ANTRAG: Simplifying and Improving FALCON Without Compromising Security

Thomas Espitau[1], Jade Guiton[2], Thi Thu Quyen Nguyen[3], Chao Sun[4], Mehdi Tibouchi[5], and Alexandre Wallet[6]

[1] PQShield SAS, France
thomas@espitau.com
[2] CentraleSupélec, France
guiton.jade@gmail.com
[3] IDEMIA & Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC, France
thi-thu-quyen.nguyen@inria.fr
[4] Osaka University, Japan
c-sun@ist.osaka-u.ac.jp
[5] NTT Social Informatics Laboratories, Japan
mehdi.tibouchi@ntt.com
[6] IRISA, Univ Rennes 1, Inria, Bretagne-Atlantique Center, France
alexandre.wallet@inria.fr

Abstract. This talk proposal presents ANTRAG, a trapdoor generation technique for Prest's hybrid sampler over NTRU lattices which we recently introduced at Asiacrypt 2023. Prest's sampler is used in particular in the recently proposed MITAKA signature scheme (Eurocrypt 2022), a variant of FALCON. MITAKA was introduced to address FALCON's main drawback, namely the fact that the lattice Gaussian sampler used in its signature generation is highly complex, difficult to implement correctly, to parallelize or protect against side-channels, and to instantiate over rings of dimension not a power of two to reach intermediate security levels. Prest's hybrid sampler is considerably simpler and solves these various issues, but when applying the same trapdoor generation approach as FALCON, the resulting signatures have far lower security in equal dimension. The MITAKA paper showed how certain randomness-recycling techniques could be used to mitigate this security loss, but the resulting scheme is still substantially less secure than FALCON (by around 20 to 50 bits of CoreSVP security depending on the parameters), and has much slower key generation.

Our new trapdoor generation techniques solves all of those issues satisfactorily: it gives rise to a much simpler and faster key generation algorithm than MITAKA's (achieving similar speeds to FALCON), and is able to comfortably generate trapdoors reaching the same NIST security levels as FALCON as well. It can also be easily adapted to rings of intermediate dimensions, in order to support the same versatility as MITAKA in terms of parameter selection. All in all, this new technique combines all the advantages of both FALCON and MITAKA (and more) with none of their drawbacks.

1 Introduction

1.1 Hash-and-sign lattice-based signatures

From GGH to FALCON. FALCON [PFH+22] is one of the three signature schemes already selected for standardization in the NIST post-quantum competition. It represents the state of the art in *hash-and-sign* lattice-based signatures, one of the two main paradigms for constructing lattice-based signatures alongside Lyubashevsky's Fiat–Shamir with aborts [Lyu09, Lyu12] (which is also represented among the final selected candidates of the NIST competition in the form of DILITHIUM [LDK+22]).

This makes FALCON the culmination of a long line of research in constructing signature schemes from *lattice trapdoors*. The basic idea, which dates back to the late 1990s with the GGH [GGH97] and NTRUSign [HHP+03] signature schemes, is to use as the signing key a "good" basis (the *trapdoor*) of a certain lattice allowing to approximate the closest vector problem within a good factor, and as the verification key a "bad" basis which allows to test membership but not decode large errors. The signature algorithm then hashes a given message to a vector in the ambient space of the lattice, and uses the trapdoor to find a relatively close lattice point to that vector. The difference is the signature, which is verified by checking that it is small and that its difference with the hashed vector does indeed belong to the lattice.

The GGH scheme, as well as several successive variants of NTRUSign, were eventually broken by statistical attacks [GS02, NR06, DN12]: it turned out that signatures would leak partial information about the secret trapdoor, that could then be progressively recovered by an attacker. This problem was finally solved in 2008, when Gentry, Peikert and Vaikuntanathan (GPV) [GPV08] showed how to use Gaussian sampling in the lattice in order to guarantee that signatures would reveal no information about the trapdoor.

GPV signatures over NTRU lattices. In order to instantiate the GPV framework efficiently in practice, one then needs lattices with compact representation and efficiently computable trapdoors, which has so far been achieved using module lattices over rings—in fact, mostly rank-2 modules over cyclotomic rings, exactly corresponding to NTRU lattices (although higher rank modules, namely ModNTRU lattices, have been shown to be usable as well in certain ranges of parameters [CPS+20]). This was first carried out by Ducas, Lyubashevsky and Prest (DLP) [DLP14], who analyzed trapdoor generation for power-of-two cyclotomic ring NTRU lattices and constructed corresponding GPV-style signatures. DLP signatures are compact, but the signing algorithm is rather slow: quadratic in the dimension $2d$ of the lattice. This is because the lattice Gaussian sampling algorithm that forms the core of its signing procedure (namely Klein–GPV sampling, in essence a randomized version of Babai's nearest plane algorithm for approximate CVP) cannot directly take advantage of the algebraic structure of the lattice, and thus operates on the full $(2d) \times (2d)$ matrix of the lattice basis as well as its Gram–Schmidt orthogonalization.

Falcon is a direct descendant of the DLP scheme, that replaces the generic, quadratic complexity Klein–GPV sampler in signature generation by an efficient, quasilinear complexity lattice Gaussian sampler that *does* take advantage of the ring structure. Specifically, that new algorithm is constructed by randomizing the Fast Fourier Orthogonalization (FFO) algorithm of Ducas and Prest [DP16], and operates in a tree-like fashion traversing the subfields of the power-of-two cyclotomic field over which the NTRU lattice is defined. This makes Falcon particularly attractive in various ways: it offers particularly compact signatures and keys (providing the best bandwidth requirements of all signature schemes in the NIST competition), achieves high security levels in relatively small lattice dimensions, and has both fast signing and very efficient verification speeds.

However, the FFO-based Gaussian sampler is also the source of Falcon's main drawbacks: it is a really contrived algorithm that is difficult to implement correctly, parallelize or protect against side-channels. It is also really difficult to adapt to other rings than power-of-two cyclotomics, which drastically limits Falcon's versatility in terms of parameter selection: in fact, recent versions of Falcon in the NIST competition only target either the lowest NIST security level (using cyclotomic fields of dimension 512) or the highest (using fields of dimension 1024) and nothing in-between.[7]

1.2 The hybrid sampler and Mitaka

The Peikert and hybrid samplers. After the publication of the DLP paper, Ducas and Prest explored and analyzed other approaches for lattice Gaussian sampling over NTRU lattices, as discussed in depth in Prest's Ph.D. thesis [Pre15], with a view towards overcoming the quadratic complexity of the naive Klein–GPV sampler. While the introduction of the FFO sampler was the final step of that exploration, they also considered two other major approaches along the way, which also achieve quasilinear complexity (see also [DP15]).

The first approach was not actually novel: it was the ring version of Peikert's lattice Gaussian sampler [Pei10], which is the randomization of the Babai rounding algorithm for approximate CVP, just like Klein–GPV is the randomization of Babai's nearest plane. For NTRU lattices, this algorithm consists of independent one-dimensional Gaussian samplings for each vector component (hence a linear number in total), as well as 2×2 matrix-vector products over the ring, amounting to a constant number of ring multiplications, that are all quasilinear when using FFT-based fast arithmetic. Thus, Peikert's sampler for NTRU lattices is quasilinear as required. However, Ducas and Prest analyzed the *quality* of NTRU trapdoors (generated in the same way as DLP) with respect to Peikert's sampler, and found that it was much worse than for Klein–GPV, both concretely and asymptotically. In other words, for the same choice of parameters, it would

[7] The earliest version of the Falcon specification [PFH+17] also included an intermediate parameter set of dimension 768, but the corresponding algorithms were so complicated that it was eventually dropped.

reduce security considerably to instantiate DLP with Peikert's sampler instead of Klein–GPV (and to recover the same security, a large increase in the dimension of the underlying ring, and hence the size of keys and signatures, would be required).

As a kind of middle ground between Peikert (fast but less secure) and Klein–GPV (secure but much slower), they introduced as a second approach the *hybrid sampler*, which uses the same structure as Klein–GPV (a randomized nearest plane algorithm) but over the larger ring instead of over $\mathbb{Z}$. In the rank-2 case of NTRU, this reduces to just two "nearest plane" iterations consisting of Gaussian sampling over the ring, which is itself carried out using Peikert's sampler with respect to a short basis of the ring. This algorithm remains quasilinear, but achieves a significantly better quality than Peikert for DLP-style NTRU trapdoors, although not as good as Klein–GPV. Concretely, for those NTRU trapdoors over the cyclotomic ring of dimension 512 (resp. 1024), signatures instantiated with the hybrid sampler achieve a little over 80 bits (resp. 200 bits) of classical CoreSVP security, compared to over 120 bits (resp. 280 bits) for Klein–GPV.

Pros and cons of hybrid vs. FFO. This substantial security loss is presumably the main reason that led to the hybrid sampler being abandoned in favor of the FFO sampler (which achieves the same quality as Klein–GPV but with quasilinear complexity) in the Falcon scheme. Indeed, security aside, the hybrid sampler has a number of advantages compared to the FFO sampler of Falcon: it is considerably simpler to implement, somewhat more efficient in equal dimension, easily parallelizable and less difficult to protect against side-channels; it also has an online-offline structure that can be convenient for certain applications, and it is easier to instantiate over non power-of-two cyclotomics, making it easier to reach intermediate security levels.

For these reasons, the use of the hybrid sampler to instantiate signatures over NTRU lattices was recently revisited by Espitau et al. as part of their proposed scheme Mitaka [EFG+22]. One of the key contributions of that paper is an optimization of trapdoor generation for the hybrid sampler that mitigates the security loss by making it possible to construct better quality trapdoor in reasonable time. Combined with the various advantages of the hybrid sampler, this allows the authors of Mitaka to achieve a trade-off between simplicity and security that they argue can be more attractive than Falcon. However, despite their efforts, Mitaka remains substantially less secure than Falcon in equal dimension (it loses over 20 bits of classical CoreSVP security over rings of dimension 512, and over 50 bits over rings of dimension 1024), with a much slower and more contrived key generation algorithm as well. In particular, Mitaka falls short of NIST security level I in dimension 512 and of level V in dimension 1024, making it less than ideal from the standpoint of parameter selection.

1.3 Contributions and technical overview

In this talk proposal based on our recent Asiacrypt 2023 paper [ENS+23a], we introduce a novel trapdoor generation technique for Prest's hybrid sampler that solves the issues faced by Mitaka in a natural and elegant fashion. Our technique gives rise to a much simpler and faster key generation algorithm than Mitaka's (achieving similar speeds to Falcon), and it is able to comfortably generate trapdoors reaching the same NIST security levels as Falcon. It can also be easily adapted to rings of intermediate dimensions, in order to support the same versatility as Mitaka in terms of parameter selection (just with better security). All in all, this new technique achieves in some sense the best of both worlds between Falcon and Mitaka.

NTRU trapdoors and their quality. In order to give a overview of the technical ideas involved, we need to recall a few facts about NTRU trapdoors and their quality with respect to the Klein–GPV and hybrid samplers. For simplicity, we concentrate on the special case of power-of-two cyclotomic rings $\mathscr{R} = \mathbb{Z}[x]/(x^d+1)$. Over such a ring, an NTRU lattice is simply a full-rank submodule lattice of $\mathscr{R}^2$ generated by the columns of a matrix of the form:

$$\mathbf{B}_h = \begin{bmatrix} 1 & 0 \\ h & q \end{bmatrix}$$

for some rational prime number q and some ring element h coprime to q. Note that this can also be described as a lattice of pairs $(u, v) \in \mathscr{R}^2$ such that $uh - v = 0 \bmod q$.

A trapdoor for this lattice is a relatively short basis:

$$\mathbf{B}_{f,g} = \begin{bmatrix} f & F \\ g & G \end{bmatrix}$$

where the basis vectors (f,g) and (F,G) are not much larger than the normalized volume $\sqrt{\det \mathbf{B}_h} = \sqrt{q}$ of the lattice. Since those vectors belong to the lattice, we have in particular that $g/f = G/F = h \bmod q$. Moreover, since the determinants are equal up to a unit of $\mathscr{R}$, we can impose without loss of generality that $fG - gF = q$.

Using the trapdoor $\mathbf{B}_{f,g}$, lattice Gaussian samplers are able to output lattice vectors following a Gaussian distribution on the lattice of standard deviation[8] a small multiple $\alpha\sqrt{q}$ of the normalized volume $\sqrt{q}$. The factor α is the *quality*, and depends both on the trapdoor and on the sampler itself. The lower the quality, the better the trapdoor, and the higher the security level of the resulting signature scheme. For the Klein–GPV sampler, one can show that the quality α is ($1/\sqrt{q}$ times) the maximum norm of a vector in the Gram–Schmidt orthogonalization of the basis $\mathbf{B}_{f,g}$ regarded as a $(2d)\times(2d)$ matrix over $\mathbb{Z}$, whereas for the hybrid sampler, it is similar but with the Gram–Schmidt orthogonalization over $\mathscr{R}$ itself.

Those quantities admit a simple expression in terms of the *embeddings* of the ring elements f and g. Recall that the embeddings are the d ring homomorphisms $\varphi_i\colon \mathscr{R} \to \mathbb{C}$; when elements of $\mathscr{R}$ are seen as polynomials, these embeddings are simply the evaluation morphisms $\varphi_i(u) = u(\zeta_i)$ where the ζ_i's are the d primitive $2d$-th roots of unity in $\mathbb{C}$. Then, quality of the basis $\mathbf{B}_{f,g}$ with respect to the Klein–GPV sampler admits the following simple expression:

$$(\alpha_{\mathrm{GPV}})^2 = \max\Big(\frac{1}{d}\sum_{i=1}^{d}\frac{|\varphi_i(f)|^2+|\varphi_i(g)|^2}{q},\ \frac{1}{d}\sum_{i=1}^{d}\frac{q}{|\varphi_i(f)|^2+|\varphi_i(g)|^2}\Big).$$

Similarly, the quality with respect to the hybrid sampler satisfies:

$$(\alpha_{\mathrm{hybrid}})^2 = \max_{1\leqslant i\leqslant d}\bigg(\max\Big(\frac{|\varphi_i(f)|^2+|\varphi_i(g)|^2}{q},\ \frac{q}{|\varphi_i(f)|^2+|\varphi_i(g)|^2}\Big)\bigg).$$

Note that $|\varphi_i(f)|^2+|\varphi_i(g)|^2 = \varphi_i(ff^*+gg^*)$ where the star denotes the complex conjugation automorphism of $\mathscr{R}$ (defined by $x^* = 1/x = -x^{d-1}$). Thus, put differently, one can say that a trapdoor $\mathbf{B}_{f,g}$ achieves quality α or better for the Klein–GPV sampler if and only if the embeddings of $(ff^*+gg^*)/q$ and of its inverse are at most α *on average*, whereas quality α or better is obtained for the hybrid sampler if *all* of the embeddings of these values are at most α. This shows in particular that the quality of a given trapdoor is always at least as good for Klein–GPV as it is for the hybrid sampler, which explains why it may be easier in practice to construct good quality trapdoors for the former than for the latter.

Trapdoor generation in Falcon and Mitaka. Now, the way trapdoors are generated in Falcon is by sampling f and g according to a discrete Gaussian in $\mathscr{R}$ (which can easily be done by sampling the coefficients as discrete Gaussians over $\mathbb{Z}$) so that their expected length is a bit over $\sqrt{q}$, and verifying using the condition above that the quality with respect to the Klein–GPV (or equivalently Falcon's) sampler is $\alpha_{\mathrm{Falcon}} = 1.17$ or better, and restarting otherwise (the value 1.17 here is chosen roughly as small as possible while keeping the number of repetitions relatively small).

The approach to generate trapdoors in Mitaka is similar using the quality formula for the hybrid sampler, and a target quality of $\alpha_{\mathrm{Mitaka}} = 2.04$ in dimension 512 (and slightly increasing as the dimension becomes larger). Doing so directly would take too many repetitions, however; therefore, the candidates for f and g are actually obtained by linear combinations of smaller Gaussian vectors and by applying Galois automorphisms to generate many candidate vectors (f,g) from a limited number of discrete Gaussian samples. Using that approach, Mitaka achieves the stated quality with a comparable number of discrete Gaussian samples as Falcon; its key generation algorithm is much slower, however, as it has to carry out an exhaustive search on a much larger set of possible candidates.

The Antrag strategy: annular NTRU trapdoor generation. In both Falcon and Mitaka, however, the overall strategy is to generate random-looking candidates (f,g) of plausible length, and repeat until the target quality is reached. In the Antrag paper [ENS+23a], we suggest a completely different strategy that is in some sense much simpler and more natural: just pick the pair (f,g) uniformly at random in the set

[8] The actual standard deviation also includes an additional factor (the smoothing parameter of the ring) which we omit in this overview for simplicity's sake.

of vectors that satisfy the desired quality level. We propose and analyze this approach specifically for the hybrid sampler.[9]

Concretely, yet another way of reformulating the quality condition for the hybrid sampler is to say that the quality is α or better if and only if for all the embeddings φ_i, one has:

$$q/\alpha^2 \leqslant |\varphi_i(f)|^2 + |\varphi_i(g)|^2 \leqslant \alpha^2 q.$$

In other words, for each embedding, the pair $\big(|\varphi_i(f)|, |\varphi_i(g)|\big)$ lies in the *annulus* $A\big(\sqrt{q}/\alpha, \alpha\sqrt{q}\big)$ bounded by the circles of radii $\sqrt{q}/\alpha$ and $\alpha\sqrt{q}$—or more precisely, in the *arc* $A_\alpha^+ = A^+\big(\sqrt{q}/\alpha, \alpha\sqrt{q}\big)$ of that annulus located in the upper-right quadrant of the plane since those absolute values are non-negative numbers. Our approach is then to sample f and g by their embeddings (i.e., directly in the Fourier domain), and select those embeddings uniformly and independently at random in the desired space. Namely, we sample $d/2$ pairs (x_i, y_i) in the arc of annulus A_α^+, and set the i-th embedding of f (resp. g) to a uniformly random complex number of absolute value x_i (resp. of absolute value y_i).

An obvious issue is that the elements f and g constructed in this way will generally not lie in the ring itself: after mapping back to the coefficient domain by Fourier inversion, their coefficients are *a priori* arbitrary real numbers instead of integers. But this is easy to address: we simply round coefficient-wise to obtain an actual ring element.

A second issue is that this rounding step will not necessarily preserve the quality property we started from: the embeddings of the rounded values do not necessarily remain in the correct domain. In fact, the probability that *all* embeddings remain in the correct domain after rounding is very low. But there is again a simple workaround: we just carry out our original continuous sampling in the Fourier domain from a slightly smaller annulus than the target one. Instead of picking the pairs (x_i, y_i) in A_α^+ as above, we sample them uniformly in some $A^+(r, R)$ with r slightly larger than $\sqrt{q}/\alpha$ and R slightly smaller than $\alpha\sqrt{q}$. This considerably increases the probability that, after rounding, all of the pairs $\big(|\varphi_i(f)|, |\varphi_i(g)|\big)$ will in fact end up in A_α^+.

And voilà: the description above is essentially a complete trapdoor generation algorithm for the hybrid sampler, that easily reaches the same NIST security levels as Falcon. Concretely, we target $\alpha = 1.15$ in dimension 512 (even better than Falcon's 1.17) and $\alpha = 1.23$ in dimension 1024 (which comfortably exceeds the 256 bits of classical CoreSVP security corresponding to NIST level V), and with those numbers, we achieve key generation speeds close to Falcon's, while benefiting of all the advantages of Mitaka in terms of simplicity of implementation, efficiency, parallelizability and so on as far as signing in concerned.

Contributions of Antrag. The main contribution of our Asiacrypt 2023 paper [ENS+23a] is to introduce, analyze and implement the Antrag trapdoor generation algorithm for the hybrid sampler described above.

The analysis includes a heuristic estimate of the success probability of sampling in the required domain, as well as a discussion of possible attacks on the resulting keys (and even though our security analysis is in a very optimistic model for the attacker, we find no weakness as long as the original sampling domain $A^+(r, R)$ is not chosen to be extremely narrow), and concrete parameters to instantiate a signature scheme.

We also provide a full portable C implementation of the corresponding signature scheme [Tib23] based on those of Falcon and Mitaka. In fact, since the C implementation of Mitaka did not include the key generation algorithm, our implementation is the first complete implementation of the corresponding paradigm. This implementation lets us compare the performance of our key generation with Falcon's, and we find that they are quite close.

Although most of the previous discussion was in the context of power-of-two cyclotomics, our approach also extends with little change to other base rings such as the cyclotomic rings with 3-smooth conductors considered in Mitaka (and we actually provide an analysis in a more general setting still). In particular, it is still possible to map candidate continuous random values generated in the Fourier domain to the ring

[9] One could consider doing so for Klein–GPV as well, but this appears less relevant for two reasons. First, since 1.17 is already quite close to the theoretical optimal quality of 1, and since the number of repetitions in Falcon's key generation is fairly modest, there is not much to gain in the Klein–GPV setting. Second, the space of key candidates has a less elegant geometric description, making it more difficult to sample uniformly in it. Extending the approach to ModFalcon [CPS+20], however, could be an interesting, albeit challenging, avenue for future research.

Table 1. Comparison of Antrag with Falcon and Mitaka for the same dimensions 512 and 1024 and the same modulus $q = 12289$ (excerpt from Table 2).

	Falcon [PFH+22]		Mitaka [EFG+22]		Antrag [ENS+23a]	
d	512	1024	512	1024	512	1024
Quality α	1.17	1.17	2.04	2.33	1.15	1.23
Classical sec.	123	284	102	233	124	264
Key size (bytes)	896	1792	896	1792	896	1792
Sig. size (bytes)	666	1280	713	1405	646	1260

by coefficient-wise rounding (we could consider other decoding techniques, but this one is sufficient for our purposes; it was in fact already used in the original ternary version of Falcon: see [PFH+17, Algorithm 10]). This only changes the distribution of the "rounding error" and hence the success probability slightly, but the analysis carries over easily. It follows that our approach supports the same versatility as Mitaka in terms of parameter selection.

2 Preliminaries

For two real numbers $0 \leqslant r \leqslant R$, we denote by $A(r, R)$ the *annulus* limited by radii r and R, i.e. the following subset of the plane $\mathbb{R}^2$: $A(r, R) := \{(x, y) \in \mathbb{R}^2 \mid r^2 \leqslant x^2 + y^2 \leqslant R^2\}$. We also denote by $A^+(r, R)$ the arc of annulus in the upper-right quadrant of the plane, i.e., $A^+(r, R) := \{(x, y) \in A(r, R) \mid x, y \geqslant 0\}$.

We note $\lfloor\cdot\rceil$ the rounding of a real number to its closest integer. We extend this notation for the coefficient-wise rounding of polynomials. If $\mathbf{x} = (x_1, \ldots, x_k)$ is a random variable, we let $\mathbb{E}[\mathbf{x}]$ the expected vector and $\mathrm{Cov}(\mathbf{x})$ its covariance matrix. The variance of a scalar random variable x is denoted by $\mathrm{Var}[x]$.

Write $\mathbf{A}^t$ for the transpose of any matrix $\mathbf{A}$. A lattice $\mathscr{L}$ is a discrete additive subgroup in a Euclidean space. When the space is $\mathbb{R}^m$, and if it is generated by (the columns of) $\mathbf{B} \in \mathbb{R}^{m\times d}$, we also write $\mathscr{L}(\mathbf{B}) = \{\mathbf{B}x | x \in \mathbb{Z}^d\}$. If $\mathbf{B}$ has full column rank, then we call $\mathbf{B}$ a basis and d the rank of $\mathscr{L}$. When the ambient space is equipped with a norm $||\cdot||$, the volume of $\mathscr{L}$ is $\mathrm{vol}(\mathscr{L}) = \det(\mathbf{B}^t\mathbf{B})^{1/2} = |\det(\mathbf{B})|$ for any basis $\mathbf{B}$.

2.1 Power-of-two cyclotomic fields

We focus on cyclotomic fields with the conductor m being a power of 2, i.e. $m = 2^l$ for a positive integer l. Let ζ to be a m-th primitive root of 1. Then for a fixed m, $\mathscr{K} := \mathbb{Q}(\zeta)$ is the cyclotomic field associated with the cyclotomic polynomial $x^d + 1$ where $d = m/2 = 2^{l-1}$, and its ring of algebraic integers is $\mathscr{R} := \mathbb{Z}[\zeta]$. We have $\mathscr{K} \simeq \mathbb{Q}[x]/(x^d + 1)$ and $\mathscr{R} \simeq \mathbb{Z}[x]/(x^d + 1)$, and both are contained in $\mathscr{K}_\mathbb{R} := \mathscr{K} \otimes \mathbb{R} = \mathbb{R}[x]/(x^d + 1)$. Each $f = \sum_{i=0}^{d-1} f_i\zeta^i \in \mathscr{K}_\mathbb{R}$ can be identified with its coefficient vector $(f_0, \cdots, f_{d-1}) \in \mathbb{R}^d$. The field automorphism induced by $\zeta \mapsto \zeta^{-1} = \bar{\zeta}$ corresponds to the complex conjugation, and we write $f^* = (f_0, -f_{d-1}, \ldots, -f_1)$ the image of $f \in \mathscr{K}$ under this automorphism. The complex conjugation operation extends naturally to $\mathscr{K}_\mathbb{R}$, and $\mathscr{K}_\mathbb{R}^+$ is the subspace of elements satisfying $f^* = f$. The cyclotomic field $\mathscr{K}$ comes with d complex field embeddings $\varphi_i : \mathscr{K} \to \mathbb{C}$ that maps f seen as a polynomial to its evaluations at ζ^k where $k < m$ is an odd integer.This defines the so-called canonical embedding $\varphi(f) := (\varphi_1(f), \ldots, \varphi_d(f))$. It extends straightforwardly to $\mathscr{K}_\mathbb{R}$ and identifies it to the space $\mathcal{H} = \{v \in \mathbb{C}^d : v_i = \overline{v_{d/2+i}}, 1 \leqslant i \leqslant d/2\}$. Note that $\varphi(fg) = (\varphi_i(f)\varphi_i(g))_{0<i\leqslant d}$. When needed, this embedding extends entry-wise to vectors or matrices over $\mathscr{K}_\mathbb{R}$. The next technical lemma is useful in our analyses, and is obtained by elementary trigonometric identities.

Lemma 1. *Let ζ be a m-th primitive root of the unity. Then we have $\sum_{j=0}^{d-1} \zeta^{2j} = 0$.*

2.2 NTRU lattices

This work deals with free $\mathscr{R}$-modules of rank 2 in $\mathscr{K}^2$, or in other words, groups of the form $\mathscr{M} = \mathscr{R}\mathbf{x} + \mathscr{R}\mathbf{y}$ where $\mathbf{x} = (x_1, x_2), \mathbf{y} = (y_1, y_2)$ span $\mathscr{K}^2$. Given $f, g \in \mathscr{R}$ such that f is invertible modulo some prime

$q \in \mathbb{Z}$, we let $h = f^{-1}g \mod q$. The NTRU module determined by h is $\mathscr{L}_{\mathrm{NTRU}} = \{(u,v) \in \mathscr{R}^2 : uh - v = 0 \mod q\}$. Two bases of this free module are of particular interest:

$$\mathbf{B}_h = \begin{bmatrix} 1 & 0 \\ h & q \end{bmatrix} \quad \text{and} \quad \mathbf{B}_{f,g} = \begin{bmatrix} f & F \\ g & G \end{bmatrix},$$

where $F, G \in \mathscr{R}$ are such that $fG - gF = q$ and (F, G) should be relatively small. This module is usually seen as a lattice of volume q^d in $\mathbb{R}^{2d}$ in the coefficient embedding.

We equip the ambient space $\mathscr{K}_{\mathbb{R}}^2$ with the inner product $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathscr{K}} = x_1^* y_1 + x_2^* y_2$. The well-known Gram-Schmidt orthogonalization procedure for a pair of linearly independent vectors $\mathbf{b}_1, \mathbf{b}_2 \in \mathscr{K}^2$ is defined as

$$\widetilde{\mathbf{b}}_1 := \mathbf{b}_1, \widetilde{\mathbf{b}}_2 := \mathbf{b}_2 - \frac{\langle \mathbf{b}_1, \mathbf{b}_2 \rangle_{\mathscr{K}}}{\langle \mathbf{b}_1, \mathbf{b}_1 \rangle_{\mathscr{K}}} \cdot \widetilde{\mathbf{b}}_1.$$

One readily checks that $\langle \widetilde{\mathbf{b}}_1, \widetilde{\mathbf{b}}_2 \rangle = 0$. The Gram-Schmidt matrix with columns $\widetilde{\mathbf{b}}_1, \widetilde{\mathbf{b}}_2$ is denoted by $\widetilde{\mathbf{B}}$ and we have $\det \widetilde{\mathbf{B}} = \det \mathbf{B}$. We also let $|\mathbf{B}|_{\mathscr{K}} = \max(||\varphi(\langle \widetilde{\mathbf{b}_1}, \widetilde{\mathbf{b}_1} \rangle)||_\infty, ||\varphi(\langle \widetilde{\mathbf{b}_2}, \widetilde{\mathbf{b}_2} \rangle)||_\infty)^{1/2}$.

Lemma 2. *Let $\mathbf{B}_{f,g}$ be a basis of an NTRU module and $\mathbf{b}_1 = (f, g)$. We have $\sqrt{q} \leqslant |\mathbf{B}_{f,g}|_{\mathscr{K}}$ and*

$$|\mathbf{B}_{f,g}|_{\mathscr{K}}^2 = \max\left(||\varphi(\langle \mathbf{b}_1, \mathbf{b}_1 \rangle_{\mathscr{K}})||_\infty, \left\| \frac{q^2}{\varphi(\langle \mathbf{b}_1, \mathbf{b}_1 \rangle_{\mathscr{K}})} \right\|_\infty \right).$$

2.3 Gaussian and chi-squared distributions

For $\mu \in \mathbb{R}$ and $\sigma > 0$ we let $\mathcal{N}(\mu, \sigma^2)$ be the normal distribution of mean μ and standard deviation σ, that is, the continuous distribution over $\mathbb{R}$ with density proportional to $\exp\big(-(x-\mu)^2/(2\sigma^2)\big)$. In higher dimensions, for Σ a positive definite matrix and a vector $\mu \in \mathbb{R}^k$, we let $\mathcal{N}(\mu, \Sigma)$ be the normal distribution of density proportional to $\exp\big(-\frac{1}{2}(x-\mu)^t \Sigma^{-1} (x-\mu)\big)$.

Let $T \sim \mathcal{N}(\mu, \sigma^2 \mathbf{I}_k)$ be a k-dimensional spherical normal random vector. The random variable $\|T\|^2$ follows a *non central chi-squared distribution of degree k, non-centrality $c := \|\mu\|^2$ and scaling* σ^2, denoted by $\chi^2(k, \sigma^2; c)$. Its expectation, variance and cumulative distribution function are described by the following classical result.

Lemma 3. *Let U be a random variable distributed as $\chi^2(k, \sigma^2; c)$. We have $\mathbb{E}[U] = \sigma^2 k + c$ and $\mathrm{Var}[U] = 2\sigma^2(\sigma^2 k + 2c)$. For $0 \leqslant a < b$, we have $\mathbb{P}[a \leqslant U \leqslant b] = Q_{k/2}(\sqrt{c}/\sigma, \sqrt{a}/\sigma) - Q_{k/2}(\sqrt{c}/\sigma, \sqrt{b}/\sigma)$, where $Q_{k/2}$ is the Marcum Q-function of order $k/2$.*

Moreover, the Marcum Q-function Q_m of integer order m satisfies the following inequalities.

Lemma 4 ([SA00, AT01]). *For integer m and $u, v \geqslant 0$, the following inequalities hold:*

$$Q_m(u,v) \geqslant 1 - \frac{1}{2} e^{-(u-v)^2/2} \qquad \text{if } u \geqslant v;$$
$$Q_m(u,v) \leqslant e^{-(v-u)^2/2} \cdot \left(1 + \frac{(v/u)^{m-1} - 1}{\pi \cdot (1 - u/v)} \right) \qquad \text{if } u \leqslant v.$$

3 New trapdoor algorithms for hybrid sampling

3.1 Hash-then-sign over lattices in a nutshell

The rationale behind this design is that a signature corresponds to a *short* Gaussian vector in a lattice $\mathscr{L}_{\mathrm{NTRU}}$ centered at the hash of a (salted) message. On the one hand, these vectors can only be generated efficiently with the knowledge of a trapdoor $\mathbf{B}_{f,g}$, that is, a basis with good quality for a given sampling method. On the other hand, verifying amounts to checking lattice membership and that the vector is indeed shorter than a threshold. For the sake of completeness, we recap this design in the form of high-level, generic algorithms `KeyGen`, `Sign`, `Verify` corresponding to the current efficient instantiations.

Algorithm 1: Signing

Input: A message m, a trapdoor $\mathbf{B}_{f,g}$, a standard deviation parameter σ
Result: the first component s_0 of $\mathbf{s} = (s_0, s_1) \in \mathscr{R}^2$ such that $\mathbf{c} - \mathbf{s}$ has a distribution close to $D_{\mathscr{L}_{\mathrm{NTRU}},\mathbf{c},\sigma}$.

$r \leftarrow^{\$} \{0,1\}^{320}$
$\mathbf{c} \leftarrow (0, \mathsf{H}(\mathtt{m}||r))$
$\mathbf{v} \leftarrow \mathtt{Sample}(\mathbf{B}_{f,g}, \mathbf{c}, \sigma)$
$(s_0, s_1) \leftarrow \mathbf{c} - \mathbf{v}$
return s_0

Algorithm 2: Verification

Input: A message m, a salt r, $s_0 \in \mathscr{R}$, a public key h and a threshold β
Result: Accept or reject

$s_1 \leftarrow \mathsf{H}(\mathtt{m}||r) + s_0 h \bmod q$
if $\|(s_0, s_1)\| > \beta$ **then**
  Reject.
end if
Accept.

In Algorithm 1, the procedure `Sample` differs from Falcon to Mitaka. The former relies on the FFO sampler (a Fast-Fourier-like version of the GPV sampler [GPV08], while the latter prefers the simpler hybrid sampler of Ducas-Prest [DP15]. Lattice membership is implicitly checked at the first step of Algorithm 2. We finish the section with a high-level description of `KeyGen` in Algorithm 3. Its purpose is to generate a pair $(h, \mathbf{B}_{f,g})$ where $\mathbf{B}_{f,g}$ should have a good quality with respect to the selected instantiation of `Sample`. For simplicity, we omit in its description the additional secret data related to the sampler. The procedure `GoodPair`, our focus in this work, outputs $(f, g) \in \mathscr{R}^2$ with the guarantee that the basis $\mathbf{B}_{f,g}$ output by `NTRUSolve` will have quality α or better for the choice of `Sample`.

3.2 NTRU trapdoors in Falcon and Mitaka

With respect to Prest's hybrid sampler, an NTRU trapdoor $\mathbf{B}_{f,g}$ has a quality α defined as

$$\alpha = |\mathbf{B}_{f,g}|_{\mathscr{K}} / \sqrt{q}, \tag{1}$$

where we recall that $|\mathbf{B}_{f,g}|^2_{\mathscr{K}} = \max\left(\|\varphi(ff^* + gg^*)\|_\infty, \left\|\frac{q^2}{\varphi(ff^*+gg*)}\right\|_\infty\right)$. The quality with respect to the Klein–GPV sampler admits a similar expression.

In hash-and-sign signatures, security against forgery attacks is driven by the standard deviation of the sampler, which is essentially $\alpha\sqrt{q}$. As the smaller the value of α, the harder forgery becomes, the goal of `KeyGen` in schemes such as DLP [DLP14], Falcon [PFH+22] and Mitaka [EFG+22] is to construct in reasonable time bases $\mathbf{B}_{f,g}$ with α as small as possible (and in particular, smaller than a given threshold related to the acceptance radius of signature verification). In other words, the goal is to instantiate efficiently the procedure `GoodPair`.

An important observation regarding NTRU trapdoors is that the knowledge of the first basis vector (f, g) alone is sufficient to determine the quality of the whole basis (see for example Lemma 2 for Mitaka). As a result, to test if a vector (f, g) can be completed into a trapdoor $\mathbf{B}_{f,g}$ reaching the desired quality threshold, it is not necessary to compute the second vector (F, G), which is a notoriously costly operation, even accounting for optimizations such as [PP19].

In DLP, Falcon and Mitaka, `GoodPair` is a trial-and-error routine, generating many potential candidate first vectors (f, g) and testing whether they satisfy the required quality threshold. The candidates themselves are generated as discrete Gaussian vectors in $\mathscr{R}^2$ with the correct expected length. In that way, Falcon reaches quality $\alpha = 1.17$ with respect to its FFO-based sampler (that admits the same quality metric as

Algorithm 3: Generic NTRU Trapdoor generator

Input: A degree d, a modulus q, a target quality α
Result: a public key $h \in \mathscr{R}$ and the trapdoor $\mathbf{B}_{f,g}$

$(f, g) \leftarrow \texttt{GoodPair}(d, q, \alpha)$
$\mathbf{B}_{f,g} \leftarrow \texttt{NTRUSolve}(f, g, q)$
$h \leftarrow gf^{-1} \bmod q$
return $(h, \mathbf{B}_{f,g})$.

Algorithm 4: Candidate pairs from uniform annulus sampling

Input: $0 < r < R$, the radii of $A^+(r, R)$
Result: $z, z' \in \mathbb{C}$ such that $\big(|z|, |z'|\big)$ is uniformly distributed in $A^+(r, R)$

$u \hookleftarrow \mathcal{U}\left([r^2, R^2]\right)$
$\rho \leftarrow \sqrt{u}$
$\theta \hookleftarrow \mathcal{U}\left([0, \pi/2]\right)$
$(x, y) \leftarrow \big(\rho\cos\theta, \rho\sin\theta\big)$ /* $(x, y) \hookleftarrow \mathcal{U}\left(A^+(r, R)\right)$ */
$\omega, \omega' \hookleftarrow \mathcal{U}\left([0, 2\pi]\right)$
$(z, z') \leftarrow \left(x \cdot e^{i\omega}, y \cdot e^{i\omega'}\right)$
return (z, z')

Klein–GPV). Doing this directly for the hybrid sampler, as discussed in [Pre15], only achieves quality $\gtrsim 3$ in dimension 512, and even larger in higher dimensions. As a result, the Mitaka paper has to introduce randomness recycling and other techniques on top of this general approach in order to increase the number of candidates and improve the achievable quality; with those improvements, Mitaka reaches $\alpha = 2.04$ in dimension 512 (which translates to 20 fewer bits of security compared to Falcon, and is thus unfortunately not sufficient to reach NIST security level I).

3.3 Antrag: annular NTRU trapdoor generation

The main contribution of the Antrag paper [ENS+23a] is a novel instantiation of `GoodPair` for the hybrid sampler, resulting in a NTRU trapdoor generation algorithm achieving much better quality than Mitaka, while reaching the same security NIST levels as Falcon.

The intuition behind our new approach stems from the following observation. For a fixed $\alpha \geqslant 1$, requiring a trapdoor $\mathbf{B}_{f,g}$ to satisfy $|\mathbf{B}_{f,g}|_{\mathscr{K}} \leqslant \alpha\sqrt{q}$ is equivalent to enforcing that for all $1 \leqslant i \leqslant d$, we have

$$\frac{q}{\alpha^2} \leqslant |\varphi_i(f)|^2 + |\varphi_i(g)|^2 \leqslant \alpha^2 q, \tag{2}$$

(where we recall that the $\varphi_i(f)$ are the *embeddings* of f in $\mathbb{C}$, and similarly for g). Equivalently, this means that for all i, the pair $\big(|\varphi_i(f)|, |\varphi_i(g)|\big)$ belongs to the arc of annulus $A^+_\alpha := A^+\big(\sqrt{q}/\alpha, \alpha\sqrt{q}\big)$.

It is thus natural to try and sample f and g *from their embeddings* (i.e., in the Fourier domain), by picking the pairs $\big(\varphi_i(f), \varphi_i(g)\big)$ as *uniform* random pairs of complex numbers such that satisfying the condition that the pair of their magnitudes belongs to A^+_α: in other words, pick (x_i, y_i) uniformly at random in A^+_α and then sample $\varphi_i(f)$ and $\varphi_i(g)$ as uniform complex numbers of magnitudes x_i and y_i respectively. Note that only $d/2$ pairs are needed, as the remaining ones are determined by conjugation.

Moreover, sampling uniformly in an annulus (or, as in our case, an arc of annulus) in polar coordinates (ρ, θ) is easy: it suffices to sample the angle θ and the *square* ρ^2 of the radial coordinate uniformly in their respective ranges. This is because the area element in polar coordinates is $\rho\, d\rho\, d\theta = \frac{1}{2} d(\rho^2)\, d\theta$. This gives rise to Algorithm 4 for sampling the pairs of embeddings.

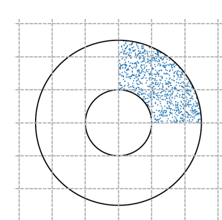


Fig. 1. $(|z|, |w|)$ is sampled uniformly in the annulus $A^+(r, R)$.

Algorithm 5: ANTRAG trapdoor generation

Input: The degree d, the modulus q, a target quality α, and starting radii r, R such that $\sqrt{q}/\alpha < r < R < \alpha\sqrt{q}$.

Result: $f, g \in \mathscr{R}^2$ such that $\frac{q}{\alpha^2} \leqslant |\varphi_i(f)|^2 + |\varphi_i(g)|^2 \leqslant \alpha^2 q$ for all i.

repeat
- **for** $1 \leqslant i \leqslant d/2$ **do**
  - using Algorithm 4, sample $(z_i, w_i) \in \mathbb{C}^2$ uniformly such that $\big(|z_i|, |w_i|\big) \in A^+(r, R)$.
- **end for**
- $\tilde{f} \leftarrow \varphi^{-1}(z_1, \ldots, z_{d/2}) \in \mathscr{K}_\mathbb{R}$
- $\tilde{g} \leftarrow \varphi^{-1}(w_1, \ldots, w_{d/2}) \in \mathscr{K}_\mathbb{R}$
- $f \leftarrow \lfloor \tilde{f} \rceil$
- $g \leftarrow \lfloor \tilde{g} \rceil$

until $\big(|\varphi_i(f)|, |\varphi_i(g)|\big) \in A^+\big(\sqrt{q}/\alpha, \alpha\sqrt{q}\big)$ *for all* $i = 1, \ldots, d/2$

return (f, g)

However, one soon realizes that the real polynomials $\tilde{f}, \tilde{g}$ corresponding to the embeddings generated by the Algorithm 4 (via the inverse Fourier transform φ^{-1}) do not always have integer coefficients, and hence do not generally correspond to ring elements. In general, they are elements of the $\mathbb{R}$-algebra $\mathscr{K}_\mathbb{R}$.

In order to obtain actual ring elements, a natural solution is to round those real polynomials $\tilde{f}, \tilde{g}$ coefficient-wise. This yields $f = \lfloor \tilde{f} \rceil$ and $g = \lfloor \tilde{g} \rceil$ in $\mathscr{R}$, which are potential candidates for a trapdoor. It turns out, however, that if one starts from $\tilde{f}, \tilde{g}$ uniform with their embeddings of magnitude in A^+_α, the resulting rounded ring elements are very unlikely to also have their embeddings of magnitude in that arc of annulus. Thus, they do not typically give rise to a trapdoor of the desired quality. This is because rounding adds an additive term (essentially uniformly distributed in $[-1/2, 1/2)$) to each coefficient, which translates to an additive "error" on each embedding, making it unlikely that the embeddings all remain in the desired domain.

A straightforward workaround is to compensate this decoding error by sampling the embeddings of $\tilde{f}, \tilde{g}$ from a narrower annulus $A^+(r, R)$ for some radii r, R such that $\sqrt{q}/\alpha < r < R < \alpha\sqrt{q}$. This yields Algorithm 5, which is our proposed ANTRAG trapdoor generation algorithm.

Remark 1. One could consider carrying out the decoding to the ring differently, for example by sampling discrete Gaussians f and g in $\mathscr{R}$ centered at $\tilde{f}$ and $\tilde{g}$ respectively. The resulting algorithm would be simpler to analyze in some ways, and might be seen as better behaved in a certain sense, but it does have a major drawback: it introduces a much larger decoding error (on the order of the smoothing parameter $\eta_\varepsilon(\mathbb{Z})$ of $\mathbb{Z}$ on each coefficient, instead of the standard deviation $1/\sqrt{12}$ of the uniform distribution in $[-1/2, 1/2)$, so about 4 times larger). As a result, in this work, we focus on the rounding approach.

3.4 On the distribution of embeddings

We have mentioned above that taking the magnitudes of the embeddings of $\tilde{f}$ and $\tilde{g}$ in A^+_α was very unlikely to result in f and g of the required quality α after rounding, but that the probability increased greatly when choosing $\tilde{f}$ and $\tilde{g}$ with embedding magnitudes in a narrower arc of annulus $A^+(r, R)$. We choose the bounds r and R as complementary convex combinations of $\alpha\sqrt{q}$ and $\sqrt{q}/\alpha$; in other words, we set:

$$r = \frac{1-\xi}{2}\alpha\sqrt{q} + \frac{1+\xi}{2} \cdot \frac{\sqrt{q}}{\alpha} \quad \text{and} \quad R = \frac{1+\xi}{2}\alpha\sqrt{q} + \frac{1-\xi}{2} \cdot \frac{\sqrt{q}}{\alpha} \tag{3}$$

for some constant $\xi \in (0, 1)$, so that $A^+(r, R)$ corresponds to the middle ξ-fraction of A^+_α. We will later specifically choose $\xi = 1/3$ (i.e., $A^+(r, R)$ as the "middle third" of A^+_α) to fix ideas, and because it yields the following expression for r and R with minimal coefficient height:

$$r = \Big(\frac{1}{3}\alpha + \frac{2}{3} \cdot \frac{1}{\alpha}\Big)\sqrt{q} \quad \text{and} \quad R = \Big(\frac{2}{3}\alpha + \frac{1}{3} \cdot \frac{1}{\alpha}\Big)\sqrt{q}.$$

In this section, we would like to provide a model on two-power cyclotomic base fields $\mathscr{K}$ allowing us to quantify the claim that sampling $\tilde{f}$ and $\tilde{g}$ in this $A^+(r,R)$ increases success probability (see the full version of the ANTRAG paper [ENS+23b] for non-two-power $\mathscr{K}$). To that end, write $e = (e_f, e_g) = \big(f - \tilde{f}, g - \tilde{g}\big) \in \mathscr{K}_\mathbb{R}^2$ for the error term introduced by rounding. We would like to control the distribution of the embeddings of e_f and e_g in order to estimate the likelihood that the condition $\big(|\varphi_\theta(f)|, |\varphi_\theta(g)|\big)$ will be satisfied for all θ.

In the polynomial basis, we write $e_f = \sum_{j=0}^{d-1} e_f^{(j)} x^j$ and similarly for e_g. Heuristically, we expect the coefficients $e_f^{(j)}$ and $e_g^{(j)}$ to behave essentially like independent uniform random variables in $[-1/2, 1/2)$.[10] This is well-supported by experiments (see the full version of the ANTRAG paper [ENS+23b]).

Now consider a single embedding φ_θ, and recall that we are interested in an *a priori* arbitrary cyclotomic base ring, so that φ_θ is defined by the evaluation at some primitive m-th root of unity $\zeta = e^{i\theta}$. We therefore have:

$$\varphi_\theta(e_f) = x_\theta + i y_\theta \quad \text{with} \quad x_\theta = \sum_{j=0}^{d-1} e_f^{(j)} \cos(j\theta) \quad \text{and} \quad y_\theta = \sum_{j=0}^{d-1} e_f^{(j)} \sin(j\theta).$$

This expresses the real and imaginary parts x_θ, y_θ of $\varphi_\theta(e_f)$ as the sum of d independent random variables, with d relatively large, so by the central limit theorem, $\varphi_\theta(e_f)$ should essentially behave[11] like a normal random variable in $\mathbb{C}$, essentially determined by its expectation and covariance.

Now since $e_f^{(j)}$ has mean 0 and variance 1/12 for all j, we obtain that $\mathbb{E}[x_\theta] = \mathbb{E}[y_\theta] = 0$. Therefore, the pair (x_θ, y_θ) has mean 0, and its covariance matrix is easily computed from Lemma 1 as follows:

$$\Sigma_\theta = \frac{d}{24}\mathbf{I}_2 =: \Sigma.$$

We thus expect that $\varphi_\theta(e_f)$ follows the normal distribution $\mathcal{N}(0, \Sigma)$, and the same argument applies to $\varphi_\theta(e_g)$ as well. Moreover, heuristically, those two normal distributions should be independent (this is again well-verified in practice), therefore, we can write

$$(\varphi_\theta(e_f), \varphi_\theta(e_g)) \sim \mathcal{N}\left(0, \begin{pmatrix} \Sigma & 0 \\ 0 & \Sigma \end{pmatrix}\right) \tag{4}$$

This leads us to model the distribution of the embeddings of secret keys as follows.

Heuristic 1. Let $(f, g) \in \mathscr{K}^2$ a pair output by Algorithm 5, corresponding to $(\tilde{f}, \tilde{g}) \in \mathscr{K}_\mathbb{R}^2$ obtained from the executions of Algorithm 4. For the embedding φ_θ corresponding to the primitive root of unity $e^{i\theta}$, $\big(\varphi_\theta(f), \varphi_\theta(g)\big)$ is distributed as

$$\big(\varphi_\theta(f),\ \varphi_\theta(g)\big) \sim \mathcal{N}\big(\,\big(\varphi_\theta(\tilde{f}),\ \varphi_\theta(\tilde{g})\big)\,,\ \mathbf{I}_2 \otimes \Sigma\big).$$

Moreover, the pairs $(\varphi_\theta(f), \varphi_\theta(g))$ as φ_θ ranges through all the embeddings of $\mathscr{K}$ are independently distributed.

Note that this heuristic considers the pair $(\varphi_\theta(f), \varphi_\theta(g))$, which is actually supported on dense but countable subgroup of $\mathbb{C}^2$, as following a *continuous* distribution. This has the merit of allowing an analysis while being an accurate representation of the situation according to our experiments.

This heuristic also allows us to express the expected length of the embeddings of secret keys and related elements. This is used the security analysis; more details are provided in the full version of the ANTRAG paper [ENS+23b].

4 Success probability and security analysis

In this section, we concentrate on the case of a power-of-two cyclotomic base ring, in which, under Heuristic 1, all the embeddings of f and g are simply modeled as independent and identically distributed isotropic normal variates, which simplifies the analysis somewhat. In this context, we analyze the success probability of Algorithm 5 as well as the security of the resulting scheme, which lets us derive concrete parameters.

[10] This is equivalent to saying that the distribution of $\tilde{f}$ and $\tilde{g}$ is uniform modulo $\mathscr{R}$ in $\mathscr{K}_\mathbb{R}$, which should indeed happen as soon as we have sufficient width (i.e., if we exceed a regularity metric analogous to the smoothing parameters for Gaussians).

[11] This can in fact be made rigorous with the Berry–Esseen theorem.

4.1 Success probability over power-of-two cyclotomics

Suppose that $\mathscr{K}$ is a cyclotomic field of conductor a power of two, and let $(\tilde{f}, \tilde{g}) \in \mathscr{K}_{\mathbb{R}}^2$ and $(f, g) \in \mathscr{R}^2$ be generated as in Steps 5–6 and Steps 7–8 of Algorithm 5 respectively.

We first fix one embedding $\varphi_\theta \colon \mathscr{K} \to \mathbb{C}$ of $\mathscr{K}$, and try to determine the probability with which the test of Step 10 of Algorithm 5 is satisfied with respect to that particular embedding. In other words, we want to estimate the probability that:

$$q/\alpha^2 \leqslant |\varphi_\theta(f)|^2 + |\varphi_\theta(g)|^2 \leqslant \alpha^2 q. \tag{5}$$

Now, according to Heuristic 1, the pair $\big(\varphi_\theta(f), \varphi_\theta(g)\big) \in \mathbb{C}^2$ follows a normal distribution centered at $\big(\varphi_\theta(\tilde{f}), \varphi_\theta(\tilde{g})\big)$ of scalar covariance $\frac{d}{24}\mathbf{I}_4$. Therefore, for fixed $(\tilde{f}, \tilde{g})$ and following the definitions of Section 2.3, the squared norm:

$$\big\|\big(\varphi_\theta(f), \varphi_\theta(g)\big)\big\|^2 = |\varphi_\theta(f)|^2 + |\varphi_\theta(g)|^2$$

follows a non central chi-squared distribution $\chi^2(4, \sigma^2; c)$ of degree 4, non-centrality $c = |\varphi_\theta(\tilde{f})|^2 + |\varphi_\theta(\tilde{g})|^2$ and scaling $\sigma^2 = d/24$. In particular, the probability that condition (5) does not depend on the exact position of the pair $\big(\varphi_\theta(\tilde{f}), \varphi_\theta(\tilde{g})\big)$, but only on its squared norm c, or equivalently on:

$$\beta \coloneqq \frac{1}{\sqrt{q}}\big\|\big(\varphi_\theta(\tilde{f}), \varphi_\theta(\tilde{g})\big)\big\|.$$

We denote the probability that condition (5) is satisfied for a certain value β by $p_{\text{succ}}(\beta)$. According to Lemma 3, the probability $p_{\text{succ}}(\beta)$ can be expressed in terms of the Marcum Q-function Q_2 as follows:

$$p_{\text{succ}}(\beta) = Q_2(\tau\beta, \tau/\alpha) - Q_2(\tau\beta, \tau\alpha) \quad \text{where} \quad \tau = \sqrt{\frac{24q}{d}}.$$

Based on this result, we will first provide a simple but loose lower bound of the success probability of Algorithm 5, and then derive a more complicated but tight estimate that we can use for numerical estimates and parameter selection.

Bounding the success probability below. The *overall* success probability $p_{\text{succ-one}}$ for a single embedding (which is the probability that condition (5) holds when the starting embedding pair $\big(|\varphi_\theta(\tilde{f})|, |\varphi_\theta(\tilde{g})|\big)$ is sampled uniformly in $A^+(r, R)$) and similarly lower bounded as:

$$p_{\text{succ-one}} \geqslant 1 - K_\xi u_\tau\big((1-\xi)\delta\big) \quad \text{with } K_\xi = \frac{3}{2} + \frac{2/\pi}{1-\xi} \text{ and } u_\tau(x) = \exp\big(-\frac{\tau^2}{2}x^2\big) \tag{6}$$

and under our independence heuristic, the success probability $p_{\text{succ-all}}$ for all $d/2$ embeddings at the same time satisfies:

$$p_{\text{succ-all}} \geqslant \Big(1 - K_\xi u_\tau\big((1-\xi)\delta\big)\Big)^{d/2}.$$

To reach an overall success probability of $1/M$ (i.e., M repetitions on average), it therefore suffices to have:

$$\frac{d}{2}\log\Big(1 - K_\xi u_\tau\big((1-\xi)\delta\big)\Big) \geqslant -\log M.$$

Finally, a quality α is achievable (with repetition rate up to M) as long as:

$$\alpha \geqslant \sqrt{A} + \sqrt{1+A} \quad \text{where} \quad A = \frac{d}{12(1-\xi)^2 q}\log\frac{K_\xi d}{2\log M}. \tag{7}$$

In particular, we see that, as long as $q = \Omega(d\log d)$, quality measures $\alpha = O(1)$ are achievable with any constant repetition rate. This is similar to Falcon and unlike Mitaka [EFG+21, Appendix C] and the original approach for the Peikert and hybrid samplers [Pre15], where α increases as a power function of the dimension independently of q.

As discussed in the previous section, we choose $\xi = 1/3$ to fix ideas, so that the starting annulus becomes the "middle third" of the target annulus (we will see below that this choice is very safe). Condition (7) above with $M = 4$ and $q = 12289$ shows that one can reach quality at least $\alpha = 1.24$ in dimension 512 and $\alpha = 1.38$ in dimension 1024 with this modulus q and repetition rate up to 4. This is already much better than the quality parameters achievable by Mitaka, but since we have used loose inequalities throughout, these are actually rough lower bounds.

More precise expression of success probability. For concrete parameter selection, and also to test the validity of our heuristic assumptions, it is useful to write down the exact expression of success probability according to our model.

Recall that the success probability $p_{\text{succ-one}}$ for a single embedding is the probability that condition (5) holds when the starting embedding pair $\big(|\varphi_\theta(\tilde{f})|, |\varphi_\theta(\tilde{g})|\big)$ is sampled uniformly in $A^+(r,R)$. In other words, $p_{\text{succ-one}}$ is the expected value of $p_{\text{succ}}(\beta)$ for β^2 uniformly distributed in $[r^2/q, R^2/q]$. Then by carrying out the change of variables, we obtain the dependence of $p_{\text{succ-one}}$ on the quality α:

$$p_{\text{succ-one}} = \frac{1}{2\xi}\int_{-\xi}^{\xi} F(\alpha,t)\cdot\Big(1 + t\frac{\alpha-\frac{1}{\alpha}}{\alpha+\frac{1}{\alpha}}\Big)\,dt$$

where

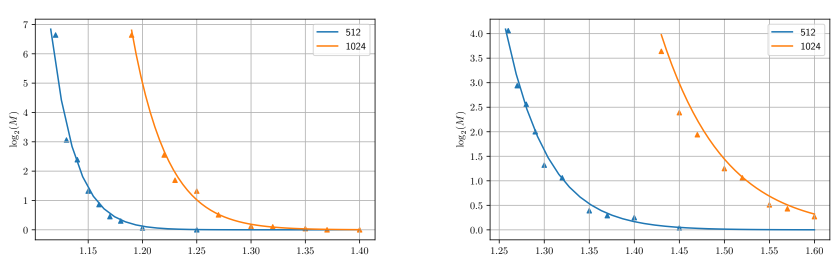
$$F(\alpha,t) = Q_2\bigg(\tau\Big(\frac{\alpha+\frac{1}{\alpha}}{2} + t\frac{\alpha-\frac{1}{\alpha}}{2}\Big), \tau/\alpha\bigg) - Q_2\bigg(\tau\Big(\frac{\alpha+\frac{1}{\alpha}}{2} + t\frac{\alpha-\frac{1}{\alpha}}{2}\Big), \tau\alpha\bigg).$$

Therefore, $1/M = p_{\text{succ-all}} = p_{\text{succ-one}}^{d/2}$. This makes it easy to solve numerically for α in order to reach a certain repetition rate. Again for $q = 12289$, we find that we reach repetition rate $M = 4$ for $\alpha \approx 1.143$ in dimension $d = 512$, and for $\alpha \approx 1.229$ for $d = 1024$. For $q = 3329$, the same repetition rate is reached for $\alpha \approx 1.290$ for $d = 512$ and $\alpha \approx 1.478$ for $d = 1024$. Moreover, this allows us to confirm that our model very closely matches experiments, as demonstrated on Fig. 2.

4.2 Security analysis

In order to assess the concrete security of the resulting signature scheme, the usual cryptanalytic methodology amounts to estimating the complexity of the best attacks against *key recovery attacks* on the one hand, and *signature forgery* on the other. In the hash-and-sign paradigm, the security of the forgery is a function of the standard deviation of the lattice Gaussian sampler used in the signature function, which itself depends on the quality α of the trapdoor. A first straightforward observation is that, since our work has only modified *which* trapdoors are used for signing, and not *how* they are used in signing, our modifications cannot have a negative impact on the resilience against forgery. On the contrary, we have shown how to increase the trapdoor quality, and therefore our new approach increases the security against forging attackers.

As such, only the resilience to key recovery attacks could be impacted by this work. In our setting, lattice reduction approaches would give the best result (as combinatorial or hybrid attacks are irrelevant in our setting, with dense, non-ternary keys). Nevertheless, by changing the sampling of the good trapdoors, we might have restricted to a possibly smaller set of secret keys, or to a possibly much more geometrically constrained set of keys. Indeed, all their complex embeddings must lie in a publicly described annulus, so an adversary could use this additional information to gather more power for an attack. Informally, they could



(a) $q = 12289$ (b) $q = 3329$

Fig. 2. Base 2 logarithm of the repetition rate M of Algorithm 5 as a function of α, for $d \in \{512, 1024\}$ and $q \in \{12289, 3329\}$. The continuous lines are obtained based on our model, and the triangle data points are measured by simulations (averaging 100 iterations of the algorithm for each data point).

Table 2. Performance comparison of ANTRAG with FALCON and MITAKA.

	FALCON [PFH+22]		MITAKA [EFG+22]		ANTRAG [ENS+23a]	
d	512	1024	512	1024	512	1024
Quality α	1.17	1.17	2.04	2.33	1.15	1.23
Classical sec.	123	284	102	233	124	264
Key size (bytes)	896	1792	896	1792	896	1792
Sig. size (bytes)	666	1280	713	1405	646	1260
`keygen` speed (Mcycles)	—	—	—	—	9.5	33.2
`keygen` speed (ms)	4.2	12.4	*1657* *	*6214* *	3.5	12.3
`sign` speed (kcycles)	—	—	299	584	298	586
`sign` speed (μs)	184	371	111	217	111	218
`verif` speed (kcycles)	—	—	20	41	20	40
`verif` speed (μs)	18	36	8	16	8	15

* Timings for the optimized SageMath implementation (excluding `NTRUSolve`), since no C implementation exists.

try to recover $ff^* + gg^*$ — note that this field element carries all the length information of the vector (f, g) — from the a good approximation of $\tilde{f}\tilde{f}^* + \tilde{g}\tilde{g}^*$.

In the full version [ENS+23b], we have analyzed such an attack in details and showed its irrelevance for power-of-two cyclotomic rings. More precisely, the cost of this attack far exceeded the cost of direct a lattice reduction approach to recover (f, g). An intuitive but informal argument for why this happens is as follows: while the location of $(|\tilde{f}|, |\tilde{g}|)$ can be known quite accurately by the attacker, the rounding step acts as adding noise to each components. Since there are many embeddings, it remains computationally expensive to recover $(|f|, |g|)$ from that knowledge only.

It follows from this analysis that the only relevant attack in our setting is the one reducing to an approximate SVP instance in the module lattice. This can then be used to derive bit security estimates accoring to the standard Core-SVP methodology. These estimates form the basis of the parameter selection provided in the next section.

5 ANTRAG implementation and further optimizations

5.1 Basic implementation

An initial implementation of the ANTRAG trapdoor generation algorithm as well as the resulting complete signature scheme in portable C is provided alongside our original paper, and available on GitHub [Tib23].

It builds upon the source codes of FALCON and MITAKA. Signing and verification are almost identical to those of MITAKA, since the lattice Gaussian sampler and the verification equation are the same. Key generation consists of ANTRAG's algorithms to generate the first basis vector (f, g), along with code to solve the NTRU equation to find (F, G), for which we basically reuse FALCON's `NTRUSolve` [PP19]. The Fast Fourier transform and the resulting code for ring arithmetic are similarly borrowed from FALCON.

We note that, since the C code of MITAKA itself did not include a key generation algorithm (only precomputed fixed keys obtained using separate Python scripts), that implementation constitutes, to the best of our knowledge, the first full C implementation of a hybrid sampler-based signature.

In view of the simplicity of our trapdoor generation, the code is fairly straightforward. In particular, since the floating point uniform distributions we generate for the absolute values of the embeddings are bounded away from zero, there is no subtlety related to precision loss for values close to zero (this is unlike the Box–Muller algorithm used in signing, for which we reuse MITAKA's code that behaves properly in that respect). The only trick worth mentioning is the use of the Conway–Sloane D_n-lattice decoder in the generation of (f, g), to ensure that the algebraic norms of f and g are odd (a condition imposed by FALCON's `NTRUSolve`).

A performance comparison with FALCON and MITAKA is provided in Table 2, using the same modulus $q = 12289$ for consistency. Compilation is carried out with `gcc` 13.2.1 with `-O3 -march=native` optimizations enabled. Timings are collected on a single core of an AMD Ryzen 7 PRO 6860Z @ 2.7 GHz laptop with hyperthreading and frequency scaling disabled. Cycle counts are not provided for Falcon, since the Falcon benchmarking tool only measures clock time.

As noted previously, the Mitaka C implementation does not include a key generation procedure. For reference, we provide the timings for the `numpy`-based SageMath implementation of the Mitaka key generation procedure instead, *not* including the cost of `NTRUSolve`, so that only the highly optimized `GoodPair` code is accounted for. As expected from the fact that Mitaka needs to explore a search space of millions of key candidates, the timings are orders of magnitude worse than Falcon and Antrag.

The running time of our key generation is close to that of Falcon. Signing speeds are basically identical to Mitaka since we mostly reuse that code (up to very minor optimizations). Verification is consistent across all three schemes.

5.2 Possible extensions and optimizations

Obtaining more compact signatures. Several techniques to improve the key and signature sizes of lattice-based hash-and-sign signatures are introduced in [ETWY22]. In principle, they can all be applied to the Antrag signature scheme. One of these techniques is a fine-tuned encoding approach for discrete Gaussian vectors, and is oblivious to the actual structure of the secret keys; we can thus consider it done by default when estimating the bit size of signatures. The two other techniques are choosing a smaller modulus q than the popular choice $q = 12289$, and elliptical sampling. They have more impact on the key generation step, and although they were shown somewhat equivalent when applied to schemes such as Falcon and Mitaka, the situation is different for Antrag.

We first discuss smaller moduli. From our analysis in Section 3 and Section 4, the annulus where candidate pairs are sampled becomes relatively smaller as q decreases, which noticeably impacts the success probability of Algorithm 5. To keep a small rejection rate in practice, we are led to decrease the quality of the key pairs, or in other words, to use a larger parameter α. Fortunately, it was pointed out in [ETWY22] that there is a range for such smaller q where, in fixed dimension, the key recovery becomes harder. This actually means that reducing q and increasing α does not necessarily translate to a substantially lower security level. We note however that q cannot be chosen arbitrarily small, as attacks exist for very small q [DEP23].

The situation for elliptical sampling is less attractive for the following reason. Candidates should now be sampled in well-chosen elliptic annuli rather than circular ones. We can easily sample continuously uniformly in such annuli, but when carrying out the decoding back to the ring (e.g., by coefficient-wise rounding), we still incur an error term on embeddings that behaves like an isotropic normal distribution of standard deviation $\Omega(\sqrt{q})$. After the addition of the error term, embeddings sampled more towards the direction of the major axis of the ellipse are more likely than in a spherical case to end up in the target elliptical annulus, but embeddings sampled in the direction of the minor axis have much lower probability of success, and this has a much greater effect on overall success probability, constraining the choice of the quality parameter α. In the end, we find that rather than using elliptical sampling in our setting with a certain skewing factor γ, it is essentially just as effective to reduce the modulus q by the same factor γ instead (which additionally has the advantage of reducing public key size). As a result, we omit the detailed analysis of this less attractive approach.

Parameter selection versatility. One of the advantages of Prest's hybrid sampler compared to the Falcon sampler, as already pointed out in the Mitaka paper, is the ease with which it can be used over base rings different from power-of-two cyclotomics, such as cyclotomics with 3-smooth conductor. Such rings can in turn be used to reach essentially arbitrary security levels (whereas the current incarnation of Falcon only targets either NIST–I or NIST–V).

This advantage naturally carries over to Antrag, as discussed extensively in the full version of that paper [ENS+23b]; it does so, moreover, without Mitaka's substantial loss in terms of sampling quality. This makes it easy to target all NIST security levels with good performance.

We present our parameter selection in Table 3 for power-of-two cyclotomics, and Table 4 for the 3-smooth case. For all parameter sets, we set the quality α with two decimal places in such a way as to reach a repetition rate M of around 3 to 4. For the moduli, we give both the choices of q found in the literature as well as smaller candidates that also have close to optimal splitting in the ambient ring, should one wish to rely on NTT multiplication to slightly speed up verification.

Table 3. Practical parameter selection, power-of-two case

	$q = 12289$		$q = 3329$	
d	512	1024	512	1024
Quality α	1.15	1.23	1.23	1.48
Repetition rate M	3	4	4	4
Bit security (C/Q)	124/113	264/240	121/110	265/240
Verification key size (bytes)	896	1792	768	1536
Signature size (bytes)	646	1260	591	1176

Table 4. Practical parameter selection for Antrag, 3-smooth conductor case.

(a) Modulus $q = 12289$

d	648	768	864	972
Quality α	1.17	1.19	1.21	1.22
Repetition rate M	4	3	3	4
Bit security (C/Q)	166/151	196/178	222/201	251/227
Verification key size (bytes)	1134	1344	1512	1701
Signature size (bytes)	808	952	1069	1200

(b) Various moduli. For $d = 768, 864, 972$, the right column shows moduli of [EFG+22].

	$d = 648$		$d = 768$		$d = 864$		$d = 972$	
Modulus q	3889	9721	3329	18433	3727	10369	4373	17497
Quality α	1.32	1.19	1.39	1.16	1.40	1.23	1.40	1.18
Expected repetitions	4	4	4	3	4	3	4	4
Bit security (C/Q)	159/144	164/149	192/174	195/177	220/200	222/201	254/230	250/227
Verification key size (bytes)	972	1134	1152	1440	1296	1512	1580	1823
Signature size (bytes)	747	796	883	977	1000	1058	1133	1225

Implementation of these optimizations and extensions. We recently extended [GT23] the original C implementation of Antrag discussed in the previous section in order to support the various optimizations and extensions discussed above, included smaller moduli q as well as arbitrary 3-smooth cyclotomic base rings.

This involved modifying the FFT, NTT, and `NTRUSolve` implementations, which all recurse on the dimension. Ternary recursion steps were introduced to eliminate the factors of 3 from the dimension, before proceeding with the usual binary recursion.

Moreover, supporting suitable smaller moduli (such as $q = 3583$ in dimension 512) involved adding support for $\mathbb{F}_{q^2}$ arithmetic in the NTT (similar to the Kyber KEM), allowing for the existence of an m-th root of unity for lower values of q. These smaller moduli come with shorter keys and signatures, as shown in Tables 3 and 4.

The original Falcon-based code included contrived, parameter-specific optimizations, which rendered it difficult to modify directly. This led to a complete rewrite of the three algorithms mentioned above from first principles, accompanied by new unit tests to ensure their correctness at every step.

To replace the hard-coded constants and constant tables used in the initial implementation of these algorithms, a SageMath script was written, which runs at build time, in order to recompute these values from the selected parameters dynamically, and store them in separate header files, which are then linked with the source code. This should simplfy any further expansions of the parameter space.

References

AT01. A. Annamalai and C. Tellambura. Cauchy–Schwarz bound on the generalized Marcum-Q function with applications. *Wireless Communications and Mobile Computing*, 1(2):243–253, 2001.

CPS+20. Chitchanok Chuengsatiansup, Thomas Prest, Damien Stehlé, Alexandre Wallet, and Keita Xagawa. ModFalcon: Compact signatures based on module-NTRU lattices. In Hung-Min Sun, Shiuh-Pyng Shieh, Guofei Gu, and Giuseppe Ateniese, editors, *ASIACCS 20*, pages 853–866. ACM Press, October 2020.

DEP23. Léo Ducas, Thomas Espitau, and Eamonn W. Postlethwaite. Finding short integer solutions when the modulus is small. In Helena Handschuh and Anna Lysyanskaya, editors, *CRYPTO 2023, Part III*, volume 14083 of *LNCS*, pages 150–176. Springer, Heidelberg, August 2023.

DLP14. Léo Ducas, Vadim Lyubashevsky, and Thomas Prest. Efficient identity-based encryption over NTRU lattices. In Palash Sarkar and Tetsu Iwata, editors, *ASIACRYPT 2014, Part II*, volume 8874 of *LNCS*, pages 22–41. Springer, Heidelberg, December 2014.

DN12. Léo Ducas and Phong Q. Nguyen. Learning a zonotope and more: Cryptanalysis of NTRUSign countermeasures. In Xiaoyun Wang and Kazue Sako, editors, *ASIACRYPT 2012*, volume 7658 of *LNCS*, pages 433–450. Springer, Heidelberg, December 2012.

DP15. Léo Ducas and Thomas Prest. A hybrid Gaussian sampler for lattices over rings. Cryptology ePrint Archive, Report 2015/660, 2015. https://eprint.iacr.org/2015/660.

DP16. Léo Ducas and Thomas Prest. Fast Fourier orthogonalization. In Sergei A. Abramov, Eugene V. Zima, and Xiao-Shan Gao, editors, *ISSAC 2016*, pages 191–198. ACM, 2016.

EFG+21. Thomas Espitau, Pierre-Alain Fouque, François Gérard, Mélissa Rossi, Akira Takahashi, Mehdi Tibouchi, Alexandre Wallet, and Yang Yu. Mitaka: a simpler, parallelizable, maskable variant of falcon. Cryptology ePrint Archive, Report 2021/1486, 2021. https://eprint.iacr.org/2021/1486.

EFG+22. Thomas Espitau, Pierre-Alain Fouque, François Gérard, Mélissa Rossi, Akira Takahashi, Mehdi Tibouchi, Alexandre Wallet, and Yang Yu. Mitaka: A simpler, parallelizable, maskable variant of falcon. In Orr Dunkelman and Stefan Dziembowski, editors, *EUROCRYPT 2022, Part III*, volume 13277 of *LNCS*, pages 222–253. Springer, Heidelberg, May / June 2022.

ENS+23a. Thomas Espitau, Thi Thu Quyen Nguyen, Chao Sun, Mehdi Tibouchi, and Alexandre Wallet. Antrag: Annular NTRU trapdoor generation. In Jian Guo and Ron Steinfeld, editors, *ASIACRYPT 2023*, volume 14444 of *LNCS*, pages 3–36. Springer, 2023.

ENS+23b. Thomas Espitau, Thi Thu Quyen Nguyen, Chao Sun, Mehdi Tibouchi, and Alexandre Wallet. Antrag: Annular NTRU trapdoor generation. Cryptology ePrint Archive, Paper 2023/1335, 2023. https://eprint.iacr.org/2023/1335.

ETWY22. Thomas Espitau, Mehdi Tibouchi, Alexandre Wallet, and Yang Yu. Shorter hash-and-sign lattice-based signatures. In Yevgeniy Dodis and Thomas Shrimpton, editors, *CRYPTO 2022, Part II*, volume 13508 of *LNCS*, pages 245–275. Springer, Heidelberg, August 2022.

GGH97. Oded Goldreich, Shafi Goldwasser, and Shai Halevi. Public-key cryptosystems from lattice reduction problems. In Burton S. Kaliski Jr., editor, *CRYPTO'97*, volume 1294 of *LNCS*, pages 112–131. Springer, Heidelberg, August 1997.

GPV08. Craig Gentry, Chris Peikert, and Vinod Vaikuntanathan. Trapdoors for hard lattices and new cryptographic constructions. In Richard E. Ladner and Cynthia Dwork, editors, *40th ACM STOC*, pages 197–206. ACM Press, May 2008.

GS02. Craig Gentry and Michael Szydlo. Cryptanalysis of the revised NTRU signature scheme. In Lars R. Knudsen, editor, *EUROCRYPT 2002*, volume 2332 of *LNCS*, pages 299–320. Springer, Heidelberg, April / May 2002.

GT23. Jade Guiton and Mehdi Tibouchi. Companion implementation of this paper. GitHub repository `mti/antrag_opt`, 2023. https://github.com/mti/antrag_opt.

HHP+03. Jeffrey Hoffstein, Nick Howgrave-Graham, Jill Pipher, Joseph H. Silverman, and William Whyte. NTRUSIGN: Digital signatures using the NTRU lattice. In Marc Joye, editor, *CT-RSA 2003*, volume 2612 of *LNCS*, pages 122–140. Springer, Heidelberg, April 2003.

LDK+22. Vadim Lyubashevsky, Léo Ducas, Eike Kiltz, Tancrède Lepoint, Peter Schwabe, Gregor Seiler, Damien Stehlé, and Shi Bai. CRYSTALS-DILITHIUM. Technical report, National Institute of Standards and Technology, 2022. available at https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022.

Lyu09. Vadim Lyubashevsky. Fiat-Shamir with aborts: Applications to lattice and factoring-based signatures. In Mitsuru Matsui, editor, *ASIACRYPT 2009*, volume 5912 of *LNCS*, pages 598–616. Springer, Heidelberg, December 2009.

Lyu12. Vadim Lyubashevsky. Lattice signatures without trapdoors. In David Pointcheval and Thomas Johansson, editors, *EUROCRYPT 2012*, volume 7237 of *LNCS*, pages 738–755. Springer, Heidelberg, April 2012.

NR06. Phong Q. Nguyen and Oded Regev. Learning a parallelepiped: Cryptanalysis of GGH and NTRU signatures. In Serge Vaudenay, editor, *EUROCRYPT 2006*, volume 4004 of *LNCS*, pages 271–288. Springer, Heidelberg, May / June 2006.

Pei10. Chris Peikert. An efficient and parallel Gaussian sampler for lattices. In Tal Rabin, editor, *CRYPTO 2010*, volume 6223 of *LNCS*, pages 80–97. Springer, Heidelberg, August 2010.

PFH+17. Thomas Prest, Pierre-Alain Fouque, Jeffrey Hoffstein, Paul Kirchner, Vadim Lyubashevsky, Thomas Pornin, Thomas Ricosset, Gregor Seiler, William Whyte, and Zhenfei Zhang. FALCON. Technical report, National Institute of Standards and Technology, 2017. available at https://csrc.nist.gov/projects/post-quantum-cryptography/post-quantum-cryptography-standardization/round-1-submissions.

PFH+22. Thomas Prest, Pierre-Alain Fouque, Jeffrey Hoffstein, Paul Kirchner, Vadim Lyubashevsky, Thomas Pornin, Thomas Ricosset, Gregor Seiler, William Whyte, and Zhenfei Zhang. FALCON. Technical report, National Institute of Standards and Technology, 2022. available at https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022.

PP19. Thomas Pornin and Thomas Prest. More efficient algorithms for the NTRU key generation using the field norm. In Dongdai Lin and Kazue Sako, editors, *PKC 2019, Part II*, volume 11443 of *LNCS*, pages 504–533. Springer, Heidelberg, April 2019.

Pre15. Thomas Prest. *Gaussian Sampling in Lattice-Based Cryptography*. PhD thesis, École Normale Supérieure, Paris, France, 2015.

SA00. M.K. Simon and M.-S. Alouini. Exponential-type bounds on the generalized Marcum Q-function with application to error probability analysis over fading channels. *IEEE Trans. Commun.*, 48(3):359–366, 2000.

Tib23. Mehdi Tibouchi. Original antrag implementation. GitHub repository `mti/antrag`, 2023. https://github.com/mti/antrag.