

A Simple Noncommutative UOV Scheme

Lih-Chung Wang^{*1}, Po-En Tseng^{†1}, Yen-Liang Kuan^{‡1}, and Chun-Yen Chou^{§1}

¹Department of Applied Mathematics, National Dong Hwa University,
No. 1, Sec. 2, Da Hsueh Rd., Shoufeng, Hualien 974301, 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, multivariate cryptosystem, digital signature, UOV, ring UOV, SNOVA

1 Introduction

The Unbalanced Oil and Vinegar (UOV) signature scheme [25] is a slight modification of the Oil and Vinegar (OV) [33] signature scheme, proposed by Patarin in 1997.

*Corresponding author: Lih-Chung Wang, Email: lcwang@gms.ndhu.edu.tw

†briantseng0320@gmail.com

‡ylkuan@gms.ndhu.edu.tw

§choucy@gms.ndhu.edu.tw

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 [40, 14, 5].

On the other hand, fundamental public key compression methods have been proposed. A. Petzoldt [34, 35] 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 $O(m^3 \cdot \log q)$. For the modern parameters of UOV which aiming at NIST security level I [30], 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 non-commutative 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 [34] 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 settings aiming for NIST security levels I, III, and V. For security level I, one of our parameter settings 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 non-commutative rings, we hope to open up new possibilities for designing signature schemes.

2 Preliminaries

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

Table 1: The table of symbols fixed with specific meaning in this paper.

Symbol	Description
\mathbb{F}_q	finite field of order q
\mathcal{R}	$\text{Mat}_{l \times l}(\mathbb{F}_q)$, matrix ring consisting of $l \times l$ matrices over \mathbb{F}_q
v	number of vinegar variables
o	number of oil variables
S	symmetric matrix in \mathcal{R} with its characteristic polynomial irreducible over \mathbb{F}_q
$n = v + o$	number of variables
$m = o$	number of equations
$F = [F_1, \dots, F_m]$	central map of the ring UOV scheme
$[F_i]$	matrix corresponding to F_i in F
$\tilde{F} = [\tilde{F}_1, \dots, \tilde{F}_m]$	central map of the SNOVA scheme
T	invertible linear map in signature scheme
$[T]$	matrix corresponding to T
$[T^{-1}]$	matrix corresponding to the map T^{-1}
$P = [P_1, \dots, P_m]$	public map of the ring UOV scheme
$[P_i]$	matrix corresponding to P_i in P
$\tilde{P} = [\tilde{P}_1, \dots, \tilde{P}_m]$	public map of the SNOVA scheme
D	document to be signed
$\text{Hash}(D)$	hash value of the document D
\mathcal{O}	oil space
$\text{MQ}(N, M, q)$	complexity of an MQ system of M equations in N variables over \mathbb{F}_q

Table 2: The table of conventions on notations in this paper.

Description	The font denoted with	Example
Integers	lower case letters	n, m and l
Elements in \mathcal{R}	upper case letters	A, S and Q
Variables over \mathcal{R}	upper case letters	X_1, \dots, X_n
Elements in \mathbb{F}_q	lower case letters	a_0, \dots, a_{l-1}
Variables over \mathbb{F}_q	lower case letters	x_1, \dots, x_n
Vectors of any dimension	boldface letters with an arrow on top	\vec{X} and \vec{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 $P_{i,jk}$
Block form of matrices $[T]$	upper case letters	$[T] = \begin{bmatrix} I^{11} & T^{12} \\ 0 & I^{22} \end{bmatrix}$
Block form of matrices $[F_i]$	upper case letters	$[F_i] = \begin{bmatrix} F_i^{11} & F_i^{12} \\ F_i^{21} & 0 \end{bmatrix}$
Block form of matrices $[P_i]$	upper case letters	$[P_i] = \begin{bmatrix} P_i^{11} & P_i^{12} \\ P_i^{21} & P_i^{22} \end{bmatrix}$

2.1 Basic Notions

MQ problem. Let \mathbb{F}_q be a finite field of order q . Given a multivariate quadratic map $P(\vec{x}) = [P_1(\vec{x}), \dots, P_M(\vec{x})]$ of M components in N variables $\vec{x} = (x_1, \dots, x_N)$ and a vector $\vec{y} \in \mathbb{F}_q^M$, to find a vector $\vec{u} \in \mathbb{F}_q^N$ such that $P(\vec{u}) = [P_1(\vec{u}), \dots, P_M(\vec{u})] = \vec{y}$. 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 finite fields such as F_4 [17], F_5 [18] and XL variants [12, 41].

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

$$P'(\vec{x}, \vec{y}) = [P'_1(\vec{x}, \vec{y}), \dots, P'_M(\vec{x}, \vec{y})] \quad (2.1)$$

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

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

Note that each P'_i is symmetric and bilinear. 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

$$[P'_i] = [P_i] + [P_i]^t. \quad (2.3)$$

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 $F = [F_1, \dots, F_m] : \mathbb{F}_q^n \rightarrow \mathbb{F}_q^m$ where each F_i is of the form

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

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

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

Oil space, \mathcal{O} . The central map F of UOV scheme vanishes on the linear space $\mathcal{O} = \{\vec{\mathbf{x}} \in \mathbb{F}_q^n : x_1 = \dots = x_v = 0\}$ 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, 45].

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

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

where “ \setminus ” denotes the set subtraction operation.

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

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

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, \dots, X_v are called the vinegar variables and X_{v+1}, \dots, 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 $\mathcal{R} = \text{Mat}_{l \times l}(\mathbb{F}_q)$ 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 $i \in \{1, 2, \dots, m\}$, of the central map $F = [F_1, \dots, F_m] : \mathcal{R}^n \rightarrow \mathcal{R}^m$ becomes

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

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

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

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

$$[F_i] = [F_{i,jk}] = \begin{bmatrix} F_i^{11} & F_i^{12} \\ F_i^{21} & 0 \end{bmatrix}, \quad (3.5)$$

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 $P = [P_1, \dots, P_m]$ is the composition of central map F and an invertible ring linear map $T : \mathcal{R}^n \rightarrow \mathcal{R}^n$, 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 $i \in \{1, 2, \dots, m\}$. The map T is defined by its matrix representation

$$[T] = \begin{bmatrix} I^{11} & T^{12} \\ 0 & I^{22} \end{bmatrix}. \quad (3.6)$$

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 $i \in \{1, \dots, m\}$, we have

$$P_i(\vec{\mathbf{U}}) = (F_i \circ T)(\vec{\mathbf{U}}) = \vec{\mathbf{U}}^t \left([T]^t [F_i] [T] \right) \vec{\mathbf{U}}. \quad (3.7)$$

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

$$[P_i] = [P_{i,d_j d_k}] = [T]^t [F_i] [T] \quad (3.8)$$

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

Some notes on a (v, o, q, l) ring UOV. Since we are working with the $l \times l$ matrix ring \mathcal{R} , we can observe that the components of ring map F_i will give us $l^2 o$ polynomials over the ring variables' entry elements in \mathbb{F}_q . Therefore, a (v, o, q, l) ring UOV not only can be considered as a UOV-like signature scheme over the matrix ring \mathcal{R} but also an $(l^2 v, l^2 o, q)$ UOV over \mathbb{F}_q .

However, when we consider a (v, o, q, l) ring UOV as an $(l^2 v, l^2 o, q)$ UOV over \mathbb{F}_q , through the computations illustrated below, we can observe that the central map and public map of this special UOV become sparse (coefficients of many quadratic terms are zero). To see this, say $l = 2$, we consider a ring coefficient $H \in \mathcal{R}$ and two ring variables X and Y . Then,

$$X^t H Y \quad (3.9)$$

$$= \begin{bmatrix} x_1 & x_3 \\ x_2 & x_4 \end{bmatrix} \begin{bmatrix} h_1 & h_2 \\ h_3 & h_4 \end{bmatrix} \begin{bmatrix} y_1 & y_2 \\ y_3 & y_4 \end{bmatrix} \quad (3.10)$$

$$= \begin{bmatrix} h_1 x_1 y_1 + h_2 x_1 y_3 + h_3 x_3 y_1 + h_4 x_3 y_3 & h_1 x_1 y_2 + h_2 x_1 y_4 + h_3 x_3 y_2 + h_4 x_3 y_4 \\ h_1 x_2 y_1 + h_2 x_2 y_3 + h_3 x_4 y_1 + h_4 x_4 y_3 & h_1 x_2 y_2 + h_2 x_2 y_4 + h_3 x_4 y_2 + h_4 x_4 y_4 \end{bmatrix} \quad (3.11)$$

gives us $4 = 2^2$ quadratic sparse polynomials over \mathbb{F}_q , with each having at most 4 out of 16 possible quadratic terms $x_i y_j$. As a result, the efficiency of forgery attacks targeting the public map might increase. To prevent such a possibility, we will make some modifications to eliminate such sparsity in the public map of ring UOV to get SNOVA.

On the other hand, for each $i \in \{1, \dots, l^2 o\}$, the i -th polynomial of the public map of the $(l^2 v, l^2 o, q)$ UOV vanishes on a linear space \mathcal{W}_i such that $\mathcal{O} \subseteq \mathcal{W}_i$ and $\dim \mathcal{O} \leq \dim \mathcal{W}_i$ (herein, \mathcal{O} denotes the oil space of this sparse $(l^2 v, l^2 o, q)$ UOV). However, the intersection of \mathcal{W}_i is still the oil space \mathcal{O} unless the coefficients of some monomial $x_j x_k$ vanish in all $l^2 o$ polynomials, which is very unlikely (we already check this phenomenon in our implementations). Therefore, we conclude that this $(l^2 v, l^2 o, q)$ UOV will not vanish on a linear space which is larger than the oil space \mathcal{O} . This observation provides us with some confidence in the security of this UOV against traditional key recovery attacks on oil space such as KS attack [26] and intersection attack [4] in the sense that this sparse UOV induced from ring UOV has the same size of oil space as the original UOV.

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 $\mathbb{F}_q = \text{GF}(16)$ 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

$$\mathbb{F}_q[S] = \{a_0 + a_1S + \cdots + a_{l-1}S^{l-1} : a_0, a_1, \cdots, a_{l-1} \in \mathbb{F}_q\} \quad (4.1)$$

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 $\tilde{F} = [\tilde{F}_1, \cdots, \tilde{F}_m] : \mathcal{R}^n \rightarrow \mathcal{R}^m$ and, for $i \in \{1, \cdots, m\}$, F_i is defined to be

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

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 $i \in \{1, \dots, m\}$, we define

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

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 $A_\alpha, B_\alpha, Q_{\alpha 1}$ and $Q_{\alpha 2}$), its ring coefficients can still be recorded by the matrix representation of its core part, i.e., the matrices

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

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

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

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

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 $i \in \{1, 2, \dots, m\}$, $\tilde{P}_i = \tilde{F}_i \circ T$. The relation $\vec{X} = [T] \cdot \vec{U}$ where $\vec{U} = (U_1, \dots, U_n) \in \mathcal{R}^n$ implies that

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

where $P_{i, d_j d_k} = \sum_{\Omega} T_{j, d_j} \cdot F_{i, j k} \cdot T_{k, d_k}$ 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

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

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

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

for $i \in \{1, \dots, m\}$. 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 for the core part of the public map \tilde{P} and the matrices $A_{\alpha}, B_{\alpha}, Q_{\alpha 1}$ and $Q_{\alpha 2}$ for $\alpha = 1, 2, \dots, l^2$, or simply the seed $\mathbf{s}_{\text{public}}$ which generates them. By utilizing matrices $\left[core(\tilde{P}_i) \right]$ and the seed $\mathbf{s}_{\text{public}}$, 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 $i = 1, 2, \dots, m$. Note that we can use the private seed $\mathbf{s}_{\text{private}}$ to generate T .

Signature. Let D be the document to be signed and $Hash(D) = \vec{Y} = (Y_1, \dots, Y_m) \in \mathcal{R}^m$ be its hash value. We compute the signature \vec{U} step by step. First, We assign

values to vinegar variables X_1, \dots, 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}, \dots, 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 $\vec{U} = T^{-1}(\vec{X}) \in \mathcal{R}^n$.

Verification. Let $\vec{U} = (U_1, \dots, U_n) \in \mathcal{R}^n$ be the signature to be verified. If $\text{Hash}(D) = \tilde{P}(\vec{U})$, 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 [34]. Note that, in SNOVA scheme, \mathbb{F}_q is of the characteristic 2.

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

$$[P_i] = [T]^t [F_i] [T]. \quad (4.9)$$

Then, we have the following

$$P_i^{11} = F_i^{11} \quad (4.10)$$

$$P_i^{12} = F_i^{11} T^{12} + F_i^{12} \quad (4.11)$$

$$P_i^{21} = (T^{12})^t F_i^{11} + F_i^{21} \quad (4.12)$$

$$P_i^{22} = (T^{12})^t \cdot (F_i^{11} T^{12} + F_i^{12}) + F_i^{21} T^{12}. \quad (4.13)$$

Therefore, to get $[P_i]$, we generate the matrices $[F_i]$, $[T]$ from a seed $\mathbf{s}_{\text{private}}$ at first and then compute $[P_i]$ for $i \in \{1, \dots, m\}$ 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 $i \in \{1, \dots, m\}$, and $[T]$ from two seeds $\mathbf{s}_{\text{public}}$ and $\mathbf{s}_{\text{private}}$ respectively. We also generate the matrices A_α , B_α , $Q_{\alpha 1}$ and $Q_{\alpha 2}$ for $\alpha = 1, 2, \dots, l^2$ from $\mathbf{s}_{\text{public}}$.

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

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

$$[F_i] = [T^{-1}]^t [P_i] [T^{-1}]. \quad (4.14)$$

Therefore, the following equations hold

$$F_i^{11} = P_i^{11} \quad (4.15)$$

$$F_i^{12} = P_i^{11}T^{12} + P_i^{12} \quad (4.16)$$

$$F_i^{21} = (T^{12})^t P_i^{11} + P_i^{21} \quad (4.17)$$

$$0 = F_i^{22} = (T^{12})^t \cdot (P_i^{11}T^{12} + P_i^{12}) + P_i^{21}T^{12} + P_i^{22}. \quad (4.18)$$

In other words, we then have

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

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

$$\text{Size}_{\text{SNOVA}} = m \cdot m^2 \cdot l^2 \quad (4.20)$$

field elements of \mathbb{F}_q . Note that the key size here does not include the size of the public seed $\mathbf{s}_{\text{public}}$ 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 assess their feasibility. The key point is that SNOVA is based on a quadratic form over ring, its private key T is shared with the ring UOV scheme corresponding to its core part whose structure is much simpler. To conduct a comprehensive and prudent security analysis, we start with the following observations.

Forgery attacks. 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 .

Key recovery attacks. 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. For key recovery attacks, the security of SNOVA will be evaluated by analyzing the complexity of such attacks against the associated ring UOV scheme which has a much simpler structure. To our best knowledge, we do not find a complete key recovery attack against the ring UOV corresponding to the core part. However, this ring UOV still induces a special UOV over field. We will point out some structures about this UOV scheme. We hope that the analysis of these structures will make the security analysis more comprehensive and enhance the understanding of the SNOVA scheme and the concept of ring UOV.

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 [12, 9, 41].

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

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

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

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

is non-positive.

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

$$HMQ(N, M, q) = q^k \cdot MQ(N-k, M, q) \quad (5.3)$$

field multiplications for the classical case and

$$q^{k/2} \cdot MQ(N-k, M, q) \quad (5.4)$$

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

Methods solving underdetermined MQ. On the other hand, several methods [39, 20, 23] 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 [39, 20, 23].)

Hence, the main term of complexity of solving MQ system under this technique is given by

$$\min_k q^k \cdot MQ(M - k - \alpha_k, M - \alpha_k, q) \quad (5.5)$$

field multiplications in the classical case and

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

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 [6], making it unable to meet NIST security levels. We want to point out that the parameter settings we have chosen still satisfies the security requirements. When solving an underdetermined MQ system, our complexity estimations consider the method with the lowest complexity.

Algorithms for super-underdetermined MQ. Note that, [26, 11, 28, 10] 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 settings 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 [31].

Signature. Let D be the document to be signed, we randomly choose **salt** and then generate a signature for the hash value $\vec{Y} = Hash(Hash(D)||\mathbf{salt})$. Therefore, the corresponding signature is of the form $\vec{\sigma} = (\vec{U}||\mathbf{salt})$ where \vec{U} is the signature of \vec{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 $P(\vec{U}) = Hash(Hash(D)||\mathbf{salt})$, 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.

5.3.1 Direct attack

For a quadratic multivariate polynomial system $P = [P_1, \dots, P_m]$ consisting of m equations in n variables over \mathbb{F}_q and an intended $\vec{y} \in \mathbb{F}_q^m$, an attacker can directly try to solve the solution \vec{u} of the system $P(\vec{u}) = \vec{y}$ algebraically with Gröbner basis approach such as [17, 18, 12, 9, 41]. We can assign values to $n - m$ variables in the system $P(\vec{u}) = \vec{y} = \text{Hash}(\text{digest}||\text{salt})$ randomly and then obtain an MQ system of m equations in n variables which can be solved with high probability. Once the system is solved, the solution \vec{u} will be a valid fake signature that satisfies $P(\vec{u}) = \vec{y}$.

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 $\mathcal{R} = \text{Mat}_{l \times l}(\mathbb{F}_q)$ yields l^2 equations over \mathbb{F}_q , the system over ring \mathcal{R} , $P(\vec{U}) = \vec{Y}$, 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.

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 F4/5 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 [42]. 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]

$$\text{Comp}_{\text{Direct}; \text{Classical}} \text{SNOVA} \tag{5.7}$$

$$= (l^2m - \alpha - k + 1) \text{HM}Q(\alpha, \alpha, q) \tag{5.8}$$

$$+ q^k (\text{HM}Q(\alpha - 1, \alpha - 1, q) + \text{HM}Q(l^2m - \alpha - k, l^2m - \alpha, q)). \tag{5.9}$$

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

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 $\mathcal{R} = \text{Mat}_{l \times l}(\mathbb{F}_q)$ behaves like a random systems consisting of $l^2 m$ equations in $l^2 n$ variables over a \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

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 \vec{U}_j where $j = 1, \dots, M$, and N hash values $\text{Hash}(\text{digest}||\text{salt}_k)$ where $k = 1, \dots, N$, whether there exists a collision $P(\vec{U}_j) = \text{Hash}(\text{digest}||\text{salt}_k)$. 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

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

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

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

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

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

and the collision exists with probability

$$1 - \left(\frac{q^{l^2 m} - M}{q^{l^2 m}} \right)^N = 1 - \left(\frac{MN - M}{MN} \right)^N \quad (5.13)$$

$$= 1 - \left(1 - \frac{1}{N} \right)^N \quad (5.14)$$

$$\approx 1 - e^{(-\frac{1}{N})N} \quad (5.15)$$

$$= 1 - e^{-1}. \quad (5.16)$$

5.4 Key Recovery Attacks

In this subsection, we will analyze the structure of ring UOV corresponding to the core part of SNOVA. Recall that when we consider this ring UOV as a UOV over field, its oil space is same as the original case. (We have verified this on some small parameter settings through some experiments.) In other words, for the key recovery attacks against the UOV induced from the ring UOV, we consider the situation as the same as the original UOV.

5.4.1 Quadratic forms over ring

No multi-layer structure. One may worry that the sparsity of the UOV induced by the ring UOV corresponding to SNOVA will lead to some structures that are prone to be broken, such as, multi-layer structure [14]. In [3, 4], Beullens proposed a series of MinRank attack against Rainbow [14] scheme based on its multi-layer structure. Such multi-layer structure will result in nested structure of oil spaces [4] and the low-rankness can be used to find a vector in the linear space $T^{-1}(\mathcal{O})$ and hence an equivalent key.

Another attack that makes use of multi-layer structure is the MinRank attack proposed by Thomae [38] against NC-Rainbow [45] which is a variant of Rainbow based on

Quaternion ring over a finite field \mathbb{F}_q of characteristic 2. If an attacker regards an NC-Rainbow scheme as a Rainbow scheme over \mathbb{F}_q , then the rank of the corresponding matrix to the first layer of central map F of NC-Rainbow will be lower than original Rainbow. Therefore, this low-rankness makes the MinRank attacks taking advantage of multi-layer structure more efficient.

Recall that for each $i \in \{1, \dots, l^2 o\}$, although the i -th polynomial of the public map of the $(l^2 v, l^2 o, q)$ UOV induced by the ring UOV corresponding to the core part of SNOVA is sparse and it vanishes on a linear space \mathcal{W}_i containing oil space \mathcal{O} , there is no multi-layer structure among these linear subspace \mathcal{W}_i from direct computations and evaluations. Therefore, we can not regard the sparse $(l^2 v, l^2 o, q)$ UOV scheme induced by the ring UOV corresponding to the core part of SNOVA as a Rainbow scheme. Consequently, attacks [4, 3, 38] that rely on the multi-layer structure have no security impact on the UOV induced by the quadratic form of the ring UOV, and thus will not affect the security of SNOVA.

Intersection of the null spaces of public key differential. In [32], Park broke the Matrix-based UOV scheme [37] which is proposed by Tan and Tang. The central map of matrix-based UOV is constructed using rotation sequence or linearly recurring sequence and then result in a special shape. The main insight of this attack is: when an attacker regards the matrices of the differential of the central map and the public map of Matrix-based UOV as linear operators, the sparsity of some of these matrices makes the intersections of the corresponding null spaces non-trivial, while general UOV do not have this phenomenon. Then, by using this structure, Park showed that any basis of this non-trivial intersection can be used to build an equivalent private key.

Since the structure of the ring UOV corresponding to the core part is different from the central map of matrix-based UOV, the null spaces of the differential have different structure from those for the Matrix-based UOV. In fact, in our case, direct computation shows that the intersection of the null space corresponding to the public differential is the same as the plain UOV in actual experiments of small parameter sets. Therefore, the attack in [32] is not applicable to this ring UOV, and hence will not affect the security of SNOVA.

Reconciliation Attack. The reconciliation attack proposed by [16] against UOV is trying to find a vector $\vec{\sigma} \in T^{-1}(\mathcal{O})$ by solving the system $P(\vec{\sigma}) = 0$ and hence the basis of $T^{-1}(\mathcal{O})$ can be recovered. This implies that $P(\vec{\sigma}) = 0$ 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 $\vec{\sigma}$. 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 UOV, thus we are in the case of solving the quadratic system of $l^2 m$ equations in $l^2 v \geq l^2 m$ variables. Hence the reconciliation attack usually will not outperform the direct attack on the public map of SNOVA in which the complexity comes from solving an induced

system of l^2m quadratic equations in l^2m variables.

Sparsity. Although we may observe that sparsity seems to have negative impacts and weaknesses on the scheme, we still want to point out the differences in sparsity between our ring UOV and other schemes. The sparsity of other schemes often arises from the design of their central map. These schemes aim to gain some advantages by introducing such sparsity [37, 45]. However, the public map of ring UOV corresponding to the core part is not sparse when it is considered over \mathcal{R} . The sparsity of UOV induced by the ring UOV comes entirely from matrix multiplication. The public map and the central map of the sparse UOV both are sparse (both caused by matrix multiplication) which is very different from other schemes. Our sparsity will not provide more algebraic equations or relationships. On the other hand, even if it is sparse, we still believe that the rank is sufficiently high to meet security requirements. We will interpret this through the analysis of equivalent attack in the next section.

5.4.2 Equivalent key attack

An attacker may try to find the submatrix $(T^{-1})^{12}$ of matrix $[T^{-1}]$ in the top right corner by algebraic attacks. Once the matrix $[T^{-1}]$ is found, the central map F can be recovered. This can be done by considering the system $P(T^{-1}\vec{x}) = F(\vec{x})$ and solve for $[T^{-1}]$ by comparing both sides of equation at ring level. Then it induces a system of $m \cdot m^2 \cdot l^2$ quadratic equations in lvo variables over \mathbb{F}_q and hence can be solved by F_4, F_5 and XL, and the complexity is

$$\text{Comp}_{T^{-1}}\text{SNOVA} = MQ(lvo + 1, m^3l^2, q) \quad (5.17)$$

field multiplications. Note that the multivariate quadratic system constructed by this attack is overdetermined, hence [25, 11, 28, 10, 39, 20, 23] are not applicable.

On the other hand, one may consider executing equivalent key attack that regards a (v, o, q, l) ring UOV corresponding to the core part of SNOVA as an (l^2v, l^2o, q) UOV over \mathbb{F}_q , then inducing a quadratic system of $M = (l^2m) \cdot (l^2m) \cdot \frac{l^2m+1}{2}$ equations in $N = lvo$ variables over \mathbb{F}_q . However, our experiments show that this formulation does not increase the number of independent equations. With table 4 below, we hope to provide some information about this equivalent key attack against our parameters aiming that NIST security level I.

Table 4: Trend table of changes in degree of regularity.

(v, o, q, l)	N	M	d_{reg}
(28, 17, 16, 2)	952	19652	11
(25, 8, 16, 3)	600	4608	16
(24, 5, 16, 4)	480	2000	23

5.4.3 Kipnis-Shamir attack (UOV attack)

The KS attack [26] 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 [26] shows that $T^{-1}(\mathcal{O})$, the oil subspace of the public key P of UOV, induces an equivalent key.

In [26, 4], 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^{n-2m} 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} . From the design of central map F of the ring UOV corresponding to the core part of SNOVA and the noncommutativity of \mathcal{R} , there does not exist the notion of oil space of F over \mathcal{R} analogous to the space \mathcal{O} of UOV and hence the notion of $T^{-1}(\mathcal{O})$ in the sense that regarding $T^{-1}(\mathcal{O})$ as a left-module or a right-module over \mathcal{R} . Such a requirement is necessary for KS attack since to execute KS attack over \mathcal{R} , the consistency of multiplication over \mathcal{R} given by a left-module or a right-module over \mathcal{R} is needed. Therefore, KS attack is not applicable to SNOVA over \mathcal{R} . Note that [33] 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} .

However, an attacker may treat the ring UOV which is corresponding to the core part of a (v, o, q, l) SNOVA scheme over \mathcal{R} as an (l^2v, l^2o, q) UOV system over \mathbb{F}_q and then carry out the KS attack over \mathbb{F}_q .

Then we have

$$\text{Comp}_{\text{KS}; \text{classical}}^{\text{SNOVA}} = q^{l^2n-2l^2m} \quad (5.18)$$

field multiplications for classical attack and

$$\text{Comp}_{\text{KS}; \text{quantum}}^{\text{SNOVA}} = q^{(l^2n-2l^2m)/2} \quad (5.19)$$

field multiplications for quantum attack.

5.4.4 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, $P' = [P'_1, \dots, P'_m]$ with $P'_i(\vec{\mathbf{u}}_1, \vec{\mathbf{u}}_2) = \vec{\mathbf{u}}_1^t [P'_i] \vec{\mathbf{u}}_2$ where $[P'_i] = [P_i] + [P_i]^t$.

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

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

$$\begin{cases} P\left([P'_i]^{-1})\vec{\mathbf{y}}\right) = 0 \\ P\left([P'_j]^{-1})\vec{\mathbf{y}}\right) = 0 \\ P'\left([P'_i]^{-1})\vec{\mathbf{y}}, ([P'_j]^{-1})\vec{\mathbf{y}}\right) = 0 \end{cases} \quad (5.20)$$

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 $\vec{\mathbf{u}}_1^t [P'_i] \vec{\mathbf{u}}_2$, when considered as over \mathcal{R} . Thus, the implementation of intersection attack still face the non-commutativity, that is, there is no efficient algorithm like F_4 , F_5 and XL to compute. Therefore, from this perspective, to implement intersection attack against SNOVA, the only possible strategy is to regard the ring UOV corresponding to the core part of SNOVA as a UOV system over \mathbb{F}_q and then to solve a system over \mathbb{F}_q . Hence the complexity is estimated by the following

Whenever $\mathbf{n} < 2.5\mathbf{m}$. If $n < 2.5m$, we have $N = (l^2n)k - (2k - 1)(l^2m)$, $M = \binom{k+1}{2}(l^2m) - 2\binom{k}{2}$, and

$$\text{Comp}_{\text{Intersection}}^{\text{SNOVA}} = \text{MQ}(N + 1, M, q) \quad (5.21)$$

field multiplications.

Whenever $2.5\mathbf{m} < \mathbf{n} < 3\mathbf{m}$. In the case $2.5m < n < 3m$, $N = 2(l^2n) - 3(l^2m)$, $M = 3(l^2m)$, and

$$\text{Comp}_{\text{Intersection}}^{\text{SNOVA}} = \text{MQ}(N + 1, M, q) \quad (5.22)$$

field multiplications.

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

$$\text{Comp}_{\text{Intersection}}^{\text{SNOVA}} = q^{(l^2n) - 3(l^2m) + 1} \cdot \text{MQ}(N + 1, M, q) \quad (5.23)$$

field multiplications where $N = l^2n$, $M = 3(l^2m)$.

Our experiment shows that the quadratic system induced by intersection attack on ring UOV will not degenerate, e.g., in the toy example $(v, o, q, l) = (3, 2, 16, 2)$. That is, it behaves like a semi-regular system.

6 Implementation and Parameters

In [29], 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 [30] 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} quantum gates for a quantum attacker, respectively.

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

$$\#gates = \#field\ multiplication \cdot (2 \cdot (\log_2 q)^2 + \log_2 q) \quad (6.1)$$

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 Settings

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 [36, 27, 24] is given.

When it comes to parameter selection, our goal is to achieve a higher level of system security, with a preference for smaller public key size and signature size. In general, system security tends to increase with the increase of o and $n = v + o$. However, the public key size and signature size also increase with the increase of o and n . Therefore, we need to strike a balance and find smaller values of o and n that meet the security requirements.

The following table shows the complexity of respective attacks against our parameters. “Dir.”, “KS.”, “Int.”, “ T^{-1} .”, and “Col.” denote direct attack in Sec. 5.3.1, KS attack in Sec. 5.4.3, intersection attack in Sec. 5.4.4 and equivalent key attack in Sec. 5.4.2 and the collision attack in Sec. 5.3.2, respectively.

Table 5: Table of complexity in $\log_2(\#\text{gates})$. 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 quantum attacker is given by the estimation 5.6.

SL	(v, o, q, l)	Dir.	K-S.	Int.	T^{-1} .	Col.
I	(28, 17, 16, 2)	167/124	181/93	275	192/192	151
	(25, 8, 16, 3)	171/126	617/311	819	231/231	159
	(24, 5, 16, 4)	184/134	1221/613	1439	286/286	175
III	(43, 25, 16, 2)	236/175	293/149	439	279/ 279	215
	(49, 11, 16, 3)	226/162	1373/689	1631	530/530	213
	(37, 8, 16, 4)	287/214	1861/933	2192	424/424	271
V	(61, 33, 16, 2)	304/224	453/229	727	386/386	279
	(66, 15, 16, 3)	302/220	1841/923	2178	707/707	285
	(60, 10, 16, 4)	350/255	3205/1605	3602	812/812	335

The complexity of KS attack and intersection attack for each parameter set is already far beyond the security level. Therefore, we only include the remaining complexity in Table 5. The key-size and the length of the signature are shown in Table 6.

Table 6: Table of key-sizes and lengths of the signature of 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	(28, 17, 16, 2)	9826	90(+16)
	(25, 8, 16, 3)	2304	148.5(+16)
	(24, 5, 16, 4)	1000	232(+16)
III	(43, 25, 16, 2)	31250	136(+16)
	(49, 11, 16, 3)	5989.5	270(+16)
	(37, 8, 16, 4)	4096	360(+16)
V	(61, 33, 16, 2)	71874	188(+16)
	(66, 15, 16, 3)	15187.5	364.5(+16)
	(60, 10, 16, 4)	8000	560(+16)

Table 7 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 system. Note that the 16 Bytes **salt** is also indicated in the size of SNOVA signature.

Table 7: 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 [27]	1312	2420
Falcon-512 [36]	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(28, 17, 16, 2)	9826	90(+16)

In [43, 44], 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 [44] 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.

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 International 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., Westerman, 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] 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.
- [26] 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.
- [27] Lyubashevsky, V., Ducas, L., Kiltz, E., Lepoint, T., Schwabe, P., Seiler, G., Stehlè, D., Bai, S.: **CRYSTALS-DILITHIUM.** Technical report, National Institute of Standards and Technology, 2020. available at <https://csrc.nist.gov/projects/post-quantum-cryptography/round-3-submissions>.
- [28] Miura, H., Hashimoto, Y., Takagi, T.: **Extended algorithm for solving underdefined multivariate quadratic equations.** In PQCryoto'13, LNCS 7932 (2013), pp.118–135.
- [29] NIST: **Post-quantum cryptography CSRC.** Available at <https://csrc.nist.gov/Projects/post-quantum-cryptography/post-quantum-cryptography-standardization>
- [30] NIST: **Post-Quantum Cryptography: Digital Signature Schemes.** Available at <https://csrc.nist.gov/projects/pqc-dig-sig/standardization/call-for-proposals>
- [31] 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>
- [32] 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>.
- [33] Patarin, J.: **The oil and vinegar signature scheme.** In Dagstuhl Workshop on Cryptography September, 1997.

- [34] Petzoldt, A.: **Selecting and reducing key sizes for multivariate cryptography.**
- [35] 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.
- [36] 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>.
- [37] 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.
- [38] Thomae, E.: **Quo Vadis Quaternion? Cryptanalysis of Rainbow over non-commutative rings.** In SCN’12, Lect. Notes Comput. Sci. 7485, pp.361–363, 2012.
- [39] Thomae, E., Wolf, C.: **Solving underdetermined systems of multivariate quadratic equations, revisited.** In PKC’12, LNCS 7293 (2012), pp.156–171.
- [40] 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.
- [41] 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
- [42] Wiedemann, D.: **Solving sparse linear equations over finite fields.** IEEE Trans. Inf. Theory IT-32, pp. 54-62, 1986.
- [43] Wiggers, T.: **Making protocols post-quantum.** In the Cloudflare blog. Available at <https://blog.cloudflare.com/making-protocols-post-quantum/>
- [44] Westerbaan, B.: **Sizing Up Post-Quantum Signatures.** In the Cloudflare blog. Available at <https://blog.cloudflare.com/sizing-up-post-quantum-signatures/>
- [45] 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.