Accurate Score Prediction for Dual-Sieve Attacks

Léo Ducas[1,2] and Ludo N. Pulles[1]

[1] CWI, Cryptology Group, Amsterdam, the Netherlands
[2] Mathematical Institute, Leiden University, Leiden, the Netherlands

Abstract. The Dual-Sieve Attack on Learning with Errors (LWE), or more generally Bounded Distance Decoding (BDD), has seen many improvements in the recent years, and ultimately led to claims that it outperforms the primal attack against certain lattice-based schemes in the PQC standardization process organised by NIST. However, the work of Ducas–Pulles (Crypto '23) revealed that the so-called "Independence Heuristic", which all recent dual attacks used, leads to wrong predictions in a contradictory regime, which is relevant for the security of cryptoschemes. More specifically, the stated distributions of scores for the actual solution and for incorrect candidates were both incorrect.

In this work, we propose to use the weaker heuristic that the output vectors of a lattice sieve are uniformly distributed in a ball. Under this heuristic, we give an analysis of the score distribution in the case of an error of fixed length. Integrating over this length, we extend this analysis to any radially distributed error, in particular the gaussian as a fix for the score distribution of the actual solution. This approach also provides a prediction for the score of incorrect candidates, using a ball as an approximation of the Voronoi cell of a lattice.

We compare the predicted score distributions to extensive experiments, and observe them to be qualitatively and quantitatively quite accurate. This constitutes a first step towards fixing the analysis of the dual-sieve attack: we can now accurately estimate false-positives and false-negatives. Now that the analysis is fixed, one may consider how to fix the attack itself, namely exploring the opportunities to mitigate a large number of false-positives.

Keywords: Lattices · Cryptanalysis · Heuristics · Learning with Errors · Dual Attack · Bessel Functions

1 Introduction

Many post-quantum cryptoschemes base their security on the hardness of the Learning with Errors (LWE) problem, introduced by Regev in 2005 [Reg09], which is basically the Bounded Distance Decoding (BDD) problem in q-ary lattices. One possible type of attack against BDD is the so-called *dual attack* dating back to [AR04].

Specifically, the dual attack attacks the search-BDD problem by performing a reduction to the decision-BDD problem, where one needs to determine if a target point in Euclidean space is either a *uniform target*, i.e. sampled uniformly modulo the lattice, or a *BDD target*, i.e. sampled from a distribution that is concentrated around lattice points. Here, short dual vectors are used to construct a score function that assigns a high score to BDD targets and an expected score of 0 to uniform targets. Then with high probability, the score function will assign a score above a certain threshold to a BDD target, and below the threshold to a uniform target, effectively solving decision-BDD. Many short dual vectors are needed to separate the two distributions and to get a high success probability. By using a lattice sieve [NV08,MV10,BDGL16], one gets many short dual vectors rather efficiently [ADPS16]. To simplify the analysis of the score distribution, [GJ21,MAT22] used the so-called *Independence Heuristic*, which states that, given a set of dual vectors $\mathcal{W}$, the inner products

$$(\langle \mathbf{w}, \mathbf{t} \rangle)_{\mathbf{w} \in \mathcal{W}} \pmod 1,$$

are mutually independent variables [DP23, Heur. 3]. Recent works [AS22,CST22] improved the dual attack of [MAT22] further against LWE, leading to the claim that the dual attack outperforms the *primal attack* (which solely uses the primal lattice) when attacking certain NIST PQC standardization candidates, like Kyber and Dilithium.

However, the Independence Heuristic was shown to be flawed in experiments done in a regime relevant to cryptoschemes [DP23]. More specifically, [DP23] concludes that an analysis using the Independence Heuristic overestimate the success probability of the dual attack. The following two phenomena were observed in experiments, but were not accounted for by the analysis derived under the Independence Heuristic.

First, the probability of a uniform target giving a high score seems to be much higher than predicted by a normal distribution, as there appears to be a “waterfall-floor” phenomenon. It was presumed in [DP23] that this is caused by rare events in which a uniform target lies very close to the lattice. Second, the BDD score distribution for gaussian errors has a much bigger variance than the prediction, and is also not normally distributed. In particular, the probability on a low score seems more likely than expected from analyses used up to now. Overall, they concluded that, as currently parametrized, the Dual-Sieve attack would lead to a large number of false-positives, and also a somewhat larger rate of a false-negative than predicted.

1.1 Contributions

The concluding section of the work of Ducas and Pulles [DP23, Sec. 6.3] mentions various mitigation strategies to deal with the large number of false-positives, but highlights the prior requirement of making accurate predictions of all the score distributions at hand. This work aims precisely at fulfilling this prior requirement.

To this end, we propose a seemingly weaker assumption on the output distribution of a lattice sieve, to overcome mispredictions caused by the Independence

Heuristic illustrated in [DP23]. The mispredictions show there is some correlation between the inner products $\langle \mathbf{w}, \mathbf{t} \rangle \pmod 1$, so the situation can be resolved by identifying the confounding variable that causes the mutual dependence. In this case, the confounding variable seems to be the target $\mathbf{t}$, or more precisely its length $\|\mathbf{t}\|$. For example, scaling the target by a factor α, multiplies *all* the inner products by α together. Indeed, by fixing the confounding variable $\|\mathbf{t}\|$, one may more reasonably hope the inner products $\langle \mathbf{w}, \mathbf{t} \rangle \pmod 1$ become independent again. This leads very naturally to studying the score distribution for spherical errors. While the case of spherical errors is of limited interest in our context, it does allow to bootstrap an analysis for other radial distributions, e.g. uniform in a ball, or gaussian.

Another alteration done differently than the existing literature is to stop approximating the distribution of lattice sieve vectors by a gaussian distribution (cf. [MAT22, Assumption 4.4]), but instead model them as being uniform in a ball, as described in Heuristic 2. This somewhat complicates the Fourier analysis, but is feasible thanks to the well-known Bessel functions.

The overall approach remains heuristic, but the previously identified issues [DP23] have now been sidestepped, and we further confirm those score distribution predictions with extensive experiments.

Individual Score Function (Section 3.1). First, for a fixed error, we derive an *analytic expression* for the expectation value and variance of the score distribution for a single dual vector, in terms of Bessel functions, using our model of the dual vectors.

Error Uniform from a Sphere (Section 3.2). Using the Independence Heuristic, one could not add individual scores together because the expectation value of each score was not fixed, since it depends on the length of error. Here, we fix the length of the error, i.e. we look at errors drawn uniformly from a sphere. Note that for every error, each individual score has the same individual score distribution, we can resort to a central limit heuristic — which seems reasonable as there are exponentially many dual vectors — to reason about the sum of individual scores. This leads us to propose Heuristic 3 for BDD targets uniform from a sphere.

Radial Error Distributions (Section 3.3). Lattice-based schemes normally do not use errors that are sampled uniformly from a sphere, but instead according to a (discrete) gaussian distribution. Hence, we extend the score prediction to any error distribution that is *radial*, i.e. where the PDF is invariant under rotations. Any radial distribution can be seen as first sampling a radius from some distribution, and then sampling a point uniformly from a sphere of that radius. Thus, there is a natural way to predict the score distribution for any radial distribution for BDD targets, by integrating the sphere prediction as a function of the radius, taking the probability of such radius into account. In particular, examples are given for gaussian errors, and errors uniformly from the ball. The obtained predictions for the Cumulative Distribution Function (CDF) can be computed numerically.

Errors Uniform Modulo Lattice (Sections 3.4). Moreover, we provide a prediction of the score distribution for targets from the uniform distribution modulo the lattice. Alternatively, these targets are sampled uniformly from the Voronoi cell of a lattice. Although determining the exact shape of the Voronoi cell is a notoriously hard problem, we simply approximate the Voronoi cell by a ball of the same volume as that of the lattice. This allows us to predict the score distribution for uniform targets, using the prediction for radial error distributions.

Experiments (Section 4). All the predictions that are made in this paper, have been extensively tested by large experiments, to verify whether the proposed heuristics are reasonable in the context of solving decision-BDD.

The predicted score distributions for sphere, ball and gaussian BDD targets are shown in Figure 1, in dimension 90. In addition, the predicted score distributions for targets modulo the lattice is shown in Figure 2, which ran in dimensions $40, 50, 60$ and 70, each based on 2^{48} samples. Lastly, Figure 3 shows the score distribution for uniform targets, using a larger saturation radius inside the sieve. Here, our prediction claims that the "waterfall-floor" phenomenon could be observed with fewer samples than normal, which the experiments confirm.

All in all, the experiments make a strong case that the heuristic proposed in this paper are very reasonable for analysing dual attacks.

Open data. All the used `Python` and `C`/`C++` code and obtained data to generate the figures, can be found in the following GitHub repository:

`https://github.com/ludopulles/AccurateScorePredictionDualSieveAttacks`.

1.2 Concurrent Work

During the writing process of this paper, we became aware of several concurrent works [WE23,PS23a,PS23b,CDMT23].

The work of Wiemer and Ehlen [WE23] proposed an analysis of the variance of the score of the actual BDD target. They use a significantly different approach, and obtain predictions compatible with the experimental data from [DP23]. This does not directly allow to conclude on false-negative rates, as it does not fully characterize the distribution.

The work of Pouly and Shen [PS23a] proposed a provable variant of a variation of the dual attack, using discrete gaussian sampling in place of sieving vectors. Notably, their provable regime does not intersect the heuristic contradictory regime of [DP23]. In fact, and quite interestingly, both regime are separated by a constant factor of 2 on the length of the BDD target. In a follow-up work in progress, they are also considering uniform dual vectors in a ball [PS23b].

The work of Carrier, Debris, Meyer-Hilfiger and Tillich, while mostly focusing on the case of statistical decoding for codes, also includes a short section on the case of lattices [CDMT23, Sec. 8]. More specifically, they propose a prediction for the floor phenomenon for uniform targets modulo the lattice, also using Bessel functions. They use a somewhat comparable but not identical reasoning and derivations as that of our Section 3.4.

1.3 Analogies with Coding Theory

The dual attack in lattices, is called "statistical decoding" in coding theory and dates back to the work of [Al 01]. Here, with the use of many low-weight parity-check equations $\mathcal{H}$ (analogous with short dual vectors for lattices), similarly one can decide whether a target $\mathbf{t}$ in $\mathbb{F}_q^n$ is close to some codeword or not, based on a similar scoring function.

Recently, [CDMT22] claim to have an improved Statistical Decoding algorithm outperforming the state-of-the-art Information Set Decoding algorithm of [BM18] on sparse binary codes. However, this work was based on the coding-theory analogue of the Independence Heuristic [CDMT22, Assumption 3.7], which states that individual scores $(\langle \mathbf{h}, \mathbf{t} \rangle)_{\mathbf{h} \in \mathcal{H}}$ are i.i.d. Bernoulli variables. Here, similarly it was soon realized this heuristic leads to a flawed analysis, and the problems were overcome, by proposing two fixes [MT23]. One is that, for any uniformly random $\mathbf{t} \in \mathbb{F}_2^n$, they model the number of weight w elements in a coset $\mathcal{C} + \mathbf{t}$ as a Poisson process with parameter $\lambda = \binom{n}{w}/2^{n-k}$, when $\mathcal{C}$ is a $[n,k]$ code.[1] With the other fix of using Fourier analysis on the score function, they show the runtime complexity of the corrected algorithm is asymptotically only increased by logarithmic factors.

Acknowledgments. We would like to thank Kevin Carrier, Thomas Debris-Alazard, Charles Meyer-Hilfiger, Amaury Pouly, Yixin Shen en Jean-Pierre Tillich for insightful conversations, and for sharing their early drafts [CDMT23,PS23b]. Authors Léo Ducas and Ludo Pulles are supported by ERC Starting Grant 947821 (ARTICULATE).

2 Preliminaries

Notation. For any set $S \subseteq \mathbb{R}^n$, the *indicator function of* S is denoted by $\mathbf{1}_S(\mathbf{x})$, and is equal to 1 when $\mathbf{x} \in S$ and 0 for $\mathbf{x} \in \mathbb{R}^n \setminus S$. The $n-1$-dimensional sphere, embedded in n-dimensional Euclidean space, is denoted by $\mathcal{S}^{n-1} \subset \mathbb{R}^n$, and the n-dimensional ball is denoted by $\mathcal{B}^n$. The n-dimensional ball has volume

$$\mathrm{Vol}_n(\mathcal{B}^n) = \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}.$$

[1] Note that the coding theory dual attack also has an enumeration part, after which one tries to solve the decisional problem on a subcode. Here, we phrase the model with $\mathcal{C}$ as the subcode, whereas the model in [MT23] is written in terms of the $[n-s, k-s]$-subcode there.

The n-dimensional *gaussian of width* $s \in \mathbb{R}_{>0}$ is defined by

$$\rho_s(\mathbf{x}) = \exp\left(-\pi \frac{\|\mathbf{x}\|^2}{s^2}\right) \qquad (\mathbf{x} \in \mathbb{R}^n).$$

The n-dimensional gaussian (or, normal distribution), $\mathcal{N}\big(0, \sigma^2\big)^n$, with standard deviation $\sigma > 0$ is the distribution with a Probability Density Function (PDF) proportional to $\rho_{\sqrt{2\pi}\sigma}(\mathbf{x})$.

Lattices. The $\mathbb{R}$-*linear span of a set* $S \subset \mathbb{R}^n$ is denoted by $\operatorname{span}(S)$. A *basis for a lattice* $\Lambda \subset \mathbb{R}^n$ consists of $\mathbb{R}$-linearly independent column vectors $\mathbf{b}_1, \ldots, \mathbf{b}_k \in \mathbb{R}^n$ such that $\Lambda = \mathbb{Z}\mathbf{b}_1 + \cdots + \mathbb{Z}\mathbf{b}_k$. The *dual of a lattice* $\Lambda \subset \mathbb{R}^n$, denoted by $\Lambda^\vee$, consists of all vectors $\mathbf{x} \in \operatorname{span}(\Lambda)$ for which $\langle \mathbf{x}, \Lambda \rangle \subseteq \mathbb{Z}$ holds. We refer to Λ as the *primal lattice* and $\Lambda^\vee$ as the *dual lattice*. The *Voronoi cell of* Λ, denoted by $\mathcal{V}(\Lambda)$, is the set of points that are not closer to any lattice point than to the origin, i.e.

$$\mathcal{V}(\Lambda) = \{\mathbf{x} \in \mathbb{R}^n \mid \forall \mathbf{v} \in \Lambda : \|\mathbf{x}\| \leq \|\mathbf{x} - \mathbf{v}\|\}.$$

The volume of a lattice is $\det(\Lambda) = \operatorname{Vol}_n(\mathcal{V}(\Lambda))$.

Gaussian Heuristic. The *Gaussian Heuristic* states that for a lattice Λ, the number of lattice points lying in a measurable set $S \subset \mathbb{R}^n$ is approximately $\operatorname{Vol}_n(S) / \det(\Lambda)$. This leads to the following heuristic on the length of a shortest vector.

Heuristic 1 (Gaussian Heuristic) *Given a random lattice $\Lambda \subset \mathbb{R}^n$, the length of the shortest non-zero vector in Λ, denoted $\lambda_1(\Lambda)$, is approximately equal to $\mathrm{GH}(n) \cdot \sqrt[n]{\det(\Lambda)}$, where*

$$\mathrm{GH}(n) = \operatorname{Vol}_n(\mathcal{B}^n)^{-1/n}.$$

Bessel functions. The class of Bessel functions is defined as follows, and will be useful for the expected score of spherical errors later on.

Definition 1. *For any $\alpha > -\frac{1}{2}$, the* Bessel function (of the first kind) of order *α is given by*

$$J_\alpha(t) = \frac{(t/2)^\alpha}{\sqrt{\pi} \cdot \Gamma\big(\alpha + \frac{1}{2}\big)} \int_{-1}^{1} e^{its} \big(1 - s^2\big)^{\alpha - \frac{1}{2}} \mathrm{d}s,$$

for $t > 0$.

It is well known that the Bessel function of some order α has an infinite number of positive roots. Let us denote with $j_{\alpha,n}$, the nth positive root of the Bessel function of order α.

Lemma 1 ([Wat22, §15.81]). *The first positive root of the Bessel function satisfies*

$$j_{\alpha,1} = \alpha + 1.855757\ldots \cdot \sqrt[3]{\alpha} + O(\alpha^{-1/3}),$$

as $\alpha \to \infty$, where $1.855757\ldots$ can be computed numerically up to arbitrary precision. In addition, we have $J_\alpha(x) > 0$ for all $0 < x < j_{\alpha,1}$.

For convenience later on, let us define for all $x > 0$ and $\alpha > -\frac{1}{2}$,

$$\xi(\alpha, x) = \frac{\Gamma(\alpha+1)}{(\pi x)^\alpha} J_\alpha(2\pi x) = {}_0F_1\left(;\alpha+1; -\pi^2 x^2\right),$$

where ${}_0F_1$ is the confluent hypergeometric function. The function $x \mapsto \xi(\alpha, x)$ has an image within the interval $[-1, 1]$.

Remark 1. By using [AS64, 9.1.10], and the crude approximation $\Gamma(\alpha+k+1) = (\alpha+1)\cdot(\alpha+2)\cdot\dots\cdot(\alpha+k)\Gamma(\alpha+1) \approx (\alpha+1)^k\Gamma(\alpha+1)$, we get the following approximation for small x:

$$\xi(\alpha, x) = \sum_{k=0}^{\infty} \frac{(-\pi^2 x^2)^k}{k!(\alpha+1)(\alpha+2)\ldots(\alpha+k)} \approx \sum_{k=0}^{\infty} \frac{(-\pi^2 x^2)^k}{k!(\alpha+1)^k} = e^{-\frac{\pi^2 x^2}{\alpha+1}}.$$

Remark 2. In high dimensions, one may get some numerical errors when computing ξ directly with Bessel functions. These issues are circumvented by computing ξ using the confluent hypergeometric function ${}_0F_1$. For example, the Python package 'mpmath' implements ${}_0F_1$ with the function `hyp0f1`.

2.1 Lattice Sieve

Lattice sieves provide a way to efficiently produce a list of many short lattice vectors [NV08,MV10,BDGL16]. Although there are many variations, e.g. [BLS16], one may think of a sieve as initially generating a list L of random linear combinations of some basis vectors defining a lattice $\Lambda \subset \mathbb{R}^n$, and then iteratively sieving, i.e. finding reductions that replace $\mathbf{v} \in L$ by a shorter $\mathbf{v} - \mathbf{w}$ for some $\mathbf{w} \in L$.

Throughout this paper, we assume a lattice sieve will never return both $\mathbf{w}$ and $-\mathbf{w}$, since this optimization is used in most implementations, and accounting for a factor of 2 in the number of dual vectors is relevant for the dual attack, cf. [DP23, App. A.4].

To be able to work with the output of a lattice sieve in our analysis, we will make a heuristic assumption about the distribution of the lattice vectors that a lattice sieve algorithm outputs.

Heuristic 2 *Given a unit-volume lattice Λ, the output distribution of a lattice sieve with a saturation radius of $r_{\text{sat}} \geq 1$ and a saturation ratio of $f_{\text{sat}} \in (0, 1]$, is a list of vectors $\mathcal{W} \subset \mathbb{R}^n$ of size $N = \frac{1}{2} f_{\text{sat}} r_{\text{sat}}^n$, where its elements are independently sampled uniformly at random from the ball of radius $r_{\text{sat}}\mathrm{GH}(n)$.*

This heuristic makes two simplifications on the output of a lattice sieve. The first simplification made in the heuristic is that not much changes when going from a list of N vectors that are the output of a sieve, to a list of N vectors that are each sampled uniformly from $\Lambda \cap r_{\text{sat}} \mathrm{GH}(n)\, \mathcal{B}^n$. Although the heuristic allows for duplicates, many distinct values will be sampled, which gives a motivation for this simplification. As an illustration, consider the following: sampling n times uniformly from a set of size n yields in expectation a set of size $n(1 - 1/e) \approx 0.63n$ as $n \to \infty$.[2]

The second simplification is that the lattice structure $\mathcal{W} \subset \Lambda$ is ignored in the output. The Gaussian Heuristic predicts that the norm of the lattice vectors follows a similar distribution as the norm distribution of points uniformly from the ball. When running a sieve on a random lattice, this assumption seems fair, because we do not expect the lattice to be distorted in any particular direction.

Note that this heuristic does not assume all vectors are of the same length $r_{\text{sat}}\sqrt{n}$, as done in e.g. [GJ21]. Instead the heuristic predicts there may be some shorter vectors $\mathbf{w}$, albeit with smaller probability, but still most of the vectors are close to the boundary of the ball. The very short vectors are more beneficial in the dual attack than longer vectors, so the analysis will be more conservative when taking shorter vectors into account.

Normally, a lattice sieve is run with a saturation radius of $r_{\text{sat}} = \sqrt{4/3}$, because in that case enough vectors are initially generated to reduce vectors in the sieve in each step [NV08]. Although in theory, one sometimes assumes a lattice sieve finds *all* lattice vectors inside a ball of radius $r_{\text{sat}} \mathrm{GH}(n)$ (i.e. $f_{\text{sat}} = 1$), in practice a much lower saturation ratio is chosen for efficiency. For instance, the `G6K` software [ADH+19] uses a saturation ratio of 0.5 by default, and even lower saturation ratios of 0.375 were used for sieves used to break certain SVP challenges with GPUs [DSvW21].

2.2 Fourier Transformation

A function $f \colon \mathbb{R}^n \to \mathbb{R}$ is called "*square integrable*" whenever $\int_{\mathbb{R}^n} f(\mathbf{x})^2 \mathrm{d}\mathbf{x}$ is a finite value.

Definition 2. *The* Fourier transform *of a square-integrable function* $f \colon \mathbb{R}^n \to \mathbb{R}$ *is given by*

$$\widehat{f}(\mathbf{y}) = \int_{\mathbb{R}^n} e^{-2\pi i \langle \mathbf{x}, \mathbf{y} \rangle} f(\mathbf{x}) \mathrm{d}\mathbf{x},$$

for all $\mathbf{y} \in \mathbb{R}^n$.

For example, the Fourier transform of ρ_s is $\widehat{\rho_s} = s^n \rho_{1/s}$. The well-known Fourier inversion theorem states that, under certain convergence conditions, one may recover f from $\widehat{f}$ by:

$$f(\mathbf{x}) = \int_{\mathbb{R}^n} e^{2\pi i \langle \mathbf{x}, \mathbf{y} \rangle} \widehat{f}(\mathbf{y}) \mathrm{d}\mathbf{y}.$$

[2] One can easily derive this result using linearity of expectation.

When f is a *radial function*, i.e. there exists $g\colon \mathbb{R} \to \mathbb{R}$ such that $f(\mathbf{x}) = g(\|\mathbf{x}\|)$, then $\widehat{f}$ is also a radial function, and vice versa [SW71, Thm. 3.3].

Theorem 1 (Poisson Summation Formula, [SW71, Chapter VII]). *Given a full-rank lattice Λ and a function $f\colon \mathbb{R}^n \to \mathbb{R}$ satisfying certain decaying conditions, for all $\mathbf{t} \in \mathbb{R}^n$ one has*

$$\sum_{\mathbf{v}\in\Lambda} f(\mathbf{v}+\mathbf{t}) = \frac{1}{\det(\Lambda)} \sum_{\mathbf{w}\in\Lambda^\vee} e^{2\pi i\langle \mathbf{t},\mathbf{w}\rangle} \widehat{f}(\mathbf{w}).$$

Bessel functions occur very naturally in the context of Fourier transformations, in particular as the Fourier transformation of the indicator function of a sphere or a ball.

Lemma 2. *The Fourier transform of $f = \mathbf{1}_{r\mathcal{B}^n}$ is given by*

$$\widehat{f}(\mathbf{x}) = \left(\frac{r}{\|\mathbf{x}\|}\right)^{\frac{n}{2}} \cdot J_{n/2}(2\pi r\|\mathbf{x}\|) = \mathrm{Vol}_n(r\mathcal{B}^n)\cdot \xi\Big(\frac{n}{2}, r\|\mathbf{x}\|\Big).$$

The proof is based on [Fis13].

Proof. First, since f is radial, $\widehat{f}$ is also radial. Hence, let us do the proof only for $\mathbf{x} = (s, 0, \dots, 0)$. Then,

$$\begin{aligned}\widehat{f}(\mathbf{x}) &= \int_{r\mathcal{B}^n} e^{-2\pi i s\mathbf{y}_1}\mathrm{d}\mathbf{y} = \int_{-r}^{r} e^{-2\pi i s t}\mathrm{Vol}_{n-1}\Big(\sqrt{r^2-t^2}\mathcal{B}^{n-1}\Big)\,\mathrm{d}t \\ &= \frac{r^n\pi^{\frac{n-1}{2}}}{\Gamma\big(\frac{n}{2}+\frac{1}{2}\big)}\int_{-1}^{1} e^{2\pi i r s u}(1-u^2)^{\frac{n-1}{2}}\mathrm{d}u = \Big(\frac{r}{s}\Big)^{n/2} J_{\frac{n}{2}}(2\pi r s).\end{aligned}$$

For the last equality, note that we have

$$\begin{aligned}\left(\frac{r}{\|\mathbf{x}\|}\right)^{\frac{n}{2}} J_{\frac{n}{2}}(2\pi r\|\mathbf{x}\|) &= \left(\frac{r}{\|\mathbf{x}\|}\right)^{\frac{n}{2}} \frac{(\pi r\|\mathbf{x}\|)^{\frac{n}{2}}}{\Gamma\big(\frac{n}{2}+1\big)}\xi\Big(\frac{n}{2}, r\|\mathbf{x}\|\Big) \\ &= \mathrm{Vol}_n(r\mathcal{B}^n)\,\xi\Big(\frac{n}{2}, r\|\mathbf{x}\|\Big).\end{aligned}$$

2.3 Dual Attack

The hard problem that we consider is the so-called Bounded Distance Decoding problem.

Definition 3. *Let $\chi\colon \mathbb{R}^n \to [0,1]$ be some distribution and $\Lambda \subset \mathbb{R}^n$ be a full-rank lattice.*

- Decision χ-BDD problem *is the problem of deciding correctly (with high probability) whether an input* $\mathbf{t}$ (mod Λ) *is sampled from χ or $U(\mathbb{R}^n/\Lambda)$.*
- Search χ-BDD problem *is the problem of finding $\mathbf{v} \in \Lambda$ on input $\mathbf{t} = \mathbf{v} + \mathbf{e}$ and Λ, where $\mathbf{v} \in \Lambda$ and $\mathbf{e} \leftarrow \chi$.*

The distribution χ is usually chosen as one that is concentrated around zero, for example a uniform distribution on a sphere or ball of some radius r, or a gaussian of some width s. For these distributions, the hardness of the problem depends on the radius or width, because taking this value too small makes the problems easy to solve, while taking it too large makes the decision-BDD problem impossible to solve [DP23, Sec. 4.1].

Decision-BDD. The most common way to solve decision-BDD is to use a *distinguisher*, i.e. a Λ-periodic function $f\colon \mathbb{R}^n \to \mathbb{R}$ for which $f(\mathbf{t})$ is large when $\mathbf{t} \leftarrow \chi$, while $f(\mathbf{t}) \approx 0$ whenever $\mathbf{t} \leftarrow \mathbb{R}^n/\Lambda$. One can then decide correctly from which distribution $\mathbf{t}$ comes based on $f(\mathbf{t})$ with a high success probability.

For example, the periodic gaussian, i.e. $\sum_{\mathbf{v}\in\Lambda} \rho_s(\mathbf{v}+\mathbf{t})$, would be a very good distinguisher, since the gaussian decays rapidly, but evaluating this function is hard. However, using Theorem 1 we can approximate the periodic gaussian with the following identity.

$$\sum_{\mathbf{v}\in\Lambda} \rho_s\left(\mathbf{v}+\mathbf{t}\right) = \frac{s^n}{\det(\Lambda)} \sum_{\mathbf{w}\in\Lambda^\vee} \rho_{1/s}(\mathbf{w}) e^{2\pi i \langle \mathbf{t},\mathbf{w}\rangle}.$$

Now, by summing the right hand side for a large set $\mathcal{W}$ of short dual vectors, one can efficiently compute a decent approximation of the above, allowing one to solve decision-BDD with high probability [AR04].

One can use the output of a lattice sieve algorithm on $\Lambda^\vee$ [ADPS16] to efficiently obtain such a large set $\mathcal{W}$. In such a situation, all the nonzero dual vectors have approximately the same weight, and [LW21] shows that the score function

$$f_{\mathcal{W}}(\mathbf{t}) = \sum_{\mathbf{w}\in\mathcal{W}} \cos\left(2\pi \left\langle \mathbf{w}, \mathbf{t}\right\rangle\right), \tag{1}$$

which is easier to compute, works asymptotically almost as good as the [AR04] original score function $\sum_{\mathbf{w}\in\mathcal{W}} \rho_{1/s}(\mathbf{w}) \cos(2\pi \langle \mathbf{w}, \mathbf{t}\rangle)$ for distinguishing between BDD and uniform targets.

By linearity of expectation we get $\mathbb{E}_{\mathbf{t}}\left[f_{\mathcal{W}}(\mathbf{t})\right] = 0$ for uniform targets $\mathbf{t}$. On the other hand, for BDD targets $\mathbf{t}$, the expected score is approximately $\xi(n/2, r_d \left\|\mathbf{t}-\mathbf{v}_1\right\|)$ where $\mathbf{t}$ is closest to $\mathbf{v}_1 \in \Lambda$.

Search-BDD to Decision-BDD. Recent dual attacks [EJK20,GJ21,MAT22] and all follow-up works reduce search-BDD on Λ to decision-BDD on a sublattice as follows. First, a sparsification $\Lambda' \subset \Lambda$ is chosen. Then decision-BDD is solved on the sublattice Λ' with the targets $(\mathbf{t}-\mathbf{g})_{\mathbf{g}\in\Lambda/\Lambda'}$. If decision-BDD was successful, the solution $\mathbf{v} \in \Lambda$ to the search problem is solved modulo Λ', i.e. $\mathbf{v} \equiv \mathbf{g}' \pmod{\Lambda'}$, where $\mathbf{t}-\mathbf{g}'$ was a BDD target according to the decision-BDD solver. Having found the "correct guess" $\mathbf{g}'$, one can then solve the easier search-BDD problem on Λ' with target $\mathbf{t}-\mathbf{g}'$, and after some iterations, $\mathbf{v}$ is finally found.

However, to solve decision-BDD, many of the recent dual attacks (for example [GJ21,MAT22] and follow-up works) run a sieve on a lattice $L \subset (\Lambda')^\vee$ of much lower rank, from which a set of dual vectors $\mathcal{W} \subset L$ is obtained. Observe that the distinguisher function now satisfies

$$f_{\mathcal{W}}(\mathbf{t}) = f_{\mathcal{W}}(\pi_{\mathrm{span}(L)}(\mathbf{t})),$$

where π_V denotes the projection map onto the vector space V. This shows that f is now merely a distinguisher for $L^\vee$: it assigns high scores to all targets $\mathbf{t} \in \mathbb{R}^n$ which have a projection $\pi_{\mathrm{span}(L)}(\mathbf{t})$ close to $L^\vee$.

More specifically, in the dual attacks first some lattice reduction is performed on Λ' to obtain some (BKZ-reduced) basis $\mathbf{D}$ for $(\Lambda')^\vee$. Then, L is chosen to be the lattice generated by the first k (column) vectors of $\mathbf{D}$. By duality, $L^\vee$ is then obtained by projecting Λ' away from the last $n-k$ (column) vectors of $\mathbf{D}^{-\mathsf{T}}$. Note, when $\mathbf{e}$ follows an n-dimensional gaussian distribution of width s, then $\pi_{\mathrm{span}(L)}(\mathbf{e})$ also follows a k-dimensional gaussian distribution of width s.

On the other hand, suppose we are given a target $\mathbf{t} = \mathbf{v} + \mathbf{e}$ and we have a guess $\mathbf{g} \notin \mathbf{v} + \Lambda'$. Then, the reduction from search-BDD to decision-BDD yields a target $\mathbf{t} - \mathbf{g} \equiv \mathbf{e} + (\mathbf{v} - \mathbf{g}) \pmod{\Lambda'}$ in the lattice Λ'. In that case, the score $f_{\mathcal{W}}(\mathbf{t} - \mathbf{g})$ is high when the point

$$\pi_{\mathrm{span}(L)}(\mathbf{e}) + \pi_{\mathrm{span}(L)}(\mathbf{v} - \mathbf{g}), \tag{2}$$

is close to $L^\vee$. Since the lower rank lattice L was only picked after some BKZ reduction, and $\mathbf{v}-\mathbf{g}$ is nonzero, we presume that $\pi_{\mathrm{span}(L)}(\mathbf{v}-\mathbf{g})$ acts as a uniform point in $\mathrm{span}\,(L)\,/L$. Now, because the distinguisher is only distinguishing for $L^\vee$, we see that the target in (2) corresponds to a BDD sample if $\mathbf{g} \in \mathbf{v} + \Lambda'$, and else to a uniform sample modulo $L^\vee$.

We note that the work of [PS23a] sidestepped the uniform model for incorrect targets, by making the a-priori assumption that $\|\mathbf{e}\| < \frac{1}{2}\lambda_1(\Lambda)$; in that case the incorrect target can be proved to be far enough from the lattice in the worst-case, without any statistical nor heuristic argument (as it was already the case in [AR04]). This is however not the regime used in recent concrete attack claims [GJ21,MAT22], and is in fact separated from the contradictory regime of [DP23] by a factor 2 on the error length $\|\mathbf{e}\|$.

3 Score Distribution Models

Experiments in [DP23] reveal that the Independence Heuristic, which is used in [GJ21,MAT22], leads to an incorrect prediction for the score distribution of both BDD and uniform targets. In particular, it leads to an overestimation of the probability that a BDD target is distinguished from uniform targets. More precisely, there are two issues with the modelling of the score, based on the Independence Heuristic:

- Uniform targets have a much larger probability to have a high score than what is predicted under the Independence Heuristic. The root cause is that the probability to get a high score because a uniform point may be close to the lattice, is not considered.

- The score distribution for gaussian BDD targets was shown to not be normally distributed, because the median score was significantly smaller than the mean score. Moreover, the variance of the score was much larger than predicted under assumption of the Independence Heuristic.

The goal of this section is to develop new score predictions that match practice more accurately than what was previously achieved under the Independence Heuristic.

In Subsection 3.1, we derive analytic expressions for the mean and variance of $f_{\mathbf{u}}(\mathbf{v})$. This is the first dual attack paper giving an exact expression for this. In Subsection 3.2, we get the exact mean and variance for the score distribution of spherical BDD targets, based on Heuristic 2. Here, we present a new heuristic that says the spherical targets *do* give a normally distributed score distribution. This then allows one to derive score distributions for different target distributions, such as uniform in a ball and gaussian, which is done in Subsection 3.3. In particular, we use the prediction for errors uniformly from a ball, to arrive at a prediction for uniform errors, because the Voronoi cell can be approximated by a ball of volume $\det(\Lambda)$.

Throughout this section, we will assume $\det(\Lambda) = 1$ without loss of generality. Note that this is not a restriction on the class of lattices to which the model applies, because one can reduce to a unit-volume lattice by rescaling Λ.

3.1 Individual Score Function

First, let us look at the score distribution $f_{\mathbf{u}}(\mathbf{v}) = \cos(2\pi \langle \mathbf{u}, \mathbf{v} \rangle)$, when given some fixed vector $\mathbf{u} \in \mathbb{R}^n$. The score distribution for points uniform on a sphere can be easily numerically computed with the following lemma.

Lemma 3. *Given a fixed* $\mathbf{u} \in \mathbb{R}^n$*, when the random variable* $\mathbf{v}$ *follows the uniform distribution over the sphere of radius* r*, we have*

$$\mathbb{E}\left[f_{\mathbf{u}}(\mathbf{v})\right] = \xi\left(\frac{n}{2} - 1, \|\mathbf{u}\|\, r\right),$$
$$\mathbb{V}\left[f_{\mathbf{u}}(\mathbf{v})\right] = \frac{1}{2} + \frac{1}{2}\xi\left(\frac{n}{2} - 1, 2\|\mathbf{u}\|\, r\right) - \xi\left(\frac{n}{2} - 1, \|\mathbf{u}\|\, r\right)^2.$$

Proof. The expectation value follows directly from a classic result from Fourier Analysis, see e.g. [GS64, p. 198] or [SW71, p. 154], as we have

$$\mathbb{E}\left[f_{\mathbf{u}}(\mathbf{v})\right] = \frac{1}{\mathrm{Vol}_n(r\mathcal{S}^{n-1})} \int_{r\mathcal{S}^{n-1}} e^{2\pi i \cdot \langle \mathbf{u}, \mathbf{v} \rangle} \mathrm{d}\mathbf{v} = \frac{\Gamma(\frac{n}{2}) \cdot J_{\frac{n}{2}-1}\left(2\pi \|\mathbf{u}\|\, r\right)}{\left(\pi \|\mathbf{u}\|\, r\right)^{\frac{n}{2}-1}}.$$

Making use of the trigonometric identity $f_{\mathbf{u}}(\mathbf{v})^2 = \frac{1}{2} + \frac{1}{2}\cos\left(4\pi \langle \mathbf{u}, \mathbf{v} \rangle\right) = \frac{1}{2} + \frac{1}{2} f_{2\mathbf{u}}(\mathbf{v})$, the variance is then given by,

$$\mathbb{V}\left[f_{\mathbf{u}}(\mathbf{v})\right] = \mathbb{E}\left[f_{\mathbf{u}}(\mathbf{v})^2\right] - \mathbb{E}\left[f_{\mathbf{u}}(\mathbf{v})\right]^2 = \frac{1}{2} + \frac{1}{2}\mathbb{E}\left[f_{2\mathbf{u}}(\mathbf{v})\right] - \mathbb{E}\left[f_{\mathbf{u}}(\mathbf{v})\right]^2.$$

Remark 3. Note that spherical errors were studied before [LW21, Section 5.1]. However, there they actually approximate an error uniformly from the $(n-1)$-dimensional sphere as a sample from the gaussian with parameter $\sigma = n^{-1/2}$. Here, we give exact expressions. For small arguments $0 < x \ll \sqrt{\alpha+1}$, the Bessel function has the approximation $J_\alpha(x) \sim (x/2)^\alpha / \Gamma(\alpha+1)$, which leads to the same approximations for the expectation and variance.

This result can be translated easily to a uniform distribution over a ball, because sampling uniformly from the n-dimensional ball can be done by first sampling from a $(n+2)$-dimensional sphere and then dropping the last two coordinates [VGS17].

Corollary 1. *Given a fixed $\mathbf{u} \in \mathbb{R}^n$, when the random variable $\mathbf{v}$ follows the uniform distribution over the ball of radius r, we have*

$$\begin{aligned}
\mathbb{E}\left[f_{\mathbf{u}}(\mathbf{v})\right] &= \xi\Big(\frac{n}{2}, \|\mathbf{u}\|\, r\Big),\\
\mathbb{V}\left[f_{\mathbf{u}}(\mathbf{v})\right] &= \frac{1}{2} + \frac{1}{2}\xi\Big(\frac{n}{2}, 2\,\|\mathbf{u}\|\, r\Big) - \xi\Big(\frac{n}{2}, \|\mathbf{u}\|\, r\Big)^2.
\end{aligned}$$

Let us now look into three possible error distributions: uniform from a sphere, uniform from a ball and gaussian.

3.2 Error Uniform from a Sphere

Assuming Heuristic 2, we can now use the above corollary with all the dual vectors that one gets by running a sieve, since these are assumed to be uniform in the ball of radius $r_d = r_{\text{sat}}\text{GH}(n)$. Then the independence of the dual vectors from the sieve allows us to say something about the score distribution. This corollary allows us to model the score $f_{\mathcal{W}}(\mathbf{t})$, when we have dual vectors $\mathcal{W}$ from a sieve.

Consider an error $\mathbf{t}$ uniformly sampled from a sphere of a radius r_p, and dual vectors acquired from a sieve with saturation radius r_{sat}. Under assumption of Heuristic 2, we can apply Corollary 1 on the fixed $\mathbf{t}$ and with random variable $\mathbf{w}$ for each dual vector $\mathbf{w} \in \mathcal{W}$ that is assumed to be an i.i.d. sample from $U(r_d\mathcal{B}^n)$. Then, as there are N dual vectors, the expected score is given by

$$E_S(r_p) = N \underset{\mathbf{w} \leftarrow U(r_d\mathcal{B}^n)}{\mathbb{E}}\left[f_{\mathbf{w}}(\mathbf{t})\right] = N\xi\Big(\frac{n}{2}, r_p \cdot r_d\Big),$$

and the variance is

$$V_S(r_p) = N\left(\frac{1}{2} + \frac{1}{2}\xi\Big(\frac{n}{2}, 2r_p r_d\Big) - \xi\Big(\frac{n}{2}, r_p r_d\Big)^2\right).$$

Based on experiments, we propose the following heuristic.

Heuristic 3 *Fix a set $\mathcal{W}$ of dual vectors from a lattice sieve. Then, errors $\mathbf{t} \leftarrow U\big(r_d\mathcal{S}^{n-1}\big)$ have a score distribution $f_{\mathcal{W}}(\mathbf{t})$ that is* gaussian *with mean $E_S(r_p)$*

and variance $V_S(r_p)$. In particular, the score distribution is described by the following CDF,

$$\underset{\mathbf{t} \leftarrow U(r_p \mathcal{S}^{n-1})}{\mathbb{P}} [f_{\mathcal{W}}(\mathbf{t}) \leq x] = \frac{1}{2} + \frac{1}{2} \operatorname{erf}\left(\frac{x - E_S(r_p)}{\sqrt{2 V_S(r_p)}} \right). \tag{3}$$

Heuristic Justification. By combining the linearity of expectation with Corollary 1, for a particular $\mathbf{t}$, the mean and variance of $f_{\mathcal{W}}(\mathbf{t})$ are respectively $E_S(r_p)$ and $V_S(r_p)$, taken over the randomness of the dual vectors.

Since for all $\mathbf{t} \in r_p \mathcal{B}^n$, we have $\|\mathbf{t}\| = r_p$, the mean and variance are the same for all $\mathbf{t} \in r_p \mathcal{B}^n$. Therefore, the score distribution over the randomness of $\mathbf{t}$, is expected to be gaussian by a central limit heuristic.

Interestingly, the expected score is very similar to the score when one takes $\mathcal{W} = r_d \mathcal{B}^n \cap \Lambda^\vee$. In this case, Theorem 1 and Lemma 2 yield the following:

$$f_{\mathcal{W}}(\mathbf{t}) = \sum_{\mathbf{w} \in \Lambda^\vee} \mathbf{1}_{r_d \mathcal{B}^n}(\mathbf{w})\, e^{2\pi i \langle \mathbf{w}, \mathbf{t} \rangle} = \operatorname{Vol}_n(r_d \mathcal{B}^n) \cdot \sum_{\mathbf{v} \in \Lambda} \xi\Big(\frac{n}{2}, r_d \|\mathbf{v} + \mathbf{t}\|\Big).$$

For $r_p \ll \mathrm{GH}(n)$, the vector $\mathbf{v} = \mathbf{0}$ gives a main contribution of $\xi\left(\frac{n}{2}, r_d r_p\right)$ to the summation, because $\xi(n/2, -)$ is a rapidly decaying function.

At this point, one could argue that one still assumes a central limit heuristic to obtain the score distribution. However, the crucial difference between Heuristic 3 and the Independence Heuristic of [DP23], is that Heuristic 3 takes a central limit of *i.i.d.* $f_{\mathcal{W}}(\mathbf{t})$ over the choice of $\mathbf{t} \in r_p \mathcal{B}^n$. However, irrespective of the target distribution, the Independence Heuristic uses a central limit heuristic for the score $f_{\mathcal{W}}(\mathbf{t})$, and neglects the dependence of $\|\mathbf{t}\|$ on the mean score. This makes Heuristic 3 a sensible heuristic, although large experiments are ultimately needed to gain confidence in this heuristic. This can be found in Section 4.2.

Lastly, the following result shows that there is a limit to the radius of a sphere where reasonable distinguishing can still be expected.

Lemma 4. *Given a set $\mathcal{W}$ of dual vectors in a ball of radius $R_d = r_{\mathrm{sat}} \mathrm{GH}(n)$, obtained according to Heuristic 2, we have $\mathbb{E}[f_{\mathcal{W}}(\mathbf{t})] > 0$ as $n \to \infty$ when a target is drawn uniformly from the sphere of radius $R_p = r_p \mathrm{GH}(n)$ satisfying $R_p R_d \leq \frac{n}{4\pi}$ (e.g. when $r_p \cdot r_{\mathrm{sat}} \leq \frac{e}{2}$).*

Proof. First note that $r_p \cdot r_{\mathrm{sat}} \leq \frac{e}{2}$ implies we have

$$R_p \cdot R_d = r_p r_{\mathrm{sat}} \mathrm{GH}(n)^2 \leq r_p r_{\mathrm{sat}} \cdot \frac{n}{2\pi e} \leq \frac{n}{4\pi}.$$

Based on Heuristic 2 and Corollary 1, we get $\mathbb{E}[f_{\mathcal{W}}(\mathbf{t})] = N \xi\big(\frac{n}{2}, R_p R_d\big) > 0$, since the first zero of $\xi(\frac{n}{2}, x)$ is at $\frac{1}{2\pi} j_{n/2,1} > \frac{n}{4\pi}$, using Lemma 1.

Note that this Corollary is sharp, i.e. if you take $r_p > \frac{e}{2 r_{\mathrm{sat}}}$, there exists some $n \in \mathbb{N}$ such that $R_p R_d > \frac{1}{2\pi} j_{n/2,1}$, resulting in a negative expected score.

3.3 Radial Error Distributions

The result for spherical errors can now be extended to any error distribution χ that is radial. Consider the case where we have the following PDF $f(r)$ indicating the density function of getting a particular radius r when $\mathbf{e}$ is sampled from some radial distribution. Note that we have,

$$f(r) = \frac{\mathrm{d}}{\mathrm{d}r} \underset{\mathbf{t} \leftarrow \chi}{\mathbb{P}} \left[\|\mathbf{t}\| \leq r \right].$$

Based on Heuristic 3 we arrive at the following claim.

Heuristic Claim 1 *Fix a set $\mathcal{W}$ of dual vectors from a lattice sieve, and a radial error distribution χ, such that the norm distribution of samples from χ has PDF $f(r)$ at radius $r \geq 0$. Then, errors $\mathbf{t} \leftarrow \chi$ have a score distribution $f_{\mathcal{W}}(\mathbf{t})$ with the following CDF:*

$$\underset{\mathbf{t} \leftarrow \chi}{\mathbb{P}} \left[f_{\mathcal{W}}(\mathbf{t}) \leq x \right] = \frac{1}{2} + \frac{1}{2} \int_0^\infty \operatorname{erf}\left(\frac{x - E_S(r)}{\sqrt{2 V_S(r)}} \right) f(r) \mathrm{d}r. \tag{4}$$

There are now two radial distributions of particular interest: gaussian and the uniform distribution on the ball.

Gaussian Error. Let us use the score distribution for spherical errors as a stepping stone for that of gaussian errors. The norm distribution of samples from $\mathcal{N}(0,1)^n$, follows the χ-distribution,[3] which has a PDF given by:

$$f(r; n) = \frac{r^{n-1} \exp\left(-\frac{r^2}{2}\right)}{2^{\frac{n}{2}-1} \Gamma\left(\frac{n}{2}\right)} \qquad (r \in \mathbb{R}_{>0}).$$

Now consider the error distribution $\mathcal{N}\big(0, \sigma^2\big)^n$ for some $\sigma > 0$. In this case, under Heuristic 3, the score distribution is as follows for gaussian errors.

Heuristic Claim 2 *Fix some $\sigma > 0$, and a set $\mathcal{W}$ of dual vectors from a lattice sieve. Then, errors $\mathbf{t} \leftarrow \mathcal{N}\big(0, \sigma^2\big)^n$ have a score distribution $f_{\mathcal{W}}(\mathbf{t})$ with the following CDF:*

$$\underset{\mathbf{t} \leftarrow \mathcal{N}(0,\sigma)^n}{\mathbb{P}} \left[f_{\mathcal{W}}(\mathbf{t}) \leq x \right] = \frac{1}{2} + \frac{1}{2} \int_0^\infty \operatorname{erf}\left(\frac{x - E_S(r)}{\sqrt{2 V_S(r)}} \right) \cdot f\Big(\frac{r}{\sigma}; n\Big) \frac{\mathrm{d}r}{\sigma}. \tag{5}$$

Note that the χ-distribution has most weight concentrated around $\sigma\sqrt{n}$, so the best numerical approximations are obtained when the numerical integration is giving special attention to the region around $\sigma\sqrt{n}$ (see Section 4.1).

[3] More well-known is the PDF of the square norm, given by the χ^2-distribution.

Error Uniform from a Ball. Consider the distribution $\chi = U(r_p\mathcal{B}^n)$ for some $r_p > 0$. Here, $\mathbb{P}_{\mathbf{t}\leftarrow\chi}\left[\|\mathbf{t}\| \le x\right] = x^n/r_p^n$ for any $x \in [0, r_p]$, and equals 1 for $x \ge r_p$. Strictly speaking, this CDF is not differentiable at r_p, so the PDF is not defined there. Still, one can mitigate this issue by integrating until r_p in (4). This yields the following claim.

Heuristic Claim 3 *Fix some $r_p > 0$, and a set $\mathcal{W}$ of dual vectors from a lattice sieve. Then, errors $\mathbf{t} \leftarrow U(r_p\mathcal{B}^n)$ have a score distribution $f_{\mathcal{W}}(\mathbf{t})$ with the following CDF:*

$$\mathop{\mathbb{P}}_{\mathbf{t}\leftarrow U(r_p\mathcal{B}^n)}\left[f_{\mathcal{W}}(\mathbf{t}) \le x\right] = \frac{1}{2} + \frac{1}{2}\int_0^{r_p} \operatorname{erf}\left(\frac{x - E_S(r)}{\sqrt{2V_S(r)}}\right) \cdot \frac{nr^{n-1}\mathrm{d}r}{r_p^n}. \tag{6}$$

3.4 Error Uniform Modulo Lattice

Given N dual vectors, prior analysis on the score distribution of targets uniform in the torus $\mathbb{R}^n/\Lambda$ modelled it as a gaussian with mean 0 and variance $\frac{1}{2}N$. However, when a dual attack is in the contradictory regime of [DP23], the "floor" phenomenon should be taken into account. This means that one cannot assume the score distribution is normally distributed, as the probability of getting a high score is much higher than what a normal gaussian tail would say.

The higher score is mainly driven by the high average score for points that lie close to the lattice, and therefore as a fix, we consider the probability of a point $\mathbf{t}$ uniform modulo Λ to be at some distance r from Λ, and use the score distribution of the sphere of radius r for such points. In particular, by Heuristic 2, the rotational invariance of the dual vectors allows us to focus on the distribution of the norm of uniform points in the Voronoi cell $\mathcal{V}(\Lambda)$. Specifically, if we would know the function

$$F(r) = \mathop{\mathbb{P}}_{\mathbf{t}\leftarrow U(\mathbb{R}^n/\Lambda)}\left[d(\mathbf{t}, \Lambda) \le r\right] = \mathrm{Vol}_n(\mathcal{V}(\Lambda) \cap r\mathcal{B}^n)\,, \tag{7}$$

for any lattice Λ, we would be able to make a prediction for scores of uniform targets by integrating the score distribution of points uniformly from a sphere over the radius r, ranging from 0 to $\mu(\Lambda)$, where $\mu(\Lambda)$ is the covering radius of the lattice. Note that we have $\frac{1}{2}\lambda_1\mathcal{B}^n \subseteq \mathcal{V}(\Lambda) \subseteq \mu(\Lambda)\mathcal{B}^n$. Hence, $F(r) = \mathrm{Vol}_n(r\mathcal{B}^n)$ for $r \le \frac{1}{2}\lambda_1$. For radii $r \in (\frac{1}{2}\lambda_1(\Lambda), \mu(\Lambda))$, there is no easy expression for $f(r)$ because the ball of radius r is not necessarily contained in the Voronoi cell. However, we still have the upper bound,

$$F(r) \le \mathrm{Vol}_n(r\mathcal{B}^n)\,. \tag{8}$$

In [DP23, Heuristic Claim 4], the following equation is derived from the Gaussian Heuristic, which holds for all $r \in (0, \mathrm{GH}(n))$:

$$F(r) = \mathrm{Vol}_n(r\mathcal{B}^n)\left(1 - n^{O(1)} \cdot \mathrm{Vol}_n(r\mathcal{B}^n)\right). \tag{9}$$

The upper bound in (8) can thus be seen as a first order approximation of $F(r)$. The following heuristic implies $F(r)$ equals the first order approximation of (9).

Heuristic 4 *Let $\mathit{\Lambda} \subset \mathbb{R}^n$ be a random full-rank (unit-volume) lattice. Then,*

$$\mathcal{V}(\mathit{\Lambda}) = \mathrm{GH}(n) \cdot \mathcal{B}^n.$$

Based on Heuristics 3 and 4, we get the following claim by taking $r_p = \mathrm{GH}(n)$ in Heuristic Claim 3:

Heuristic Claim 4 *Fix a set $\mathcal{W}$ of dual vectors from a lattice sieve. Then, errors $\mathbf{t} \leftarrow U(\mathbb{R}^n/\mathit{\Lambda})$ have a score distribution $f_{\mathcal{W}}(\mathbf{t})$ with the following CDF:*

$$\underset{\mathbf{t}\leftarrow U(\mathbb{R}^n/\mathit{\Lambda})}{\mathbb{P}}[f_{\mathcal{W}}(\mathbf{t}) \leq x] = \frac{1}{2} + \frac{1}{2}\int_0^{\mathrm{GH}(n)} \mathrm{erf}\left(\frac{x - E_S(r)}{\sqrt{2V_S(r)}}\right) \cdot \frac{nr^{n-1}\mathrm{d}r}{\mathrm{GH}(n)^n}. \tag{10}$$

This claim will predict the "waterfall-floor" phenomenon in the score distribution of uniform targets:

- For scores $x \approx 0$, the integrand is biggest around $r \approx \mathrm{GH}(n)$ because then $E_S(r) \approx 0$, but $E_S(r) \gg x$ when $r \ll \mathrm{GH}(n)$. Because here $V_S(r) \approx N/2$, we expect *approximately* a normal distribution around 0 with variance $N/2$, which is precisely the "waterfall" part of the score distribution.
- When x becomes somewhat large, i.e. $c\sqrt{N/2}$ for some constant c, the integrand is negligible for $r \approx \mathrm{GH}(n)$, as x is many standard deviations above the expected score for a target of radius r. Instead, the integrand is biggest around r with $E_S(r) \approx x$, because the error function is close to $\frac{1}{2}$ here. This is exactly the point where we expect the "floor phenomenon": at the point where being closer to the lattice is much more likely than predicted by a simple normal distribution. The value for r where the integrand is biggest in (10) may be somewhere in the range $(\frac{1}{2}\mathrm{GH}(n)\,, \mathrm{GH}(n))$, and strongly depends not only on x but also on the dimension.

Note here that, informally speaking, assuming Heuristic 4 incorrectly predicts a larger fraction of the points to be at distance at most some r from the lattice, instead of the actual $F(r)$. Because $E_S(r)$ is decreasing as a function of r, Heuristic Claim 4 predicts the probability to get a score above some x to be larger than what it actually is. That is, the right hand side of (10) is a lower bound for the CDF on the left hand side. This is beneficial for establishing a lower bound on the dual attack, as it upper bounds the probability of a uniform target being a false-positive, i.e. having a score above a certain threshold value.

4 Experiments

In this section, we provide further substantiation of the concrete predictions made in Section 3, in particular the predictions in (3), (6) and (5), with experimental support. With the experiments, we want to verify whether the used heuristics lead to conclusions that precisely match practice.

In the experiments, we take the full output of a lattice sieve on $\mathit{\Lambda}^\vee$, containing almost as many as $\frac{1}{2}\left(\frac{4}{3}\right)^{n/2}$ dual vectors of length at most $\sqrt{4/3}\mathrm{GH}(n)$, similar to [DP23].

We will compare three possible BDD distributions with their respective prediction from Section 3: uniform from a sphere, uniform from a ball and gaussian.

Moreover, we compare the distribution of uniform scores from [DP23], which was obtained by running extensive experiments, with the prediction that was made in Subsection 3.4 with (10).

4.1 Implementation Details.

We used the `G6K` software [ADH$^+$19] for running the experiments, using `Python` on a high-level, but with a binding to some `C/C++` code for speeding up BDD sampling.

The script `bdd_sample.py` samples the three BDD score distributions by first sampling a random q-ary lattice from a matrix of dimension $n \times n/2$ (n is only even), then running a lattice sieve to acquire $\frac{1}{2}(4/3)^{n/2}$ vectors, and finally computing scores for many samples. Here, $q = 3329$ and a saturation ratio of 0.99 was used, so $\lceil 0.495(4/3)^{n/2} \rceil$ vectors were taken from the `G6K` database, since only one of $\mathbf{w}, -\mathbf{w}$ is used.

For a gaussian sample, we simply sample $\mathbf{x} \leftarrow N(0, f_{\text{GH}} \cdot \text{GH}(n) / \sqrt{n \cdot q})$, where we have considered values of $f_{\text{GH}} \in \{0.1, 0.2, \ldots, 1.0\}$. Note the normalization $1/\sqrt{n}$ is needed to get target length $f_{\text{GH}} \cdot \text{GH}(n) / \sqrt{q}$ and the normalization $1/\sqrt{q}$ is to compensate for the scaling of the primal lattice, which has determinant $q^{n/2}$. We took $q = 3329$. For a sample uniform on the sphere, we reuse the same gaussian sample $\mathbf{x}$, and create the sample

$$\mathbf{y} = \mathbf{x} \cdot \frac{f_{\text{GH}} \cdot \text{GH}(n)}{\sqrt{nq}\, \|\mathbf{x}\|},$$

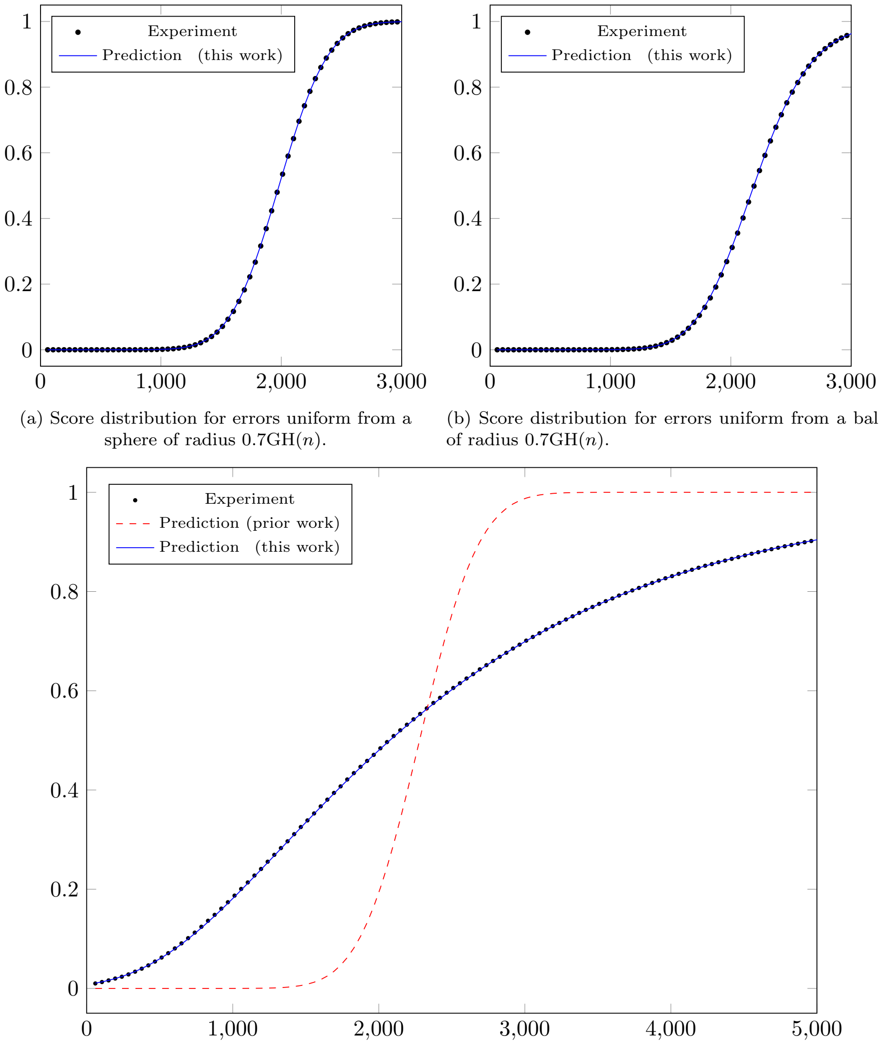
which is then uniformly on the sphere of radius $f_{\text{GH}} \cdot \text{GH}(n) / \sqrt{nq}$. For a sample uniform in the ball, we take a sample uniformly from the $n+1$-dimensional sphere in dimension $n + 2$ and drop the last two coordinates, making use of [VGS17]. We only generated 2 more gaussian samples, and reused the n from $\mathbf{x}$ here. The script `bdd_sample.py` ran to collect 100 000 samples.

This data was then used by the script `bdd_predict.py` which plotted the both experimental data, and the predictions. The predictions require evaluating the Bessel function and an integration, for which we used `hyp0f1` and `quad` respectively from the Python package `mpmath`. Specifically, the gaussian prediction was numerically more accurate when the interval $(0, \infty)$ was split into two with the split happening at the expected length.[4] The integration for the uniform ball was performed from 0.001 up to 1 times the radius of the ball, to prevent the singularity around 0.

4.2 Experiments for BDD Targets

The experiments for the BDD score are in Figure 1. It is clear that the predictions made in Section 3 give accurate estimates on the score distribution for BDD

[4] For more details, see “Highly variable functions” from the documentation (`https://mpmath.org/doc/current/calculus/integration.html`)



(a) Score distribution for errors uniform from a sphere of radius 0.7GH(n).

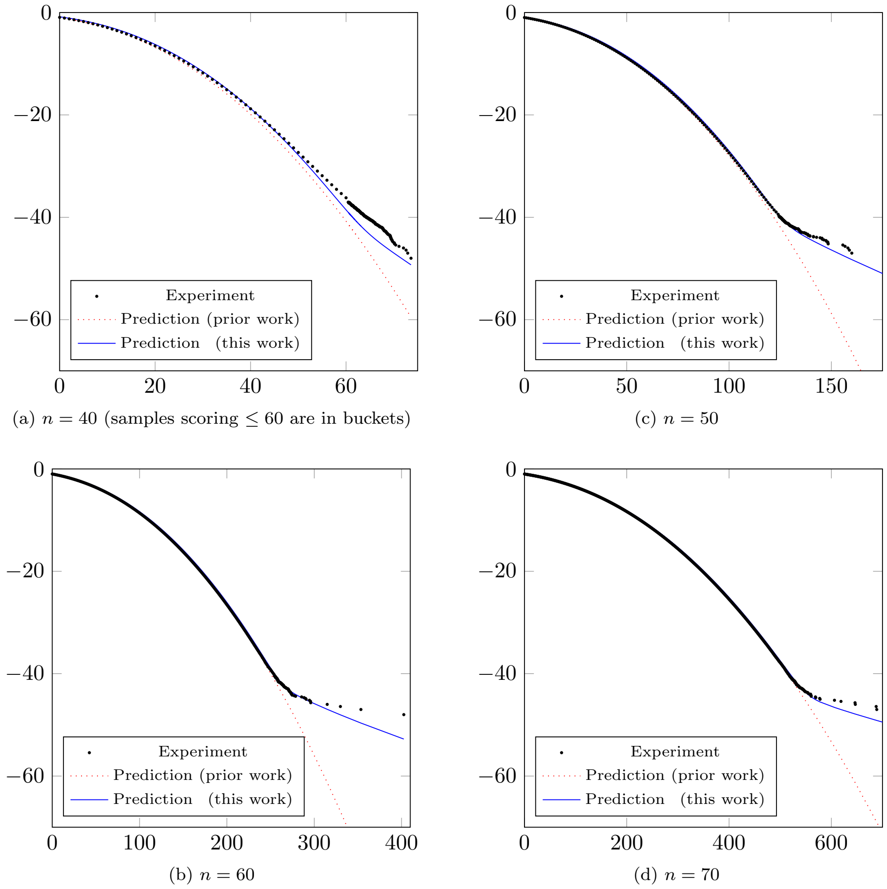
(b) Score distribution for errors uniform from a ball of radius 0.7GH(n).

(c) Score distribution for gaussian errors with $\sigma = 0.7\mathrm{GH}(n)/\sqrt{n}$.

Fig. 1. The CDF of the BDD score distribution for sphere, ball and gaussian error distributions, in dimension $n = 90$. The heuristic prediction for gaussian errors is based on the Independence Heuristic, used in prior works [EJK20,GJ21,MAT22]. Experimental data is based on 10^5 samples, and $\lceil \frac{1}{2} 0.99 \left(\frac{4}{3}\right)^{45} \rceil = 207419$ dual vectors.

targets that are from a sphere, from a ball or gaussian. It also shows again that the heuristic is completely off and enormously underestimates the probability on a low score when the target distribution is gaussian.

4.3 Experiments for Uniform Targets



(a) $n = 40$ (samples scoring ≤ 60 are in buckets)

(c) $n = 50$

(b) $n = 60$

(d) $n = 70$

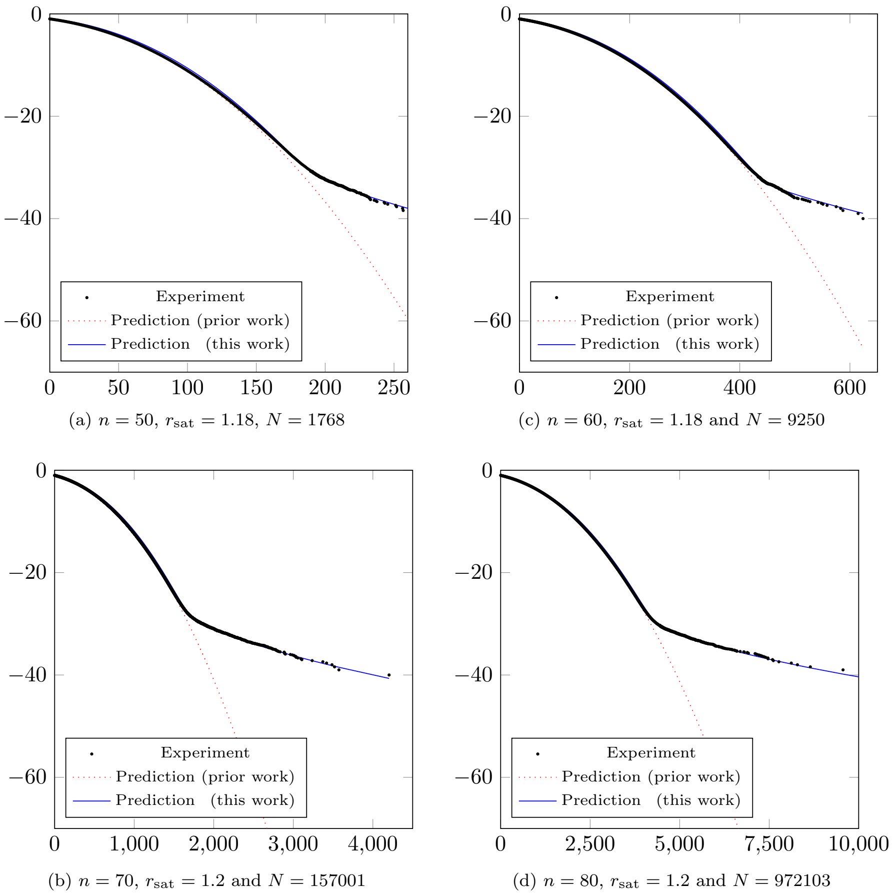
Fig. 2. Score distribution for uniform targets. x-axis: score, y-axis: $\log_2(1 - \mathrm{CDF}(x))$. Experimental data consists of $T = 2^{48}$ uniform samples, saturation radius for dual sieve was $r_{\mathrm{sat}} = \sqrt{4/3}$.

Figure 2 compares the prediction of the score distribution for uniform targets, versus experiments. Here, a saturation ratio of $f_{\mathrm{sat}} = 0.9$ and saturation radius of $r_{\mathrm{sat}} = \sqrt{4/3}$ was used. Thus, $0.5 f_{\mathrm{sat}} r_{\mathrm{sat}}^n$ (pairs of) dual vectors were used in the calculation of a score.

The experimental data for the uniform scores was acquired independently from the earlier work [DP23].

Note that in dimension $50, 60$ and 70, the uniform prediction seems to be very close to the experimental data. The right tail of the experimental data depends on extremely rare events (happening once in 2^{48} trials), so a small number of the most-right data points can change a little bit in another run. We believe if more samples were taken, the experimental data would get closer to the prediction, however, the experiment would have to run for multiple days in that case on our server.

In dimension 40, it seems that the experiments give scores higher than expected by our prediction. Because 142 dual vectors were used, we believe that this is only a low-dimensional phenomenon.



(a) $n = 50$, $r_{\text{sat}} = 1.18$, $N = 1768$

(c) $n = 60$, $r_{\text{sat}} = 1.18$ and $N = 9250$

(b) $n = 70$, $r_{\text{sat}} = 1.2$ and $N = 157001$

(d) $n = 80$, $r_{\text{sat}} = 1.2$ and $N = 972103$

Fig. 3. Score distribution for uniform targets, with different saturation radius. x-axis: score, y-axis: $\log_2(1-\text{CDF}(x))$. Experimental data consists of $T = 2^{40}$ uniform samples.

Each of the subfigures of Figure 2 has experimental data, that is obtained on a server with 40 physical CPUs running for a bit over one day. Hence, we also considered running a lattice sieve with a different saturation radius, which produces more (and therefore a bit longer) dual vectors than the standard saturation radius of $\sqrt{4/3} \approx 1.1547$. In these scenarios, the floor phenomenon happens with a much higher probability. Thus fewer samples are needed to observe the floor phenomenon in the experimental data. This eases the computational power required to get the experiments, but the analysis should of course still hold in a situation where the saturation radius is not $\sqrt{4/3}$. This motivates Figure 3. This figure shows the predictions for a larger saturation radius. In these cases the floor phenomenon can be seen after a decent computation time, and it shows that the new predictions match the experimental data very accurately.

5 Conclusion

This paper proposed heuristics that can be used in the analysis of attacks against the decision-BDD problem. Specifically, Heuristic Claim 2 and Heuristic Claim 4 experimentally show better predictions regarding BDD and uniform targets than the Independence Heuristic, which was used previously. Thus, we can conclude that these heuristics may be used confidently in the analysis of dual attacks on search-BDD.

However, the effectiveness of the state-of-the-art dual attacks remains as future work. In particular, note that existing dual attacks, e.g. [GJ21], will most likely require a different parametrization to achieve the best runtime, while having a constant success probability. Moreover, there will be other aspects to the costing of the attack that may still require some attention, as highlighted in [DP23, App. A].

In addition to handling false-positives as suggested in [DP23, Sec. 6.3] or adapting the strategy for the same issue in statistical decoding [MT23, Alg. 4.1], one could potentially improve the dual attack further by applying different weights to the individual scores, to achieve better separation between the BDD and uniform distributions [AR04,LW21,PS23b].

On another note, theoretically it would be interesting to predict the scores of uniform targets from Section 3.4: instead of relying on a ball approximation of the Voronoi cell, it could be more satisfactory to adapt the Poisson model of [MT23, Assumption 8]. Indeed, one expects on average to have $\mathrm{Vol}_n(r\mathcal{B}^n) \,/\, \det(\Lambda)$ many lattice points at distance at most r from a target $\mathbf{t}$ sampled uniformly modulo a lattice. Whereas the code had a Poisson process for each weight $w = 0, 1, \ldots, n$, the situation is a bit more complex for lattices, as there are is a continuum of processes to consider.

References

ADH^{+}19. Martin R. Albrecht, Léo Ducas, Gottfried Herold, Elena Kirshanova, Eamonn W. Postlethwaite, and Marc Stevens. The general sieve kernel and new records in lattice reduction. In Yuval Ishai and Vincent Rijmen, editors, *EUROCRYPT 2019, Part II*, volume 11477 of *LNCS*, pages 717–746. Springer, Heidelberg, May 2019.

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

Al 01. A. Kh. Al Jabri. A statistical decoding algorithm for general linear block codes. In Bahram Honary, editor, *8th IMA International Conference on Cryptography and Coding*, volume 2260 of *LNCS*, pages 1–8. Springer, Heidelberg, December 2001.

AR04. Dorit Aharonov and Oded Regev. Lattice problems in NP cap coNP. In *45th FOCS*, pages 362–371. IEEE Computer Society Press, October 2004.

AS64. Milton Abramowitz and Irene A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series.* U.S. Government Printing Office, 1964. `https://archive.org/details/AandS-mono600`.

AS22. Martin R. Albrecht and Yixin Shen. Quantum augmented dual attack, 2022. `https://arxiv.org/abs/2205.13983`.

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

BLS16. Shi Bai, Thijs Laarhoven, and Damien Stehlé. Tuple lattice sieving. Cryptology ePrint Archive, Report 2016/713, 2016. `https://eprint.iacr.org/2016/713`.

BM18. Leif Both and Alexander May. Decoding linear codes with high error rate and its impact for LPN security. In Tanja Lange and Rainer Steinwandt, editors, *Post-Quantum Cryptography - 9th International Conference, PQCrypto 2018*, pages 25–46. Springer, Heidelberg, 2018.

CDMT22. Kevin Carrier, Thomas Debris-Alazard, Charles Meyer-Hilfiger, and Jean-Pierre Tillich. Statistical decoding 2.0: Reducing decoding to LPN. In Shweta Agrawal and Dongdai Lin, editors, *ASIACRYPT 2022, Part IV*, volume 13794 of *LNCS*, pages 477–507. Springer, Heidelberg, December 2022.

CDMT23. Kevin Carrier, Thomas Debris-Alazard, Charles Meyer-Hilfiger, and Jean-Pierre Tillich. Reduction from sparse LPN to LPN, dual attack 3.0. Private communication, 2023.

CST22. Kevin Carrier, Yixin Shen, and Jean-Pierre Tillich. Faster dual lattice attacks by using coding theory. Cryptology ePrint Archive, Paper 2022/1750, 2022. `https://eprint.iacr.org/2022/1750`.

DP23. Léo Ducas and Ludo N. Pulles. Does the dual-sieve attack on learning with errors even work? In Helena Handschuh and Anna Lysyanskaya, editors, *CRYPTO 2023, Part III*, volume 14083 of *LNCS*, pages 37–69. Springer, Heidelberg, August 2023.

DSvW21. Léo Ducas, Marc Stevens, and Wessel P. J. van Woerden. Advanced lattice sieving on GPUs, with tensor cores. In Anne Canteaut and François-Xavier Standaert, editors, *EUROCRYPT 2021, Part II*, volume 12697 of *LNCS*, pages 249–279. Springer, Heidelberg, October 2021.

EJK20. Thomas Espitau, Antoine Joux, and Natalia Kharchenko. On a dual/hybrid approach to small secret LWE - A dual/enumeration technique for learning with errors and application to security estimates of FHE schemes. In Karthikeyan Bhargavan, Elisabeth Oswald, and Manoj Prabhakaran, editors, *INDOCRYPT 2020*, volume 12578 of *LNCS*, pages 440–462. Springer, Heidelberg, December 2020.

Fis13. Daniel Fischer. Fourier transform of the indicator of the unit ball. Mathematics Stack Exchange, 2013. `https://math.stackexchange.com/a/492055` (version: 2013-09-12).

GJ21. Qian Guo and Thomas Johansson. Faster dual lattice attacks for solving LWE with applications to CRYSTALS. In Mehdi Tibouchi and Huaxiong Wang, editors, *ASIACRYPT 2021, Part IV*, volume 13093 of *LNCS*, pages 33–62. Springer, Heidelberg, December 2021.

GS64. Israel Moiseevich Gel'fand and Georgi Evgen'evich Shilov. *Generalized functions. Volume I: Properties and operations.* Academic Press, New York and London, 1964. Translated by Eugene Saletan.

LW21. Thijs Laarhoven and Michael Walter. Dual lattice attacks for closest vector problems (with preprocessing). In Kenneth G. Paterson, editor, *CT-RSA 2021*, volume 12704 of *LNCS*, pages 478–502. Springer, Heidelberg, May 2021.

MAT22. MATZOV. Report on the security of LWE: Improved dual lattice attack, April 2022.

MT23. Charles Meyer-Hilfiger and Jean-Pierre Tillich. Rigorous foundations for dual attacks in coding theory. Cryptology ePrint Archive, Paper 2023/1460, 2023. `https://eprint.iacr.org/2023/1460`.

MV10. Daniele Micciancio and Panagiotis Voulgaris. Faster exponential time algorithms for the shortest vector problem. In Moses Charika, editor, *21st SODA*, pages 1468–1480. ACM-SIAM, January 2010.

NV08. Phong Q. Nguyen and Thomas Vidick. Sieve algorithms for the shortest vector problem are practical. *Journal of Mathematical Cryptology*, 2(2):181–207, 2008.

PS23a. Amaury Pouly and Yixin Shen. Provable dual attacks on learning with errors. Cryptology ePrint Archive, Paper 2023/1508, 2023. `https://eprint.iacr.org/2023/1508`.

PS23b. Amaury Pouly and Yixin Shen. Provable dual attacks on LWE with sieving. Private communication, 2023.

Reg09. Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. *Journal of the ACM*, 56(6):1–40, 2009. Preliminary version in STOC'05.

SW71. Elias M. Stein and Guido Weiss. *Introduction to Fourier Analysis on Euclidean Spaces.* Number 32 in Princeton Mathematical Series. Princeton University Press, Princeton, New Jersey, 1971.

VGS17. Aaron R. Voelker, Jan Gosmann, and Terrence C. Stewart. Efficiently sampling vectors and coordinates from the n-sphere and n-ball. *Centre for Theoretical Neuroscience-Technical Report*, 2017.

Wat22. George Neville Watson. *A treatise on the theory of Bessel functions.* Cambridge University Press, 1922. `https://archive.org/details/treatiseontheory00watsuoft`.

WE23. Andreas Wiemers and Stephan Ehlen. A remark on the independence heuristic in the dual attack. Cryptology ePrint Archive, Paper 2023/1238, 2023. `https://eprint.iacr.org/2023/1238`.