_q$.

5.3.1 Direct attack

For a quadratic multivariate polynomial system $P = [P_1, \cdots, P_m]$ consisting of m equations in n variables over $\mathbb{F}_q$ and an intended $\vec{\mathbf{y}} \in \mathbb{F}_q^m$, an attacker can directly try to solve the solution $\vec{\mathbf{u}}$ of the system $P(\vec{\mathbf{u}}) = \vec{\mathbf{y}}$ algebraically with Gröbner basis approach such as [9, 12, 17, 18, 43]. We can assign values to $n-m$ variables in the system $P(\vec{\mathbf{u}}) = \vec{\mathbf{y}} = Hash(\mathbf{digest}||\mathbf{salt})$ randomly and then obtain an MQ system of m equations in m variables which can be solved with high probability. Once the system is solved, the solution $\vec{\mathbf{u}}$ will be a valid fake signature that satisfies $P(\vec{\mathbf{u}}) = \vec{\mathbf{y}}$.

In the case of SNOVA, to produce a fake signature, an attacker need to regard a (v, o, q, l)-SNOVA public map as an (l^2v, l^2o, q)-UOV public map over $\mathbb{F}_q$ and then forge a signature for this UOV. Since each equation over $\mathcal{R} = \mathrm{Mat}_{l\times l}(\mathbb{F}_q)$ yields l^2 equations over $\mathbb{F}_q$, the system over ring $\mathcal{R}$, $P(\vec{\mathbf{U}}) = \vec{\mathbf{Y}}$, with m equations and n ring variables will result in an MQ system consisting of l^2m equations in l^2n field variables.

Table 3 gives comparison of the degree at the first step degree falls or goes flat using F_4 algorithm [17], which is strongly connected to the degree of regularity [15], in Magma algebra system [7] that starts to go either down or flat among all step degrees of the quadratic system obtained by SNOVA and a random quadratic system respectively.

Table 3: Table of comparison of the degree at the first step degree falls or goes flat between SNOVA and random systems. Our experiment shows that in the case of small size parameter sets such a quadratic system over field induced by SNOVA public key with m equations in n variables over $\mathcal{R} = \mathrm{Mat}_{l\times l}(\mathbb{F}_q)$ behaves like a random system consisting of l^2m equations in l^2n variables over $\mathbb{F}_q$.

(v, o, q, l, k)	SNOVA system	random system
$(6, 1, 16, 2, 1)$	3	3
$(6, 2, 16, 2, 1)$	5	5
$(6, 2, 16, 2, 2)$	4	4
$(6, 2, 16, 2, 3)$	3	3
$(6, 3, 16, 2, 1)$	7	7
$(6, 3, 16, 2, 2)$	6	6
$(6, 3, 16, 2, 3)$	5	5
$(6, 4, 16, 2, 2)$	7	7
$(6, 4, 16, 2, 3)$	6	6
$(6, 1, 16, 3, 2)$	4	4
$(6, 1, 16, 3, 3)$	4	4
$(6, 1, 16, 3, 4)$	3	3
$(6, 2, 16, 3, 3)$	7	7
$(6, 2, 16, 3, 4)$	6	6
$(6, 2, 16, 3, 5)$	5	5
$(6, 1, 16, 4, 1)$	9	9
$(6, 1, 16, 4, 2)$	7	7
$(6, 1, 16, 4, 3)$	6	6
$(6, 1, 16, 4, 4)$	5	5
$(6, 1, 16, 4, 5)$	5	5

In random systems, the first fall step degree is generally equal to the degree of regularity. Table 3 indicates that the first fall step degrees of SNOVA systems and random systems are identical for small size parameter sets. Thus, we can expect that the degree of regularity of SNOVA systems, the first fall step degree, and the degree of regularity of random systems are the same. For Gröbner bases algorithms such as F4/F5 and XL, the size of the Macaulay matrix employed in solving quadratic systems is determined by the degree of regularity. The complexity of solving quadratic systems is determined by the difficulty of solving the sparse Macaulay matrix using the Wiedemann solver [44]. As a result, the complexity of a direct attack on SNOVA is estimated by the complexity of a direct attack on random systems.

The complexity of classical direct attack is given by the estimation in [23]

$$
\begin{aligned}
&\mathrm{Comp}_{\mathrm{Direct;\ Classical}}{}^{\mathrm{SNOVA}} && (5.7)\\
=&(l^2m-\alpha-k+1)HMQ(\alpha,\alpha,q) && (5.8)\\
&+q^k\left(HMQ(\alpha-1,\alpha-1,q)+HMQ(l^2m-\alpha-k,l^2m-\alpha,q)\right). && (5.9)
\end{aligned}
$$

provided that $l^2n \geq \max\{(\alpha+1)(l^2m-k-\alpha+1), \alpha(l^2m-k)-(\alpha-1)^2+k\}$ holds.

Note that not only do the first fall degrees of SNOVA and a random system coincide, but the numbers of columns and ranks of Macaulay matrices also exhibit the same correspondence.

5.3.2 Collision attack

To forge a fake signature, an attacker can also try to check M intended signatures $\overrightarrow{\mathbf{U}_j}$ where $j=1,\cdots,M$, and N hash values $Hash(\mathbf{digest}||\mathbf{salt}_k)$ where $k=1,\cdots,N$, whether there exists a collision $P(\overrightarrow{\mathbf{U}_j}) = Hash(\mathbf{digest}||\mathbf{salt}_k)$. And if it does, then the attacker has a valid fake signature. Thus, M signature computations and N hash values computations are involved. Therefore, according to the estimation of [8], the cost of such a collision attack would be

$$M\cdot(l^2m)\cdot\left(2(\log_2 q)^2+3\cdot\log_2 q\right)+N\cdot 2^{17} \tag{5.10}$$

gates in the sense that regarding SNOVA as a UOV scheme over $\mathbb{F}_q$. Note that a lower bound of the complexity of collision attack is

$$2\cdot\left(M(l^2m)\left(2(\log_2 q)^2+3\cdot\log_2 q\right)\cdot N\cdot 2^{17}\right)^{1/2} \tag{5.11}$$

gates. If $MN = q^{l^2m}$, then this lower bound turns into

$$2\cdot\left(q^{l^2m}(l^2m)\left(2(\log_2 q)^2+3\cdot\log_2 q\right)\cdot 2^{17}\right)^{1/2}, \tag{5.12}$$

and the collision exists with probability

$$1-\left(\frac{q^{l^2m}-M}{q^{l^2m}}\right)^N = 1-\left(\frac{MN-M}{MN}\right)^N \tag{5.13}$$
$$= 1-\left(1-\frac{1}{N}\right)^N \tag{5.14}$$
$$\approx 1-e^{\left(\frac{-1}{N}\right)N} \tag{5.15}$$
$$= 1-e^{-1}. \tag{5.16}$$

5.4 Key Recovery Attacks

In this subsection, we analyze the structure of ring UOV defined by the quadratic form over $\mathcal{R}$ corresponding to $core(\tilde{F}_i)$ and the related (lv, lo, q)-UOV and discuss the key recovery attacks against this (lv, lo, q)-UOV.

5.5 UOV Induced From the Core Part of SNOVA

To conduct a comprehensive and prudent security analysis, we start with the following observations. Note that the structure mentioned in this section has similar discussions in [25, 28].

Quadratic form over ring. Let $\overrightarrow{X} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$ where $X_i = \begin{bmatrix} x_i^{(11)} & \cdots & x_i^{(1l)} \\ \vdots & \ddots & \vdots \\ x_i^{(l1)} & \cdots & x_i^{(ll)} \end{bmatrix}$, $i = 1, \ldots, n$, are ring variables. Then, for $i \in \{1, \ldots, m\}$, we have

$$core(\tilde{F}_i)(\overrightarrow{X}) := F_i(\overrightarrow{X}) = \sum_{(j,k)\in\Omega} X_j^t F_{i,jk} X_k = \overrightarrow{X}^t \cdot [F_i] \cdot \overrightarrow{X}. \tag{5.17}$$

Since $\mathrm{Mat}_{n\times n}(\mathcal{R}) = \mathrm{Mat}_{ln\times ln}(\mathbb{F}_q)$, the $n \times n$ matrix $[F_i]$ over $\mathcal{R}$ can be regarded as a $ln \times ln$ matrix over $\mathbb{F}_q$. Hence the quadratic form 5.17 over $\mathcal{R}$

can be viewed as over $\mathbb{F}_q$

$$
\begin{aligned}
& core(\tilde{F}_i)(\overrightarrow{X}) && (5.18)\\
&= \left(\overrightarrow{X}\right)^t \cdot [F_i] \cdot \left(\overrightarrow{X}\right) && (5.19)\\
&= \begin{bmatrix} x_1^{(11)} & \cdots & x_1^{(l1)} & \cdots & x_n^{(11)} & \cdots & x_n^{(1l)} \\ \vdots & & \vdots & & \vdots & & \vdots \\ x_1^{(1l)} & \cdots & x_1^{(ll)} & \cdots & x_n^{(l1)} & \cdots & x_n^{(ll)} \end{bmatrix} \cdot [F_i] \cdot \begin{bmatrix} x_1^{(11)} & \cdots & x_1^{(1l)} \\ \vdots & & \vdots \\ x_1^{(l1)} & \cdots & x_1^{(ll)} \\ \vdots & & \vdots \\ x_n^{(11)} & \cdots & x_n^{(1l)} \\ \vdots & & \vdots \\ x_n^{(l1)} & \cdots & x_n^{(ll)} \end{bmatrix} && (5.20)
\end{aligned}
$$

and $[F_i]$'s are viewed as $ln \times ln$ matrix over $\mathbb{F}_q$ in the last equality. Therefore, if the matrices $[F_1], \ldots, [F_m]$ and $[P_1], \ldots, [P_m]$ are viewed as $ln \times ln$ matrices over $\mathbb{F}_q$, then there exists an (lv, lo, q)-UOV defined on $\mathbb{F}_q$ corresponding to them. Note that the matrices $[F_1], \ldots, [F_m]$ are not symmetric when they are regarded as $ln \times ln$ matrices over $\mathbb{F}_q$. From this point of view, the core part $core(\tilde{F}_i)$ can be related to the central map of the (lv, lo, q)-UOV which is defined by the $ln \times ln$ matrices $[F_i]$ over $\mathbb{F}_q$. Our security analysis will mainly focus on this (lv, lo, q)-UOV, especially for key recovery attacks.

Oil space and Oil Vector. For key recovery attacks against UOV scheme and its variants, the most important task is to find the oil space $T^{-1}(\mathcal{O})$. Similarly, in SNOVA case, the task is to find the oil space of the public map

$$T^{-1}\left(\left\{\overrightarrow{X} \in \mathcal{R}^n : X_1 = \ldots = X_v = 0\right\}\right), \tag{5.21}$$

and then it suffices to find the oil space of the (lv, lo, q)-UOV induced from $core(\tilde{F}_i)$ according to the observation 5.20. In conclusion, once the oil space of the related (lv, lo, q)-UOV, $T^{-1}(\mathcal{O})$ is found, then an equivalent key of SNOVA can be recovered. Here, the space $\mathcal{O}$ is defined by

$$\mathcal{O} = \{\overrightarrow{x} = (x_1, \ldots, x_{ln}) \in \mathbb{F}_q^{ln} : x_1 = \cdots = x_{lv} = 0\} \tag{5.22}$$

On the other hand, since the components of $[T]$ are in $\mathbb{F}_q[S]$, the private key $[T]$ satisfying the identity over $\mathcal{R}$

$$[T]\,[D] = [D]\,[T] \tag{5.23}$$

where $[D] = \begin{bmatrix} S & & \\ & \ddots & \\ & & S \end{bmatrix} = S \cdot I_n$ is a $n \times n$ matrix over $\mathcal{R}$.

If we identify $[T], [D]$ as an $ln \times ln$ matrix over $\mathbb{F}_q$ then

$$[T]^{-1}(\mathcal{O}) = [T]^{-1}[D](\mathcal{O}) = [D][T]^{-1}(\mathcal{O}) \tag{5.24}$$

Therefore, for each oil vector $\vec{x} \in [T]^{-1}(\mathcal{O})$, we have

$$[D] \cdot \vec{x}, \ldots, [D]^{l-1} \cdot \vec{x} \in T^{-1}(\mathcal{O}). \tag{5.25}$$

In particular, for any $\vec{x} \in [T]^{-1}(\mathcal{O})$ and $j, k \in \{0,\ldots, l-1\}$, we then have

$$\vec{x}^t \cdot [D]^j [P_i] [D]^k \cdot \vec{x} = 0, \ (i = 1, \ldots, m). \tag{5.26}$$

5.5.1 Kipnis-Shamir attack (UOV attack)

The KS attack [27] is trying to find an equivalent private key by finding an equivalent invertible linear map T and hence the corresponding matrix $[T]$. Once we have an equivalent $[T]$, we can recover equivalent $[F_i]$ by the relation $[F_i] = [T^{-1}]^t [P_i] [T^{-1}]$. Note that [27] shows that $T^{-1}(\mathcal{O})$, the oil subspace of the public key P of UOV, induces an equivalent key.

In [4, 27], it shows that $T^{-1}(\mathcal{O})$ is an invariant subspace of $[P_i']^{-1} [P_j']$. The KS attack is trying to find a vector in $T^{-1}(\mathcal{O})$. Once one such vector is found, then we expect that the whole space $T^{-1}(\mathcal{O})$ can be recovered efficiently by using method in [4]. A vector in $T^{-1}(\mathcal{O})$ can be expected to be found with q^{v-o} attempts. Note that if there are $[P_i']$'s not invertible, then we can replace $[P_i']$ with invertible linear combinations of $[P_i']$'s randomly chosen and the cryptanalysis of KS attack remains the same.

First of all, we discuss the feasibility of the execution of KS attack over $\mathcal{R}$. For KS attack to be executed, the consistency of multiplication over $\mathcal{R}$ given by a left-module or a right-module over $\mathcal{R}$ is necessary. The KS attack is difficult to execute over $\mathcal{R}$ due to the design of central map F of the ring UOV corresponding to the core part of SNOVA and the noncommutativity of $\mathcal{R}$. Therefore, KS attack is not applicable to SNOVA over $\mathcal{R}$. Note that [35] also proposes two methods to find an invariant subspace: the Linearization method and the Characteristic Polynomial method. These two methods become invalid over $\mathcal{R}$ since they still suffer from the noncommutativity of $\mathcal{R}$.

On the other hand, an attacker can execute KS attack on the (lv, lo, q)-UOV induced from the core part of SNOVA, $core(\tilde{F}_i)$. Then, the complexity is

$$\text{Comp}_{\text{KS; classical}}\text{SNOVA} = q^{lv-lo} \tag{5.27}$$

field multiplications for the classical case and

$$\text{Comp}_{\text{KS; quantum}}\text{SNOVA} = q^{(lv-lo)/2} \tag{5.28}$$

field multiplications for the quantum case.

5.5.2 Reconciliation Attack

The reconciliation attack proposed by [16] against UOV is trying to find a vector $\vec{\mathbf{o}} \in \mathcal{O}$ by solving the system $P(T^{-1}(\vec{\mathbf{o}})) = 0$ and hence the basis of $T^{-1}(\mathcal{O})$ can be recovered. This implies that $P(T^{-1}(\vec{\mathbf{o}})) = 0$ is a quadratic system that having a solution space of dimension m. To expect a unique solution, we can impose m linear constraints with respect to the components of $\vec{\mathbf{o}}$. Hence the complexity of this attack is mainly given by that of solving the quadratic system of m equations in v variables.

A reconciliation attack on SNOVA, if considered over field, is as an attack on an (lv, lo, q)-UOV which trying to find a vector $\vec{x} \in T^{-1}(\mathcal{O})$. Thus, we are in the case of solving the quadratic system

$$\vec{x}^t \cdot [D]^j \cdot [P_i] \cdot [D]^k \cdot \vec{x} = 0, \ (i = 1, \dots, m) \tag{5.29}$$

where $j, k \in \{0, \dots, l-1\}$, which results in $l^2 m$ equations in $lv + 1 = ln - (lo - 1)$ variables. Hence the complexity of reconciliation attack is

$$\text{Comp}_{\text{Reconciliation; classical}}\text{SNOVA} = HMQ(lv + 1, l^2 m, q) \tag{5.30}$$

field multiplications for the classical attacker.

5.5.3 Intersection attack

In [4], Beullens proposed the intersection attack to attack UOV scheme. It uses the polar form of the public key P, that is, $P' = [P'_1, \cdots, P'_m]$ with

$P'_i(\overrightarrow{\mathbf{u_1}}, \overrightarrow{\mathbf{u_2}}) = \overrightarrow{\mathbf{u_1}}^t [P'_i] \overrightarrow{\mathbf{u_2}}$ where $[P'_i] = [P_i] + [P_i]^t$. The intersection attack is trying to first find a vector $\overrightarrow{\mathbf{y}}$ in the subspace, namely the intersection $\Big([P'_i](T^{-1}\mathcal{O})\Big) \cap \Big([P'_j](T^{-1}\mathcal{O})\Big)$ where $[P'_i], [P'_j]$ are invertible, and then to obtain an equivalent key by recovering the subspace $T^{-1}(\mathcal{O})$.

Since $([P'_i]^{-1})\overrightarrow{\mathbf{y}}$, $([P'_j]^{-1})\overrightarrow{\mathbf{y}} \in T^{-1}(\mathcal{O})$, we obtain the following system.

$$\begin{cases} P\Big(([P'_i]^{-1})\,\overrightarrow{\mathbf{y}}\Big) = \overrightarrow{0} \\ P\Big(([P'_j]^{-1})\overrightarrow{\mathbf{y}}\Big) = \overrightarrow{0} \\ P'\Big(([P'_i]^{-1})\overrightarrow{\mathbf{y}}, ([P'_j]^{-1})\overrightarrow{\mathbf{y}}\Big) = \overrightarrow{0} \end{cases} \tag{5.31}$$

In case of intersection attack against SNOVA, due to our construction, we can not write the public polynomial P_i of SNOVA in quadratic form, namely $\overrightarrow{\mathbf{u_1}}^t [P'_i] \overrightarrow{\mathbf{u_2}}$, when considered as over $\mathcal{R}$. Thus, the implementation of intersection attack still face the noncommutativity, that is, there is no efficient algorithm like F_4, F_5 and XL to compute. Therefore, from this perspective, to implement intersection to attack against SNOVA, the possible strategy is attack the (lv, lo, q)-UOV corresponding to the core part of SNOVA [25].

The attacker is trying to find a vector $\overrightarrow{\mathbf{y}} \in \Big([L_1](T^{-1}\mathcal{O})\Big) \cap \Big([L_2](T^{-1}\mathcal{O})\Big)$ where $[L_1], [L_2]$ are two invertible linear combinations of the matrices $[P_i]$'s of size $ln \times ln$ over $\mathbb{F}_q$. Then, since $[L_1]^{-1}\overrightarrow{\mathbf{y}}$, $[L_2]^{-1}\overrightarrow{\mathbf{y}} \in T^{-1}(\mathcal{O})$, we have

$$\begin{cases} \left([L_1]^{-1}\overrightarrow{\mathbf{y}}\right)^t \cdot \left([D]^j [P_i] [D]^k\right) \cdot \left([L_1]^{-1}\overrightarrow{\mathbf{y}}\right) = 0 \\ \left([L_1]^{-1}\overrightarrow{\mathbf{y}}\right)^t \cdot \left([D]^j [P_i] [D]^k\right) \cdot \left([L_2]^{-1}\overrightarrow{\mathbf{y}}\right) = 0 \\ \left([L_2]^{-1}\overrightarrow{\mathbf{y}}\right)^t \cdot \left([D]^j [P_i] [D]^k\right) \cdot \left([L_1]^{-1}\overrightarrow{\mathbf{y}}\right) = 0 \\ \left([L_2]^{-1}\overrightarrow{\mathbf{y}}\right)^t \cdot \left([D]^j [P_i] [D]^k\right) \cdot \left([L_2]^{-1}\overrightarrow{\mathbf{y}}\right) = 0 \end{cases} \tag{5.32}$$

The case v < 2o. Since $\dim\Big([L_1](T^{-1}\mathcal{O})\Big) \cap \Big([L_2](T^{-1}\mathcal{O})\Big) \geq 2lo - lv > 0$, then the system 5.32 reduces to a homogeneous quadratic system of $M = 4l^2o - 2l$ equations in $N = ln - (2lo - lv - 1) = 2lv - lo + 1$ variables. Hence the complexity is

$$\text{Comp}_{\text{Intersection}}\text{SNOVA} = HMQ(N, M, q) \tag{5.33}$$

field multiplications for classical attacker.

The case $\mathbf{v \geq 2o}$. If $n \geq 3m$, then there is no guarantee that the intersection $\Big([P'_i](T^{-1}\mathcal{O})\Big) \cap \Big([P'_j](T^{-1}\mathcal{O})\Big)$ will exist. Therefore, the intersection attack becomes a probabilistic attack against SNOVA. In this case, the complexity is

$$\mathrm{Comp}_{\mathrm{Intersection}}\mathrm{SNOVA} = \min q^{lv-2lo+1} \cdot q^k \cdot MQ(N-k+1, M, q) \quad (5.34)$$

field multiplications where $N = ln, M = l^2o - 2l$ for the classical case.

6 Implementation and Parameters

In [31], NIST suggested several security levels for post-quantum cryptosystem design. In the new call for additional digital signature scheme project, NIST slightly modified their security level request. In this section, we propose our parameters aiming at three security levels in the new call of NIST PQC project [32] levels I, III and V, respectively.

6.1 NIST Security Level

Herein, We focus on levels I, III, and V. The NIST security levels I, III and V require that a classical attacker needs 2^{143}, 2^{207} and 2^{272} classical gates to break the scheme, and 2^{61}, 2^{125} and 2^{189} gates for a quantum attacker, respectively.

The number of gates required for an attack against digital signature scheme can be computed by

$$\sharp\text{gates} = \sharp\text{field multiplication} \cdot (2 \cdot (\log_2 q)^2 + \log_2 q) \quad (6.1)$$

with the assumption that one field multiplication in the field $\mathbb{F}_q$ needs about $(\log_2 q)^2$ bit multiplications and same for bit additions and, for each field multiplication in the computation, an addition of field elements taking $\log_2 q$ bit additions.

6.2 Proposed Parameter Sets

In this section, we give our proposed parameters and the corresponding sizes of public key and signature respectively. Finally, the comparison table of

SNOVA with NIST finalists [24, 29, 38] is given.

The following table shows the complexity of respective attacks against our parameters. "Dir.", "KS.", "Rec.", "Int." and "Col." denote direct attack in Sec. 5.3.1, KS attack in Sec. 5.5.1, Reconciliation attack in Sec. 5.5.2, intersection attack in Sec. 5.5.3 and the collision attack in Sec. 5.3.2, respectively.

Table 4: Table of complexity in $\log_2(\sharp\text{gates})$. In any pair of complexity, the left one denotes the complexity in classical gates and the right one denotes in quantum gates, respectively. The lowest complexity is marked in bold fonts. The complexity of direct attack against a quantum attacker is given by the estimation 5.6.

SL	(v, o, q, l)	Dir.	KS.	Rec.	Int.	Col.
I	$(37, 17, 16, 2)$	165/123	165/85	203	153	**151**
	$(25, 8, 16, 3)$	171/126	209/107	200	221	**159**
	$(24, 5, 16, 4)$	184/134	309/157	269	353	**175**
III	$(56, 25, 16, 2)$	234/173	253/129	297	221	**215**
	$(49, 11, 16, 3)$	226/162	461/233	438	529	**213**
	$(37, 8, 16, 4)$	287/214	469/237	387	506	**271**
V	$(75, 33, 16, 2)$	302/222	341/173	389	288	**279**
	$(66, 15, 16, 3)$	302/220	617/311	574	690	**285**
	$(60, 10, 16, 4)$	350/255	805/405	695	922	**335**

The key-size and the length of the signature are shown in Table 5.

Table 5: Table of key-sizes and lengths of the signature for SNOVA parameter settings. Herein, the notation Size_{pk} denotes the public key size and Size_{sig} denotes the signature size.

Security Level	(v, o, q, l)	Size_{pk} (Bytes)	Size_{sig} (Bytes)
I	$(37, 17, 16, 2)$	9826	108(+16)
	$(25, 8, 16, 3)$	2304	148.5(+16)
	$(24, 5, 16, 4)$	1000	232(+16)
III	$(56, 25, 16, 2)$	31250	162(+16)
	$(49, 11, 16, 3)$	5989.5	270(+16)
	$(37, 8, 16, 4)$	4096	360(+16)
V	$(75, 33, 16, 2)$	71874	216(+16)
	$(66, 15, 16, 3)$	15187.5	364.5(+16)
	$(60, 10, 16, 4)$	8000	560(+16)

Table 6 gives the comparison of SNOVA of 3 sets of parameters with those NISTPQC signature finalists that aim at the security level I. Based on the public key sizes and signature sizes of SNOVA, we consider SNOVA to be a competitive signature scheme. Note that the 16 Bytes **salt** is also indicated in the size of SNOVA signature.

Table 6: A comparison table of SNOVA with the NISTPQC signature finalists aims at NIST security level I.

Signature Scheme	Size of public key (Bytes)	Size of signature (Bytes)
Dilithium-2 [29]	1312	2420
Falcon-512 [38]	897	666
SPHINCS$^+$-128s [24]	32	7856
SPHINCS$^+$-128f [24]	32	17088
SNOVA$(24, 5, 16, 4)$	1000	232(+16)
SNOVA$(25, 8, 16, 3)$	2304	148.5(+16)
SNOVA$(37, 17, 16, 2)$	9826	108(+16)

In [45, 46], they both pointed out that the protocol TLS, which we used to protect our web browsing, is no longer secure due to the impact of the quantum computer. Making TLS post-quantum is an important task, but such a fundamental change could take years and be quite costly if we do not have a quantum-resistant signature that is relatively well compatible with the existing framework. Note that [46] gives the corresponding condition: six times signature size and two times of public key size fit in 9KB. According to the specification of SNOVA, SNOVA could be a more practical general-purpose signature scheme than others.

7 Conclusion

SNOVA has shown that multivariate signature schemes over noncommutative rings could be beneficial to security and key size reduction. With tremendous efforts on security analysis, to our best, we are confident that the SNOVA scheme is capable of resisting all known attacks for multivariate cryptosystems. By comparison with other post-quantum signature schemes, SNOVA is a practical secure signature scheme which is relatively efficient on both public key size and signature size.

Acknowledgement

The first author would like to express thanks to Prof. Jintai Ding for his encouragement. The first author and the second author are partially supported by MOST110-2122-M-259-001. The third author is partially supported by MOST111-2115-M-259-001-MY2.

The authors would like to thank Prof. Yasuhiko Ikematsu, Prof. Rika Akiyama for their friendly communication and discussions and the referee(s) of this paper for their kind review.

References

[1] Bardet, M., Faugère, J. C., Salvy, B., Yang, B. Y.: **Asymptotic behavior of the index of regularity of quadratic semi-regular polynomial systems.** In 8th Interna- tional Symposium on Effective Methods in Algebraic Geometry (MEGA), pp. 1–14 (2005).

[2] Bettale, L., Faugère, J.-C., Perret, L.: **Hybrid approach for solving multivariate systems over finite fields.** Journal of Mathematical Cryptology 3, pp. 177–197 (2009).

[3] Beullens, W.: **Breaking Rainbow Takes a Weekend on a Laptop.** Cryptology ePrint Archive, Report 2022/214, 2022. https://eprint.iacr.org/2022/214.pdf.

[4] Beullens, W.: **Improved cryptanalysis of UOV and Rainbow.** Cryptology ePrint Archive, Report 2020/1343, 2020. https://eprint.iacr.org/2020/1343.pdf.

[5] Beullens, W.: **MAYO: Practical Post-Quantum Signatures from Oil-and-Vinegar Maps.** Cryptology ePrint Archive, Report 2021/1144, 2021. https://eprint.iacr.org/2021/1144.pdf.

[6] Beullens, W., Campos, F., Celi, S., Hess, B., Kannwischer, M. J.: **MAYO.** https://pqmayo.org/assets/specs/mayo.pdf (version at June 1, 2023).

[7] Bosma, W., Cannon, J., Playoust, C.: **The Magma algebra system. I. The user language.** Journal of Symbolic Computation 24(3-4), pp. 235–265 (1997)

[8] Bouillaguet, C., Chen, H.C., Cheng, C.M., Chou, T., Niederhagen, R., Shamir, A., Yang., B.Y.: **Fast exhaustive search for polynomial systems in $\mathbb{F}_2$.** In Stefan Mangard and François-Xavier Standaert, editors, CHES 2010, volume 6225 of LNCS, pages 203–218, Santa Barbara, CA, USA, August 17–20, 2010. Springer, Heidelberg, Germany.

[9] Cheng, C.M., Chou, T., Niederhagen, R., Yang, B.Y.: **Solving quadratic equations with XL on parallel architectures.** In Emmanuel Prouff and Patrick Schaumont, editors, CHES 2012, volume 7428 of LNCS, pages 356–373. Springer, Heidelberg, September 2012.

[10] Cheng, C.M., Hashimoto, Y., Miura, H., Takagi, T.: **A polynomial-time algorithm for solving a class of underdetermined multivariate quadratic equations over fields of odd characteristics**. In PQCrypto'14, LNCS 8772 (2014), pp.40–58.

[11] Courtois, N., Goubin, L., Meier, W., Tacier, J.-D.: **Solving underdefined systems of multivariate quadratic equations**. In PKC'02, LNCS 2274 (2002), pp.211–227.

[12] Courtois, N., Klimov, A., Patarin, J., Shamir, A.: **Efficient algorithms for solving overdefined systems of multivariate polynomial equations.** In Bart Preneel, editor, EUROCRYPT 2000, volume 1807 of LNCS, pages 392–407, Bruges, Belgium, May 14–18, 2000. Springer, Heidelberg, Germany.

[13] Ding, J., Chen, M.S., Kannwischer, M., Patarin, J., Petzoldt, A., Schmidt, D., Yang, B.Y.: **Rainbow. NIST Post-Quantum Cryptography Standardization Round 3 Submissions**, available at https://csrc.nist.gov/Projects/post-quantum-cryptography/round-3-submissions

[14] Ding, J., Schmidt, D.: **Rainbow, a new multivariable polynomial signature scheme.** In International Conference on Applied Cryptography and Network Security, pages 164–175. Springer, 2005.

[15] Ding, J., Schmidt, D.: **Solving Degree and Degree of Regularity for Polynomial Systems over a Finite Fields.** In: Fischlin, M., Katzenbeisser, S. (eds) Number Theory and Cryptography. Lecture Notes in Computer Science, vol 8260. Springer, Berlin, Heidelberg, 2013. https://doi.org/10.1007/978-3-642-42001-6_4.

[16] Ding, J., Yang, B.Y., Chen, C.-O., Chen, M., Cheng, C.: **New differential-algebraic attacks and reparametrization of Rainbow.** In: ACNS 2008, LNCS, vol. 5037, pp. 242–257. Springer (2008).

[17] Faugère, J.C.: **A new efficient algorithm for computing Gröbner bases (F4).** Journal of Pure and Applied Algebra, 139:61–88 (1999).

[18] Faugère, J.C.: **A new efficient algorithm for computing Gröbner bases without reduction to zero (F5).** In Proceedings of the 2002 international symposium on Symbolic and algebraic computation, pages 75–83, 2002.

[19] Furue, H., Ikematsu, Y., Kiyomura, Y., Takagi, T. (2021). **A New Variant of Unbalanced Oil and Vinegar Using Quotient Ring: QR-UOV.** In: Tibouchi, M., Wang, H. (eds) Advances in Cryptology – ASIACRYPT 2021. ASIACRYPT 2021. Lecture Notes in Computer Science(), vol 13093. Springer, Cham. https://doi.org/10.1007/978-3-030-92068-5_7.

[20] Furue, H., Nakamura, S., Takagi, T.: **Improving Thomae-Wolf algorithm for solving underdetermined multivariate quadratic polynomial problem**. In PQC'21, LNCS 12841 (2021), pp.65–78.

[21] Garey, M.-R., Johnson, D.-S.: **Computers and intractability: a guide to the theory of NP-completeness.** W. H. Freeman (1979).

[22] Grover, L.-K.: **A fast quantum mechanical algorithm for database search.** In STOC 1996, pp. 212–219. ACM (1996).

[23] Hashimoto, Y.: **An improvement of algorithms to solve under-defined systems of multivariate quadratic equations.** Available at https://eprint.iacr.org/2021/1045.pdf.

[24] 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+.** Technical report, National Institute of Standards and Technology, 2020. available at https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[25] Ikematsu, Y., Akiyama, R.: **Revisiting the security analysis of SNOVA.** Cryptology ePrint Archive. Available at https://eprint.iacr.org/2024/096.pdf

[26] Kipnis, A., Patarin, J., Goubin, L.: **Unbalanced oil and vinegar signature schemes.** In Jacques Stern, editor, EUROCRYPT'99, volume 1592 of LNCS, pages 206–222. Springer, Heidelberg, May 1999.

[27] Kipnis, A., Shamir, A.: **Cryptanalysis of the oil and vinegar signature scheme.** In Hugo Krawczyk, editor, CRYPTO'98, volume 1462 of LNCS, pages 257–266. Springer, Heidelberg, August 1998.

[28] Li, P., Ding, J.: **Cryptanalysis of the SNOVA signature scheme.** Cryptology ePrint Archive. Available at https://eprint.iacr.org/2024/110.pdf

[29] Lyubashevsky, V., Ducas, L., Kiltz, E., Lepoint, T., Schwabe, P., Seiler, G., Stehlè, D., Bai, S.: **CRYSTALS-DILITHIUM.** Technical report, National Insti-tute of Standards and Technology, 2020. available at https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[30] Miura, H., Hashimoto, Y., Takagi, T.: **Extended algorithm for solving underdefined multivariate quadratic equations.** In PQCryoto'13, LNCS 7932 (2013), pp.118–135.

[31] NIST: **Post-quantum cryptography CSRC.** Available at https://csrc.nist.gov/Projects/post-quantum-cryptography/post-quantum-cryptography-standardization

[32] NIST: **Post-Quantum Cryptography: Digital Signature Schemes.** Available at https://csrc.nist.gov/projects/pqc-dig-sig/standardization/call-for-proposals

[33] NIST: **Submission Requirements and Evaluation Criteria for the Post-Quantum Cryptography Standardization Process.** Available at https://csrc.nist.gov/CSRC/media/Projects/Post-Quantum-Cryptography/documents/call-for-proposals-final-dec-2016.pdf

[34] Park, C.M.: **Cryptanalysis of Matrix-based UOV.** In Finite Fields and Their Applications, Volume 50, 2018, Pages 209-221, ISSN 1071-5797, https://doi.org/10.1016/j.ffa.2017.11.012.

[35] Patarin, J.: **The oil and vinegar signature scheme.** In Dagstuhl Workshop on Cryptography September, 1997.

[36] Petzoldt, A.: **Selecting and reducing key sizes for multivariate cryptography.**

[37] Petzoldt, A., Thomae, E., Bulygin, S., Wolf, C.: **Small public keys and fast verification for Multivariate Quadratic public key systems.** In Bart Preneel and Tsuyoshi Takagi, editors, CHES 2011, volume 6917 of LNCS, pages 475–490, Nara, Japan, September 28–October 1, 2011. Springer, Heidelberg, Germany.

[38] Prest, T., Fouque, P. A., Hoffstein, J., Kirchner, P., Lyubashevsky, V., Pornin, T., Ricosset, T., Seiler, G., Whyte, W., Zhang, Z.: **FALCON.** Technical report, National Institute of Standards and Technology, 2020. available at https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions.

[39] Tan, Y., Tang, S.: **Two Approaches to Build UOV Variants with Shorter Private Key and Faster Signature Generation.** In: Lin, D., Wang, X., Yung, M. (eds) Information Security and Cryptology. Inscrypt 2015. Lecture Notes in Computer Science(), vol 9589. Springer, Cham. https://doi.org/10.1007/978-3-319-38898-4_4.

[40] Thomae, E.: **Quo Vadis Quaternion? Cryptanalysis of Rainbow over non-commutative rings.** In SCN'12, Lect. Notes Comput. Sci. 7485, pp.361–363, 2012.

[41] Thomae, E., Wolf, C.: **Solving underdetermined systems of multivariate quadratic equations**, revisited. In PKC'12, LNCS 7293 (2012), pp.156–171.

[42] Wang, L.C., Hu, Y.H., Lai, F., Chou, C.Y., Yang, B.Y.: **Tractable rational map signature.** In PKC, Serge Vaudenay, ed., Public Key Cryptography — PKC 2005, (2005), pages 244–257. ISBN 3-540-24454-9.

[43] Wang, L.C., Wei, T.J., Shih, J.M., Hu, Y.H., Hsieh, C.C.:**An algorithm for solving over-determined multivariate quadratic systems over finite fields.** doi: 10.3934/amc.2022001

[44] Wiedemann, D.: **Solving sparse linear equations over finite fields.** IEEE Trans. Inf. Theory IT-32, pp. 54-62, 1986.

[45] Wiggers, T.: **Making protocols post-quantum.** In the Cloudflare blog. Available at https://blog.cloudflare.com/making-protocols-post-quantum/

[46] Westerbaan, B.: **Sizing Up Post-Quantum Signatures.** In the Cloudflare blog. Available at https://blog.cloudflare.com/sizing-up-post-quantum-signatures/

[47] Yasuda, T., Sakurai, K., Takagi, T.: **Reducing the Key Size of Rainbow Using Non-Commutative Rings.** In CT-RSA, volume 7178 of Lecture Notes in Computer Science, pages 68-83. Springer, 2012.