A Note on the SNOVA Security

SNOVA Team

Lih-Chung Wang, Chun-Yen Chou, Jintai Ding, Yen-Liang Kuan, Jan Adriaan Leegwater, Ming-Siou Li, Bo-Shu Tseng, Po-En Tseng*, Chia-Chun Wang

Abstract

SNOVA is one of the submissions in the NIST Round 1 Additional Signature of the Post-Quantum Signature Competition. SNOVA is a UOV variant that uses the noncommutative-ring technique to reduce the size of the public key. SNOVA's public key size and signature size are well-balanced and have good performance. Recently, Beullens proposed a forgery attack against SNOVA, pointing out that the parameters of SNOVA can be attacked. Beullens also argued that with some slight adjustments his attacks can be prevented. In this note, we explain Beullens' forgery attack and show that the attack can be invalid by two different approaches. Finally, we show that these two approaches do not increase the sizes of the public keys or signatures and the current parameters satisfy the security requirement of NIST.

Keywords: PQC, MQ, SNOVA

1 Introduction

SNOVA is a member of the NIST Additional Digital Signature Competition [7]. SNOVA has a relatively small public key and signature and has been performing well at the same time. Recently, there has been some discussion about the security of SNOVA [1, 3, 4, 5]. In this note, we would like to point out that SNOVA still meets the security requirements with minor adjustments.

In [3, 4, 5], the (lv, lo, q)-UOV or (v, o, q^l)-UOV structure was used to perform key recovery attacks on SNOVA. The security of the current SNOVA parameter sets against these attacks was confirmed in [5, 9]. In this note, we propose two different approaches in Section 3 and Section 4 to resist the Beullens attack. The adjustments made in both approaches do not affect the analyses presented in [3, 4, 5]. Therefore, the current parameter sets of SNOVA remain secure

*Corresponding author: Po-En Tseng, Email: briantseng0320@gmail.com

against key recovery attacks under the adjustments outlined in Section 3 and Section 4. As such, the primary focus of this note is on countering the Beullens attack.

Table 1: Table of current SNOVA parameter sets with key-sizes and lengths of the signature.

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)

2 Beullens Attack on SNOVA

In [1], Beullens interprets the public map of SNOVA with bilinear form formulation. Under this formulation, he discovered that if the $A_\alpha, B_\alpha, Q_{\alpha 1}, Q_{\alpha 2}$ matrices in SNOVA are generated randomly then the matrices $\mathbf{E}_{i,j}$ in the formulation may have low rank linear combination and this gives a forgery attack. The attack shows that the matrix $\mathbf{E}_{i,j} \in \mathbb{F}_q^{ml^2 \times ml^2}$ is block diagonal matrix with m identical blocks of size $l^2 \times l^2$ on the diagonals. Moreover, these $\mathbf{E}_{i,j}$ matrices are determined by the $A_\alpha, B_\alpha, Q_{\alpha 1}, Q_{\alpha 2}$ matrices and hence, as in Beullens' paper, this attack can be avoided with suitable adjustment on $A_\alpha, B_\alpha, Q_{\alpha 1}, Q_{\alpha 2}$.

We propose two approaches, both make the current parameters meet NIST security requirements [6, 7], do not change the public key and signature size, and maintain the original verification efficiency.

2.1 Notation

Let o and v denote the number of Oil and Vinegar variables. Let $n = o + v$ and $m = o$ denote the number of matrix variables and the number of matrix equations in the SNOVA public map. Let $\mathbb{F}_q$ be a finite field of order q. For a matrix $\mathbf{A}$ and a positive integer n, $\mathbf{A}^{\otimes n}$ denotes the block diagonal matrix with n copies of A on the block diagonal.

SNOVA public map is a multivariate quadratic map characterized with $l \times l$ matrix ring $\mathcal{R} = \text{Mat}_{l \times l}(\mathbb{F}_q)$ and a symmetric matrix S with irreducible characteristic polynomial. More details can be found in [8, 9].

2.2 Description

The attack uses the following formulation of SNOVA public map.

Bilinear formulation. For SNOVA public map, $P(\vec{\mathbf{U}}) = \left(P_1(\vec{\mathbf{U}}), \ldots, P_m(\vec{\mathbf{U}})\right) : \mathcal{R}^n \to \mathcal{R}^m$, it can be expressed as, for $i \in \{1, 2, \ldots, m\}$,

$$P_i(\vec{\mathbf{U}}) = \sum_{\alpha=1}^{l^2} \sum_{j=1}^{n} \sum_{k=1}^{n} A_\alpha \cdot U_j^t (Q_{\alpha 1} P_{i,jk} Q_{\alpha 2}) U_k \cdot B_\alpha \tag{2.1}$$

$$= \sum_{\alpha=1}^{l^2} A_\alpha \cdot \vec{\mathbf{U}}^t \cdot Q_{\alpha 1}^{\otimes n} \cdot [P_i] \cdot Q_{\alpha 2}^{\otimes n} \cdot \vec{\mathbf{U}} \cdot B_\alpha \tag{2.2}$$

where $[P_i]$ are the public keys of SNOVA, $Q_{\alpha 1}^{\otimes n}$ is a $n \times n$ diagonal matrix over $\mathcal{R}$ with identical blocks $Q_{\alpha 1}$ and similarly for $Q_{\alpha 2}^{\otimes n}$. Here, the vector $\vec{\mathbf{U}} = (U_1, \cdots, U_n)^t \in \mathcal{R}^n$ is a matrix of height nl and width l when we regard it as over $\mathbb{F}_q$.

In [1], it can be seen that

$$P(\vec{\mathbf{U}}) = \sum_{j=1}^{l} \sum_{j=1}^{l} \mathbf{E}_{j,k} \cdot \mathcal{B}(\mathbf{u}_j, \mathbf{u}_k) \tag{2.3}$$

where $\mathbf{E}_{j,k}$ is block diagonal matrix with identical blocks, i.e., $\mathbf{E}_{j,k} = \tilde{\mathbf{E}}_{j,k}^{\otimes m}$, $\tilde{\mathbf{E}}_{j,k}$ is an $l^2 \times l^2$ matrix determined by matrices $A_\alpha, B_\alpha, Q_{\alpha 1}, Q_{\alpha 2}$ and $\mathbf{u}_j$ is the j-th column of $\vec{\mathbf{U}}$. Here, $\mathcal{B} : \mathbb{F}_q^{nl} \times \mathbb{F}_q^{nl} \to \mathbb{F}_q^{ml^2}$ is a bilinear form defined by SNOVA public keys $[P_1], \ldots, [P_m]$ and the matrix S. In the equation (4.5) below we derive an explicit formula for $\mathbf{E}_{j,k}$ in terms of the A, B and Q matrices.

We then briefly describe the attack by Beullens. For other details, we refer to [1].

Attack. This attack attempts to forge a signature by solving for $\vec{\mathbf{U}}$ satisfy that the columns $\mathbf{u}_j = a_j u_1 + v_j$ where $v_j \in \mathbb{F}_q^{ln}$ is randomly chosen for all $j \in \{2, \ldots, l\}$, for some $a_1, \ldots, a_l \in \mathbb{F}_q$. Under the formulation (2.3), this implies that the quadratic part of public map $P(\vec{\mathbf{U}})$ is $\mathbf{E}_\alpha \cdot \mathcal{B}(\mathbf{u}_1, \mathbf{u}_1)$ where

$$\mathbf{E}_\alpha = \sum_{j=1}^{l} \sum_{j=1}^{l} a_j a_k \mathbf{E}_{jk}. \tag{2.4}$$

The attack is divided into three steps:

- Since $\mathbf{E}_{j,k} = \tilde{\mathbf{E}}_{j,k}^{\otimes m}$, the linear combination $\mathbf{E}_\alpha$ is also a block diagonal matrix of size $l^2 m \times l^2 m$ with identical $l^2 \times l^2$ blocks on diagonal. Therefore, if the linear combination of matrices $\tilde{\mathbf{E}}_{jk}$ have rank defect d then the corresponding linear combination $\mathbf{E}_\alpha$ will have rank defect md. This gives a generalized MinRank problem.

- Following with the first step, if $d = l^2 - r$ then we have $\text{rank}(\mathbf{E}_\alpha) = mr$. Therefore, for an attacker who wants to forge a fake signature, the remaining is to solving an MQ system of mr equations in $nl - m(l^2 - r)$ variables.
- Using the structure of $\mathbb{F}_q[S]$, the generalized MinRank problem in the first step can be extended to a generalized MinRank problem with $l(l-1)$ variables that trying to find the low rank of $\mathbf{E}_\mathbf{R}$ which is the quadratic part of $P(\vec{\mathbf{U}})$ under the setting that $u_j = \mathbf{R}_j^{\otimes n} u_1 + \mathbf{v}_j$ for $\mathbf{R}_j \in \mathbb{F}_q[S]$. This will allow the attackers to find matrices with lower rank. Hence, the number of variables in step 2 can be further reduced. Then, the attack becomes more efficient.

Complexity and Countermeasures. The complexity of Beullens attack is dominated by the cost of solving the MQ system of mr equations in $nl - m(l^2 - r)$ variables. This MQ system can be very underdetermined if the rank defect d at step 1 becomes larger, or equivalently, the rank defect of $\mathbf{E}_\mathbf{R} = \tilde{\mathbf{E}}_\mathbf{R}^{\otimes m}$ becomes lager where $\tilde{\mathbf{E}}_\mathbf{R}$ is an $l^2 \times l^2$ matrix induced by $A_\alpha, B_\alpha, Q_{\alpha 1}, Q_{\alpha 2}$ in SNOVA. We can observe that Beullens's attack is based on the rank defect of the matrix $\mathbf{E}_\mathbf{R}$. Therefore, we realize that we can make non-trivial linear combination of matrices $\mathbf{E}_{jk}$ all are of full rank or make the case that this rank defect is not enough to affect the security of the scheme. We believe that there are a number of minor mistakes in Beullens paper [1]. In particular the complexity formula of Hashimoto's algorithm mentioned by Beullens deviates from that in the original [2]. As a result, we have not been able to reproduce the complexity estimates. We do agree that the attack is valid and needs to be addressed.

In the following section, we propose two approaches to keep the security the current parameters of SNOVA.

3 Varying A, B and Q matrices in SNOVA

For Beullens attack, the key point is that when $\tilde{\mathbf{E}}_\mathbf{R}$ has rank defect d then $\mathbf{E}_\mathbf{R} = \tilde{\mathbf{E}}_\mathbf{R}^{\otimes m}$ has rank defect md. This rank defect is caused by the m identical blocks structure of $\mathbf{E}_\mathbf{R}$. Note that this identical blocks structure of $\mathbf{E}_\mathbf{R} = \tilde{\mathbf{E}}_\mathbf{R}^{\otimes m}$ comes from the fact that $P_1, \ldots, P_m$ share this set of matrices

$$\{A_1, \ldots, A_{l^2}, B_1, \ldots, B_{l^2}, \ldots, Q_{1,1}, \ldots, Q_{l^2,1}, Q_{2,1}, \ldots, Q_{l^2,2}\}. \tag{3.1}$$

Therefore, a direct countermeasure is to make $P_1, \ldots, P_m$ do not share the same set of A, B, Q matrices in SNOVA. In other words, for $i \in \{1, 2, \ldots, m\}$, we let

$$P_i(\vec{\mathbf{U}}) = \sum_{\alpha=1}^{l^2} \sum_{j=1}^{n} \sum_{k=1}^{n} A_{i,\alpha} \cdot U_j^t (Q_{i,\alpha 1} P_{i,jk} Q_{i,\alpha 2}) U_k \cdot B_{i,\alpha}. \tag{3.2}$$

Note that, the matrices $A_{i,\alpha}$, $B_{i,\alpha}$, $Q_{i,\alpha 1}$ and $Q_{i,\alpha 2}$ now are "varying" with index i. We call such adjustment to be the "A,B,Q varying". Since $A_{i,\alpha}$, $B_{i,\alpha}$,

$Q_{i,\alpha 1}$ and $Q_{i,\alpha 2}$ are generated from a random seed, this adjustment does not affect the public key size and signature size of SNOVA and does not have much influence on the efficiency of SNOVA .

The result is that the matrix $\mathbf{E_R} \in \mathbb{F}_q^{ml^2 \times ml^2}$ will no longer a block diagonal matrix with identical blocks but a block diagonal matrix with different diagonal blocks in general. The effectiveness of the Beullens attack comes from the fact that the MinRank problem in the first step significantly reduces the complexity of solving the MQ in the second step. More precisely, because every diagonal block of $\mathbf{E_R}$ is identical then the solution of MinRank problem of $\mathbf{E_R} = \tilde{\mathbf{E}}_{\mathbf{R}}^{\otimes m}$ share the same solution of MinRank problem of $\tilde{\mathbf{E}}_{\mathbf{R}}$ which is much smaller in size. However, it is different in the case of A, B, Q varying, the MinRank problem in the first step will become a MinRank problem of more general matrices due to the fact that now the blocks of $\mathbf{E_R}$ are different. This makes Beullens' attack much less effective. The following table records the lower bound of the rank of $\mathbf{E_R}$ that makes SNOVA current parameter set in Table 1 satisfy the NIST security requirements.

Table 2: Table of lower bound of the rank of $\mathbf{E_R}$ (in the generalized MinRank problem) that makes SNOVA current parameter set satisfy the NIST security requirements and the cost of corresponding MQ system in step 3 (in gates).

Security Level	(v, o, q, l)	lower bound of the rank	complexity of the corresponding MQ
I	$(37, 17, 16, 2)$	52	144
	$(25, 8, 16, 3)$	49	143
	$(24, 5, 16, 4)$	52	146
III	$(56, 25, 16, 2)$	79	208
	$(49, 11, 16, 3)$	83	210
	$(37, 8, 16, 4)$	78	210
V	$(75, 33, 16, 2)$	106	272
	$(66, 15, 16, 3)$	110	272
	$(60, 10, 16, 4)$	113	273

Note that the numbers in Table 2 coincide with the analysis in [1]. However, since $\mathbf{E_R}$ now is a block diagonal matrix with different diagonal blocks, the probability that rank of $\mathbf{E_R}$ will be lower than the lower bound in Table 2 is negligible in practice. Therefore, with P_i of the form (3.2), the minimal rank of $\mathbf{E_R}$ all are higher than the lower bound that makes SNOVA current parameter set satisfy the NIST security requirements in general. In other words, with A, B, Q varying, the current parameters of SNOVA in Table 1 remain its security. To demonstrate that the probability that rank of $\mathbf{E_R}$ will be lower than the lower bound in Table 2 is negligible in practice. With 10000 trials, we exhaustively search the solution of MinRank problem on the matrix $\mathbf{E_R}$ in the case of A, B, Q varying. The results are presented in the following Table 3, Table 4 and Table 5.

Table 3: Table of the distribution of the minimal rank of $l = 2$ parameter set for the SL I with 10000 trials.

(v, o, q, l)	lower bound in Table 2	minimal rank	No. of trials
$(37, 17, 16, 2)$	52	57	1
		58	2
		59	35
		60	101
		61	410
		62	1699
		63	5083
		64	2650
		65	19

Table 4: Table of the distribution of the minimal rank of $l = 2$ parameter set for the SL III with 10000 trials.

(v, o, q, l)	lower bound in Table 2	minimal rank	No. of trials
$(56, 25, 16, 2)$	79	87	5
		88	14
		89	35
		90	149
		91	360
		92	943
		93	2456
		94	4371
		95	1652
		96	15

Table 5: Table of the distribution of the minimal rank of $l = 2$ parameter set for the SL V with 10000 trials.

(v, o, q, l)	lower bound in Table 2	minimal rank	No. of trials
$(75, 33, 16, 2)$	106	116	5
		117	8
		118	22
		119	68
		120	159
		121	326
		122	696
		123	1339
		124	2535
		125	3556
		126	1267
		127	19

Note that, as q decreases, the rank drop tends to more in general. We compare the distribution of minimal rank of $\mathbf{E_R}$ on $(37, 17, 16, 2)$ and $(37, 17, 5, 2)$ parameter sets in Table 6 to demonstrate this phenomenon. With 10000 trials, the smallest minimal rank of $\mathbf{E_R}$ for $q = 5$ is 44, which is much lower than 57, the smallest minimal rank of $\mathbf{E_R}$ for $q = 16$.

Table 6: The comparison of the distribution of the minimal rank of $\mathbf{E_R}$ on $(37, 17, 16, 2)$ and $(37, 17, 5, 2)$ parameter sets with 10000 trials. In each trial, we exhaustively search the solution of MinRank problem on the matrix $\mathbf{E_R}$.

(v, o, q, l)	minimal rank	No. of trials ($q = 16$)	No. of trials ($q = 5$)
$(37, 17, q, 2)$	44	0	1
	45	0	4
	46	0	13
	47	0	24
	48	0	96
	49	0	205
	50	0	442
	51	0	713
	52	0	991
	53	0	1350
	54	0	1540
	55	0	1453
	56	0	1263
	57	1	874
	58	2	570
	59	35	307
	60	101	120
	61	410	33
	62	1699	1
	63	5083	0
	64	2650	0
	65	19	0

Since the computation complexity of the exhaustively search of the minimal rank of $\mathbf{E_R}$ is too large in the case of $q = 16$ and $l = 3$, we exhaustively search the minimal rank of $\mathbf{E_R}$ with $q = 5$. In Table 7, we can see that the smallest minimal ranks of $\mathbf{E_R}$ are much higher than the lower bound in Table 2 even when $q = 5$. Therefore, in the case $q = 16$, we can expect that the distribution of minimal rank of $\mathbf{E_R}$ are far away from the lower bound in Table 2.

Table 7: Table of the distribution of the minimal rank with 1000 trials when $l = 3$. In each trial, we exhaustively search the solution of MinRank problem on the matrix $\mathbf{E_R}$. In this table, with $q = 5$, we can see that the smallest minimal rank are higher than the lower bound in Table 2.

(v, o, q, l)	lower bound in Table 2	minimal rank	No. of trials
$(25, 8, 5, 3)$	49	61	1
		62	6
		63	27
		64	367
		65	599

Table 8: Table of the distribution of the minimal rank with 1000 trials when $l = 3$. In each trial, we exhaustively search the solution of MinRank problem on the matrix $\mathbf{E_R}$. In this table, with $q = 5$, we can see that the smallest minimal rank are higher than the lower bound in Table 2.

(v, o, q, l)	lower bound in Table 2	minimal rank	No. of trials
$(49, 11, 5, 3)$	83	87	11
		88	47
		89	244
		90	624
		91	74

Table 9: Table of the distribution of the minimal rank with 1000 trials when $l = 3$. In each trial, we exhaustively search the solution of MinRank problem on the matrix $\mathbf{E_R}$. In this table, with $q = 5$, we can see that the smallest minimal rank are higher than the lower bound in Table 2.

(v, o, q, l)	lower bound in Table 2	minimal rank	No. of trials
$(66, 15, 5, 3)$	110	118	2
		119	7
		120	16
		121	34
		122	89
		123	239
		124	518
		125	95

Therefore, we conclude that our parameter set with $l = 3$ satisfy the security requirement since for our current parameters we set $q = 16$.

In [1], we can see that the attack has almost no impact on $l = 4$ parameter sets. As A, B, Q varying, we could expect that the minimal rank of $\mathbf{E_R}$ will be way higher than the lower bound in Table 2 on $l = 4$ parameter sets.

4 Full rank SNOVA

As show by Beullens [1] attention must be paid to a good choice of the matrices A_α, B_α, $Q_{\alpha 1}$ and $Q_{\alpha 2}$ or the scheme will be vulnerable to forgery attacks. Before we address this we make explicit the relation between the SNOVA form the equation (2.1) and the MAYO structure as described by Beullens.

Recall the bilinear formulation of the SNOVA public map, the equation (2.1), for $i \in \{1, \ldots, m\}$

$$P_i(\vec{\mathbf{U}}) = \sum_{\alpha=1}^{l^2} \sum_{j=1}^{n} \sum_{k=1}^{n} A_\alpha \cdot U_j^t (Q_{\alpha 1} P_{i,jk} Q_{\alpha 2}) U_k \cdot B_\alpha \tag{4.1}$$

Denote the components of the $nl \times l$ matrix $\vec{\mathbf{U}}$ as $U_{k,j}$ Adding explicit matrix indices, the equation (4.1) can be written as

$$P_{i,i_1,j_1}(\vec{\mathbf{U}}) = \sum_{\alpha} \sum_{\substack{i_2,j_2,k_1 \\ k_2,k_3,k_4}} A_{\alpha,i_1,i_2} U_{k_1,i_2} Q^{\otimes n}_{1(\alpha,k_1,k_2)} P_{i,k_2,k_3} Q^{\otimes n}_{2(\alpha,k_3,k_4)} U_{k_4,j_2} B_{\alpha,j_2,j_1} \tag{4.2}$$

Here, and in the following, we use $i_\bullet, j_\bullet = 0, ..., l-1$ and $k_\bullet = 0, ..., nl-1$. As the Q matrices are in $\mathbb{F}_q[S]$, the $Q^{\otimes n}$ matrices can be expressed in terms of its coefficients $q_{1(\alpha,a)}$ as

$$Q^{\otimes n}_{1\alpha} = \sum_{a=0}^{l-1} q_{1(\alpha,a)} \left(S^a\right)^{\otimes n}$$

and similarly $Q^{\otimes n}_{2\alpha}$ and $q_{2(\alpha,b)}$. In terms of these coefficients, the equation (4.2) can be expressed as

$$P_{i,i_1,j_1}(\mathbf{U}) = \sum_{(i_2,a),(j_2,b)} E_{i_1,j_1,(i_2,a),(j_2,b)} D_{i,(i_2,a),(j_2,b)}(\mathbf{U}) \tag{4.3}$$

where

$$D_{i,(i_2,a),(j_2,b)}(\mathbf{U}) = \sum_{k_1,k_2,k_3,k_4} U_{k_1,i_2} \left(S^a\right)^{\otimes n}_{k_1,k_2} P_{i,k_2,k_3} \left(S^b\right)^{\otimes n}_{k_3,k_4} U_{k_4,j_2} \tag{4.4}$$

and

$$E_{i_1,j_1,(i_2,a),(j_2,b)} = \sum_{\alpha} q_{1(\alpha,a)} q_{2(\alpha,b)} A_{\alpha,i_2,i_1} B_{\alpha,j_1,j_2}. \tag{4.5}$$

Note that

$$D_{i,(i_2,a),(j_2,b)}(\mathbf{U}) = \mathbf{u}_{i_2}^t \left(\mathbf{S}^a\right)^{\otimes n} P_i \left(\mathbf{S}^b\right)^{\otimes n} \mathbf{u}_{j_2} = \mathcal{B}_{i,a,b}(\mathbf{u}_{i_2}, \mathbf{u}_{j_2}) \tag{4.6}$$

where $\mathcal{B}_{i,a,b}(\mathbf{u}_{i_2}, \mathbf{u}_{j_2})$ is the (a,b) entries of bilinear map $\mathcal{B}(\mathbf{u}_{i_2}, \mathbf{u}_{j_2})$ [1].

Note that the $\{(i_2,a),(j_2,b)\}$-th entries of the $l^2 \times l^2$ matrix $\tilde{\mathbf{E}}_{i_1,j_1}$ in the bilinear formulation (2.3) is $E_{i_1,j_1,(i_2,a),(j_2,b)}$, i.e.,

$$\left[E_{i_1,j_1,(i_2,a),(j_2,b)}\right]_{i_2,j_2,a,b \in \{0,\ldots,l-1\}} = \tilde{\mathbf{E}}_{i_1,j_1} \tag{4.7}$$

the equation (4.3) was first arrived at by Beullens [1]. the equation (4.5) makes the relation between the SNOVA form the equation (4.2) and the MAYO formulation the equation (4.3) explicit. If the index α is allowed to run up to l^4 the relation can always be reversed; for any set of $\tilde{\mathbf{E}}_{i_1,j_1}$ matrices, the equation (4.3) can also be put in the SNOVA form the equation (4.2). See appendix A for an example. This may not be the case if the sum over α contains only l^2 elements as currently specified for SNOVA.

The attack formulated by Beullens in [1] depends on the possibility of finding coefficients $c_{(a,b)}$ for which the rank of the sum $\sum c_{i_1,j_1}\tilde{\mathbf{E}}_{i_1,j_1}$ is less than l^2. For random matrices this will usually be possible so the set of $\mathbf{E}$ matrices needs to constructed carefully. For full rank SNOVA we use a $l^2 \times l^2$ matrix $\mathbf{E}$ with an irreducible characteristic polynomial. As $\mathbb{F}_q[\mathbf{E}]$ is a field, all non-zero elements are invertible. It follows that $\mathbf{E}_{i_1,j_1} = \mathbf{E}^{i_1 l + j_1}$ for $i_1, j_1 = 0, ..., l-1$ is a set of matrices that satisfy the full rank condition. See appendix B for a choice of $\mathbf{E}$ matrices.

Using the equation (4.3) it is much easier to find a set of matrices that satisfy full rank condition than using the equation (4.2). But there is no fundamental change to SNOVA whether we use A, B, and Q or the equivalent formulation in terms of $\mathbf{E}_{i_1,j_1}$. The main difference to the original SNOVA is that the sum over α has to run op to l^4 in order to use an irreducible set of $\mathbf{E}$ matrices.

In Table 10 we report the performance for an AVX2 optimized implementation of full rank SNOVA.

Table 10: Performance comparison at SL I between the AVX2 optimized implementation of full rank SNOVA and the AVX2 optimized implementation of SNOVA with ABQ fixed as scheme parameters.

(v, o, l)	XOF	Sig.	PK	KeyGen	Sign SSK	Sign ESK	Verify
Full Rank							
(37, 17, 2)	AES	124	9842	975,448	1,040,270	470,151	183,097
(25, 8, 3)	AES	165	2320	472,859	958,048	608,151	206,443
(24, 5, 4)	AES	248	1016	407,477	1,301,711	990,112	252,739
Fixed ABQ							
(37, 17, 2)	AES	124	9842	994,606	979,662	408,244	133,514
(25, 8, 3)	AES	165	2320	461,836	909,411	567,129	206,905
(24, 5, 4)	AES	248	1016	271,411	898,118	711,862	156,897

5 ABQ matrices as a scheme parameters

In the current specification of SNOVA the A, B, and $Q_{1,2}$ matrices are generated from the public key seed, just like the P_i matrices. As some of these matrices yield a weakened scheme, a possible response is to fix A, B, and $Q_{1,2}$ to well-chosen values and leave the rest unchanged.

Preferably the well-chosen values result in a full rank scheme, but we have not been able to find such values for $l \geq 3$. Beullens has argued that the expected minrank of a random matrix is $l^2 - l + 1$. We use this value as a lower bound for the definition of a "good enough" value when $l \geq 3$. For $l = 2$ an exhaustive brute force search is feasible. We have imposed the additional constraint that all l^2 resulting A_α and B_α matrices are invertible. The consequence of this constraint is that the corrections specified in Algorithms 1 and 2 of the SNOVA specification [8] are not needed. This brute force search has resulted in multiple full rank matrices for $l = 2$. We use shake256("SNOVA_2_89047137", 24) for $l = 2$. For higher ranks we have not found a set of ABQ matrices that results in a full rank scheme. For $l = 3$ we selected shake256("SNOVA_3_15", 108) and for $l = 4$ we use shake256("SNOVA_4_52", 320). These shake256 seeds may change as a result of ongoing analysis.

The advantage of using a fixed set of ABQ matrices over the full rank scheme described in Section 4 is that the sum over α has l^2 terms whereas the full rank resolution requires a sum over l^4 terms. Comparing to Varying ABQ, using fixed values for ABQ allows for a higher level of vectorization. This results in faster signing and verification. Table 10 shows the performance of the schemes with fixed ABQ and compares it to the Full Rank resolution. Security estimates of the described schemes are presented in Table 11. Both the full rank approach and fixing ABQ schemes have estimated security levels satisfying the NIST SL I requirement.

Table 11: Security estimates for Beullens attack for the SL I parameter sets. Full Rank is the estimate for Beullens attack with the full rank resolution described in section 4. Fixed ABQ is the estimate for Beullens attack using the apperoach of fixing ABQ of section 5. For comparison the security estimate for the collision attack [9] has been included. The collision attack is the most efficient known attack for all parameter sets at SL I.

(v, o, l)	Collision	Full Rank	Fixed ABQ
(37, 17, 2)	**151**	188	188
(25, 8, 3)	**159**	224	163
(24, 5, 4)	**175**	230	184

6 Conclusion

In this note, we propose two approaches to address the rank drop problem. The security claims of SNOVA are upheld with the current set of parameters for both approaches. Note that for parameter sets with $l = 4$, the security claim was never in doubt, even after the Beullens attack.

The first approach, which varies the A, B, and Q matrices, may not increase the computational cost for signing or verifying, meaning its performance could be comparable to the original SNOVA. The second approach involves creating a full-rank version of SNOVA by increasing the sum over α from l^2 to l^4. While

this alternative approach may appear to significantly increase computational complexity, optimizations ensure that both signing and verifying remain only marginally slower than the original SNOVA. Therefore, this alternative approach remains a viable option for adjusting SNOVA.

References

[1] Beullens, W.: **Improved Cryptanalysis of SNOVA.** Cryptology ePrint Archive, Report 2024/1297, 2024. https://eprint.iacr.org/2024/1297.pdf.

[2] Hashimoto, Y.: **An improvement of algorithms to solve underdefined systems of multivariate quadratic equations.** *JSIAM Letters*, 15:53–56, 2023. doi:10.14495/jsiaml.15.53.

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

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

[5] Nakamura, S., Tani, Y., Furue, H.: **Lifting approach against the SNOVA scheme.** Cryptology ePrint Archive, Paper 2024/1374, 2024. Available at https://eprint.iacr.org/2024/1374.

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

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

[8] Wang, L.C., Chou, C.Y., Ding, J., Kuan, Y.L., Li, M.S., Tseng, B.S., Tseng, P.E., Wang, C.C.: **SNOVA.** Technical report, National Institute of Standards and Technology, 2023. Available at https://csrc.nist.gov/Projects/pqc-dig-sig/round-1-additional-signatures.

[9] Wang, L.C., Tseng, P.E., Kuan, Y.L., Chou, C.Y.: **A simple noncommutative UOV scheme.** Cryptology ePrint Archive, Report 2022/1742, 2022. https://eprint.iacr.org/2022/1742.

A Generating the A,B,Q matrices from a set of E matrices

The following code snippet converts an arbitrary set of **E** matrices to A,B,Q matrices:

```
/**
 * Generate ABQ matrices from a given set of E_(i1, j1) matrices
 */
void gen_ABQ_from_E(map_group1 *map, const uint8_t *E)
{
  assert(nalpha_SNOVA == (lsq_SNOVA * lsq_SNOVA));

  for (int alpha = 0; alpha < nalpha_SNOVA; alpha++)
  {
    int a = alpha % l_SNOVA;
    int b = (alpha / l_SNOVA) % l_SNOVA;
    int j1 = (alpha / lsq_SNOVA) % l_SNOVA;
    int j2 = (alpha / lsq_SNOVA / l_SNOVA) % l_SNOVA;

    for (int i1 = 0; i1 < l_SNOVA; ++i1)
      for (int i2 = 0; i2 < l_SNOVA; ++i2)
      {
        map->Aalpha[alpha][i1 * l_SNOVA + i2] =
             E[i1 * l_SNOVA + j1][i2 * l_SNOVA + a][j2 * l_SNOVA + b];
        map->Balpha[alpha][i1 * l_SNOVA + i2] = (i1 == j2) * (i2 == j1);
        map->Qalpha1[alpha][i1 * l_SNOVA + i2] = S[a][i1 * l_SNOVA + i2];
        map->Qalpha2[alpha][i1 * l_SNOVA + i2] = S[b][i1 * l_SNOVA + i2];
      }
  }
}
```

B E matrices

For the irreducible polynomials generating $\mathbf{E}$ we use the following:

Table 12: Irreducible polynomials $p(z)$ use to generate $\mathbf{E}$ for all ranks l

Rank	**Irreducible polynomial** $p(z)$ **over** $\mathbb{F}_{16} = \mathbb{F}_2[x]/(x^4 + x + 1)$
$l = 2$	$z^4 + z^2 + x^3z + 1$
$l = 3$	$z^9 + z + 1$
$l = 4$	$z^{16} + z^3 + xz^2 + (x^3 + x + 1)z + 1$

To make the specification complete, the entries to the companion matrix $\mathbf{E}$ of p are taken to be:

$$\begin{aligned}\mathbf{E}_{i,j} &= 1 \text{ if } i > 0 \text{ and } j = i - 1 \\ &= \text{ coefficient of } z^{l^2-1-j} \text{ in } p(z) \text{ if } i = 0 \\ &= 0 \text{ otherwise}\end{aligned}$$