ANTRAG: Annular NTRU Trapdoor Generation
Making MITAKA As Secure As FALCON

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

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

Abstract. In this paper, we introduce a novel trapdoor generation technique for Prest's hybrid sampler over NTRU lattices. Prest's sampler is used in particular in the recently proposed MITAKA signature scheme (Eurocrypt 2022), a variant of the FALCON signature scheme, one of the candidates selected by NIST for standardization. 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 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 by 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 the drawbacks.

Keywords: Post-quantum cryptography · Hash-and-sign lattice-based signatures · NTRU trapdoors · Discrete Gaussian sampling

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.[6]

[6] 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.

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 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 of this paper

In this paper, 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[7] 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:

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

[7] 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.

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

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

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 $\alpha_{\text{Falcon}} = 1.17$ 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 $\alpha_{\text{Mitaka}} = 2.04$ 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.

Our 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 this paper, 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 of vectors that satisfy the desired quality level. We propose and analyze this approach specifically for the hybrid sampler.[8]

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).

[8] 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.

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.

Our contributions. The main contribution of this paper 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, as supplementary material, a full portable C implementation of the corresponding signature scheme 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 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 $\mathbb{R}^2$: $A(r, R) := \big\{(x, y) \in \mathbb{R}^2 \mid r^2 \leqslant x^2 + y^2 \leqslant R^2\big\}$. We also denote by $A^+(r, R)$ the arc of annulus in the upper-right quadrant of the plane, i.e., $A^+(r, R) := \big\{(x, y) \in A(r, R) \mid x, y \geqslant 0\big\}$.

When f is a real-valued function over a countable set S, we write $f(S) = \sum_{s \in S} f(s)$ assuming that this sum is absolutely convergent. We note $\lfloor \cdot \rceil$ the rounding of a real number to

Table 1. Comparison with Falcon and Mitaka for the same dimensions 512 and 1024 and the same modulus $q = 12289$ (excerpt from Table 4).

	Falcon [PFH+22]		Mitaka [EFG+22]		This paper	
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

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 Cyclotomic fields

Let m be a positive integer, and $d = \phi(m)$ be the degree of the m-th cyclotomic polynomial Φ_m (ϕ is the Euler totient function). 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 Φ_m, and its ring of algebraic integers is $\mathscr{R} := \mathbb{Z}[\zeta]$. The field automorphism induced by $\zeta \mapsto \zeta^{-1} = \bar{\zeta}$ corresponds to the complex conjugation, and we write f^* the image of $f \in \mathscr{K}$ under this automorphism. We have $\mathscr{K} \simeq \mathbb{Q}[x]/(\Phi_m(x))$ and $\mathscr{R} \simeq \mathbb{Z}[x]/(\Phi_m(x))$, and both are contained in $\mathscr{K}_\mathbb{R} := \mathscr{K} \otimes \mathbb{R} = \mathbb{R}[x]/(\Phi_m(x))$. 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 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 $\gcd(k, m) = 1$. 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 $\mathscr{K}_\mathbb{R}$. We let $\mathscr{K}_\mathbb{R}^{++}$ be the subset of $\mathscr{K}_\mathbb{R}^+$ which have all positive coordinates in the canonical embedding. We have a partial ordering over $\mathscr{K}_\mathbb{R}^+$ by $f \succ g$ if and only if $f - g \in \mathscr{K}_\mathbb{R}^{++}$. The algebra $\mathscr{K}_\mathbb{R}$ is also equipped with a norm $N_\mathscr{K}(x) = \prod_i \varphi(x)$, which extends the standard field norm.

The next technical lemma is useful in our analyses, and is obtained by elementary trigonometric identities.

Lemma 1. *Let $\zeta = \exp(i\theta)$ with $\theta = \frac{2k\pi}{m}$ and $gcd(k, m) = 1$ be a m-th primitive root of the unity, and $d = \phi(m)$. Let $S(\theta) = \sum_{j=0}^{d-1} \zeta^{2j}$. We have $S(\theta) = \frac{\sin(\theta d)}{\sin\theta} e^{i\theta(d-1)}$ and*

$$\mathrm{Re}\, S(\theta) = \frac{1}{2} + \frac{\sin((2d-1)\theta)}{2\sin\theta} \quad \textit{and} \quad \mathrm{Im}\, S(\theta) = \frac{\sin(d\theta)\sin((d-1)\theta)}{\sin\theta}.$$

Remark 1. If m is a power of 2 then $2d = m$ so we always have $S(\theta) = 0$.

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:*

$$Q_m(u,v) \geqslant 1 - \frac{1}{2} e^{-(u-v)^2/2} \qquad \text{if } u \geqslant v;$$

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

We also note that the independent sum of a $\chi^2(k, \sigma^2; c)$ variable and a $\chi^2(k', \sigma^2; c')$ variable, for the same scaling σ^2, follows a $\chi^2(k + k', \sigma^2; c + c')$ distribution.

In the general case where $T \sim \mathcal{N}(\mu, \Sigma)$, let $\lambda_i > 0$ be the eigenvalues of the positive definite symmetric matrix Σ. If P is an orthogonal matrix that diagonalizes Σ, let $\nu = (\nu_1, \ldots, \nu_k) := P\mu$. Then $\|T\|^2 \sim \chi^2(1, \lambda_1; \nu_1^2) + \cdots + \chi^2(1, \lambda_k; \nu_k^2)$. This distribution is called the *weighted sum of k independent non central chi-squared variables*. There is no known closed form for its cumulative distribution function, but there exist tools to evaluate it numerically (e.g., the Python package `chi2comb`).

Algorithm 1: Signing

Input: A message $\mathtt{m}$, 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 $\mathtt{m}$, 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.

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.

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 $(f, g) \in \mathscr{R}^2$ such that f and g are coprime in $\mathscr{R}$, and with the guarantee that the basis $\mathbf{B}_{f,g}$ output by `NTRUSolve` will have quality α 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

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 $(h, \mathbf{B}_{f,g})$.

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 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 this paper 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.

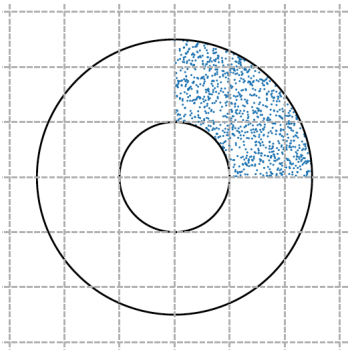


Fig. 1. $(|z|, |w|)$ is sampled uniformly in the annulus $A^+(r, R)$.

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 $A^+(r, R)$

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

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.

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 2. 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

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)

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 allowing us to quantify the claim that sampling $\tilde{f}$ and $\tilde{g}$ in this $A^+(r, R)$ increases success probability. To that end, write $e = (e_f, e_g) = \big(f - \tilde{f}, g - \tilde{g}\big) \in \mathscr{K}^2_{\mathbb{R}}$ 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 $\big(|\varphi_i(f)|, |\varphi_i(g)|\big)$ will be satisfied for all i.

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 $[-1/2, 1/2)$.[9] This is well-supported by experiments (see Fig. 4(a) in Appendix C).

[9] 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).

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:

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

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[10] 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 expressed as follows:

$$\Sigma_\theta = \frac{d}{24}\mathbf{I}_2 + E(\theta) \quad \text{where} \quad E(\theta) = \frac{1}{24}\begin{bmatrix} \operatorname{Re} S(\theta) & \operatorname{Im} S(\theta) \\ \operatorname{Im} S(\theta) & -\operatorname{Re} S(\theta) \end{bmatrix}.$$

Note that Σ_θ has eigenvalues $\lambda_+^\theta = \frac{d+|S(\theta)|}{24}$ and $\lambda_-^\theta = \frac{d-|S(\theta)|}{24}$. We thus expect that $\varphi_\theta(e_f)$ follows the normal distribution $\mathcal{N}(0, \Sigma_\theta)$, 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: see Fig. 4(b)), therefore, we can write

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

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

Heuristic 1. Let $(f, g) \in \mathcal{K}^2$ a pair output by Algorithm 5, corresponding to $(\tilde{f}, \tilde{g}) \in \mathcal{K}_\mathbb{R}^2$ 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

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

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

Under this heuristic we can express the expected length of the embeddings of secret keys and related elements, which will be useful in the security analysis. The proof is provided in Appendix A.

Proposition 1 (Heuristic). *Keeping the notation of Algorithm 5, let (f, g) be a random variable following the distribution of its output. Let θ be an argument of a primitive m-th root of unity, and let φ_θ be the corresponding embedding. Then:*

$$\mathbb{E}\big[|\varphi_\theta(f)|^2 + |\varphi_\theta(g)|^2\big] = \frac{d}{6} + \frac{r^2 + R^2}{2}.$$

Let $\|\cdot\|_\theta$ be the norm induced by the quadratic form Σ_θ. Then we also have:

$$\mathbb{E}\big[|\varphi_\theta(f)|^4 + |\varphi_\theta(g)|^4\big] = \frac{5}{8}(R^4 + r^4) + R^2 r^2 + \frac{d}{12}(R^2 + r^2) + \frac{d^2}{36} + T(\theta),$$

where $T(\theta) \coloneqq |S(\theta)|^2/72 + 4 \cdot \mathbb{E}\big[\|\varphi_\theta(\tilde{f})\|_\theta^2 + \|\varphi_\theta(\tilde{g})\|_\theta^2\big]$.

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

4 Success probability and security analysis

In this section, we first 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.

At the end of the section, we also briefly describe how the analysis extends to the more general setting of cyclotomic rings with conductor $m = 2^k p^\ell$, with further details provided as supplementary material.

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 $\frac{d}{24}\mathbf{I}_4$ (since over power-of-two cyclotomic fields, $E(\theta) = 0$ for all θ). 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. According to Lemma 4, the following bounds on the Marcum Q function hold for any $1/\alpha \leqslant \beta \leqslant \alpha$:

$$Q_2(\tau\beta, \tau/\alpha) \geqslant 1 - \frac{1}{2}\exp\big(-\frac{\tau^2}{2}(\beta - 1/\alpha)^2\big)$$
$$Q_2(\tau\beta, \tau\alpha) \leqslant \Big(1 + \frac{\alpha/\beta}{\pi}\Big)\exp\big(-\frac{\tau^2}{2}(\alpha - \beta)^2\big)$$

from which it follows that:

$$p_{\text{succ}}(\beta) \geqslant 1 - \frac{1}{2}u_\tau(\beta - 1/\alpha) - \big(1 + \frac{\alpha/\beta}{\pi}\big)u_\tau(\alpha - \beta) \quad \text{where} \quad u_\tau(x) = \exp\big(-\frac{\tau^2}{2}x^2\big). \tag{6}$$

We write $\beta = (\alpha + 1/\alpha)/2 + t(\alpha - 1/\alpha)/2$ for some $t \in (-1, 1)$. Recall furthermore from Eq. (3) that we have set:

$$\frac{r}{\sqrt{q}} = (\alpha + 1/\alpha)/2 - \xi(\alpha - 1/\alpha)/2 \quad \text{and} \quad \frac{R}{\sqrt{q}} = (\alpha + 1/\alpha)/2 + \xi(\alpha - 1/\alpha)/2$$

so that t actually varies in $[-\xi, \xi]$. In particular, we have:

$$\frac{\alpha}{\beta} \leqslant \frac{\alpha}{r} = \frac{\alpha}{\frac{1-\xi}{2}\alpha + \frac{1+\xi}{2}\frac{1}{\alpha}} = \frac{2}{1-\xi} \cdot \frac{1}{1 + \frac{1+\xi}{1-\xi}\frac{1}{\alpha^2}} \leqslant \frac{2}{1-\xi}.$$

Thus, inequality (6) becomes:

$$p_{\text{succ}}(\beta) \geqslant 1 - \frac{1}{2}u_\tau\big((1-t)\delta\big) - \big(1 + \frac{2/\pi}{1-\xi}\big)u_\tau\big((1+t)\delta\big)\big) \quad \text{for} \quad \delta = \frac{\alpha - 1/\alpha}{2}.$$

Since u_τ is a decreasing function, both $u_\tau\big((1-t)\delta\big)$ and $u_\tau\big((1+t)\delta\big)$ are bounded above by $u_\tau\big((1-\xi)\delta\big)$, so that:

$$p_{\text{succ}}(\beta) \geqslant 1 - K_\xi u_\tau\big((1-\xi)\delta\big) \quad \text{with} \quad K_\xi = \frac{3}{2} + \frac{2/\pi}{1-\xi}$$

holds for all $\beta \in [r/\sqrt{q}, R/\sqrt{q}]$.

As a result, 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 $A^+(r, R)$) is similarly lower bounded as:

$$p_{\text{succ-one}} \geqslant 1 - K_\xi u_\tau\big((1-\xi)\delta\big) \tag{7}$$

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:

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

Using the usual first order approximation $\log(1-x) \approx -x$, this yields $\frac{d}{2}K_\xi u_\tau\big((1-\xi)\delta\big) \lesssim \log M$, or equivalently:

$$\frac{\alpha - 1/\alpha}{2} \gtrsim \frac{d}{12(1-\xi)^2 q}\log\frac{K_\xi d}{2\log M}.$$

This shows that a quality α is achievable (with repetition rate up to M) as long as:

$$\alpha \geqslant \sqrt{A} + \sqrt{1+A} \quad \text{where} \quad A = \frac{d}{12(1-\xi)^2 q}\log\frac{K_\xi d}{2\log M}. \tag{8}$$

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 (8) 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 $\left(|\varphi_\theta(\tilde{f})|, |\varphi_\theta(\tilde{g})|\right)$ 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 $[r^2/q, R^2/q]$. Therefore:

$$p_{\text{succ-one}} = \frac{q}{R^2 - r^2} \int_{r^2/q}^{R^2/q} p_{\text{succ}}(\sqrt{B})\, dB = \frac{2q}{R^2 - r^2} \int_{r/\sqrt{q}}^{R/\sqrt{q}} p_{\text{succ}}(\beta)\, \beta\, d\beta.$$

Carrying out the change of variables $\beta = (\alpha + 1/\alpha)/2 + t(\alpha - 1/\alpha)/2$ and plugging in the expression of $p_{\text{succ}}(\beta)$ in terms of Q_2, we finally get:

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

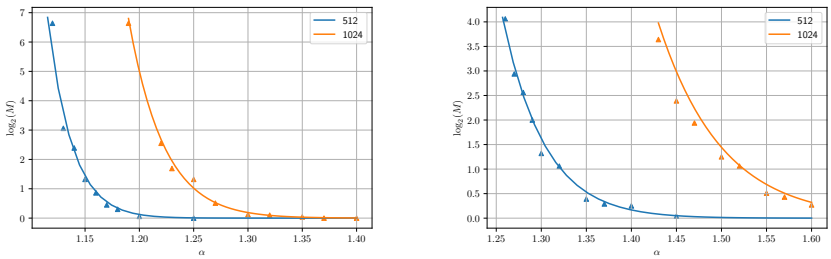
where

$$F(\alpha, t) = Q_2\left(\tau\left(\frac{\alpha + \frac{1}{\alpha}}{2} + t\frac{\alpha - \frac{1}{\alpha}}{2}\right), \tau/\alpha\right) - Q_2\left(\tau\left(\frac{\alpha + \frac{1}{\alpha}}{2} + t\frac{\alpha - \frac{1}{\alpha}}{2}\right), \tau\alpha\right),$$

and $1/M = p_{\text{succ-all}} = p_{\text{succ-one}}^{d/2}$. 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 for power-of-two cyclotomics

In order to assess the concrete security of the resulting signature scheme, we proceed using the usual cryptanalytic methodology of 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



(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).

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 our focus will now be the resilience to key recovery attacks.

In Section 4.2.1, we go through a short review of the general lattice reduction approach for key recovery, which is the current best attack when no additional information is provided to the attacker (seeing 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.

In Section 4.2.2 we present a new approach exploiting this additional geometric information. It is reminiscent of the subfield attacks [ABD16, CJL16], however here we stop the descent in the subfields at the totally real subfield $\mathscr{K}^+$ (the set of elements satisfying $f = f^*$). Indeed, this subfield encodes the length information of the pair (f, g) in the elements (ff^*, gg^*) and its collection of embeddings. In the extreme (unlikely) case where the annulus would be a circle, an adversary would know this element exactly, and could use the Gentry-Szydlo attack [GS02] to recover f or g. Our situation could be summed-up as an "approximate" Gentry-Szydlo attack, where too much proximity of all the embeddings to a known circle could be exploited by an attacker through lattice reduction.

Our trapdoor generator could output keys with embeddings that would all be close to some circle, and we call these temporarily *potentially weak keys*. Our analysis will show that these potentially weak keys are in fact not so weak, or in other words, that we have some freedom for parametrization with respect to the available space $(\alpha - \alpha^{-1})\sqrt{q}$. This ensures a good success rate for Antrag. Ultimately, the attack will use lattice reduction but on a different lattice than in the direct, standard key-recovery context, and will try to recover (ff^*, gg^*).

For the context of Section 4.2.2, we need the expected length of (f, g) and (ff^*, gg^*). These two properties are gathered in the next result. The proof, a direct application of Proposition 1, is provided in Appendix A.

Corollary 1. *With the notation of Algorithm 5, let (f, g) be a random variable following the distribution of its output. Then we have $\mathbb{E}[\|(f, g)\|^2] = \frac{d}{6} + \frac{R^2+r^2}{2}$ and $\mathbb{E}[\|(ff^*, gg^*)\|^2] = \frac{5}{8}(R^4 + r^4) + R^2r^2 + \frac{d}{6}(R^2 + r^2) + \frac{d^2}{36}$.*

4.2.1 Classical attack against NTRU keys. The key recovery in this context consists in constructing the algebraic lattice over $\mathscr{R}$ spanned by the vectors $(0, q)$ and $(1, h)$ (i.e. the public basis attached to the NTRU key) and retrieving the lattice vector $\mathbf{s} = (f, g)$ among all possible lattice vectors of norm bounded by $\|s\|$ (or a functionally equivalent vector, for instance $(\mu \cdot f, \mu \cdot g)$ for any unit μ of the ring of integer of the number field). From Corollary 1 we obtain $\mathbb{E}[\|\mathbf{s}\|^2] \leqslant qA$, where $A = \frac{d}{6q} + \frac{1}{9}\left(\frac{5\alpha^2}{2} + \frac{5}{\alpha^2} + 4\right)$. Since the attack is easier when the key to recover is longer, we take the value qA acting as $\mathbb{E}[\|\mathbf{s}\|^2]$. In order to avoid enumerating and testing all integer vectors in the sphere of radius $\sqrt{q}S$, which would contain a large number of vectors under the Gaussian heuristic[11], namely around $\left(\frac{qA}{q}\right)^d = A^d$, we make use of the projection trick (see also [EFG+22, ETWY22]). This technique involves reducing the public basis with some lattice reduction algorithm,

[11] The Gaussian heuristic predicts the number of vectors of length at most ℓ in a random lattice Γ of volume V to be a $v_\Gamma(\ell)/V + o(1)$ for large enough ℓ, where $v_\Gamma(\ell)$ is the volume of the sphere of radius ℓ for the measure induced by the inner product on Γ.

and seeking for the projection of the secret key onto the lattice spanned by the few last Gram-Schmidt vectors of this reduced basis. If we find the projection of the secret key, we can retrieve the full key by using the Babai nearest plane algorithm to lift it to a lattice vector of the desired norm.

More precisely we proceed as follows. Set β to be the block size parameter of the DBKZ algorithm [MW16] and start by reducing the public basis with this latter algorithm. Call $[\mathbf{b}_1, \ldots, \mathbf{b}_{2d}]$ the resulting vectors. Then if we can recover the *projection* of the secret key onto $\mathcal{P}$, the orthogonal space to $\mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_{2d-\beta-1})$, then we can retrieve in polynomial time the full key by *Babai nearest plane* algorithm to lift it to a lattice vector of the desired norm. Hence it suffices to be able find the projection of the secret key among the shortest vector of the lattice generated by the last β vectors projected onto $\mathcal{P}$. Classically, sieving on this projected lattice will recover all vectors of norm smaller than $\sqrt{\frac{4}{3}}\ell$, where ℓ is the norm of the $2d-\beta$-th Gram-Schmidt vector $\widetilde{\mathbf{b}}_{2d-\beta}$ of the reduced basis.

The expected length of the projection is usually estimated under the *Geometric Series Assumption* (GSA). Instantiated on NTRU lattices, it states that the Gram-Schmidt vectors of the basis outputted by DBKZ with block-size β satisfy the relations (see Cor 2. of [MW16]):

$$\|\widetilde{\mathbf{b}}_i\| = \delta_\beta^{2(d-i)+1}\sqrt{q} \quad \text{where} \quad \delta_\beta = \left(\frac{(\pi\beta)^{1/\beta}\cdot\beta}{2\pi e}\right)^{\frac{1}{2(\beta-1)}}.$$

Therefore, we expect that $\ell = \delta_\beta^{-2(d-\beta)+1}\sqrt{q} \approx \sqrt{q}\cdot\left(\frac{\beta}{2\pi e}\right)^{1-\frac{d}{\beta-1}}$. Moreover, assuming that $\mathbf{s}$ behaves as a random vector, and using the GSA to bound the norm of the Gram-Schmidt vectors $[\widetilde{\mathbf{b}}_1, \ldots, \widetilde{\mathbf{b}}_{2d-\beta}]$, the (squared) norm of its projection over $\mathcal{P}$ concentrates around $\frac{\beta}{2d}\cdot\mathbb{E}\left[\|\mathbf{s}\|^2\right] = \frac{Aq\beta}{2d}$. Hence, we will retrieve the projection among the sieved vectors if $\frac{Aq\beta}{2d} \leqslant \frac{4}{3}\ell^2$, that is if the following condition is fulfilled:

$$A \leqslant \frac{8d}{3\beta}\delta_\beta^{4(\beta-d)+2}. \tag{9}$$

Remark 3. (*On the use of GSA*) In order to make a more accurate assessment of potential attacks, numerical models of the profile of the Gram-Schmidt length derived from simulations of the behavior of (D)BKZ can be utilized instead of relying solely on the Gaussian heuristic approximation (GSA). While this section focuses on using the GSA for the purpose of simplifying the formulae and presenting the information in a clear manner, it is important to note that predictive models that generate a "Z-shaped" profile are employed in the estimation scripts.

(*On the size of the enumeration window*) In the previous description we only considered the space $\mathcal{P}$, orthogonal to span $(b_1, \ldots, b_{2d-\beta-1})$. It is natural to want to extend its dimension, and choose the optimal one. It appears that for the specific parameters of our work, this optimization would only result in a difference of less than a single bit of security. Besides, on the one hand, by using the exact block size beta we can extract the vectors we need to sieve for free from the preliminary run of DBKZ, avoiding the need for an additional sieving pass. On the other hand, using a larger dimension for the additional sieving pass adds a non-negligible cost. Note that this is a consequence of the Core-SVP methodology, which we discuss in more length in Section 4.3 which ignores the polynomial overhead cost of (D)BKZ.

4.2.2 Towards a subfield attack. Given the knowledge of the relative norm $M = ff^* + gg^*$, the structure of NTRU keys allows an attacker to determine both ff^* and gg^*. Note that (ff^*, gg^*) is in the NTRU lattice of hh^* over the totally real subfield $\mathscr{K}^+$, meaning that $ff^*\cdot hh^* \equiv gg^* \pmod q$. Thus, we deduce that $gg^* = \frac{Mhh^*}{1+hh^*} \bmod q$ and $fg^* = \frac{Mh^*}{1+hh^*} \bmod q$ over $\mathscr{R}$ — a step we refer to as "algebra". As observed in [FKT+20], since

f and g are chosen to be co-prime, the attacker can recover a $\mathbb{Z}$-basis of the principal ideal (g) in addition to gg^* through a greatest common divisor computation between the ideals (fg^*) and (gg^*). The attacker can finally retrieve g modulo units through the application of either the Gentry-Szydlo algorithm for power-of-two cyclotomic number fields or its extension for arbitrary cyclotomics, as demonstrated in the attack of Espitau et al. in [EFGT17].

Now if the attacker does not know the value of M exactly, but has a fairly good approximation of it, the preliminary "algebra" can be replaced by lattice reduction. Indeed, write $ff^* + gg^* = qN + E$ for a known[12] N and a small E, so that (ff^*, gg^*, E) is a rather short solution of the linear system

$$\begin{cases} HX - Y = 0 \bmod q, \\ X + Y - E = qN, \end{cases} \tag{10}$$

where $H = hh^*$. More precisely, this value would not correspond to an element of the ring R, but solving such a system amounts to finding a short vector inside the coset $(0, 0, qN) + \mathscr{L}$ (considered inside the extended NTRU lattice in $(\mathscr{K}^+)^3$ corresponding to $\{(u, v, w) | uH = v \pmod q\}$). A (row) basis of the lattice $\mathscr{L}$ corresponding to (10) is given by:

$$L = \begin{pmatrix} 1 & H & H+1 \\ 0 & q & q \end{pmatrix}.$$

and the most efficient known algorithms to solve this problem are essentially variations of lattice reduction and decoding (see for instance [EK20]), and amount in estimating the hardness of retrieving a vector of a given norm inside $\mathscr{L}$. We now give the details to find lower bound on the parameters of the key generation algorithm to make such attacks infeasible.

Distribution of the relative norm vector. We now want to estimate the expected length of (ff^*, gg^*, E). By Corollary 1, we know already $\mathbb{E}[\|ff^*, gg^*\|^2]$. To determine the remaining term $\mathbb{E}[\|E\|^2]$, we must select a convenient value for qN. For this, fix an embedding φ_θ, and let $(F, G) = (\varphi_\theta(ff^*), \varphi_\theta(gg^*))$ and $(\tilde{F}, \tilde{G}) = (|\varphi_\theta(\tilde{f})|^2, |\varphi_\theta(\tilde{g})|^2)$ as in Proposition 1, so that $\mathbb{E}[F + G] = \frac{d}{6} + \frac{R^2+r^2}{2}$. Since each embedding of $ff^* + gg^*$ averages around this (public!) value, we conveniently choose it for qN. From $ff^* + gg^* - qN = E$ and the definition of the variance, we obtain $\mathbb{E}[\|E\|^2] = \mathrm{Var}[F + G]$. The law of total variance with Lemmas 3 and 5 and the fact that $\tilde{F} + \tilde{G}$ is uniform in $[r^2, R^2]$ yield

$$\begin{aligned} \mathbb{E}[\|E\|^2] &= \mathrm{Var}\left[\frac{d}{6} + \tilde{F} + \tilde{G}\right] + \mathbb{E}\left[\frac{d}{12}\left(\frac{d}{6} + 2\tilde{F} + 2\tilde{G}\right)\right] \\ &= \frac{(R^2 - r^2)^2}{12} + \frac{d}{12}(R^2 + r^2) + \frac{d^2}{72}. \end{aligned}$$

For convenience in the next paragraphs, we write $\mathbb{E}[\|ff^*, gg^*\|^2] = 2q^2x$ and $\mathbb{E}[\|E\|^2] = q^2y$, then:

$$\begin{aligned} x \cdot q^2 &= \frac{5}{16}(R^4 + r^4) + \frac{1}{2}(R \cdot r)^2 + \frac{d}{12}(R^2 + r^2) + \frac{d^2}{72}, \\ y \cdot q^2 &= \frac{1}{12}\left(R^2 - r^2\right)^2 + \frac{d}{12}\left(R^2 + r^2\right) + \frac{d^2}{72}. \end{aligned}$$

Mounting the lattice attack. In order to find a short solution for the system in Equation (10), it is known that $\|ff^*\|^2$ and $\|gg^*\|^2$ approximate to xq^2 and $\|E\|^2$ concentrates to yq^2. This results in the vector (ff^*, gg^*, E) being unbalanced with the first two coefficients being significantly larger than the third one. To address this issue, we can utilize a technique similar to the rescaling approach proposed in [BG14, ETWY22].

[12] A typical "known" N would be the radius of a well-chosen circle inside the annulus. This value would not correspond to a ring element in general, but one can reduce to this case in a similar way as SIS and ISIS relate.

It has been observed that in the estimation procedure outlined in Section 4.2.1, the ratio of the length of the secret vector to the normalized volume of the lattice is the only relevant quantity. As such, we can run the same attack under any quadratic twist of the norm of the lattice, by replacing the ℓ_2 norm with any quadratic form of determinant 1, and selecting the one that minimizes the desired ratio. By following the proof technique in [ETWY22], we can restrict ourselves to quadratic forms corresponding to diagonal matrices.

Therefore, to view the corresponding lattice problem in a more suitable manner, we want to analyze it under the twisted (Euclidean) norm encoded by the Gram matrix (clearly of determinant 1) $G_\eta = \mathrm{diag}(\eta, \eta, 1/\eta^2)$ with for $\eta = \left(\frac{y}{x}\right)^{\frac{1}{3}}$. Then under this new norm $\|\cdot\|_\eta$, we find that:

$$\mathbb{E}\left[\|(ff^*, gg^*, E)\|_\eta^2\right] = \eta\mathbb{E}\left[\|ff^*\|^2\right] + \eta\mathbb{E}\left[\|gg^*\|^2\right] + \frac{\mathbb{E}\left[\|E\|^2\right]}{\eta^2} = 3q^2\left(x^2y\right)^{\frac{1}{3}}.$$

Under this norm the lattice $\mathscr{L}$ has $\mathscr{K}^+$-volume:

$$\det(LG_\eta L^T) = \left|\begin{bmatrix} \eta H^2 + \eta + \frac{(H+1)^2}{\eta^2} & \eta Hq + \frac{(H+1)q}{\eta^2} \\ \eta Hq + \frac{(H+1)q}{\eta^2} & \eta q^2 + \frac{q^2}{\eta^2} \end{bmatrix}\right| = q^2\left(\eta^2 + \frac{2}{\eta}\right),$$

giving a lattice of normalized volume being $\sqrt{q}(\eta^2 + \frac{2}{\eta})^{\frac{1}{4}}$ as of $\mathscr{K}^+$-rank 2. The attack is then similar as the one in Section 4.2.1 but where we want to recover a vector of squared norm $3q^2(x^2y)^{\frac{1}{3}}$ in a $\mathbb{Z}$-lattice[13] of normalized (squared) volume $2q(\eta^2 + \frac{1}{\eta})^{\frac{1}{2}}$ of rank $2\frac{d}{2} = d$, yielding a condition of the form:

$$\frac{\beta}{d}3q^2\left(x^2y\right)^{\frac{1}{3}} \leqslant 2q\left(\eta^2 + \frac{2}{\eta}\right)^{\frac{1}{2}}\delta_\beta^{2(2\beta-d+1)} \tag{11}$$

simplifying into:

$$q \leqslant \frac{2d}{3\beta}\sqrt{\frac{y+2x}{x^2y}}\delta_\beta^{2(2\beta-d+1)}.$$

4.2.3 Further optimizations. Beyond the projection trick and the rescaling, we can apply a final standard optimization to this lattice reduction part as there is an unbalance between the size of the secret vector we want to recover and the normalized volume of the lattice. Instead of working with the full lattice coming from the descent of $\mathscr{L}$ over $\mathbb{Z}$, we can instead consider the lattice spanned by a subset of the vectors of the public basis and perform the decoding within this sublattice. The only interesting subset seems to consists in forgetting the $k \leqslant \frac{d}{2}$ first vectors (dropping the so-called q-vectors would not be beneficial as it would actually sparsify the lattice, making the attack worst). Doing so, the rank is of course reduced by k, at the cost of working with a lattice with covolume proportionally $q^{\frac{k}{2(d-k)}}$ bigger. The condition of (11) updates into[14]:

$$\frac{\beta(d-k)}{(d-k)d}3q^2\left(x^2y\right)^{\frac{1}{3}} \leqslant 2q^{\frac{n}{2n-2k}}\left(\eta^2 + \frac{2}{\eta}\right)^{\frac{1}{2}}\delta_\beta^{2(2\beta-d+k+1)},$$

for all $k \in \left\{0, \ldots, \frac{d}{2}\right\}$, which in turn simplifies to:

$$q \leqslant \min_{0\leqslant k\leqslant\frac{n}{2}}\left(\frac{2d}{3\beta}\sqrt{\frac{y+2x}{x^2y}}\delta_\beta^{2(2\beta-d+1)}\right)^{\frac{2n-2k}{n-2k}}. \tag{12}$$

[13] The factor 2 accounting here for the normalized discriminant of the totally real subfield.

[14] This assumes the coefficients of s are balanced, which is a reasonable assumption after the rescaling by η.

The right-hand-side term increases as y becomes smaller making the attack easier and easier, recovering the intuition presented that knowing exactly the value of $ff^* + gg^*$) leads to a complete key recovery in polynomial time. However, because of the rounding to the ring of integer this term cannot be 0: it converges to a term which is greater than $\frac{d^2}{72} + \frac{dq}{6}$. Thus, the condition is *never* satisfied for cryptographically relevant parameters.

Remark 4 (On other subfield type attacks and related). We can also approach the problem as solving a *noisy-ring SIS* instance (namely $(1 + H)F = N + E \pmod q$) or as solving a NTRU instance with a hint, in the spirit of [DDGR20]). In both cases, we are *in fine* decoding a lattice point at distance $\|E\|$ inside a lattice of normalized volume comparable to q. Up to some minor unessential constants, all three approaches give comparable results.

It is tempting to go further and try projection to other subfields, but the ratio secret size to normalized volume is increasing, worsening the attack. It indicates that we shall only focus on the plain NTRU and on the totally real subfield.

4.3 Practical security assessment

This analysis translates into concrete bit-security estimates following the methodology of NewHope [ADPS16] (so-called "core-SVP methodology"). In this model [BDGL16], the bit complexity of lattice sieving (which is asymptotically the best SVP oracle) is taken as $\lfloor 0.292\beta \rfloor$ in the classical setting and $\lfloor 0.259\beta \rfloor$ in the quantum setting in dimension β. Using the analysis presented, we can tailor the radius α of the final annulus to match the desired security level (NIST-I and NIST-V). The size of the signature is then derived similarly as in [ETWY22].

4.4 Extension to more general cyclotomic rings

As discussed at the beginning of this section, the analysis so far has concentrated on base fields $\mathscr{K}$ that are cyclotomic with power-of-two conductor for the sake of simplicity, but it extends with relatively few changes to a more general setting. Specifically, in Appendix B, we show that both the success probability estimates and the security analysis carry over to cyclotomic conductors of the form $m = 2^\ell p^k$ for some odd prime p. This setting encompasses in particular the case of 3-smooth conductors $m = 2^\ell 3^k$ for which parameters are proposed in the Mitaka paper [EFG+22] (and for which we also propose parameters below), and provides plenty of leeway to reach essentially any desired security level.[15]

While the analysis in this more general setting closely mimics the one presented so far, we briefly highlight the ways in which it does differ. The key change is that, for these conductors, the covariance matrix in Heuristic 1 is no longer scalar, making the estimation of the meaningful quantities more subtle. We give a high-level description of the situation here, referring to Appendix B for details.

First, for the success probability of Algorithm 5, the conditional distribution of the embeddings of (f, g) becomes the sum of two non-central χ^2 distributions with different scaling parameters, each corresponding to the eigenvalues $\lambda_+^\theta, \lambda_-^\theta$ of Σ_θ. This complicates the analysis somewhat, but counterparts to the results of Section 4.1 can still be obtained, either through numerical computations or by upper and lower bounding Σ_θ by scalar matrices independent of θ. See in particular Fig. 3.

Second, regarding the security analysis with respect to key recovery attacks, the length of the secret keys is also impacted by the additional error term $T(\theta)$ in Proposition 1. Qualitatively speaking, the behavior in the case $m = 2^\ell p^k$ is however quite close to the

[15] One could in principle generalize the analysis even further (e.g., to arbitrary cyclotomic conductors), but this would introduce additional technicalities (such as the need to replace the power basis by the so-called powerful basis in order to obtain a well-behaved matrix for the canonical embedding), and would really be of theoretical interest at best.

power-of-two case, since for most embeddings, $|S(\theta)| = \left|\frac{\sin(d\theta)}{\sin\theta}\right|$ is small compared to d: only a handful of embeddings have a phase θ close to a multiple of π. We use the worst of these embeddings to bound from above the magnitude of $T(\theta)$, and find that even this pessimistic estimate only has negligible impact on the security level. Lastly, while we could rely on the identity $d\|x\|^2 = \|\varphi(x)\|^2$ in the power-of-two case, this is not true anymore for general conductors; we rely on upper bounds instead. Nevertheless, the geometry of the power basis for 3-smooth conductors remains quite good, acting at worst as an additional $\sqrt{2}$ factor.

5 Antrag in practice

5.1 Optimization and parameter selection

In [ETWY22] new techniques to compress lattice-based hash-then-sign schemes were presented. Theoretically, they can all be applied to Antrag's signatures as well. One of these technique is a fine-tuned encoding approach for discrete Gaussian vectors, and is oblivious to the actual structure of the secret keys — we 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 $q = 12289$ on the one hand, and elliptical sampling on the other hand. They have more impact on the key generation step, and although they were shown somewhat equivalent when applied to scheme such as Falcon or 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, at 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.

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.

We present our parameter selection in Table 2 for power-of-two cyclotomics, and Table 3 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 2. 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 3. 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

5.2 Implementation results

We have implemented our trapdoor generation algorithm Antrag as well as the resulting complete signature scheme in portable C based on the source codes of Falcon and Mitaka. The code archive is attached as supplementary material to this submission.

Since the signature scheme arising from Antrag is essentially identical to Mitaka for signing and verification, we largely reuse the code of Mitaka for those parts. Key generation consists of the original algorithm presented in this paper to generate the first basis vector (f, g), along with code to solve the NTRU equation in order to deduce (F, G), for which we basically reuse the code of Falcon, which follows the techniques presented in [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), our 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 a check in the generation of (f, g) which rejects early the pairs such that the cyclotomic integer prime above 2 divides both f and g (this is a necessary condition for the later computation of F and G to succeed, so it saves some time to test it early).

As explained above, dimension 512 and 1024 are supported, and our `GoodPair` algorithm naturally extends to other conductors such as the 3-smooth cyclotomics considered in Mi-

Table 4. Performance comparison with Falcon and Mitaka.

	Falcon [PFH+22]		Mitaka [EFG+22]		This paper	
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)	—	—	—	—	15.4	55.2
`keygen` speed (ms)	4.7	13.8	*1657**	*6214**	5.7	20.5
`sign` speed (kcycles)	—	—	340	661	334	655
`sign` speed (μs)	204	412	127	246	124	243
`verif` speed (kcycles)	—	—	23	46	23	46
`verif` speed (μs)	21	43	9	17	9	17

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

taka to reach intermediate dimensions, as well as the signing and verification procedure. However, suitably optimized FFT code is needed for those intermediate rings, and more importantly, the NTRUSolve code of [PP19] needs to be adapted as well, in the spirit of, e.g., [LS19]. Neither of those steps are difficult in principle, but they represent a significant engineering effort left as future work.

A performance comparison with Falcon and Mitaka is provided in Table 4, using the same modulus $q = 12289$ for consistency. Compilation is carried out with `gcc` 13.1.1 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.

Acknowledgments

We are indebted to Léo Ducas for the idea of the attack considered in Section 4.2.2, and for invaluable comments and discussions. We thank anonymous reviewers for numerous comments and suggestions for improvement.

Chao Sun was supported by the project "Research and development on IoT malware removal / make it non-functional technologies for effective use of the radio spectrum" among "Research and Development for Expansion of Radio Wave Resources (JPJ000254)", which was supported by the Ministry of Internal Affairs and Communications, Japan. Thi Thu Quyen Nguyen and Alexandre Wallet were supported by PEPR quantique France 2030 programme (ANR-22-PETQ-0008) and by the ANR ASTRID project AMIRAL (ANR-21-ASTR-0016).

References

ABD16. Martin R. Albrecht, Shi Bai, and Léo Ducas. A subfield lattice attack on overstretched NTRU assumptions - cryptanalysis of some FHE and graded encoding schemes. In Matthew Robshaw and Jonathan Katz, editors, *CRYPTO 2016, Part I*, volume 9814 of *LNCS*, pages 153–178. Springer, Heidelberg, August 2016.

ADPS16. Erdem Alkim, Léo Ducas, Thomas Pöppelmann, and Peter Schwabe. Post-quantum key exchange - A new hope. In Thorsten Holz and Stefan Savage, editors, *USENIX Security 2016*, pages 327–343. USENIX Association, August 2016.

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.

BDGL16. Anja Becker, Léo Ducas, Nicolas Gama, and Thijs Laarhoven. New directions in nearest neighbor searching with applications to lattice sieving. In Robert Krauthgamer, editor, *27th SODA*, pages 10–24. ACM-SIAM, January 2016.

BG14. Shi Bai and Steven D. Galbraith. Lattice decoding attacks on binary LWE. In Willy Susilo and Yi Mu, editors, *ACISP 14*, volume 8544 of *LNCS*, pages 322–337. Springer, Heidelberg, July 2014.

CJL16. Jung Hee Cheon, Jinhyuck Jeong, and Changmin Lee. An algorithm for NTRU problems and cryptanalysis of the GGH multilinear map without a low-level encoding of zero. *LMS Journal of Computation and Mathematics*, 19:255–266, 2016.

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.

DDGR20. Dana Dachman-Soled, Léo Ducas, Huijing Gong, and Mélissa Rossi. LWE with side information: Attacks and concrete security estimation. In Daniele Micciancio and Thomas Ristenpart, editors, *CRYPTO 2020, Part II*, volume 12171 of *LNCS*, pages 329–358. Springer, Heidelberg, August 2020.

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.

EFGT17. Thomas Espitau, Pierre-Alain Fouque, Benoît Gérard, and Mehdi Tibouchi. Side-channel attacks on BLISS lattice-based signatures: Exploiting branch tracing against strongSwan and electromagnetic emanations in microcontrollers. In Bhavani M. Thuraisingham, David Evans, Tal Malkin, and Dongyan Xu, editors, *ACM CCS 2017*, pages 1857–1874. ACM Press, October / November 2017.

EK20. Thomas Espitau and Paul Kirchner. The nearest-colattice algorithm. Cryptology ePrint Archive, Report 2020/694, 2020. https://eprint.iacr.org/2020/694.

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.

FKT+20. Pierre-Alain Fouque, Paul Kirchner, Mehdi Tibouchi, Alexandre Wallet, and Yang Yu. Key recovery from Gram-Schmidt norm leakage in hash-and-sign signatures over NTRU lattices. In Anne Canteaut and Yuval Ishai, editors, *EUROCRYPT 2020, Part III*, volume 12107 of *LNCS*, pages 34–63. Springer, Heidelberg, May 2020.

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.

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.

LPR13. Vadim Lyubashevsky, Chris Peikert, and Oded Regev. A toolkit for ring-LWE cryptography. In Thomas Johansson and Phong Q. Nguyen, editors, *EUROCRYPT 2013*, volume 7881 of *LNCS*, pages 35–54. Springer, Heidelberg, May 2013.

LS19. Vadim Lyubashevsky and Gregor Seiler. NTTRU: Truly fast NTRU using NTT. *IACR TCHES*, 2019(3):180–201, 2019. https://tches.iacr.org/index.php/TCHES/article/view/8293.

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.

MW16. Daniele Micciancio and Michael Walter. Practical, predictable lattice basis reduction. In Marc Fischlin and Jean-Sébastien Coron, editors, *EUROCRYPT 2016, Part I*, volume 9665 of *LNCS*, pages 820–849. Springer, Heidelberg, May 2016.

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.

SSS20. Antonio Scala, Carlo Sanna, and Edoardo Signorini. On the condition number of the vandermonde matrix of the nth cyclotomic polynomial. *Journal of Mathematical Cryptology*, 15:174–178, 11 2020.

A Additional proofs

Lemma 5. *Let (z, z') be distributed according to the output of Algorithm 4. Then, the pair $s = \big(|z|^2, |z'|^2\big)$ has the following covariance:*

$$\mathrm{Cov}(s) = \frac{1}{16}\begin{pmatrix} R^4 + r^4 & -\frac{1}{3}(R^4 + 4R^2r^2 + r^4) \\ -\frac{1}{3}(R^4 + 4R^2r^2 + r^4) & R^4 + r^4 \end{pmatrix}$$

and its expectation is:

$$\mathbb{E}[s] = \Big(\frac{R^2 + r^2}{2}, \frac{R^2 + r^2}{2}\Big).$$

Proof. Using the notation of Algorithm 4, the magnitudes $|z|^2$ and $|z'|^2$ satisfy:

$$|z|^2 = \rho^2 \cos^2\theta = u \cdot \cos^2\theta \quad \text{and} \quad |z'|^2 = \rho^2 \sin^2\theta = u \cdot \sin^2\theta$$

where $u = \rho^2$ is uniformly distributed in $[r^2, R^2]$ and θ is uniformly distributed in $[0, \pi/2]$ independently of u. Now let $v = |z|^2 + |z'|^2$ and $w = |z|^2 - |z'|^2$. We thus have:

$$v = u \cdot \big(\cos^2\theta + \sin^2\theta\big) = u \quad \text{and} \quad w = u \cdot \big(\cos^2\theta - \sin^2\theta\big) = u \cdot \cos(2\theta).$$

As a result, v is uniform in $[r^2, R^2]$, and in particular:

$$\mathbb{E}[v] = \frac{R^2 + r^2}{2} \quad \text{and} \quad \mathrm{Var}[v] = \frac{(R^2 - r^2)^2}{12}.$$

Since u and θ are independently distributed, we also have:

$$\mathbb{E}[w] = \mathbb{E}[u] \cdot \mathbb{E}[\cos(2\theta)] = 0 \quad \text{and} \quad \mathbb{E}[vw] = \mathbb{E}[u^2] \cdot \mathbb{E}[\cos(2\theta)] = 0$$

since the mean of cos over $[0, \pi]$ vanishes. Finally:

$$\begin{aligned}\mathrm{Var}[w] = \mathbb{E}[w^2] = \mathbb{E}[u^2] \cdot \mathbb{E}[\cos^2(2\theta)] &= \big(\mathrm{Var}[v] + \mathbb{E}[v]^2\big) \cdot \mathbb{E}\Big[\frac{1 + \cos(4\theta)}{2}\Big] \\ &= \frac{(R^2 - r^2)^2 + 3(R^2 + r^2)^2}{12} \cdot \Big(\frac{1}{2} + 0\Big) = \frac{4R^4 + 4R^2r^2 + 4r^4}{24}.\end{aligned}$$

As a result, the covariance $D = \mathrm{Cov}(v, w)$ satisfies:

$$D = \mathrm{diag}\Big(\frac{(R^2 - r^2)^2}{12}, \frac{R^4 + R^2r^2 + r^4}{6}\Big).$$

To conclude, we observe that the pair s is given by:

$$s = \frac{1}{2}\begin{pmatrix} v + w \\ v - w \end{pmatrix} = T\begin{pmatrix} v \\ w \end{pmatrix} \quad \text{where} \quad T = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}.$$

The claimed value for the expectation of s follows directly, and its covariance satisfies $\mathrm{Cov}(s) = TDT^t$, which yields the claimed result. □

Proof (of Proposition 1). Let us write $(F, G) = (\varphi_\theta(ff^*), \varphi_\theta(gg^*))$ and $(\tilde{F}, \tilde{G}) = (|\varphi_\theta(\tilde{f})|^2, |\varphi_\theta(\tilde{g})|^2)$. Denote $\lambda_+ > \lambda_-$ the eigenvalues of Σ_θ. Let finally Q be the orthogonal matrix such that $Q(\mathbf{I}_2 \otimes \Sigma_\theta)Q^t = \mathrm{diag}(\lambda_+, \lambda_+, \lambda_-, \lambda_-)$. Write $Q(\varphi_\theta(f), \varphi_\theta(g)) = (x, y) \in \mathbb{R}^4$ and $Q(\varphi_\theta(\tilde{f}), \varphi_\theta(\tilde{g})) = (\tilde{x}, \tilde{y})$. By Heuristic 1, the variable $X = \|x\|^2$ is distributed as $\chi^2(2, \lambda_+; \tilde{X})$

where $\tilde{X} = \|\tilde{x}\|^2$, so from the law of total expectation and Lemma 3, we have $\mathbb{E}[X] = 2\lambda_+ + \mathbb{E}[\tilde{X}]$. Similarly, $Y = \|y\|^2$ is distributed as $\chi^2(2, \lambda_-; \tilde{Y})$ and we obtain its expected value in a similar manner. Because Q is orthogonal, we can write

$$\mathbb{E}[F + G] = \mathbb{E}[X] + \mathbb{E}[Y] = 2 \cdot \operatorname{Tr} \Sigma_\theta + \mathbb{E}\big[\tilde{F} + \tilde{G}\big].$$

From Algorithm 4, the variable $\tilde{F} + \tilde{G}$ is uniformly distributed in $[r^2, R^2]$. The first claim follows.

Next, let P be the orthogonal matrix such that $P^t \Sigma_\theta P = \operatorname{diag}(\lambda_+, \lambda_-)$. Our goal is to calculate

$$\mathbb{E}[\|F, G\|^2] = \operatorname{Tr}(\operatorname{Cov}(F, G)) + \big\|\mathbb{E}[(F, G)]\big\|^2,$$

and we will proceed term by term, reducing the situation to $\tilde{F}, \tilde{G}$. Writing $P\varphi(\tilde{f}) = (\tilde{x}_f, \tilde{y}_f)$, we see from Heuristic 1 that F is the sum of two independent non-central chi-squared distributions, namely $F \sim \chi^2(1, \lambda_+; \tilde{x}_f^2) + \chi^2(1, \lambda_-; \tilde{y}_f^2)$. Then by Lemma 3, the conditional expectation and variance of F express as

$$\begin{aligned} \mathbb{E}[F|\tilde{F}] &= \operatorname{Tr} \Sigma_\theta + \tilde{F}, \\ \operatorname{Var}[F|\tilde{F}] &= 2 \cdot (\operatorname{Tr} \Sigma_\theta^2 - 2 \det \Sigma_\theta) + 4 \cdot \|\varphi_\theta(\tilde{f})\|_\theta^2, \end{aligned}$$

and the situation is analogous for G with respect to $P\varphi_\theta(\tilde{g})$. On the one hand, the law of total expectation then gives us $\mathbb{E}[F] = \operatorname{Tr} \Sigma_\theta + \mathbb{E}[\tilde{F}]$ and similarly for $\mathbb{E}[G]$, from which we get

$$\big\|\mathbb{E}[(F, G)]\big\|^2 = 2 \cdot \operatorname{Tr} \Sigma_\theta^2 + 2 \cdot \operatorname{Tr} \Sigma_\theta \cdot \mathbb{E}[\tilde{F} + \tilde{G}] + \big\|\mathbb{E}[(\tilde{F}, \tilde{G})]\big\|^2. \tag{13}$$

On the other hand, the law of total covariance yields $\operatorname{Var}[F] = \operatorname{Var}[\tilde{F}] + 2(\operatorname{Tr} \Sigma_\theta^2 - 2 \det \Sigma_\theta) + 4 \cdot \mathbb{E}\big[\|\varphi_\theta(\tilde{f})\|_\theta^2\big]$ with a similar expression for $\operatorname{Var}[G]$. We deduce

$$\operatorname{Tr}(\operatorname{Cov}(F, G)) = \operatorname{Tr}(\operatorname{Cov}(\tilde{F}, \tilde{G})) + 4 \cdot (\operatorname{Tr} \Sigma_\theta^2 - 2 \det \Sigma_\theta) + 4 \cdot \mathbb{E}\big[\|\varphi_\theta(\tilde{f})\|_\theta^2 + \|\varphi_\theta(\tilde{g})\|_\theta^2\big]. \tag{14}$$

The second claim follows by combining Equations (13) and (14) with the expression of Σ_θ and Lemma 5, as well as that $\tilde{F} + \tilde{G}$ is uniformly distributed in $[r^2, R^2]$. □

Proof (of Corollary 1). Recall that $\varphi_\theta(ff^*) = |\varphi_\theta(f)|^2$ for any argument θ of a cyclotomic root, and that $S(\theta) = 0$ in the power-of-two case. Moreover, we have $T(\theta) = \frac{d}{12}(R^2 + r^2)$ in Proposition 1 for such rings. The claimed expressions follow from the identity $d\,\|x\|^2 = \|\varphi(x)\|^2$ and our assumption that all embeddings behave independently. □

Proof (of Corollary 2). Let V be the linear transformation such that $\varphi(f, g) = V(f, g)$, we want to know the square of its smallest singular value. With [SSS20, Lemma 2.1 and 2.2], the eigenvalues of V^*V are those of the Vandermonde for the prime case scaled by $\frac{d}{p-1}$. Combining with e.g. [LPR13, Lemma 4.3] for the prime case, we obtain that the spectrum (without multiplicity) of V^*V is $\{\frac{m}{2}, \frac{d}{p-1}\}$ if m is even and $\{m, \frac{d}{p-1}\}$ is odd. This means that we have $\|x\|^2 \leqslant \frac{p-1}{d}\|\varphi(x)\|^2$ for all $x \in \mathscr{K}$. The inequalities now follow from Proposition 1, Lemma 7 and the properties of the expectation. □

B Analysis over cyclotomic fields of conductor $m = 2^\ell p^k$

As mentioned in Section 4.4, the analysis in the general case of cyclotomic base rings of conductor $m = 2^\ell p^k$ is slightly more involved, although it proceeds in a largely similar way at the power-of-two cyclotomic case.

For the success probability of Algorithm 5, the squared length of a pair of (fixed) embeddings now follows a sum of two non-central χ^2 distributions of degree 2 and distinct scaling in general. The cumulative distribution function of such a law does not admit a simple expression in terms of classically defined special functions, so giving a manageable analytic

expression of the success probability in that case is difficult. One *can* however provide a loose lower bound of success probability analogous to Eq. (7). This is done by overestimating the probability of the embedding falling outside of the outer circle of the target annulus or inside of the inner circle using the more pessimistic eigenvalue of Σ_θ in each case, thus reducing the estimate to the analysis of a single non-central χ^2 distribution of degree 4 again. This makes it possible to prove that one can achieve $\alpha = O(1)$ in this setting as well as long as $q = \Omega(d \log d)$.

Since such a result is of little concrete interest in practice, however, we omit the details of that analysis, and instead observe that the description above of the distribution of the squared length of the pair of embedding is in fact sufficient to carry out precise numerical computations of the success probability, analogous to the results presented at the end of Section 4.1. This lets us analyze the behavior of the success probability as a function of α for all parameters of interest.

For the security against key recovery attacks, the situation is made simpler by Proposition 1 and the fact that we can accept giving more power to an attacker (or equivalently, overestimate lengths related to secret keys) as long as it does not significantly impact our security level. We find that pessimistic overestimates of all of the quantities involved do not in fact result in a meaningful security loss, so that the security analysis is mostly unchanged in this setting.

B.1 Success probability

We start with some notation to describe the necessary quantities. For some fixed embedding φ_θ, let $X = (\varphi_\theta(f), \varphi_\theta(g))$. Conditioned on $\tilde{x} := (\varphi_\theta(\tilde{f}), \varphi_\theta(\tilde{g}))$, we have $X_{\tilde{x}} \sim \mathcal{N}(\tilde{x}, \mathbf{I}_2 \otimes \Sigma_\theta)$ from Heuristic 1. Now, Σ_θ has two distinct eigenvalues λ_+^θ and λ_-^θ. Let $Q \in \mathbb{R}^{4\times 4}$ be the orthogonal matrix such that $\mathbf{I}_2 \otimes \Sigma_\theta = Q^t \Lambda Q$, where $\Lambda := \mathrm{diag}(\lambda_+^\theta, \lambda_+^\theta, \lambda_-^\theta, \lambda_-^\theta)$. Then the variable $QX \in \mathbb{R}^4$ is distributed as $\mathcal{N}(Q\tilde{x}, \Lambda)$. Writing $Q\tilde{x} = (\nu_+, \nu_-)$ with $\nu_+, \nu_- \in \mathbb{R}^2$ satisfying that $\|\nu_+\|^2 + \|\nu_-\|^2 = \|\tilde{x}\|^2$ or rather $\|\nu_+\|^2 + \|\nu_-\|^2 = \beta^2 q$ (since $\beta = \|\tilde{x}\|/\sqrt{q}$ as defined in Section 4.1). Thanks to Q being orthogonal and Λ being diagonal, we have

$$\|X_{\tilde{x}}\|^2 \sim \chi^2(2, \lambda_+^\theta; \|\nu_+\|^2) + \chi^2(2, \lambda_-^\theta; \|\nu_-\|^2). \tag{15}$$

Lastly, let $z := \|\nu_+\|^2/(\beta^2 q) \in [0,1]$, then $\|\nu_-\|^2/(\beta^2 q) = 1 - z$. We rewrite (15) as

$$\frac{1}{q}\|X_{\tilde{x}}\|^2 \sim \chi^2\Big(2, \frac{\lambda_+^\theta}{q}; \beta^2 z\Big) + \chi^2\Big(2, \frac{\lambda_-^\theta}{q}; \beta^2(1-z)\Big), \quad \text{where} \quad \frac{\lambda_\pm^\theta}{q} = \frac{d \pm \left|\frac{\sin(d\theta)}{\sin\theta}\right|}{24q}. \tag{16}$$

Let $p_{\mathrm{succ}}^\theta(z, \beta)$ be the probability that $X_{\tilde{x}}$ satisfying the condition (5) given the value of (z, β), meaning that $1/\alpha^2 \leqslant \|X_{\tilde{x}}\|^2/q \leqslant \alpha^2$. This probability is therefore:

$$p_{\mathrm{succ}}^\theta(z, \beta) = F_{z,\beta}(\alpha^2) - F_{z,\beta}(1/\alpha^2)$$

for $F_{z,\beta}$ the cumulative distribution function of the combination of non-central χ^2 distribution from (16), which can be computed numerically, e.g., by the Python package `chi2comb`. Let furthermore f be the probability density function of the random variable z. Then, recalling that $\beta^2 q$ is uniformly distributed in $[r^2, R^2]$, the success probability $p_{\text{succ-one}}^\theta$ for the embedding φ_θ is expressed as

$$\begin{aligned} p_{\text{succ-one}}^\theta &= \frac{1}{R^2 - r^2} \int_{r^2}^{R^2} \left(\int_0^1 p_{\mathrm{succ}}^\theta(z, \beta) f(z)\, dz \right) d(\beta^2 q) \\ &= \frac{2q}{R^2 - r^2} \int_{r/\sqrt{q}}^{R/\sqrt{q}} \int_0^1 p_{\mathrm{succ}}^\theta(z, \beta) f(z)\, dz\, d\beta. \end{aligned}$$

Accordingly, under our heuristic assumption that the embeddings are independently distributed, the success rate $p_{\text{succ-all}}$ for all $d/2$ embeddings is $p_{\text{succ-all}} = \prod_\theta p_{\mathrm{succ}}^\theta$, and the

repetition rate M is $M = p_{\text{succ-all}}^{-1}$. All of these values can thus be evaluated numerically provided that we find the PDF $f(z)$ of the random variable z. It is given by the following lemma.

Lemma 6. *The probability density function f of the random variable z defined above is given by:*

$$f(z) = \frac{8}{\pi^3} K(z)K(1-z),$$

where K is the complete elliptic integral of the first kind (with the parameter *notation,* not *the elliptic modulus notation):*

$$K(m) = \int_0^{\pi/2} \frac{dt}{\sqrt{1 - m\sin^2 t}}.$$

Proof. The embeddings $\varphi_\theta(\tilde{f})$ and $\varphi_\theta(\tilde{g})$ are sampled by independently sampling $\rho^2 = \beta^2 q$ uniform in $[r^2, R^2]$, a uniform in $[0, \pi/2]$ and ω, ω' uniform in $[0, 2\pi]$ and letting:

$$\tilde{x} = \big(\varphi_\theta(\tilde{f}), \varphi_\theta(\tilde{g})\big) = \big(\rho\cos(a)e^{i\omega}, \rho\sin(a)e^{i\omega'}\big) \in \mathbb{C}^2.$$

Moreover, Σ_θ is diagonalized as $R^t \cdot \mathrm{diag}(\lambda_+^\theta, \lambda_-^\theta) \cdot R$ for some rotation matrix R (acting on $\mathbb{C}$ by multiplication by $e^{i\psi}$, say). Then:

$$R \cdot \varphi_\theta(\tilde{f}) = \rho\cos(a)e^{i(\omega+\psi)} \quad \text{and} \quad R \cdot \varphi_\theta(\tilde{g}) = \rho\sin(a)e^{i(\omega'+\psi)}.$$

It follows that, as a vector in $\mathbb{R}^4$:

$$Q\tilde{x} = \big(\rho\cos(a)\cos(\omega+\psi), \rho\sin(a)\cos(\omega'+\psi), \rho\cos(a)\sin(\omega+\psi), \rho\sin(a)\sin(\omega'+\psi)\big),$$

as the first two components correspond to the eigenvalue λ_+^θ. In particular:

$$\begin{aligned}
\|\nu_+\|^2 &= \big\|\big(\rho\cos(a)\cos(\omega+\psi), \rho\sin(a)\cos(\omega'+\psi)\big)\big\|^2 \\
&= \rho^2\big[\cos^2 a \cdot \cos^2(\omega+\psi) + \sin^2 a \cdot \cos^2(\omega'+\psi)\big], \\
z &= \frac{\|\nu_+\|^2}{\rho^2} = \cos^2 a \cdot \cos^2(\omega+\psi) + \sin^2 a \cdot \cos^2(\omega'+\psi).
\end{aligned}$$

Now $\omega + \psi$ and $\omega' + \psi$ are independent and uniformly distributed modulo 2π, and since, moreover, $\cos^2$ is even and π-periodic, we see that the distribution of z is the distribution of $\cos^2 a \cdot \cos^2 b + \sin^2 a \cdot \cos^2 c$ for independent uniform random variables $a, b, c \in [0, \pi/2]$.

Let $u = \cos^2 a$, $v = \cos^2 b$ and $w = \cos^2 c$. These are independent and uniformly distributed random variables with values in $[0, 1]$, whose common probability density function h is given by:

$$h(u) = \frac{d}{du}\Pr[\cos^2 a \leqslant u] = \frac{d}{du}\Pr[a \geqslant \arccos\sqrt{u}] = \frac{d}{du}\Big(1 - \frac{2}{\pi}\arccos\sqrt{u}\Big) = \frac{1/\pi}{\sqrt{u}\sqrt{1-u}}.$$

The density $g_{v,w}$ of $z = uv + (1-u)w$ conditioned on fixed values v, w, where we assume without loss of generality that $v > w$, is then given by:

$$\begin{aligned}
g_{v,w}(z) &= \frac{d}{dz}\Pr[(v-w)u - w \leqslant z] = \frac{d}{dz}\Pr\Big[u \leqslant \frac{z-w}{v-w}\Big] \\
&= \frac{d}{dz}H\Big(\frac{z-w}{v-w}\Big) = \frac{1}{v-w}h\Big(\frac{z-w}{v-w}\Big) \\
&= \frac{1/\pi}{\sqrt{v-z}\sqrt{z-w}}
\end{aligned}$$

for all $z \in (w, v)$, where H denotes the CDF of u (i.e., the antiderivative of h vanishing at 0). Overall, it follows that the density f of z (without conditioning) satisfies:

$$f(z) = 2 \iint_{0 \leqslant w < z < w \leqslant 1} g_{v,w}(z) h(v) h(w) \, dv \, dw$$
$$= \frac{2}{\pi^3} \int_0^z \frac{dw}{\sqrt{w}\sqrt{z-w}\sqrt{1-w}} \int_z^1 \frac{dv}{\sqrt{v}\sqrt{v-w}\sqrt{1-v}}$$

by Fubini (note that we can indeed take the integral over the domain $0 \leqslant w < z < v \leqslant 1$ since for fixed v, w the density vanishes for z not between v and w, and the two cases $v > w$ and $v < w$ are clearly equivalent by symmetry, hence the factor 2). In this last formula, the integral in w is readily seen to be equal to $2K(z)$ by the change of variable $w = z \sin^2 t$, and the integral in v takes the same form as the first one by the change of variable $v \mapsto 1 - v$. This concludes the proof. □

We can numerically confirm that this approach correctly models the repetition rate observed in practice for non-power of two cyclotomic fields, as shown in Fig. 3.

B.2 Security analysis

We want to understand the expected values of both $\|f, g\|^2$ and $\|ff^*, gg^*\|^2$. Qualitatively speaking, the behavior in the case $m = 2^\ell p^k$ is in fact quite close to the power-of-two case, since for most embeddings, $|S(\theta)| = \left|\frac{\sin(d\theta)}{\sin\theta}\right|$ is small compared to d: $\sin\theta$ is bounded away from zero except possibly for just a handful of embeddings with θ close to a multiple of π.

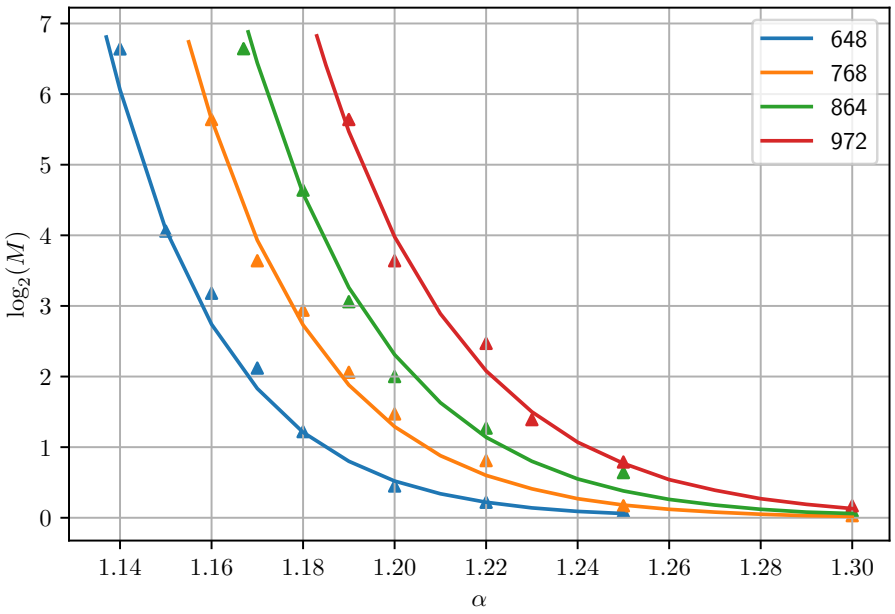


Fig. 3. Base 2 logarithm of the repetition rate M of Algorithm 5 as a function of α, for $d \in \{648, 768, 864, 972\}$, $q = 12289$ and $\xi = \frac{1}{3}$. The continuous lines are obtained based on our model, and the triangle data points are measured by simulations (averaging 100 iterations of the algorithm for each data point).

For those few embeddings, success probability tends to become slightly worse. Nevertheless, even in the worst case, we have:

$$S(\theta) = \frac{\sin(2\pi/3)}{\sin(2\pi/3d)} \approx \frac{\sqrt{3}/2}{2\pi/3d} = \frac{3\sqrt{3}}{4\pi} d \approx 0.413d.$$

We shall see shortly that the quantity $\|f, g\|^2$ is not difficult to estimate for the conductors we are interested in, although exact formulas cannot be obtained. For the quantity $\|ff^*, gg^*\|^2$, compared to the power-of-two case, we now have the additional term depending on the embedding considered, namely

$$T(\theta) = \frac{|S(\theta)|^2}{72} + 4 \cdot \mathbb{E}\left[\|\varphi_\theta(\tilde{f})\|_\theta^2 + \|\varphi_\theta(\tilde{g})\|_\theta^2\right]$$

as seen in Proposition 1. The next result gives a clean (yet not tight) estimate.

Lemma 7. *We have* $T(\theta) \leqslant \frac{d^2}{288} + \frac{d}{4}(R^2 + r^2)$.

Proof. From the observations above, we have $|S(\theta)| \leqslant \frac{d}{2}$, and hence $d + |S(\theta)| \leqslant \frac{3}{2}d$. By Lemma 5 and the monotonicity of expectation, we obtain $\mathbb{E}\left[\|\varphi_\theta(\tilde{f})\|_\theta^2 + \|\varphi_\theta(\tilde{g})\|_\theta^2\right] \leqslant \lambda_+^\theta \cdot (R^2 + r^2)$. The inequality follows using $\lambda_+^\theta = \frac{d+|S(\theta)|}{24}$. □

On the one hand, the quantity $T(\theta)$ is involved in the attack over the relative subfield $\mathscr{K}^{++}$, and therefore has to be compared to the normalized volume of the corresponding relative lattice, which is of the order of q^2. Recall that R^2, r^2 are both of the order of q, and observe in the next result that (ff^*, gg^*) is on the other hand of the order of magnitude of q^2. As claimed, this pessimistic estimate on $T(\theta)$ should therefore have a limited negative impact on the attack.

Corollary 2 (of Proposition 1, heuristic). *Let K be a cyclotomic field of conductor $m = 2^\ell p^k$ with p odd and $\ell \geqslant 0, k \geqslant 1$. With the notation of Algorithm 5, let (f, g) be a random variable following the distribution of its output. Then we have*

$$\mathbb{E}[\|f, g\|^2] \leqslant (p-1)\left(\frac{d}{6} + \frac{R^2 + r^2}{2}\right),$$
$$\mathbb{E}[\|ff^*, gg^*\|^2] \leqslant (p-1)\left(\frac{5}{8}(R^4 + r^4) + R^2 r^2 + \frac{d}{3}(R^2 + r^2) + \frac{d^2}{32}\right).$$

The proof of this corollary is a straightforward norm computation and is provided in Appendix A.

With these estimates in mind, we can carry out the same analysis as in Section 4.2.2, replacing the expected values by their pessimistic upper bound. Keeping the notation of that section, the last ingredient is to express $\mathbb{E}[\|E\|^2]$, where $E = ff^* + gg^* - qN$ for some carefully chosen qN. We will follow the same roadmap as in the power-of-two case. From Proposition 1, we note that the average embedding of $ff^* + gg^*$ does not depend on the phase θ, and we let $qN = \mathbb{E}[\varphi_\theta(ff^*) + \varphi_\theta(gg^*)] = \frac{d}{6} + \frac{R^2+r^2}{2}$. Following the proof of Corollary 2 and using also the definition of the variance, we then obtain

$$\mathbb{E}\left[\|E\|^2\right] \leqslant (p-1) \cdot \max_\theta \mathrm{Var}\left[\varphi_\theta(ff^*) + \varphi_\theta(gg^*)\right].$$

For a fixed embedding φ_θ, the law of total variance and the calculations done in Proposition 1 give $\mathrm{Var}\left[\varphi_\theta(ff^*) + \varphi_\theta(gg^*)\right] = \frac{(R^2-r^2)^2}{12} + \frac{d^2}{72} + T(\theta)$. Lemma 7 then yields

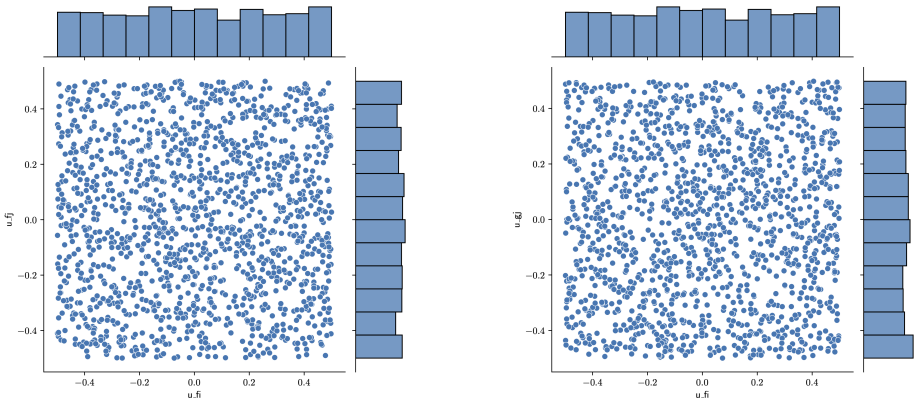
$$\mathbb{E}\left[\|E\|^2\right] \leqslant (p-1) \cdot \left(\frac{(R^2 - r^2)^2}{12} + \frac{d}{4}(R^2 + r^2) + \frac{5d^2}{288}\right).$$

C Experimental data

This supplementary material collects data aimed at justifying the heuristics of Section 3, in the form of the following figures.

Figure 4 shows that the embeddings of the error e_f and e_g do behave as independent and uniform in $[-1/2, 1/2)$. We additionally confirm this graphical observation of the uniformity and independence using statistical χ^2 tests. Multiple experiments of goodness of fit (with bin size 20 and 1500 samples each) and independence (with bin size 400 and 1500 samples each) yield expectedly random-looking p-values, consistently above 0.05.

Figure 5 shows that the error magnitude on an embedding has the expected distribution (namely, a scaled $\chi(4)$) in the power-of-two cyclotomic case, and Fig. 6 illustrates the situation of 3-smooth base fields, again showing that experimental results match our model.



(a) $\left(e_f^{(21)}, e_f^{(50)}\right)$ (b) $\left(e_f^{(101)}, e_g^{(3)}\right)$

Fig. 4. Empirical joint distributions of two randomly coefficients of e_f (resp. a randomly chosen coefficient of e_f and another of e_g). The data is collected from 1500 samples (f, g) of degree $d = 512$.

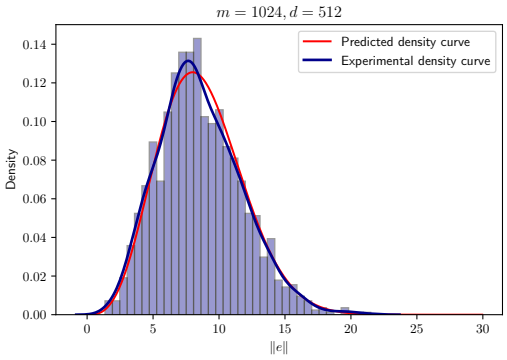


Fig. 5. Statistical density of $\|\varphi_i(e)\|$ in case $m = 1024, d = 512$ (1500 samples).

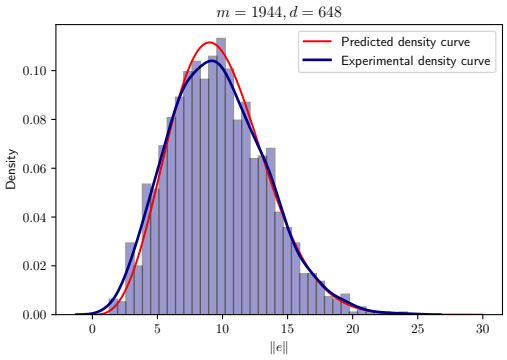


Fig. 6. Statistical density of $\|\varphi_i(e)\|$ in case $m = 1944, d = 648$ (1500 samples).