Polytopes in the Fiat-Shamir with Aborts Paradigm

Henry Bambury[1,2], Hugo Beguinet[1,3], Thomas Ricosset[3], and Éric Sageloli[1,3,4]

[1] DIENS, École normale supérieure, CNRS, Inria, PSL University, Paris, France
{henry.bambury, hugo.beguinet, eric.sageloli}@ens.fr
[2] DGA
[3] Thales, Gennevilliers, France
{hugo.beguinet, thomas.ricosset, eric.sageloli}@thalesgroup.com
[4] École polytechnique, Institut polytechnique de Paris, Palaiseau, France

Abstract. The Fiat-Shamir with Aborts paradigm (FSwA) uses rejection sampling to remove a secret's dependency on a given source distribution. Recent results revealed that unlike the uniform distribution in the hypercube, both the continuous Gaussian and the uniform distribution within the hyperball minimise the rejection rate and the size of the proof of knowledge. However, in practice both these distributions suffer from the complexity of their sampler. So far, those three distributions are the only available alternatives, but none of them offer the best of all worlds: competitive proof of knowledge size and rejection rate with a simple sampler.
We introduce a new generic framework for FSwA using polytope based rejection sampling to enable a wider variety of constructions. As a matter of fact, this framework is the first to generalise these results to integral distributions. To complement the lack of alternatives, we also propose a new polytope construction, whose uniform sampler approaches in simplicity that of the hypercube. At the same time, it provides competitive proof of knowledge size compared to that obtained from the Gaussian distribution. Concurrently, we share some experimental improvements of our construction to further reduce the proof size. Finally, we propose a signature based on the FSwA paradigm using both our framework and construction. We prove it to be competitive with Haetae in signature size and with Dilithium on sampler simplicity.

Keywords: Zero-Knowledge Proofs · Lattice-based Cryptography · Fiat-Shamir with Aborts · Rejection Sampling · Integral Polytope Uniform Sampling

1 Introduction

Lattice-based cryptography offers numerous advantages over traditional number-theoretic public-key cryptography. These advantages span from conjectured resistance to quantum attacks to the capability of performing arbitrary computations on encrypted data, all while maintaining comparable or even superior

efficiency. However, a notable challenge persists: the need to reduce the size of transmittable elements, including zero-knowledge proofs of knowledge (ZKPoK).

Even when using algebraic lattices, zero-knowledge proofs still tend to be at least an order of magnitude larger than their traditional counterparts. Consequently, the transition towards this so-called post-quantum cryptography, driven by the release of the first standards, presents a series of challenges. These challenges include a substantial increase in bandwidth consumption. Presently, these issues serve as barriers to the widespread adoption of lattice-based cryptography.

Zero-knowledge. There exists a wide variety of lattice-based ZKPoK constructions, starting with [KTX08] and seeing improvements in [LNSW13]. This evolution has led to multiple lines of work, in particular to the birth of the Fiat-Shamir with Aborts [Lyu09] paradigm. In this paradigm, the crucial zero-knowledge step is done through a rejection sampling algorithm in order to remove any sort of secret dependency from the output distribution. This led to a plethora of different improved constructions [BLNS20, LNS20, LNS21a, LNP22], from basic signatures [DKL+21, CCD+23] to blind [BLNS23a] and group signatures [dLS18, LNS21b] as well as anonymous credentials [BLNS23b].

Fiat-Shamir with aborts. We particularly focus on the recent [DFPS22] as a foundational work in the study of FSwA, specifically regarding the rejection sampling theorem from [Lyu12]. Informally, it studies the rejection rates and proof of knowledge sizes obtained from the following existing distributions: Gaussian distribution, bimodal distribution [DDLL13], uniform distribution inside a hypercube, and uniform distribution inside a hyperball with its bimodal counterpart [CCD+23]. The authors obtain generic optimal bounds for these two metrics and prove that both the Gaussian and the hyperball uniform distributions achieve them, whereas the hypercube shows poor results in these aspects. This raises a natural question regarding the widespread usage of uniform distributions within the hypercube. The explanation of its attractiveness comes from its simplicity and most notably the simplicity of its sampler. Which is, indeed, another important metric for such primitives in their applications. In this aspect, both the Gaussian and the hyperball uniform distributions suffer from their respective samplers compared to the trivial sampler inside a hypercube. Ensuring the secure generation of these crucial samples has proven to be surprisingly challenging in an efficient and provably resistant against side-channel attacks fashion [BHLY16, EFGT17, PBY17, GMRR22, Pre23]. Moreover, the analysis of [DFPS22] is conducted in the continuous setting, leaving a blurry gap between the theory (studying volumes) and the application (restricting the volume to its integral points). In most lattice-based cases, the primitives are handled both with the samplers and the operations over integers.

By definition, at some point the need to restrict results to integers is necessary, however it appears that generic constructions dealing with this restriction part is missing in the literature.
We highlight the different missing parts in the literature as well as some existing problem in the case of FSwA:

Is it possible to have a generic approach on rejection sampling directly with the restriction to integers? If yes, how?
Would it leads to overall practical constructions with few to no inconvenient?
In other words, is there a construction that lives in the best of both the hyperball and hypercube worlds (simple sampler, competitive proof of knowledge size, simple characterisation)?

Contributions. In this paper, we address the aforementioned problems through a comprehensive analysis of rejection sampling using integral uniform distributions for cryptographic applications. We provide an improvement in the trade-off between the proof of knowledge size and the simplicity of the samplers using a new construction.

Before diving into technical details, let's review the main (although not exhaustive) choices for distributions with regards to rejection sampling in cryptography: the Gaussian and uniform distributions. Due to the shortcomings of its sampler and its rejection condition, we choose to exclude the Gaussian distribution from the scope of our study. Additionally, part of our results are generalisable to L_p balls and in particular hold for L_2 balls. However, the necessary bridge between rejection sampling inside a continuous L_2 ball and rejection sampling inside its discrete restriction to integers appears to be false in general. Thus, we make the choice to focus on more structured convex spaces, and deal only with polytopes in order to obtain a main rejection sampling theorem on uniform distributions over integral sets.

Our contributions can be summarised as follows:

- We provide a generic study over polytopes to describe the rejection sampling procedure when source and target distributions are uniform distributions on polytopes. Informally, for a fixed polytope $\mathcal{P}$ and chosen rejection rate M (expected number of repetitions), we give a closed form to obtain the minimal couple (R, r) such that: for any translation $\mathbf{v} \in \mathbb{Z}^n$, with source distribution $\mathcal{U}(R \cdot \mathcal{P} + \mathbf{v})$ and target distribution $\mathcal{U}(r \cdot \mathcal{P})$, the average rejection rate is M. More importantly, we prove that this result extends to its restriction to integers (*i.e.* on $\mathcal{P} \cap \mathbb{Z}^n$) under some specific constraints on $\mathcal{P}$ to answer the first interrogation above.
- By summing-up previous remarks, a polytope is attractive for FSwA because of its sampler simplicity and the estimation of its proof of knowledge size. For the first part this can be done through a considerable number of sub-metrics (randomness usage, running time ...) but for the latter, an appropriate way to measure the size of the knowledge proof is to compute the ratio between the radius of the sphere circumscribed to $\mathcal{P}$ with the radius of its inscribed sphere (in L_2 metric). For example if n is the dimension, this ratio for a hypercube is equal to $\sqrt{n}$. In this part we introduce $\mathcal{H}$, defined as the intersection of a hypercube with its dual L_1 ball. In comparison with the hypercube, $\mathcal{H}$ has a ratio of $\sqrt[4]{n}$ which, for cryptographic parameters, varies between approximately 33 for the hypercube to approximately 6 for $\mathcal{H}$. Additionally, we define a uniform and isochronous sampler in $\mathcal{H} \cap \mathbb{Z}^n$ using only uniform

sampling (to be compared with the hyperball sampler that uses Gaussian sampling).
- Lastly we share some applications to both our main theorem on rejection sampling and our construction $\mathcal{H}$. We start by introducing an experimental improvement over $\mathcal{H}$ using an additional fine-tuned Euclidean norm bound to go from a ratio of 6 using cryptographic parameters to 1.5. Then, we wrap up our contributions by constructing a FSwA signature called Patronus which is an attractive choice over existing FSwA schemes, as much for its signature size as for its sampler simplicity.

Technical details. This work focuses on identifying sets of vectors that are highly compatible with the Fiat-Shamir with Aborts (FSwA) paradigm, with a particular emphasis on optimising vector sizes after rejection and facilitating implementation. Within the FSwA paradigm, an element $\mathbf{z} = \mathbf{y} + c\mathbf{s}$ is retained only if it reveals no information regarding the possible values of $\mathbf{y}$ and $c\mathbf{s}$; otherwise, it is rejected, repeating the process. If we denote by V_Y the set of possible $\mathbf{y}$ values (from which $\mathbf{y}$ is uniformly sampled) and $V_{c\mathbf{s}}$ as the set of possible $c\mathbf{s}$ values, it is clear that $\mathbf{z}$ avoids information leakage if and only if:

$$V_Z \subseteq \bigcap_{\mathbf{x} \in V_{c\mathbf{s}}} (V_Y + \mathbf{x}).$$

Furthermore, V_Z minimises the number of rejects if and only if:

$$V_Z = \bigcap_{\mathbf{x} \in V_{c\mathbf{s}}} (V_Y + \mathbf{x}).$$

To achieve this, we must identify sets V_Y and $V_{c\mathbf{s}}$ that satisfy several essential constraints:

1. **Restriction to $\mathbb{Z}^n$:** V_Y and $V_{c\mathbf{s}}$ must be subsets of $\mathbb{Z}^n$.
2. **Simple membership test:** We must have a straightforward way to determine membership in V_Z, as it is essential for characterising V_Z and for efficient implementation.
3. **Minimising aborts:** To minimise the occurrence of restarts, our objective is to have V_Z closely approximate the set encompassing all possible $\mathbf{y} + c\mathbf{s}$ values.
4. **Hyperball approximation:** Given that the hyperball offers optimal proof of knowledge sizes, our objective is to identify a set V_Z that closely approximates it.
5. **Efficient uniform sampling:** Since $\mathbf{y}$ is uniformly sampled from V_Y, we need an efficient method to sample from V_Y.

Interestingly, no existing solutions excel in addressing these five challenges:

– **Hyperball:** The hyperball approach, as presented in [CCD+23], inherently solves the hyperball approximation and has been proved to minimise the number of aborts in [DFPS22]. Additionally, membership verification only

needs a computation of the L_2 norm. However, sampling in the integral points of the hyperball is complex. In [CCD+23], they manage to avoid dealing with floating points contrarily to the Gaussian approach. But, they still rely on a sub-procedure that uses Gaussian sampling with 128-bit precision fixed-point arithmetic.

– **Hypercube:** In contrast, sampling within the hypercube, as in [DKL+21], is straightforward and has been extensively studied for various rejection sampling applications. Nevertheless, in high dimensions, the hypercube's vertices are distant from the inscribed sphere, resulting in a ratio of $\sqrt{n}$ as mentioned previously.
– **Gaussian:** The discrete Gaussian distribution, as in [Lyu12], is more or less the opposite of the hypercube distribution. It offers the lowest proof of knowledge size and minimises the number of aborts. However, it entails the most complex rejection sampler: the rejection step involves computing a transcendental function on input dependent on a secret, and as previously noted, Gaussian samplers pose significant challenges in terms of efficiently and provably resisting side-channel attacks. While Gaussian distributions are ideal theoretical tools, their practical implementation remains more challenging compared to the previous distributions.

To answer these fives challenges we first build a generic framework for FSwA proofs of knowledge using uniform distributions in polytopes. We use the following notation $\mathcal{P}^n_{r,\mathbf{v},\mathbb{Z}} = (\mathbf{v} + r \cdot \mathcal{P}) \cap \mathbb{Z}^n$ for a polytope $\mathcal{P}$ (we omit $\mathbf{v}$ in the subscript if $\mathbf{v} = 0$) and define $\mathcal{R}_\infty$ as the ∞-Rényi divergence. Then, by associating $V_Y = \mathcal{P}^n_{R,\mathbb{Z}}$ and $V_{\mathbf{cs}} = \mathcal{P}^n_{\beta,\mathbb{Z}}$, we summarise our generic FSwA result with the following simplification of our main theorem :

Theorem (Rejection sampling on polytopes). *Let $\mathcal{P}$ be a polytope, $M > 1$, $\beta > 0$ and h a probability distribution such that $\mathsf{Supp}(h) \subseteq \mathcal{P}^n_{\beta,\mathbb{Z}}$. There exist $r_M, R \in \mathbb{Z}_{>0}$, with r_M computable and $R \geq r + \beta$. Let $\varepsilon_{n,r_M}, \varepsilon_{n,R} \in \mathbb{R}_{>0}$ and $M' = \frac{1+\varepsilon_{n,R}}{1+\varepsilon_{n,r}} \cdot M$. Let $\mathbf{v} \in \mathcal{P}^n_\beta$ and define $\rho^n_{r,\mathbf{v}} := \mathcal{U}\left(\mathcal{P}^n_{r,\mathbf{v},\mathbb{Z}}\right)$. If $M' > 1$ then:*

$$\mathcal{R}_\infty \left[\mathcal{U}\left(\mathcal{P}^n_{r,\mathbb{Z}}\right) \parallel \mathcal{U}\left(\mathcal{P}^n_{R,\mathbf{v},\mathbb{Z}}\right) \right] = \left(\frac{R}{r}\right)^n \cdot \frac{1+\varepsilon_{n,R}}{1+\varepsilon_{n,r}} = M',$$

and the two algorithms $\mathcal{A}$ and $\mathcal{F}$ below have indistinguishable output distributions.

$\mathcal{A}$	$\mathcal{F}$
$\mathbf{v} \hookleftarrow h$	$\mathbf{v} \hookleftarrow h$
$\mathbf{z} \hookleftarrow \rho^n_{R,\mathbf{v}}$	$\mathbf{z} \hookleftarrow \rho^n_r$
output $(\mathbf{z}, \mathbf{v})$ *if* $\mathbf{z} \in \mathcal{P}^n_{r,\mathbb{Z}}$, *else* $\perp$	*output* $(\mathbf{z}, \mathbf{v})$ *with probability* $1/M'$, *else* $\perp$

Furthermore, $\mathcal{A}$ outputs $(\mathbf{z}, \mathbf{v})$ with probability $1/M'$.

Theorem description. This theorem informally provides a lot of complementary information. First, it allows to build a procedure such that algorithms $\mathcal{A}$ and $\mathcal{F}$ are indistinguishable. This crucial part ensures the zero-knowledge property induced by this rejection sampling. It implies that after executing $\mathcal{A}$ then the output distribution induced by $\mathbf{z}$ is independent of $\mathbf{v}$. In practice $\mathbf{v}$ depends on a secret and therefore algorithm $\mathcal{A}$ must hide $\mathbf{v}$ and its associated secret from the output distribution induced by $\mathbf{z}$. Now, by itself this result is a pretty standard one. In addition our theorem proves two other things. First, contrarily to the state-of-the-art, it uses generic polytopes (and uniform distributions inside these polytopes), making it the first generalisation of the rejection sampling theorem to this setting. Then, in most cases the number of integral points inside a polytope is hard to manipulate because its closed form can be complicated. Consequently, instead of directly proving the rejection sampling over integral sets, we first work with volumes to compute the Rényi divergence. Then we tweak this result appropriately in order to obtain the necessary result over the restriction to integers of the volume as well as being able to compute the appropriate values of r_M and R. Last but not least, behind it hides some nice properties such as minimisation of the rejection rate. Alternatively, in most practical cases the rejection rate is fixed beforehand and thus this theorem minimises the size of r_M. These two minimisations are dual depending on which parameter is fixed.

At this point, we still have no improvement to resolve the five challenges defined above. However, the above theorem allows to study a wider range of constructions. Currently, in lattice-based cryptography the only distributions used are: Gaussian distributions, uniform distributions over hypercube and newly over hyperball and to a lesser extent bimodal distributions. This phenomenon is not due to the lack of practical distributions but the lack of a dedicated framework. Now that we have given one, we propose a new construction defined by the polytope $\mathcal{H}_r^n = \mathcal{B}_\infty^n(r) \cap \mathcal{B}_1^n(\sqrt{n}r)$.

By definition, test of membership is trivial as it involves computing a maximum and a sum. Additionally, most of the work involved for restriction to $\mathbb{Z}^n$ and abort minimisation has been handled through the previous theorem. More details are provided in Section 3.

The last two challenges remain: having an efficient sampler and accurately approximating this hyperball *i.e* for $\mathcal{H}_r^n$ having a circumradius close to r. We provide a positive answer to both challenges. First, we prove that all the vertices of $\mathcal{H}_r^n$ are at distance $r\sqrt[4]{n}$ of its centre, which implies that the circumradius is exactly $r\sqrt[4]{n}$. To compare with the hypercube, many lattice-based applications use $n \approx 1024$. As such we diminish the appropriate radius from $33r$ with the hypercube to less than $6r$ with $\mathcal{H}_r^n$.

We propose an isochronous uniform sampler on $\mathcal{H}$. Informally, an isochronous sampler is a sampler whose running time is independent of its sensitive inputs and outputs (see [HPRR20] for details). To achieve it we use two main tricks linked to the L_1 ball. First, we use the fact that the volume of an L_1 ball is mostly concentrated inside this inscribed $\mathcal{H}$. Furthermore this result can be extended

to restrictions to integral points, meaning that most points inside a L_1 ball are inside $\mathcal{H}_{\mathbb{Z}}$. Therefore, at the cost of a negligible additional amount of rejection, one can directly sample inside the integral points of an L_1 ball. However, this still remains a challenge as the quest for efficiency and practicality reduces the number of possible approaches. To circumvent this, we use some link between an L_1 ball and L_1 sphere. By denoting $\mathcal{B}_1^n(r)$ the L_1 ball of radius r in dimension n and $\mathcal{S}_1^n(r)$ its surface then there is a bijection between any 1-dimensional projection on the canonical basis of $\mathcal{S}_1^{n+1}(r)$ with $\mathcal{B}_1^n(r)$. Since sampling on the L_1 sphere is well-known and doable using solely basic uniform sampling, we build and prove a variant that respects the isochronous property in order to sample directly inside $\mathcal{H}_{\mathbb{Z}}$ at the cost of a negligible amount of restarts.

Experimentally it appears that $\mathcal{H}$ can be made even more compact. The first intuition can be given when computing a θ_n such that $\mathsf{Vol}\left(\mathcal{H}_r^n\right) = \mathsf{Vol}\left(\mathcal{B}_2^n(\theta_n \cdot r)\right)$. For large enough n we obtain $\theta_n \approx \sqrt{\frac{2e}{\pi}} \approx 1.315$, which is a small radius. However, when sampling in $\mathcal{H}_{r,\mathbb{Z}}^n$ the number of vectors that have Euclidean norm bigger than $\theta_n \cdot r$ is non-negligible which leads to a high rejection rate. By carefully increasing a parameter θ from θ_n upwards, it appears that after $\theta = 1.5$, most vectors from $\mathcal{H}_{r,\mathbb{Z}}^n$ have Euclidean norm less than $\theta \cdot r$. This leads to some nice improvements in the approximation of the hyperball. More formally this boils down to directly working on $\mathcal{H}_{r,\mathbb{Z}}^n \cap \mathcal{B}_{2,\mathbb{Z}}^n(\theta \cdot r)$. As this does not define a polytope, only some results from the overall framework carry over, and others remain to be adapted.

We wrap all our contributions into one possible application of FSwA by giving birth to a new signature scheme: Patronus. We prove its basic properties such as correctness and security properties such as UF-CMA in the QROM using [KLS18, BBD+23, DFPS23]. This signature has non-negligibly shorter signatures compared to Dilithium [DKL+21] (around 25% shorter) while providing shorter public keys (around 13%). Regarding sampling, since it is based on uniform distributions in integral intervals and not fixed-point Gaussian distributions, firstly it uses less randomness and secondly it should be much easier to protect against side-channel attacks than Haetae [CCD+23]. Given classical bit security targets 120, 180, and 260, we show in Table 1 a comparison between our public elements sizes with Dilithium and Haetae. Dilithium-G [DLL+17] is excluded from the comparison because its use of the infinity-norm with Gaussian distributions results in a non-negligible security loss, leading to a misleading comparison.

Overview. In Section 2 we introduce necessary preliminaries, including notations and lemmas for the following parts. We then present our generic framework on polytopes in Section 3. We complement it in Section 4 by a comprehensive study of $\mathcal{H}$. Lastly, in Section 5 we present experimental upgrades to $\mathcal{H}$ which further improve the proof of knowledge size, and conclude Section 5 by introducing a signature scheme based on this whole study of FSwA.

Edit. Addition of Corollary 1 to properly apply our main polytope theorem to Fiat-Shamir with Aborts in practice (considering the support's shape of $c\mathbf{s}$).

Table 1: Comparison of both the signature and verification key sizes in bytes for Dilithium, Haetae, and Patronus, where "II", "III" and "V" represent respectively 120, 180 and 260 bits of classical security.

(a) Signature size comparison.

	II	III	V
Haetae	1,463	2,337	2,908
Patronus (this work)	2,070	2,575	3,721
Dilithium	2,420	3,293	4,595

(b) Verification key size comparison.

	II	III	V
Haetae	992	1,472	2,080
Patronus (this work)	832	1,152	1,632
Dilithium	1,312	1,952	2,592

In Section 5.2, the signature security proof is now based on the infinity norm. As a brief explanation, the bits-cut engaged in the security proof lies within a hypercube. It leads to the equivalence factor $\sqrt{n}$ when transposing the bound from the initial L_∞ norm to the L_2 norm in our initial study implying a direct security estimation loss. Directly basing security on MSIS using L_∞ leads to slightly larger signatures however the public keys and the number of restarts are drastically improved.

2 Preliminaries

The non-negative integers, integers and reals are respectively denoted $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{R}$. Matrices are written as bold capital letters and vectors as low-case bold letters. Vectors should be understood as column vectors. Unless otherwise specified, n will denote the dimension of the ambient space and integral will mean $\subset \mathbb{Z}^n$.

The coordinates of a vector $\mathbf{x} \in \mathbb{R}^n$ will be written $\mathbf{x} = (x_1, \cdots, x_n)$. For $a, b \in \mathbb{R}$, we define $[\![a, b]\!] = [a, b] \cap \mathbb{Z}$ and $]\!]a, b]\!] = (a, b] \cap \mathbb{Z}$. For a predicate P, we write $[\![P]\!] = 1$ if P is true and 0 otherwise.

For $\mathcal{D}$ a distribution, we define $\mathbf{z} \leftarrow\!\!\$\ \mathcal{D}$ as $\mathbf{z}$ sampled according to distribution $\mathcal{D}$. In case $\mathcal{D}$ is not a distribution but a set, we use the convention that $\mathbf{z} \leftarrow\!\!\$\ \mathcal{D}$ means uniformly sampling $\mathbf{z}$ inside $\mathcal{D}$.

Operations on sets. Given a set $\mathcal{P} \subset \mathbb{R}^n$ we note $\mathsf{Vol}(\mathcal{P}) \in \mathbb{R}_{\geq 0} \cup \{\infty\}$ its volume when appropriate, and $|\mathcal{P}| \in \mathbb{N} \cup \{\infty\}$ its cardinality. For $X \subset \mathbb{R}^n$, $X_\mathbb{Z} = X \cap \mathbb{Z}^n$ denotes its restriction to the integers and $\mathsf{conv}(X)$ denotes its convex hull, the smallest convex region containing X, which we define as:

$$\mathsf{conv}(X) := \left\{ \sum_{\mathbf{x} \in X} \lambda_\mathbf{x} \mathbf{x} \;\middle|\; \begin{array}{c} \sum_{\mathbf{x} \in X} \lambda_\mathbf{x} = 1, \forall \mathbf{x} \in X, \lambda_\mathbf{x} \in [0, 1], \\ \text{and only a finitely many } \lambda_\mathbf{x} \text{ are non-zero.} \end{array} \right\}.$$

We specifically focus our study on convex spaces and in particular to their volumes or to the cardinal of their restriction to integers. For this we use a common result on the volume of a convex space homothety.

Lemma 1. *Let $S \subset \mathbb{R}^n$ be a measurable set, and let $r \geq 0$. Recall $S_r = \{r\mathbf{s} : \mathbf{s} \in S\}$. Then, we have $\mathsf{Vol}(S_r) = r^n \cdot \mathsf{Vol}(S)$.*

L_p norms, balls and spheres. For $\mathbf{x} \in \mathbb{R}^n$ and $p \in \mathbb{R}_{>0} \cup \{\infty\}$, we denote by $\|\mathbf{x}\|_p$ its L_p norm. The p-ball (*resp.* p-sphere) of radius r centred at $\mathbf{c}$ in dimension n is denoted $\mathcal{B}_p^n(r, \mathbf{c})$ (*resp.* $\mathcal{S}_p^n(r, \mathbf{c})$) or $\mathcal{B}_p^n(r)$ (*resp.* $\mathcal{S}_p^n(r)$) for $\mathbf{c} = \mathbf{0}$. In Table 2 we provide a brief summary of closed forms for the volume of common L_p balls as well as their number of integral points. [1]

Table 2: Volume and cardinality of L_1, L_2 and L_∞ balls.

	$\mathcal{B}_1^n(r)$	$\mathcal{B}_2^n(r)$	$\mathcal{B}_{\infty,\mathbb{Z}}^n(r)$
$\mathsf{Vol}(\mathcal{B})$	$\frac{(2r)^n}{n!}$	$\frac{\pi^{\frac{n}{2}} r^n}{\Gamma(1+\frac{n}{2})}$	$(2r)^n$
$\mathsf{Vol}(\mathcal{B}_\mathbb{Z})$	$\sum_{i=0}^{\min(r,n)} \binom{n}{i}\binom{r}{i} 2^i$	$\mathsf{Vol}(\mathcal{B}_2^n(r))$	$(2r+1)^n$

Definition 1 (Polytope). *Let n and v be integers, and let $(\mathbf{x}_i)_{1\leq i\leq v} \in (\mathbb{R}^n)^v$ be a family of vectors. A subset $\mathcal{P} \subset \mathbb{R}^n$ is a polytope in dimension n with v vertices $(\mathbf{x}_i)_{1\leq i\leq v}$ if $\mathcal{P}$ is the convex hull of $(\mathbf{x}_i)_{1\leq i\leq v}$ and if no strict sub-family of $(\mathbf{x}_i)_{1\leq i\leq v}$ has a convex hull equal to $\mathcal{P}$. If in addition, the linear span of the vertices is $\mathbb{R}^n$, the polytope is referred to as full-rank. Lastly, a polytope $\mathcal{P}$ with integral vertices is said* integral.

Unless stated otherwise, in the rest of this paper, by polytope we mean full-rank polytope. In Proposition 2, we show that the vertices of a polytope are unique up to ordering, and we write $\mathcal{V}(\mathcal{P}) := \{(\mathbf{x}_i)_{1\leq i\leq v}\}$ for the set of vertices of $\mathcal{P}$.

Definition 2 (Translation and dilation of polytopes). *For a polytope (or any subset of $\mathbb{R}^n$) $\mathcal{P} \subseteq \mathbb{R}^n$, a centre $\mathbf{c} \in \mathbb{R}^n$, and a dilation factor $r \in \mathbb{R}$, we define $\mathcal{P}_{r,\mathbf{c}} := \{r\mathbf{x} + \mathbf{c} : \mathbf{x} \in \mathcal{P}\}$. We will omit r if $r = 1$ and $\mathbf{c}$ if $\mathbf{c} = 0$. It follows that $\mathcal{V}(\mathcal{P}_{r,\mathbf{c}}) = \{r\mathbf{x} + \mathbf{c} : \mathbf{x} \in \mathcal{V}(\mathcal{P})\}$.*

Definition 3 (Symmetric and inscribed polytopes). *A full-rank polytope $\mathcal{P}$ is* symmetric *if $\mathcal{P} = \mathcal{P}_{-1}$ (or equivalently if $\mathcal{V}(\mathcal{P}) = -\mathcal{V}(\mathcal{P})$). A full-rank polytope $\mathcal{P}$ is an* inscribed *polytope if for all $\mathbf{x}, \mathbf{y} \in \mathcal{V}(P)$, $\|\mathbf{x}\|_2 = \|\mathbf{y}\|_2$. The radius of an inscribed polytope is the L_2 norm of its vertices.*

The two following Propositions can be found in [Brø83].

[1] Precise estimates are notoriously difficult to obtain in general. For r large enough we refer to the estimates of [Ste17].

Proposition 1 (Intersection of polytopes [Brø83, (Section 1)]). *The intersection of two (full-rank) polytopes is a (not always full-rank) polytope. If the intersection contains a non-trivial open ball, then it is also full-rank.*

Proposition 2 (Polytope vertices characterisation [Brø83, (Theorem 7.2)]). *Let $\mathcal{P}$ be a polytope. Then, $\mathbf{x}$ is not a vertex of $\mathcal{P}$ if and only if there exist vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathcal{P}$ such that $\mathbf{x} \in (\mathbf{a}, \mathbf{b})$, where*

$$(\mathbf{a}, \mathbf{b}) = [\mathbf{a}, \mathbf{b}] - \{\mathbf{a}, \mathbf{b}\} = \{t\mathbf{a} + (1-t)\mathbf{b} : t \in [0,1]\} - \{\mathbf{a}, \mathbf{b}\}.$$

Rényi divergence. We present the Rényi divergence, while it is defined for any order between, we only focus on the relevant Rényi with $a = \infty$.

Definition 4 (Rényi divergence). *Let $\mathcal{P}, \mathcal{Q}$ be two distributions such that $\mathsf{Supp}(\mathcal{P}) \subseteq \mathsf{Supp}(\mathcal{Q})$. The Rényi divergence of order ∞ is defined as follows:*

$$R_\infty(\mathcal{P} \parallel \mathcal{Q}) = \max_{x \in \mathsf{Supp}(\mathcal{P})} \frac{\mathcal{P}(x)}{\mathcal{Q}(x)}.$$

Remark 1. For $a = \infty$, the Rényi divergence between two uniform distributions on measurable sets X_s and X_t, $X_t \subseteq X_s$, with nonzero volume (*resp.* on finite sets) is exactly the ratio of their volumes (*resp.* cardinalities).

Rejection sampling. Rejection sampling is a technique used to generate samples from a target distribution D_t based on samples from a source distribution D_s under the condition that the support of D_t is (almost) contained within the support of D_s.

Lemma 2 (Rejection sampling (from [Lyu12, (Lemma 4.7)])). *Let D_s be a source distribution and D_t be a target distribution with $\mathsf{Supp}(D_t) \subseteq \mathsf{Supp}(D_s)$. If there exists $M > 1$ such that $\mathcal{R}_\infty[\, D_t \parallel D_s \,] \leq M$ then the output distribution of the following algorithm $\mathcal{A}$ is statistically equivalent to the output distribution of the following algorithm $\mathcal{F}$.*

$\mathcal{A}$

1 : $\mathbf{z} \hookleftarrow D_s$
2 : *with probability* $\min\left(\frac{D_t(\mathbf{z})}{M \cdot D_s(\mathbf{z})}, 1\right)$:
3 : **return z**

$\mathcal{F}$

1 : $\mathbf{z} \hookleftarrow D_t$
2 : *with probability* $1/M$:
3 : **return z**

Notably, $\mathcal{A}$ outputs $\mathbf{z}$ with probability $\frac{1}{M}$.

Modular arithmetic. For any $p \in \mathbb{Z}_{>0}$ and $x \in \mathbb{Z}_p := \mathbb{Z}/p\mathbb{Z}$, we write $x \bmod p$ the unique representative in $[\![0, p[\![$. For any even (*resp.* odd) $p \in \mathbb{Z}_{>0}$ and any $x \in \mathbb{Z}_p$, we will denote by $x \bmod^{\pm} p$ the unique representative in $]\!] -p/2, p/2]\!]$ (*resp.* $[\![-(p-1)/2, (p-1)/2]\!]$).

We extend this definition to vectors entrywise. For $x \in \mathbb{Z}_p$, we define $|x| := \left|x \bmod^{\pm} p\right|$. For any $p, n \in \mathbb{Z}_{>0}$ and $\mathbf{x} \in \mathbb{Z}_p^n$, we define $\|\mathbf{x}\|_p$ as the L_p norm of $|\mathbf{x}| \in \mathbb{R}^n$, where $|\cdot|$ is taken componentwise.

The ring $\mathcal{R}_q$. We will work in $\mathcal{R} = \mathbb{Z}[X]/(X^n+1)$ and $\mathcal{R}_q = \mathbb{Z}_q[X]/(X^n+1)$ for q a prime and n a power of two that will be clear from context.

We extend the definition of L_p norms to $\mathcal{R}_q$ by identifying each $\mathbf{x} \in \mathcal{R}_q$ (seen as a polynomial of degree less than n) with the vector $\mathbf{x} \in \mathbb{Z}_q^n$ of its coefficients. We will consider $\mathcal{R}_2$ the subset of $\mathcal{R}$ and $\mathcal{R}_q$ of binary polynomials, with coefficients 0 or 1.

Support of a probability distribution. For $\mathcal{D}$ a probability distribution with values in a set X, we define $\mathsf{Supp}(\mathcal{D}) = \{x \in X : \Pr[\mathcal{D} = x] \neq 0\}$.

3 Rejection sampling on polytopes

The authors of [DFPS22] formally prove and characterise optimal distributions in the continuous setting for zero-knowledge proofs using the Fiat-Shamir with Aborts (FSwA) paradigm. They define optimality through two notions: minimal rejection rates, and compact distributions in the proof of knowledge. In practice, this equates to sandwiching the support S of a uniform distribution between its inscribed and circumscribed spheres as tightly as possible. However, these distributions studies are done in the continuous setting instead of doing in their restriction to $\mathbb{Z}^n$ as it is done in practice.

On the broader front, uniform distributions appear to allow for overall simpler sampling algorithms. This, however, is not always true as uniform distribution does not rhyme with uniform sampling. In the case of the hyperball, its state-of-the-art uniform sampler [CCD+23] uses Gaussian samplers. That is why we focus our study on uniform distributions over structured convex supports, *i.e.* polytopes. We propose a different framework for FSwA that relies specifically on those uniform distributions. We thereby translate the framework of distribution indistinguishability through its implications to intersections of volumes over which probability densities are taken to be homogeneous.

In this section we propose a study in the continuous and more importantly in the discrete setting (restricting to $\mathbb{Z}^n$) that minimises the rejection rate given a target uniform distribution over a polytope. For this, we first prove an important characterisation of a special intersection of polytopes, and then we share our main general theorem on rejection sampling using discrete uniform distributions. Throughout this section, $n \in \mathbb{N}$ refers to the dimension of our ambient space.

3.1 Intersection of polytopes

In this section we study intersections of polytopes that will help us find the largest volume that hides the secret. Proposition 3 is the main tool for our general result on polytope rejection sampling. We prove it through a couple of lemmas.

Lemma 3. *For any symmetric convex $\mathcal{S}$, let $r, R \in \mathbb{R}$ be two radii such that $R > r > 0$. Then:*

$$\mathcal{S}_{R-r} = \bigcap_{\mathbf{c} \in \mathcal{S}_r} \mathcal{S}_{R,\mathbf{c}}.$$

Proof. We first prove the direct inclusion. For this, given $\mathbf{x} \in \mathcal{S}_{R-r}$, and $\mathbf{c} \in \mathcal{S}_r$. There exists $\mathbf{y}, \mathbf{v} \in \mathcal{S}$ such that $\mathbf{x} = (R-r)\mathbf{y}$ and $\mathbf{c} = r\mathbf{v}$. Since $\mathcal{S}$ is symmetric there exists $\mathbf{c}^* = -\mathbf{c} \in \mathcal{S}_r$ such that $\mathbf{x} - \mathbf{c} = \mathbf{x} + \mathbf{c}^*$. By convexity, $\mathbf{x} + \mathbf{c}^* = R\mathbf{y} \cdot (1 - \frac{r}{R}) + R\mathbf{v} \cdot \frac{r}{R}$ which implies $\mathbf{x} - \mathbf{c} \in \mathcal{S}_R$.

We prove the reverse inclusion by contraposition. Namely, any vector not in $\mathcal{S}_{R-r}$ is not in $\bigcap_{\mathbf{c}\in\mathcal{S}_r} \mathcal{S}_{R,\mathbf{c}}$. Let $\mathbf{z} \notin \mathcal{S}_{R-r}$. There exists unique $\varepsilon > 0$ and $\mathbf{x} \in \mathcal{S}$ of maximal L_2 norm such that $\mathbf{z} = (R - r + \varepsilon)\mathbf{x}$. By taking $\mathbf{c} = -r\mathbf{x}$ which is in $\mathcal{S}_r$ by symmetry of $\mathcal{S}$, we obtain $\mathbf{z} - \mathbf{c} = (R + \varepsilon)\mathbf{x}$. By maximality of $\mathbf{x}$, $\mathbf{z} - \mathbf{c} \notin \mathcal{S}_R$, and this concludes our proof.

Lemma 3 is the starting point of the whole section. The focus of this subsection is to extend it to integral restrictions.

Lemma 4. *Let $\mathcal{P}$ be a convex region, and $\mathbf{a} \in \mathbb{R}^n$ a vector. Then we have $\mathcal{P} \cap (\mathcal{P} + \mathbf{a}) = \bigcap_{t\in[0,1]} (\mathcal{P} + t\mathbf{a})$. As a consequence, for any convex region $\mathcal{P}$ and polytope $\mathcal{Q}$:*

$$\bigcap_{\mathbf{c}\in\mathcal{Q}} (\mathcal{P} + \mathbf{c}) = \bigcap_{\mathbf{c}\in\delta\mathcal{Q}} (\mathcal{P} + \mathbf{c}) = \bigcap_{\mathbf{c}\in\mathcal{V}(\mathcal{Q})} (\mathcal{P} + \mathbf{c}).$$

Proof. We start with the first equation, ie $\mathcal{P}$ is convex. Let $\mathbf{x} \in \mathcal{P} \cap (\mathcal{P} + \mathbf{a})$. Then there exists $\mathbf{y} \in \mathcal{P}$ such that $\mathbf{x} = \mathbf{y} + \mathbf{a}$. Let $t \in [0, 1]$. Then $\mathbf{x} = (\mathbf{y} + (1-t)\mathbf{a}) + t\mathbf{a}$, where $\mathbf{y} + (1 - t)\mathbf{a} = (1 - t)\mathbf{x} + t\mathbf{y}$ and therefore lives in $\mathcal{P}$ by convexity. Thus $\mathbf{x} \in \mathcal{P} + t\mathbf{a}$ and we have proved the direct inclusion. The reverse inclusion is trivial. We now establish the second statement by induction on the number of vertices of $\mathcal{Q}$. The case $|\mathcal{V}(\mathcal{Q})| = 2$ has been dealt with. Now let $\mathcal{P}$ be a convex region and $\mathcal{Q}$ a polytope with $m + 1$ vertices $\mathbf{a}_1, \ldots, \mathbf{a}_{m+1}$, and suppose the result holds for convex hulls with fewer vertices. Let $\mathbf{c} = \sum_{i=1}^{m+1} t_i\mathbf{a}_i \in \mathcal{Q}$, then $\mathbf{c} = (1 - t_{m+1})\mathbf{c}' + t_{m+1}\mathbf{a}_{m+1}$, where $\mathbf{c}' \in \mathcal{Q}'$ and $\mathcal{V}(\mathcal{Q}') = (\mathbf{a}_i)_{i\leq m}$. Reciprocally, any point of $\mathbf{c}' \in \mathcal{Q}'$ gives a segment $[\mathbf{c}', \mathbf{a}_{m+1}] \subseteq \mathcal{Q}$. Now using the first identity multiple times,

$$\begin{aligned}\bigcap_{\mathbf{c}\in\mathcal{Q}} (\mathcal{P} + \mathbf{c}) &= \bigcap_{\mathbf{c}'\in\mathcal{Q}'} \bigcap_{t\in[0,1]} (\mathcal{P} + (1 - t)\mathbf{c}' + t\mathbf{a}_{m+1}) = \bigcap_{\mathbf{c}'\in\mathcal{Q}'} (\mathcal{P} + \mathbf{c}') \cap (\mathcal{P} + \mathbf{a}_{m+1}) \\ &= (\mathcal{P} + \mathbf{a}_{m+1}) \cap \bigcap_{\mathbf{c}'\in\mathcal{Q}'} (\mathcal{P} + \mathbf{c}').\end{aligned}$$

We conclude by our induction hypothesis. Note that the vertices are contained in the boundary so we don't bother with the middle term of the last statement. □

The following result remains true for integral polytopes.

Proposition 3 ($\mathcal{P}$-ception: Intersection of polytopes). *Let $\mathcal{P}$ be a symmetric inscribed polytope. Let $r, R \in \mathbb{R}$ such that $R > r > 0$. Then:*

$$\bigcap_{\mathbf{c}\in\mathcal{P}_r} \mathcal{P}_{R,\mathbf{c}} = \bigcap_{\mathbf{c}\in\delta\mathcal{P}_r} \mathcal{P}_{R,\mathbf{c}} = \bigcap_{\mathbf{c}\in\mathcal{V}(\mathcal{P}_r)} \mathcal{P}_{R,\mathbf{c}} = \mathcal{P}_{R-r}.$$

In particular, if $\mathcal{V}(\mathcal{P}_r) \subset \mathbb{Z}^n$, $\bigcap_{\mathbf{c}\in\mathcal{P}_{r,\mathbb{Z}}} \mathcal{P}_{R,\mathbf{c},\mathbb{Z}} = \mathcal{P}_{R-r,\mathbb{Z}}$.

Proof. The proof follows immediately from the statements of Lemma 3 and Lemma 4. □

However, practically the set of points defining the translation of the main polytope lies within a hyperball. As such, we need a stronger result to model truthfully $\mathcal{P}$-ception for practical applications. For this, we complement Proposition 3 to achieve an equivalent result using facets instead of vertices so that our previous result remains true on perfectly circumscriptible polytopes.

Corollary 1. *Let $\mathcal{P}$ be a convex region, $\mathcal{Q}$ a polytope, then for any function $h : \mathcal{F}(\mathcal{Q}) \to \mathbb{R}^n$ such that $\mathbf{f} \overset{h}{\mapsto} \mathbf{h_f} \in \mathcal{F}(\mathcal{Q})$:*

$$\bigcap_{\mathbf{f}\in\mathcal{F}(\mathcal{Q})} (\mathcal{P} + \mathbf{h_f}) = \bigcap_{\mathbf{v}\in\mathcal{V}(\mathcal{Q})} (\mathcal{P} + \mathbf{v}).$$

Proof. Let $h : \mathcal{F}(\mathcal{Q}) \to \mathbb{R}^n$, for any point $\mathcal{F}$ in $\mathcal{F}(\mathcal{Q})$ then $\mathbf{h_f}$ is in a facet of $\mathcal{Q}$ defined by k different vertices $(\mathbf{c}_i)_{i\leq k} \subset \mathcal{V}(\mathcal{Q})$. In other words, by convexity there exist $(\lambda_i)_{i\leq k} \subset [0,1]^k$ such that $\mathbf{h_f} = \sum_{i=1}^{k} \lambda_i \mathbf{c}_i$. Then using the first equality of Lemma 3, we obtain the desired result:

$$\begin{aligned}\bigcap_{\mathbf{f}\in\mathcal{F}(\mathcal{Q})} (\mathcal{P} + \mathbf{h_f}) &= \bigcap_{\mathbf{f}\in\mathcal{F}(\mathcal{Q})} (\mathcal{P} + \sum_i \lambda_i \mathbf{c}_i) \\ &= \bigcap_{\mathbf{c}\in\mathcal{V}(\mathcal{Q})} \mathcal{P} + \mathbf{c}.\end{aligned}$$

3.2 Rejection sampling in the FSwA paradigm

Proposition 3 is a very useful tool for the study of rejection sampling, most notably in the FSwA paradigm. Informally, it gives the exact characterisation of the largest set of points that does not leak any information on the secret engaged in the zero-knowledge proof. In this subsection, we propose a formalisation of this idea using our framework.

This implies computing a Rényi divergence between uniform distributions over a polytope. In practice, estimating the volume of a generic polytope can be a delicate task, and precisely counting its integral points might be almost impossible. Luckily, it is proven in [DF88] that it can be done in polynomial time on integral polytopes with algorithms such as [CCF22]. We propose a step-by-step approach that provides a path from the continuous setting to the desired theorem on rejection sampling in the discrete setting. We use an ε-approximation of the ratio between the cardinal of a set discretization and its volume. Using lemma from Appendix Section B.3, ε should be small for a cryptographic instance.

Lemma 5. *Let $\mathcal{P}$ be a symmetric inscribed polytope, $\mathbf{v} \in \mathbb{R}^n$ and let $\beta, r, R > 0$ such that $R \geq r + \beta$, and $\mathbf{v} \in \mathcal{P}_\beta^n$. We have:*

$$\mathcal{R}_\infty \left[\mathcal{U}\left(\mathcal{P}_r^n\right) \,\|\, \mathcal{U}\left(\mathcal{P}_{R,\mathbf{v}}^n\right) \right] = \left(\frac{R}{r}\right)^n.$$

In particular if $\beta = \|\mathbf{v}\|_2$ *and* $R = r + \beta$*, then for any* $M > 1$*, the inequality* $\left(\frac{R}{r}\right)^n \leq M$ *holds if and only if* r *satisfies the condition* $r \geq \frac{\|\mathbf{v}\|_2}{M^{\frac{1}{n}} - 1}$.

Proof. Let $\mathcal{P}$ be a symmetric inscribed polytope and $\mathbf{v} \in \mathcal{P}_\beta^n$, by applying Proposition 3 we have: $\mathcal{P}_{R-\beta}^n = \bigcap_{\mathbf{c} \in \mathcal{P}_\beta^n} \mathcal{P}_{R,\mathbf{c}}^n \subset \mathcal{P}_{R,\mathbf{v}}^n$. With $r = R - \beta$ the Rényi divergence is well-defined. The desired Rényi divergence is then obtained through a direct application of Lemma 1. Finally, by fixing $R = r + \beta$ we can derive the following equivalences:

$$\left(\frac{R}{r}\right)^n \leq M \quad \Leftrightarrow \quad r + \beta \leq r \cdot M^{\frac{1}{n}} \quad \Leftrightarrow \quad r \geq \frac{\beta}{M^{\frac{1}{n}} - 1}.$$

□

Proposition 4. *Let* $\mathcal{P}$ *be a symmetric inscribed polytope,* $\mathbf{v} \in \mathbb{R}^n$ *and let* $M > 1$*, and* $\mathbf{v} \in \mathcal{P}_{\beta,\mathbb{Z}}^n$*. For any* $s \in \mathbb{R}$*, define* $\varepsilon_{n,s} \in \mathbb{R}$ *by* $\varepsilon_{n,s} = \left|\mathcal{P}_{s,\mathbb{Z}}^n\right| / \mathsf{Vol}\left(\mathcal{P}_s^n\right) - 1$*. If* $\beta = \|\mathbf{v}\|_2$*,* $r \geq \frac{\beta}{M^{\frac{1}{n}} - 1}$*,* $R = r + \beta$*,* $\mathcal{V}\left(\mathcal{P}_\beta^n\right) \subseteq \mathbb{Z}^n$ *and* $M' = \frac{1+\varepsilon_{n,r}}{1+\varepsilon_{n,R}} \cdot M$ *then:*

$$\mathcal{R}_\infty\left[\,\mathcal{U}\left(\mathcal{P}_{r,\mathbb{Z}}^n\right) \,\|\, \mathcal{U}\left(\mathcal{P}_{R,\mathbf{v},\mathbb{Z}}^n\right)\,\right] = \left(\frac{R}{r}\right)^n \cdot \frac{1+\varepsilon_{n,R}}{1+\varepsilon_{n,r}} \leq M'.$$

Proof. As $\mathcal{P}$ is a symmetric inscribed polytope, $\mathbf{v} \in \mathcal{P}_\beta^n$ and $\mathcal{P}_\beta^n$ has integral vertices, we can apply Proposition 3 to obtain $\mathcal{P}_{R-\beta,\mathbb{Z}}^n = \bigcap_{\mathbf{c} \in \mathcal{P}_{\beta,\mathbb{Z}}^n} \mathcal{P}_{R,\mathbf{c},\mathbb{Z}}^n \subset \mathcal{P}_{R,\mathbf{v},\mathbb{Z}}^n$. Thus the Rényi divergence in the statement is well-defined. It is then obtained by simply rewriting:

$$\mathcal{R}_\infty\left[\,\mathcal{U}\left(\mathcal{P}_{r,\mathbb{Z}}^n\right) \,\|\, \mathcal{U}\left(\mathcal{P}_{R,\mathbf{v},\mathbb{Z}}^n\right)\,\right] = \frac{\left|\mathcal{P}_{R,\mathbb{Z}}^n\right|}{\left|\mathcal{P}_{r,\mathbb{Z}}^n\right|} = \frac{\mathsf{Vol}\left(\mathcal{P}_R^n\right)}{\mathsf{Vol}\left(\mathcal{P}_r^n\right)} \cdot \frac{\left|\mathcal{P}_{R,\mathbb{Z}}^n\right|}{\mathsf{Vol}\left(\mathcal{P}_R^n\right)} \cdot \frac{\mathsf{Vol}\left(\mathcal{P}_r^n\right)}{\left|\mathcal{P}_{r,\mathbb{Z}}^n\right|}.$$

The rest of the proof follows the one of Lemma 5. □

Proposition 3 contributes in two major ways to Proposition 4. First, it proves the existence of our Rényi divergence. Second, it minimises M for fixed r and R. Alternatively, for fixed M, it reduces the distance between r and R allowing for shorter choices of polytope circumradii in practice. This leads to our main generic theorem enabling FSwA rejection sampling using uniform distributions over polytopes:

Theorem 1 (Rejection sampling for $\mathcal{U}\left(\mathcal{P}_\mathbb{Z}^n\right)$). *Let* $M > 1$*,* $\beta > 0$ *and* h *a probability distribution such that* $\mathsf{Supp}\left(h\right) \subset \mathcal{P}_{\beta,\mathbb{Z}}^n$*. Let* $r \geq \frac{\beta}{M^{\frac{1}{n}} - 1}$*,* $R \geq r + \beta$*, and define* $\varepsilon_{n,s} = \left|\mathcal{P}_{s,\mathbb{Z}}^n\right| / \mathsf{Vol}\left(\mathcal{P}_s^n\right) - 1$ *for* $s \in \mathbb{R}$ *and* $M' = \frac{1+\varepsilon_{n,R}}{1+\varepsilon_{n,r}} \cdot M$*. Let* $\mathbf{v} \in \mathcal{P}_\beta^n$ *and define* $\rho_{r,\mathbf{v}}^n := \mathcal{U}\left(\mathcal{P}_{r,\mathbf{v},\mathbb{Z}}^n\right)$*. If* $M' > 1$ *then:*

$$\mathcal{R}_\infty\left[\,\mathcal{U}\left(\mathcal{P}_{r,\mathbb{Z}}^n\right) \,\|\, \mathcal{U}\left(\mathcal{P}_{R,\mathbf{v},\mathbb{Z}}^n\right)\,\right] = \left(\frac{R}{r}\right)^n \cdot \frac{1+\varepsilon_{n,R}}{1+\varepsilon_{n,r}} = M',$$

and the two algorithms $\mathcal{A}$ *and* $\mathcal{F}$ *below have indistinguishable output distributions:*

$\mathcal{A}$	$\mathcal{F}$
$\mathbf{v} \hookleftarrow\$ h$	$\mathbf{v} \hookleftarrow\$ h$
$\mathbf{z} \hookleftarrow\$ \rho^n_{R,\mathbf{v}}$	$\mathbf{z} \hookleftarrow\$ \rho^n_r$
output $(\mathbf{z}, \mathbf{v})$ *if* $\mathbf{z} \in \mathcal{P}^n_{r,\mathbb{Z}}$, *else* $\perp$	*output* $(\mathbf{z}, \mathbf{v})$ *with probability* $1/M'$, *else* $\perp$

Furthermore, $\mathcal{A}$ *outputs* $(\mathbf{z}, \mathbf{v})$ *with probability* $1/M'$.

Proof. Apart from its prerequisites, this proof follows the common approach of the Rényi divergence applied to rejection sampling. We start by proving that, given $\mathbf{z} \in \mathcal{P}^n_{R,\mathbf{v},\mathbb{Z}}$, then $\min\left(\frac{\rho^n_r(\mathbf{z})}{M' \cdot \rho^n_{R,\mathbf{v}}(\mathbf{z})}, 1\right)$ is either equal to 0 if $\mathbf{z} \notin \mathcal{P}^n_r$ or 1 otherwise. Indeed, we know by Proposition 4 that:

$$\frac{\rho^n_r(\mathbf{z})}{M' \cdot \rho^n_{R,\mathbf{v}}(\mathbf{z})} = \begin{cases} \geq 1 \text{ if } \rho^n_r(\mathbf{z}) \neq 0 \\ = 0 \text{ if } \rho^n_r(\mathbf{z}) = 0 \end{cases}.$$

Using distributions $D_s = \{(\mathbf{z}, \mathbf{v}) : \mathbf{v} \hookleftarrow\$ h \wedge \mathbf{z} \hookleftarrow\$ \rho^n_{R,\mathbf{v}}\}$ and its counterpart $D_t = \{(\mathbf{z}, \mathbf{v}) : \mathbf{v} \hookleftarrow\$ h \wedge \mathbf{z} \hookleftarrow\$ \rho^n_r\}$, we aim to apply Lemma 2. First we verify that they both satisfy conditions of this lemma.

We highlight that $\mathsf{Supp}(D_t) \subset \mathsf{Supp}(D_s)$ as, in Proposition 3, we know that for each $\mathbf{v} \in \mathsf{Supp}(h) \subset \mathcal{P}^n_{\beta,\mathbb{Z}}$, $\mathsf{Supp}(\rho^n_r) \subset \mathsf{Supp}(\rho^n_{R,\mathbf{v}})$. By hypothesis $M > 1$ and as such given Proposition 4: $r \geq \frac{\beta}{M^{\frac{1}{n}} - 1}$ and $R \geq r + \beta$. We compute that:

$$\begin{aligned} R_\infty(\mathcal{D}_s \parallel \mathcal{D}_t) &= \max_{\mathbf{z} \in \mathcal{P}^n_{r,\mathbb{Z}}, \mathbf{v} \in \mathsf{Supp}(h)} \frac{\mathcal{D}_s((\mathbf{z}, \mathbf{v}))}{\mathcal{D}_t((\mathbf{z}, \mathbf{v}))} = \max_{\mathbf{z} \in \mathcal{P}^n_{r,\mathbb{Z}}, \mathbf{v} \in \mathsf{Supp}(h)} \frac{h(\mathbf{v})\, \rho^n_r(\mathbf{z})}{h(\mathbf{v})\, \rho^n_{R,\mathbf{v}}(\mathbf{z})} \\ &= \max_{\mathbf{v} \in \mathsf{Supp}(h)} \left(\max_{\mathbf{z} \in \mathcal{P}^n_{r,\mathbb{Z}}} \frac{\rho^n_r(\mathbf{z})}{\rho^n_{R,\mathbf{v}}(\mathbf{z})} \right) = \max_{\mathbf{v} \in \mathsf{Supp}(h)} \mathcal{R}_\infty \left[\rho^n_r \parallel \rho^n_{R,\mathbf{v}} \right] \\ &\leq \left(\frac{R}{r}\right)^n \cdot \frac{1 + \varepsilon_{n,R}}{1 + \varepsilon_{n,r}} \leq M \cdot \frac{1 + \varepsilon_{n,R}}{1 + \varepsilon_{n,r}} = M' \end{aligned}$$

Where the last line uses Proposition 4 and the definitions of r, R and M'. Now under the assumption that $M' > 1$, we can thus apply Lemma 2. □

4 One polytope to rule them all

We supplement Section 3 with a new construction. Before going into details, we first recall existing distributions that are used in practice: the Gaussian distribution over the hyperball, as well as the uniform distributions over the hyperball or the hypercube. Both the Gaussian distribution and the uniform distribution in an hyperball are optimal with respect to two aspects: minimal rejection rate and proof of knowledge size compactness (from [DFPS22] framework). The fact that one can characterise optimality using the L_2 norm comes from the original purpose of FSwA: lattice-based zero-knowledge proofs. Most lattice based security assumptions are linked to the Euclidean norm, making it a good estimator

of the quality of a distribution. In this respect, the uniform distribution in an hypercube is a poor distribution: a hypercube of side length $2r$ has an inscribed ball of radius only r, whereas its circumradius is $n^{1/2}r$, a much larger quantity. Why then has the FSwA signature Dilithium [DKL+21] gained such attention and why has it been standardised by the NIST? The answer lies in the impracticality of Gaussian and uniform samplers in the hyperball. Both need fixed-point arithmetic at best and use a lot of randomness. Meanwhile, uniformly sampling in an hypercube is trivial and requires less randomness.

Back to our new construction, we propose a polytope with nearly the best of both aforementioned worlds: better L_2 norm compared to the hypercube (order of $n^{1/4}r$ instead of $n^{1/2}r$) and a sampler using only uniform distributions on polytopes implying a friendlier sampler compared to both Gaussian and spherical uniform distributions. We divide this section into three parts: we first define and characterise our special polytope $\mathcal{H}$, then we use Theorem 1 to successfully apply the framework from Section 3 and last we share an efficient isochronous algorithm for uniform sampling in $\mathcal{H}$.

4.1 Characterisation of $\mathcal{H}$

Definition 5. *For $n \in \mathbb{Z}_{>0}$ and $r \in \mathbb{R}_{>0}$, we define $\mathcal{H}_r^n = \mathcal{B}_\infty^n(r) \cap \mathcal{B}_1^n(r\sqrt{n})$.*

As $\mathcal{H}$ is defined as an intersection of full-rank polytopes that contains an open ball, Proposition 1 tells us that $\mathcal{H}$ is also a full-rank polytope.

Remark 2. We use the simplified notation $\mathcal{H}$ when both the dimension n and radius r can be omitted without ambiguity. Additionally and in practice, we restrict r to the positive integers, as this will always be true in our setup; this distinction will be useful for our application of Theorem 1.

Proposition 5. *For $n \in \mathbb{Z}_{>0}$ and $r \in \mathbb{R}_{>0}$, $\mathcal{H}_r^n$ is a symmetric inscribed polytope with radius $r\sqrt{\lfloor\sqrt{n}\rfloor + (\sqrt{n} - \lfloor\sqrt{n}\rfloor)^2} \leq r\sqrt[4]{n}$. Equality is achieved when $\sqrt{n} \in \mathbb{Z}$. The vertices of this polytope are all of the form $(\underbrace{r, \ldots, r}_{\lfloor\sqrt{n}\rfloor}, \Delta_n r, 0, \ldots, 0)$ up to signed permutation, where $\Delta_n = (\sqrt{n} - \lfloor\sqrt{n}\rfloor)$.*

Proof. From its definition, the polytope $\mathcal{H}_r^n$ is stable under signed permutation of coordinates, in particular it is symmetric. Let $\mathbf{v} \in \mathcal{V}(\mathcal{H}_r^n)$ be a vertex of $\mathcal{H}_r^n$. Without any loss of generality we can assume its coordinates are non-negative and sorted in decreasing order. By Proposition 2, it is impossible that $\mathbf{v}$ has two coordinates $0 < v_j < v_i < r$ for $i < j$. Indeed if that were the case, we could then write $\mathbf{v} = \frac{\mathbf{v}_1 + \mathbf{v}_2}{2}$, where $\mathbf{v}_1 = \mathbf{v} + \varepsilon\mathbf{e}_i - \varepsilon\mathbf{e}_j$ and $\mathbf{v}_2 = \mathbf{v} - \varepsilon\mathbf{e}_i + \varepsilon\mathbf{e}_j$ are both in $\mathcal{H}_r^n$ for a small enough $\varepsilon > 0$ (here $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$ denotes the canonical basis for $\mathbb{R}^n$). Using the L_1 norm condition, this shows that $\mathbf{v}$ is of the form $(r, \ldots, r, \Delta_n r, 0, \ldots, 0)$, where $\Delta_n = (\sqrt{n} - \lfloor\sqrt{n}\rfloor)$ and the first $\lfloor\sqrt{n}\rfloor$ coordinates are equal to r. This proves the first part of the statement, ie $\mathcal{H}_r^n$ is inscribed with radius $r\sqrt{\lfloor\sqrt{n}\rfloor + \Delta_n^2}$. The inequality follows from the fact that $\Delta_n^2 \leq \Delta_n$, with

equality if and only if $\Delta_n = 0$, exactly when n is a perfect square. A little extra effort with Proposition 2 can also show that all points of the aforementioned form are indeed vertices of $\mathcal{H}_r^n$. By contradiction if wlog $\mathbf{v} = (r, \dots, r, \Delta_n r, 0, \dots, 0)$ is not a vertex, then there we can write $\mathbf{v} = t\mathbf{v}_1 + (1-t)\mathbf{v}_2$ for a $t \in (0,1)$, and $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{H}_r^n$ not equal to $\mathbf{v}$. The first $\lfloor\sqrt{n}\rfloor$ coordinates of both $\mathbf{v}_1$ and $\mathbf{v}_2$ must be r, otherwise one would escape $\mathcal{B}_\infty^n(r)$. The $(1+\lfloor\sqrt{n}\rfloor)$-th coordinate of $\mathbf{v}_1$ and $\mathbf{v}_2$ must also be $\Delta_n r$, otherwise one of them escapes from $\mathcal{B}_1^n(r\sqrt{n})$. This forces the last coordinates to be 0, and $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{v}$, this is a contradiction. □

In Proposition 5 we prove two different properties of $\mathcal{H}_r^n$. First, its circumscribed circle has radius approximately $r\sqrt[4]{n}$. Second, $\mathcal{H}$ is an *inscribed* and *symmetric* polytope, as required by Proposition 3. One thing remains, for the discrete restriction we need $\Delta_n r$ to be an integer. This can be circumvented because of how flexible the L_1 norm is, using the following elementary Lemma.

Lemma 6. *Let $\varepsilon \in [0,1)$, $n, r \in \mathbb{Z}_{\geq 0}$: $\mathcal{B}_{1,\mathbb{Z}}^n(r) = \mathcal{B}_{1,\mathbb{Z}}^n(r+\varepsilon)$*

From Proposition 5, $\mathcal{H}$ is most of the time not an integral polytope. However, we prove an equivalence between its restriction to $\mathbb{Z}^n$ ($\mathcal{H}_\mathbb{Z}$) and the restriction to $\mathbb{Z}^n$ of an integral polytope:

Corollary 2. *Let $n, r \in \mathbb{Z}_{\geq 0}$, and $\Delta_n r = (\sqrt{n} - \lfloor\sqrt{n}\rfloor)$. Then:*

$$\mathcal{H}_{r,\mathbb{Z}}^n = \mathcal{B}_{\infty,\mathbb{Z}}^n(r) \cap \mathcal{B}_{1,\mathbb{Z}}^n(\lfloor\sqrt{n}\rfloor r + \lfloor\Delta_n r\rfloor).$$

In particular by Proposition 3, for positive integers R and β, $\bigcap_{\mathbf{c}\in\mathcal{H}_{\beta,\mathbb{Z}}^n} \mathcal{H}_{R,\mathbf{c},\mathbb{Z}}^n = \mathcal{H}_{R-\beta,\mathbb{Z}}^n$. Indeed, $\mathcal{B}_{1,\mathbb{Z}}^n(\lfloor\sqrt{n}\rfloor\beta + \lfloor\Delta_n\beta\rfloor)$ is an integral polytope.

4.2 Rejection sampling on $\mathcal{H} \cap \mathbb{Z}^n$

We recall our main theorem Theorem 1 and apply it directly on $\mathcal{H}_\mathbb{Z}$ since we prove its inscribed, symmetric and integral properties in Section 4.1. In this subsection we focus on its practical implementation. we first discuss the magnitude of $\varepsilon_\mathcal{H}$ in Corollary 3 meaning studying both the volume of $\mathcal{H}$ and the cardinal of $\mathcal{H}_\mathbb{Z}$. Then, we compare $\mathcal{H}_\mathbb{Z}$ to other existing constructions.

Corollary 3. *Let $M > 1$, $\beta > 0$ and h a probability distribution such that $\mathsf{Supp}(h) \subset \mathcal{H}_{\beta,\mathbb{Z}}^n$. Let $r \geq \frac{\beta}{M^{\frac{1}{n}}-1}$, $R \geq r+\beta$, and define $\varepsilon_{n,s} = \left|\mathcal{H}_{s,\mathbb{Z}}^n\right| / \mathsf{Vol}(\mathcal{H}_s^n) - 1$ for $s \in \mathbb{R}$ and $M' = \frac{1+\varepsilon_{n,R}}{1+\varepsilon_{n,r}} \cdot M$. Let $\mathbf{v} \in \mathcal{H}_\beta^n$ and define $\rho_{r,\mathbf{v}}^n := \mathcal{U}\left(\mathcal{H}_{r,\mathbf{v},\mathbb{Z}}^n\right)$. If $M' > 1$ then:*

$$\mathcal{R}_\infty\left[\,\mathcal{U}\left(\mathcal{H}_{r,\mathbb{Z}}^n\right) \,\|\, \mathcal{U}\left(\mathcal{H}_{R,\mathbf{v},\mathbb{Z}}^n\right)\,\right] = \left(\frac{R}{r}\right)^n \cdot \frac{1+\varepsilon_{n,R}}{1+\varepsilon_{n,r}} = M',$$

and the two algorithms $\mathcal{A}$ and $\mathcal{F}$ below have indistinguishable output distributions:

$\mathcal{A}$	$\mathcal{F}$
$\mathbf{v} \hookleftarrow\$ h$	$\mathbf{v} \hookleftarrow\$ h$
$\mathbf{z} \hookleftarrow\$ \rho^n_{R,\mathbf{v}}$	$\mathbf{z} \hookleftarrow\$ \rho^n_r$
output $(\mathbf{z}, \mathbf{v})$ *if* $\mathbf{z} \in \mathcal{H}^n_{r,\mathbb{Z}}$, *else* $\perp$	*output* $(\mathbf{z}, \mathbf{v})$ *with probability* $1/M'$, *else* $\perp$

$\mathcal{A}$	$\mathcal{F}$
$\mathbf{v} \hookleftarrow\$ h$	$\mathbf{v} \hookleftarrow\$ h$
$\mathbf{z} \hookleftarrow\$ \rho^n_{R,\mathbf{v}}$	$\mathbf{z} \hookleftarrow\$ \rho^n_r$
output $(\mathbf{z}, \mathbf{v})$ *if* $\mathbf{z} \in \mathcal{H}^n_{r,\mathbb{Z}}$, *else* $\perp$	*output* $(\mathbf{z}, \mathbf{v})$ *with probability* $1/M'$, *else* $\perp$

Furthermore, $\mathcal{A}$ *outputs* $(\mathbf{z}, \mathbf{v})$ *with probability* $1/M'$.

Theoretically we use the existence of M' in Corollary 3 zero-knowledge proof. Yet, in practice, we estimate the number of reject using M to obtain r and R then, given these two, we verify that $M' > 1$. This approach, detailed in the first half of Corollary 3, let us compute the volume (resp. cardinal) of $\mathcal{H}$ (resp. $\mathcal{H}_\mathbb{Z}$) for both r and R. Then, we obtain $\varepsilon_{\mathcal{H}^n_r}$ and $\varepsilon_{\mathcal{H}^n_R}$ to get M'. Directly trying to find M', without using first a M coming from the Rényi study on the associated volumes is not doable for two reasons: first, there are, at the moment, no closed form for the cardinal of $\mathcal{H}_\mathbb{Z}$. Then, using the Rényi study on volumes force us to use the intermediate M in order to avoid cyclic definition as ε can not bet defined if r and R are not defined and vice-versa.

Lemma 7 ([Fel71, (I.9 Th.3)]). *Define* α_n *as the probability that* $\mathbf{x}$ *belongs to* $\mathcal{H}^n_r$ *given that* $\mathbf{x}$ *is uniformly sampled from* $\mathcal{B}^n_1(r\sqrt{n})$. *Then:*

$$\alpha_n = \sum_{i=0}^{\lfloor\sqrt{n}\rfloor} (-1)^i \binom{n}{i} \left(1 - i\frac{1}{\sqrt{n}}\right)^{n+1}.$$

Corollary 4 (Volume of $\mathcal{H}^n_r$). *For* $n \in \mathbb{Z}_{\geq 0}$, $r \in \mathbb{R}_{>0}$ *and* α_n *as defined in Lemma 7,* $\mathsf{Vol}\left(\mathcal{H}^n_r\right) = \alpha_n \cdot \frac{(2r\sqrt{n})^n}{n!}$.

Proof. Consequence of $\Pr\left[\mathbf{x} \in \mathcal{H}^n_r \mid \mathbf{x} \hookleftarrow\$ \mathcal{B}^n_1(r\sqrt{n})\right] = \frac{\mathsf{Vol}(\mathcal{H}^n_r)}{\mathsf{Vol}\left(\mathcal{B}^n_1(r\sqrt{n})\right)}$ and Lemma 7.

We now present a trick for counting the exact number of points in $\mathcal{H}_\mathbb{Z}$. Similar techniques have been used in [DEP23, (Section 3.3)].

Lemma 8. *Let* $n, r \in \mathbb{Z}_{\geq 0}$, *and* $e_{n,r} = \lfloor\sqrt{n}\rfloor r + \lfloor\Delta_n r\rfloor$ *using notations from Corollary 2. We define* $\Omega \in \mathbb{Z}[X]$ *as* $\Omega = \sum_{i=0}^{\infty} \omega_i X^i = \left(1 + \sum_{i=1}^{r} 2X^i\right)^n$. *Then* $\left|\mathcal{H}^n_{r,\mathbb{Z}}\right| = \sum_{i=0}^{e_{n,r}} \omega_i$.

Proof. If $\delta_i = \begin{cases} 1 \text{ if } i = 0 \\ 2 \text{ otherwise} \end{cases}$, then

$$\left(1 + 2\sum_{i=0}^{r} X^i\right)^n = \sum_{l=0}^{\infty} \sum_{\substack{\mathbf{v}\in\mathbb{N}^n \\ \|\mathbf{v}\|_\infty \leq r}} \left[\delta_{v_1}\delta_{v_2}\dots\delta_{v_n} X^{\|\mathbf{v}\|_1}\right] \text{ with } \mathbf{v} = (v_1, v_2, \dots, v_n)$$

$$= \sum_{l=0}^{\infty} X^l \sum_{\substack{\mathbf{v}\in\mathbb{N}^n \\ \|\mathbf{v}\|_\infty \leq r \\ \|\mathbf{v}\|_1 = l}} \delta_{v_1}\delta_{v_2}..\delta_{v_n} = \sum_{l=0}^{\infty} \sum_{\substack{\mathbf{v}\in\mathbb{Z}^n \\ \|\mathbf{v}\|_\infty \leq r \\ \|\mathbf{v}\|_1 = l}} 1 \cdot X^l = \sum_{l=0}^{\infty} |I_{r,l}| X^l,$$

where $I_{r,l} = \{\mathbf{v} \in \mathbb{Z}^n : \|\mathbf{v}\|_\infty \leq r \text{ and } \|\mathbf{v}\|_1 = l\}$. □

Lemma 8 does not give a closed form for the cardinality of $\mathcal{H}_\mathbb{Z}$ but only an algorithm to compute it. As foreshadowing for the next section, we use Lemma 8 to estimate the magnitude of $\varepsilon_{\mathcal{H}_r^n}$ for explicit choices of parameters (n, r), and expose our results in Table 3.

Table 3: Explicit log-computations of: the volume induced by $\mathcal{H}_r^n$, the cardinality of $\mathcal{H}_{r,\mathbb{Z}}^n$, and the ratio $\varepsilon_{\mathcal{H},n,r}$.

(n, r)	$\log(\mathsf{Vol}(\mathcal{H}_r^n))$	$\log(\lvert\mathcal{H}_{r,\mathbb{Z}}^n\rvert)$	$\log(\epsilon_{\mathcal{H},n,r})$
$(1024, 165907)$	1.71756029770040	1.71756012540027	$-1.723001301634497e-07$
$(1280, 252748)$	1.72347943010027	1.72339927759822	$-8.015250204995716e-05$

To link Table 3 with Corollary 3, we compute the difference between M and M' in the case where $M = 6$ and $n = 1280$. With Corollary 3, we obtain $r = 252,748$ and $R = 252,984$. According to Table 3, $\log(\varepsilon_{\mathcal{H}_r^n})$ is approximately equal to 10^{-4} for this couple (n, r). This leads to a value of M' that is very close to M (a difference of approximately 10^{-5}). This enables us to conclude as M is a good approximation of M' in practice.

Lastly, in Table 4 we compare our construction to other well-used ones in zero-knowledge proofs in the FSwA paradigm (specifically for uniform distributions), namely: the hyperball, the hypercube and the support of the bimodal setting on hyperballs. As a reminder, r corresponds to the circumradius of the support from which the target distribution is defined (Corollary 3). Similarly, R corresponds to the circumradius of the support within which the source distribution is defined in Corollary 3.

Table 4: Comparative table between the existing construction on uniform distribution on rejection sampling with $\mathcal{B}_2^n(r, \mathbf{c}) \cup \mathcal{B}_2^n(r, -\mathbf{c})$ highlighting the bimodal approach of Haetae.

	$\mathcal{B}_2^n(r)$	$\mathcal{H}_r^n$	$\mathcal{B}_\infty^n(r)$	$\mathcal{B}_2^n(r, \mathbf{c}) \cup \mathcal{B}_2^n(r, -\mathbf{c})$
$\max_x \|x\|_2$	r	$\sqrt[4]{n} \cdot r$	$\sqrt{n} \cdot r$	$\sqrt{r^2 + \|\mathbf{c}\|_2^2}$
r	$\frac{\|\mathbf{c}\|_2}{M^{1/n}-1}$	$\frac{\|\mathbf{c}\|_2 \cdot \sqrt[4]{n}}{M^{1/n}-1}$	$\frac{\|\mathbf{c}\|_2 \cdot \sqrt{n}}{M^{1/n}-1}$	$\frac{\|\mathbf{c}\|_2}{\sqrt{(M/2)^{2/n}-1}}$
R	$r + \|\mathbf{c}\|_2$	$r + \|\mathbf{c}\|_2$	$r + \|\mathbf{c}\|_2$	$\sqrt{r^2 + \|\mathbf{c}\|_2^2}$

4.3 An isochronous sampler on $\mathcal{H} \cap \mathbb{Z}^n$

Recall our aim is to strike a balance between simplicity and optimality. In this subsection, we present an isochronous uniform sampling algorithm for $\mathcal{H}_{r,\mathbb{Z}}^n$ (as detailed in Figure 1), which relies solely on uniform sampling without replacement. This approach eliminates the need for Gaussian sampling, albeit at the cost of a low rejection rate.

A sampler is considered (perfectly) isochronous [HPRR20, (Definition 5)] when its running time is independent of any sensitive variable. We establish our main claim in Theorem 2, demonstrating that our sampler is both uniform in $\mathcal{H}_{r,\mathbb{Z}}^n$ and isochronous.

Our sampler for $\mathcal{H}_{r,\mathbb{Z}}^n$ is based on a uniform sampler in the discrete L_1-ball. We explain how to extend this approach to obtain an isochronous uniform sampler for the discrete L_1-ball of dimension n when we already have one for the discrete L_1-sphere of dimension $n+1$.

Theorem 2. *For $r \in \mathbb{Z}_{>0}$, $\mathsf{SampleH}(n, r)$ is isochronous and uniformly samples from the set $\mathcal{H}_{r,\mathbb{Z}}^n$.*

Proof. Direct consequence of Proposition 6, Lemma 9, and Proposition 7.

Furthermore, the probabilities of restarting at step 11 of $\mathsf{SampleSphere}_1(n, r)$ and step 7 of $\mathsf{SampleH}$ are provided in Proposition 6 and Proposition 7, respectively.

To achieve uniform sampling on $\mathcal{H}_{r,\mathbb{Z}}^n$, we can rely on Lemma 7, which shows that a significant portion of samples from the L_1-ball with an appropriate radius already belong to $\mathcal{H}_{r,\mathbb{Z}}^n$. Hence, we only need to reject samples that are not in the corresponding L_∞-ball.

Proposition 6. *For $r \in \mathbb{Z}_{>0}$, the sampler $\mathsf{SampleH}(n, r)$ is isochronous and provides uniform samples in $\mathcal{H}_{r,\mathbb{Z}}^n$ if $\mathsf{SampleBall}_1(n, r)$ is isochronous and uniform. Additionally, the probability of $\mathbf{Y} \neq \perp$ in Step 7 of $\mathsf{SampleH}$ is given by:*

$$\sum_{i=0}^{\lfloor \sqrt{n} \rfloor} (-1)^i \binom{n}{i} \left(1 - i\frac{1}{\sqrt{n}}\right)^{n+1}.$$

SampleSphere$_1(n, r)$

1 : $x_0 \leftarrow 0, x_n \leftarrow r + n$
2 : // $\mathcal{S} = \{X \subset [\![1, r + n - 1]\!] : \#X = n - 1\}$
3 : $\mathbf{X} \leftarrow\!\!\$\ \mathcal{U}(\mathcal{S})$
4 : $\mathbf{X} \leftarrow \mathbf{X} \cup \{x_0, x_n\}$
5 : $\mathbf{X}$.sort()
6 : // $x_0, \cdots, x_n$ the ordered elements of X
7 : **for** $i \in [\![1, n]\!]$:
8 : $b \leftarrow\!\!\$\ \{0, 1\}$
9 : $y_i \leftarrow (x_i - x_{i-1} - 1)$
10 : **if** $y_i + b = 0$ **then**
11 : **restart**
12 : $y_i \leftarrow (-1)^b y_i$
13 : **return** $\mathbf{Y} := (y_i)_{1 \le i \le n}$

SampleBall$_1(n, r)$

1 : $(y_i)_{n+1} \leftarrow\!\!\$$ SampleSphere$_1(n + 1, r)$
2 : **return** $(y_1, \cdots, y_n)$

SampleH(n, r)

1 : $\Delta_n \leftarrow (\sqrt{n} - \lfloor \sqrt{n} \rfloor)$
2 : $r' \leftarrow \lfloor \sqrt{n} \rfloor r + \lfloor \Delta_n r \rfloor$
3 : $\mathbf{Y} \leftarrow \perp$
4 : **while** $\mathbf{Y} = \perp$ **do**
5 : $\mathbf{Y} \leftarrow$ SampleBall$_1(n, r')$
6 : **if** $\|\mathbf{Y}\|_\infty > r$ **then**
7 : $\mathbf{Y} \leftarrow \perp$
8 : **return** $\mathbf{Y}$

Fig. 1: Sampling algorithm on $\mathcal{H}_\mathbb{Z}$ using intermediate samplers on the L_1 ball and sphere for $r \in \mathbb{Z}_{>0}$.

Proof. Given r' in step 2 of SampleH(n, r), we know that $\mathcal{H}^n_{r,\mathbb{Z}} \subset \mathcal{B}^n_{1,\mathbb{Z}}(r')$. If SampleBall$_1(n, r)$ is called again due to an abort ($\mathbf{Y} = \perp$ in step 7), we can conclude that SampleH(n, r) is uniform and isochronous, provided that SampleBall$_1(n, r)$ is uniform and isochronous.

Using Lemma 7, we find that its acceptance rate is exactly:

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} \left(1 - i\frac{1}{\sqrt{n}}\right)_+^{n+1},$$

where $\left(1 - i\frac{1}{\sqrt{n}}\right)_+ = 0$ when $\left(1 - i\frac{1}{\sqrt{n}}\right) \le 0$. □

In Figure 2, we share some simulations on the rejection rate for vectors in $\mathcal{H}^n_{r,\mathbb{Z}}$ when they have been uniformly sampled in $\mathcal{B}^n_{1,\mathbb{Z}}(r)$. Additionally the evolution of α_n can be found in Section B.1.

To sample in the discrete ball in dimension n using a sampler on the discrete sphere in dimension $n + 1$, we claim that there is a direct bijection between $\mathcal{S}^{n+1}_{1,\mathbb{Z}}(r\sqrt{n})$ and $\mathcal{B}^n_{1,\mathbb{Z}}(r\sqrt{n})$ by using a projection on the canonical basis in lower dimension.

Lemma 9. *Let $r \in \mathbb{Z}_{>0}$. For all $i \le n$, let $\mathbf{x} = (x_i)_{1 \le i \le n} \in \mathbb{Z}^n_{>0}$ and $r > 0$ such that $\|\mathbf{x}\|_1 = r$. If $p_i(x_1, x_2, ..., x_n) = (x_1, ..., x_{i-1}, x_{i+1}, ..., x_n)$ defines the i-th projection, then p_i is bijective, i.e. $|\mathcal{S}^{n+1}_{1,\mathbb{Z}}(r)| = |\mathcal{B}^n_{1,\mathbb{Z}}(r)|$. Additionally, if $\mathbf{X}$ is a random variable with distribution $\mathcal{U}\left(\mathcal{S}^{n+1}_{1,\mathbb{Z}}(r)\right)$, then $p_i(X)$ has distribution $\mathcal{U}\left(\mathcal{B}^n_{1,\mathbb{Z}}(r)\right)$.*

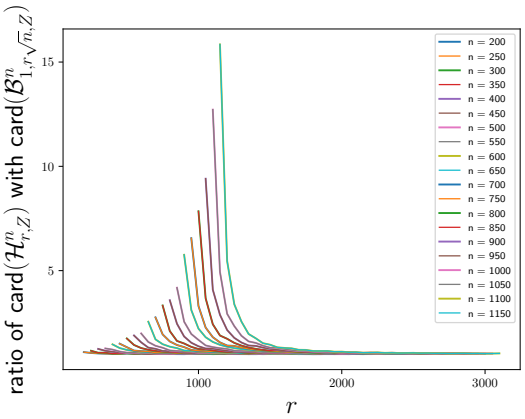

Fig. 2: Membership probability on $\mathcal{H}^n_{r,\mathbb{Z}}$ using random elements from $\mathcal{B}^n_{1,\mathbb{Z}}(r\sqrt{n})$ for $r \in [\![n, n\log(n)]\!]$ and $n \in [\![200, 1150]\!]$.

This shows that $\mathsf{SampleBall}_1(n, r)$ is both uniform and isochronous, because $\mathsf{SampleSphere}_1(n+1, r)$ is also uniform and isochronous.

Proposition 7. *For any integer $r \in \mathbb{Z}_{>0}$, the $\mathsf{SampleH}(n, r)$ algorithm of Figure 1 is both isochronous and uniform in $\mathcal{B}^n_{1,\mathbb{Z}}(r)$. Furthermore, the probability of an abort (triggering the* **restart** *instruction) is equal to:*

$$\frac{\sum_{k=1}^{n} \binom{n}{k}(1-2^{-k})\left|\mathcal{S}^{n-k}_{1,\mathbb{Z}}(r-n+k)\right|}{|\mathsf{Simplex}^n_{\mathbb{Z}}(r)|},$$

where, for $m \in \mathbb{Z}_{>0}$, $\mathsf{Simplex}^m_{\mathbb{Z}}(r) = \{\mathbf{x} \in \mathbb{N}^m : \|\mathbf{x}\|_1 = r\}$.

Proof. All operations within this algorithm, including the uniform selection of $\mathbf{X}$, can be completed in constant time except the sorting algorithm. We claim trivially that even knowing the order of each unknown variable does not help recovering them. The number of aborts is independent of the outputted value since $\mathbf{X}$ is resampled at each restart.

Let's demonstrate that output follows the uniform distribution in $\mathcal{S}^n_{1,\mathbb{Z}}(r)$. We define:

$$\mathsf{S}_{\mathsf{source}} := \{((b_1, y_1), \ldots, (b_n, y_n)) \in (\{0,1\} \times [\![0, r]\!])^n \ : \ \sum_i y_i = r\};$$

$$\mathsf{S}_{\mathsf{target}} := \{((b_1, y_1), \ldots, (b_n, y_n)) \in \mathsf{S}_{\mathsf{source}} \ : \ y_i = 0 \Rightarrow b_i = 1\}.$$

A direct analysis reveals that $\mathsf{SampleSphere}_1(n, r)$ can be reformulated as follows:

$\mathsf{SampleSphere}_1(n, r)$

1 : $\mathbf{A} = ((b_1, y_1), \ldots, (b_n, y_n)) \xleftarrow{\$} \mathcal{U}(\mathsf{S}_{\mathsf{source}})$
2 : **if** $\mathbf{A} \notin \mathsf{S}_{\mathsf{target}}$ **then goto** 1
3 : **return** $\left((-1)^{b_1} y_1, \ldots, (-1)^{b_n} y_n\right)$

Here, the (b_i, y_i) of the step 1 correspond to the b and y_i values computed in steps 8 and 9 of $\mathsf{SampleSphere}_1(n, r)$ in Figure 1.

Furthermore, we can observe that the mapping:

$$((b_1, y_1), \ldots, (b_n, y_n)) \to \left((-1)^{b_1} y_1, \ldots, (-1)^{b_n} y_n\right)$$

is a bijection between $\mathsf{S}_{\mathsf{target}}$ and $\mathcal{S}^n_{1,\mathbb{Z}}(r)$. This establishes the uniformity of the sampler.

Finally, the probability of abort is equal to the probability of sampling an element from $\mathsf{S}_{\mathsf{source}} - \mathsf{S}_{\mathsf{target}}$. By denoting $\mathcal{S}^{*,n}_{1,\mathbb{Z}}(r)$ the elements of $\mathcal{S}^n_{1,\mathbb{Z}}(r)$ without any coefficients equal to 0, this can be expressed as:

$$\begin{aligned}
&\Pr\left[\exists i \in [\![1, n]\!] \text{ with } y_i, b_i = 0 \ : \ ((b_1, y_1), \ldots, (b_n, y_n)) \leftarrow\!\$\, \mathcal{U}\left(\mathsf{S}_{\mathsf{source}}\right)\right] \\
&= \frac{1}{|\mathsf{Simplex}^n_{\mathbb{Z}}(r)|} \cdot \sum_{k=1}^{n} \binom{n}{k} 2^{-k} \sum_{j=1}^{k} \binom{k}{j} \left|\mathcal{S}^{*,n-k}_{1,\mathbb{Z}}(r)\right| \\
&= \frac{1}{|\mathsf{Simplex}^n_{\mathbb{Z}}(r)|} \cdot \sum_{k=1}^{n} \binom{n}{k} (1 - 2^{-k}) \left|\mathcal{S}^{*,n-k}_{1,\mathbb{Z}}(r)\right| \\
&= \frac{1}{|\mathsf{Simplex}^n_{\mathbb{Z}}(r)|} \cdot \sum_{k=1}^{n} \binom{n}{k} (1 - 2^{-k}) \left|\mathcal{S}^{n-k}_{1,\mathbb{Z}}(r - n + k)\right|.
\end{aligned}$$

Here, k represents the number of indices i for which $y_i = 0$, and l represents the number of indices i for which $y_i = 0$ and $b_i = 0$. □

We end this section with brief note on samplers for the hypercube and the hyperball. The randomness necessary to sample in the hypercube can be obtained directly ($n \log(2R + 1)$ with (n, R) as in Theorem 1). In this case, the sampling mechanism is easy and direct, without rejections. On the contrary, state-of-the-art samplers in the hyperball are based on Corollary 5 and use continuous Gaussian samplers. Simulating continuous Gaussian sampling with discrete Gaussian sampling leads to a large overhead in randomness usage. In addition, to sample inside the integer restriction of the ball, one needs to add specific constraints which lead to rejection (see [CCD+23] for more insight). Our algorithm to uniformly sample in $\mathcal{H}_{\mathbb{Z}}$ only uses uniform samplers and has a negligible rejection rate for practical parameters. Combined with the result of Table 4, this makes $\mathcal{H}$ an ideal candidate for zero-knowledge proofs in the FSwA paradigm.

5 The wizardry of polytopes in application

In this last section we describe applications of Section 3 and Section 4. Firstly, we give an experimental improvement of $\mathcal{H}$ by highlighting that for practical parameters most points in $\mathcal{H}_{\mathbb{Z}}$ have limited Euclidean norm. Secondly, we propose

an example application to zero-knowledge proofs in the FSwA paradigm using Lyubashevsky-like signatures. Our construction improves on Dilithium [DKL+21] by a large margin while being more practical than its counterpart Haetae [CCD+23] because of the simpler sampler.

5.1 Experimental improvement over $\mathcal{H}$

Definition 6. *For an integer $n \in \mathbb{Z}_{>0}$ and reals $r, \theta \in \mathbb{R}_{>0}$, we define $\mathcal{C}^n_{\theta,r} := \mathcal{H}^n_r \cap \mathcal{B}^n_2(\theta \cdot r)$. Furthermore, for any $u > 0$ and vector $\mathbf{w} \in \mathbb{R}^n$, recall $\mathcal{C}^n_{\theta,u,\mathbf{w}} = u\mathcal{C}^n_\theta + \mathbf{w}$.*

Remark 3. Sampling on $\mathcal{C}_\mathbb{Z}$ (with the usual abuse of notation) is done by sampling in $\mathcal{H}_\mathbb{Z}$ using Figure 1 and rejecting if the Euclidean norm is out of bounds. In our application, r is at least 2 and an integer.

Before going further we emphasise an important aspect of this construction with respect to rejection sampling in the FSwA paradigm. As a brief informal reminder, in this paradigm we sample a base vector $\mathbf{y}$ and translate it by another vector $\mathbf{s}$ to obtain $\mathbf{z}$. The goal of this procedure is to reject $\mathbf{z}$ until its distribution is independent of $\mathbf{s}$. It is necessary to understand the implications of this procedure. Namely if $\mathbf{y}$ is taken from $\mathcal{C}_\mathbb{Z}$ then the rejection sampling theorem *has* to be done using $\mathcal{C}_\mathbb{Z}$. Meanwhile, if we take $\mathbf{y}$ in $\mathcal{H}_\mathbb{Z}$ and force $\mathbf{z}$ to be in $\mathcal{C}_\mathbb{Z}$ then we only need to apply Corollary 3 with $\mathcal{H}_\mathbb{Z}$.
Why use $\mathbf{y} \in \mathcal{C}_\mathbb{Z}$ from the get go? The gains depends solely on the choice of θ for $\mathcal{C}_\mathbb{Z}$. If θ is chosen aggressively, so as to lower the radius of the circumscribed sphere of $\mathcal{C}_\mathbb{Z}$, then some additional rejections are bound to happen. In this particular case, it is more attractive to sample directly on $\mathcal{C}_\mathbb{Z}$ to avoid redoing computations.

Lemma 10. *Let $S \subset \mathbb{R}^n$ be a convex region. Define $\mathbf{v} \in \mathbb{Z}^n$ and $R, r, c > 0$ such that $R \geq r + c$ and $\mathbf{0}, -\mathbf{v}/c \in S$, then $S_r \subset S_{R,\mathbf{v}}$.*

Proof. In Section A.2 □

Because $\mathcal{C}$ is convex, the inclusion from Lemma 10 is central to define the Rényi divergence from Theorem 1. What remains to do is to estimate $\varepsilon_\mathcal{C}$. However, both the volume of $\mathcal{C}$ and the cardinality of $\mathcal{C}_\mathbb{Z}$ are non-trivial to obtain. As a matter of fact for most polytope just getting a closed form for the volume is difficult [DF88]. Our approach is twofold. We first give a volumetric argument that explains why intersecting with a Euclidean ball doesn't lose too many points. Secondly, we use experimentation to get an approximation of this ε factor as we only need it to be sufficiently small to not drastically change the rejection rate of our sampler.

The next proposition formalises the intuition that because the L_1 ball concentrates towards its center, so cutting the corners of $\mathcal{H}$ with a L_2 ball with radius a constant times larger will not affect its volume too much.

Proposition 8. *Let $n \in \mathbb{Z}_{>0}$ be a positive integer, and $r \in \mathbb{R}_{>0}$ be a positive real number. Then there exists $c > 0$ such that for any real number $\theta > 1/c$:*

$$1 - \alpha_n^{-1} \exp(-c\theta\sqrt{n}) \leq \mathsf{Vol}\left(\mathcal{C}^n_{\theta,r}\right) / \mathsf{Vol}\left(\mathcal{H}^n_r\right) \leq 1.$$

Proof. The upper bound is clear by definition of $\mathcal{C}^n_{\theta,r}$. The lower bound is more involved. First note that:

$$\mathcal{H}^n_r = (\mathcal{B}^n_\infty(r) \cap \mathcal{B}^n_1(r\sqrt{n}) - \mathcal{B}^n_2(\theta r)) \cup \mathcal{C}^n_{\theta,r} \subset (\mathcal{B}^n_1(r\sqrt{n}) - \mathcal{B}^n_2(\theta r)) \cup \mathcal{C}^n_{\theta,r}.$$

Therefore with volumes:

$$\begin{aligned}\mathsf{Vol}\left(\mathcal{C}^n_{\theta,r}\right) &\geq \mathsf{Vol}\left(\mathcal{H}^n_r\right) - \mathsf{Vol}\left(\mathcal{B}^n_1(r\sqrt{n}) - \mathcal{B}^n_2(\theta r)\right) \\ &\geq \mathsf{Vol}\left(\mathcal{H}^n_r\right) - \mathsf{Vol}\left(\mathcal{B}^n_1(r\sqrt{n})\right) + \mathsf{Vol}\left(\mathcal{B}^n_1(r\sqrt{n}) \cap \mathcal{B}^n_2(\theta r)\right).\end{aligned}$$

The last volume is computed by direct application of a theorem by Schechtman and Zinn [SZ90] restated as theorem 5.1 in [PTT18]. Taking $p = 1$ and $q = 2$, we obtain:

$$\mathsf{Vol}\left(\mathcal{B}^n_1(r\sqrt{n}) \cap \mathcal{B}^n_2(\theta r)\right) \geq (1 - \exp(-c\theta\sqrt{n}))\mathsf{Vol}\left(\mathcal{B}^n_1(r\sqrt{n})\right).$$

Using the fact that $\mathsf{Vol}\left(\mathcal{H}^n_r\right) = \alpha_n \mathsf{Vol}\left(\mathcal{B}^n_1(r\sqrt{n})\right)$, we conclude. □

Procedures to compute $|\mathcal{C}_\mathbb{Z}|$ exist however they are not memory efficient, we present an example of such a procedure using bivariate generating series in Section A.2. We now comment on the choice of the parameter θ. One natural choice would be to take $\theta = \theta_n$ where we define θ_n in such a way that we obtain $\mathsf{Vol}\left(\mathcal{B}^n_2(\theta_n r)\right) = \mathsf{Vol}\left(\mathcal{H}^n_r\right)$. Using Stirling's approximation, this amounts to taking $\theta_n = (\alpha_n/\sqrt{2})^{1/n}\sqrt{\frac{2e}{\pi}} \approx 1.315$.

We provide in Figure 3 an estimation of the proportion of rejects added by using $\mathcal{C}_\mathbb{Z}$ over $\mathcal{H}_\mathbb{Z}$ at fixed dimension with differents θ. This experiment show that there is a range of possible θ from 1.35 to 1.5 that enable a trade-off between aggressiveness (smaller proof of knowledge) and additional rejection cost (as the Euclidean norm filter gets tighter). In what follows, we use $\theta = 1.5$ as a conservative choice that still leads to major improvements.

5.2 An improved signature scheme: Patronus

This subsection highlights a concrete application of our contributions from Section 3, Section 4 and Section 5.1 through a signature scheme, Patronus, using the FSwA paradigm. We further compare it to Dilithium and Haetae [DKL+21, CCD+23], two FSwA signatures. In order to do so, we compare them practically using signature sizes. However, we may study them directly with framework from Lemma 3 (resp. Corollary 1) using the hyperball (resp. the hypercube) for Haetae (resp. Dilithium). We divide this subsection in five different parts: first, we introduce signature security notions as well as necessary lattice-based

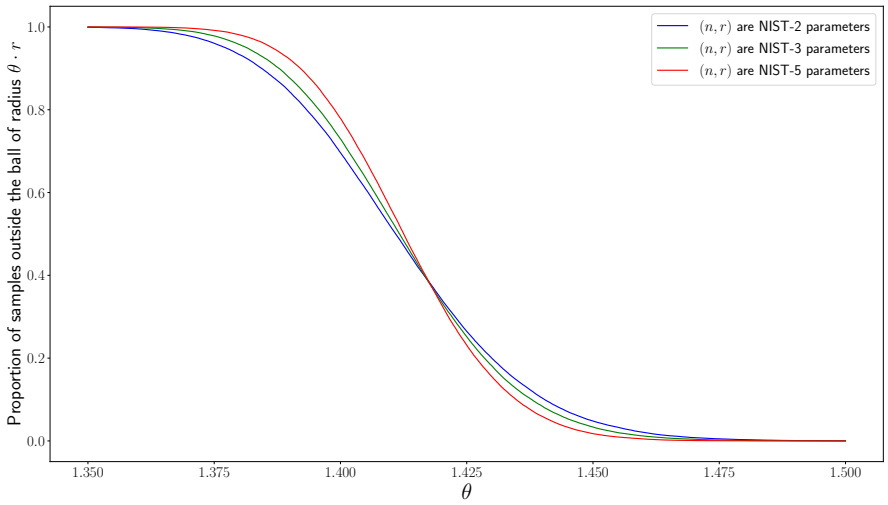

Fig. 3: Proportion of samples from $\mathcal{H}_{\mathbb{Z}}$ outside of $\mathcal{C}_\theta$ for varying θ. Experiment over 100000 samples for each parameter set, using our sampler defined in Figure 1.

assumptions. We describe Patronus and verify its correctness to then prove its security. We then present the cost of known attacks to propose different sets of parameters for 120, 180 and 260 classical bits of targeted security. Ultimately, we discuss the pointlessness of a bimodal variant. We provide competitive sets of parameters using the signature security proof as well as standard estimators derived from the cost of known attacks. Note that while this is not used in our analysis, our sets of parameters in Table 5 are chosen NTT-friendly.

Security properties and hardness assumptions. To prove Patronus secure, we prove that no adversary can forge signatures even when choosing messages. This notion is called Unforgeability under Chosen Message Attacks or alternatively UF-CMA.

Definition 7 (Unforgeability under chosen message attack (UF-CMA)). *Let $\mathcal{S} = (\mathsf{KeyGen}, \mathsf{Sign}, \mathsf{Verify})$ be a signature scheme that uses a quantum random oracle H and $\mathcal{O}_{\mathsf{sign}}(\mathsf{sk})$ an oracle which on input m computes $\sigma \leftarrow \mathsf{Sign}(\mathsf{sk}, m)$ and returns (m, σ).*

We define the advantage $\mathsf{Adv}^{\mathsf{ufcma}}_{\mathcal{S},\mathsf{q_H},\mathsf{q_S}}(\mathcal{A})$ of a quantum adversary that uses at most $\mathsf{q_H}$ quantum queries to H and $\mathsf{q_S}$ signatures queries to $\mathcal{O}_{\mathsf{sign}}(\mathsf{sk})$ against the UF-CMA security game as:

$$\mathsf{Adv}^{\mathsf{ufcma}}_{\mathcal{S},\mathsf{q_H},\mathsf{q_S}}(\mathcal{A},\mathcal{S}) = \Pr\left[\begin{array}{l}\mathsf{Verify}(\mathsf{vk}, m, \sigma) = 1 \\ \wedge (m,\sigma) \text{ not given by } \mathcal{O}_{\mathsf{sign}}(\mathsf{sk})\end{array} : \begin{array}{l}(\mathsf{sk},\mathsf{vk}) \leftarrow \mathsf{KeyGen}() \\ \wedge\, (m,\sigma) \leftarrow \mathcal{A}^{\mathcal{O}_{\mathsf{sign}}(\mathsf{vk}),|\mathsf{H}\rangle}\end{array}\right].$$

While proving UF-CMA is fairly standard for FSwA signatures in the Random Oracle Model (ROM), its proof counterpart in the Quantum Random Ora-

cle Model (QROM) is not trivial to obtain. This gave birth to a nice line of work [KLS18, DFPS23, BBD+23] to prove UF-CMA with QROM specifically for signatures based in the FSwA paradigm. For this, they prove that an alternative version of UF-CMA for the specific case of No Message Attack (UF-NMA) in the QROM implies UF-CMA in the QROM.

Definition 8 (Unforgeability under no message attack (UF-NMA)). *Let* $\mathcal{S} = (\mathsf{KeyGen}, \mathsf{Sign}, \mathsf{Verify})$ *a signature scheme that uses a quantum random oracle* H.

We define the advantage $\mathsf{Adv}^{\mathsf{ufnma}}_{\mathcal{S},\mathsf{q_H}}(\mathcal{A})$ *of a quantum adversary that uses at most* $\mathsf{q_H}$ *quantum queries to* H *against the UF-NMA security game as:*

$$\mathsf{Adv}^{\mathsf{ufcma}}_{\mathcal{S},\mathsf{q_H},\mathsf{q_S}}(\mathcal{A},\mathcal{S}) = \Pr\left[\mathsf{Verify}(\mathsf{vk}, m, \sigma) = 1 \ : \ \begin{array}{l}(\mathsf{sk},\mathsf{vk}) \leftarrow \mathsf{KeyGen}() \\ \wedge\, (m,\sigma) \leftarrow \mathcal{A}^{|\mathsf{H}\rangle}\end{array}\right].$$

Both UF-CMA and UF-NMA are proven accordingly to the signature structure and more specifically on underlying hardness assumptions given the signature design. In our case, Patronus uses lattice-based hardness assumptions (MLWE, MSIS and SelfTargetMSIS). We first define MLWE and MSIS as they are common assumptions in lattice-based cryptography. For the sake of readability, we omit the explicit mention of the modulus q and the dimension $n = 256$ associated with $\mathcal{R}_q = \mathbb{Z}_q[X]/(X^n+1)$ when defining the parameters of the problems. We refer to elements of $\mathcal{R}_q^k$ as elements of the $\mathcal{R}_q$-module of rank k and consider them with their embedding in $\mathbb{Z}_q^{k\cdot n}$. For $m, k \in \mathbb{Z}_{>0}$ and a distribution χ with $\mathsf{Supp}(\chi) \subseteq \mathcal{R}_q$, we define the distribution $\mathcal{D}^{\mathsf{mlwe}}_{m,k,\chi}$ on $\mathcal{R}_q^{m\times k} \times \mathcal{R}_q^m$ as follows:

$$(\mathbf{A},\mathbf{b}) \leftarrow\!\$\ \mathcal{D}^{\mathsf{mlwe}}_{m,k,\chi} \ \Leftrightarrow\ \mathbf{A} \leftarrow\!\$\ \mathcal{R}_q^{m\times k}, \mathbf{b} = \mathbf{A}\mathbf{s} + \mathbf{e} \text{ for } (\mathbf{s},\mathbf{e}) \leftarrow\!\$\ \chi^{k+m}.$$

Definition 9 (Decisional MLWE problem). *Given a set of parameters* $m, k \in \mathbb{Z}_{>0}$ *and a distribution* χ *with* $\mathsf{Supp}(\chi) \subseteq \mathcal{R}_q$*, the advantage* $\mathsf{Adv}^{\mathsf{d\text{-}mlwe}}_{m,k,\chi}(\mathcal{A})$ *of any probabilistic polynomial time algorithm* $\mathcal{A}$ *in solving the decisional* d-MLWE *problem over* $\mathcal{R}_q$ *is:*

$$\left|\Pr[\mathcal{A}(\mathbf{A},\mathbf{b}) = 1 : (\mathbf{A},\mathbf{b}) \leftarrow \mathcal{D}^{\mathsf{mlwe}}_{m,k,\chi}] - \Pr[\mathcal{A}(\mathbf{A},\mathbf{b}) = 1 : (\mathbf{A},\mathbf{b}) \leftarrow\!\$\ \mathcal{R}_q^{m\times k} \times \mathcal{R}_q^m]\right|.$$

Definition 10 (Search MLWE problem). *Given a set of parameters* $m, k \in \mathbb{Z}_{>0}$ *and a distribution* χ *with* $\mathsf{Supp}(\chi) \subseteq \mathcal{R}_q$*, the advantage* $\mathsf{Adv}^{\mathsf{s-mlwe}}_{m,k,\chi}(\mathcal{A})$ *of any probabilistic polynomial time algorithm* $\mathcal{A}$ *in solving the search* s-MLWE *problem over* $\mathcal{R}_q$ *is:*

$$\Pr[\mathcal{A}(\mathbf{A}, \mathbf{A}\cdot\mathbf{s}+\mathbf{e}) = \mathbf{s} : \mathbf{s} \leftarrow\!\$\ \chi^k, \mathbf{A} \leftarrow\!\$\ \mathcal{R}_q^{m\times k}, \mathbf{e} \leftarrow\!\$\ \chi^m].$$

Definition 11 (MSIS problem). *Given a set of parameters* $l, k \in \mathbb{Z}_{>0}$ *and* $\beta > 0$*, the advantage* $\mathsf{Adv}^{\mathsf{msis}}_{l,k,\beta}(\mathcal{A})$ *of any probabilistic polynomial algorithm* $\mathcal{A}$ *in solving the* MSIS *problem over* $\mathcal{R}_q$ *is:*

$$\Pr\left[(\ \mathbf{A} \mid \mathbf{I}_k)\,\mathbf{y} = \mathbf{0} \ \wedge\ 0 < \|\mathbf{y}\|_\infty < \beta : \mathbf{A} \leftarrow \mathcal{R}_q^{k\times l} \ \wedge\ \mathbf{y} \in \mathcal{R}_q^{k+l} \leftarrow \mathcal{A}(\mathbf{A})\right].$$

To prove our signature Patronus secure, we still need one more common assumption for the specific case of FSwA signatures called Self-Target MSIS (SelfTargetMSIS). We use the same SelfTargetMSIS problem as Dilithium ([DKL+21, (Section 4.1)]), except that we consider the L_2 norm instead of the L_1 norm:

Definition 12 (SelfTargetMSIS problem). *Suppose that* $\mathsf{H} : \{0,1\}^* \times \mathcal{M} \to \mathsf{SetChall} = \{c \in \mathcal{R}_q : \|c\|_1 = \tau \wedge \|c\|_\infty = \tau\}$ *is a quantum random hash oracle for some* $\tau \in \mathbb{N}$*. For positive integers* k, l *and a positive real number* β*, the advantage* $\mathsf{Adv}^{\mathsf{stmsis}}_{\mathsf{H},k,l,\beta,\mathsf{q_H}}(\mathcal{A})$ *against the* SelfTargetMSIS *problem of an adversary making at most* $\mathsf{q_H}$ *quantum queries to* $|\mathsf{H}\rangle$ *is:*

$$\Pr\left[\begin{array}{c} 0 < \|\mathbf{y}\|_2 < \beta \\ \wedge \mathsf{H}\left((\mathbf{Id} \mid \mathbf{A})\,\mathbf{y}, M\right) = c \end{array} : \begin{array}{c} \mathbf{A} \leftarrow \mathcal{R}_q^{k\times l} \\ (\mathbf{y} = (\mathbf{r}, c), M) \leftarrow \mathcal{A}^{|\mathsf{H}(\cdot)\rangle}(\mathbf{A}) \\ (\mathbf{y} = (\mathbf{r}, c), M) \in \mathcal{R}_q^{l+k-1} \times \mathsf{SetChall} \times \{0,1\}^* \end{array}\right].$$

In the classical setting, there exists a classical reduction from SelfTargetMSIS to MSIS that exhibits an adversary $\mathcal{B}$ such that $\mathsf{Adv}^{\mathsf{stmsis}}_{\mathsf{H},k,l,\beta,\mathsf{q_H}}(\mathcal{A}) \approx \sqrt{\frac{\mathsf{Adv}^{\mathsf{msis}}_{k,l,2\beta}(\mathcal{B})}{\mathsf{q_H}}}$. More details are given in [KLS18, (Section 4.5.1)].

Patronus scheme. We give a high level description of the different element of the signature from Figure 4 and a proof of its correctness. Coefficients of $(\mathbf{e}, \mathbf{s})$ follow centred binomials of parameter $(2, 0.5)$.

An analysis of the entropy is given in Section A.4 in appendix. η is a bound on the norm of $\mathbf{s}$ to tailor the Euclidean norm of $\mathbf{s}c$ such that $\|\mathbf{s}c\|_2 \leq \beta = \eta\sqrt{\tau}$ with τ the number of ± 1 in the challenge c. We assert this bound using to the detailed analysis of [CCD+23, (Section 3.1)]. This bound ensures that our main theorem on rejection sampling works on $r + \beta$ with r the radius for the target distribution after the rejection step. The parameter r is defined as in Theorem 1 but with an additional factor $\theta = 1.5$ chosen from Section 5.1. ξ in step 5 of Sign is an artificial sample to get sufficient entropy in the challenge c. In step 14 of Sign, we separate $\mathbf{z}$ into its lowbits and highbits part, the lowbits part follows approximately a uniform distribution therefore we do not apply any compression, however we apply the compression from [Dud09] on the highbits part to obtain a compressed signature. Due to the cut constraints we need $q > 2\gamma$ and $\gamma|(q-k)$. Parameter ω has been introduced in Dilithium [DKL+21], it represents the number of carry introduced by $\mathsf{MakeHint}_m$ or alternatively the number of its coefficients equal to 1. This is an interesting information since it allows to obtain a tighter bound in the Patronus security reduction. Differently to Dilithium, in this work we directly use the L_2 bound which implies from the second point in [DKL+21, (Lemma 1)] a tighter bound in the security reduction of Patronus.

We first define necessary primitives to build Patronus.

Definition 13. *Let* $r \in \mathbb{Z}$*,* $d \in \mathbb{N}^*$ *and* γ *a power of two. We define* Highbits, Lowbits *and* Power2round *as:*

$\mathsf{Power2round}(r, d)$

$r := r \bmod^+ q$

$r_0 := r \bmod^\pm 2^d$

return $((r - r_0)/2^d, r_0)$

$\mathsf{Highbits}(r, \gamma)$

return $\left\lfloor \frac{r}{\gamma} + \frac{1}{2} \right\rfloor$

$\mathsf{Lowbits}(r, \gamma)$

return $r \bmod^\pm \gamma$

Definition 14. *Let $r \in \mathbb{Z}$. Let q be a prime number and $\gamma | (q-k)$ a power of two. Let $m = (q-k)/\gamma$ and $\mathsf{s}(x)$ the function that returns 1 if $x - 1 \geq 0$ and -1 otherwise. Finally, we define $\mathsf{MakeHint}_m$ and $\mathsf{UseHint}_m$ and the subroutines $\mathsf{Highbits}_m$, $\mathsf{Lowbits}_m$ as:*

$\mathsf{Highbits}_m(r, \gamma)$

$r_1 \leftarrow \mathsf{Highbits}(r \bmod^+ q, \gamma)$

$r_0 \leftarrow \mathsf{Lowbits}(r \bmod^+ q, \gamma)$

if $r_1 = m$ **then**

  return 0

return r_1

$\mathsf{Lowbits}_m(r, \gamma)$

$r_1 \leftarrow \mathsf{Highbits}(r \bmod^+ q, \gamma)$

$r_0 \leftarrow \mathsf{Lowbits}(r \bmod^+ q, \gamma)$

if $r_1 = m$ **then**

  return $r_0 - k \bmod^\pm \gamma$

return r_0

$\mathsf{UseHint}_m(h, r, \gamma)$

$r_1 \leftarrow \mathsf{Highbits}_m(r, \gamma)$

$r_0 \leftarrow \mathsf{Lowbits}_m(r, \gamma)$

if $h = 1$ **then**

  return $(r_1 + \mathsf{s}(r_0)) \bmod^+ m$

return r_1

$\mathsf{MakeHint}_m(z, r, \gamma)$

$r_1 \leftarrow \mathsf{Highbits}_m(r, \gamma)$

$v_1 \leftarrow \mathsf{Highbits}_m(r + z, \gamma)$

return $[\![r_1 \neq v_1]\!]$

Correctness is proven using the properties of the functions defined in Definition 13 and Definition 14 in a similar way to Dilithium [DKL+21].

The proofs of these lemmas are postponed to Section A.4 in appendix.

Lemma 11. *Let $a, b \in \mathbb{Z}$ such that $a \geq 0$ and $b > 0$. It holds that:*

$$a = \left\lfloor \frac{a}{b} + \frac{1}{2} \right\rfloor \cdot b + (a \bmod^\pm b),$$

this form is the unique $a = bq + r$ with $r \in \left(-\frac{b}{2}, \frac{b}{2}\right]$.

Lemma 11 ensures the well definition of Definition 14 with its existence and unicity. As for the remaining lemmas, they are mandatory milestones to prove Patronus correctness.

Lemma 12. *Let $r \in \mathbb{Z}$. Let q a prime, $\gamma | (q-k)$ a power of two. Let $m = (q-k)/\gamma$. It holds that:*

$$r = \mathsf{Highbits}_m(r, \gamma) \cdot \gamma + \mathsf{Lowbits}_m(r, \gamma) \bmod q$$
$$\mathsf{Lowbits}_m(r, \gamma) \in [-\gamma/2, \gamma/2]$$
$$\mathsf{Highbits}_m(r) \in [0, m-1].$$

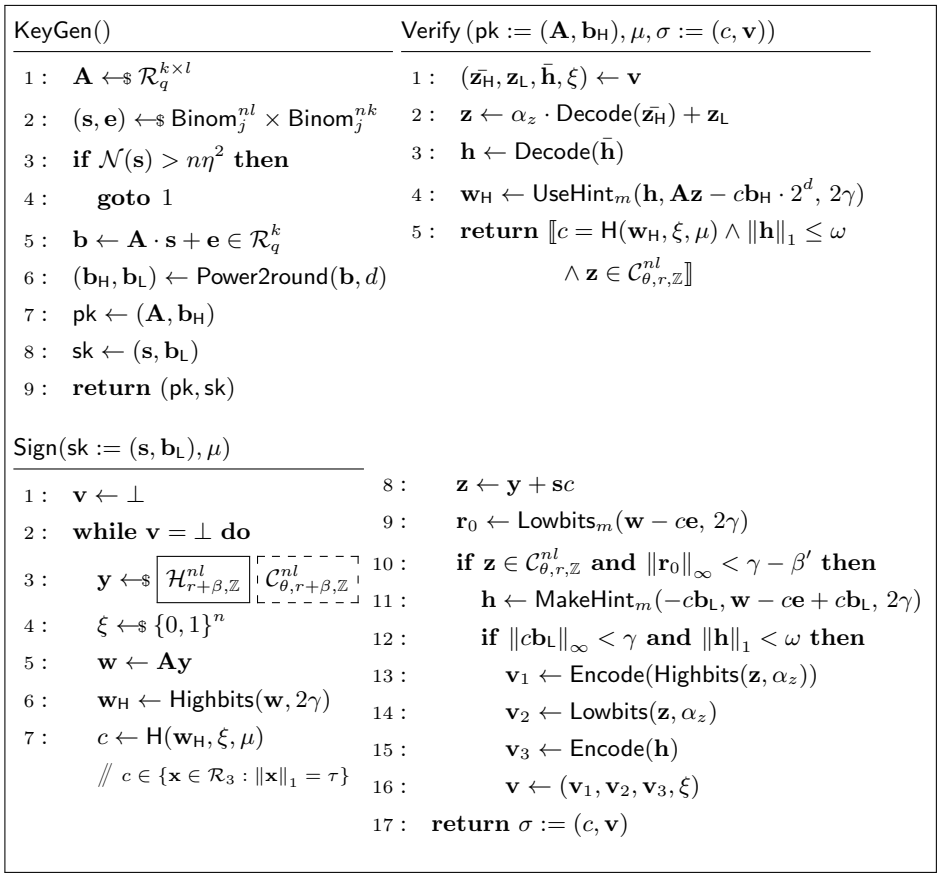

Fig. 4: Two variants of the Patronus signature. The more conservative variant is defined by using $\mathcal{H}$ in the full box while the more aggressive one uses $\mathcal{C}$ in the dashed box. $\mathcal{N}$ is defined as in [CCD+23] and $\beta' = \tau$

Lemma 13. *Let $r, z \in \mathbb{Z}_q$ and $\|z\|_\infty \le \gamma/2$:*

$$\mathsf{UseHint}_m(\mathsf{MakeHint}_m(z,r,\gamma),r,\gamma) = \mathsf{Highbits}_m(r+z,\gamma).$$

Lemma 14. *If $\|\mathbf{s}\|_\infty \le \beta$, $\|\mathsf{Lowbits}_m(\mathbf{r},\gamma)\|_\infty < \gamma/2-\beta$ then:*

$$\mathsf{Highbits}_m(\mathbf{r},\gamma) = \mathsf{Highbits}_m(\mathbf{r}+\mathbf{s},\gamma).$$

Lastly we use Lemma 12, Lemma 13 and Lemma 14 to obtain the following correctness proof for Patronus.

Proposition 9 (Correctness). *Let $(\mathsf{pk},\mathsf{sk}) \leftarrow \mathsf{KeyGen}()$, $m \in \{0,1\}^*$ and $\sigma \leftarrow \mathsf{Sign}(\mathsf{sk},M)$. Then, $\mathsf{Verify}(\mathsf{pk},M,\sigma) = 1$.*

Proof. Let's consider the elements $\mathbf{z}$, $\mathbf{w}$, $\mathbf{w}_{\mathsf{H}}$, c, ξ, and $\mathbf{v} = (\mathsf{Encode}(\mathsf{Highbits}(\mathbf{z})), \mathsf{Lowbits}(\mathbf{z}), \mathsf{Encode}(\mathbf{h}), \xi)$ computed by Sign.

It is clear by definition of $\mathsf{Encode}, \mathsf{Decode}, \mathsf{Highbits}$ and $\mathsf{Lowbits}$ that $\mathbf{z} = \alpha_z \cdot \mathsf{Decode}\left(\mathsf{Encode}\left(\mathsf{Highbits}(\mathbf{z})\right)\right) + \mathsf{Lowbits}(\mathbf{z})$ and $\mathbf{h} = \mathsf{Decode}\left(\mathsf{Encode}(\mathbf{h})\right)$. We thus only need to show that:

$$\mathbf{w}_{\mathsf{H}} = \mathsf{UseHint}_m(\mathbf{h}, \mathbf{Az} - c\mathbf{b}_{\mathsf{H}} \cdot 2^d, 2\gamma) \ \wedge \ \|\mathbf{h}\|_p \leq \omega \ \wedge \ \mathbf{z} \in \mathcal{H}^{nl}_{r,\mathbb{Z}}.$$

The two last conditions are trivially verified by definition of the signature algorithm that abort is they are not verified.

Let's show the first equation. The fact, provided by the signature algorithm, that $\|c\mathbf{b}_{\mathsf{L}}\|_\infty < \gamma$, and $\mathbf{h} = \mathsf{MakeHint}_m(-c\mathbf{b}_{\mathsf{L}}, \mathbf{w} - c\mathbf{e} + c\mathbf{b}_{\mathsf{L}}, 2\gamma)$ implies by Lemma 13 that:

$$\begin{aligned}
\mathsf{UseHint}_m(\mathbf{h}, \mathbf{Az} - c\mathbf{b}_{\mathsf{H}} \cdot 2^d, 2\gamma) &= \mathsf{Highbits}(\mathbf{Az} - c\left(\mathbf{b}_{\mathsf{H}} \cdot 2^d + \mathbf{b}_{\mathsf{L}}\right), 2\gamma) \\
&= \mathsf{Highbits}(\mathbf{Az} - c\mathbf{b}, 2\gamma) \quad \text{By Lemma 12} \\
&= \mathsf{Highbits}(\mathbf{w} - c\mathbf{e}, 2\gamma) \quad \text{by definition of } \mathbf{z} \text{ and } \mathbf{b}.
\end{aligned}$$

Finally, the fact that $\|\mathsf{Lowbits}(\mathbf{w} - c\mathbf{e}, 2\gamma)\|_\infty < \gamma - \beta$ (provided by the signature algorithm), and lemma Lemma 14 allow us to conclude. By assumption $\|c\mathbf{b}_{\mathsf{L}}\|_\infty \leq \gamma$, by Lemma 13:

$$\mathsf{UseHint}_m(\mathsf{MakeHint}_m(c\mathbf{b}_{\mathsf{L}}, \mathbf{w} - c\mathbf{e} + c\mathbf{b}_{\mathsf{L}}, \gamma), \mathbf{Az} - c\mathbf{b}_{\mathsf{H}} \cdot 2^d, 2\gamma) = \mathsf{Highbits}(\mathbf{w}, 2\gamma),$$

because $\|\mathbf{e}\|_\infty \leq (j/2)$ and $\mathsf{Lowbits}(\mathbf{w} - c\mathbf{e}, 2\gamma) < \gamma - \beta'$ with $\beta' = \max\|c\mathbf{e}\|_\infty = \tau(j/2)$. With j being the parameter of the binomial distribution. □

Secret key constraint. The rejection on $\mathbf{s}$ in Figure 4 is an important step to define and use properly $c\mathbf{s}$ in the following security proof paragraph. It ensures a theoretical bound directly on $\|c\mathbf{s}\|_2$ following directly the study from [CCD+23, (Lemma 6)], which improving drastically the study of the euclidian norm on $c\mathbf{s}$ leading to better signatures.

Lemma 15 ([CCD+23]). *For any $c \in \{0,1\}^n$ with hamming weight τ and a secret $\mathbf{s} \in \mathcal{R}^{k+l}$, $n\|c\mathbf{s}\|_2$ is bounded by*

$$\mathcal{N}(\mathbf{s}) = \tau^2 \cdot \sum_{i=1}^{m} \underset{j}{\overset{i\text{-}th}{\mathsf{max}}} \|\mathbf{s}(\omega_j)\|_2^2 + r \cdot \tau \cdot \underset{j}{\overset{(m+1)\text{-}th}{\mathsf{max}}} \|\mathbf{s}(\omega_j)\|_2^2 \|\mathbf{s}(\omega_j)\|_2^2,$$

with $m = \lfloor n/\tau \rfloor$ and $r = n \mod \tau$

Using Lemma 15, by simply fixing $\mathcal{N}(\mathbf{s}) \leq n\eta^2$ in the rejection rate, we obtain a direct upper bound on $c\mathbf{s}$.

Security of Patronus. We apply [BBD+23, (Theorem 2)] to reduce UF-CMA security to UF-NMA security.

This theorem relies on an analysis of the commitment min-entropy, the property of accepting honest-verifier zero knowledge, and the abort probability inherent in the associated identification protocol, as described in Figure 5.

$\mathsf{Com}(\mathsf{sk})$	$\mathsf{Resp}(\mathrm{com}, c, \mathrm{st})$
$\mathbf{s} \leftarrow \mathsf{sk}$	$\mathbf{y} \leftarrow \mathrm{st}$
$\mathbf{v} \leftarrow \perp$	$(\mathbf{w}_{\mathsf{H}}, \xi) \leftarrow \mathrm{com}$
$\mathbf{y} \leftarrow\!\!\$\ \mathcal{C}^{nl}_{\theta, r+\beta, \mathbb{Z}}$	$\mathbf{z} \leftarrow \mathbf{y} + \mathbf{s}c$
$\xi \leftarrow\!\!\$\ \{0,1\}^n$	$\mathbf{r}_0 \leftarrow \mathsf{Lowbits}(\mathbf{w} - c\mathbf{e},\ 2\gamma)$
$\mathbf{w} \leftarrow \mathbf{A}\mathbf{y}$	$\mathbf{v} \leftarrow \perp$
$\mathbf{w}_{\mathsf{H}} \leftarrow \mathsf{Highbits}(\mathbf{w}, 2\gamma)$	**if** $\mathbf{z} \in \mathcal{C}^{nl}_{\theta, r}$ **and** $\|\mathbf{r}_0\|_\infty < \gamma - \beta'$ **then**
$\mathrm{com} \leftarrow (\mathbf{w}_{\mathsf{H}}, \xi)$	$\quad \mathbf{h} \leftarrow \mathsf{MakeHint}_m(-c\mathbf{b}_{\mathsf{L}}, \mathbf{w} - c\mathbf{e} + c\mathbf{b}_{\mathsf{L}},\ 2\gamma)$
$\mathrm{st} \leftarrow \mathbf{y}$	$\quad$ **if** $\|c\mathbf{b}_{\mathsf{L}}\|_\infty < \gamma$ **or** $\|\mathbf{h}\|_1 < \omega$ **then**
return $(\mathrm{com}, \mathrm{st})$	$\quad\quad \mathbf{v} = (\mathsf{Encode}(\mathsf{Highbits}(\mathbf{z})), \mathsf{Lowbits}(\mathbf{z}), \mathsf{Encode}(\mathbf{h}), \xi)$
	return $(c, \mathbf{v})$

Fig. 5: Identification scheme associated to the signature algorithm.

Zero-knowledge. The underlying identification protocol has ε bits of min-entropy when the following condition is met for any $(\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{KeyGen}$ and $\mathbf{y} \leftarrow \mathcal{H}_{r_s}$:

$$\forall (\mathbf{w}, \xi), \Pr_{\mathbf{y}} \left[(\mathsf{Highbits}(\mathbf{A}\mathbf{y}, 2\gamma), \varepsilon) = (\mathbf{w}, \xi) \right] \leq 2^{-\varepsilon}.$$

Since ξ represents a uniformly random binary vector of length n, the probability is bounded by 2^{-n}, regardless of the chosen $(\mathsf{pk}, \mathsf{sk})$. Consequently, there is a minimum of 256 bits of entropy.

We must now demonstrate that the underlying Σ-protocol of the signature scheme, in Figure 5, satisfies the naHVZK property [BBD+23, (Definition 1)]. This property stipulates that non-aborting transcripts generated by the simulator must have a minimal statistical distance from real transcripts. In this context, Theorem 1 implies that the statistical distance is 0, the simulator being described in Figure 6.

For a given key pair $(\mathsf{pk}, \mathsf{sk})$, the underlying identification protocol has an abort probability associated to $(\mathsf{pk}, \mathsf{sk})$ equal to:

$$p_{\mathsf{pk},\mathsf{sk}} = \Pr_{(\mathrm{com},\mathrm{st}) \leftarrow \mathsf{Com}(\mathsf{sk}), c \leftarrow \mathsf{SetChall}} \left[\mathsf{Resp}(\mathrm{com}, c, \mathrm{st} = \perp) \right].$$

It is imperative to upper-bound every $p_{\mathsf{pk},\mathsf{sk}}$ with a constant p, except possibly for a negligible number of key pairs. Following the approach outlined in [BBD+23, (Theorem 2)], we do not rigorously prove this bound but rather estimate it using a heuristic. We follow the methodology detailed in [DKL+21, (Section 3.2)]. There are three distinct reasons to abort:

- When $\mathbf{z} \notin \mathcal{H}^{nl}_r$, which happens with probability $\frac{M-1}{M}$, as derived from Theorem 1.
- When $\|\mathsf{Lowbits}(\mathbf{w} - c\mathbf{e},\ 2\gamma)\|_\infty < \gamma - \beta'$. In [DKL+21, (Section 3.2)], it is heuristically assumed that $\mathsf{Lowbits}(\mathbf{w} - c\mathbf{e},\ 2\gamma)$ is uniformly distributed, leading to an estimation of the abort probability as $1 - e^{-256\frac{\beta' k}{\gamma}}$ with $\beta' = \tau$.

$\mathsf{Sim}(\mathbf{A}, c)$

$\xi \leftarrow\!\$\ \{0,1\}^n$
$\mathsf{flag} \leftarrow \top$
$\mathbf{z} \leftarrow\!\$\ \mathcal{C}^{nl}_{\theta,r}$
with probability $\dfrac{M-1}{M}$ **return** $\perp$
$\mathbf{r}_0 \leftarrow \mathsf{Lowbits}(\mathbf{Az} - c\mathbf{b},\ 2\gamma)$
// We use the fact that in the signature
// algorithm ,$\mathbf{w} - c\mathbf{e} = \mathbf{Az} - c\mathbf{b}$
if $\|\mathbf{r}_0\|_\infty \geq \gamma - \beta'$ **then return** $\perp$
$\mathbf{h} \leftarrow \mathsf{MakeHint}_m(-c\mathbf{b}_\mathsf{L}, \mathbf{Az} - c\mathbf{b} + c\mathbf{b}_\mathsf{L},\ 2\gamma)$
if $\|c\mathbf{b}_\mathsf{L}\|_\infty \geq \gamma$ **and** $\|\mathbf{h}\|_1 \geq \omega$:
  return $\perp$
$\mathbf{v} = (\mathsf{Encode}(\mathsf{Highbits}(\mathbf{z})), \mathsf{Lowbits}(\mathbf{z}),$
  $\mathsf{Encode}(\mathbf{h}), \xi)$
$\mathbf{w}_\mathsf{H} \leftarrow \mathsf{UseHint}_m(\mathbf{h}, \mathbf{Az} - c\mathbf{b}_\mathsf{H} \cdot \alpha,\ 2\gamma)$
return $((\mathbf{w}_\mathsf{H}, \xi), c, \mathbf{v})$

Fig. 6: Simulator for the naHVZK property.

– When $\|c\mathbf{b}_\mathsf{L}\|_\infty < \gamma$ or $\|\mathbf{h}\|_1 < \omega$. According to [DKL+21, (Section 3.2)], the parameters suggest an abort probability of less than 0.005 for this scenario.

Based on these considerations, we can estimate that $p \leq \frac{M-1}{M} + 1 - e^{-256\frac{\beta' k}{\gamma}} + \frac{1}{200}$.

UF-NMA security. The UF-NMA security is established following a procedure analogous to that of Dilithium or Haetae. The only distinction lies in our usage of a different LWE distribution.

Proposition 10. *For any quantum adversary $\mathcal{A}$ targeting the UF-NMA security with at most q_H queries to the random oracle $|H(\cdot)\rangle$, we can establish the existence of quantum adversaries $\mathcal{B}$ and $\mathcal{C}$ such that:*

$$\mathsf{Adv}^{\mathsf{ufnma}}_{\mathsf{Patronus}}(\mathcal{A}) \leq \mathsf{Adv}^{\mathsf{d\text{-}mlwe}}_{k,l,\mathcal{U}(\mathsf{Binom}^n)}(\mathcal{B}) + \mathsf{Adv}^{\mathsf{stmsis}}_{\mathsf{H},k+1,l,B_{\mathsf{NMA}},\mathsf{q_H}}(\mathcal{C}),$$

where $B_{\mathsf{NMA}} = \max\left(2\gamma + 1 + 2^{d-1}\tau, r\right)$

Proof. We call PatronumUnif the signature scheme Patronus where the vector $\mathbf{b}$ computed in KeyGen is uniformly taken on $\mathcal{R}_q^k$.

We can directly see that there exists an adversary $\mathcal{B}$ such that

$$\left|\mathsf{Adv}^{\mathsf{ufnma}}_{\mathsf{Patronus},\mathsf{q_H}}(\mathcal{A}) - \mathsf{Adv}^{\mathsf{ufnma}}_{\mathsf{PatronumUnif},\mathsf{q_H}}(\mathcal{A})\right| \leq \mathsf{Adv}^{\mathsf{d\text{-}mlwe}}_{k,l,\mathcal{U}(\mathsf{Binom}^n)}(\mathcal{B}).$$

We now study the UF-NMA security of PatronumUnif.

Let's consider a matrix $\mathbf{M} = (\mathbf{A} \mid \mathbf{b})$ uniformly taken in $\mathcal{R}_q^{k\times(l+1)}$ for the MSIS problem. We compute $(\mathbf{b}_\mathsf{H}, \mathbf{b}_\mathsf{L}) \leftarrow \mathsf{Power2round}(\mathbf{b}, d)$ and set $\mathsf{pk} = (\mathbf{a}, \mathbf{b}_\mathsf{H})$. This pk is indistinguishable from a real public key of PatronumUnif.

Suppose that $\mathcal{A}$ finds a valid signature $\left(c, \mathbf{v} = (\bar{\mathbf{z}_\mathsf{H}}, \mathbf{z}_\mathsf{L}, \bar{\mathbf{h}}, \xi)\right)$ of a message m. We define $\mathbf{z} = \alpha_z \cdot \mathsf{Decode}(\bar{\mathbf{z}_\mathsf{H}}) + \mathbf{z}_\mathsf{L}$ and $\mathbf{h} = \mathsf{Decode}(\bar{\mathbf{h}})$. By definition of Verify, we have $\mathbf{z} \in \mathcal{C}^{nl}_{\theta,r,\mathbb{Z}}$ and $c = \mathsf{H}(\mathsf{UseHint}_m(\mathbf{h}, \mathbf{Az} - c\mathbf{b}_\mathsf{H} \cdot \gamma,\ 2\gamma), \xi, m)$. We set $\mathbf{u} = \mathsf{UseHint}_m(\mathbf{h}, \mathbf{Az} - c\mathbf{b}_\mathsf{H} \cdot \gamma,\ 2\gamma) \cdot 2\gamma - \mathbf{Az} - c\mathbf{b}$, so we have

$$c = \mathsf{H}((\mathbf{A} \mid \mathbf{b} \mid \mathbf{Id}_{k+1})(\mathbf{z}, c, \mathbf{u}),\ \xi, m) \Leftrightarrow c = \mathsf{H}((\mathbf{M} \mid \mathbf{Id}_{k+1})(\mathbf{z}, c, \mathbf{u}),\ \xi, m).$$

Moreover, using properties of $\mathsf{MakeHint}_m$ and [DKL+21, (Lemma 1)], we obtain:

$$\begin{aligned}\|\mathbf{u}\|_\infty &\leq \|\mathsf{UseHint}_m(\mathbf{h}, \mathbf{Az} - c\mathbf{b}_\mathsf{H} \cdot \gamma,\, 2\gamma) - \mathbf{Az} - c\mathbf{b}_\mathsf{H} \cdot \gamma\|_\infty + \|c\,(\mathbf{b} - \mathbf{b}_\mathsf{H} \cdot \gamma)\|_\infty \\ &\leq 2\gamma + 1 + 2^{d-1}\tau.\end{aligned}$$

Thus, using $\|\mathbf{z}\|_\infty \leq r$ which is implied by $\mathbf{z} \in \mathcal{C}^{nl}_{\theta,r,\mathbb{Z}}$, we have:

$$\|(\mathbf{z}, c, \mathbf{u})\|_\infty \leq \max\left(2\gamma + 1 + 2^{d-1}\tau, r\right) = B_\mathsf{NMA},$$

□

Parameters and cost of known attacks. To provide concrete parameters and estimate the cost of the best-known attacks against our signature scheme, we adapt the concrete security analysis conduced in Haetae [CCD+23] and use an adapted version of the scripts [DS21] and `HAETAE-helper-scripts/HAETAE_security_estimates.py`, included in the Haetae reference implementation version 2023.05.02.v1.0.[5] using the $\|\cdot\|_\infty$ MSIS estimator. Lastly, to obtain shorter signatures we use [Dud09] as compression algorithm.

In Table 5, we propose concrete parameters for Patronus and present estimated security levels and sizes. These parameters aim for a similar rejection rate M to Dilithium, but the rejection rates for Patronus are slightly larger to improve public key sizes. Then, in Table 1, we compare Patronus to Dilithium and Haetae, demonstrating that it offers a compelling trade-off in terms of signature size between these two constructions.

We follow the established *core-SVP* methodology as in Haetae [CCD+23] to estimate the number of gates required to solve MLWE, MSIS and SelfTargetMSIS. Since we do not currently know of any way of exploiting the ring structure to solve MLWE and MSIS problems, we are simply viewing these problems as LWE and SIS problems. We consider the *primal* and *dual* attacks against LWE, and *plain BKZ* attacks for SIS and SelfTargetSIS. Replacing vectors $\mathbf{v}$ with $\mathrm{vec}(\mathbf{v})$ the vector obtained by concatenating the coefficients of its coordinates, and matrix entries $\mathbf{a}_{ij} \in \mathcal{R}_q$ by the 256×256 matrix whose i-th column is $\mathrm{vec}(\mathbf{x}^{i-1} \cdot \mathbf{a}_{ij})$.

The security of $\mathsf{SelfTargetMSIS}_{\mathsf{H},k,l,B_\mathsf{NMA},\mathsf{q}_\mathsf{H}}$ is estimated based on the security of $\mathsf{MSIS}_{k,l,B_\mathsf{NMA}}$ with the same bound B_NMA, following the analysis in [DKL+21, (Section 6.2.1)]. While it could have been possible to use the non-tight reduction from $\mathsf{SelfTargetMSIS}_{\mathsf{H},k,l,B_\mathsf{NMA},\mathsf{q}_\mathsf{H}}$ to $\mathsf{MSIS}_{k,l,2B_\mathsf{NMA}}$, as described under the definition of SelfTargetMSIS, we note that this choice aligns with neither Dilithium nor Haetae's approaches for the security property UF-CMA that we consider.

Sampler implementation. To demonstrate the practical viability of our sampler, we developed an unoptimised, isochronous, and portable implementation in C for SampleH and conducted experiments using the parameter sets presented in Table 5. The source code is publicly available at:

[5] Accessible at https://www.kpqc.cryptolab.co.kr/haetae.

Table 5: Patronus parameter sets for NIST security levels II, III and V.

Security target	120	180	260
q	523,777	523,777	1,047,041
(k, l)	(5,4)	(7,5)	(10,7)
η	31	34	37
j	2	2	2
τ	39	49	60
r	180,350	210,424	467,345
M	3	4.25	3
$\beta = \eta\sqrt{\tau}$	194	238	287
γ	104,755	104,755	349,013
d	14	14	15
ω	70	80	50
Forgery			
BKZ block-size b	412	625	899
Classical hardness	120	182	262
Quantum hardness	105	160	231
Key Recovery			
BKZ block-size b	475	630	902
Classical hardness	138	184	263
Quantum hardness	122	161	231
Size			
vk (with seed)	832	1,152	1,632
sign	2,070	2,575	3,721

https://github.com/patronus-signature/patronus-signature.

In Table 6, we present the competitive performance of our sampler, tested on an i5-1021U CPU. We utilised the same SHAKE-256 code as provided in Dilithium's reference implementation[6] to generate the pseudo-randomness. One should note that, on average, the sampling time constitutes less than 10% of the total signature generation time for the Dilithium reference implementation.

A bimodal variant of Patronus? In this section, we study the feasibility of a bimodal version of our scheme. Using the same notations as in the introduction, recall that in the FSwA paradigm, an element $\mathbf{z} = \mathbf{y} + c\mathbf{s}$ is rejected when it leaks information about $c\mathbf{s}$. It was shown in [DDLL13] in the case of Gaussian distributions that if one generates $\mathbf{z}$ as $\mathbf{z} = \mathbf{y} + bc\mathbf{s}$ instead, where b is a secret uniform random bit $b \in \{\pm 1\}$, then one can rescale the initial target distribution for $\mathbf{z}$ and obtain substantially shorter sizes. This technique has been widely reused in modern signature schemes and with various distributions, such as Haetae with uniform distributions over hyperballs.

[6] Accessible at https://github.com/pq-crystals/dilithium/tree/master/ref.

Table 6: Running time (cycles) and randomness consumption (bytes) for Patronus and Dilithium samplers utilizing SHAKE-256.

(a) SampleH (this work).

Speed	II	III	V
median	420,721	575,430	1,028,036
average	453,294	594,168	1,111,171
Randomness			
median	16,048	10,064	24,208
average	16,827	11,087	25,221

(b) ExpandMask (Dilithium).

Speed	II	III	V
median	24,152	29,732	42,262
average	24,173	29,943	41,968
Randomness	2,720	3,400	4,760

We use the following proposition to argue that a direct application of this technique does not improve the results of this paper.

Proposition 11. *Let $R, r \in \mathbb{R}_{>0}$ be two radii with $R > r$. Let $n \geq 4$ denote the ambient dimension. Then:*

$$\max\left\{\rho \in \mathbb{R}_{\geq 0} : \mathcal{H}^n_\rho \subseteq \bigcap_{\mathbf{c}\in\mathcal{H}^n_r} \left(\mathcal{H}^n_{R,\mathbf{c}} \cup \mathcal{H}^n_{R,-\mathbf{c}}\right)\right\} = R - r.$$

Proof. Cleary the bimodal intersection contains the unimodal intersection so $\mathcal{H}^n_{R-r}$ is contained in both. Therefore $\rho \geq R - r$. Assume by contradiction that there exists a radius $r' > R-r$ such that $\mathcal{H}^n_{r'}$ is included in $\bigcap_{\mathbf{c}\in\mathcal{H}^n_r} \left(\mathcal{H}^n_{R,\mathbf{c}} \cup \mathcal{H}^n_{R,-\mathbf{c}}\right)$. Alternatively, for all $\mathbf{z} \in \mathcal{H}^n_{r'}$ and all $\mathbf{c} \in \mathcal{H}^n_r$, at least one of $\mathbf{z}+\mathbf{c}$ and $\mathbf{z}-\mathbf{c}$ should be an element of $\mathcal{H}^n_R$. Recall the notation $\Delta_n = \sqrt{n} - \lfloor\sqrt{n}\rfloor$, and take:

$$\mathbf{c} = (0, \ldots, 0, \underbrace{r, \ldots, r}_{\lfloor\sqrt{n}\rfloor}, \Delta_n r) \qquad \mathbf{z} = (\underbrace{r', \ldots, r'}_{\lfloor\sqrt{n}\rfloor}, \Delta_n r', 0, \ldots, 0).$$

$\mathbf{z}$ is one of the vertices of $\mathcal{H}^n_{r'}$ and $\mathbf{c}$ is a vertex of $\mathcal{H}^n_r$. We have $\|\mathbf{z}+\mathbf{c}\|_1 = \|\mathbf{z}-\mathbf{c}\|_1 = \|\mathbf{z}\|_1 + \|\mathbf{c}\|_1$ as both have disjoint coordinate support because $n \geq 4$. This leads to $\|\mathbf{z}+\mathbf{c}\|_1 = \|\mathbf{z}-\mathbf{c}\|_1 = \sqrt{n}\cdot(r+r')$. Since $r' > R-r$, $r'+r > R$, which means that neither $\mathbf{z}+\mathbf{c}$ nor $\mathbf{z}-\mathbf{c}$ are in $\mathcal{H}^n_R$. We get our contradiction and this proves that $\rho \leq R-r$, which concludes.

Proposition 11 shows that in the case of uniform distribution rejection sampling, the largest version of $\mathcal{H}$ contained in the set of possible values of $\mathbf{z}$ that leak no information on the secret is the same as that obtained in the unimodal case, and therefore using a bimodal variant would not improve our results. A similar -although slightly different- limitation appears in the case of Dilithium, with hypercubes. Indeed in our case, there might still exist another polytope between $\mathcal{H}_{R-r}$ and the bimodal intersection. However, it is not hard to see that this bimodal intersection is not convex, so a convenient polytope is unlikely to exist, would complicate the analysis, and would lead to imperfect rejection sampling, which falls out of the scope of our analysis.

Acknowledgements. This work has been partially supported by the European Union - Next Generation EU through the France Relance program and by the French government through the France 2030 program under the project RESQUE. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 885394). The second author has also been supported by ANRT under the program CIFRE No 2021/0645. This work has been further supported by the French Agence Nationale de la Recherche through the France 2030 program under grant agreement No ANR-22-PETQ-0008 PQ-TLS.

References

BBD+23. Manuel Barbosa, Gilles Barthe, Christian Doczkal, Jelle Don, Serge Fehr, Benjamin Grégoire, Yu-Hsuan Huang, Andreas Hülsing, Yi Lee, and Xiaodi Wu. Fixing and mechanizing the security proof of fiat-shamir with aborts and dilithium. Cryptology ePrint Archive, Report 2023/246, 2023. https://eprint.iacr.org/2023/246.

BBRS23. Henry Bambury, Hugo Beguinet, Thomas Ricosset, and Eric Sageloli. Polytopes in the fiat-shamir with aborts paradigm. Cryptology ePrint Archive, Report 2024/411, 2023. https://eprint.iacr.org/2024/411.

BGMN05. Franck Barthe, Olivier Gué don, Shahar Mendelson, and Assaf Naor. A probabilistic approach to the geometry of the ℓpn-ball. *The Annals of Probability*, 33(2), mar 2005.

BHLY16. Leon Groot Bruinderink, Andreas Hülsing, Tanja Lange, and Yuval Yarom. Flush, gauss, and reload - A cache attack on the BLISS lattice-based signature scheme. In Benedikt Gierlichs and Axel Y. Poschmann, editors, *CHES 2016*, volume 9813 of *LNCS*, pages 323–345. Springer, Heidelberg, August 2016.

BLNS20. Jonathan Bootle, Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Gregor Seiler. More efficient amortization of exact zero-knowledge proofs for LWE. Cryptology ePrint Archive, Report 2020/1449, 2020. https://eprint.iacr.org/2020/1449.

BLNS23a. Ward Beullens, Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Gregor Seiler. Lattice-based blind signatures: Short, efficient, and round-optimal. Cryptology ePrint Archive, Report 2023/077, 2023. https://eprint.iacr.org/2023/077.

BLNS23b. Jonathan Bootle, Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Alessandro Sorniotti. A framework for practical anonymous credentials from lattices. In Helena Handschuh and Anna Lysyanskaya, editors, *CRYPTO 2023, Part II*, volume 14082 of *LNCS*, pages 384–417. Springer, Heidelberg, August 2023.

Brø83. Arne Brøndsted. *An Introduction to Convex Polytopes*. Springer New York, NY, 1983.

CCD+23. Jung Hee Cheon, Hyeongmin Choe, Julien Devevey, Tim Güneysu, Dongyeon Hong, Markus Krausz, Georg Land, Junbum Shin, Damien Stehlé, and MinJune Yi. HAETAE algorithm specifications and supporting documentation. Submission to the NIST's post-quantum cryptography standardization process, 2023.

CCF22. Augustin Chevallier, Frédéric Cazals, and Paul Fearnhead. Efficient computation of the volume of a polytope in high-dimensions using piecewise deterministic markov processes, 2022.

DDLL13. Léo Ducas, Alain Durmus, Tancrède Lepoint, and Vadim Lyubashevsky. Lattice signatures and bimodal Gaussians. In Ran Canetti and Juan A. Garay, editors, *CRYPTO 2013, Part I*, volume 8042 of *LNCS*, pages 40–56. Springer, Heidelberg, August 2013.

DEP23. Léo Ducas, Thomas Espitau, and Eamonn W. Postlethwaite. Finding short integer solutions when the modulus is small. In Helena Handschuh and Anna Lysyanskaya, editors, *Advances in Cryptology – CRYPTO 2023*, pages 150–176, Cham, 2023. Springer Nature Switzerland.

DF88. Martin E. Dyer and Alan M. Frieze. On the complexity of computing the volume of a polyhedron. *SIAM J. Comput.*, 17(5):967–974, 1988.

DFPS22. Julien Devevey, Omar Fawzi, Alain Passelègue, and Damien Stehlé. On rejection sampling in lyubashevsky's signature scheme. In Shweta Agrawal and Dongdai Lin, editors, *ASIACRYPT 2022, Part IV*, volume 13794 of *LNCS*, pages 34–64. Springer, Heidelberg, December 2022.

DFPS23. Julien Devevey, Pouria Fallahpour, Alain Passelègue, and Damien Stehlé. A detailed analysis of fiat-shamir with aborts. Cryptology ePrint Archive, Report 2023/245, 2023. https://eprint.iacr.org/2023/245.

DKL+21. Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, Peter Schwabe, Gregor Seiler, and Damien Stehlé. CRYSTALS–Dilithium: A lattice-based digital signature scheme. Submission to the NIST's post-quantum cryptography standardization process (update from February 2021), 2021.

DLL+17. Léo Ducas, Tancrède Lepoint, Vadim Lyubashevsky, Peter Schwabe, Gregor Seiler, and Damien Stehlé. CRYSTALS–Dilithium: Digital Signatures from Module Lattices. Cryptology ePrint Archive, Paper 2017/633, Version 20170627:201152, 2017. https://eprint.iacr.org/archive/2017/633/20170627:201152.

dLS18. Rafael del Pino, Vadim Lyubashevsky, and Gregor Seiler. Lattice-based group signatures and zero-knowledge proofs of automorphism stability. Cryptology ePrint Archive, Report 2018/779, 2018. https://eprint.iacr.org/2018/779.

DS21. Léo Ducas and John Schanck. pq-crystals/security-estimates. https://github.com/pq-crystals/security-estimates, 2021.

Dud09. Jarek Duda. Asymmetric numeral systems. *CoRR*, abs/0902.0271, 2009.

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.

Fel71. William Feller. *An introduction to probability theory and its applications. Vol. II.* Second edition. John Wiley & Sons Inc., New York, 1971.

GMRR22. Morgane Guerreau, Ange Martinelli, Thomas Ricosset, and Mélissa Rossi. The hidden parallelepiped is back again: Power analysis attacks on falcon. *IACR TCHES*, 2022(3):141–164, 2022.

HPRR20. James Howe, Thomas Prest, Thomas Ricosset, and Mélissa Rossi. Isochronous gaussian sampling: From inception to implementation. In Jintai Ding and Jean-Pierre Tillich, editors, *Post-Quantum Cryptography - 11th International Conference, PQCrypto 2020*, pages 53–71. Springer, Heidelberg, 2020.

KLS18. Eike Kiltz, Vadim Lyubashevsky, and Christian Schaffner. A concrete treatment of Fiat-Shamir signatures in the quantum random-oracle model. 10822:552–586, April / May 2018.

KTX08. Akinori Kawachi, Keisuke Tanaka, and Keita Xagawa. Concurrently secure identification schemes based on the worst-case hardness of lattice problems. In Josef Pieprzyk, editor, *ASIACRYPT 2008*, volume 5350 of *LNCS*, pages 372–389. Springer, Heidelberg, December 2008.

KV97. Ravi Kannan and Santosh S. Vempala. Sampling lattice points. In Frank Thomson Leighton and Peter W. Shor, editors, *Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing, El Paso, Texas, USA, May 4-6, 1997*, pages 696–700. ACM, 1997.

LNP22. Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Maxime Plançon. Lattice-based zero-knowledge proofs and applications: Shorter, simpler, and more general. In Yevgeniy Dodis and Thomas Shrimpton, editors, *CRYPTO 2022, Part II*, volume 13508 of *LNCS*, pages 71–101. Springer, Heidelberg, August 2022.

LNS20. Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Gregor Seiler. Practical lattice-based zero-knowledge proofs for integer relations. Cryptology ePrint Archive, Report 2020/1183, 2020. https://eprint.iacr.org/2020/1183.

LNS21a. Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Gregor Seiler. Shorter lattice-based zero-knowledge proofs via one-time commitments. In Juan Garay, editor, *PKC 2021, Part I*, volume 12710 of *LNCS*, pages 215–241. Springer, Heidelberg, May 2021.

LNS21b. Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Gregor Seiler. SMILE: Set membership from ideal lattices with applications to ring signatures and confidential transactions. Cryptology ePrint Archive, Report 2021/564, 2021. https://eprint.iacr.org/2021/564.

LNSW13. San Ling, Khoa Nguyen, Damien Stehlé, and Huaxiong Wang. Improved zero-knowledge proofs of knowledge for the ISIS problem, and applications. In Kaoru Kurosawa and Goichiro Hanaoka, editors, *PKC 2013*, volume 7778 of *LNCS*, pages 107–124. Springer, Heidelberg, February / March 2013.

LS18. Vadim Lyubashevsky and Gregor Seiler. Short, invertible elements in partially splitting cyclotomic rings and applications to lattice-based zero-knowledge proofs. In Jesper Buus Nielsen and Vincent Rijmen, editors, *EUROCRYPT 2018, Part I*, volume 10820 of *LNCS*, pages 204–224. Springer, Heidelberg, April / May 2018.

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.

PBY17. Peter Pessl, Leon Groot Bruinderink, and Yuval Yarom. To BLISS-B or not to be: Attacking strongSwan's implementation of post-quantum signatures. In Bhavani M. Thuraisingham, David Evans, Tal Malkin, and Dongyan Xu, editors, *ACM CCS 2017*, pages 1843–1855. ACM Press, October / November 2017.

Pre23. Thomas Prest. A key-recovery attack against mitaka in the t-probing model. In Alexandra Boldyreva and Vladimir Kolesnikov, editors, *PKC 2023, Part I*, volume 13940 of *LNCS*, pages 205–220. Springer, Heidelberg, May 2023.

PTT18. Joscha Prochno, Christoph Thäle, and Nicola Turchi. Geometry of ℓ_p^n-balls: Classical results and recent developments, 2018.

Ste17. Noah Stephens-Davidowitz. *On the Gaussian measure over lattices*. Phd thesis, New York University, 2017.

SZ90. Gideon Schechtman and Joel Zinn. On the volume of the intersection of two ln p balls. *Proceedings of the American Mathematical Society*, 110(1):217–224, 1990.

A Supplementary materials

A.1 Volume of a measurable set.

Lemma 1. *Let $S \subset \mathbb{R}^n$ be a measurable set, and let $r \geq 0$. Recall $S_r = \{r\mathbf{s} : \mathbf{s} \in S\}$. Then, we have $\mathsf{Vol}(S_r) = r^n \cdot \mathsf{Vol}(S)$.*

Proof. For a set $A \subset \mathbb{R}^n$ and $x \in \mathbb{R}^n$, we define $1_A(x) := 1$ if $x \in A$ and 0 else. It is clear that 1_S is positive and Lebesgue integrable because S is measurable. The function $f : \mathbb{R}^n \to \mathbb{R}^n$ defined by $f(\mathbf{x}) = r\mathbf{x}$ is clearly a $\mathcal{C}^1$ diffeomorphism with determinant of the Jacobian equal to r^n. We thus compute by change of variable that

$$\begin{aligned}
\mathsf{Vol}(S_r) &= \int_{\mathbb{R}^n} 1_{S_r}(x) \quad \text{by definition of volume} \\
&= \int_{\mathbb{R}^n} r^n 1_{S_r}(f(x)) \quad \text{by change of variable} \\
&= \int_{\mathbb{R}^n} r^n 1_S(x) \quad \text{because } 1_{S_r}(f(x)) = 1_S(s) \\
&= r^n \mathsf{Vol}(S) \quad \text{by definition of volume}
\end{aligned}$$

A.2 Exact point-counting algorithm for $\mathcal{C}_{\mathbb{Z}}$

Section 5.1 We prove a point-counting algorithm for $\mathcal{C}_{\mathbb{Z}}$, however we emphasise that on ordinary computers this is highly inefficient and we could not apply it for cryptographic parameters.

Lemma 16. *Let $n, r \in \mathbb{N}$, and $e_{n,r} = \lfloor \sqrt{n} \rfloor r + \lfloor \Delta_n r \rfloor$. We define $W \in \mathbb{Z}[X]$ as $W = \sum_{i=0,k=0}^{\infty} w_{i,k} X^i Y^{k^2} = \left(1 + \sum_{i=1}^{r} 2X^i Y^{i^2}\right)^n$. Then :*

$$\left|\mathcal{C}^n_{\theta,r,\mathbb{Z}}\right| = \sum_{i=0}^{e_{n,r}} \sum_{k=0}^{\lfloor (\theta \cdot r)^2 \rfloor} w_{i,k}.$$

Proof. Define $\delta_i = \begin{cases} 1 \text{ if } i = 0 \\ 2 \text{ otherwise} \end{cases}$, then :

$$\begin{aligned}
\left(1 + 2\sum_{i=0}^{r} X^i Y^{i^2}\right)^n &= \sum_{l=0}^{\infty} \sum_{\substack{\mathbf{v} \in \mathbb{N}^n \\ \|\mathbf{v}\|_\infty \leq r}} \left[\delta_{v_1}\delta_{v_2}\ldots\delta_{v_n} X^{\|\mathbf{v}\|_1} Y^{\|\mathbf{v}\|_2}\right] \\
&\quad \text{with } \mathbf{v} = (v_1, v_2, \ldots, v_n) \\
&= \sum_{l=0,k=0}^{\infty} \sum_{\substack{\mathbf{v} \in \mathbb{Z}^n, \|\mathbf{v}\|_\infty \leq r \\ \|\mathbf{v}\|_1 = l, \|\mathbf{v}\|_2^2 = k}} 1 \cdot X^l Y^k. = \sum_{l=0,k=0}^{\infty} |Y_{r,l,k}| \cdot X^l Y^k,
\end{aligned}$$

With $Y_{r,l,k} = \{\mathbf{v} \in \mathbb{Z}^n; \|\mathbf{v}\|_\infty \leq r \wedge \|\mathbf{v}\|_1 = l \wedge \|\mathbf{v}\|_2^2 = k\}$. □

A.3 Inclusion constraints for the Rényi divergence study of Section 5.1.

In order to use $\mathcal{C}_{\mathbb{Z}}$ in Theorem 1 we need to prove the following elementary inclusion lemma.

Lemma 10. *Let $S \subset \mathbb{R}^n$ be a convex region. Define $\mathbf{v} \in \mathbb{Z}^n$ and $R, r, c > 0$ such that $R \geq r + c$ and $\mathbf{0}, -\mathbf{v}/c \in S$, then $S_r \subset S_{R,\mathbf{v}}$.*

Proof. Let $\mathbf{x} \in S_r$. We want to show that $\mathbf{x} \in S_{R,\mathbf{v}}$. We have:

$$\mathbf{x} - \mathbf{v} = \frac{r}{R}\bar{\mathbf{x}} + \frac{c}{R}\tilde{\mathbf{v}} + \frac{R-r-c}{R} \cdot \mathbf{0} \text{ with } \bar{\mathbf{x}} = \frac{R}{r}\mathbf{x} \in S_R \text{ and } \tilde{\mathbf{v}} = -\frac{R}{c}\mathbf{v} \in S_R$$
$$\in S_R \text{ since } R - c - r \geq 0 \text{ and } \bar{\mathbf{x}}, \mathbf{0}, \tilde{\mathbf{v}} \in S.$$

□

A.4 Signature preliminaries

In this section we provide some additional but not core preliminaries as well as some additional lemmas in order to prove the correctness of Patronus.

Entropy of a discrete distribution. Let X be a discrete random variable over a discrete distribution $\mathcal{D}$ with probability mass function $P(X) = \{p_1, p_2, \ldots, p_n\}$, where p_i represents the probability of outcome x_i for $i = 1, 2, \ldots, n$. The entropy of the distribution $P(X)$ is defined as:

$$H(X) = -\sum_{i=1}^{n} p_i \log(p_i).$$

Centred Binomial. We denote by Binom the probability distribution of the centred binomial of parameter $(2, 0.5)$ on $\{-1, 0, 1\}$ such that $\Pr[\mathsf{Binom} = 0] = 1/2$ and $\Pr[\mathsf{Binom} = -1] = \Pr[\mathsf{Binom} = 1] = 1/4$. We define Binom^n as the probability distribution of a dimension n vector for which each coefficient follows Binom.

For cryptographic purposes, we have to compute the entropy of Binom^n for some n. A brief entropy computation gives: $H(\bar{B}_2) = \frac{\log_2(4)}{2} + \frac{\log_2(2)}{2} \approx 0.452$. For $n = 768$ as baseline for the entropy of this distribution. We obtain: $H(\bar{B}_2^{768}) = 768 \cdot H(\bar{B}_2) \approx 346$ which is larger than what is required by all proposed security parameters of our signature scheme Patronus.

On a side note, by sampling a vector $\mathbf{v}$ of dimension n following Binom^n, we consider its canonical embedding in $\mathcal{R}_q$ by considering directly $\mathbf{v} \in \mathcal{R}_q$.

Properties of the decomposition functions

Lemma 17. *Let γ a power of two. Let $q > 2$ a prime and k such that $k < \frac{\gamma}{2} + 1$ and $\gamma | q - k$ and $r \in \mathbb{Z}$. Then:*

$$
\begin{aligned}
&r = \gamma \cdot \mathsf{Highbits}(r, \gamma) + \mathsf{Lowbits}(r, \gamma);\\
&\mathsf{Lowbits}(r, \gamma) \in \left(-\frac{\gamma}{2}, \frac{\gamma}{2}\right];\\
&r \in [0, q-1] \Rightarrow \mathsf{Highbits}(r, \gamma) \in \left[0, \frac{q-k}{\gamma}\right].
\end{aligned}
$$

Proof. The first two claim are direct by definition of both Highbits and Lowbits with Lemma 11. Then by definition of Highbits, it is a non-decreasing function in r. We therefore only need to verify the case $r = q - 1$:

$$
\begin{aligned}
\mathsf{Highbits}(q-1, \gamma) &= \left\lfloor \frac{q-1}{\gamma} + \frac{1}{2} \right\rfloor\\
&= \left\lfloor \frac{q-k}{\gamma} + \frac{k-1}{\gamma} + \frac{1}{2} \right\rfloor\\
&= \frac{q-k}{\gamma} \text{ , since } k < \frac{\gamma}{2} + 1 \text{ and } \gamma | q - k \text{ .}
\end{aligned}
$$

□

Lemma 12. *Let $r \in \mathbb{Z}$. Let q a prime, $\gamma | (q - k)$ a power of two. Let $m = (q - k)/\gamma$. It holds that:*

$$
\begin{aligned}
&r = \mathsf{Highbits}_m(r, \gamma) \cdot \gamma + \mathsf{Lowbits}_m(r, \gamma) \bmod q\\
&\mathsf{Lowbits}_m(r, \gamma) \in [-\gamma/2, \gamma/2]\\
&\mathsf{Highbits}_m(r) \in [0, m-1].
\end{aligned}
$$

Proof. If $r'_0 = r_0$ and $r'_1 = r_1$ then the equality holds, if not $r'_0 = r_0 - k$ and $r'_1 = r_1 - (q-k)/\gamma$ which leads to $r'_1 \cdot \gamma + r'_0 = r_1 \cdot \alpha_h + r_0 - q$ implying the first equality with Lemma 17.
The belonging of $\mathsf{Lowbits}_m(r)$ in $[-\gamma/2, \gamma/2]$ and of $\mathsf{Highbits}_m(r)$ in $[0, m-1]$ are a direct implication of the same Lemma 17. □

Lemma 13. *Let $r, z \in \mathbb{Z}_q$ and $\|z\|_\infty \leq \gamma/2$:*

$$
\mathsf{UseHint}_m(\mathsf{MakeHint}_m(z, r, \gamma), r, \gamma) = \mathsf{Highbits}_m(r + z, \gamma).
$$

Proof. Let $(r_1, r_0) = (\mathsf{Highbits}_m(r, \gamma), \mathsf{Lowbits}_m(r, \gamma))$.
By Lemma 12, $r_1 \in [0, \frac{q-k}{\gamma})$ and $\|r_0\|_\infty \leq \gamma/2$. Since $\|\mathbf{z}\|_\infty \leq \gamma/2$:

$$
\begin{cases}
r_0 + z \in [-\gamma, \gamma/2] & \text{if } r \leq 0\\
r_0 + z \in (-\gamma/2, \gamma] & \text{if } r > 0
\end{cases}
\Rightarrow
\begin{cases}
v_1 = r_1 \bmod m \vee v_1 = r_1 - 1 \bmod m\\
v_1 = r_1 \bmod m \vee v_1 = r_1 + 1 \bmod m
\end{cases}
$$

Then depending on $h = \mathsf{MakeHint}_m(z, r, \gamma)$ either $v_1 = r_1 \bmod m$ or $v_1 = r_1 \pm 1 \bmod m$. □

Lemma 18 (Lemma 4.1 of [KLS18]). *If* $\|\mathbf{s}\|_\infty \leq \beta$ *and* $\|\mathsf{Lowbits}(\mathbf{r},\gamma)\|_\infty < \gamma/2 - \beta$ *then*

$$\mathsf{Highbits}(\mathbf{r},\gamma) = \mathsf{Highbits}(\mathbf{r}+\mathbf{s},\gamma)$$

Lemma 19. *Let* $(h,r) \in \{0,1\} \times \mathbb{Z}_q$ *and* $v_1 = \mathsf{UseHint}_m(h,r,\gamma)$*, if* $h = 0$*, then* $\|r - v_1 \cdot \gamma\|_\infty \leq \gamma/2$ *else* $\|r - v_1 \cdot \gamma\|_\infty \leq \gamma + 1$

Proof. We define and prove it comparatively to [DKL+21]. Let $(r_1, r_0) := (\mathsf{Highbits}_m(r,\gamma), \mathsf{Lowbits}_m(r,\gamma))$.
If $h = 0$: then $v_1 \cdot \gamma - r_1 = r_0 \leq \gamma/2$ using Lemma 12.
If $h = 1$ and $r_0 > 0$, $v_1 = r_1 + 1 \bmod \frac{q-1}{\gamma}$ therefore $r - v_1 \cdot \gamma = r_0 - \gamma \bmod q - 1$. Since $r_0 > 0$ by Lemma 12, $r - v_1 \cdot \gamma \leq \gamma$.
If $h = 1$ and $r_0 < 0$, the result is the same as above by changing signs.
Finally, the case $r_0 = 0$ is direct. □

Lemma 14. *If* $\|\mathbf{s}\|_\infty \leq \beta$*,* $\|\mathsf{Lowbits}_m(\mathbf{r},\gamma)\|_\infty < \gamma/2 - \beta$ *then:*

$$\mathsf{Highbits}_m(\mathbf{r},\gamma) = \mathsf{Highbits}_m(\mathbf{r}+\mathbf{s},\gamma).$$

Proof. By supposing that $\|\mathsf{Lowbits}_m(\mathbf{r},\gamma)\|_\infty < \gamma/2 - \beta$, then by definition we have $r_0 \in [-\gamma/2 + \beta, \gamma/2 + \beta]$ which implies $r_0 + s \in [-\gamma/2, \gamma/2]$ and leads to $\mathsf{Highbits}_m(\mathbf{r},\gamma) = \mathsf{Highbits}_m(\mathbf{r}+\mathbf{s},\gamma)$. □

B Additional supplementary materials

B.1 Evolution of α_n

In this subsection we show the evolution of α_n. As can be seen in Figure 7, it converges swiftly to 1, indicating that most of the volume of a L_1 ball is actually inside its inscribed $\mathcal{H}$.

B.2 General results about projections from $\mathcal{S}_p^{n+p}(r)$ to $\mathcal{B}_p^n(r)$

In our sampler, we used a bijection between a discrete L_1 sphere of dimension $n+1$ and a discrete L_1 ball or dimension n and the fact that a uniform distribution on the first set implies a uniform distribution on the second. We describe the situation for a general $p \in \mathbb{Z}_{>0}$, in the case of continuous balls. It is a particularly interesting result because it allows to reduce some problems such as sampling in a L_p ball to sampling on the L_p sphere of higher dimension (or vice versa).

Lemma 20 ([BGMN05]). *Let* $n, m \in \mathbb{Z}$ *be dimensions and* $p \in \mathbb{Z}_{>0}$ *the norm indicator, then the orthogonal projection of the uniform distribution on* $\mathcal{S}_p^n(n+m)$ *on the first* n *coordinates has density:*

$$p(x) = \frac{\Gamma(\frac{n+m}{p})}{\Gamma(\frac{m}{p})\left[2\Gamma(\frac{1}{p}+1)\right]^n} \cdot (1 - \|x\|_p^p)^{\frac{m}{p}-1}.$$

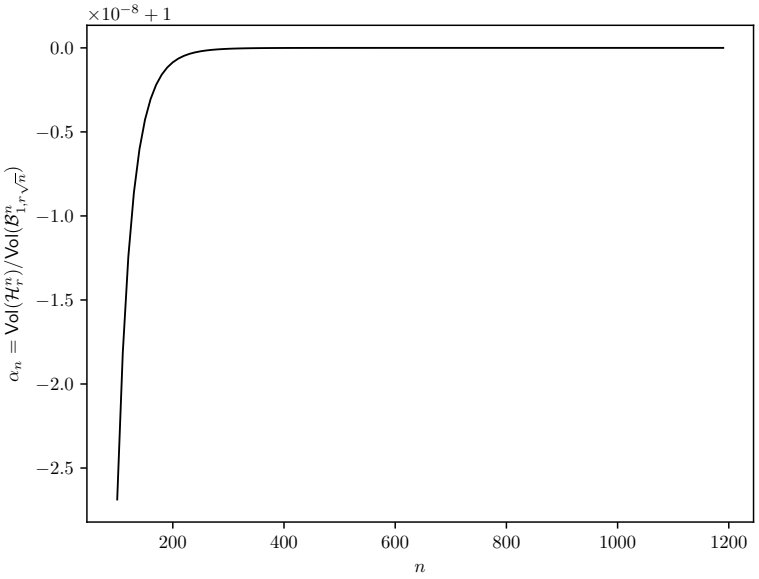

Fig. 7: Evolution of α_n.

Corollary 5. *Let $n, p \in \mathbb{Z}$. The orthogonal projection of the uniform distribution on $\mathcal{S}_p^n(n+p)$ on the first n coordinates has density:*

$$p(x) = \frac{\Gamma(\frac{n+p}{p})}{\left[2\Gamma(\frac{1}{p}+1)\right]^n},$$

which corresponds exactly to the uniform distribution in $\mathcal{B}_p^n(n)$.

While the discretisation for L_1 works fine as it is just a projection, it does not seem to hold for more general L_p balls with $p > 1$. Particularly in the interesting case of the L_2 ball, this theorem can not be extended to facilitate uniform sampling on a discretised L_2 ball.

B.3 Convergence of the cardinal of a set

This last section bounds the number of integral points in $\mathcal{H}_r^n$ and $\mathcal{C}_{\theta,r}^n$ using their volumes. In some sense, this translates the fact that for a large enough radius, the Gaussian heuristic holds. Our bounds are not tight.

We denote by $\mathcal{P}(\mathbf{x}, 1/2)$ the elementary hypercube of radius 1 centered on an arbitrary vector $\mathbf{x}$.

Lemma 21. *Let $r \in \mathbb{R}_{>0}$, then $\mathcal{H}_{r-\sqrt{n}/2}^n \subseteq \cup_{\mathbf{y}\in\mathcal{H}_{r,\mathbb{Z}}^n} \mathcal{P}(\mathbf{y}, 1/2)$.*

Proof. Let $\mathbf{x}$ be a vector of $\mathcal{H}^n_{r-\sqrt{n}/2}$, we denote $\mathbf{y} = \lfloor \mathbf{x} \rceil$. Then, there exists $\alpha \in \mathcal{P}(\mathbf{0}, 1/2)$ such that $\mathbf{x} = \mathbf{y} + \alpha$. By definition of the rounding operator $\lfloor \cdot \rceil$, $\|\mathbf{y}\|_\infty \leq \|\mathbf{x}\|_\infty + 1/2$ and $\|\mathbf{y}\|_1 \leq \|\mathbf{x}\|_1 + n/2$, which implies by definition of $\mathcal{H}^n_{r-\sqrt{n}}$ that $\mathbf{y} \in \mathcal{H}^n_r$. Additionally, $\alpha \in \mathcal{P}(\mathbf{0}, 1/2)$ directly implies $\mathbf{x} \in \mathbf{y} + \mathcal{P}(\mathbf{0}, 1/2) \in \cup_{\mathbf{y} \in \mathcal{H}^n_{r,\mathbb{Z}}} \mathcal{P}(\mathbf{y}, 1/2)$. □

Lemma 22. *Let $r \in \mathbb{R}_{>0}$, then $\cup_{\mathbf{y} \in \mathcal{H}^n_{r,\mathbb{Z}}} \mathcal{P}(\mathbf{y}, 1/2) \subseteq \mathcal{H}^n_{r+\sqrt{n}/2}$.*

Proof. Let $\mathbf{x} \in \cup_{\mathbf{y} \in \mathcal{H}^n_{r,\mathbb{Z}}} \mathcal{P}(\mathbf{y}, 1/2)$. There exists $\mathbf{y} \in \mathcal{H}^n_{r,\mathbb{Z}}$ and $\alpha \in \mathcal{P}(\mathbf{0}, 1/2)$ such that $\mathbf{x} = \mathbf{y} + \alpha$. By the triangle inequality, $\|\mathbf{x}\|_\infty \leq r + 1/2$ and $\|\mathbf{x}\|_1 \leq r\sqrt{n} + n/2$ which implies that $\mathbf{x} \in \mathcal{H}^n_{r+\sqrt{n}/2}$. □

Proposition 12. *If $r > \sqrt{n}/2$, then $\mathsf{Vol}\left(\mathcal{H}^n_{r-\sqrt{n}/2}\right) \leq \left|\mathcal{H}^n_{r,\mathbb{Z}}\right| \leq \mathsf{Vol}\left(\mathcal{H}^n_{r+\sqrt{n}/2}\right)$.*

Proof. Direct using Lemma 21 and Lemma 22 and the fact that $\left|\mathcal{H}^n_{r,\mathbb{Z}}\right| = \mathsf{Vol}\left(\cup_{\mathbf{y} \in \mathcal{H}^n_{r,\mathbb{Z}}} \mathcal{P}(\mathbf{y}, 1/2)\right)$.

Corollary 6. *If $r > \sqrt{n}/2$, then $\mathsf{Vol}\left(\mathcal{C}^n_{\theta,r-\sqrt{n}}\right) \leq \left|\mathcal{C}^n_{\theta,r,\mathbb{Z}}\right| \leq \mathsf{Vol}\left(\mathcal{C}^n_{\theta,r+\sqrt{n}}\right)$.*

Proof. We use the same proof as that of proposition 12 and verify the additional necessary constraints on the L_2 norm.

Lemma 23. *If $r \gg \sqrt{n}$ then:*

$$\left|\mathcal{H}^n_{r,\mathbb{Z}}\right| \approx \mathsf{Vol}\left(\mathcal{H}^n_r\right) \text{ and } \left|\mathcal{C}^n_{\theta,r,\mathbb{Z}}\right| \approx \mathsf{Vol}\left(\mathcal{C}^n_{\theta,r}\right).$$

Proof. Using Proposition 12 (or similarly Corollary 6 for $\mathcal{C}$) we have:

$$\mathsf{Vol}\left(\mathcal{H}^n_{r-\sqrt{n}/2}\right) \leq \left|\mathcal{H}^n_{r,\mathbb{Z}}\right| \leq \mathsf{Vol}\left(\mathcal{H}^n_{r+\sqrt{n}/2}\right).$$

By dividing the above inequalities by $\mathsf{Vol}\left(\mathcal{H}^n_{r+\sqrt{n}/2}\right)$, we compute:

$$\begin{aligned}
\frac{\mathsf{Vol}\left(\mathcal{H}^n_{r-\sqrt{n}/2}\right)}{\mathsf{Vol}\left(\mathcal{H}^n_{r+\sqrt{n}/2}\right)} &= \frac{(r-\sqrt{n}/2)^n}{(r+\sqrt{n}/2)^n} \\
&= \frac{(1-\frac{\sqrt{n}}{2r})^n}{(1+\frac{\sqrt{n}}{2r})^n} \\
&\sim \frac{e^{-\frac{n\sqrt{n}}{2r}}}{e^{\frac{n\sqrt{n}}{2r}}} \text{ since } r \gg \sqrt{n} \\
&\sim e^{-\frac{n\sqrt{n}}{r}}.
\end{aligned}$$

□