J Cryptol (2026) 39:8
https://doi.org/10.1007/s00145-025-09560-7

Journal of CRYPTOLOGY

Research Article

Accurate Score Prediction for Dual-Sieve Attacks*

Léo Ducas
Cryptology Group, CWI, Amsterdam, The Netherlands
Mathematical Institute, Leiden University, Leiden, The Netherlands

Ludo N. Pulles
Cryptology Group, CWI, Amsterdam, The Netherlands
lnp@cwi.nl

Communicated by Shweta Agrawal

Received 22 May 2024 / Revised 27 August 2025 / Accepted 10 October 2025

Abstract. Guo and Johansson (ASIACRYPT 2021), and MATZOV (tech. report 2022) have independently claimed improved attacks against various NIST lattice candidates by using a Fast Fourier Transform (FFT) on top of the so-called Dual-Sieve attack. However, we will show that a heuristic used in above works not only theoretically contradicts with both formal theorems and well-tested heuristics in certain regimes, but also provides incorrect predictions experimentally. We conclude that this heuristic significantly overestimates the success probability of the Dual-Sieve attack. Alternatively, we propose a seemingly weaker heuristic for the output of a lattice sieve. When determining part of the secret in the Dual-Sieve attack, we derive predictions for the score distribution associated to candidates using this heuristic: for correct candidates with noise drawn from any radial distribution, we derive score predictions using a central limit heuristic; for incorrect candidates, we derive score predictions by approximating the Voronoi cell by a ball. In the process, we show that the use of the FFT is not specific to Learning with Errors (LWE) but is more generally useful against the Bounded Distance Decoding problem (BDD). Ultimately, we compare the predicted score distributions with extensive experiments, and observe these predictions to be qualitatively and quantitatively quite accurate. This makes it possible to accurately estimate the number of false positives and false negatives, opening the door for a sound analysis of the Dual-Sieve attack. In particular, one may consider exploring the opportunities to mitigate a large number of false positives.[1]

Keywords. Lattices, Cryptanalysis, Heuristics, Learning with Errors, Dual attack, Fast Fourier Transform, Bessel functions.

[1] This paper was reviewed by Shi Bai and Alice Pellet-Mary

*This work is based on two articles from the same authors [19,20]. Sections 3, 4 and 5 are based on the former work [20]. Sections 6 and 7 are based on the latter work [19]

Published online: 20 November 2025

1. Introduction

Many post-quantum cryptoschemes base their security on the hardness of the Learning with Errors (LWE) problem, introduced by Regev in 2005 [43], 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*, which uses the idea dating back to [2] that short dual vectors allow distinguishing points close to a lattice from points far away from the lattice. This idea can even be traced back in a pure geometric context to earlier work by Håstad [28], and is not even limited to lattices, as it was already implicit in the very construction of Low-Density parity-Check codes dating back to [25]. In contrast, a lattice-based attack operating solely on the *primal* lattice is called a *primal attack*. The best dual and primal attacks are then typically used by the lattice cryptanalyst to instantiate LWE cryptosystems.

More specifically, the dual attack attempts to solve the search-BDD problem by performing a reduction to the decision-BDD problem, where one is asked to determine whether 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, a short dual vector provides a score function that assigns a high score to BDD targets and an expected score of 0 to uniform targets. Only by using many dual vectors and considering the sum of individual scores [31], one can hope for distinguishing between the two target distributions with a high success probability. To get exponentially many short dual vectors, Alkim, Ducas, Pöppelmann and Schwabe initially suggested [7] to use a lattice sieve on the dual lattice [10,36,38].

We refer to this style of attack as a *Dual-Sieve* attack. The concrete cryptanalytic impact of this idea can be further improved by guessing multiple coordinates of the secret rather than just one [4,23], and then finding the right solution among these candidates rather than just a few.

Another development to the dual attack is also reminiscent from a cryptanalytic technique of code-based cryptography by Levieil and Fouque [32]. The idea is to batch the score evaluation of a large number of algebraically related candidates via a Fast Fourier Transform (FFT). For carefully crafted parameters, the cost of getting all those scores is barely larger than the cost of naïvely computing a single score. This led Guo and Johansson [27] to claim an improved attack on various NIST post-quantum standardization candidates, followed quickly by an independent technical report of MATZOV [33]. We refer to this style of attack as a *Dual-Sieve-FFT* attack. The latter has already been followed up upon, with a quantum variant [6] and a coding-theoretic enhanced variant [14]. In their analysis of the Dual-Sieve attack, all these four works use a heuristic saying that individual score functions given by a set of dual vectors are mutually independent variables.

1.1. *Contributions*

Generalization of the Dual-Sieve-FFT Attack (Sect. 3). The original principle of the dual attack [2,28] is stated for the BDD problem in any lattice. However, the recent instances of the Dual-Sieve attack [4,7,23] and the Dual-Sieve-FFT attack [6,14,27,33]

are described in a specialized way to Learning with Errors. One exception is the work of [31], which is geometrically enlightening but limited to the Dual-Sieve attack.

Our first contribution is therefore to generalize the FFT trick of [27] to the setting of BDD. Beyond the theoretical satisfaction of abstracting the technique to its mathematical core, this generalization also offers further improvement over the work of [27]: for the same algorithmic price, we can further improve the shortness of the dual vectors and therefore their distinguishing power.

Contradictions to the Independent Score Heuristic (Sect. 4). A second observation regarding this literature is that the analysis of the Dual-Sieve attack (with or without FFT) relies on one specific heuristic, which has received essentially no attention so far. Namely, it is assumed that all the individual scores, given by each dual vector, are mutually independent.

We approach the analysis of this heuristic by looking at the conclusions it leads to. The geometric point of view offered by the work of Laarhoven and Walter [31] is pivotal in that respect: judging the reasonability of a heuristic conclusion is very much enabled by the language of geometry. In particular, [31] concludes with a heuristic algorithm that solves decision-BDD, even when the noise slightly exceeds the Gaussian Heuristic (GH), i.e., the expected minimal distance of a random lattice. This should raise suspicion, as a random point is on average not expected to be further than GH away from the lattice. We show this suspicion is fair: it easily follows from a result of Debris, Ducas, Resch, and Tillich [16] that the above task is statistically impossible to solve, even to an unbounded attacker.

The contradiction above is, however, limited to a rather theoretical regime of the Dual-Sieve attack, which is not that of the recent concrete cryptanalytic claims [6,14,23,27, 33], where the noise is below GH. Still, here the attack needs to distinguish between a BDD sample and a uniform sample with a very high success probability. We will show that this situation also leads to a contradiction.

Namely, we argue that the heuristic probability of having a distinguishing failure is much smaller than the probability of a uniform target lying closer to the lattice than a BDD target. The claimed high success probability of distinguishing between BDD and uniform contradicts the basic principle that targets closer to the lattice get a higher score in expectation. It turns out that the parameters used in [27,33] specifically fall into that contradictory regime that requires a distinguisher with a very high success probability.

Experimental invalidation of the Prior Model (Sect. 5). To develop a more precise understanding, we zoom in on the score distribution for BDD targets and uniform targets. Extensive experiments were performed to discover that both distributions deviate from the predictions made under the Independent Score Heuristic. First, the *body* of the distribution of scores for uniform targets is properly predicted, but not its *tail*: after a predicted rapid decrease, visually resembling a *waterfall*, this distribution hits a *floor*. This is perfectly in line with our second contradictory regime: some uniform targets will be close to the lattice, and should therefore have a high score.

The score distribution for BDD targets is also mispredicted, and this is no longer just a matter of the tail. Contrary to prediction, this distribution is not Gaussian-like. It is in fact not even symmetric around its average, and its variance appears exponentially larger than predicted. In particular, a BDD target will more likely have a low score than what is heuristically predicted.

Accurate Score Distribution models (Sect. 6) The mispredictions caused by the Independent Score Heuristic show there is some correlation between the inner products $\langle \mathbf{w}, \mathbf{t} \rangle \pmod 1$, so the situation may be resolved by identifying the confounding variable that causes dependencies. In this case, the confounding variable seems to be the target $\mathbf{t}$, or more precisely its norm $\|\mathbf{t}\|$. 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.

In contrast to existing literature, we do not approximate the distribution of lattice sieve vectors by a Gaussian distribution [33, Assumption 4.4]. Instead, we model them in Heuristic 2.13 as being uniform in a ball, which describes lattice sieve vectors much more accurately. This somewhat complicates the Fourier analysis, but is feasible thanks to the well-known Bessel functions.

Fixing the target $\mathbf{t}$ leads very naturally to studying the score distribution of $\mathbf{t}$ taken over the randomness of the lattice and the set of short dual vectors. For this target, we derive a novel *analytic expression* for the expectation value and variance of the score distribution. Note that in this case, each individual score has the same individual score distribution, so 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, being normally distributed.

While the case of a particular target is of limited interest in our context, the analysis can be easily extended to any other radial distribution, e.g., Gaussian or uniform in a ball.

Although it is notoriously hard to determine the exact shape of the Voronoi cell [30], one can roughly approximate the Voronoi cell by a ball of same volume. This in turn approximates the uniform target distribution by a radial distribution, which allows predicting the score distribution for uniform targets using the above prediction.

Validating the New Model (Sect. 7). The overall approach remains heuristic, but the previously identified issues have now been sidestepped, and we further confirm those score distribution predictions with extensive experiments. All the predictions that are made throughout this work 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 Fig. 6, in dimension 90. In addition, the predicted score distribution for uniform targets is shown in Fig. 8. This figure contains data of experiments in dimensions 40, 50, 60 and 70, each based on 2^{48} samples. Lastly, Fig. 9 shows the score distribution for uniform targets when using a larger saturation radius inside lattice sieving. Here, our predictions show that the "waterfall-floor" phenomenon could be observed with fewer samples, compared to the default saturation radius of $\sqrt{4/3}$, which the experiments confirm.

Open data. All the written code and the obtained data to generate the figures, can be found in the two following GitHub repositories:

1. https://github.com/ludopulles/DoesDualSieveWork,
2. https://github.com/ludopulles/AccurateScorePredictionDualSieveAttacks

The experiments are written in `Python` and use the G6K and FPyLLL libraries [5,47], and there is a binding to some `C` code to accelerate the FFT.

1.2. *Conclusion*

Our theoretical contradictions and our experiments both demonstrate that the Independent Score Heuristic, underlying the prior standard analysis of the Dual-Sieve attack, is invalid. Two independent phenomena uncovered by the experiments point to the success probability of the Dual-Sieve attack (with or without FFT) being presumably overestimated by this Independent Score Heuristic, at least in certain regimes of interest.

In particular, the concrete cryptanalytic claims of numerous works, including [6,7,14,23,27,33], should be considered at least unsubstantiated, as these are currently based on this flawed heuristic. Still, some of those claims might not be that far from reality, but those of [6,14,27,33] are so deep in the contradictory regime that these are presumably significantly far away from reality.

In addition, after identifying the norm of the target $\|\mathbf{t}\|$ as a confounding variable in the individual scores, we propose alternative heuristics in which this confounding variable is fixed. This leads us to predict the score distributions for both BDD targets and uniform targets in the Dual-Sieve attack. The newly derived heuristic predictions are tested in extensive experiments, which reveal that the predictions are rather accurate compared to experimental data.

In conclusion, the experiments strongly suggest that Heuristics 2.13, 6.5 and 6.11 are very reasonable to use while analyzing dual attacks. The precise effectiveness of the state-of-the-art dual attack remains as future work.

1.3. *Concurrent Work*

During the writing process of this paper, we became aware of several concurrent works [13,40,41,49].

The work of Wiemers, Ehlen and Bashiri [49] proposed an analysis of the variance of the score of the actual BDD target and the uniform targets. They use a significantly different approach, and obtain predictions compatible with the experimental data from Sect. 5.3. This does not directly allow to conclude on false-negative nor false-positive rates, as it does not fully characterize the distribution.

The work of Pouly and Shen [41] proposed a provable variant of the dual attack, using discrete Gaussian sampling in place of sieving vectors. Notably, their provable regime does not intersect the heuristic contradictory regime in Sect. 4.3. In fact, and quite interestingly, both regimes 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 [40].

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 [13, Sect. 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 Sect. 6.4.

1.4. *Analogies with Coding Theory*

The dual attack in lattices is called "statistical decoding" in coding theory and dates back to the works of [3,25,39]. 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, [12] claimed to have an improved statistical decoding algorithm outperforming the state-of-the-art information set decoding algorithm of [11] on sparse binary codes. However, this work was based on the coding-theory analog of the Independent Score Heuristic [12, 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 [34]. 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.[2] 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.

2. Preliminaries

In this paper, we will make clear which heuristics are used by referring to these as *Heuristics*. Any statement that is derived using one or more heuristics will be called a *Heuristic Claim*, which will be motivated by a *Heuristic Justification* explaining why it is believed to be true.

Notation. For any set $S \subseteq \mathbb{R}^n$, the *indicator function of* S, denoted by $\mathbf{1}_S$, evaluates to 1 on S, while $\mathbf{1}_S(\mathbf{x}) = 0$ for $\mathbf{x} \in \mathbb{R}^n \backslash S$. The n-dimensional *gaussian of width* $s \in \mathbb{R}_{>0}$ is defined by

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

Geometric objects. The n-dimensional (closed) ball of radius 1 is denoted by $\mathcal{B}^n$. The $(n-1)$-dimensional sphere (residing in the n-dimensional ambient space) is denoted by $\mathcal{S}^{n-1}$. In particular, the *unit circle*, denoted by $\mathcal{S}^1$, is naturally a subgroup of $\mathbb{C}^*$. The n-dimensional ball has volume

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

Probabilities and Distributions. The probability of event E happening is denoted by $\mathbb{P}[E]$, and the expectation value of a variable X is denoted by $\mathbb{E}[X]$. The *variance* of a random variable X is $\mathbb{V}[X] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$, and its *standard deviation* is $\sigma_X = \sqrt{\mathbb{V}[X]}$. The *cumulative distribution function* (CDF) at $x \in \mathbb{R}$ of a distribution $\mathcal{D}$ is $\mathbb{P}_{X \leftarrow \mathcal{D}}[X \leq x]$, while the *probability density function* (PDF) at $x \in \mathbb{R}$ is the derivative of $\mathbb{P}_{X \leftarrow \mathcal{D}}[X \leq x]$ with respect to x. The *survival function* (SF) of $\mathcal{D}$ is $\mathbb{P}_{X \leftarrow \mathcal{D}}[X \geq x] = 1 - \mathbb{P}_{X \leftarrow \mathcal{D}}[X < x]$.

The *uniform distribution on a set* X is denoted by $U(X)$.

[2] 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 [34] is written in terms of the $[n-s, k-s]$-subcode there.

The *gaussian* (or, normal distribution) $\mathcal{N}\left(c, \sigma^2\right)$ with center $c \in \mathbb{R}$ and standard deviation $\sigma \in \mathbb{R}_{>0}$ has a probability density at $x \in \mathbb{R}$ proportional to $\rho_{\sqrt{2\pi}\sigma}(x-c)$. The CDF of a Gaussian $\mathcal{N}\left(c, \sigma^2\right)$ is expressible in terms of the *error function*, denoted by erf, as follows:

$$\mathop{\mathbb{P}}_{X \leftarrow \mathcal{N}(c,\sigma^2)}[X \le x] = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{x-c}{\sigma\sqrt{2}}\right) \qquad (x \in \mathbb{R}). \tag{1}$$

The *complementary error function* is given by $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$, and has the following asymptotic decay rate.

Lemma 2.1. ([1, 7.1.24]) *For all $x > 0$,*

$$\mathrm{erfc}(x) \le \frac{1}{\sqrt{\pi}x} \cdot e^{-x^2}.$$

The *Central Limit Theorem* states that, given *n independent and identically distributed* (i.i.d.) random variables $X_1, \ldots, X_n$, each having mean μ and variance σ^2, the distribution $\frac{\sum_i X_i/n - \mu}{\sigma/\sqrt{n}}$ tends toward $\mathcal{N}(0,1)$ as $n \to \infty$. In concrete settings, heuristically we expect that this distribution is close to a Gaussian already.

Heuristic 2.2. *(Central Limit Heuristic) Given a large number of i.i.d. random variables $X_1, \ldots, X_n$, each having mean μ and variance σ^2, the sum of these variables is distributed as such:*

$$X_1 + X_2 + \cdots + X_n \sim \mathcal{N}\left(n\mu, n\sigma^2\right).$$

Lattices. The *$\mathbb{R}$-linear span of a set $S \subset \mathbb{R}^n$* is denoted by $\mathrm{span}(S)$. A *lattice* Λ is a discrete subgroup of $\mathbb{R}^n$, its *rank* is the dimension of $\mathrm{span}(\Lambda)$, and its *volume* is $\det(\Lambda) = \mathrm{Vol}_k(\mathrm{span}(\Lambda)/\Lambda)$. The length of a shortest nonzero vector of a lattice $\Lambda \subset \mathbb{R}^n$ is denoted by $\lambda_1(\Lambda)$. We speak of a *full-rank* lattice if its rank is n, and a *unit-volume* lattice if $\det(\Lambda) = 1$.

A *basis for a lattice* $\Lambda \subset \mathbb{R}^n$ is a matrix $\mathbf{B} \in \mathbb{R}^{n \times k}$ that consists of $\mathbb{R}$-linearly independent column vectors $\mathbf{b}_1, \ldots, \mathbf{b}_k \in \mathbb{R}^n$ such that $\Lambda = \mathbb{Z}\mathbf{b}_1 + \cdots + \mathbb{Z}\mathbf{b}_k$. Note, given a basis $\mathbf{B}$ for Λ, we have $\det(\Lambda) = \sqrt{\det(\mathbf{B}^\mathsf{T}\mathbf{B})}$.

The *Voronoi cell of* Λ, denoted by $\mathcal{V}(\Lambda)$, is the set of points that are not closer to any lattice point other than to the origin, i.e.,

$$\mathcal{V}(\Lambda) = \left\{\mathbf{x} \in \mathrm{span}(\Lambda) \;\middle|\; \forall \mathbf{v} \in \Lambda : \|\mathbf{x}\| \le \|\mathbf{x} - \mathbf{v}\|\right\}.$$

Note that we have $\det(\Lambda) = \mathrm{Vol}_k(\mathcal{V}(\Lambda))$ for any lattice Λ rank k.

Given a basis $\mathbf{B} \in \mathbb{R}^{n \times k}$ and $1 \le \ell \le r \le k$, we write $\mathbf{B}_{[\ell,r]}$ for the basis consisting of the vectors $\pi_\ell(\mathbf{b}_\ell), \ldots, \pi_\ell(\mathbf{b}_r)$, where π_ℓ is the projection map that projects orthogonally away from $\mathbf{b}_1, \ldots, \mathbf{b}_{\ell-1}$.

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 $\mathrm{Vol}_n(S) / \det(\Lambda)$. This leads to the following two heuristics on the length of short vectors.

Heuristic 2.3. *Given a random lattice $\Lambda \subset \mathbb{R}^n$, $\lambda_1(\Lambda)$ is approximately equal to $\mathrm{GH}(n) \cdot \det(\Lambda)^{1/n}$, where*

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

Note that Minkowski's theorem states $\lambda_1(\Lambda) \leq 2 \cdot \mathrm{GH}(n) \cdot \det(\Lambda)^{1/n}$. Moreover,

$$\mathrm{GH}(n) \sim \sqrt{n/(2\pi e)} \qquad (\text{as } n \to \infty).$$

Heuristic 2.4. *Given a random lattice $\Lambda \subset \mathbb{R}^n$ and some $r > 1$, we have*

$$\left|\left\{\mathbf{v} \in \Lambda \;\middle|\; \|\mathbf{v}\| \leq r \cdot \mathrm{GH}(n) \det(\Lambda)^{1/n}\right\}\right| \approx r^n.$$

In particular, the i^{th} shortest lattice point $\mathbf{v}$ has a length of approximately $\|\mathbf{v}\| \approx \mathrm{GH}(n) \det(\Lambda)^{1/n} i^{1/n}$.

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

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

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 2.6. ([48, §15.81]) *The first positive root of the Bessel function satisfies*

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

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

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

However, note we will not use this approximation, since ξ can be computed easily in many programming languages.

Remark 2.8. 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, `Python` supports ${}_0F_1$ with the function `hyp0f1` from the `mpmath` package.

2.1. *Duality*

There are two ways to define the dual of a lattice. The first one is inherited from groups, while the second one is geometric and specific to lattices. In the context of Fourier transformations, it is useful to consider both definitions, and relate them. Note that characters are only taken for full-rank lattices.

Definition 2.9. For a locally compact group G (e.g., a finite group or a torus $\mathbb{R}^n/\Lambda$ of a full-rank lattice Λ), the *group of characters on* G, is denoted by

$$\widehat{G} = \left\{\chi : G \to \mathcal{S}^1 \mid \chi \text{ continuous group homomorphism}\right\}.$$

Note that when G is a finite abelian group, any homomorphism $\chi : G \to \mathcal{S}^1$ is continuous.

Definition 2.10. The *dual of a lattice* $\Lambda \subset \mathbb{R}^n$, denoted by $\Lambda^\vee$, consists of all vectors $\mathbf{x} \in \text{span}(\Lambda)$ 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 following lemma shows that we may interchange these two notions of duality, i.e., any dual vector defines a character and vice versa.

Lemma 2.11. *The map from* $\Lambda^\vee$ *to* $\widehat{\mathbb{R}^n/\Lambda}$ *that sends a dual vector* $\mathbf{w}$ *to the character*

$$\chi_{\mathbf{w}} : \mathbf{t} \mapsto \exp(2\pi i \cdot \langle \mathbf{t}, \mathbf{w}\rangle), \tag{2}$$

is a group isomorphism.

Proof. **Well-definedness:** this follows directly from the definition of $\Lambda^\vee$.

Injectivity: $(\forall \mathbf{t} \in \mathbb{R}^n : \chi_{\mathbf{w}}(\mathbf{t}) = 1) \iff \langle \mathbf{w}, \mathbb{R}^n \rangle \subseteq \mathbb{Z} \iff \mathbf{w} = \mathbf{0}$.

Surjectivity: Let $\chi \in \widehat{\mathbb{R}^n/\Lambda}$ be given. By continuity, there is an open ball $U \subseteq \mathbb{R}^n$ centered at $\mathbf{0}$ such that we have $\chi(U + \Lambda) \subseteq \{z \in \mathcal{S}^1 \mid \mathrm{Re}(z) > 0\}$. On this ball, we can find a linear map $\varphi\colon U \to (-\frac{1}{4}, \frac{1}{4})$ such that $\chi(\mathbf{x} + \Lambda) = \exp(2\pi i \varphi(\mathbf{x}))$ holds for all $\mathbf{x} \in U$ as there is exactly one value for $\varphi(\mathbf{x})$ that is valid. The character χ is completely determined by its values on U, i.e., for any $\mathbf{x} \in \mathbb{R}^n$ there exists $m \in \mathbb{Z}_{\geq 1}$ such that $\mathbf{x}/m \in U$ so $\chi(\mathbf{x}) = \chi(\mathbf{x}/m)^m$. We can now extend φ to a linear function $\varphi\colon \mathbb{R}^n \to \mathbb{R}$, which will satisfy $\chi(\mathbf{x} + \Lambda) = \exp(2\pi i \varphi(\mathbf{x}))$ for all $\mathbf{x} \in \mathbb{R}^n$. Now there exists some $\mathbf{w} \in \mathbb{R}^n$ such that $\varphi = \langle -, \mathbf{w} \rangle$ as any linear map is of this form. Note that $\langle \mathbf{w}, \Lambda \rangle \subset \mathbb{Z}$, i.e., $\mathbf{w} \in \Lambda^\vee$ because $\chi(\Lambda) = 1$ and therefore $\chi = \chi_{\mathbf{w}}$. □

For a sublattice $\Lambda' \subset \Lambda$, the dual of the finite group Λ/Λ' has a natural connection to the dual lattices of Λ' and Λ by the following lemma.

Lemma 2.12. *For two full-rank lattices $\Lambda_1 \subset \Lambda_2 \subset \mathbb{R}^n$, there is a canonical group isomorphism of abelian groups,*

$$\left(\widehat{\mathbb{R}^n/\Lambda_1}\right) \Big/ \left(\widehat{\mathbb{R}^n/\Lambda_2}\right) \to \widehat{\Lambda_2/\Lambda_1},$$

given by restricting a character $\chi : \mathbb{R}^n/\Lambda_1 \to \mathcal{S}^1$ (modulo $\widehat{\mathbb{R}^n/\Lambda_2}$) to Λ_2/Λ_1.

Proof. **Well-definedness**: any Λ_1-periodic character χ can be multiplied by a Λ_2-periodic character $\psi \in \widehat{\mathbb{R}^n/\Lambda_2}$ as the function stays the same on Λ_2/Λ_1.

Injectivity: any character $\chi \in \widehat{\mathbb{R}^n/\Lambda_1}$ that is 1 on Λ_2/Λ_1 is coming from a function $\mathbb{R}^n \to \mathcal{S}^1$ that is Λ_2-periodic, i.e., a character from $\widehat{\mathbb{R}^n/\Lambda_2}$.

Surjectivity: left hand side has size

$$\left|\left(\widehat{\mathbb{R}^n/\Lambda_1}\right) \Big/ \left(\widehat{\mathbb{R}^n/\Lambda_2}\right)\right| = \left|\Lambda_1^\vee/\Lambda_2^\vee\right| = |\Lambda_2/\Lambda_1|,$$

and right hand side has size $\left|\widehat{\Lambda_2/\Lambda_1}\right| = |\Lambda_2/\Lambda_1|$ where we use that a finite abelian group G is isomorphic to its dual. □

Dual basis and dual blocks. Given a basis $\mathbf{B}$ of the primal lattice Λ, one can construct an associated dual basis $\mathbf{B}^\vee = \mathbf{B} \cdot (\mathbf{B}^\mathsf{T} \cdot \mathbf{B})^{-1}$ of the dual lattice $\Lambda^\vee$. Consider the *reversed dual basis* ${}^\vee\mathbf{B} = \left[\mathbf{b}_n^\vee, \ldots, \mathbf{b}_1^\vee\right]$ in which the ordering of the basis vectors is reversed. It is known that $\tau(\mathbf{b}_r^\vee), \ldots, \tau(\mathbf{b}_\ell^\vee)$ forms a basis for the dual of the lattice with basis $\mathbf{B}_{[\ell,r]}$, where τ is the map that projects away from $\mathbf{b}_{r+1}^\vee, \ldots, \mathbf{b}_n^\vee$, denoted by $({}^\vee\mathbf{B})_{[n+1-r,n+1-\ell]}$. Informally, this shows that projecting in the primal lattice corresponds to sectioning in the dual lattice. More details on dual bases can be found in the course of Micciancio [35].

2.2. *Lattice Sieve*

Lattice sieves provide a way to efficiently produce a list of many short lattice vectors [10,36,38]. Although there are many variations, e.g., [8], 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. Moreover, this factor of 2 needs to be taken into account when analyzing the dual attack, cf. [18, 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 2.13. *Given a full-rank unit-volume lattice $\Lambda \subset \mathbb{R}^n$, the output distribution of a lattice sieve with a saturation radius of $r_{\text{sat}} \geq 1$ and a saturation ratio of $f_{\text{sat}} \in (0, 1]$, denoted by* $\text{Sieve}(\Lambda, r_{\text{sat}}, f_{\text{sat}})$*, 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 $r_d = r_{\text{sat}}\text{GH}(n)$.*

This heuristic makes two simplifications on the output of a lattice sieve. The first simplification made 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 $\Lambda \cap r_{\text{sat}}\text{GH}(n)\,\mathcal{B}^n$. Although the heuristic allows for duplicates, still many distinct values will be sampled, which motivates 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 $n \to \infty$.[3]

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 $r_{\text{sat}}\text{GH}(n)$, as done in, e.g., [27]. Instead, the heuristic predicts there may be some shorter vectors $\mathbf{w}$, albeit with smaller probability. 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 a security analysis will be more conservative when taking shorter vectors into account.

Normally, a lattice sieve runs 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 [38]. Although in theory, one sometimes assumes a lattice sieve finds *all* lattice vectors inside a ball of radius $r_{\text{sat}}\text{GH}(n)$, i.e., $f_{\text{sat}} = 1$, in practice a much lower saturation ratio is chosen for efficiency. For instance, the `G6K` software [5] 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 [21].

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

One can also sieve on a lattice that is not full-rank. Sieve algorithms require both time and memory exponential in the rank of Λ in general [10].

2.3. *Fourier Transformation*

Definition 2.14. The *Fourier transform* of a function $f\colon \mathbb{R}^n \to \mathbb{R}$, for which the integral $\int_{R^n} f(\mathbf{x})^2 dx$ is a finite value, 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:

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

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 [45, Thm. 3.3].

Theorem 2.15. (Poisson Summation Formula, [45, 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 ball.

Lemma 2.16. *The Fourier transform of $f = \mathbf{1}_{r\mathcal{B}^n}$ is given by*

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

The proof is based on [24].

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

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

For the last equality, note that we have

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

□

Discrete Fourier Transforms. For any set S let $\mathbb{C}^S$ be the group of sequences $(x_s)_{s\in S}$ having complex coefficients x_s, where the group operation is given by pointwise addition. Given a finite group G, the *Discrete Fourier Transform* (DFT) of a sequence $(x_g)_{g\in G} \subset \mathbb{C}$ is the $\mathbb{C}$-linear map

$$\begin{aligned} \mathrm{DFT}_G\colon \quad & \mathbb{C}^G \to \mathbb{C}^{\widehat{G}}, \\ & (x_g)_{g\in G} \mapsto \left(\sum_{g\in G} x_g \cdot \overline{\chi(g)}\right)_{\chi\in\widehat{G}}. \end{aligned} \tag{3}$$

The m-dimensional *Fast Fourier Transform* (FFT) is an algorithm that, upon input a group G, given as $n_1, \ldots, n_m \in \mathbb{Z}_{\geq 2}$ such that $G \cong \bigoplus_{j=1}^m (\mathbb{Z}/n_j\mathbb{Z})$, and $(x_g)_{g\in G}$, outputs $\mathrm{DFT}_G\left((x_g)_{g\in G}\right)$ in time $O(|G|\log|G|)$. There are various FFTs known for any finite group G (even when an n_i is a large prime) [22,42]. When the group G is not cyclic, the algorithm is often referred to as a multi-dimensional FFT. When $G \cong (\mathbb{Z}/2\mathbb{Z})^k$, the algorithm is a *Walsh–Hadamard Transform* (WHT), which is more efficient in practice. For a finite group G, the inverse of DFT_G is given by

$$\mathrm{DFT}_G^{-1}\left((y_\chi)_{\chi\in\widehat{G}}\right) = \frac{1}{|G|} \cdot \left(\sum_{\chi\in\widehat{G}} y_\chi \chi(g)\right)_{g\in G}.$$

Identifying an element $g \in G$ with the evaluation map $\mathrm{ev}_g\colon \chi \mapsto \chi(g)$ gives the canonical isomorphism $G \cong \widehat{\widehat{G}}$, so an inverse DFT is basically a DFT, up to some reordering.

2.4. *Dual Attack*

The following computational problems are considered hard for specific parameters, on which the security of LWE cryptosystems is based.

Problem 2.17. *(Search-BDD, Lattice Form) For $r > 0$,* Search Bounded Distance Decoding *is the task of, given a lattice Λ and a target $\mathbf{t} \in \mathbb{R}^n$ that is the sum of a randomly sampled lattice point $\mathbf{v} \in \Lambda$ and a randomly sampled error $\mathbf{e}$ of norm at most $r\lambda_1(\Lambda)$, finding this lattice point $\mathbf{v}$.*

By considering $\mathbf{t}$ modulo the lattice and demanding $\mathbf{t} - \mathbf{v}$ as a result, we get the syndrome form.

Problem 2.18. *(Search-BDD, Syndrome Form) For $r > 0$, (syndrome)* Search Bounded Distance Decoding *is the task of, given a lattice Λ and target $\bar{\mathbf{t}} = \mathbf{e} + \Lambda \in \mathbb{R}^n/\Lambda$ in the torus where $\mathbf{e}$ is randomly sampled with norm at most $r\lambda_1(\Lambda)$, finding this error $\mathbf{e}$.*

Concretely, to solve a BDD instance, one is given some basis $\mathbf{B}$ of the lattice together with the target $\mathbf{t}$ being expressed in terms of the basis $\mathbf{B}$ with coefficients in the interval $[0, 1)$.

When search-BDD is instantiated with $r < \frac{1}{2}$, it is guaranteed that there is only one lattice point close enough to $\mathbf{t}$. For random lattices, there is still one lattice point close enough with high probability when instantiating search-BDD with $r < 1$ by the following heuristic.

Heuristic Claim 2.19. *Let $\Lambda \subset \mathbb{R}^n$ be a random unit-volume lattice, $r \in (0, 1)$, and $R = r\,\mathrm{GH}(n)$. The probability that a target $\mathbf{t} \leftarrow U(R\mathcal{B}^n)$ is at a distance of at most R from some* nonzero *lattice point $\mathbf{v} \in \Lambda$ is at most $O(n\sqrt{n})r^n$.*

This Heuristic can be justified with the Gaussian Heuristic and an upper bound on spherical domes, cf. [36, Lem. 4.1].

Heuristic Justification. Note that only lattice points $\mathbf{v} \in \Lambda\backslash\{\mathbf{0}\}$ are relevant with $\|\mathbf{v}\| \leq 2R = 2r\,\mathrm{GH}(n)$ by the triangle inequality. For such a $\mathbf{v} \in \Lambda\backslash\{\mathbf{0}\}$, we are interested in $\mathrm{Vol}_n(R\mathcal{B}^n \cap (\mathbf{v} + R\mathcal{B}^n))$, which is twice the volume of the spherical dome $\{\mathbf{t} \in R\mathcal{B}^n \,|\, \langle \mathbf{t}, \mathbf{v}\rangle \geq \frac{1}{2}\|\mathbf{v}\|^2\}$. Set $\alpha = \|\mathbf{v}\|/2R$. This spherical dome is contained in a cylinder with base $R\sqrt{1-\alpha^2} \cdot \mathcal{B}^{n-1}$ and height $R(1-\alpha)$, which has volume at most $R^n(1-\alpha^2)^{n/2}\mathrm{Vol}_{n-1}\left(\mathcal{B}^{n-1}\right)$.

One can show that $\mathrm{Vol}_{n-1}\left(\mathcal{B}^{n-1}\right) \leq \frac{\sqrt{en}}{2}\mathrm{Vol}_n(\mathcal{B}^n)$ holds, which implies

$$\mathrm{Vol}_n\left(R\mathcal{B}^n \cap (\mathbf{v} + R\mathcal{B}^n)\right) \leq O(\sqrt{n})r^n\left(1-\alpha^2\right)^{n/2}.$$

The Gaussian Heuristic predicts approximately ℓ^n lattice points in a ball of radius $\ell\,\mathrm{GH}(n)$. By using this estimate for $\ell \in (1, 2r)$, the volume of all the spherical domes is roughly

$$\int_1^{2r} n\ell^{n-1} \cdot O(\sqrt{n})r^n\left(1 - \frac{\ell^2}{4r^2}\right)^{n/2} \mathrm{d}\ell \leq O(n\sqrt{n})r^n \int_1^{2r}\left(\ell^2 - \frac{\ell^4}{4r^2}\right)^{n/2} \mathrm{d}\ell. \quad (4)$$

The integrand reaches the maximum r^n at $\ell = \sqrt{2}r$ so Eq. (4) can be upper bounded by $O(n\sqrt{n})r^{2n}\,(2r-1)$. For the desired probability, we consider the ratio of volumes, which is at most $O(n\sqrt{n})r^{2n}/\mathrm{Vol}_n(R\mathcal{B}^n) = O(n\sqrt{n})r^n$.

Alternatively, there is also a decisional version of BDD.

Problem 2.20. *(Decision-BDD) For $r > 0$, let $\chi : \mathbb{R}^n \to [0, 1]$ be some distribution with norm concentrated around $r\lambda_1(\Lambda)$.* Decision Bounded Distance Decoding *is the*

task of, given a lattice Λ, deciding correctly with high probability whether a target $\mathbf{t}$ *is sampled from* $\chi \pmod{\Lambda}$ *or* $U(\mathbb{R}^n/\Lambda)$.

Usually, in search-BDD and decision-BDD, the error distribution χ is a uniform distribution on a sphere or ball of some radius R, or a Gaussian of some width s. The hardness of search-BDD and decision-BDD then depends on the radius or width, i.e., the bigger the radius or width is, the larger the average error is and thus the harder the problem instances are.

Solving Decision-BDD. The idea of using short dual vectors for solving decision-BDD can be traced back at least to [2] in the lattice literature, and can be viewed as a lattice analog to an old decoding technique [3,25,39]. Given a BDD sample $\mathbf{t} = \mathbf{v} + \mathbf{e}$ with $\mathbf{v} \in \Lambda$, for any dual vector $\mathbf{w} \in \Lambda^\vee$ one has,

$$\langle \mathbf{t}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{e}, \mathbf{w} \rangle \equiv \langle \mathbf{e}, \mathbf{w} \rangle \pmod{1}. \tag{5}$$

In particular, if the error $\mathbf{e}$ and the dual vector $\mathbf{w}$ are of small enough ℓ_2-norm, $\langle \mathbf{t}, \mathbf{w} \rangle$ should be close to an integer. Moreover, $\langle \mathbf{t}, \mathbf{w} \rangle \pmod 1$ is thus well-defined for any $\mathbf{t} \in \mathbb{R}^n/\Lambda$. A natural score to consider as some indication of the target $\mathbf{t}$ being close to the lattice Λ is therefore given by,

$$f_{\mathbf{w}}(\mathbf{t}) := \frac{\chi_{\mathbf{w}}(\mathbf{t}) + \chi_{-\mathbf{w}}(\mathbf{t})}{2} = \cos(2\pi \cdot \langle \mathbf{t}, \mathbf{w} \rangle), \tag{6}$$

reusing $\chi_{\mathbf{w}}$ from Lemma 2.11.

If a target $\mathbf{t}$ is indeed close to the lattice, $f_{\mathbf{w}}(\mathbf{t})$ should be close to 1, but the converse does not need to be true. To improve confidence in the fidelity of this score, one may naturally consider the total score over many dual vectors $\mathcal{W} \subset \Lambda^\vee$ given by,

$$f_{\mathcal{W}}(\mathbf{t}) := \sum_{\mathbf{w} \in \mathcal{W}} f_{\mathbf{w}}(\mathbf{t}). \tag{7}$$

This function is referred to as the simple distinguisher f_{simple} by Laarhoven and Walter [31], and resembles the Aharonov–Regev [2] distinguisher closely which is given by $f^{\text{AR}}_{\mathbf{w}}(\mathbf{t}) := \rho_{1/s}(\mathbf{w}) f_{\mathbf{w}}(\mathbf{t})$. By checking whether $f_{\mathcal{W}}(\mathbf{t})$ is above or below some optimally chosen threshold, one can say from which distribution $\mathbf{t}$ is drawn, effectively solving decision-BDD with a high success probability.

In carefully crafted circumstances, in particular regarding the construction of the set of short dual vectors $\mathcal{W}$, this approach can give a provable worst-case distinguisher [2] or certificate [28].

More recent works [27,31,33] have reused this idea more heuristically, in a context where $\mathcal{W}$ simply is the set of all the dual vectors smaller than a certain radius (typically given by running a sieve algorithm [10]). In such a situation, all the nonzero dual vectors have approximately the same weight, and [31] shows that the simple distinguisher $f_{\mathcal{W}}(\mathbf{t})$ works asymptotically almost as good as the Aharonov–Regev one $f^{\text{AR}}_{\mathcal{W}}$, although the former can be computed faster in practice.

Reducing Search-BDD to Decision-BDD. Recent dual attacks [23,27,33] and follow-up works reduce search-BDD on Λ to decision-BDD on a (projected) sublattice as follows. Suppose we want, given a target $\mathbf{t}$ sampled by adding a small noise $\mathbf{e}$ to a lattice point $\mathbf{v} \in \Lambda$, to recover $\mathbf{v}$. First, a sparsification $\Lambda' \subset \Lambda$ is chosen. Then, we try to solve decision-BDD on the sublattice Λ' with $\left|\Lambda/\Lambda'\right|$ many targets: $(\mathbf{t} - \mathbf{g})_{\mathbf{g}+\Lambda' \in \Lambda/\Lambda'}$. There is exactly one target drawn from a BDD distribution, namely the target $\mathbf{t}-\mathbf{g}$ with $\mathbf{g} \in \mathbf{v}+\Lambda'$. If decision-BDD is solved successfully, i.e., it has marked only this $\mathbf{g} \in \mathbf{v}+\Lambda'$ as coming from a BDD distribution, search-BDD is solved modulo Λ', i.e., $\mathbf{v} \equiv \mathbf{g} \pmod{\Lambda'}$. One can then solve the easier search-BDD problem on Λ' with target $\mathbf{t} - \mathbf{g}$. Ultimately, after some iterations, $\mathbf{v}$ is completely recovered.

Remark 2.21. Many of the recent dual attacks (for example [27,33] and follow-up works), however, run a sieve on a lattice $L \subset (\Lambda')^\vee$ of much lower rank $\beta_{\text{sieve}} < n$, from which a set of dual vectors $\mathcal{W} \subset L$ is obtained, to solve decision-BDD. Observe that the distinguisher function now satisfies

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

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 $\pi_{\text{span}(L)}(\mathbf{t})$ close to $L^\vee$.

More specifically, in the dual attacks first some lattice reduction (e.g., BKZ reduction) is performed on Λ' to obtain some reduced basis $\mathbf{D}$ for $(\Lambda')^\vee$. Then, $\mathbf{D}_{[1,\beta_{\text{sieve}}]}$ is a basis for L. By duality, then $({}^\vee\mathbf{D})_{[n-\beta_{\text{sieve}}+1,n]}$ is a basis for $L^\vee$, obtained by projecting Λ' away from the first $n - \beta_{\text{sieve}}$ vectors of ${}^\vee\mathbf{D}$.

Note, when $\mathbf{e}$ follows an n-dimensional Gaussian distribution of width s, then $\pi_{\text{span}(L)}(\mathbf{e})$ follows a β_{sieve}-dimensional Gaussian distribution of width s. On the other hand, if we make an incorrect guess $\mathbf{g}$, i.e., having $\mathbf{g} \notin \mathbf{v} + \Lambda'$, we compute the score for the target $\mathbf{t} - \mathbf{g} \equiv \mathbf{e} + (\mathbf{v} - \mathbf{g}) \pmod{\Lambda'}$. Now, the score $f_{\mathcal{W}}(\mathbf{t} - \mathbf{g})$ is high when the point

$$\pi_{\text{span}(L)}(\mathbf{e}) + \pi_{\text{span}(L)}(\mathbf{v} - \mathbf{g}), \tag{8}$$

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 $\pi_{\text{span}(L)}(\mathbf{v} - \mathbf{g})$ acts as a uniform point in $\text{span}(L)/L$. In conclusion, because the distinguisher is only distinguishing for $L^\vee$, we see that the target in (8) corresponds to a BDD sample if $\mathbf{g} \in \mathbf{v} + \Lambda'$, and else to a uniform sample modulo $L^\vee$.

Remark 2.22. We note that the work of [41] sidestepped the uniform model for incorrect targets, by using dual vectors from the full-rank lattice Λ', and making the a priori assumption that $\|\mathbf{e}\| < \frac{1}{2}\lambda_1(\Lambda)$. In this 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 [2]). This is, however, not the regime used in recent concrete attack claims [27,33], and is in fact separated from the contradictory regime in Fig. 3 by a factor 2 on the error length $\|\mathbf{e}\|$.

2.5. *Prior Model*

In this subsection, we explain how the Dual-Sieve attack is analyzed in the works of [27,31,33], and explicitly state where they have used heuristics.

The Analysis of [31].

First, Laarhoven and Walter analyze in [31, Lem. 6] the distribution of the score $f_{\mathbf{w}}(\mathbf{t})$ for BDD targets $\mathbf{t}$ with a distance of exactly r to the primal lattice. In the derivation, however, they approximate this distribution by targets sampled from a continuous Gaussian of parameter $\sigma = r/\sqrt{n}$.

Lemma 2.23. ([31, Lem. 6]) *Let $\Lambda \subset \mathbb{R}^n$ be a full-rank lattice and $\mathbf{w} \in \Lambda^\vee$ be a dual vector.*

(a) If $\mathbf{t} \leftarrow U(\mathbb{R}^n/\Lambda)$, then $\mathbb{E}[f_{\mathbf{w}}(\mathbf{t})] = 0$, and $\mathbb{V}[f_{\mathbf{w}}(\mathbf{t})] = 1/2$,
(b) If $\mathbf{t} \leftarrow \mathcal{N}(0, \sigma^2)^n \pmod{\Lambda}$ with $\sigma \in \mathbb{R}_{>0}$, then

$$\mathbb{E}[f_{\mathbf{w}}(\mathbf{t})] = e^{-2\pi^2\sigma^2\|\mathbf{w}\|^2}, \quad \text{and} \quad \mathbb{V}[f_{\mathbf{w}}(\mathbf{t})] = \frac{1}{2} - \Theta\left(e^{-4\pi^2\sigma^2\|\mathbf{w}\|^2}\right). \tag{9}$$

Proof. For the variance, we will use $f_{\mathbf{w}}(\mathbf{t})^2 = \frac{1}{2} + \frac{1}{2}\cos(4\pi\langle \mathbf{w}, \mathbf{t}\rangle) = \frac{1}{2} + \frac{1}{2} f_{2\mathbf{w}}(\mathbf{t})$.

(a) Integrating over a fundamental region $\mathbf{B} \cdot [0,1]^n$ shows $\mathbb{E}[f_{\mathbf{w}}(\mathbf{t})] = 0$ since $\int_0^1 \cos(\alpha + 2\pi k x)\mathrm{d}x = 0$ holds for all $k \in \mathbb{Z}$ and $\alpha \in \mathbb{R}$. Then by the above, it readily follows,

$$\mathbb{V}[f_{\mathbf{w}}(\mathbf{t})] = \frac{1}{2} + \frac{1}{2}\mathbb{E}[f_{2\mathbf{w}}(\mathbf{t})] = \frac{1}{2}.$$

(b) Because a Gaussian is a radial distribution,

$$\mathbb{E}[f_{\mathbf{w}}(\mathbf{t})] = \mathop{\mathbb{E}}_{x \leftarrow \mathcal{N}(0,\sigma^2)}[\cos(2\pi x\,\|\mathbf{w}\|)] = \exp\left(-2\pi^2\sigma^2\,\|\mathbf{w}\|^2\right),$$

and $\mathbb{V}[f_{\mathbf{w}}(\mathbf{t})] = \frac{1}{2} + \frac{1}{2}\mathbb{E}[f_{2\mathbf{w}}(\mathbf{t})] - \mathbb{E}[f_{\mathbf{w}}(\mathbf{t})]^2 = \frac{1}{2} + \frac{1}{2}\varepsilon^4 - \varepsilon^2$, where $\varepsilon = \exp(-2\pi^2\sigma^2\,\|\mathbf{w}\|^2) \in (0,1)$. □

However, to conclude on the behavior of the total score $f_{\mathcal{W}}(\mathbf{t})$, one needs to resort to some heuristic. In [31], they resort to the following heuristic.

Heuristic 2.24. (Independent Score Heuristic) *For any fixed set $\mathcal{W} \subset \Lambda^\vee$ and for any distribution of targets $\mathbf{t}$ considered in Lemma 2.23, the random variables $(\langle \mathbf{w}, \mathbf{t}\rangle \bmod 1)_{\mathbf{w}\in\mathcal{W}}$ are mutually independent.*

By combining the above heuristic with a central limit approximation—which looks fair, given the exponential size of $\mathcal{W}$ in the context of interest—one can model the total score $f_{\mathcal{W}}(\mathbf{t})$ of each type of sample as a Gaussian of center $|\mathcal{W}| \cdot E$ and variance $|\mathcal{W}| \cdot V$, where E and V are the expectation and variance given by the above Lemma 2.23.

One may then deduce the success probability of distinguishing with the score function using the following lemma.

Lemma 2.25. *Let $X \leftarrow \mathcal{N}(E_X, V_X))$ and $Y \leftarrow \mathcal{N}(E_Y, V_Y)$ be independent Gaussian random variables. Then,*

$$\mathbb{P}[X > Y] = \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{E_X - E_Y}{\sqrt{2(V_X + V_Y)}}\right). \tag{10}$$

Proof. Consider the variable $Z = Y - X$, which is also Gaussian, specifically $Z \sim \mathcal{N}(E_Z, V_Z)$ where $E_Z = E_Y - E_X$ and $V_Z = V_X + V_Y$. The event $X > Y$ is equivalent to $Z < 0$, so the result follows from Eq. (1). □

This lemma leads Laarhoven and Walter to the following heuristic claim.

Heuristic Claim 2.26. (cf. [31, Lem. 9]) *Let $\Lambda \subset \mathbb{R}^n$ be a random unit-volume lattice, $r > 0$ and $\mathcal{W} \subset \Lambda^\vee$ a set consisting of the α^n shortest vectors of $\Lambda^\vee$, where $\alpha = \min\{\beta \mid e^2 \ln(\beta) = \beta^2 r^2\}$. Then, we have*

$$\mathbb{P}[f_{\mathcal{W}}(\mathbf{t}_{\mathrm{BDD}}) > f_{\mathcal{W}}(\mathbf{t}_{\mathrm{unif}})] \geq \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{1}{\sqrt{2}}\right) \approx 0.84,$$

where $\mathbf{t}_{\mathrm{unif}} \leftarrow U(\mathbb{R}^n/\Lambda)$ and $\mathbf{t}_{\mathrm{BDD}} \leftarrow U\left(r\mathrm{GH}(n)\,\mathcal{S}^{n-1}\right)$ are sampled independently.

Heuristic Justification. First, we approximate the BDD sample $\mathbf{t}_{\mathrm{BDD}}$ by a Gaussian distribution of parameter $\sigma = r\mathrm{GH}(n)/\sqrt{n}$. The lengths of the vectors in $\mathcal{W}$ are concentrated around $\alpha \cdot \mathrm{GH}(n)$ according to the Gaussian Heuristic. By the Independent Score Heuristic and Lemma 2.23, the score function for the BDD sample follows a Gaussian distribution $\mathcal{N}(E_X, V_X)$, where

$$E_X = \alpha^n \exp\left(-\frac{2\pi^2\alpha^2 r^2 \cdot \mathrm{GH}(n)^4}{n}\right) = \alpha^n \exp\left(-\frac{\alpha^2 r^2 \cdot n}{2e^2}\right) = \alpha^{n/2},$$

by construction of α. The variance is $V_X \approx \frac{1}{2} \cdot \alpha^n$.

On the other hand, uniform samples give a score distribution Y following a Gaussian distribution $\mathcal{N}(E_Y, V_Y)$ where $E_Y = 0$ and $V_Y = \alpha^n/2$ by case (a) of Lemma 2.23.

Hence by Lemma 2.25, the probability of having $X > Y$ equals

$$\frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{\alpha^{n/2}}{\sqrt{2 \cdot \alpha^n}}\right) \approx 0.84.$$

The analysis of [27] *and* [33]. The analysis proposed by Guo–Johansson [27] is somewhat less explicit. Instead of analyzing the score directly, they consider the statistical distance between the distribution of $\langle \mathbf{t}, \mathbf{w} \rangle \bmod 1$ for $\mathbf{t}$ uniform and Gaussian.

Using the same Independent Score Heuristic, they then conclude on the statistical distance between the tuples $(\langle \mathbf{t}, \mathbf{w} \rangle \bmod 1)_{\mathbf{w} \in \mathcal{W}}$ for $\mathbf{t}$ uniform and Gaussian. While

there exist optimal distinguishers (introduced in [31] using a lemma dating back to Neyman-Pearson [37]), it differs from the score function $f_{\mathcal{W}}$, but they seem to assume that the scoring function is not that far from optimal. An argument for such a statement is given by Laarhoven and Walter [31, Corollary 2], but it is not mentioned by Guo and Johansson [27].

The analysis of MATZOV [33] is on the contrary quite explicit on computing the distribution of scores, while taking into account the severe technical complications introduced by their *modulus switching*. Namely, they increase the number of dual vectors with a factor D_{round} in [33, Section 5] to account for the effect of rounding the Fourier coefficients after performing a modulus switch. Another factor D_{arg} is also introduced, but we view it as dubious, see [18, App. A.4].

In any case, we can essentially recover the key claims of [27,33] directly from the above analysis as well. Note that we express the result more generally in terms of BDD in an arbitrary lattice rather than a specific LWE instance. Moreover, we state this key claim here without loss of generality for a unit-volume, full-rank lattice, because of renormalization and because the dual vectors used in distinguishing actually come from the dual of a projected sublattice. Indeed, the general setting of the Dual-Sieve attack [7,23,27,33] first applies BKZ reduction on the dual lattice $\Lambda_{\text{LWE}}{}^{\vee}$ of dimension d, and then runs a sieve on a lower-rank sublattice $\Lambda^{\vee} \subset \Lambda_{\text{LWE}}{}^{\vee}$, see Remark 2.21.

Heuristic Claim 2.27. (Key claim of [27,33], reconstructed) *Let $\Lambda \subset \mathbb{R}^n$ be a random unit-volume lattice, $\mathcal{W} \subset \Lambda^{\vee}$ the set consisting of the $(4/3)^{n/2}$ shortest vectors of $\Lambda^{\vee}$, and $\sigma > 0$. Taking $\ell = \sqrt{4/3} \cdot \mathrm{GH}(n)$ and $\varepsilon = \exp(-2\pi^2\sigma^2\ell^2)$, we then have*

$$\mathbb{P}\left[f_{\mathcal{W}}\left(\mathbf{t}_{\text{BDD}}\right) > f_{\mathcal{W}}(\mathbf{t}_{\text{unif}})\right] \geq 1 - e^{-\frac{1}{2}|\mathcal{W}|\varepsilon^2},$$

where $\mathbf{t}_{\text{unif}} \leftarrow U(\mathbb{R}^n/\Lambda)$ and $\mathbf{t}_{\text{BDD}} \leftarrow \mathcal{N}\left(0, \sigma^2\right)^n$ are sampled independently.

Heuristic Justification. Justification is similar to that of Heuristic Claim 2.26. The score distribution for $\mathbf{t}_{\text{BDD}}$ is approximately $X \sim \mathcal{N}\left(\varepsilon \cdot |\mathcal{W}|, \frac{1}{2}|\mathcal{W}|\right)$, as lengths of vectors in $\mathcal{W}$ are concentrated around $\ell = \sqrt{4/3} \cdot \mathrm{GH}(n)$ by the Gaussian Heuristic. The uniform sample $\mathbf{t}_{\text{unif}}$ has a score distribution of approximately $Y \sim \mathcal{N}\left(0, \frac{1}{2}|\mathcal{W}|\right)$. Thus by Lemma 2.25,

$$\mathbb{P}\left[f_{\mathcal{W}}\left(\mathbf{t}_{\text{BDD}}\right) > f_{\mathcal{W}}\left(\mathbf{t}_{\text{unif}}\right)\right] = 1 - \frac{1}{2}\operatorname{erfc}\left(\sqrt{|\mathcal{W}|/2} \cdot \varepsilon\right).$$

If $\varepsilon\sqrt{|\mathcal{W}|} \leq 1/\sqrt{2\pi}$, the above probability is trivially at least $\frac{1}{2} \geq 1 - e^{-\frac{1}{2}|\mathcal{W}|\varepsilon^2}$. In the other case, Lemma 2.1 yields the claim:

$$\mathbb{P}\left[f_{\mathcal{W}}\left(\mathbf{t}_{\text{BDD}}\right) > f_{\mathcal{W}}\left(\mathbf{t}_{\text{unif}}\right)\right] \geq 1 - \frac{1}{\sqrt{2\pi\,|\mathcal{W}|} \cdot \varepsilon} e^{-\frac{1}{2}|\mathcal{W}|\varepsilon^2} > 1 - e^{-\frac{1}{2}|\mathcal{W}|\varepsilon^2}.$$

An important corollary of the above claim is that one can distinguish one BDD sample from $O(e^{\frac{1}{2}|\mathcal{W}|\varepsilon^2})$ uniform samples with constant probability, by marking the sample that evaluates to the highest score as the BDD sample. In the case of search-BDD, then with

constant probability, one can recover the lattice vector $\mathbf{v}$ closest to $\mathbf{t}$ modulo the sparsified lattice Λ', if the sparsification factor is bounded by the above quantity.

3. Generalization of the Dual-Sieve-FFT Attack

As established by the literature [2,28,31], scoring target points to obtain information about their distance to a primal lattice Λ using short dual vectors is very general, and not limited to LWE lattices. In this section, we will show that this is also the case of the extra FFT trick as proposed in recent work of Guo and Johansson [27].

3.1. *Abstracting the Dual-Sieve-FFT Attack of Guo–Johansson*

The general idea is as follows. Given a lattice Λ, one first crafts a sparsification Λ' of Λ, i.e., a sublattice $\Lambda' \subset \Lambda$ of finite index. This gives rise to a finite abelian group of cosets $G := \Lambda/\Lambda'$. Now, to solve BDD for a target $\mathbf{t} \in \mathbb{R}^n/\Lambda$ on the lattice Λ, one solves BDD for the target $\mathbf{t}$ on all the cosets $\Lambda' + g$, or equivalently, solve BDD for all the targets $\mathbf{t} - g$ with $g + \Lambda' \in G$ on the sublattice Λ'. For the correct choice of coset $g + \Lambda'$, the distance to $\mathbf{t}$ is the same as in the initial BDD problem, but the sublattice is sparser, making this BDD problem easier than the original one. However, we now have $|G|$ instances to consider.

With the help of the DFT, the score function $f_{\mathcal{W}}$ can be computed for all targets $\mathbf{t} - g$ in a batch. That is, applying the DFT_G in Eq. (3) on a sequence $(\chi_{\mathbf{w}'}(g - \mathbf{t}))_{g\in G}$ for some $\mathbf{w}' \in (\Lambda')^\vee$ gives another sequence that at index $\chi_{\mathbf{w}} \in \widehat{G}$ has a value of,

$$\sum_{g\in G} \chi_{\mathbf{w}'}(g - \mathbf{t})\overline{\chi_{\mathbf{w}}(g)} = \chi_{\mathbf{w}'}(-\mathbf{t}) \cdot \sum_{g\in G} \frac{\chi_{\mathbf{w}'}(g)}{\chi_{\mathbf{w}}(g)}, \tag{11}$$

where $\mathbf{w} + \Lambda^\vee \in (\Lambda')^\vee/\Lambda^\vee$ as $\widehat{G}$ is isomorphic to $(\Lambda')^\vee/\Lambda^\vee$ by Lemmata 2.11 and 2.12. Note that $\chi_{\mathbf{w}'}(g)$ is well-defined for $g \in \Lambda/\Lambda'$ as $\chi_{\mathbf{w}'}$ is Λ'-periodic. By the orthogonality of characters, note that

$$\sum_{g\in G} \frac{\chi_{\mathbf{w}'}(g)}{\chi_{\mathbf{w}}(g)} = \begin{cases} |G|, & \text{if } \mathbf{w}' \in \mathbf{w} + \Lambda^\vee, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, Eq. (11) is zero everywhere except at index $\chi_{\mathbf{w}'}$, where it is equal to $|G| \cdot \chi_{\mathbf{w}'}(-\mathbf{t})$.

By $\mathbb{C}$-linearity of the DFT, one can obtain an expression for the DFT of $f_{\mathcal{W}}(g - \mathbf{t})$ for any finite set of dual vectors $\mathcal{W} \subset (\Lambda')^\vee$. More specifically, if for all $\mathbf{w} \in \mathcal{W}$ we have $-\mathbf{w} \in \mathcal{W}$, i.e., it is *symmetric*, we have

$$\mathrm{DFT}_G\left(\left(f_{\mathcal{W}}(\mathbf{t} - g)\right)_{g\in G}\right) = |G| \cdot \Big(\sum_{\mathbf{w}'\in\mathcal{W}\cap(\mathbf{w}+\Lambda^\vee)} f_{\mathbf{w}'}(\mathbf{t})\Big)_{\mathbf{w}+\Lambda^\vee\in\widehat{G}}.$$

Neglecting this scalar $|G|$, we therefore construct a batch of score functions, by performing an inverse FFT on the sequence $\sum_{\mathbf{w}'\in\mathbf{w}+\Lambda^\vee} f_{\mathbf{w}'}(-\mathbf{t})$ for all (dual) cosets $\mathbf{w} + \Lambda^\vee \in$

$(\Lambda')^\vee/\Lambda^\vee$. Then, the entry with $g + \Lambda' \in G$ that has the highest score is most likely the coset $g + \Lambda'$ containing the lattice point in Λ that is closest to $\mathbf{t}$.

3.2. *Implementation of the General Dual-Sieve-FFT Attack*

In this section, we will give a concrete implementation of an algorithm that performs the general Dual-Sieve-FFT attack on a lattice Λ.

Concretely, the lattice Λ is specified by a basis $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_n]$, and one can take a simple sparsification Λ' being specified by the basis $[d_1\mathbf{b}_1, \ldots, d_n\mathbf{b}_n]$ for suitable $d_1, \ldots, d_n \in \mathbb{N}$.

In fact, any sparsification is, after a basis change, of this shape. When the sublattice $\Lambda' = \mathbf{B}' \cdot \mathbb{Z}^n \subset \mathbf{B} \cdot \mathbb{Z}^n = \Lambda$ is described by a matrix $\mathbf{B}'$, we can express the basis $\mathbf{B}'$ in terms of $\mathbf{B}$, i.e., find the matrix $\mathbf{A} \in \mathbb{Z}^{n\times n}$ such that $\mathbf{B}' = \mathbf{B} \cdot \mathbf{A}$.

Then, put $\mathbf{A}$ in the Smith Normal Form, i.e., find matrices $\mathbf{S}, \mathbf{T} \in \mathrm{GL}_n(\mathbb{Z})$ and a diagonal matrix $\mathbf{D}$ such that $\mathbf{A} = \mathbf{SDT}$ and thus, we have $\mathbf{B}'\mathbf{T}^{-1} = (\mathbf{BS}) \cdot \mathbf{D}$. As $\mathbf{A}$ was full-rank, $\mathbf{D}$ is full-rank, i.e., invertible over $\mathbb{Q}$, so here Λ' is described by the basis $\left[d_1\mathbf{b}'_1, \ldots, d_n\mathbf{b}'_n\right]$, where $\mathrm{diag}(d_1, \ldots, d_n) = \mathbf{D}$ and $\mathbf{BS} = \left[\mathbf{b}'_1, \ldots, \mathbf{b}'_n\right]$ is a basis for Λ. Note that the Smith Normal Form is efficiently computable [46].

Hence, without loss of generality, we have a sparsification $\Lambda' \subset \Lambda$, where $\mathbf{B}$ is a basis for Λ and $\mathbf{B}' = [d_1\mathbf{b}_1, \ldots, d_n\mathbf{b}_n]$ is a basis for Λ'. Then Algorithm 1 will find the coset of Λ' that is most likely to contain a lattice vector closest to target $\mathbf{t}$.

Algorithm 1 DualFFT($\mathbf{B}, \mathbf{B}', \mathcal{W}, \boldsymbol{t}$)

Require:
a basis $\mathbf{B}$ of a full-rank lattice $\Lambda \subset \mathbb{R}^n$,
a basis $\mathbf{B}' = \mathbf{B} \cdot \mathrm{diag}(d_1, \ldots, d_n)$ of $\Lambda' \subset \Lambda$,
a set of short dual vectors $\mathcal{W} \subset (\Lambda')^\vee$,
a target $\boldsymbol{t} \in \mathbb{R}^n/\Lambda'$.
Ensure: a lattice coset $g \in \Lambda/\Lambda'$ closest to $\boldsymbol{t}$

Initialize a table T with zeros of dimension $d_1 \times d_2 \times \cdots \times d_n$
for $\boldsymbol{w} \in \mathcal{W}$ **do**
  Write $\boldsymbol{w} \equiv \frac{j_1}{d_1}\boldsymbol{b}_1^\vee + \cdots + \frac{j_n}{d_n}\boldsymbol{b}_n^\vee \pmod{\Lambda^\vee}$, where $0 \le j_i < d_i$
  $\mathsf{T}[j_1, j_2, \ldots, j_n] \leftarrow \mathsf{T}[j_1, j_2, \ldots, j_n] + \cos(2\pi\,\langle \boldsymbol{w}, \boldsymbol{t}\rangle)$
end for
$\mathsf{S} = \mathrm{DFT}^{-1}_{\mathbb{Z}/d_1\mathbb{Z}\times\cdots\times\mathbb{Z}/d_n\mathbb{Z}}(\mathsf{T})$
$(j_1, j_2, \ldots, j_n) \leftarrow \underset{\substack{k_1,k_2,\ldots,k_n \\ 0\le k_i<d_i}}{\mathrm{argmax}} \{\mathsf{S}[k_1, k_2, \ldots, k_n]\}$
return $j_1\boldsymbol{b}_1 + \cdots + j_n\boldsymbol{b}_n$

Structure of the Quotient Group. From a geometric perspective, concerning the length of the vectors in $\mathcal{W}$, the structure of the group $G = \Lambda/\Lambda'$ does not appear to matter at all, only its size does. On the other hand, while asymptotically all group structures allow to compute DFT_G in time $O(|G| \log |G|)$, the structure of the group matters quite a lot in practice and the case $G = (\mathbb{Z}/2\mathbb{Z})^k$, i.e., the Walsh–Hadamard Transform, should definitely be the best choice. That is, one should construct the sublattice Λ' as generated by $\mathbf{B}' = \left[2\mathbf{b}_1, \ldots, 2\mathbf{b}_k, \mathbf{b}_{k+1}, \ldots, \mathbf{b}_n\right]$, which has index 2^k in Λ.

Randomized Sparsification. Note that the analysis of the length of the vectors in $\mathcal{W}$ requires applying Gaussian Heuristic to the densification of the dual, induced by the sparsification of the primal.

This might require care. Indeed, if the basis is well reduced before we apply the dual densification $\mathrm{diag}\left(\frac{1}{d_1}, \ldots, \frac{1}{d_n}\right)$, this might create a dual lattice which is not random-looking; in particular it might contain a few vectors shorter than predicted by Gaussian Heuristic, with an unclear impact on the rest of $\mathcal{W}$.

We do not expect this to be an issue if the basis $\mathbf{B}$ is adequately randomized before constructing the sparsification.

3.3. *Advantages of the Generalization*

Not only is it theoretically more satisfying to apply the FFT trick to the general decoding problem rather than to the specific LWE problem, it also makes recursion more straightforward.

Shorter Dual Vectors. The algorithm of Guo and Johansson [27] seems to perfectly fit this formalization, where the basis $\mathbf{B}$ is the standard basis associated with the q-ary representation of the lattice, and $\mathbf{B}' = \mathbf{B} \cdot \mathrm{diag}(\gamma, \ldots, \gamma, 1, \ldots 1)$ with k many γ's. The set of short dual vectors is obtained by first running BKZ reduction with block size β_{BKZ} on the dual of $\mathbf{B}'$, and then sieving in the sublattice generated by the first β_{sieve} vectors of this reduced dual basis. The impact of the sparsification $\mathbf{B}' = \mathbf{B} \cdot \mathrm{diag}(\gamma, \ldots, \gamma, 1, \ldots 1)$ on the length of the vectors in $\mathcal{W}$ is that these are shortened by a factor $\gamma^{k/n}$. That is, the sparsification has been *diluted* over n many dimension.

Instead, consider first applying dual-BKZ reduction with block size β_{BKZ} on $\mathbf{B}$, and then taking $\mathbf{B}' = \mathbf{B} \cdot \mathrm{diag}(1, \ldots 1, \gamma, \ldots, \gamma)$. In this way, the densification of the reversed dual basis is given by ${}^{\vee}(\mathbf{B}') = {}^{\vee}\mathbf{B} \cdot \mathrm{diag}\left(\frac{1}{\gamma}, \ldots, \frac{1}{\gamma}, 1, \ldots 1\right)$ remains concentrated on the k first dual vectors![4] Therefore, the sparsification now makes the vectors in $\mathcal{W}$ shorter by a factor $\gamma^{k/\beta_{\mathrm{sieve}}}$ (assuming $k \leq \beta_{\mathrm{sieve}}$).

We leave the concrete exploitation of this improvement to the Dual-Sieve-FFT attack as future work, because the next section will invalidate the analysis of [27,33], so we first need a fixed analysis before thinking about this possible improvement. A visualization of this improvement can be found in Fig. 1.

Remark 3.1. The algorithm of MATZOV [33] differs quite a bit from that of Guo–Johansson [27] by resorting to a modulus switching technique, and it is claimed that this technique allows to decrease dimension at the cost of some extra error. We remain rather circumspect regarding a perceived superiority of the variant of MATZOV [33] over that of Guo–Johansson [27]. For more details, see [18, App. A.6].

[4] The warning on randomized sparsification from Sect. 3.2 still applies here; the basis randomization should be applied to the block $\mathbf{B}_{[n-\beta_{\mathrm{sieve}}+1,n]}$.

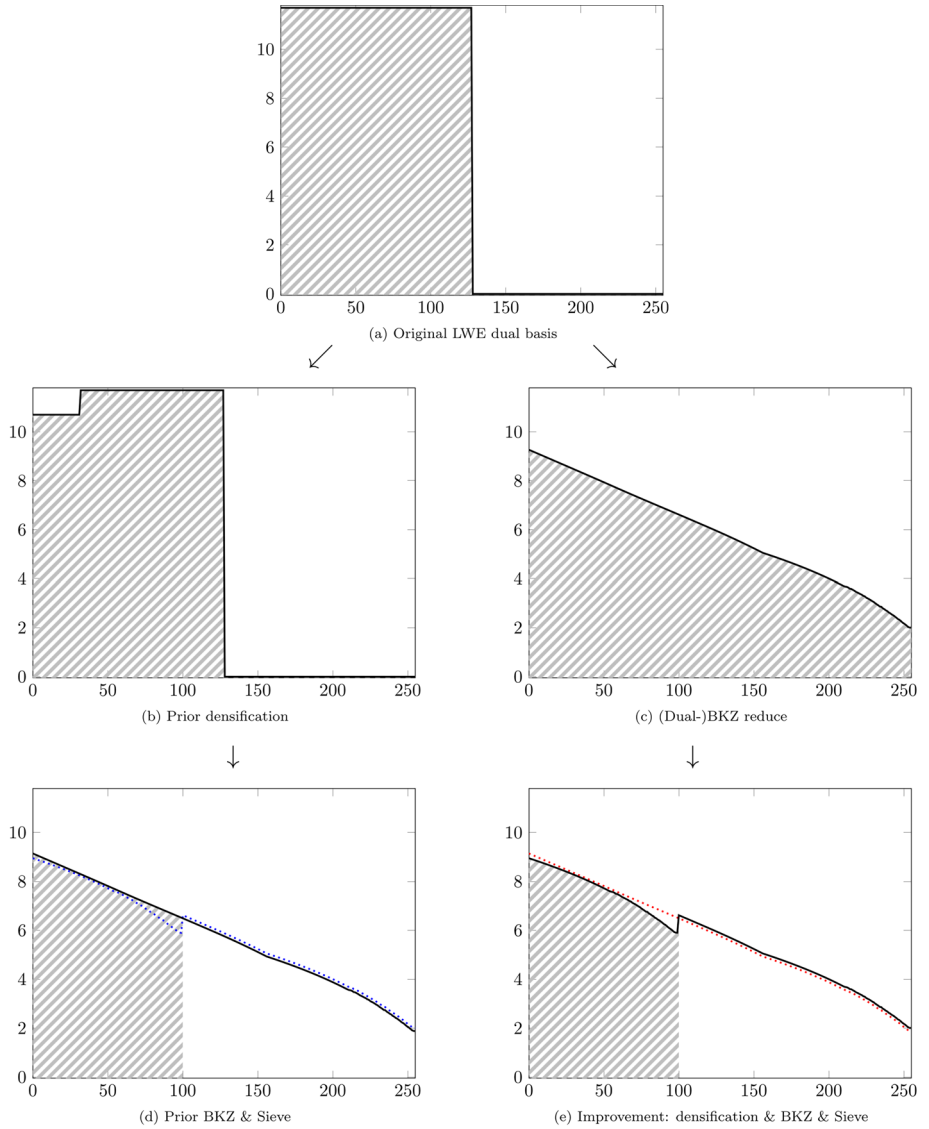


Fig. 1. Simulation of dual basis profile, i.e., $(\log_2 \|\mathbf{d}_i^*\|)_{i=0}^{n-1}$ as a function of the basis index i using the BKZ 2.0 simulator [15]. The used parameters are $n = 128$, $\gamma = 2$, $k = 32$, $\beta_{\mathrm{BKZ}} = \beta_{\mathrm{sieve}} = 100$, $q = 3329$. Prior work follows the path (a) → (b) → (d); the suggested improvement follows the path (a) → (c) → (e). The filled regions in (d) and (e) represent the log volume of the sublattice used for sieving, and the dotted line shows (e) and (d), respectively, for comparison. *Note: the densification in (e) is done by shortening* $({}^{\vee}\mathbf{B})_{[\beta-k,\beta-1]}$ *by a factor* γ. *See Sect.* 3.2 §*Randomized Sparsification regarding the lengths of dual vectors in* $\mathcal{W}$.

4. Contradictions to the Independent Score Heuristic

In the following, we will show two regimes where the analyses of [27,31,33] give rise to absurd conclusions, when using the Independent Score Heuristic. Both are concerned with distinguishing between a BDD target and a uniform target.

In the first regime, we consider BDD samples where the expected distance to the lattice exceeds the Gaussian Heuristic, a task that was recently proved statistically impossible in a random lattice [16].

The second one is concerned with the case of finding a planted solution among many candidates. For certain parameters, the analysis of [27,33] predicts a successful guess of the desired target, despite the existence of many other candidates that are even closer to the lattice than the planted solution. We would like to stress that this contradiction is independent of whether one uses the FFT trick or not, but merely arises from the large number of candidates that is used. Phrased differently, the works of [27,33] predict a much lower failure probability on distinguishing between a BDD and uniform target than what one can reasonably assume.

Recall that in this section, as in [31], we implicitly renormalize the lattice Λ to have volume 1.

4.1. *Distinguishing the Indistinguishable*

Set Up. Recall that the main result of Laarhoven and Walter [31, Lem. 9], reformulated as Heuristic Claim 2.26, provides an algorithm to distinguish between a BDD instance at distance $r \cdot \mathrm{GH}(n)$, and a uniform target modulo the lattice. After a precomputation depending solely on the lattice Λ, this algorithm has exponential complexity α^n, where α satisfies $e^2 \ln(\alpha) = \alpha^2 r^2$, and the heuristic analysis claims that it is successful with constant probability $p > \frac{1}{2}$. That is, the algorithm outputs "BDD" with probability p if its input is a sampled BDD instance $\mathbf{t}$, and it outputs "uniform" with probability p if its input $\mathbf{t}$ is sampled uniformly modulo the lattice.

Remark 4.1. The algorithm here only gets the sample $\mathbf{t}$ as input, whereas Heuristic Claim 2.26 considers an algorithm that gets as input $\mathbf{t}_{\mathrm{BDD}}$ *and* $\mathbf{t}_{\mathrm{unif}}$ without knowing which is which. By saying "BDD" if the score for $\mathbf{t}$ is at least the threshold value $\frac{1}{2}\mathbb{E}[f_{\mathcal{W}}(\mathbf{t}_{\mathrm{BDD}})]$ and "uniform" otherwise, one derives a success probability of $\frac{1}{2} + \frac{1}{2}\,\mathrm{erf}(\frac{1}{2}) \approx 0.76$.

Lemma 4.2. *The equation $e^2 \ln(x) = x^2 r^2$ admits a real solution in x if and only if $r^2 \leq e/2$.*

Proof. The statement and its proof are illustrated by Fig. 2. First, note that $x \mapsto e^2 \ln(x)$ is concave, while $x \mapsto x^2 r^2$ is convex for any $r \in \mathbb{R}$. We discuss three cases.
Case 1: $r^2 = e/2$. The parabola $y = r^2 x^2$ intersects tangentially the curve $y = e^2 \ln x$ at $(x, y) = (\sqrt{e}, \frac{1}{2}e^2)$, with slope $\frac{dy}{dx}\Big|_{x=\sqrt{e}} = e^{3/2}$. By convexity, we have $e^2 \ln(x) < x^2 r^2$ for any $x \neq \sqrt{e}$.

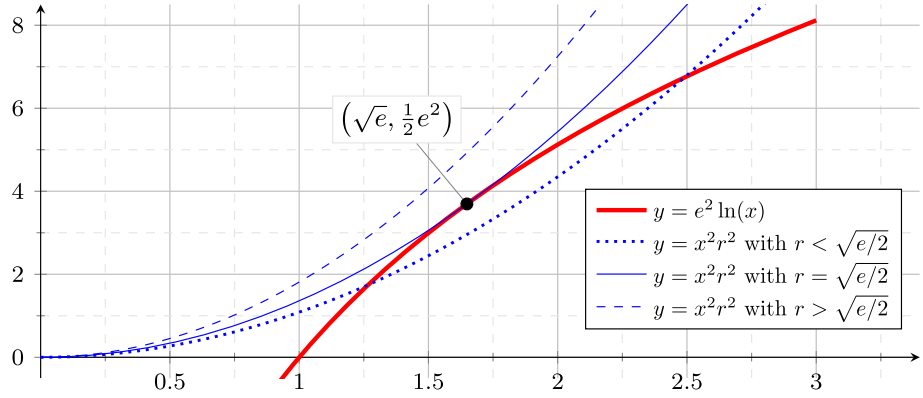


Fig. 2. The equation $e^2 \ln(x) = x^2r^2$ for various r.

Case 2: $r^2 > e/2$. Note that r^2x^2 is strictly increasing in r^2 for any x. Reusing Case 1, we see that $e^2 \ln(x) < x^2r^2$ for all x: there are no solution in that case.
Case 3: $r^2 < e/2$. We have $e^2 \ln(x) > x^2r^2$ at $x = \sqrt{e}$. We also have $e^2 \ln(x) < x^2r^2$ when $x = 1$. The intermediate value theorem then tells there is a solution with $x \in (1, \sqrt{e})$. □

Note that $\sqrt{e/2} \approx 1.1658 > 1$. The fact that the algorithm is supposed to work beyond $r > 1$ raises suspicion: the average number of points at distance at most $r \cdot \mathrm{GH}(n)$ for a uniform target is exactly r^n by the Gaussian Heuristic. More formally, it is a theorem that for any measurable set $V \subset \mathbb{R}^n$, it holds that

$$\mathop{\mathbb{E}}_{\mathbf{t} \leftarrow U(\mathbb{R}^n/\Lambda)} \Big[|(V + \mathbf{t}) \cap \Lambda| \Big] = \frac{\mathrm{Vol}_n(V)}{\det(\Lambda)},$$

where the difference with the Gaussian Heuristic is the presence of a uniformly random shift $\mathbf{t} \leftarrow U(\mathbb{R}^n/\Lambda)$.

Still, one could imagine a scenario where with small probability a target has few close vectors, but most likely it will not, making distinguishing statistically possible.

The Contradiction. It has been shown recently by Debris et al. [16] that the above scenario does not occur with random lattices. More specifically, for a formally defined notion of random lattices and a fixed $r > 1$, it is proved that for errors from a uniform distribution on the ball of radius $r\mathrm{GH}(n)$, the statistical distance between the error modulo the lattice and $U(\mathbb{R}^n/\Lambda)$ is exponentially small as a function of the dimension [16, Prop. 4.3].

That is, for all $r > 1$, no algorithm, whatever its complexity, can even succeed with probability greater than $\frac{1}{2} + O(1) \cdot r^{-n/2}$. Yet, [31, Lem. 9] (reformulated as Heuristic Claim 2.26) claims a constant success probability $p > \frac{1}{2}$. ↯

Discussion. One could counter-argue that the claim [31, Lem. 9] is given for a uniform distribution on a sphere, a case not contradicted by Debris et al. [16]. However, the actual

analysis in [31] is done for a Gaussian distribution, a case which *is* also covered by Debris et al. [16, Thm. 4.6].

The Suspect Heuristic. We note that this counter-argument applies only to the (heuristic) [31, Lem. 9], that is given a single sample, and not to [31, Lem. 8]. Indeed, in the context of the (heuristic) [31, Lem. 8] where exponentially many samples, either all uniform or all BDD, are given, the exponentially small statistical distance can be compensated for with a large number of samples, as discussed between both lemmata.

We note in particular that [31, Lem. 8] does not require the Independent Score Heuristic, as it uses only one dual vector. In fact, after close inspection of the reasoning behind [31, Lem. 8], we could not identify any step that should be too hard to make formally provable, up to minor conditions and small losses in the concrete efficiency of the distinguisher. Indeed, with enough effort, it appears that all the other heuristics and approximations could be dealt with formally.

This sets the Independent Score Heuristic as the prime suspect leading to the erroneous [31, Lem. 9].

4.2. *Candidates Closer than the Solution (Asymptotic)*

Set Up. Recall that the key claim of [27] and [33], reconstructed as Heuristic Claim 2.27 considers the case where the set of dual vectors comes from a lattice sieve [10,36, 38], i.e., it consists of $N = (4/3)^{n/2}$ dual vectors of length $\ell = \sqrt{4/3} \cdot \mathrm{GH}(n)$. Given one uniform sample and one BDD sample with an error sampled from a Gaussian of parameter σ, it was claimed that one can distinguish with a failure probability that is exponentially small in $N\varepsilon^2$, whenever $N \geq 1/\varepsilon^2$, where $\varepsilon = \exp(-2\pi^2\sigma^2\ell^2)$. Let us consider the constraint the other way around by asking the question:

How large can one take σ to have a failure probability of at most p_{fail}?

It can be seen that the failure probability in Heuristic Claim 2.27 is at most p_{fail} when $\varepsilon \geq \sqrt{\frac{\ln(1/p_{\mathrm{fail}}^2)}{N}}$. This constrains σ to satisfy

$$\sigma = \sqrt{\frac{\ln(1/\varepsilon)}{2\pi^2\ell^2}} \leq \frac{\sqrt{\ln N - \ln\ln(1/p_{\mathrm{fail}}^2)}}{2\pi\ell} = \frac{\sqrt{\frac{n}{2}\ln\frac{4}{3} - \ln\ln(1/p_{\mathrm{fail}}^2)}}{2\pi\sqrt{\frac{4}{3}}\cdot\mathrm{GH}(n)}. \tag{12}$$

Using the approximation $\mathrm{GH}(n) \approx \sqrt{\frac{n}{2\pi e}}$ for the Gaussian Heuristic, leads to $\sigma \leq \sqrt{C - \frac{C'\ln\ln(1/p_{\mathrm{fail}}^2)}{n}}$ for some constants $C = \frac{3e\ln(4/3)}{16\pi} \approx 0.047$ and $C' = \frac{3e}{8\pi} \approx 0.32$. This means that Heuristic Claim 2.27 supposedly still distinguishes a BDD sample with an expected distance of $\sqrt{C\cdot n} \approx 0.89\,\mathrm{GH}(n)$ from a uniform sample with a failure probability that is *double-exponentially small*: $p_{\mathrm{fail}} = \exp(-\frac{1}{2}\exp(n^{.99}))$.

The Contradiction (Asymptotic). We will show that the above heuristic claim leads to a contradiction, already for a single exponentially small failure probability of $p_{\mathrm{fail}} = (0.48)^n$.

Lemma 4.3. *Let Λ be a unit-volume lattice, and $r > 0$ such that $r < \frac{\lambda_1(\Lambda)}{2\,\mathrm{GH}(n)}$. Then, for a target $\mathbf{t}$ uniform in $\mathbb{R}^n/\Lambda$, it holds with probability r^n that $\mathbf{t}$ is at distance at most $r\,\mathrm{GH}(n)$ from the lattice.*

Proof. Note that the volume of the ball of radius $r \cdot \mathrm{GH}(n)$ is exactly r^n by definition of $\mathrm{GH}(n)$. Furthermore, because $r \cdot \mathrm{GH}(n) < \lambda_1(\Lambda)/2$, all translations of this ball by points in Λ are disjoint. Said otherwise, this ball does not intersect itself modulo the lattice. More formally, its projection onto the torus $\mathbb{R}^n/\Lambda$ is injective. Hence, the ball modulo the lattice also has volume r^n in $\mathbb{R}^n/\Lambda$. The probability of a uniform target $\mathbf{t}$ falling into that ball is therefore $r^n/\mathrm{Vol}_n(\mathbb{R}^n/\Lambda) = r^n/\det(\Lambda) = r^n$. □

Let us use this lemma in the case of a random unit-volume lattice, or more specifically, one where we expect $\lambda_1(\Lambda) \approx \mathrm{GH}(n)$. Using Lemma 4.3 with $r = 0.49$, the probability that $\mathbf{t}_{\mathrm{unif}}$ lies in the ball of radius $r\,\mathrm{GH}(n)$ is r^n. Assume this event happens. In this case, the uniform target lies at a distance $r\,\mathrm{GH}(n)$ from the lattice, which is much shorter than the distance of $0.89\,\mathrm{GH}(n)$ the BDD sample was away from the lattice. However, the score function $f_{\mathcal{W}}$ is precisely meant to associate a larger score to closer targets, so in this case we expect to have $f_{\mathcal{W}}(\mathbf{t}_{\mathrm{unif}}) > f_{\mathcal{W}}(\mathbf{t}_{\mathrm{BDD}})$, with at least a constant probability. Thus, in fact the failure probability of distinguishing is close to 0.49^n, which is much more likely than the claimed 0.48^n. ↯

Discussion. One might counter-argue that $f_{\mathcal{W}}$ only probabilistically classifies vectors by their distance to the lattice and might somehow still give the particularly close uniform sample a lower score than the BDD sample. However, the probability that the uniform target $\mathbf{t}$ lies at distance $\mathrm{GH}(n)/n^2 = \Theta(n^{-3/2})$ away from the lattice, is n^{-2n}, which is (only) super-exponentially small. In this case we have $\langle \mathbf{t}, \mathbf{w} \rangle \leq O(1/n)$ for any $\mathbf{w} \in \mathcal{W}$ output by a sieve, and approximating the cosine we know the score $f_{\mathcal{W}}(\mathbf{t})$ for this target $\mathbf{t}$ will be at least $N(1 - O(1/n^2))$, which is essentially maximal.

The Suspect Heuristic. The discussion above points to the same suspect as Sect. 4.1, namely, the Independent Score Heuristic. Indeed, under independence the probability of one uniform target reaching a constant fraction of the maximal score N should decrease as fast as $\exp(-\Theta(N))$, but we have shown that this probability is in fact at least $\exp(-\Theta(n \log n))$. Independence cannot hold for such large choices of $N = \omega(n \log n)$, and this should be visible in the tail of the score distribution for uniform targets.

4.3. *Candidates Closer than the Solution (Concrete)*

Set Up. In the contradiction above, we choose p_{fail} as small as 0.48^n to be able to invoke Lemma 4.3 to have the uniform sample at distance $r\,\mathrm{GH}(n) < \frac{1}{2}\lambda_1(\Lambda)$, quantifying the probability that a random target in $\mathbb{R}^n/\Lambda$ falls close to the lattice. This leads to an *overcontradiction*: we analyzed the probability that the uniform target $\mathbf{t}_{\mathrm{unif}}$ is at a distance $0.49\,\mathrm{GH}(n)$ from the lattice, much closer than $\mathbf{t}_{\mathrm{BDD}}$ at distance $0.89\,\mathrm{GH}(n)$. However, even when there is a uniform sample at distance $0.88\,\mathrm{GH}(n)$ from the lattice, $\mathbf{t}_{\mathrm{unif}}$ can get a higher score than the $\mathbf{t}_{\mathrm{BDD}}$.

To extend Lemma 4.3 up to radii $r < 1$, we will resort to a heuristic instead. In this regime, translations of the ball by lattice points may start to intersect. In practice,

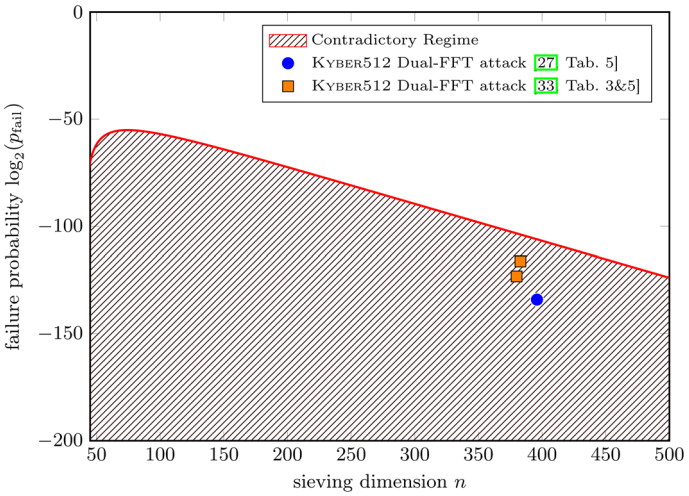


Fig. 3. The concrete contradictory regime, in which Heuristic Claim 2.27 *overpredicts* the success probability of distinguishing in dimension n, compared to a *simple* lower bound derived from Heuristic Claim 4.4. By determining the largest $\sigma \in [0, \mathrm{GH}(n)/\sqrt{n}]$ for which the contradiction arises, we know the upper boundary of the contradictory regime by computing the probability for that σ. For any pair (n, p_{fail}) falling in the concrete contradictory regime, a distinguisher cannot distinguish, with failure probability at most p_{fail}, the uniform distribution from a BDD distribution having σ satisfy Eq. (12). Obtained with script `volumetric_contr.py`.

however, the volume of this intersection remains rather small and should not affect the volume so much.

Heuristic Claim 4.4. *Let Λ be a random unit-volume lattice, and $r \in (0, 1)$. For a target $\mathbf{t}$ sampled uniformly in the torus $\mathbb{R}^n/\Lambda$, we have a probability of $r^n \cdot \left(1 - n^{O(1)} r^n\right)$ that $\mathbf{t}$ is at distance at most $r\,\mathrm{GH}(n)$ from the lattice.*

Heuristic Justification. Contrary to the proof of Lemma 4.3, we now have to subtract from r^n the probability that a target $\mathbf{t}$ has at least two lattice points at distance less than $r\,\mathrm{GH}(n)$ away, which by Heuristic Claim 2.19 happens with probability $O(n\sqrt{n})r^{2n}$.

The Contradiction. The idea is to arrive at a contradiction whenever we instantiate the claim from above with the largest possible $r > 0$ up to which it is more likely to have a uniform target within $r\,\mathrm{GH}(n)$ away from the lattice than the heuristically claimed failure probability p_{fail}.

For a given dimension n and relative distance $r \in (0, 1)$, Heuristic Claim 4.4 lower bounds the probability that we fail to distinguish a BDD sample from a uniform sample. Indeed, when the uniform sample happens to be closer to the lattice than the BDD sample, with constant probability a higher score will be assigned to it. Now by identifying all the $r \in (0, 1)$ for which the lower bound is *bigger* than the upper bound from Heuristic Claim 2.27, we have found a regime in which the two heuristics are in clear contradiction with each other. We will call this regime the "*concrete contradictory regime*," and it is depicted in Fig. 3. ↯

The Contradictory Regime, in the Context of Concrete Attacks Against LWE. Above, we have determined the contradictory regime when obtaining the set $\mathcal{W}$ by a sieve over the full lattice, and assuming its volume was 1. It is not hard to see that scaling the lattice up or down is not going to affect the conclusion. Indeed, the Gaussian Heuristic of the primal, σ and r will scale with the lattice, while the length ℓ of the dual vectors will scale inversely; this leaves ε and p unaffected.

In the context of the cryptanalytic literature [23,27,33], the set $\mathcal{W}$ does not come from the full dual lattice $\Lambda_{\mathrm{LWE}}^{\vee} \subset \mathbb{R}^n$, but comes from a dual lattice $L \subset \Lambda_{\mathrm{LWE}}^{\vee}$ of rank β_{sieve}, by Remark 2.21. Now, targets are projected onto the β_{sieve}-dimensional space spanned by all $\mathcal{W}$. Hence, we effectively run the distinguisher here over a projected sublattice of dimension β_{sieve}.

In this scenario, the contradictory regime is solely determined by β_{sieve} and the needed failure probability p_{fail}, and not by other quantities such as the LWE parameters and β_{BKZ}. Indeed, the LWE parameters and β_{BKZ} are going to influence the volume of the lattice on which we run the final sieve to obtain $\mathcal{W}$.

This might not perfectly be representative of the exact analysis of MATZOV [33] in that we do not make a special analysis of the modulus switching effect on the score distribution. Instead, this treats modulus switching as adding an implicit error, increasing σ. This remains a strong signal on the credibility of the heuristic analysis in that regime.

Another point raising discussion is the fact that our contradiction is established in the case where the uniform targets $\mathbf{t}$ are independent. This is not formally the case when those targets come from a partial enumeration, though such a heuristic has been used in the past, for example, underlying the analysis of the hybrid attack [29]. More critically, we see no mention of such dependence and how they would affect the algorithm in the existing analysis [23,27,33]. While we do not claim that it is impossible, we view the notion that such dependencies could fix the algorithm as quite doubtful, and requiring specific substantiation with analysis and experiments.

In other words, while our contradiction does not formally disprove the recent claims on the Dual-Sieve attack [23,27,33], it does invalidate the reasoning leading to these claims. Because we do not see an obvious reason why this or that detail would solve the issues raised here, it seems reasonable to presume that these claims are indeed incorrect.

The Parameters of Guo–Johansson and MATZOV. We now turn to the instantiations from [27] and [33], focusing on the Kyber512 [44] parameter set, in the "asymptotic model" for dimensions for free.[5]

In [27, Table 5], we find a sieving dimension $\beta_{\mathrm{sieve}} = 396$ (where all the dual vectors come from), a guessing dimension $t_1 = 20$, and an FFT dimension $t = 78$. The guessing part considers all 7 possible values $\{-3, \ldots, 3\}$ of each coordinate, while the FFT is done with $\gamma = 2$, giving rise to $T = 7^{20} \cdot 2^{78} \approx 2^{134.1}$ targets of which all are uniform samples except for one.

In [33, Table 3], we find a sieving dimension $\beta_{\mathrm{sieve}} = 380$ (where all the dual vectors come from), a guessing dimension $k_{\mathrm{enum}} = 19$, and an FFT dimension $k_{\mathrm{fft}} = 34$. The guessing part enumerates over $\{-3, \ldots, 3\}^{k_{\mathrm{enum}}}$ in order of decreasing probability from the used binomial distribution, while the FFT is done with $p = 5$, giving rise to, according to [33], $T = 2^{19 \cdot H(\chi_s)} \cdot 5^{34} \approx 2^{123.3}$ many targets. Using an improved

[5] This is the optimistic estimate in [17]. The other "G6K model" used in [27,33] is debated in [18, App. A.2].

cost metric, they also give [33, Table 5] another set of parameter where $\beta_2 = 383$ and $T = 2^{17 \cdot H(\chi_s)} \cdot 5^{33} \approx 2^{116.3}$.[6]

To get a constant success probability in the dual attack, the distinguisher needs to have a failure probability p_{fail} of the order $1/T$: taking $p_{\text{fail}} = 1/T$ yields a total success probability of $(1 - p_{\text{fail}})^T \geq e^{-p_{\text{fail}} T} = 1/e$ of distinguishing one BDD sample from T (independent) uniform samples.

For both instantiations [27,33] the dual attack is used rather deep in its contradictory regime, as depicted in Fig. 3. ↯

5. Experimental invalidation of the Prior Model

In this section, we provide further substantiation of the concrete contradiction from Sect. 4.3 with experimental evidence. We hope that the experiments will provide insight on what exactly is the issue with the prior analysis, and will show how the issue with the analysis can be resolved. We focus our analysis on the case where $\mathcal{W}$ is the output of a full sieve with a saturation radius of $r_{\text{sat}} = \sqrt{4/3}$ and a saturation ratio of $f_{\text{sat}} = 0.90$.

We look at two distributions: the score for uniform targets, and the score for BDD targets with a Gaussian error. There are two plausible diagnoses for how the contradictory regime appears: the BDD scores are smaller than predicted, or the uniform scores are higher than predicted.

Because the concrete contradictory regime of Fig. 3 starts when p_{fail} is very small, say 2^{-55} even in small dimension, these unpredicted high scores might be very rare. However, these events are important to determine the tail of the uniform score distribution, so the experiments require numerous samples. Naïvely, it would take a long time to run such large scale experiments, but the same FFT trick from [27] (cf. Sect. 3) makes it feasible to run experiments on this scale!

5.1. *Implementation Details*

We used the `G6K` software [5] for running the experiments, using Python on a high-level with a binding to some C code for computing the WHT. For the uniform targets we wrote the script `unif_scores.py`, which computes scores for many points sampled uniformly from $(\mathbb{Z}/q\mathbb{Z})^n$, where $\mathcal{W}$ is the output of a full sieve on the dual of $\mathbf{B} \cdot \text{diag}(2, \ldots, 2, 1, \ldots, 1)$ with the number of 2s equal to the FFT dimension. This setup allowed us to get roughly 2^{25} samples per second per CPU core. The scores were stored in buckets of width 1 while the exceptionally high scores were kept in a list. Here, we sieve using the dual mode built into `G6K` which only works with the dual basis implicitly, cf. https://github.com/fplll/fplll/wiki/FPLLL-Days-5-Summary.

In addition, the BDD scores are obtained with the script `bdd_scores.py`. The script randomly samples a q-ary lattice for the dual lattice $\Lambda^\vee$, so the script did not use the dual mode in `G6K`, making the implementation a little bit easier. Then, it computed the score function for the BDD distribution being a Gaussian of parameter $\sigma = f_{\text{GH}} \cdot \text{GH}(n) / \sqrt{nq}$

[6] We are not quite sure how this quantity was derived, and it seems incorrect. See [18, App. A.5].

for some parameter $f_{\mathrm{GH}} \in (0, 1]$, where $1/\sqrt{q}$ is a normalization factor needed because the dual lattice has determinant $q^{n/2}$.

For the uniform scores, we extracted the $2N = f_{\mathrm{sat}} r_{\mathrm{sat}}^n$ shortest dual vectors from the database. Because `G6K` returns at most one out of $\{\mathbf{w}, -\mathbf{w}\}$, `G6K` provides only N dual vectors of length below $r_{\mathrm{sat}} \cdot \mathrm{GH}(n)$. The other N dual vectors that were extracted, are therefore slightly larger than $r_{\mathrm{sat}} \cdot \mathrm{GH}(n)$, and the whole database is rather concentrated around this length. This does *not* affect our conclusions, as the prediction under the heuristic analysis plotted in Fig. 4 does not depend on the length of dual vectors.

Remark 5.1. This halving of the number of short vectors, which is also in Heuristic 2.13, was missed in the prior works of [27,33] leading to some irrelevant complication in the analysis of [33], as discussed in [18, App. A.5].

5.2. *Distribution of Scores of Uniform Targets*

We measured the score distribution for uniform targets over lattices of various dimension, and plotted our result in Fig. 4. On each of these curves, we see a clear deviation from prediction for rare events: large scores are more likely to occur than predicted. After following a *waterfall* shape, i.e., a quadratic decay in logarithmic scale, the score probability seems to reach a *floor*, where it decays much slower than a normal distribution predicts. This is perfectly in accordance with the contradiction discussed in Sects. 4.2 and 4.3: we start encountering vectors that are quite close to the lattice, which should have a rather high score.

Conclusion. Comparing Figure 3 with the floors appearing in Fig. 4, we conclude that the floor seems to appear earlier than expected by the contradictory regime, vindicating the notion that the analysis might fail even in an earlier regime than our predicted contradiction. As we will see in the next section, the floor behavior is not the only thing that the Independent Score Heuristic mispredicts.

5.3. *Distribution of Scores for BDD Targets*

We measured the score distribution for BDD targets sampled from a Gaussian of parameter $\sigma = 0.7 \cdot \mathrm{GH}(n) / \sqrt{n}$, over lattices of various dimensions n, and plotted our result in Fig. 5. The predicted score distribution is based on the Independent Score Heuristic, but also takes into account the exact lengths of each dual vector in Eq. (9), instead of approximating all the lengths to be equal to $\sqrt{4/3} \cdot \mathrm{GH}(n)$, to make the prediction more accurate.

The first thing one notices is that the distribution is significantly more spread out than predicted. The variance is significantly higher. In fact, the ratio between the actual and predicted variance appears to grow drastically with the dimension, as visible on Table 1.

One might also want to consider the average. However, since the average is linear, its prediction does not require the Independent Score Heuristic. And indeed, the prediction is close to the measured average.

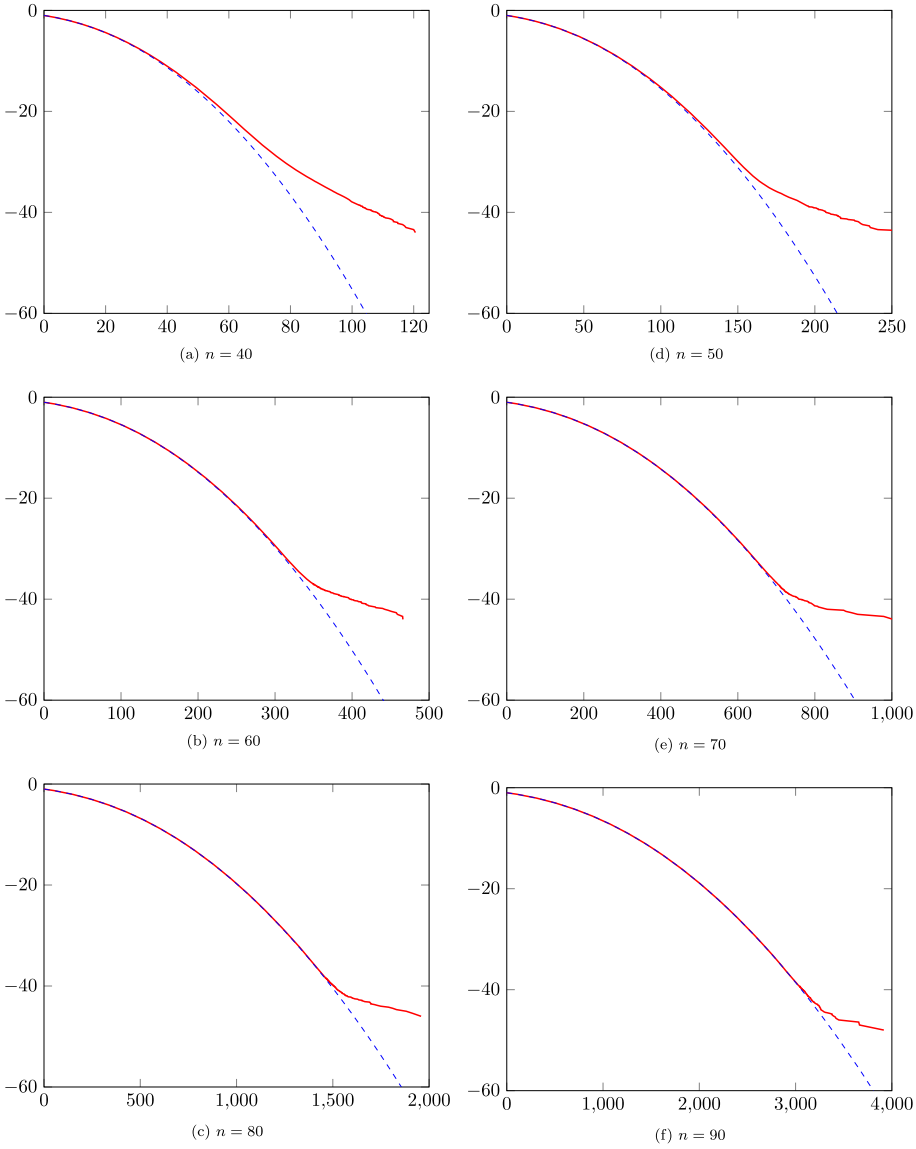


Fig. 4. Binary logarithm of survival function (SF) of the score distribution for uniform targets in various dimensions n, according to prediction from the heuristic analysis (dashed blue line), and experiments (red line). Each experimental curve has 2^{45} samples, which were obtained with `code/unif_scores.py` and are listed in `data/unif_scores_nX.csv` of the auxiliary files.

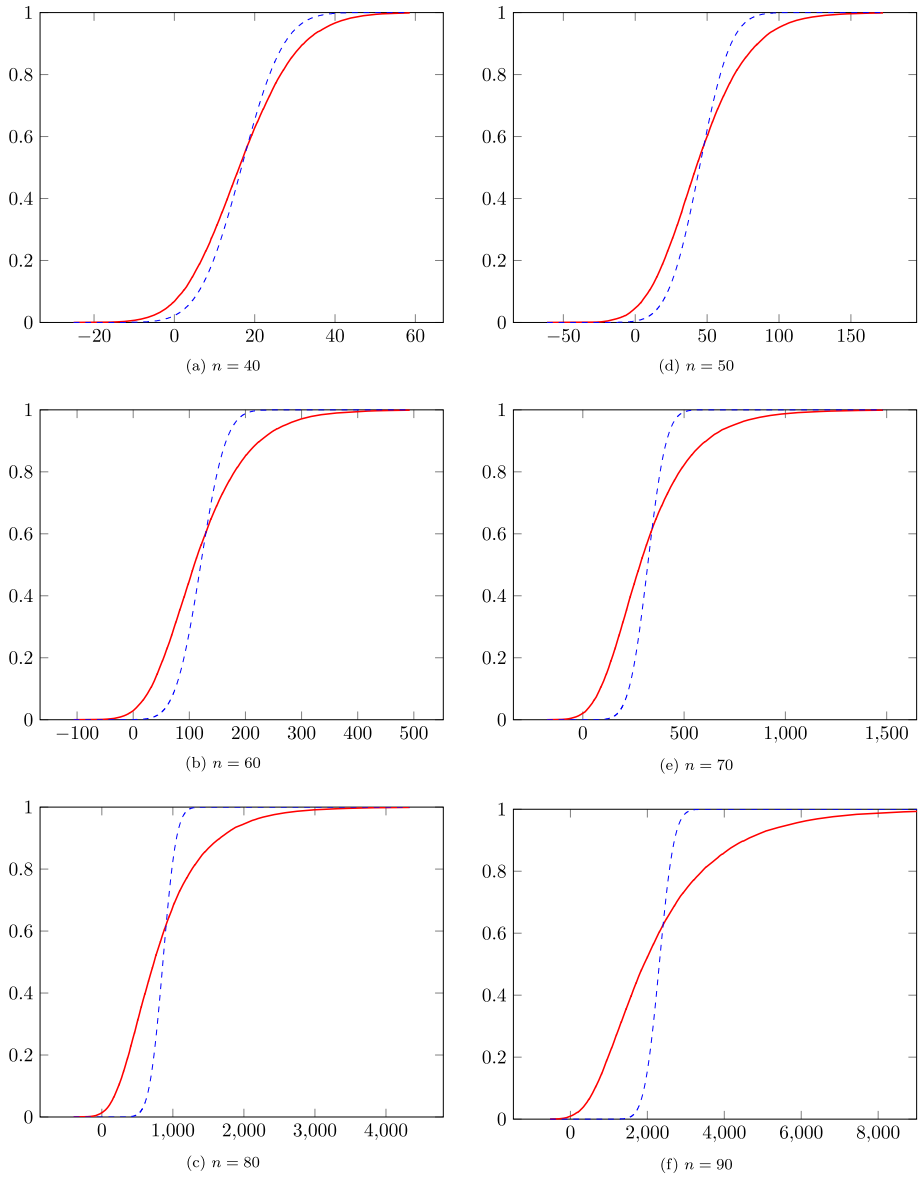


Fig. 5. CDF of the score for Gaussian BDD targets in various dimensions n, with parameter $\sigma = 0.7 \cdot \mathrm{GH}(n)/\sqrt{n}$, according to prediction from the heuristic analysis (dashed blue line), and experiments (red line). Each experimental curve has 2^{15} samples, which were obtained with `code/bdd_scores.py` and are listed in `data/bdd_scores_nX.csv` of the auxiliary files.

Table 1. Variance of the BDD score distribution.

Dimension n	40	50	60	70	80	90
predicted std. dev.	8.31	17.19	35.41	72.80	149.52	307.01
measured std. dev.	11.86	30.24	80.19	221.03	614.16	1732.72
ratio meas./pred.	1.43	1.76	2.26	3.04	4.11	5.64
predicted average	16.71	44.68	120.26	320.53	859.99	2310.28
measured average	16.75	45.32	120.60	322.51	863.15	2325.99
ratio meas./pred.	1.00	1.01	1.00	1.01	1.00	1.01
measured average	16.75	45.32	120.60	322.51	863.15	2325.99
measured median	16.13	42.27	108.74	282.54	735.47	1914.36
ratio med./avg.	0.96	0.93	0.90	0.88	0.85	0.82

A more interesting statistic is the median. According to the heuristic analyses of [31,33] reconstructed in Sect. 2.5, the median is predicted to be equal to the average. In practice, however, the median is noticeably lower. This partially implies that the distribution is quite asymmetric around its average, contrary to the analysis' prediction.

Conclusion. All in all, it is fair to say that the heuristic analyses of [31,33], reconstructed in Sect. 2.5 of the dual attack is completely off when it comes to predicting the score for BDD targets, with or without modulus switching. The distribution is definitely *not* Gaussian, nor even symmetric around its average, and its variance is hugely underestimated.

6. Score Distribution Models

The experiments in Sect. 5 reveal that the Independent Score Heuristic, as used in [27,33], leads to an incorrect prediction for the score distribution of both BDD and uniform targets. In particular, it overestimates the success probability of distinguishing between a BDD target and a uniform targets. The goal of this section is to develop new score predictions that match practice more accurately than the predictions in Sect. 2.5.

In Sect. 6.1, we derive analytic expressions for the mean and variance of an individual score $f_{\mathbf{w}}(\mathbf{t})$. In Sect. 6.2, we derive expressions for the exact mean and variance of the score distribution for a fixed BDD target based on Heuristic 2.13. Here, we present a new heuristic that says the score distribution for a single target $\mathbf{t}$ *is* Gaussian and only depends on $d(\mathbf{t}, \Lambda)$, when the probability is taken over the lattice. Using this heuristic and by integrating over the radius, we predict in Sect. 6.3 the score distribution for a radial target distribution, such as uniform in a ball and Gaussian. Moreover, by approximating the Voronoi cell by a ball of volume $\det(\Lambda)$, we can also predict the score distribution for an error uniform modulo Λ.

Throughout this section, let us assume Λ has full rank and $\det(\Lambda) = 1$ without loss of generality, because the ambient space can be restricted to span (Λ) and any lattice Λ can be rescaled to one of unit-volume.

We strongly advise the reader to think carefully for which lattice Λ to use the predictions of this section. Namely, in many of the recent dual attacks, it is a projected sublattice of Λ_{LWE} having rank β_{sieve}, see Remark 2.21.

6.1. *Individual Score Function*

First, let us fix an arbitrary vector $\mathbf{t} \in \mathbb{R}^n$, and look at the score function $f_{\mathbf{w}}(\mathbf{t}) = \cos(2\pi \langle \mathbf{w}, \mathbf{t} \rangle)$, where $\mathbf{w}$ is sampled uniformly from a sphere. In this case, the mean and variance are expressible in terms of Bessel functions (see Sect. 2).

Lemma 6.1. *Given an arbitrary vector $\mathbf{t} \in \mathbb{R}^n$ and $r \in \mathbb{R}_{>0}$, when the random variable $\mathbf{w}$ follows the uniform distribution on the $(n-1)$-dimensional sphere of radius r, we have*

$$\mathop{\mathbb{E}}_{\mathbf{w} \leftarrow U(r\mathcal{S}^{n-1})}[f_{\mathbf{w}}(\mathbf{t})] = \xi\left(\frac{n}{2} - 1, \|\mathbf{t}\| \cdot r\right),$$
$$\mathop{\mathbb{V}}_{\mathbf{w} \leftarrow U(r\mathcal{S}^{n-1})}[f_{\mathbf{w}}(\mathbf{t})] = \frac{1}{2} + \frac{1}{2}\xi\left(\frac{n}{2} - 1, 2 \cdot \|\mathbf{t}\| \cdot r\right) - \xi\left(\frac{n}{2} - 1, \|\mathbf{t}\| \cdot r\right)^2.$$

Proof. The expectation value follows directly from a classic result from Fourier Analysis, see, e.g., [26, p. 198] or [45, p. 154], as we have

$$\begin{aligned}\mathop{\mathbb{E}}_{\mathbf{w} \leftarrow U(r\mathcal{S}^{n-1})}[f_{\mathbf{w}}(\mathbf{t})] &= \frac{1}{\mathrm{Vol}_n\left(r\mathcal{S}^{n-1}\right)} \int_{r\mathcal{S}^{n-1}} e^{2\pi i \cdot \langle \mathbf{w}, \mathbf{t} \rangle} \mathrm{d}\mathbf{w} \\ &= \frac{\Gamma(\frac{n}{2}) \cdot J_{\frac{n}{2}-1}(2\pi \|\mathbf{t}\| r)}{(\pi \|\mathbf{t}\| r)^{\frac{n}{2}-1}}.\end{aligned}$$

The trigonometric identity $f_{\mathbf{w}}(\mathbf{t})^2 = \frac{1}{2} + \frac{1}{2}\cos(4\pi \langle \mathbf{w}, \mathbf{t} \rangle) = \frac{1}{2} + \frac{1}{2} f_{\mathbf{w}}(2 \cdot \mathbf{t})$ allows us to easily derive a similar expression for the variance, by observing

$$\begin{aligned}\mathop{\mathbb{V}}_{\mathbf{w} \leftarrow U(r\mathcal{S}^{n-1})}[f_{\mathbf{w}}(\mathbf{t})] &= \mathop{\mathbb{E}}_{\mathbf{w}}\left[f_{\mathbf{w}}(\mathbf{t})^2\right] - \mathop{\mathbb{E}}_{\mathbf{w}}[f_{\mathbf{w}}(\mathbf{t})]^2 \\ &= \frac{1}{2} + \frac{1}{2}\mathop{\mathbb{E}}_{\mathbf{w}}[f_{\mathbf{w}}(2 \cdot \mathbf{t})] - \mathop{\mathbb{E}}_{\mathbf{w}}[f_{\mathbf{w}}(\mathbf{t})]^2,\end{aligned}$$

and reusing the result for the expectation value. □

Remark 6.2. The individual score function has been studied before, for example, by Laarhoven and Walter [31, Sect. 4.1]. However, they approximate the distribution for $\mathbf{w}$ by $\mathcal{N}(0, r^2/n)^n$, yielding only an approximation for the expectation value and variance, whereas the above lemma has an exact expression for both. Still, if one is inclined to work with a crude approximation, by Remark 2.7, the expectation in the above lemma is for small values r approximated by $\exp\left(-\frac{2\pi^2 \|\mathbf{t}\|^2 r^2}{n}\right)$, which matches the approximated expectation in [31].

This result can be translated easily to a uniform distribution on a ball, because sampling uniformly from the n-dimensional ball can be done by first sampling from a $(n+1)$-dimensional sphere and then dropping the last two coordinates [9, Corollary 4].

As we will reuse the mean and variance for the ball, let us introduce the following notation.

Definition 6.3. For $n \in \mathbb{Z}_{\geq 1}$, $r_p, r_d \in \mathbb{R}_{>0}$ let us denote

$$E_{n,r_d}(r_p) = \xi\left(\frac{n}{2}, r_d \cdot r_p\right),$$
$$V_{n,r_d}(r_p) = \frac{1}{2} + \frac{1}{2}E_{n,r_d}(2 \cdot r_p) - \left(E_{n,r_d}(r_p)\right)^2.$$

Corollary 6.4. *Given an arbitrary vector $\mathbf{t} \in \mathbb{R}^n$ and $r \in \mathbb{R}_{>0}$, when the random variable $\mathbf{w}$ follows the uniform distribution on the n-dimensional ball of radius r, we have*

$$\mathop{\mathbb{E}}_{\mathbf{w} \leftarrow U(r\mathcal{B}^n)}[f_{\mathbf{w}}(\mathbf{t})] = E_{n,r}(\|\mathbf{t}\|), \quad \text{and} \quad \mathop{\mathbb{V}}_{\mathbf{w} \leftarrow U(r\mathcal{B}^n)}[f_{\mathbf{w}}(\mathbf{t})] = V_{n,r}(\|\mathbf{t}\|),$$

6.2. *Total Score Function*

The next step is to determine the expectation value and variance of the total score $f_{\mathcal{W}}(\mathbf{t})$, in the case that the set $\mathcal{W}$ is obtained by lattice sieving. It is now necessary to take into account the randomness with which the lattice Λ is sampled, because the lattice has influence on the set of dual vectors $\mathcal{W} \subseteq r_d\mathcal{B}^n \cap \Lambda^\vee$.

In this section, let us fix a target $\mathbf{t} \in \mathbb{R}^n$ initially. We will study the score distribution for this target $\mathbf{t}$ by considering the randomness of the lattice and the set of dual vectors $\mathcal{W}$.

Upon first reading the rest of this section, the reader is encouraged to think that the set of dual vectors is $\mathcal{W} = r_d\mathcal{B}^n \cap \Lambda^\vee$, or rather: half of them to break the negation symmetry (recall Sect. 2.2). To argue about the dual vectors, we will make use of Heuristic 2.13, which can be used for more general saturation parameters. When sieving in practice, a saturation ratio $f_{\text{sat}} < 1$ is used to obtain enough dual vectors in the ball, and it is very natural to believe that different saturation parameters do not change the score distribution significantly, other than having an effect on N, the expected size of $\mathcal{W}$. Although all of the shortest vectors are usually found, most of the missing $1 - f_{\text{sat}}$ portion of vectors is around the boundary. Still, we believe this does not noticeably affect the score distribution for reasonably large f_{sat}.

By using Heuristic 2.13, the $\mathbf{w} \in \mathcal{W}$ are i.i.d. samples and match the distribution in Corollary 6.4 too. Because heuristically the scores $(f_{\mathbf{w}}(\mathbf{t}))_{\mathbf{w}\in\mathcal{W}}$ are i.i.d., we expect that the total score is Gaussian for large $|\mathcal{W}|$ by a central limit heuristic. This argument results in the following heuristic.

Heuristic 6.5. *Fix a target $\mathbf{t} \in \mathbb{R}^n$ of norm $\|\mathbf{t}\| \in (0, \mathrm{GH}(n))$. The score distribution $f_{\mathcal{W}}(\mathbf{t})$ is Gaussian with mean $N \cdot E_{n,r_d}(\|\mathbf{t}\|)$ and variance $N \cdot V_{n,r_d}(\|\mathbf{t}\|)$ when sampling Λ and obtaining N dual vectors $\mathcal{W} \leftarrow \mathrm{Sieve}\left(\Lambda^\vee, r_{\text{sat}}, f_{\text{sat}}\right)$ from a sieve algorithm,*

where E_{n,r_d} and V_{n,r_d} are defined in Definition 6.3 and $r_d = r_{\text{sat}}\text{GH}(n)$. In particular, the CDF of the score distribution is as follows:

$$\mathop{\mathbb{P}}_{\Lambda,\mathcal{W}}[f_{\mathcal{W}}(\mathbf{t}) \le x] = \frac{1}{2} + \frac{1}{2}\,\text{erf}\left(\frac{x - N \cdot E_{n,r_d}(\|\mathbf{t}\|)}{\sqrt{2N \cdot V_{n,r_d}(\|\mathbf{t}\|)}}\right). \tag{13}$$

Heuristic Justification. First, based on Heuristic 2.13, we assume, over the randomness of Λ and $\mathcal{W}$, that $\mathcal{W}$ is a set of i.i.d. $\mathbf{w} \leftarrow U(r_d\mathcal{B}^n)$ of size N. Then, Corollary 6.4 proves that each individual score $(f_{\mathbf{w}})_{\mathbf{w}\in\mathcal{W}}$ has mean $E_{n,r_d}(\|\mathbf{t}\|)$ and variance $V_{n,r_d}(\|\mathbf{t}\|)$.

Because the $(f_{\mathbf{w}}(\mathbf{t}))_{\mathbf{w}\in\mathcal{W}}$ are i.i.d. individual scores, we can use a central limit heuristic here. In this case, the total score distribution will have mean $N \cdot E_{n,r_d}(\|\mathbf{t}\|)$ and variance $N \cdot V_{n,r_d}(\|\mathbf{t}\|)$. □

Remark 6.6. The heuristic should not be used for targets $\mathbf{t}$ of norm above $\text{GH}(n)$. Given a random lattice, one expects there exists $\mathbf{t}' \in \mathbf{t} + \Lambda$ of norm $\|\mathbf{t}'\| \le \text{GH}(n)$ by the Gaussian Heuristic. Taking $\mathbf{t}' \in \mathbf{t} + \Lambda$ of minimal norm we have $f_{\mathcal{W}}(\mathbf{t}) = f_{\mathcal{W}}(\mathbf{t}')$, but using Heuristic 2.13 and Corollary 6.4 yields a *different* expected score of $E_{n,r_d}(\|\mathbf{t}'\|)$ for $\mathbf{t}'$, which is much higher than $E_{n,r_d}(\|\mathbf{t}\|) \approx 0$ if $\|\mathbf{t}'\| < 0.5\text{GH}(n)$, say. Hence, use of Heuristic 2.13, which forgets that $\mathcal{W}$ are dual vectors, results in contradictory predictions in the case $\|\mathbf{t}\| > \text{GH}(n)$.

On the other hand, given a target $\mathbf{t}$ of norm $\alpha\text{GH}(n)$ for a constant $\alpha < 1$, we expect $\|\mathbf{t}\| = d(\mathbf{t}, \Lambda)$ holds for most random lattices as $n \to \infty$ by Heuristic Claim 2.19, avoiding this issue with Heuristic 2.13.

Interestingly, the expected score $f_{\mathcal{W}}(\mathbf{t})$ is very similar to the score when one takes $\mathcal{W} = r_d\mathcal{B}^n \cap \Lambda^\vee$. Without any heuristics, Theorem 2.15 and Lemma 2.16 yield the following identity:

$$f_{\mathcal{W}}(\mathbf{t}) = \sum_{\mathbf{w}\in\Lambda^\vee} \mathbf{1}_{r_d\mathcal{B}^n}(\mathbf{w})\, e^{2\pi i\langle \mathbf{w},\mathbf{t}\rangle} = \text{Vol}_n\left(r_d\mathcal{B}^n\right) \cdot \sum_{\mathbf{v}\in\Lambda} \xi\left(\frac{n}{2}, r_d\,\|\mathbf{v}+\mathbf{t}\|\right). \tag{14}$$

For $r_p \ll \text{GH}(n)$, the vector $\mathbf{v} = \mathbf{0}$ gives the main contribution of $\xi\left(\frac{n}{2}, r_d r_p\right)$ to the summation in (14), because $\xi(n/2, -)$ is a rapidly decaying function.

At this point, one could argue that Heuristic 6.5 is not better at predicting a score distribution than the Independent Score Heuristic because it still uses a central limit heuristic. However, observe that Heuristic 6.5 takes a central limit heuristic for i.i.d. $f_{\mathbf{w}}(\mathbf{t})$, after first fixing the target $\mathbf{t}$ independent of the lattice.

On the other hand, usage of the Independent Score Heuristic leads to a prediction that uses a central limit heuristic for the individual scores $f_{\mathbf{w}}(\mathbf{t})$ *irrespective of the target distribution for* $\mathbf{t}$, which ignores the dependency of its norm on the mean score. Heuristic 6.5, being restricted to a fixed target norm, thus seems reasonable. However, large experiments are ultimately needed to gain confidence in the heuristic, and these can be found in Sect. 7.2.

Lastly, the following lemma contains an upper bound on the norm of a BDD target such that it is distinguishable from targets uniform modulo the lattice.

Lemma 6.7. *If $r_p \cdot r_d < \frac{n}{4\pi}$, then $E_{n,r_d}(r_p) > 0$.*

Proof. The first zero of the function $x \mapsto \xi(\frac{n}{2}, x)$ is at $x = \frac{1}{2\pi} j_{n/2,1} > \frac{n}{4\pi} > r_p r_d$, using Lemma 2.6. □

Using the approximation $\mathrm{GH}(n) \approx \sqrt{\frac{n}{2\pi e}}$, note the condition on $r_p r_d$ translates to roughly $r_p r_d < \frac{e}{2}\mathrm{GH}(n)^2$ for large n.

Note that above lemma is sharp, i.e., if we take $r_p r_d = \alpha n$ for some constant $\alpha > \frac{1}{4\pi}$, then $r_p r_d = \frac{n}{4\pi} + \left(\alpha - \frac{1}{4\pi}\right) n$. Lemma 2.6 implies there exists some $N \in \mathbb{N}$ such that for all $n \geq N$ we have $r_p r_d > \frac{1}{2\pi} j_{n/2,1}$, so there will be some $n \geq N$ giving a negative expectation score.

6.3. *Radial Error Distributions*

The predictions for a particular error can now be extended to any error distribution χ that is radial, i.e., the PDF is the same at two points of the same norm. In particular, given the radial distribution χ, let us write $f(r)$ to denote the probability (density) of a sample $\mathbf{e} \leftarrow \chi$ having norm r, i.e.,

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

We refer to $f(r)$ as the *norm distribution of* χ.

Using Heuristic 6.5 we arrive at the following claim.

Heuristic Claim 6.8. *Let χ be an error distribution with norm distribution $f(r)$, such that the support of χ is mostly contained in* $\mathrm{GH}(n)\,\mathcal{B}^n$. *The score distribution $f_{\mathcal{W}}(\mathbf{t})$ for an error vector $\mathbf{t} \leftarrow \chi$ has the following CDF:*

$$\mathop{\mathbb{P}}_{\Lambda, \mathbf{t} \leftarrow \chi} [f_{\mathcal{W}}(\mathbf{t}) \leq x] = \frac{1}{2} + \frac{1}{2} \int_0^\infty \mathrm{erf}\left(\frac{x - N \cdot E_{n,r_d}(r)}{\sqrt{2N \cdot V_{n,r_d}(r)}} \right) f(r) \mathrm{d}r, \tag{15}$$

where the probability is taken over randomness from sampling Λ and obtaining N dual vectors $\mathcal{W} \leftarrow \mathrm{Sieve}\left(\Lambda^\vee, r_{\mathrm{sat}}, f_{\mathrm{sat}}\right)$ from a sieve algorithm.

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

Gaussian Error. The norm distribution of samples from $\mathcal{N}(0, 1)^n$, is the χ-distribution of order n,[7] which has a PDF given by:

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

[7] Perhaps more well known is the PDF of the square norm, given by a χ^2-distribution.

Now instantiating Heuristic Claim 6.8 for $\mathcal{N}(0,\sigma^2)^n$, the Gaussian error distribution of parameter $\sigma > 0$, yields the following claim.

Heuristic Claim 6.9. *Let $\sigma \in (0, \mathrm{GH}(n)/\sqrt{n})$, and let $f(x;n)$ be the PDF of the χ-distribution. The score distribution $f_{\mathcal{W}}(\mathbf{t})$ for an error vector $\mathbf{t} \leftarrow \mathcal{N}(0,\sigma^2)^n$ has the following CDF:*

$$\mathop{\mathbb{P}}_{\substack{\Lambda,\mathcal{W},\\ \mathbf{t}\leftarrow\mathcal{N}(0,\sigma)^n}}[f_{\mathcal{W}}(\boldsymbol{t}) \le x] = \frac{1}{2} + \frac{1}{2}\int_0^{\infty} \mathrm{erf}\left(\frac{x - N\cdot E_{n,r_d}(r)}{\sqrt{2N\cdot V_{n,r_d}(r)}}\right)\cdot f\left(\frac{r}{\sigma};n\right)\frac{\mathrm{d}r}{\sigma}, \quad (16)$$

where the probability is taken over randomness from sampling Λ and obtaining N dual vectors $\mathcal{W} \leftarrow \mathrm{Sieve}(\Lambda^\vee, r_{\mathrm{sat}}, f_{\mathrm{sat}})$ from a sieve algorithm.

Note that the χ-distribution is very 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 Sect. 7.1).

Error Uniform from a Ball. Consider the distribution $U(r_p\mathcal{B}^n)$ for some $r_p > 0$. In this case,

$$\mathop{\mathbb{P}}_{\mathbf{t}\leftarrow U(r_p\mathcal{B}^n)}[\|\mathbf{t}\| \le r] = r^n/r_p^n$$

for any $r \in [0, r_p]$, while this equals 1 for $r \ge r_p$. Strictly speaking, the CDF of the norm distribution is not differentiable at $r = r_p$, so $f(r_p)$ is not defined. Still, one can mitigate this issue by integrating from 0 to r_p in (15), yielding the following claim.

Heuristic Claim 6.10. *Let $r_p \in (0, \mathrm{GH}(n))$. The score distribution $f_{\mathcal{W}}(\mathbf{t})$ for an error vector $\mathbf{t} \leftarrow U(r_p\mathcal{B}^n)$ has the following CDF:*

$$\mathop{\mathbb{P}}_{\substack{\Lambda,\mathcal{W},\\ \mathbf{t}\leftarrow U(r_p\mathcal{B}^n)}}[f_{\mathcal{W}}(\mathbf{t}) \le x] = \frac{1}{2} + \frac{1}{2}\int_0^{r_p} \mathrm{erf}\left(\frac{x - N\cdot E_{n,r_d}(r)}{\sqrt{2N\cdot V_{n,r_d}(r)}}\right)\cdot \frac{nr^{n-1}\mathrm{d}r}{r_p^n}, \quad (17)$$

where the probability is taken over randomness from sampling Λ and obtaining N dual vectors $\mathcal{W} \leftarrow \mathrm{Sieve}(\Lambda^\vee, r_{\mathrm{sat}}, f_{\mathrm{sat}})$ from a sieve algorithm.

6.4. *Error Uniform Modulo Lattice*

Given uniform targets $\mathbf{t} \leftarrow U(\mathbb{R}^n/\Lambda)$, the inner product $\langle \mathbf{w}, \mathbf{t}\rangle \pmod 1$ follows the uniform distribution on $\mathbb{R}/\mathbb{Z}$, for any (nonzero) dual vector $\mathbf{w} \in \Lambda^\vee$. Experiments in Sect. 5, however, show that a central limit heuristic cannot be used on the individual scores $(f_{\mathbf{w}}(\mathbf{t}))_{\mathbf{w}\in\mathcal{W}}$, because the score distribution for uniform targets does not decay as fast as a Gaussian. Indeed, Sect. 4.2 shows the probability on an exceptionally high score is underestimated with a central limit heuristic, because it does not consider the probability of a uniform target landing close to Λ, to which a distinguisher assigns a high score.

In this section, we want to derive predictions for uniform targets, without using the Independent Score Heuristic. The above motivation shows that having a high score is driven by a target close to the lattice, so it seems important to know how the distance $d(\mathbf{t}, \Lambda)$ from a uniform target to a lattice is distributed to predict the score distribution.

First, because the Voronoi cell tiles the space, assume targets are sampled from $U(\mathcal{V}(\Lambda))$. The score for such targets would ideally be predicted by using Heuristic Claim 6.8, but $U(\mathcal{V}(\Lambda))$ is only a radial distribution in dimension $n = 1$. Instead, because the dual vectors already have rotational invariance when using Heuristic 2.13, we could approximate the score distribution by still using Heuristic Claim 6.8 with norm distribution $f(r) = \mathrm{d}F(r)/\mathrm{d}r$, where

$$F(r) = \mathop{\mathbb{P}}_{\mathbf{t}\leftarrow U(\mathbb{R}^n/\Lambda)}[d(\mathbf{t}, \Lambda) \leq r] = \mathrm{Vol}_n\left(\mathcal{V}(\Lambda) \cap r\mathcal{B}^n\right). \tag{18}$$

However, already computing the Voronoi cell is considered a hard task [36]. Note that we have $\frac{1}{2}\lambda_1\mathcal{B}^n \subseteq \mathcal{V}(\Lambda) \subseteq \mu(\Lambda)\mathcal{B}^n$, where $\mu(\Lambda)$ is the covering radius of the lattice. Hence, $F(r) = \mathrm{Vol}_n(r\mathcal{B}^n)$ for $r \leq \frac{1}{2}\lambda_1$. 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,

$$F(r) \leq \mathrm{Vol}_n\left(r\mathcal{B}^n\right), \tag{19}$$

which is not much smaller than what is derived in Heuristic Claim 4.4 for $r = (1 - \varepsilon)\mathrm{GH}(n)$ with $\varepsilon > 0$. The upper bound in Eq. (19) can thus be seen as a first order approximation of $F(r)$. The following heuristic implies that $F(r)$ equals the upper bound in Eq. (19).

Heuristic 6.11. *Let $\Lambda \subset \mathbb{R}^n$ be a random full-rank unit-volume lattice. Then,*

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

Using Heuristic 6.11 and Heuristic Claim 6.10 directly results in the following claim.

Heuristic Claim 6.12. *The score distribution $f_{\mathcal{W}}(\mathbf{t})$ for an error vector $\mathbf{t} \leftarrow U(\mathbb{R}^n/\Lambda)$ has the following CDF:*

$$\mathop{\mathbb{P}}_{\substack{\Lambda,\mathcal{W},\\ \mathbf{t}\leftarrow U(\mathbb{R}^n/\Lambda)}}[f_{\mathcal{W}}(\mathbf{t}) \leq x] = \frac{1}{2} + \frac{1}{2}\int_0^{\mathrm{GH}(n)} \mathrm{erf}\left(\frac{x - N \cdot E_{n,r_d}(r)}{\sqrt{2N \cdot V_{n,r_d}(r)}}\right) \cdot \frac{nr^{n-1}\mathrm{d}r}{\mathrm{GH}(n)^n}, \tag{20}$$

where the probability is taken over randomness from sampling Λ and obtaining N dual vectors $\mathcal{W} \leftarrow \mathrm{Sieve}\left(\Lambda^\vee, r_{\mathrm{sat}}, f_{\mathrm{sat}}\right)$ from a sieve algorithm.

This heuristic claim predicts the waterfall and floor shapes in the score distribution for uniform targets:

– For x close to zero, the integrand is negligible for r much smaller than $\mathrm{GH}(n)$ since those expected scores are much higher than x. For $r \approx \mathrm{GH}(n)$ the expected score is approximately zero and the variance $N \cdot V_{n,r_d}(r) \approx N/2$. Given some $\varepsilon > 0$, by using the approximation $N \cdot E_{n,r_d}(r) \approx 0$ and $N \cdot V_{n,r_d}(r) \approx N/2$ for $r > (1-\varepsilon)\mathrm{GH}(n)$, one can see the integral in (20) is approximated by $\mathrm{erf}(x/\sqrt{N})\left(1 - e^{-n/\varepsilon}\right) \to \mathrm{erf}(x/\sqrt{N})$ (as $n \to \infty$). Hence, this CDF has the waterfall shape around $x = 0$.
– For the floor phenomenon, let us consider the survival function of $f_{\mathcal{W}}(\mathbf{t})$:

$$\mathop{\mathbb{P}}_{\substack{\Lambda,\mathcal{W},\\ \mathbf{t}\leftarrow U(\mathbb{R}^n/\Lambda)}}[f_{\mathcal{W}}(\mathbf{t}) > x] = \frac{1}{2}\int_0^{\mathrm{GH}(n)} \mathrm{erfc}\left(\frac{x - N \cdot E_{n,r_d}(r)}{\sqrt{2N \cdot V_{n,r_d}(r)}}\right) \cdot \frac{nr^{n-1}\mathrm{d}r}{\mathrm{GH}(n)^n}.$$

For large x, say $c\sqrt{N}$ for some $c > 10$, the biggest contribution to the integral comes from $r \approx r_0$ where r_0 is a solution to $N \cdot E_{n,r_d}(r_0) = x$, because here erfc evaluates to ≈ 1. However, for $r \gg r_0$, the integrand drops to zero as erfc decreases much quicker than r^{n-1}, while for $r \ll r_0$, the integrand also drops to zero because r^{n-1} is simply too small.

7. Validating the New Model

In this section, we provide further substantiation of the concrete predictions made in Sect. 6, in particular the predictions in (13), (16), (17) and (20), with experimental support. With the experiments, we want to verify whether the used heuristics lead to conclusions that precisely match practice.

Similarly to Sect. 5, we take $\mathcal{W}$ to be the full output of a lattice sieve on $\Lambda^\vee$ in the experiments. We will compare three possible BDD distributions with their respective prediction from Sect. 6: uniform from a sphere, uniform from a ball and Gaussian. In addition, we compare the score distribution for uniform targets with the predicted CDF of Eq. (20), which was derived in Sect. 6.4.

7.1. *Implementation Details.*

We used the `G6K` software [5] for running the experiments, using `Python` on a high-level, but with a binding to some `C/C++` code for speeding up BDD sampling.

BDD targets.

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 many dual vectors, and finally computing scores for many samples. To make the implementation easier, the sampled q-ary lattice is used as the dual lattice $\Lambda^\vee$, because we do not work with primal lattice vectors. The prime used is $q = 3329$.

For the experiments, we sample BDD targets from a distribution with norms concentrated around $f_{\mathrm{GH}} \cdot \mathrm{GH}(n)/\sqrt{q}$ for some parameter $f_{\mathrm{GH}} \in (0, 1]$, where Λ is the primal

lattice. Note that the Gaussian Heuristic predicts $\lambda_1(\Lambda) \approx \mathrm{GH}(n)/\sqrt{q}$ holds, because the dual lattice has determinant $q^{n/2}$.

More specifically, for a Gaussian sample, we simply sample $\mathbf{x} \leftarrow \mathcal{N}(0, \sigma^2)^n$ with $\sigma = \frac{f_{\mathrm{GH}} \cdot \mathrm{GH}(n)}{\sqrt{n \cdot q}}$. Note that we need a factor of $1/\sqrt{n}$ such that $\|\mathbf{x}\|$ is concentrated around the target length $f_{\mathrm{GH}} \cdot \mathrm{GH}(n)/\sqrt{q}$. For a sample $\mathbf{y}$ uniform on the sphere, we reuse the same Gaussian sample $\mathbf{x}$, and compute

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

which is then uniformly distributed on the sphere of radius $f_{\mathrm{GH}} \cdot \mathrm{GH}(n)/\sqrt{q}$. For a sample uniform in the ball, we take a sample uniformly from the $(n+1)$-dimensional sphere in $\mathbb{R}^{n+2}$ and drop the last two coordinates, making use of [9, Corollary 4]. 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.

The script `bdd_predict.py` then uses these data to plot the experimental data, the prediction from Sect. 6 and the heuristic prediction. The predictions from Sect. 6 required evaluating the Bessel function and an integration, for which we used `hyp0f1` and `quad`, respectively, both in the Python package `mpmath`. Specifically, the Gaussian prediction was numerically more accurate when the interval $(0, \infty)$ was split in two, with the split happening at the expected length.[8] The integration for the uniform ball was performed from $0.001r$ to r, with r the radius of the ball, to prevent the singularity around 0.

Uniform targets.

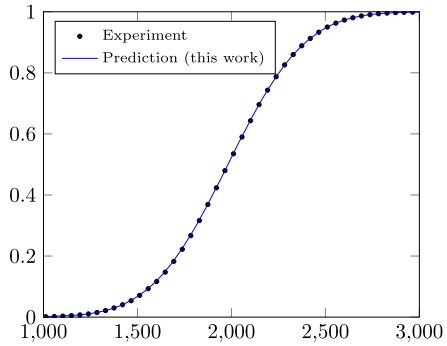
For targets uniform in the lattice, the scripts `unif_sample.py` and `unif_predict.py` were used similarly to the BDD case. One should pay special attention to the integration errors caused by numerical integration when calculating the predictions of Heuristic Claim 6.12. Namely, the numerical integration should give an answer with a small *relative error*, because the output can be very small, as can be seen in Fig. 8. Thus, for numerical integration we used the function `quad` from the Python package `SciPy` with parameters `epsabs=0` and `epsrel=0.001`.

7.2. *Experiments for BDD Targets*

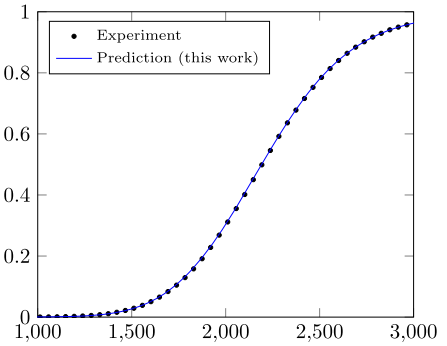
The experiments with the score for BDD targets can be found in Figs. 6 and 7. The former samples BDD targets at expected distance of $0.7\mathrm{GH}(n)$ from the lattice, while the latter samples at distance $0.5\mathrm{GH}(n)$ from the lattice.

Here, a saturation ratio of $f_{\mathrm{sat}} = 0.99$ and saturation radius of $r_{\mathrm{sat}} = \sqrt{4/3}$ was used, so $N = \lceil \frac{1}{2} f_{\mathrm{sat}} \cdot {r_{\mathrm{sat}}}^{90} \rceil = 207419$ dual vectors were used. It is clear that the predictions made in Sects. 6.2 and 6.3 give accurate estimates on the score distribution for BDD targets that are from a sphere, from a ball or Gaussian. Moreover, these experiments show that Heuristic 6.5 appears to be a reasonable heuristic to use in this regime.

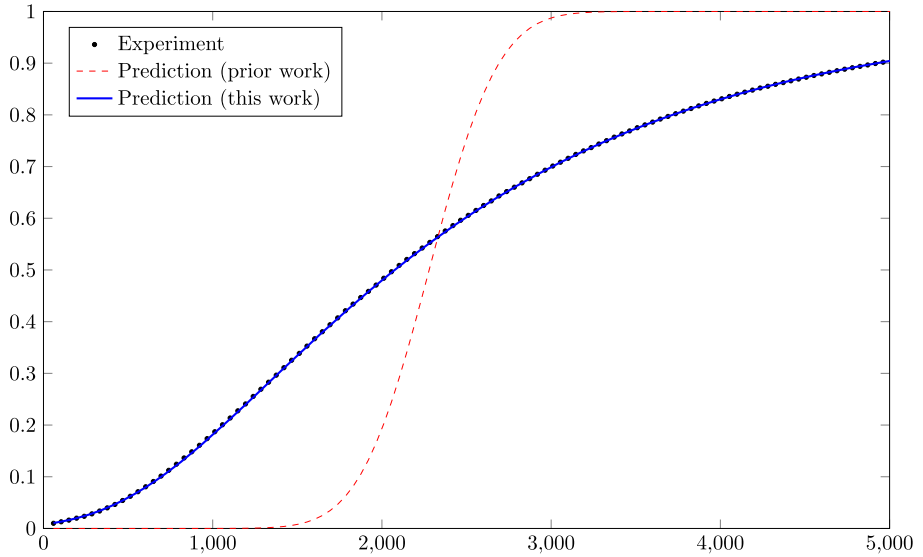
[8] For more details, see "Highly variable functions" from the documentation (https://mpmath.org/doc/current/calculus/integration.html)



(a) CDF of score for errors uniform from a sphere of radius $0.7\mathrm{GH}(n)$.



(b) CDF of score for errors uniform from a ball of radius $0.7\mathrm{GH}(n)$.

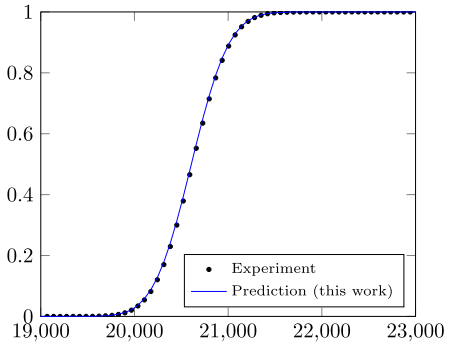


(c) CDF of score for gaussian errors with $\sigma = 0.7\mathrm{GH}(n)/\sqrt{n}$.

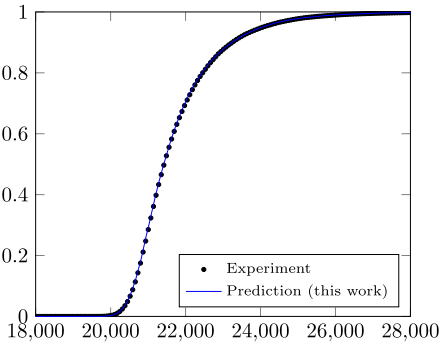
Fig. 6. CDF of the score for three BDD target distributions in dimension $n = 90$. The prior prediction uses the Independent Score Heuristic. "Prediction (this work)" is based on Sects. 6.2 and 6.3. Each experimental curve contains 10^5 samples, which are obtained with `code/bdd_sample.py` and are listed in `data/predictions_n90_ghf0.7.csv` of the auxiliary files.

7.3. *Experiments for Uniform Targets*

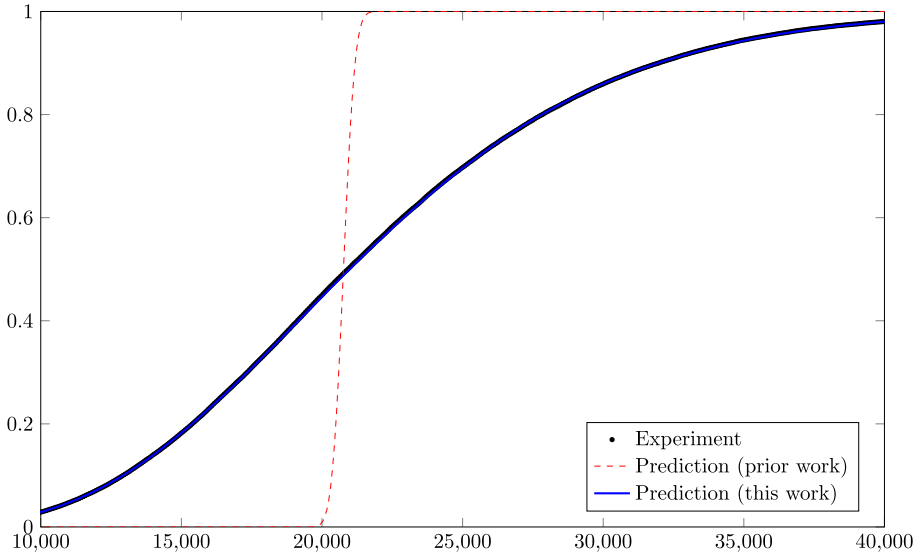
Figure 8 compares the prediction of the score distribution for uniform targets, versus experiments. Here, the lattice sieve used a saturation ratio of $f_{\mathrm{sat}} = 0.9$ and saturation radius of $r_{\mathrm{sat}} = \sqrt{4/3} \approx 1.1547$. The experimental data for uniform targets are acquired independently from that of Sect. 5.2, which used $2N$ dual vectors. For each of the four experimental curves of Fig. 8, we ran the script `code/unif_sample.py`, which took slightly more than one day on a server with 40 physical cores.



(a) CDF of score for errors uniform from a sphere of radius $0.5\mathrm{GH}(n)$.



(b) CDF of score for errors uniform from a ball of radius $0.5\mathrm{GH}(n)$.



(c) CDF of score for gaussian errors with $\sigma = 0.5\mathrm{GH}(n)/\sqrt{n}$.

Fig. 7. CDF of the score for three BDD target distributions in dimension $n = 90$. The prior prediction uses the Independent Score Heuristic. "Prediction (this work)" is based on Sects. 6.2 and 6.3. Each experimental curve contains 10^5 samples, which are obtained with `code/bdd_sample.py` and are listed in `data/predictions_n90_ghf0.5.csv` of the auxiliary files.

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 few 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, but then the experiment would require a runtime of multiple days on our server.

In dimension 40, it seems that the experiments give scores higher than expected by our prediction. However, the sieve provides only $N = \lceil \frac{1}{2} f_{\mathrm{sat}} \cdot {r_{\mathrm{sat}}}^{40} \rceil = 142$ dual vectors in this dimension, which may be too small for the heuristic prediction to be accurate,

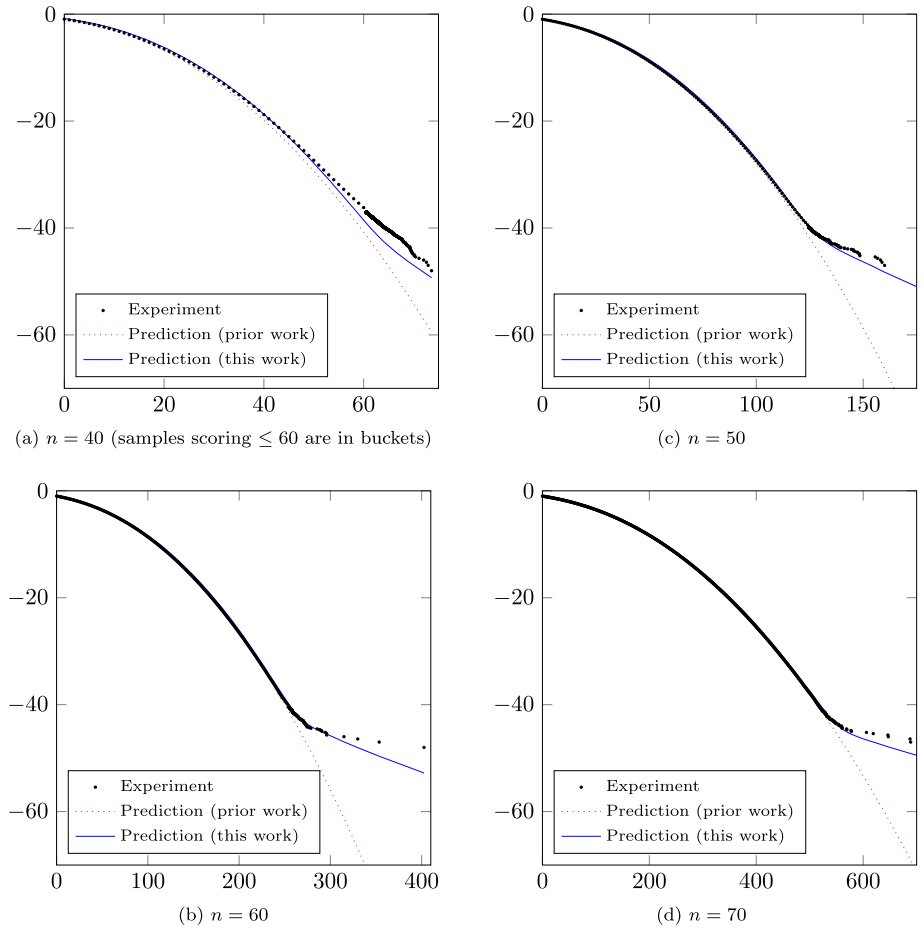
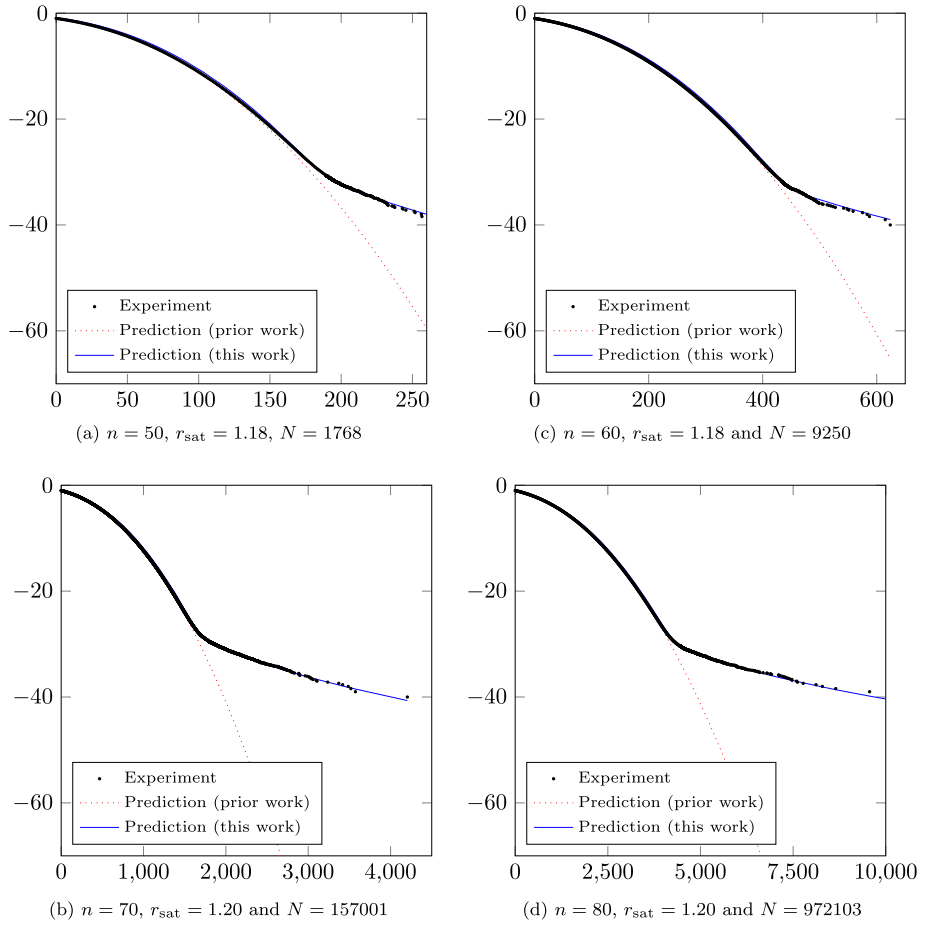


Fig. 8. Binary logarithm ($\log_2$) of the survival function (SF) of the score distribution for uniform targets in various dimensions n. Each experimental curve contains 2^{48} samples, and ran a sieve with saturation ratio $f_{\text{sat}} = 0.9$ and saturation radius $r_{\text{sat}} = \sqrt{4/3}$.

as a central limit heuristic might not apply with few dual vectors, but also the lengths of the dual vectors are small. Therefore, we believe that the predictions become more accurate when there are more dual vectors, which happens, e.g., in higher dimensions, or when sieving with a larger saturation radius.

To test the robustness of our prediction for the floor phenomenon, we also ran experiments with a lattice sieve using a different saturation radius r_{sat}, which produces more (and therefore a bit longer) dual vectors than using the standard saturation radius of $\sqrt{4/3} \approx 1.1547$. In these scenarios, the floor phenomenon is observable already at a larger failure 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 larger than $\sqrt{4/3}$. This motivates Fig. 9. This figure shows the predictions



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

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

(b) $n = 70$, $r_{\text{sat}} = 1.20$ and $N = 157001$

(d) $n = 80$, $r_{\text{sat}} = 1.20$ and $N = 972103$

Fig. 9. Binary logarithm ($\log_2$) of the survival function (SF) of the score distribution for uniform targets in various dimensions n, with a different *saturation radius*. Each experimental curve contains 2^{40} samples, and ran a sieve with saturation ratio $f_{\text{sat}} = 0.9$.

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.

8. Conclusion

As shown in the experiments, the Independent Score Heuristic seems to mispredict the cost of the dual attack. On the other hand, we provide a new heuristic model giving accurate predictions in experiments to be used in an analysis of the dual attack. We conclude that prior claims based on the Independent Score Heuristic are invalidated by the experiments, while proposing some improvements that could be added to the dual

attack, and mentioning some pitfalls to look out for if one wants to fix the analysis of the dual attack.

8.1. *The Independent Score Heuristic*

The theoretical analysis in Sect. 4 and the experiments in Sect. 5 unequivocally invalidate the analyses of the Dual-Sieve attack (with or without FFT) as found in [6,14,27,31,33], because these use the Independent Score Heuristic in their analysis. In the context of the Dual-Sieve-FFT attack, as instantiated in [6,14,27,33], the chosen parameters result in a presumably large number of false positives, i.e., incorrect guesses for the part of the secret having a higher score than that of the correct guess.

8.2. *The New Model*

This paper proposes heuristics that can be used in the analysis of dual attacks against the decision-BDD problem. Specifically, Heuristic Claim 6.9 and Heuristic Claim 6.12 experimentally show accurate predictions regarding both BDD and uniform targets, compared to experimental data.

To analyze the probability of guessing correctly in a Dual-Sieve attack, a single random lattice Λ and set of dual vectors $\mathcal{W}$ is used, so one cannot directly apply the two heuristic claims, because the CDFs take the randomness over Λ and $\mathcal{W}$. Further investigation is needed regarding the independence of the score functions between two targets (BDD or uniform) when the same Λ and $\mathcal{W}$ is used for both. Until then, the heuristic claims can already be used with a union bound. Because the probabilities are extremely small for the uniform targets, one only loses a small constant compared to when having perfect independence.

8.3. *Possible Pitfalls*

Note that pulling the parameters of an attack outside the contradictory regime in Fig. 3 is not a convincing way to prevent mispredictions while using the Independent Score Heuristic. Indeed, the contradictory regime shows a fundamental flaw with this heuristic, and experiments indicate that there may well be an impact outside the regime. That is, a sound model needs to be used with theoretically justified predictions matching experimental behavior measured in Figs. 4 and 5 and beyond.

Moreover, we warn against a flawed argument for a fix, that would consist of constructing the set of dual vectors $\mathcal{W} = \mathcal{W}_1 \cup \mathcal{W}_2$ from two BKZ reductions and sieves instead of just one, yielding the score function $f_{\mathcal{W}}(\mathbf{t}) = f_{\mathcal{W}_1}(\mathbf{t}) + f_{\mathcal{W}_2}(\mathbf{t})$. One might argue that the dual distinguishing is now happening in a lattice of dimension 2β instead of β, possibly pushing the attack far outside the contradictory regime of Fig. 3. However, considering the experiments in Fig. 4 one can see this argument is invalid. Each score distribution $f_{\mathcal{W}_i}$ is going to hit its floor, and thus $f_{\mathcal{W}}$ will also have a floor starting from essentially the same score. Indeed, it suffices to hit the floor of *at least one* of the functions to hit the floor of the aggregate.

A more credible approach, however, is indeed the following. Run two (or a few) BKZ reduction and sieves, obtain two sets of short dual vectors $\mathcal{W}_1, \mathcal{W}_2$ and then consider the

aggregate score as the *minimum* of both scores $f' = \min(f_{\mathcal{W}_1}, f_{\mathcal{W}_2})$ rather than its sum. Now to hit the floor of this new aggregate function f', a uniform target would need to hit *both* floors simultaneously. If $f_{\mathcal{W}_1}, f_{\mathcal{W}_2}$ are sufficiently independent (an assumption that would need substantiation), such event is much rarer than hitting either floor.

However, note that taking such a minimum aggregate of scores might also amplify the issues with low scores for BDD targets observed in Sect. 5.3. A robust model for all the distributions at hand is therefore still required.

Note this list of pitfalls is not comprehensive, and thus any potential fix of an attack, needs to be motivated at least by theoretical arguments or experimental validation.

8.4. *Possible Improvements & Future Work*

The experiments on the Independent Score Heuristic show that current parameterizations of the attacks in [6,14,27,33] will result in many false positives. However, if this number of false positives is still reasonable, one might mitigate the issue at hand, although the attack cost will be somewhat higher. Indeed, one may consider the Dual-Sieve-FFT technique as a first "filtering stage" in an attack with multiple stages: the remaining problem is still the problem of finding one BDD target among many candidates, but in an easier lattice (smaller in dimension, and/or sparser). If this remaining problem is sufficiently easier to accommodate all these targets, the Dual-Sieve-FFT attack might be salvaged, although the cost of such an attack will be higher than what was listed in [6,14,27,33]. Also note that first filtering and then filtering the "survivors" again, is conceptually not that far off from the idea of taking the smallest of both scores.

Alternatively, 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 [2,31,40].

The effectiveness of the state-of-the-art dual attack remains as future work. In particular, note that existing dual attacks, e.g., [27], will most likely require a different parametrization to achieve the best runtime, while having a constant success probability, when the analysis will use the new model of Sect. 6. Moreover, there will be other aspects to the costing of the attack that may still require some attention, as highlighted in [18, App. A].

On another note, theoretically it would be interesting to predict the scores for uniform targets from Sect. 6.4: instead of relying on a ball approximation of the Voronoi cell, it could be more satisfactory to adapt the Poisson model of [34, Assumption 8] to the lattice setting. 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 setting has a Poisson process for each weight $w = 0, 1, \ldots, n$, the situation is a bit more complex for lattices, as there is a continuum of processes to consider.

Acknowledgements

We would like to thank Kevin Carrier, Thomas Debris-Alazard, Charles Meyer-Hilfiger, Amaury Pouly, Yixin Shen and Jean-Pierre Tillich for insightful conversations, and for sharing their early drafts [13,40]. We would also like to thank Martin Albrecht,

Qian Guo, Thomas Johansson, Eamonn Postlethwaite, Michael Walter and Wessel van Woerden for helpful discussion and feedback. Lastly, we thank the anonymous Journal of Cryptology reviewers for their valuable feedback. Authors Léo Ducas and Ludo N. Pulles are supported by ERC Starting Grant 947821 (ARTICULATE).

References

[1] M. Abramowitz, I.A. Stegun, Handbook of mathematical functions with formulas, graphs, and mathematical tables, in *National Bureau of Standards Applied Mathematics Series*, vol. 55. (U.S. Government Printing Office, 1964). https://archive.org/details/AandS-mono600

[2] D. Aharonov, O. Regev, Lattice problems in NP cap coNP, in: *45th FOCS*. (IEEE Computer Society Press 2004), pp. 362–371.https://doi.org/10.1109/FOCS.2004.35

[3] A. Al Jabri, A statistical decoding algorithm for general linear block codes, in B. Honary (ed.), *8th IMA International Conference on Cryptography and Coding*. LNCS, vol. 2260, (Springer, Berlin, Heidelberg, 2001), pp. 1–8. https://doi.org/10.1007/3-540-45325-3_1

[4] M.R. Albrecht, On dual lattice attacks against small-secret LWE and parameter choices in HElib and SEAL, in J.S. Coron, J.B. Nielsen (eds.), *EUROCRYPT 2017, Part II*. LNCS, vol. 10211, (Springer, Cham, 2017), pp. 103–129. https://doi.org/10.1007/978-3-319-56614-6_4

[5] M.R. Albrecht, L. Ducas, G. Herold, E. Kirshanova, E.W. Postlethwaite, M. Stevens, The general sieve kernel and new records in lattice reduction, in Y. Ishai, V. Rijmen (eds.), *EUROCRYPT 2019, Part II*. LNCS, vol. 11477, (Springer, Cham, 2019), pp. 717–746. https://doi.org/10.1007/978-3-030-17656-3_25

[6] M.R. Albrecht, Y. Shen, Quantum augmented dual attack (2022). arxiv:2205.13983

[7] E. Alkim, L. Ducas, T. Pöppelmann, P. Schwabe, Post-quantum key exchange - a new hope, in T. Holz, S. Savage (eds.), *USENIX Security 2016*, (USENIX Association, 2016). pp. 327–343

[8] S. Bai, T. Laarhoven, D. Stehlé, Tuple lattice sieving. *LMS J. Comput. Math.***19**(A), 146–162 (2016). https://doi.org/10.1112/S1461157016000292

[9] F. Barthe, O. Guédon, S. Mendelson, A. Naor, A probabilistic approach to the geometry of the ℓ_p^n-ball. *Ann. Probab.***33**(2), 480–513 (2005). https://doi.org/10.1214/009117904000000874

[10] A. Becker, L. Ducas, N. Gama, T. Laarhoven, New directions in nearest neighbor searching with applications to lattice sieving, in R. Krauthgamer (ed.), *27th SODA*, (ACM-SIAM, 2016), pp. 10–24. https://doi.org/10.1137/1.9781611974331.ch2

[11] L. Both, A. May, Decoding linear codes with high error rate and its impact for LPN security, in T. Lange, R. Steinwandt (eds.), *Post-Quantum Cryptography - 9th International Conference*, PQCrypto 2018, (Springer, Cham, 2018), pp. 25–46. https://doi.org/10.1007/978-3-319-79063-3_2

[12] K. Carrier, T. Debris-Alazard, C. Meyer-Hilfiger, J.P. Tillich, Statistical decoding 2.0: reducing decoding to LPN, in S. Agrawal, D. Lin (eds.), *ASIACRYPT 2022, Part IV*. LNCS, vol. 13794, (Springer, Cham, 2022), pp. 477–507. https://doi.org/10.1007/978-3-031-22972-5_17

[13] K. Carrier, T. Debris-Alazard, C. Meyer-Hilfiger, J.P. Tillich, Reduction from sparse LPN to LPN, dual attack 3.0, in M. Joye, G. Leander (eds.), *EUROCRYPT 2024, Part VII*. LNCS, vol. 14657, (Springer, Cham, 2024), pp. 286–315. https://doi.org/10.1007/978-3-031-58754-2_11

[14] K. Carrier, Y. Shen, J.P. Tillich, Faster dual lattice attacks by using coding theory. *Cryptology ePrint Archive*, Report 2022/1750 (2022). https://eprint.iacr.org/archive/2022/1750/20221220:184709

[15] Y. Chen, P.Q. Nguyen, BKZ 2.0: better lattice security estimates, in D.H. Lee, X. Wang (eds.), *ASIACRYPT 2011*. LNCS, vol. 7073, (Springer, Berlin, Heidelberg, 2011), pp. 1–20. https://doi.org/10.1007/978-3-642-25385-0_1

[16] T. Debris-Alazard, L. Ducas, N. Resch, J.P. Tillich, Smoothing codes and lattices: systematic study and new bounds. *IEEE Trans Inf Theory* **69**(9), 6006–6027 (2023).https://doi.org/10.1109/TIT.2023.3276921

[17] L. Ducas, Shortest vector from lattice sieving: a few dimensions for free, in J.B. Nielsen, V. Rijmen (eds.), *EUROCRYPT 2018, Part I*, LNCS, vol. 10820, (Springer, Cham, 2018), pp. 125–145. https://doi.org/10.1007/978-3-319-78381-9_5

[18] L. Ducas, L. Pulles, Does the dual-sieve attack on learning with errors even work? Cryptology ePrint Archive, Report 2023/302 (2023). https://eprint.iacr.org/2023/302

[19] L. Ducas, L.N. Pulles, Accurate score prediction for dual-sieve attacks. *Cryptology ePrint Archive*, Report 2023/1850 (2023). https://eprint.iacr.org/2023/1850

[20] L. Ducas, L.N. Pulles, Does the dual-sieve attack on learning with errors even work?, in H. Handschuh, A. Lysyanskaya (eds.), *CRYPTO 2023, Part III*. LNCS, vol. 14083, (Springer, Cham, 2023), pp. 37–69. https://doi.org/10.1007/978-3-031-38548-3_2

[21] L. Ducas, M. Stevens, W.P.J. van Woerden, Advanced lattice sieving on GPUs, with tensor cores, in A. Canteaut, F.X. Standaert (eds.), *EUROCRYPT 2021, Part II*. LNCS, vol. 12697, (Springer, Cham, 2021), pp. 249–279. https://doi.org/10.1007/978-3-030-77886-6_9

[22] P. Duhamel, M. Vetterli, Fast Fourier transforms: a tutorial review and a state of the art. *Signal proc.***19**(4), 259–299 (1990). https://infoscience.epfl.ch/record/59946

[23] T. Espitau, A. Joux, N. 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 K. Bhargavan, E. Oswald, M. Prabhakaran (eds.), *INDOCRYPT 2020*. LNCS, vol. 12578, (Springer, Cham, 2020), pp. 440–462. https://doi.org/10.1007/978-3-030-65277-7_20

[24] D. Fischer, Fourier transform of the indicator of the unit ball. *Math. Stack Exch.* (2013), https://math.stackexchange.com/a/492055 (version: 2013-09-12)

[25] R. Gallager, Low-density parity-check codes. *IRE Trans. Inf. Theory* **8**(1), 21–28 (1962). https://doi.org/10.1109/TIT.1962.1057683

[26] I.M. Gel'fand, G.E. Shilov, *Generalized functions. Volume I: Properties and Operations*, (Academic Press, New York and London, 1964). translated by Eugene Saletan

[27] Q. Guo, T. Johansson, Faster dual lattice attacks for solving LWE with applications to CRYSTALS, in M. Tibouchi, H. Wang (eds.), *ASIACRYPT 2021, Part IV*. LNCS, vol. 13093, (Springer, Cham, 2021), pp. 33–62. https://doi.org/10.1007/978-3-030-92068-5_2

[28] HåJ. stad, Dual vectors and lower bounds for the nearest lattice point problem. *Combinatorica* **8**(1), 75–81 (1988). https://doi.org/10.1007/BF02122554

[29] N. Howgrave-Graham, A hybrid lattice-reduction and meet-in-the-middle attack against NTRU, in A. Menezes (ed.), *CRYPTO 2007*. LNCS, vol. 4622, (Springer, Berlin, Heidelberg, 2007). pp. 150–169, https://doi.org/10.1007/978-3-540-74143-5_9

[30] T. Laarhoven, Approximate Voronoi cells for lattices, revisited. *J. Math. Cryptol.***15**(1), 60–71 (2021). https://doi.org/10.1515/jmc-2020-0074

[31] T. Laarhoven, M. Walter, Dual lattice attacks for closest vector problems (with preprocessing), in K.G. Paterson (ed.), *CT-RSA 2021*. LNCS, vol. 12704, (Springer, Cham, 2021), pp. 478–502. https://doi.org/10.1007/978-3-030-75539-3_20

[32] É. Levieil, P.A. Fouque, An improved LPN algorithm, in R. De Prisco, M. Yung (eds.), *SCN 06*. LNCS, vol. 4116, (Springer, Berlin, Heidelberg, 2006), pp. 348–359. https://doi.org/10.1007/11832072_24

[33] MATZOV: Report on the security of LWE: improved dual lattice attack (2022). https://doi.org/10.5281/zenodo.6493704

[34] C. Meyer-Hilfiger, J.P. Tillich, Rigorous foundations for dual attacks in coding theory, in G.N. Rothblum, Wee, H. (eds.) TCC 2023, Part IV. LNCS, vol. 14372, (Springer, Cham, 2023), pp. 3–32. https://doi.org/10.1007/978-3-031-48624-1_1

[35] D. Micciancio, Lattice algorithms and applications, lecture 3: the dual lattice (2014). https://cseweb.ucsd.edu/classes/sp14/cse206A-a/lec3.pdf, available at https://cseweb.ucsd.edu/classes/sp14/cse206A-a/lec3.pdf

[36] D. Micciancio, P. Voulgaris, Faster exponential time algorithms for the shortest vector problem, in M. Charika (ed.), 21st SODA. pp. 1468–1480. (ACM-SIAM, 2010). https://doi.org/10.1137/1.9781611973075.119

[37] J. Neyman, E.S. Pearson, On the problem of the most efficient tests of statistical hypotheses. Philosophical Transactions of the Royal Society of London. *Ser. A Contain. Papers Math. Phys. Char.***231**(694-706), 289–337 (1933). https://doi.org/10.1098/rsta.1933.0009

[38] P.Q. Nguyen, T. Vidick, Sieve algorithms for the shortest vector problem are practical. *J. Math. Cryptol.***2**(2), 181–207 (2008). https://doi.org/10.1515/JMC.2008.009

[39] R. Overbeck, Statistical decoding revisited, in L.M. Batten, R. Safavi-Naini (eds.), ACISP 06. LNCS, vol. 4058, (Springer, Berlin, Heidelberg, 2006), pp. 283–294. https://doi.org/10.1007/11780656_24

[40] A. Pouly, Y. Shen, Provable dual attacks on LWE with sieving. *Priv. Commun.* (2023).

[41] A. Pouly, Y. Shen, Provable dual attacks on learning with errors. in M. Joye, G. Leander (eds.), *EUROCRYPT 2024, Part VII*. LNCS, vol. 14657, (Springer, Cham, 2024), pp. 256–285. https://doi.org/10.1007/978-3-031-58754-2_10

[42] C.M. Rader, Discrete Fourier transforms when the number of data samples is prime. *Proc. IEEE* **56**(6), 1107–1108 (1968).

[43] O. Regev, On lattices, learning with errors, random linear codes, and cryptography. *J. ACM* **56**(6), 1–40 (2009). preliminary version in STOC'05.

[44] P. Schwabe, R. Avanzi, J. Bos, L. Ducas, E. Kiltz, T. Lepoint, V. Lyubashevsky, J.M. Schanck, G. Seiler, D. Stehlé, J. Ding, *CRYSTALS-KYBER. Tech. rep., National Institute of Standards and Technology* (2022). available at https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022

[45] E.M. Stein, G. Weiss, Introduction to Fourier analysis on euclidean spaces. No. 32, in *Princeton Mathematical Series*, (Princeton University Press, Princeton, New Jersey, 1971)

[46] A. Storjohann, Computing hermite and smith normal forms of triangular integer matrices. *Linear Algebra Appl.***282**(1), 25–45 (1998). https://doi.org/10.1016/S0024-3795(98)10012-5

[47] The FPLLL development team, fpylll, a Python wrapper for the fplll lattice reduction library, Version: 0.5.9 (2023). available at https://github.com/fplll/fpylll

[48] G.N. Watson, *A Treatise on the Theory of Bessel Functions*, (Cambridge University Press, 1922). https://archive.org/details/treatiseontheory00watsuoft

[49] A. Wiemers, S. Ehlen, K. Bashiri, A remark on the independence heuristic in the dual attack. *Cryptology ePrint Archive*, Report 2023/1238 (2023). https://eprint.iacr.org/2023/1238