# On Unpadded NTRU Quantum (In)Security

Théodore Conrad-Frenkiel[1], Rémi Géraud-Stewart[2], and David Naccache[1]

[1] Département d'informatique de l'ÉNS, École normale supérieure,
CNRS, PSL Research University, Paris, France.
[2] QPSI, Qualcomm Inc., San Diego CA, USA.

**Abstract.** This paper utilizes the techniques used by Regev [Reg09] and Lyubashevsky, Peikert & Regev in the security reduction of LWE and its algebraic variants [LPR13] to exhibit a quantum reduction from the decryption of NTRU to leaking information about the secret key. Since this reduction requires decryption with the same key one wishes to attack, it renders NTRU vulnerable to the same type of attacks that affect the Rabin–Williams scheme [Ber08] – albeit requiring a quantum decryption query.
A common practice thwarting such attacks consists in applying the Fujisaki-Okamoto (FO, [FO99]) transformation before encrypting. However, not all NTRU protocols enforce this protection. In particular the DPKE version of NTRU [SXY18] is susceptible to such an attack.

## 1 Introduction

The leading post-quantum cryptographic schemes are lattice-based. They account for five among six PQC Round 3 finalists in NIST's standardization process, and two out of seven alternative candidates[3]. Within lattice-based cryptography the most common approaches are module learning with errors (MLWE, three finalists) and NTRU (two finalists and one alternative), which have both undergone substantial scrutiny [AD21].

The major appeal of MLWE schemes comes from their security reductions. In 2009, Regev [Reg09] showed a quantum reduction from the worst-case instances of lattice problems to random instances of the learning with errors (LWE) problem. In 2013 Lyubashevsky, Peikert & Regev [LPR13] extended this result to algebraic variants of LWE (such as MLWE) with a quantum reduction from worst-case algebraic lattice problems to MLWE.

Recently, Pellet-Mary and Stehle provided the first security proof for NTRU [PS21]. In their paper they present a quantum reduction from worst-case approximate Shortest Vector Problem (SVP) over ideal lattices to an average-case search variant of the NTRU problem. However, their result does not provide a complete security proof.

Indeed, in NTRU (unlike MLWE) the mathematical problem which must be solved to break the private key and the mathematical problem which must be solved to decrypt a single message are *very different*: breaking the secret key requires finding the shortest vector in a certain lattice, whereas decrypting a message requires the solution of a Bounded Distance Decoding (BDD) problem in the dual lattice.

In this paper we show that a quantum decryption oracle allows an adversary to find relatively short vectors in the key lattice. For the parameters considered for the NTRU NIST

[3] See https://csrc.nist.gov/Projects/post-quantum-cryptography/round-3-submissions

PQC candidates this does not seem to lead to a full attack, but does allow the attacker to solve an otherwise computationally difficult problem and thereby breaks indistiguishability. Note that stronger results are known for LWE [AJOP20]: a single quantum decryption query allows the adversary to recover the full secret key with constant success probability.

The result is somewhat a double-edged sword: On one hand it is another step in the direction of proving NTRU's security. On the other hand, it lends the scheme to a new set of attacks since a single quantum query to the decryption oracle would compromise it. This is similar to the security proof of the Rabin–Williams scheme, where the security proof led many to avoid it, precisely because of the threat of an adversary which might be able to use a single decryption to break the private key using the reduction of the security proof.

While NIST's original call for proposals did not require resistance to quantum decryption oracles, we think that since such an attack exists and since the foreseen advent of quantum computers is the very reason to be of the NIST's call, it is prudent to implement all possible protections to the scheme before deployment.

## 2 The NTRU Cryptosystem

Before we proceed, let us briefly remind the definition of NTRU [HPS98]. The scheme uses three public parameters: a prime $n$ and two integers $p, q$ such that $p \nmid q$. Typically $n$ and $q$ are taken in the range 250 to 2500, whereas $p$ is usually very small, e.g., $p = 3$.

There are two equivalent ways to describe NTRU operations, in terms of polynomial multiplication in the quotient ring $R = \mathbb{Z}[X]/(X^n - 1)$ (which is the usual point of view), or in terms of the convolution product in the group $\mathbb{Z}^n$.[4] Ideed a polynomial $a_0 + \ldots + a_{n-1}X^{n-1}$ is identified with the vector of its coefficients $\boldsymbol{a} = (a_0, \ldots, a_{n-1})$, which makes the equivalence immediate:

$$\boldsymbol{c} = \boldsymbol{ab} \quad \Leftrightarrow \quad c_k = \sum_{k=i+j \bmod n} a_i b_j.$$

Polynomials of $R$ with coefficients in $\{-1, 0, 1\}$ form the *small elements set* $S$, which plays a fundamental role in NTRU. Key generation consists in picking $\boldsymbol{F}, \boldsymbol{G} \in S$ and computing

$$\boldsymbol{f} \leftarrow 1 + p\boldsymbol{F}, \qquad \boldsymbol{g} \leftarrow p\boldsymbol{G}, \qquad \boldsymbol{h} \leftarrow \boldsymbol{f}^{-1}\boldsymbol{g} \bmod q,$$

where all operations are in $R$. The public key is $\mathsf{pk} = \boldsymbol{h}$, whereas the secret key is $\mathsf{sk} = (\boldsymbol{F}, \boldsymbol{G})$.

To encrypt a plaintext $\boldsymbol{m} \in S$, pick a random $\boldsymbol{s} \in S$ and compute the ciphertext

$$\boldsymbol{c} \leftarrow \boldsymbol{sh} + \boldsymbol{m} \bmod q.$$

Finally, to decrypt a ciphertext $\boldsymbol{c}$, first compute

$$\boldsymbol{a} \leftarrow \boldsymbol{fc} \bmod q,$$

then lift $\boldsymbol{a}$ to $\mathbb{Z}^n$ with coefficients $|a_i| \leq \frac{q}{2}$. The result of this operation, taken modulo $p$ retrieves $\boldsymbol{m}$.

[4] See Silverman: http://archive.dimacs.rutgers.edu/Workshops/Post-Quantum/Slides/Silverman.pdf.

## 3 Quantum Decryption Query Attacks

### 3.1 Preliminaries

*Quantum Decryption Query.* A *Quantum Decryption Query* is a superposition of inputs, i.e., a linear combination of inputs of the attacker's choice. Using the bra-ket notation, such a query is written $\sum_x \psi_x|x\rangle$. We do not make normalization explicit to maintain readability.

In fact we will consider queries of the form $\sum_x \psi_x |x\rangle |x\rangle$ (this is to ensure unitarity). A response to this query is a superposition $\sum_x \psi_x |x\rangle |\mathrm{Dec}_k(x)\rangle$ where $\mathrm{Dec}_k(x)$ computes the decryption of $x$ with key $k$ (unknown to the attacker but known to the oracle). For more information on such queries see [BZ13b,BZ13a,GHS16].

*Quantum Fourier Transform.* The *Quantum Fourier Transform* (QFT) is a unitary operator, usually defined over $\{0,1\}^N$ by

$$\mathrm{QFT} = \frac{1}{\sqrt{2^N}} \sum_{x=0}^{2^N-1} \sum_{y=0}^{2^N-1} \omega^{xy} |y\rangle \langle x| \,,$$

where $\omega = \exp\bigl(2i\pi/2^N\bigr)$. Recall that the QFT can be computed exactly in polylog time on a quantum computer [Kit95,HH00]. The above definition for the QFT is directly extended over vectors, using the dot product instead of integer product.

### 3.2 Algorithm

Inputs: $S \subset R, h \in R$.

1. Prepare initial state: $\sum_{m\in S} |m\rangle$
2. State expansion (see [Reg09]) $\sum_{m,s\in S} |m\rangle |s\rangle$
3. Apply the unitary operation $|a\rangle |b\rangle \mapsto |a\rangle |a+hb\rangle$ to obtain the state

$$\sum_{m,s\in S} |m\rangle |hs+m\rangle$$

   Notice that the second register's state is the NTRU encryption of the first register's state.
4. Apply the unitary operation $|a\rangle |b\rangle \mapsto |a - \mathrm{Dec}_k(b)\rangle |b\rangle$ to obtain the state

$$\sum_{m,s\in S} |m - \mathrm{Dec}_k(hs+m)\rangle |hs+m\rangle = \sum_{m,s\in S} |0\rangle |hs+m\rangle$$

   Note that to perform this step we execute one quantum decryption query to the oracle.
5. Compute the QFT of the right-most register, and return the result $|p\rangle$.

### 3.3 Analyzing the Results

Just before applying the QFT gate, our quantum state behaved according to the distribution $D = D_m * (hD_s)$, where $D_m$ is the distribution of short messages, $D_s$ is the distribution of short salts and $*$ denotes convolution. By the convolution theorem,

$$\begin{aligned}\mathrm{QFT}\left(D_m * (hD_s)\right) &= \mathrm{QFT}(D_m) \cdot \mathrm{QFT}(hD_s) = \widehat{D}_m \cdot \mathrm{QFT}(hD_s) \\ &= \widehat{D}_m \cdot \bar{h}\, \mathrm{QFT}(D_s) = \widehat{D}_m \cdot \bar{h}\widehat{D}_s\end{aligned}$$

We would like to express the right-hand side of this equation as a function of $\widehat{D}_s = \mathrm{QFT}(D_s)$. As we show below there exists $\bar{h}$ such that $\mathrm{QFT}(hD_s) = \bar{h}\,\mathrm{QFT}(D_s)$.

For some choices of $D_s$ and $D_m$ we can anticipate that both $D$ and $\widehat{D}$ will have a low variance (e.g., if $D$ is a Gaussian distribution with standard deviation $\sigma = \sqrt{q}$ then so is $\widehat{D}$).

Therefore, the output of our algorithm is a ring element which tends to be both short and $\bar{h}$ times a short element. If $\bar{h}$ were equal to $h^{-1}$ (the inverse of $h$ in $R$), then finding this element would be exactly the problem of finding a short vector in the lattice of the NTRU private key.

### 3.4 Dealing With the Transpose

Now we need to explain why being $\bar{h}$ times a short vector is equivalent to being $h^{-1}$ times a short vector.

Assume for simplicity that we are dealing with the ring $\mathbb{Z}[x]/(x^n+1)$, and consider the polynomial $h(x) = h_0 + h_1 x + \ldots h_{n-1}x^{n-1}$. Denote by $H$ the matrix such that for the polynomial $p(x) = p_0 + p_1 x + \ldots p_{n-1}x^{n-1}$, the coefficients of $ph$ are given by $(p_0, \ldots, p_{n-1})H$. It is easy to see that:

$$H = \begin{pmatrix} h_0 & h_1 & \ldots & h_{n-1} \\ -h_{n-1} & h_0 & \ldots & h_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ -h_1 & -h_2 & \ldots & h_0 \end{pmatrix} \quad \text{for which} \quad H^\top = \begin{pmatrix} h_0 & -h_{n-1} & \ldots & -h_1 \\ h_1 & h_0 & \ldots & -h_2 \\ \vdots & \vdots & \ddots & \vdots \\ h_{n-1} & h_{n-2} & \ldots & h_0 \end{pmatrix}$$

That is: if $H$ is the matrix corresponding to the polynomial $h(x) = h_0 + h_1 x + \ldots + h_{n-1}x^{n-1}$, then $H^\top$ corresponds to the polynomial $h^\top = h_0 - h_{n-1}x - \ldots - h_1 x^{n-1}$. We will note three important facts about this transformation:

1. It clearly preserves the $L_2$ norm of the coefficients.
2. It can be shown that it behaves well with the multiplication in the ring:

$$\forall p, h \in \mathbb{Z}[x]/(x^n+1), \quad (ph)^\top(x) = h^\top p^\top(x).$$

3. $(h^\top)^\top = h$.

Denote by $p$ (instead of $|p\rangle$) the output of the quantum algorithm described above. We showed that $p$ is short and that $h^\top p$ is short. Therefore $p^\top$ is short and such that $hp^\top = (h^\top p)^\top$ is short, so by transposing the output of the quantum algorithm we obtain a short vector in the NTRU private key lattice.

The same phenomenon happens in $R = \mathbb{Z}[X]/(X^n - 1)$ namely:

$$H = \begin{pmatrix} h_0 & h_1 & \dots & h_{n-1} \\ h_{n-1} & h_0 & \dots & h_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ h_1 & h_2 & \dots & h_0 \end{pmatrix} \quad \text{for which obviously} \quad H^\top = \begin{pmatrix} h_0 & h_{n-1} & \dots & h_1 \\ h_1 & h_0 & \dots & h_2 \\ \vdots & \vdots & \ddots & \vdots \\ h_{n-1} & h_{n-2} & \dots & h_0 \end{pmatrix}$$

Here $H$ is the matrix corresponding to the polynomial $h(x) = h_0 + h_1 x + \ldots + h_{n-1}x^{n-1}$ and $H^\top$ corresponds to the polynomial $h^\top = h_0 + h_{n-1}x + \ldots + h_1 x^{n-1}$. The three facts that we stated for the case $R = \mathbb{Z}[X]/(X^n + 1)$ hold for $R = \mathbb{Z}[X]/(X^n - 1)$ as well.

*Remark 1 (Ternary error distribution).* In most NTRU applications the secret keys, the messages and the salt are chosen to be random ternary vectors. If we look at $D_s = D_m = U(\{-1, 0, 1\})$, the indicator function has a simple Fourier transform:

$$\widehat{f}(\alpha) = \sum_{x=-1,0,1} \exp(2i\pi\alpha x/q) = 1 + 2\cos(2\pi\alpha/q).$$

Therefore, $|f(\alpha)|^2 = (1 + 2\cos(2\pi\alpha/q))^2$. This distribution attains its maximum in $\alpha = 0$ and its minimum at $\alpha = \pm q/3$.

*Remark 2.* If we now consider a vector-valued distribution, i.e, $U(\{-1, 0, 1\}^n)$, the Fourier transform becomes:

$$\widehat{f}(\boldsymbol{\alpha}) = \sum_{\boldsymbol{x} \in \{-1,0,1\}^n} \exp\left(\frac{2i\pi}{q} \boldsymbol{\alpha} \cdot \boldsymbol{x}\right) = \prod_{i=1}^{n} \left(1 + 2\cos(2\pi\alpha_i/q)\right),$$

which is also strongly biased, peaking at vectors whose coordinates are multiples of $q$.

*Remark 3.* Note that in the general case, the Fourier transform of the indicating function of $U([-\ell, \ell])$ is

$$|\hat{f}(\alpha)| = \csc\left(\frac{\ell\pi\alpha}{q}\right) \sin\left(\frac{(1 + 2\ell)\pi\alpha}{q}\right).$$

This distribution exhibits a similar behaviour as the cases discussed above.

## 4 Small-Scale Example

We can work out the computation on a small-scale example, demonstrating the claims made above. We consider $q = 3329$, $n = 3$, $f = X + 1$, $g = -X^2 + X + 1$. We get

$$h := \frac{g}{f} = 1664X^2 + 1664X + 1666.$$

The collection $\{m + hs \mid m, s \in S\}$ contains 729 elements, which are *not* uniformly distributed. Of these, 238 are unique, 90 appear twice, 54 appear three times, 18 appear 4 times, etc.

Rather that visualising directly the Fourier transform, which is a function of 3D space, we can look at it over 1D slides along each coordinate. This is represented in Figure 1, obtained using a classical FFT. This distribution favours short results, a fact that is not substantially impacted by increasing $n$.

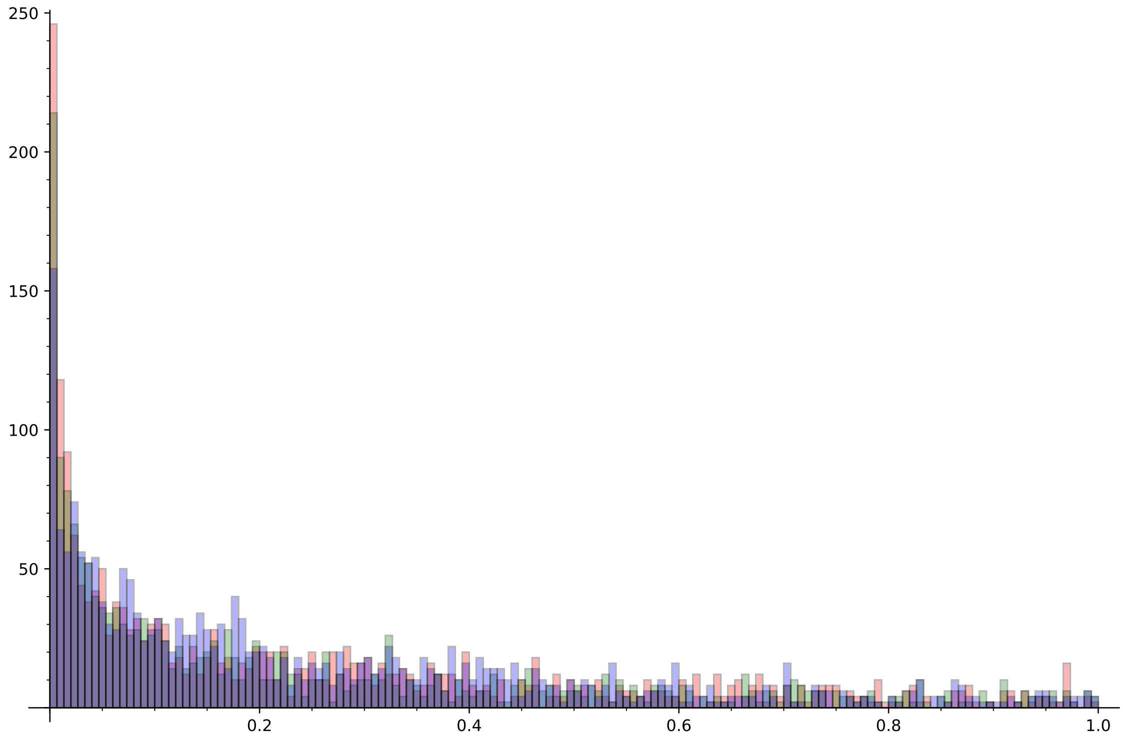


**Fig. 1.** Coordinate-wise Fourier transform. Each coordinate corresponds to a colour: red, green, blue. The $x$ scale is between 0 and a cutoff at $10^8$.

Next we can verify that $hp$ is also short. We can do this by plotting $\widehat{f}(hb)$, where $b$ spans $R$ and $\widehat{f}$ is the Fourier transform computed above. Here too we only look at 1D slices which are easier to visualise, see Figure 2. This distribution should also favor short values.

*Remark 4.* Storing $hb$ for all $b \in R$ requires $q^n \log_2 q$ bits; this barely fits within 64 GB with the example parameters.

## 5 Mitigations

One possible mitigation could be to use the NTRU encryption in a protocol which would prevent direct access to a decryption oracle. In fact, modern lattice KEMs use an external

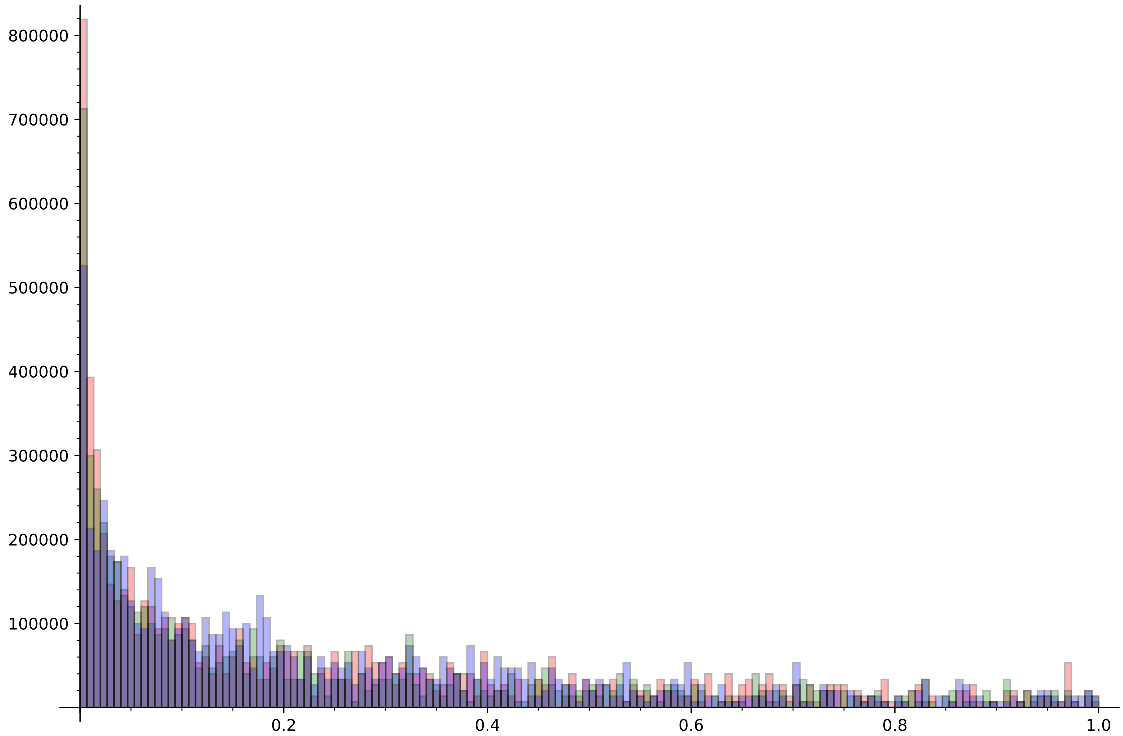


**Fig. 2.** Coordinate-wise distribution of $\widehat{f}(hb)$, with a horizontal cutoff at $10^8$.

protocol which is aimed at preventing adversaries from decrypting malicious ciphertexts. However, it is unclear whether current protocols should prevent the user from decrypting a legitimate ciphertext upon request.

The Fujisaki-Okamoto transformation [FO99], which is a good textbook practice, protects against this type of attack since it both requires the salt and the message entering the CPA encryption to be produced using a PRF, and hashes the decryption's result. Almost all lattice-based cryptographic schemes in the NIST competition use FO, except NTRU. Indeed, the NTRU candidate uses a modified protocol that defends against known decryption failure attacks; in particular their commitment is to never decrypt a ciphertext that was not produced using a message and salt pair within the message space and hence rely upon the "rigidity" of their scheme thwart decryption failure attacks.

In particular, the DPKE version of the NTRU candidate [SXY18] contains explicit instructions on how to make it CCA secure[5] but these still leave it vulnerable to our attack.

[5] Excerpt of DPKE_Decrypt: *"This implementation assumes that only the KEM interface is exprosed to users. Implementations that exposed to users. Implementations that expose the DPKE to users are required to return* $(\boldsymbol{pack\_S3}(0)\|\boldsymbol{pack\_S3}(0), 1)$ *on failure."*

## 6 Conclusions and Further Research

The reduction presented above can be seen both as a security claim for NTRU (since together with [PS21] it shows that NTRU's security reduces to the worst-case hardness of gap ideal SVP). However, it is still important to take into consideration the adverse effect it has on the security of NTRU. Note that the attack does not allow us to find the private key itself but allows us to solve a related problem that might (or rather should) otherwise be very difficult. In particular this reduction can be used by an adversary with access to a quantum decryption oracle to obtain a hint about the private key (at the very least, breaking the indistiguishability assumption).

An interesting research direction[6] would consist in running the attack described in this paper several times to collect more and more information about the secret key's lattice with the hope to eventually hand-over the extracted information to quantum lattice sieving or extreme pruning to access the secret key (or a functionally equivalent one).

A further question is that of the applicability of the techniques described in this paper to Falcon and NTRU Prime – if such extensions happen to be possible.

## References


AD21. Martin Albrecht and Léo Ducas. Lattice Attacks on NTRU and LWE: A History of Refinements. Cryptology ePrint Archive, Report 2021/799, 2021. https://ia.cr/2021/799. (cited on page 1)

AJOP20. Gorjan Alagic, Stacey Jeffery, Maris Ozols, and Alexander Poremba. On quantum chosen-ciphertext attacks and learning with errors. *Cryptogr.*, 4(1):10, 2020. (cited on page 2)

Ber08. Daniel J. Bernstein. Proving tight security for Rabin-Williams signatures. In Nigel P. Smart, editor, *Advances in Cryptology - EUROCRYPT 2008, 27th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Istanbul, Turkey, April 13-17, 2008. Proceedings*, volume 4965 of *Lecture Notes in Computer Science*, pages 70–87. Springer, 2008. (cited on page 1)

BZ13a. Dan Boneh and Mark Zhandry. Quantum-secure message authentication codes. In Thomas Johansson and Phong Q. Nguyen, editors, *Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings*, volume 7881 of *Lecture Notes in Computer Science*, pages 592–608. Springer, 2013. (cited on page 3)

BZ13b. Dan Boneh and Mark Zhandry. Secure signatures and chosen ciphertext security in a quantum computing world. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part II*, volume 8043 of *Lecture Notes in Computer Science*, pages 361–379. Springer, 2013. (cited on page 3)

FO99. Eiichiro Fujisaki and Tatsuaki Okamoto. How to enhance the security of public-key encryption at minimum cost. In Hideki Imai and Yuliang Zheng, editors, *Public Key Cryptography, Second International Workshop on Practice and Theory in Public Key Cryptography, PKC '99, Kamakura, Japan, March 1-3, 1999, Proceedings*, volume 1560 of *Lecture Notes in Computer Science*, pages 53–68. Springer, 1999. (cited on pages 1, 7)

GHS16. Tommaso Gagliardoni, Andreas Hülsing, and Christian Schaffner. Semantic security and indistinguishability in the quantum world. In Matthew Robshaw and Jonathan Katz, editors, *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 14-18, 2016, Proceedings, Part III*, volume 9816 of *Lecture Notes in Computer Science*, pages 60–89. Springer, 2016. (cited on page 3)


[6] that, if successful may lead to a complete attack or, equivalently, to a complete proof.

HH00. Lisa Hales and Sean Hallgren. An improved quantum fourier transform algorithm and applications. In *41st Annual Symposium on Foundations of Computer Science, FOCS 2000, 12-14 November 2000, Redondo Beach, California, USA*, pages 515–525. IEEE Computer Society, 2000. (cited on page 3)

HPS98. Jeffrey Hoffstein, Jill Pipher, and Joseph H. Silverman. NTRU: A ring-based public key cryptosystem. In Joe Buhler, editor, *Algorithmic Number Theory, Third International Symposium, ANTS-III, Portland, Oregon, USA, June 21-25, 1998, Proceedings*, volume 1423 of *Lecture Notes in Computer Science*, pages 267–288. Springer, 1998. (cited on page 2)

Kit95. A Yu Kitaev. Quantum measurements and the Abelian stabilizer problem. *arXiv preprint quant-ph/9511026*, 1995. (cited on page 3)

LPR13. Vadim Lyubashevsky, Chris Peikert, and Oded Regev. A toolkit for Ring-LWE cryptography. In Thomas Johansson and Phong Q. Nguyen, editors, *Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings*, volume 7881 of *Lecture Notes in Computer Science*, pages 35–54. Springer, 2013. (cited on page 1)

PS21. Alice Pellet-Mary and Damien Stehlé. On the hardness of the NTRU problem. *IACR Cryptol. ePrint Arch.*, page 821, 2021. (cited on pages 1, 8)

Reg09. Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *J. ACM*, 56(6):34:1–34:40, 2009. (cited on pages 1, 3)

SXY18. Tsunekazu Saito, Keita Xagawa, and Takashi Yamakawa. Tightly-secure key-encapsulation mechanism in the quantum random oracle model. In Jesper Buus Nielsen and Vincent Rijmen, editors, *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part III*, volume 10822 of *Lecture Notes in Computer Science*, pages 520–551. Springer, 2018. (cited on pages 1, 7)