

Does the Dual-Sieve Attack on Learning with Errors even Work?

Léo Ducas^{1,2}  and Ludo N. Pulles¹ 

¹ CWI, Cryptology Group, Amsterdam, the Netherlands

² Mathematical Institute, Leiden University, Leiden, The Netherlands

Abstract. Guo and Johansson (ASIACRYPT 2021), and MATZOV (tech. report 2022) have independently claimed improved attacks against various NIST lattice candidate by adding a Fast Fourier Transform (FFT) trick on top of the so-called Dual-Sieve attack. Recently, there was more follow up work in this line adding new practical improvements.

However, from a theoretical perspective, all of these works are painfully specific to Learning with Errors, while the principle of the Dual-Sieve attack is more general (Laarhoven & Walter, CT-RSA 2021). More critically, all of these works are based on heuristics that have received very little theoretical and experimental attention.

This work attempts to rectify the above deficiencies of the literature. We first propose a generalization of the FFT trick by Guo and Johansson to arbitrary Bounded Distance Decoding instances. This generalization offers a new improvement to the attack.

We then theoretically explore the underlying heuristics and show that these are in contradiction with formal, unconditional theorems in some regimes, and with well-tested heuristics in other regimes. The specific instantiations of the recent literature fall into this second regime.

We confirm these contradictions with experiments, documenting several phenomena that are not predicted by the analysis, including a “waterfall-floor” phenomenon, reminiscent of Low-Density Parity-Check decoding failures.

We conclude that the success probability of the recent Dual-Sieve-FFT attacks are presumably significantly overestimated. We further discuss the adequate way forward towards fixing the attack and its analysis.

Keywords: Lattices, Cryptanalysis, Heuristics, Learning with Errors, Dual Attack, Fast Fourier Transform

1 Introduction

The idea of using short dual vectors for distinguishing between points close to or far from a lattice was put forward, in a complexity theoretic context, by Aharonov and Regev [AR04], and can even be traced back in a pure geometric context to earlier work by Håstad [Hås88]. The problem at hand here is coined the decisional Bounded Distance Decoding problem (BDD). This idea is not even limited to lattices, and was already implicit in the very construction of Low-Density Parity-Check codes dating back to [Gal62].

This idea made its way into cryptanalysis of cryptosystems based on the Learning with Errors (LWE) problem in a survey of Micciancio and Regev [MR09]. Indeed, Learning with Errors is a special case of BDD, for a specific family of random lattices. An attack on LWE (or BDD) using this idea is a so-called *dual attack*, in contrast with the other type of attacks that operate solely in the *primal* lattice. The best dual and primal attacks are then typically used by the lattice cryptanalyst to instantiate LWE cryptosystems.

Since then, two fundamental developments have happened. The first, initially suggested by Alkim, Ducas, Pöppelmann, and Schwabe [ADPS16], consisted of exploiting the fact that lattice sieving [NV08, MV10, BDGL16] — a class of algorithms for finding short lattice vectors — naturally provides not only the shortest vector, but exponentially many short vectors. The hope is that the information of these exponentially many short dual vectors can be leveraged to improve a distinguisher, for example by summing a score function over them [LW21].

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 [Alb17, EJK20], and then finding the right solution among these candidates rather than just a few.

The second development is also reminiscent from a cryptanalytic technique of code-based cryptography by Leveil and Fouque [LF06]. The idea is to batch the score evaluation of a large number of algebraically related candidates via a Fast Fourier Transform. For carefully crafted parameters, the cost of getting all those scores is barely larger than the cost of naively computing a single score. This led Guo and Johansson [GJ21] to claim an improved attack on various NIST post-quantum standardization candidates, followed quickly by an independent technical report of MATZOV [MAT22]. 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 [AS22] and a coding-theoretic enhanced variant [CST22].

1.1 Contributions

Abstraction and Generalization of the FFT trick (Section 3). We note that the original principle of the dual attack [AR04, Hås88] is general: it applies to the bounded distance decoding problem (BDD) in arbitrary lattices. However, the recent instances of the Dual-Sieve attack [ADPS16, Alb17, EJK20] and the Dual-Sieve-FFT attacks [GJ21, MAT22, AS22, CST22] are described in a very specialized way to Learning with Errors. The only exception is the work of [LW21], which we find geometrically enlightening, although their work is limited to the Dual-Sieve attack.

Our first contribution (Section 3) is therefore to also generalize the FFT trick of [GJ21] to the general setting. Beyond the theoretical satisfaction of abstracting the technique to its mathematical core, this generalization also offers further improvement over the work of [GJ21]: for the same algorithmic price, we can further improve the shortness of the dual vectors and therefore their distinguishing power.

Contradictions from the Heuristic Analysis (Section 4). A second observation regarding this literature is that the analysis of the Dual-Sieve attack (with or without FFT) relies on one specific independence 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 [LW21] is pivotal in that respect: judging the reasonability of a heuristic conclusion is very much enabled by the language of geometry. In particular, their work concludes with a heuristic algorithm that distinguishes a noisy lattice point from random, even when the noise slightly exceeds the Gaussian Heuristic, *i.e.* the expected minimal distance of a random lattice. This should raise suspicion, as even random points are not expected to be much further away from the lattice than this minimal distance. This suspicion of invalidity becomes an undeniable contradiction by considering a recent result of Debris, Ducas, Resch, and Tillich [DADRT22], stating that the above task is statistically impossible, 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 [EJK20, GJ21, MAT22, AS22, CST22]. In their context, the BDD error is below the Gaussian Heuristic but the actual BDD sample needs to be discovered among a large number T of uniform samples. We will show that this also leads to a contradiction. Namely, for large T , we argue that many of those random targets will lie closer to the lattice than the BDD target itself. The claim that the BDD target can be successfully identified among so many random targets contradicts the very principle of the attack, namely that the expected score of a target increases with its closeness to the lattice. It turns out that the parameters used in [GJ21, MAT22] specifically fall into that contradictory regime that uses a large number T of targets.

Experiments (Section 5). To understand what is going on, we zoom in on the distribution of scores for random targets and BDD targets. We ran extensive experiments, and discovered that both distributions deviate from the predictions made under the independence heuristic. First, the *body* of the distribution of scores for random targets is properly predicted, but not its *tail*: after a predicted rapid decrease (visually, a *waterfall*), this distribution hits a *floor*. This is perfectly in line with our second contradictory regime: some random targets will be close to the lattice, and should therefore have a high score.

However, that is not all. The distribution of scores for BDD target 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, the probability of the score of a BDD target being low is higher than predicted.

All of the code that is used for the experiments, as well as the results of the experiments are publicly available at:

<https://github.com/ludopulles/DoesDualSieveWork>.

These experiments are implemented in `python` using the G6K and FPyLLL libraries [ADH⁺19, dt23], as well as a custom `C` library to accelerate the FFT step.

1.2 Conclusion

Our theoretical contradictions and our experiments both demonstrate that the underlying heuristic of the Dual-Sieve attack is invalid. Both phenomena uncovered by the experiments point to the success probability of the Dual-Sieve attack (with or without FFT) being presumably over-estimated by the current heuristic, at least in certain regimes of interest.

In particular, the concrete cryptanalytic claims of numerous works [ADPS16, EJK20, GJ21, MAT22, AS22, CST22] should be considered at least unsubstantiated, as these are currently based on a flawed heuristic. Still, some of those claims might not be that far from reality, but those of [GJ21, MAT22, AS22, CST22], being so deep in the contradictory regime, are presumably significantly far away from reality.

Afterthoughts (Section 6). We conclude our work with various discussions. First, we mention the prior occurrence of a similar *waterfall-floor* phenomenon in the coding literature [Ric06, VCN14, ABH⁺22] and relate to it. We then reflect on the source of the issue in the independence heuristic, and highlight the effect of these dependencies with a toy example. We finally discuss a suitable way forward in fixing the Dual-Sieve attack and its analysis.

Acknowledgments. We would like to thank Martin Albrecht, Qian Guo, Thomas Johansson, Eamonn Postlethwaite, Yixin Shen, Michael Walter, Wessel van Woerden for helpful discussion and feedback. Some of them might not endorse our conclusions. Authors Léo Ducas and Ludo Pulles are supported by ERC Starting Grant 947821 (ARTICULATE).

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.

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 is denoted by \mathcal{S}^1 and is naturally a subgroup of \mathbb{C}^* .

2.1 Probabilities and Distributions

Probabilities are denoted by \mathbb{P} , expectations by \mathbb{E} . The variance of a random variable X is $\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$, and its standard deviation is $\sigma_X = \sqrt{\mathbb{V}[X]}$. The *cumulative density function* (CDF) at $x \in \mathbb{R}$ of a distribution \mathcal{D} is $\mathbb{P}_{X \leftarrow \mathcal{D}}[X \leq x]$, and the *survival function* (SF) of \mathcal{D} is $\mathbb{P}_{X \leftarrow \mathcal{D}}[X \geq x] = 1 - \mathbb{P}_{X \leftarrow \mathcal{D}}[X < x]$, where X is drawn from \mathcal{D} .

The uniform distribution on a set X is denoted by $U(X)$. The continuous gaussian $\mathcal{N}(c, \sigma^2)$ of average $c \in \mathbb{R}$ and standard deviation $\sigma \in \mathbb{R}_{>0}$ has a probability density at $x \in \mathbb{R}$ proportional to $\rho_{c, \sigma}(x) = \exp\left(-\frac{(x-c)^2}{2\sigma^2}\right)$.

The (gaussian) *error function* is erf and the *complementary error function* is $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$, for a random variable $X \leftarrow \mathcal{N}(c, \sigma^2)$ we have for all $x \in \mathbb{R}$

$$\mathbb{P}[X < x] = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{x-c}{\sigma\sqrt{2}} \right) \right). \quad (1)$$

Lemma 1 ([AS64, 7.1.23]). *It holds that*

$$\operatorname{erfc}(x) = e^{-x^2} \cdot \left(\frac{1}{\sqrt{\pi} \cdot x} + O\left(\frac{1}{x^3}\right) \right) \quad \text{as } x \rightarrow \infty.$$

2.2 Lattices

A *lattice* Λ is a discrete subgroup of \mathbb{R}^n , its *rank* is the dimension of its \mathbb{R} -linear span, and the *volume* of a full rank lattice is $\det \Lambda = \operatorname{Vol}(\mathbb{R}^n / \Lambda)$. For $1 \leq k \leq n$, a *basis* $\mathbf{B} \in \mathbb{R}^{n \times k}$ consisting of \mathbb{R} -linearly independent column vectors $\mathbf{b}_1, \dots, \mathbf{b}_k$ defines the rank- k lattice

$$\mathcal{L}(\mathbf{B}) = \{ \mathbf{v} \in \mathbb{R}^n \mid \exists \mathbf{c} \in \mathbb{Z}^k : \mathbf{v} = c_1 \mathbf{b}_1 + \dots + c_k \mathbf{b}_k \},$$

of volume $\sqrt{\det \mathbf{B}^T \mathbf{B}}$. For $1 \leq \ell \leq r \leq n$, we use $\mathbf{B}_{[\ell, r]}$ for the basis consisting of vectors $\pi_\ell(\mathbf{b}_\ell), \dots, \pi_\ell(\mathbf{b}_r)$, where π_ℓ is the projection map that projects away from $\mathbf{b}_1, \dots, \mathbf{b}_{\ell-1}$. The length of a shortest nonzero vector of a lattice $\Lambda \subset \mathbb{R}^n$ is denoted by $\lambda_1(\Lambda)$.

Duality. There are two ways to define the dual of a lattice, the first one being geometric and specific to lattices, while the second is inherited from groups. In the context of the FFT trick, it is useful to consider both definitions, and relate them.

Definition 1. *The dual lattice Λ^\vee of a full rank lattice $\Lambda \subset \mathbb{R}^n$ is the set of all $\mathbf{w} \in \mathbb{R}^n$ such that $\langle \mathbf{w}, \Lambda \rangle \subseteq \mathbb{Z}$.*

Definition 2. *For a full rank lattice $\Lambda \subset \mathbb{R}^n$, let*

$$\widehat{\Lambda} = \{ \chi : \mathbb{R}^n / \Lambda \rightarrow \mathcal{S}^1 \mid \chi \text{ continuous group hom.} \}$$

denote the group of characters on the torus \mathbb{R}^n / Λ .

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. *The map from Λ^\vee to $\widehat{\Lambda}$ that sends a dual vector \mathbf{w} to the character*

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

is a group isomorphism.

Proof. **Well-definedness:** this follows directly from the definition of Λ^\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: note that for any character $\chi: \mathbb{R}^n/\Lambda \rightarrow \mathcal{S}^1$, by continuity, there is an open ball $U \subseteq \mathbb{R}^n$ centred at $\mathbf{0}$ such that we have $\chi(U + \Lambda) \subseteq \{z \in \mathcal{S}^1 \mid \operatorname{Re}(z) > 0\}$. On this ball, we can find a linear map $\varphi: U \rightarrow (-\frac{1}{2}, \frac{1}{2})$ 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 linearly to a function $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$, 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 $\langle \mathbf{w}, \Lambda \rangle \subseteq \mathbb{Z}$, i.e. $\mathbf{w} \in \Lambda^\vee$ because $\chi(\Lambda) = 1$ and therefore $\chi = \chi_{\mathbf{w}}$.

For a finite abelian group G , the dual group \widehat{G} is the group of the homomorphisms $\chi: G \rightarrow \mathcal{S}^1$. For a sublattice $\Lambda' \subset \Lambda$, the dual of Λ/Λ' has a natural connection to the dual lattices of Λ' and Λ by the following lemma.

Lemma 3. *For two full rank lattices $\Lambda_1 \subset \Lambda_2 \subset \mathbb{R}^n$, there is a canonical group isomorphism of abelian groups,*

$$\widehat{\Lambda_1/\Lambda_2} \rightarrow \widehat{\Lambda_2/\Lambda_1},$$

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

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

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

Surjectivity: left hand side has size $|\widehat{\Lambda_1/\Lambda_2}| = |\Lambda_1^\vee/\Lambda_2^\vee| = |\Lambda_2/\Lambda_1|$ and right hand side has size $|\widehat{\Lambda_2/\Lambda_1}| = |\Lambda_2/\Lambda_1|$ 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 $\mathbf{B}^\vee = \mathbf{B} \cdot (\mathbf{B}^\top \cdot \mathbf{B})^{-1}$ of the dual lattice Λ^\vee . Consider the *reversed dual basis* ${}^\vee\mathbf{B} = [\mathbf{b}_n^\vee, \dots, \mathbf{b}_1^\vee]$ in which the ordering of the basis vectors is reversed. A basis for the dual of $\mathcal{L}(\mathbf{B}_{[\ell, r]})$ is given by $\tau(\mathbf{b}_r^\vee), \dots, \tau(\mathbf{b}_\ell^\vee)$ where τ is the map projecting away from $\mathbf{b}_{r+1}^\vee, \dots, \mathbf{b}_n^\vee$, 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 [Mic14].

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. The *Discrete Fourier Transform* (DFT) of a sequence $(x_g)_{g \in G} \subset \mathbb{C}$ is the \mathbb{C} -linear map

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

The m -dimensional *Fast Fourier Transform* (FFT) is an algorithm that, upon input a group G , given as $n_1, \dots, n_m \in \mathbb{Z}_{\geq 2}$ such that $G \cong \bigoplus_{j=1}^m (\mathbb{Z}/n_j\mathbb{Z})$, and $(x_g)_{g \in G}$, outputs $\text{DFT}_G((x_g)_{g \in G})$ 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) [Rad68, DV90]. 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

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

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

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

Heuristic 1 (Gaussian Heuristic) *Given a random lattice $\Lambda \subset \mathbb{R}^n$ of volume 1, then $\lambda_1(\Lambda)$ is approximately*

$$\text{GH}(n) := \text{Vol}(\mathcal{B}^n)^{-1/n} = \frac{\Gamma(1 + \frac{n}{2})^{1/n}}{\sqrt{\pi}} \approx \sqrt{\frac{n}{2\pi e}} \cdot (\pi n)^{1/n},$$

where we use Stirling's formula in the approximation step.

Note that Minkowski's theorem states $\lambda_1(\Lambda) \leq 2 \cdot \text{GH}(n)$.

Heuristic 2 *Given a random lattice $\Lambda \subset \mathbb{R}^n$ of volume 1, for $r > 1$ we have*

$$|\{ \mathbf{v} \in \Lambda \mid \|\mathbf{v}\| \leq r \cdot \text{GH}(n) \}| \approx r^n.$$

In particular, the i^{th} shortest lattice point \mathbf{v} has length $\|\mathbf{v}\| \approx \text{GH}(n) \sqrt[i]{i}$.

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

Problem 1 (BDD, Lattice Form). For $r > 0$, *Bounded Distance Decoding* (BDD_r) is the task of, given a lattice Λ and a target $\mathbf{t} \in \mathbb{R}^n$ with the promise that there exists a nearby lattice point $\mathbf{v} \in \Lambda$ at distance at most $r\lambda_1(\Lambda)$ away from \mathbf{t} , finding the point $\mathbf{v} \in \Lambda$.

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

Problem 2 (BDD, Syndrome Form). For $r > 0$, (syndrome) *Bounded Distance Decoding* (BDD_r) is the task of, given a lattice Λ and target $\mathbf{t} \in \mathbb{R}^n/\Lambda$ in the torus with the promise that there exists $\mathbf{e} \in \mathbf{t}$ such that $\|\mathbf{e}\| < r\lambda_1(\Lambda)$, 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 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 you move up to $r < 1$ by the following heuristic.

Heuristic Claim 1 *Let Λ be a random lattice of volume 1, $r \in (0, 1)$. 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 $O(n\sqrt{n})r^n$, where $R = r\text{GH}(n)$.*

This Heuristic can be justified with the Gaussian Heuristic and an upper bound on spherical domes, cf. [MV10, Lem. 4.1].

Heuristic Justification. Note that only lattice points $\mathbf{v} \in \Lambda \setminus \{\mathbf{0}\}$ are relevant with $\|\mathbf{v}\| \leq 2R = 2r\text{GH}(n)$ by the triangle inequality. For such a $\mathbf{v} \in \Lambda \setminus \{\mathbf{0}\}$, we are interested in $\text{Vol}(R\mathcal{B}^n \cap (\mathbf{v} + R\mathcal{B}^n))$, which is twice the volume of the spherical dome $\{\mathbf{t} \in R\mathcal{B}^n \mid \langle \mathbf{t}, \mathbf{v} \rangle \geq \frac{1}{2}\|\mathbf{v}\|^2\}$. 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 $R^n(1 - \alpha^2)^{n/2}\text{Vol}(\mathcal{B}^{n-1})$, where $\alpha = \|\mathbf{v}\|/2R$. One can show that $\text{Vol}(\mathcal{B}^{n-1}) \leq \frac{\sqrt{en}}{2}\text{Vol}(\mathcal{B}^n)$ holds, which implies $\text{Vol}(R\mathcal{B}^n \cap (\mathbf{v} + R\mathcal{B}^n)) \leq O(\sqrt{n})r^n(1 - \alpha^2)^{n/2}$.

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

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

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

2.3 Dual Distinguishing

The idea of using short dual vectors for distinguishing between BDD samples and random samples can be traced back at least to [AR04] in the lattice literature, and can be viewed as a lattice analog to an old decoding technique [Gal62, Jab01, Ove06]. 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,

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

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 that the target \mathbf{t} is close to the lattice Λ , is therefore given by,

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

If \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 boost one's 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,

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

This function is referred to as the simple decoder f_{simple} by Laarhoven and Walter [LW21], and resembles the Aharonov-Regev [AR04] decoder closely which is given by $f_{\mathbf{w}}^{\text{AR}}(\mathbf{t}) := \rho_{1/\sigma}(\mathbf{w}) f_{\mathbf{w}}(\mathbf{t})$.

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 [AR04] or certificate [Häs88].

More recent works [LW21, GJ21, MAT22] 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 [BDGL16]), and for a random error \mathbf{e} .

The Analysis of [LW21]. First, Laarhoven and Walter analyze in [LW21, Lem. 6] the distribution of the score $f_{\mathbf{w}}(\mathbf{t})$ of BDD targets \mathbf{t} with a distance of exactly r to the primal lattice. In the derivation, they approximate this distribution by targets sampled from a continuous gaussian with $\sigma = r/\sqrt{n}$.

Lemma 4 (cf. [LW21, Lemma 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*

$$\mathbb{E}[f_{\mathbf{w}}(\mathbf{t})] = e^{-2\pi^2 \sigma^2 \|\mathbf{w}\|^2}, \quad \text{and} \quad \mathbb{V}[f_{\mathbf{w}}(\mathbf{t})] = \frac{1}{2} - \Theta\left(e^{-4\pi^2 \sigma^2 \|\mathbf{w}\|^2}\right). \quad (8)$$

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 $\int_0^1 \cos(\alpha + 2\pi kx) = 0$ 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,

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

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, one must resort to the following heuristic.

Heuristic 3 (Independence Heuristic) *For any fixed set $\mathcal{W} \subset \Lambda^V$ and any distribution for \mathbf{t} considered in Lemma 4, the random variables $(\langle \mathbf{w}, \mathbf{t} \rangle \bmod 1)_{\mathbf{w} \in \mathcal{W}}$ are mutually independent.*

We will refer to this heuristic as the Independence Heuristic.

By combining the above heuristic with a central limit approximation — which is 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 4.

One may then deduce the distinguishing advantage of the score function using the following lemma.

Lemma 5. *Let $X \leftarrow \mathcal{N}(E_X, V_X)$ and $Y \leftarrow \mathcal{N}(E_Y, V_Y)$ be independent gaussian random variables. Then*

$$\mathbb{P}[X > Y] = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{E_X - E_Y}{\sqrt{2(V_X + V_Y)}} \right) \right] \quad (9)$$

Proof. Consider the variable $Z = X - Y$, which is also gaussian, specifically $Z \sim \mathcal{N}(E_Z, V_Z)$ where $E_Z = E_X - E_Y$ and $V_Z = V_X + V_Y$. Conclude noting that the event $X > Y$ is equivalent to $Z > 0$.

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

Heuristic Claim 2 (cf. [LW21, Lem. 9]) *Let $\Lambda \subset \mathbb{R}^n$ be a random lattice of volume 1, $r > 0$ and $\mathcal{W} \subset \Lambda^V$ a set consisting of the α^n shortest vectors of Λ^V , where $\alpha = \min\{\beta \mid e^2 \ln(\beta) = \beta^2 r^2\}$. Then, we have*

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

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

Heuristic Justification. First, we approximate the uniform distribution of BDD samples $\mathbf{t}_{\text{BDD}} \leftarrow U(r\text{GH}(n)\mathcal{S}^{n-1})$ by a gaussian distribution with $\sigma = r\text{GH}(n)/\sqrt{n}$. According to the Gaussian Heuristic, the lengths of the vectors in \mathcal{W} are concentrated around $\alpha \cdot \text{GH}(n)$. By the Independence Heuristic and Lemma 4, the score function for the BDD sample follows a gaussian distribution $\mathcal{N}(E_X, V_X)$, where

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

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 4.

Hence by Lemma 5, 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 [GJ21] and [MAT22]. The analysis proposed by Guo–Johansson [GJ21] 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 Independence 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 [LW21] using a lemma dating back to Neyman–Pearson [NP33]), 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 [LW21, Corollary 2], but it is not mentioned by Guo and Johansson [GJ21].

The analysis of MATZOV [MAT22] 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 [MAT22, 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 Appendix A.4).

In any case, we can essentially recover the key claims of [GJ21, MAT22] directly from the above analysis as well, using the following lemma. Note that we express the result more generally in terms of BDD in an arbitrary lattice rather than a specific LWE instance.

Lemma 6. *Let $X \leftarrow \mathcal{N}(E_X, V_X)$ and $Y_i \leftarrow \mathcal{N}(E_Y, V_Y)$ be independent gaussian random variables for $i \in \{1, \dots, T\}$. Then*

$$\mathbb{P}\left[X > \max_{i \in \{1, \dots, T\}} (Y_i)\right] \geq 1 - \frac{T}{2} \operatorname{erfc}\left(\frac{E_X - E_Y}{\sqrt{2 \cdot (V_X + V_Y)}}\right) \quad (10)$$

Proof. Follows directly from a union bound on the complementary events with Lemma 5.

This leads to the following heuristic claim, underlying the work of [GJ21, MAT22]. Note that we state this key claim here for a *projected sublattice* with renormalization of the volume, where the dual vectors are in the dual of this projected sublattice. Indeed, the general setting of the Dual-Sieve attack [ADPS16, EJK20, GJ21, MAT22] first applies BKZ on the dual of a lattice of dimension d , obtaining a reversed dual basis ${}^\vee\mathbf{B}$. Then the set \mathcal{W} is obtained by running a sieve in dimension β_{sieve} on the first dual block, that is the β_{sieve} -dimension dual sublattice $\Lambda^\vee \subset \Lambda_{\text{LWE}}^\vee$ generated by the partial reversed dual basis $({}^\vee\mathbf{B})_{[1, \beta_{\text{sieve}}]}$. This is the dual of the projected sublattice generated by $\mathbf{B}_{[d+1-\beta_{\text{sieve}}, d]}$ of the full primal lattice Λ_{LWE} .

Heuristic Claim 3 (Key claim of [GJ21, MAT22], reconstructed) *Let $\Lambda \subset \mathbb{R}^n$ be a random lattice of volume 1, $\mathcal{W} \subset \Lambda^\vee$ the set consisting of the $(4/3)^{n/2}$ shortest vectors of Λ^\vee . For some $\sigma > 0$ and $T \in \mathbb{Z}_{\geq 1}$, consider $\mathbf{t}_{\text{BDD}} \leftarrow \mathcal{N}(0, \sigma^2)^n$ and i.i.d. $\mathbf{t}_{\text{unif}}^{(i)} \leftarrow U(\mathbb{R}^n/\Lambda)$ where $i \in \{1, \dots, T\}$. Set $\ell = \sqrt{4/3} \cdot \text{GH}(n)$ and $\varepsilon = \exp(-2\pi^2\sigma^2\ell^2)$. If $\ln T \leq |\mathcal{W}|\varepsilon^2$, we have*

$$\mathbb{P}\left[\forall i \in \{1, \dots, T\} : f_{\mathcal{W}}(\mathbf{t}_{\text{BDD}}) > f_{\mathcal{W}}(\mathbf{t}_{\text{unif}}^{(i)})\right] \geq 1 - O\left(\frac{1}{\sqrt{\ln T}}\right).$$

Heuristic Justification. Similar to the Heuristic Justification of Heuristic Claim 2, the score distribution for \mathbf{t}_{BDD} is approximately $X \sim \mathcal{N}(\varepsilon \cdot |\mathcal{W}|, \frac{1}{2}|\mathcal{W}|)$, as lengths of vectors in \mathcal{W} are concentrated around $\ell = \sqrt{4/3} \cdot \text{GH}(n)$ by the Gaussian Heuristic. The uniform samples $\mathbf{t}_{\text{unif}}^{(i)}$ each give a score distribution that is approximately $Y_i \sim \mathcal{N}(0, \frac{1}{2}|\mathcal{W}|)$.

Thus by Lemma 6, the probability that $f_{\mathcal{W}}(\mathbf{t}_{\text{BDD}})$ is bigger than every score of the uniform errors equals

$$1 - \frac{T}{2} \operatorname{erfc}\left(\frac{|\mathcal{W}| \cdot \varepsilon}{\sqrt{|\mathcal{W}|}}\right) \geq 1 - \frac{T}{2} \operatorname{erfc}\left(\sqrt{\ln T}\right).$$

Hence, with Lemma 1 we conclude,

$$\mathbb{P}\left[X > \max_{i \in \{1, \dots, T\}} (Y_i)\right] \geq 1 - \frac{1}{2\sqrt{\pi \ln T}} \left(1 + O\left(\frac{1}{\ln T}\right)\right) \geq 1 - O\left(\frac{1}{\sqrt{\ln T}}\right).$$

3 Dual-Sieve-FFT Distinguishing, Generalized

As established by the literature [Häs88, AR04, LW21], 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 [GJ21]. We further show in Section 3.4 that the version of MATZOV [MAT22] can be understood in two separate steps, the second step fitting our formalization of the FFT trick as well.

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 Equation (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,

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

where $\mathbf{w} + \Lambda' \in (\Lambda')^\vee/\Lambda'$ as \widehat{G} is isomorphic to $(\Lambda')^\vee/\Lambda'$ by Lemmata 2 and 3. 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

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

Hence, Equation (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

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

Neglecting this scalar $|G|$, we therefore construct a batch of score functions, by performing an inverse FFT on the sequence $\sum_{\mathbf{w}' \in \mathbf{w} + \Lambda'} f_{\mathbf{w}'}(-\mathbf{t})$ for all (dual) cosets $\mathbf{w} + \Lambda' \in (\Lambda')^\vee/\Lambda'$. 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 $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_n]$, and one can take a simple sparsification such as $\Lambda' = \mathcal{L}([d_1 \mathbf{b}_1, \dots, d_n \mathbf{b}_n])$ for suitable $d_1, \dots, d_n \in \mathbb{N}$.

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 $\mathbf{S}, \mathbf{T} \in \text{GL}_n(\mathbb{Z})$ and a diagonal matrix \mathbf{D} such that $\mathbf{A} = \mathbf{S}\mathbf{D}\mathbf{T}$ and thus, we have $\mathbf{B}'\mathbf{T}^{-1} = (\mathbf{B}\mathbf{S}) \cdot \mathbf{D}$. 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 $[d_1 \mathbf{b}'_1, \dots, d_n \mathbf{b}'_n]$, where $\text{diag}(d_1, \dots, d_n) = \mathbf{D}$ and $\mathbf{B}\mathbf{S} = [\mathbf{b}'_1, \dots, \mathbf{b}'_n]$ is a basis for Λ .

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

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

Require:

- 1: A basis \mathbf{B} of a full rank lattice $\Lambda \subset \mathbb{R}^n$,
- 2: A basis $\mathbf{B}' = \mathbf{B} \cdot \text{diag}(d_1, \dots, d_n)$ of $\Lambda' \subset \Lambda$,
- 3: A set of short dual vectors $\mathcal{W} \subset (\Lambda')^\vee$,
- 4: A target $\mathbf{t} \in \mathbb{R}^n / \Lambda'$

Ensure:

- 5: A lattice coset $g \in \Lambda / \Lambda'$ closest to \mathbf{t}
 - 6: _____
 - 7: Initialize a table \mathbf{T} with zeros of dimension $d_1 \times d_2 \times \dots \times d_n$
 - 8: **for** $\mathbf{w} \in \mathcal{W}$ **do**
 - 9: Write $\mathbf{w} \equiv \frac{j_1}{d_1} \mathbf{b}_1^\vee + \dots + \frac{j_n}{d_n} \mathbf{b}_n^\vee \pmod{\Lambda^\vee}$, where $0 \leq j_i < d_i$
 - 10: $\mathbf{T}[j_1, j_2, \dots, j_n] \leftarrow \mathbf{T}[j_1, j_2, \dots, j_n] + \cos(2\pi \langle \mathbf{w}, \mathbf{t} \rangle)$
 - 11: **end for**
 - 12: $\mathbf{S} = \text{DFT}_{\mathbb{Z}/d_1\mathbb{Z} \times \dots \times \mathbb{Z}/d_n\mathbb{Z}}^{-1}(\mathbf{T})$
 - 13: $(j_1, j_2, \dots, j_n) \leftarrow \underset{\substack{k_1, k_2, \dots, k_n \\ 0 \leq k_i < d_i}}{\text{argmax}} \{ \mathbf{S}[k_1, k_2, \dots, k_n] \}$
 - 14: **return** $j_1 \mathbf{b}_1 + \dots + j_n \mathbf{b}_n$
-

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 $\mathbf{B}' = [2\mathbf{b}_1, \dots, 2\mathbf{b}_k, \mathbf{b}_{k+1}, \dots, \mathbf{b}_n]$, 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 $\text{diag}\left(\frac{1}{d_1}, \dots, \frac{1}{d_n}\right)$, 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 [GJ21] 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 $\mathbf{B}' = \mathbf{B} \cdot \text{diag}(\gamma, \dots, \gamma, 1, \dots, 1)$ (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 $\mathbf{B}' = \mathbf{B} \cdot \text{diag}(\gamma, \dots, \gamma, 1, \dots, 1)$ on the length of the vectors in \mathcal{W} is to shorten them 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 $\mathbf{B}' = \mathbf{B} \cdot \text{diag}(1, \dots, 1, \gamma, \dots, \gamma)$. In this way, the densification of the reversed dual basis is given by ${}^\vee(\mathbf{B}') = {}^\vee\mathbf{B} \cdot \text{diag}\left(\frac{1}{\gamma}, \dots, \frac{1}{\gamma}, 1, \dots, 1\right)$ remains concentrated on the k first dual vectors!¹ Therefore, its effect is now to shorten the length of the vectors in \mathcal{W} by a factor $\gamma^{k/\beta_{\text{sieve}}}$ (assuming $k < \beta_{\text{sieve}}$).

We leave the concrete exploitation of this improvement to future work, given that the next section of this work will invalidate the current success probability analysis of the dual attack as a whole. Before the improved concrete cryptanalysis claims are made, the attack should first be convincingly fixed!

Open Question 1 *Produce estimates of the gain of the above improvement of the Dual-Sieve-FFT attack on relevant concrete lattice-based schemes, once the general analysis of Dual-Sieve attacks has been fixed (cf. Open Question 2).*

3.4 What about the Version of MATZOV?

The algorithm of MATZOV [MAT22] differs a bit from that of Guo–Johansson [GJ21] 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 note however, that it does not appear to sparsify the primal by a factor $|G|$. Indeed,

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

the volume of the dual lattice in [MAT22] is a power of the LWE modulus q , unrelated to the modulus p underlying the FFT group structure $G \cong (\mathbb{Z}/p\mathbb{Z})^k$.

In this section, we explain that the algorithm of MATZOV [MAT22] is best understood as split into two steps: the first step is a purely geometric transformation of the BDD instance, trading dimension for a densification of the lattice and distortion of the error, and the second step is equivalent to the Dual-Sieve-FFT described above. In the parametrization of MATZOV [MAT22], the densification and sparsification cancel out exactly. This view may allow to decorrelate those two steps, offering a larger parametrization space; in particular one may in principle choose $G \cong (\mathbb{Z}/2\mathbb{Z})^k$ to exploit the advantages of the WHT without having to give up on the dimension-reduction technique of MATZOV.

Consider an n -dimensional lattice Λ with basis \mathbf{B} , and some dimension $r < n$. Now consider the orthogonal projection π onto the span of $\mathbf{B}_{[1,r]}$, as well as the lattice Λ_0 generated by $\mathbf{B}_{[1,r]}$. In general, $\pi(\Lambda)$ is not a lattice; it is still a group in a subspace of dimension r , but it is not always a discrete. However, it *can* be a lattice in special cases. If it is the case that $p \cdot \pi(\mathbf{b}_i) \in \Lambda_0$ for every $i \in \{r+1, \dots, n\}$, then $\pi(\Lambda)$ is a lattice, and more precisely it is a densification of Λ_0 of index $|\pi(\Lambda)/\Lambda_0| \leq p^{n-r}$.

Hence, in that case, solving the initial BDD instance \mathbf{t} in Λ can be related to solving the BDD instance $\pi(\mathbf{t})$ in $\pi(\Lambda)$. That is, we have restricted the BDD problem in dimension n to BDD in dimension r . Now we are *not* working in the sublattice Λ_0 , but in a densification of it.

The modulus switching of MATZOV should be understood as applying a carefully crafted linear transformation \mathbf{T} so that $p \cdot \pi(\mathbf{T} \cdot \mathbf{b}_i) \in \Lambda_0$ does holds for every $i \in \{r+1, \dots, n\}$, to construct a distorted lattice Λ_T generated by $[\mathbf{b}_1, \dots, \mathbf{b}_r, \mathbf{T}\mathbf{b}_{r+1}, \dots, \mathbf{T}\mathbf{b}_n]$. Then, one may attempt to solve the BDD problem on $\pi(\mathbf{t})$ in the densified lattice $\pi(\Lambda_T)$ of Λ_0 , hoping that \mathbf{T} does not increase the error too much. But applying the FFT trick with the group $G = \pi(\Lambda_T)/\Lambda_0$ precisely cancels out this densification.

We do *not* claim that \mathbf{T} would be necessarily easy to construct in the general BDD setting. The purpose of this subsection is more about making geometric sense of the variant of MATZOV, than about generalizing it.

Remark. We remain rather circumspect regarding the perceived superiority of the variant of MATZOV [MAT22] over that of Guo–Johansson [GJ21]. We discuss this point in Appendix A.6.

4 Contradictions from the Heuristic Analysis

In the following, we will show two regimes where the analyses of [LW21, GJ21, MAT22] give rise to absurd conclusions.

The first one is concerned with distinguishing a target from a single uniform sample, when its expected distance to the lattice exceeds the Gaussian Heuristic, a task that was recently proven statistically impossible in a random lattice [DADRT22].

The second one is concerned with the case of finding a planted solution among many candidates. For certain parameters, the analysis of [GJ21, MAT22] 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 or not one uses the FFT trick, but merely arises from the large number of candidates that is used.

Recall that in this section, as in [LW21], we implicitly renormalize the lattice Λ to have volume 1.

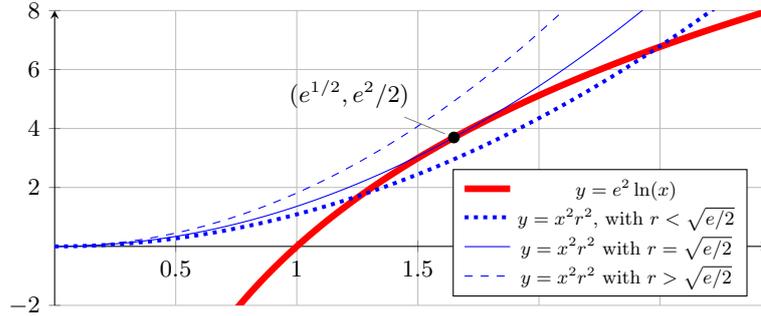


Fig. 1. The equation $e^2 \ln(x) = x^2 r^2$ for various r .

4.1 Distinguishing the Indistinguishable

Set Up. Recall that the main result of Laarhoven and Walter [LW21, Lem. 9], reformulated as Heuristic Claim 2, provides an algorithm to distinguish a BDD instance at distance $r \cdot \text{GH}(n)$ from a uniform target modulo the lattice. After a precomputation depending solely on the lattice Λ , this algorithm has exponential complexity α^n where $e^2 \ln(\alpha) = \alpha^2 r^2$, and the heuristic analysis claims that it is successful with probability ≈ 0.84 , that is, with constant advantage.

Lemma 7. *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 is illustrated by Figure 1. 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 $\left. \frac{dy}{dx} \right|_{x=\sqrt{e}} = e^{3/2}$. By convexity, we have $e^2 \ln(x) < x^2 r^2$ for any $x \neq \sqrt{e}$.

Case 2: $r^2 > e/2$. Note that $r^2 x^2$ is strictly increasing in r^2 for any x . Reusing Case 1, we see that $e^2 \ln(x) < x^2 r^2$ for all x : there are no solution in that case.

Case 3: $r^2 < e/2$. We have $e^2 \ln(x) > x^2 r^2$ at $x = \sqrt{e}$. We also have $e^2 \ln(x) < x^2 r^2$ 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 $r \cdot \text{GH}(n)$ for a uniform target is exactly r^n by the Gaussian Heuristic. More formally, it is a theorem² that for any measurable set $V \subset \mathbb{R}^n$, it holds that

$$\mathbb{E}_{\mathbf{t} \leftarrow U(\mathbb{R}^n/A)} [(V + \mathbf{t}) \cap A] = \frac{\text{Vol}(V)}{\det A}.$$

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. [DADRT22] that the above scenario does not occur with random lattices. More specifically, for a formally defined notion of random lattices, it is proven that for errors from a uniform distribution on the ball of radius $r \text{GH}(n)$ ($r > 1$), the statistical distance between the error modulo the lattice and $U(\mathbb{R}^n/A)$ is exponentially small as a function of the dimension [DADRT22, 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, [LW21, Lem. 9] (reformulated as Heuristic Claim 2) claims a constant advantage. $\not\Leftarrow$

Discussion. One could counter-argue that the claim [LW21, Lem. 9] is given for uniform distributions over a sphere, a case not contradicted by Debris et al. [DADRT22]. However, the actual analysis in [LW21] is done for a Gaussian distribution, a case which *is* also covered by Debris et al. [DADRT22, Theorem 4.6].

The Suspect Heuristic. We note that this counter-argument applies only to the (heuristic) [LW21, Lem. 9], that is given a single sample, and not to [LW21, Lem. 8]. Indeed, in the context of the (heuristic) [LW21, 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 [LW21, Lem. 8] does not require the Independence Heuristic, as it uses only one dual vector. In fact, after close inspection of the reasoning behind [LW21, 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.

² The difference with the Gaussian Heuristic being the presence of a uniform random shift $\mathbf{t} \leftarrow U(\mathbb{R}^n/A)$.

This sets the Independence Heuristic as the prime suspect leading to the erroneous [LW21, Lem. 9].

4.2 Candidates Closer than the Solution (Asymptotic)

Set Up. Recall that the key claim of [GJ21] and [MAT22], reconstructed as Heuristic Claim 3 considers the case where the set of dual vectors comes from a lattice sieve [NV08, MV10, BDGL16], that is, it consists of $N = (4/3)^{n/2}$ vectors of length $\ell = \sqrt{4/3} \cdot \text{GH}(n)$. Given one BDD instance with the error sampled from a gaussian with parameter σ and T uniform samples, the claim is made that the BDD sample can be detected with probability close to 1, whenever $\ln T \leq N\varepsilon^2$ where $\varepsilon = \exp(-2\pi^2\sigma^2\ell^2)$.

Let us consider the constraint the other way around, that is, how large can one take σ for a given number of targets T ? The condition translates to $\frac{1}{\varepsilon} \leq \sqrt{N/\ln T}$, and this constrains σ to satisfy

$$\sigma \leq \sqrt{\frac{\ln(1/\varepsilon)}{2\pi^2\ell^2}} = \sqrt{\frac{\ln N - \ln \ln T}{4\pi^2\ell^2}} = \frac{1}{2\pi\sqrt{\frac{4}{3}} \cdot \text{GH}(n)} \cdot \sqrt{\frac{n}{2} \ln \frac{4}{3} - \ln \ln T}.$$

With $\text{GH}(n) \approx \sqrt{\frac{n}{2\pi e}}$, one then arrives at $\sigma \leq \sqrt{C - \frac{C' \ln \ln T}{n}}$ 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 3 supposedly still finds a BDD sample at expected distance $\sqrt{C \cdot n} \approx 0.89 \text{GH}(n)$, even among a number of random candidates as large as doubly-exponential $T = \exp(\exp(n^{.99}))$.

The Contradiction (Asymptotic). We will show that the above claim leads to a contradiction, already for a single-exponential $T = 2.05^n$ number of random candidates.

Lemma 8. *Let Λ be a lattice of volume 1, and $r > 0$ such that $r < \frac{\lambda_1(\Lambda)}{2 \text{GH}(n)}$. Then, for a target \mathbf{t} uniform in \mathbb{R}^n/Λ , it holds with probability r^n that \mathbf{t} is at distance at most $r \text{GH}(n)$ from the lattice.*

Proof. Note that the volume of the ball of radius $r \cdot \text{GH}(n)$ is exactly r^n by definition of $\text{GH}(n)$. Furthermore, because $r \cdot \text{GH}(n) < \lambda_1(\Lambda)/2$, 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/Λ is injective. Hence, the ball modulo the lattice also has volume r^n in \mathbb{R}^n/Λ . The probability of \mathbf{t} falling into that ball is therefore $r^n/\text{Vol}(\mathbb{R}^n/\Lambda) = r^n/\det \Lambda = r^n$.

Let us use this lemma in the case of a random lattice of volume 1, or more specifically, one where we expect $\lambda_1(\Lambda) \approx \text{GH}(n)$. Using Lemma 8 with $r = 0.49$, the probability that a uniform target lies in the ball of radius $r \text{GH}(n)$ equals r^n . When taking $T = 2.05^n$ uniform samples, on expectation we have $T \cdot r^n > 1.004^n \gg 1$ of the uniform samples to fall in this ball and therefore with high

probability there will be one such target at most $r \text{GH}(n)$ away from a lattice point. More concretely, the probability that none of these targets lies in this ball is $(1 - r^n)^T \rightarrow e^{-1.004^n}$ (as $n \rightarrow \infty$), so with overwhelming probability there is a uniform target in the ball of radius $r \text{GH}(n)$.

On the other hand, recall that the actual BDD target had an expected length of $\sigma\sqrt{n} \approx 0.89 \text{GH}(n) > r \text{GH}(n)$. We note that $0.89 > r$, so we expect one uniformly sampled candidate lying closer to the lattice than the solution we are looking for. However, the score function $f_{\mathcal{W}}$ is precisely meant to associate larger score to closer targets, so we expect that this uniform sample will get a higher score than the BDD sample and thus, the algorithm gives with overwhelming probability a wrong result. ζ

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, if we consider super-exponential number of uniform samples, for example $T = n^{2n}$, with the same argument, for some constant probability, there exists a random target \mathbf{t} lying at distance $\frac{1}{n^2} \text{GH}(n) = O(n^{-3/2})$ from the lattice. 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 of \mathbf{t} will be $f_{\mathcal{W}}(\mathbf{t}) \geq N(1 - O(1/n^2))$, which is essentially maximal.

The Suspect Heuristic. The discussion above points to the same suspect as Section 4.1, namely, the Independence 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 can not hold for such large choices of $N = \omega(n \log n)$, and this should be visible in the tail of the score distribution of uniform targets.

4.3 Candidates Closer than the Solution (Concrete)

Set Up. In the contradiction above, we have chosen T as large as 2.05^n to be able to invoke Lemma 8 to have a uniform sample at distance $r \text{GH}(n) < \frac{1}{2} \lambda_1(A)$, quantifying the probability that a random target in \mathbb{R}^n/A falls close to the lattice. This leads to an *over-contradiction*: we exhibited the existence of a random target at distance $0.49 \text{GH}(n)$ from the lattice, much closer than the planted solution, at distance $0.89 \text{GH}(n)$. However, even when there is a uniform sample at distance e.g. $0.88 \text{GH}(n)$ from the lattice, this sample can get a higher score than the BDD sample resulting in an incorrect guess of which one the BDD sample was.

To extend Lemma 8 up to radii $r < 1$, we will resort to a heuristic instead. In this regime, translations of the ball by points in A may start to intersect. In practice, however, the volume of this intersection remains rather small and should not affect the volume so much.

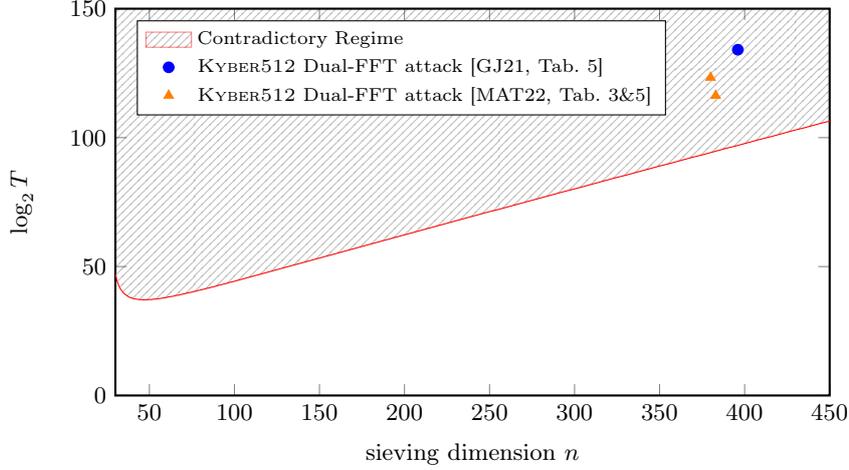


Fig. 2. Concrete Contradictory Regime: Maximum number of targets T before one is expected to be closer to a random lattice of dimension d than the planted solution. Obtained with script `volumetric_contr.py`.

Heuristic Claim 4 *Let A be a random lattice of volume 1, and $r \in (0, 1)$. For a target \mathbf{t} sampled uniformly in the torus \mathbb{R}^n/A , we have a probability of $r^n \cdot (1 - n^{O(1)}r^n)$ that \mathbf{t} is at distance at most $r \text{GH}(n)$ from the lattice.*

Heuristic Justification. Contrary to the proof of Lemma 8, 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 $r \text{GH}(n)$ away, which by Heuristic Claim 1 happens with probability $O(n\sqrt{n})r^{2n}$.

The Contradiction. The idea now is that we end up at a contradiction whenever we instantiate the claim from above with the smallest possible r such that it is likely there is such a point among the T uniform samples.

For a given number of random samples T , we will pick σ for the BDD sample as before, i.e. such that $\ln T = N\varepsilon^2$ where $\varepsilon = \exp(-2\pi^2\sigma^2\ell^2)$. By the above Heuristic Claim 4, among those T targets, with constant probability there exists a target at distance at most $r \cdot \text{GH}(n)$ from the lattice, where $r = 1/\sqrt[3]{T}$. When for a given T , the length $r \text{GH}(n)$ is smaller than the expected length of the BDD sample, i.e. $\sqrt{n} \cdot \sigma$, this contradicts Heuristic Claim 3 and we say it is in the contradictory regime. This concrete contradictory regime is depicted in Figure 2. $\not\Leftarrow$

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: 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 T unaffected.

In the context of the cryptanalytic literature [EJK20, GJ21, MAT22], the set \mathcal{W} does not come from the full dual lattice $\Lambda_{\text{LWE}}^\vee \subset \mathbb{R}^n$. Instead, the full dual basis is first BKZ-reduced with blocksize β_{BKZ} , and then a sieve is run on the lattice $\Lambda' \subset \Lambda_{\text{LWE}}^\vee$ generated by the β_{sieve} first basis vectors of that BKZ-reduced basis. Effectively, this means that the dual distinguishing is not performed with respect to the whole LWE lattice Λ_{LWE} , but with respect to the projected sublattice Λ of it. Indeed, let $W \subset \mathbb{R}^n$ be the β_{sieve} -dimensional real vector space spanned by \mathcal{W} , and let π_W denote the orthogonal projection onto W . Then, for any $\mathbf{w} \in \mathcal{W}$, and any target $\mathbf{t} \in \mathbb{R}^n / \Lambda_{\text{LWE}}$, it holds that

$$\langle \mathbf{w}, \mathbf{t} \rangle = \langle \mathbf{w}, \pi_W(\mathbf{t}) \rangle. \quad (12)$$

Therefore, $f_{\mathcal{W}}(\mathbf{t}) = f_{\mathcal{W}}(\pi_W(\mathbf{t}))$ for any target $\mathbf{t} \in \mathbb{R}^n / \Lambda_{\text{LWE}}$. Note that $\pi_W(\mathbf{t})$ now lies in the β_{sieve} -dimensional torus $W / \pi_W(\Lambda_{\text{LWE}})$. Hence we are effectively running the dual-distinguishing here over a projected sublattice of dimension β_{sieve} .

In this scenario, the contradictory regime is solely determined by β_{sieve} and T , and not by other quantities such as the LWE parameters, 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} , but if the parameters are tuned optimally, we still have $\ln T \approx N\epsilon^2$ where $\epsilon = \exp(-2\pi^2\sigma^2\ell^2)$.

This might not perfectly be representative of the exact analysis of MATZOV [MAT22] 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 comes from a partial enumeration, though such a heuristic has been used in the past, for example underlying the analysis of the hybrid attack [HG07]. More critically, we see no mention of such dependence and how they would affect the algorithm in the existing analysis [EJK20, GJ21, MAT22]. While we do not claim that it is impossible, we view the notion that such dependences could fix the algorithm as quite doubtful, and requiring specific substantiation with analysis and experiments.

In other word, while our contradiction does not formally disprove the recent claims on the Dual-Sieve attack [EJK20, GJ21, MAT22], but it does invalidate the reasoning leading to these claims. And in the absence of 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 [GJ21] and [MAT22], focusing on the KYBER512 [ABD⁺19] parameter set, in the “asymptotic model” for dimensions for free.³

In [GJ21, Table 5], we find a sieving dimension $\beta_{\text{sieve}} = 396$ (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 $\{-3, \dots, 3\}$ 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.

In [MAT22, Table 3], we find a sieving dimension $\beta_{\text{sieve}} = 380$ (where all the dual vectors come from), a guessing dimension $k_{\text{enum}} = 19$, and an FFT dimension $k_{\text{fft}} = 34$. The guessing part enumerates over $\{-3, \dots, 3\}^{\text{enum}}$ in order of decreasing probability from the used binomial distribution, while the FFT is done with $p = 5$, giving rise to, according to [MAT22], $T = 2^{19 \cdot H(\chi_s)} \cdot 5^{34} \approx 2^{123.3}$ many targets. Using an improved cost metric, they also give [MAT22, Table 5] another set of parameter where $\beta_2 = 383$ and $T = 2^{17 \cdot H(\chi_s)} \cdot 5^{33} \approx 2^{116.3}$.⁴

For both instantiations [GJ21, MAT22] the dual attack is used rather deep in its contradictory regime, as depicted in Figure 2. \neq

Remark. At this point, we clarify that it would be a mistake to consider the analysis still valid whenever it is not in the contradictory regime. The existence of the contradictory regime shows a fundamental flaw in the Independence Heuristic, that may very well have an impact beyond the contradictory regime. If it does not, this should be thoroughly substantiated with analysis and experiments.

5 Experiments

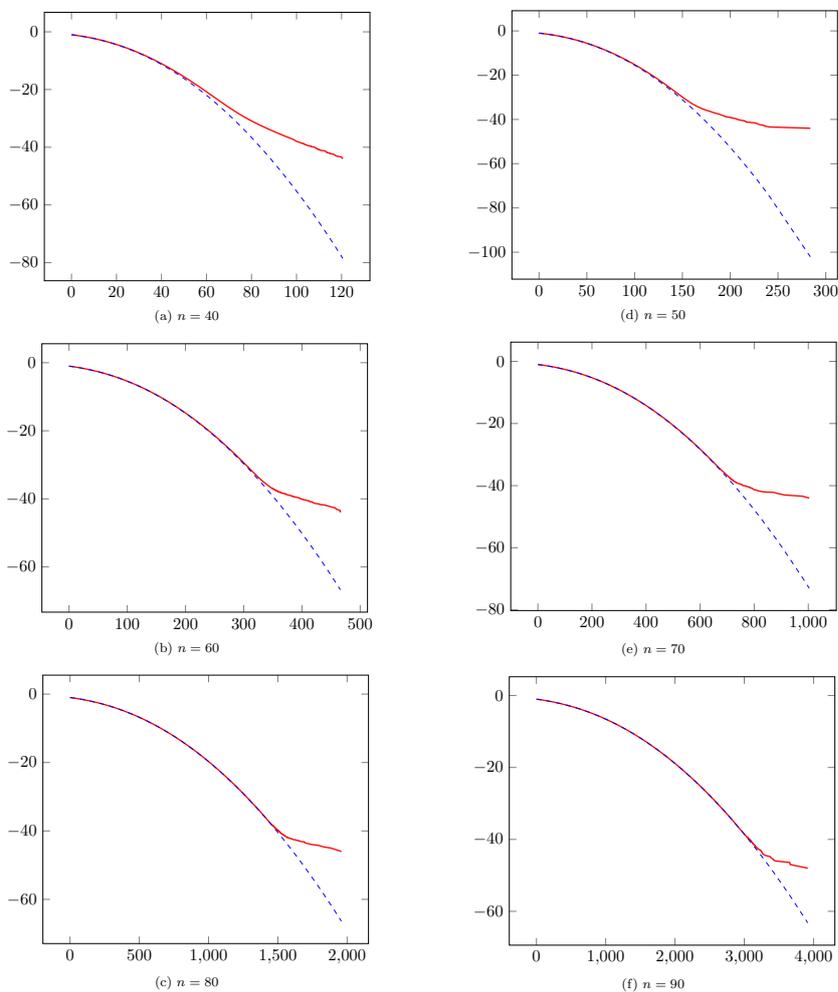
In this section, we provide further substantiation of the concrete contradiction (Sec. 4.2) with experimental evidence. We hope that the experiments will provide insight on what exactly goes wrong, and will show how the analysis can be fixed. We focus our analysis on the case where \mathcal{W} is the output of a full sieve, namely it contains all the $(4/3)^{n/2}$ vectors of length less than $\sqrt{4/3} \cdot \text{GH}(n)$.

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

Because the contradictory regime only kicks in for rather large values of T (say 2^{40} even in small dimension) these unpredicted high scores might be very rare and we are interested in the tail of that distribution. Naïvely, it would take a long time to run such large scale experiments, but the same FFT trick from [GJ21] (cf. Section 3) makes it feasible to run experiments on this scale!

³ This is the optimistic estimate in [Duc18]. The other “G6K model” used in [GJ21, MAT22] is debated in Appendix A.2.

⁴ We are not quite sure how this quantity was derived, and it seems incorrect. See Appendix A.5.



Legend: distribution of the scores on the x -axis and the logarithm (base 2) of the survival function (SF) on the y -axis. Dashed blue line: prediction from the heuristic analysis. Red line: experimental distribution. 2^{45} samples per curve.

Fig. 3. The distribution of scores according to the prediction and determined experimentally. The experimental data is obtained with `unif_scores.py` and listed in `data/unif_scores_nX.csv` of the auxiliary files.

5.1 Implementation Details

We used the G6K software [ADH⁺19] for running the experiments, using Python on a high-level but with a binding to some C 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 $\mathbf{B} \cdot \text{diag}(2, \dots, 2, 1, \dots, 1)$ 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 implicitly⁵.

In addition, the BDD scores are obtained with the script `bdd_scores.py`, which computed the score function for a BDD sample that samples from a gaussian of parameter $\sigma = \text{ghf} \text{GH}(n) / \sqrt{nq}$ where $\text{ghf} \in (0, 1)$ and \sqrt{q} is the normalization factor for the LWE lattice.

Lastly, the modulus switching scores are obtained with the script `mod_switch.py`, which performs the dual attack of [MAT22, Alg. 2]. In particular, it samples one random q -ary lattice, and then computes the score for targets that are sampled as in KYBER [ABD⁺19].

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 Figure 3. 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 Sections 4.2 and 4.3: we start encountering vectors that are quite close to the lattice, which should have a rather high score.

In this light, it seems insightful to compare the number of samples needed to reach the floor in practice, to the number of samples where the contradictory regime starts according to Section 4.3. For this, we first need to define precisely when the distribution enters the floor region: we consider the floor to start when the experimental SF of a score exceeds the predicted SF by an arbitrary factor of 2. Graphically, this corresponds in Figure 3 to a vertical gap of 1 unit between both curves. This comparison is depicted in Figure 4.

Conclusion. We notice in Figure 4 that in small dimensions, the floor begins quite earlier than the contradictory regime, vindicating the notion that the analysis might fail even in an earlier regime than our predicted contradiction. In larger dimension both curves appear to converge, but at this point, one *should not* extrapolate this behavior to higher dimensions without providing any theoretical justification. Furthermore, as we will see in the next section 5.3, this floor behaviour is not the only thing the analyses of [LW21, MAT22] mispredict.

⁵ cf. <https://github.com/fplll/fplll/wiki/FPLLL-Days-5-Summary>.

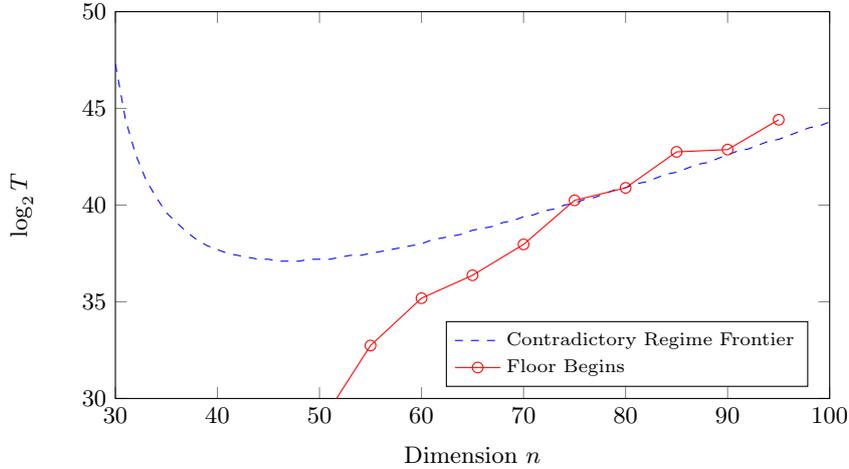


Fig. 4. Comparison of the Floor of the score of uniform samples with the Frontier of the contradictory regime of Section 4.3.

5.3 Distribution of Scores of BDD Targets

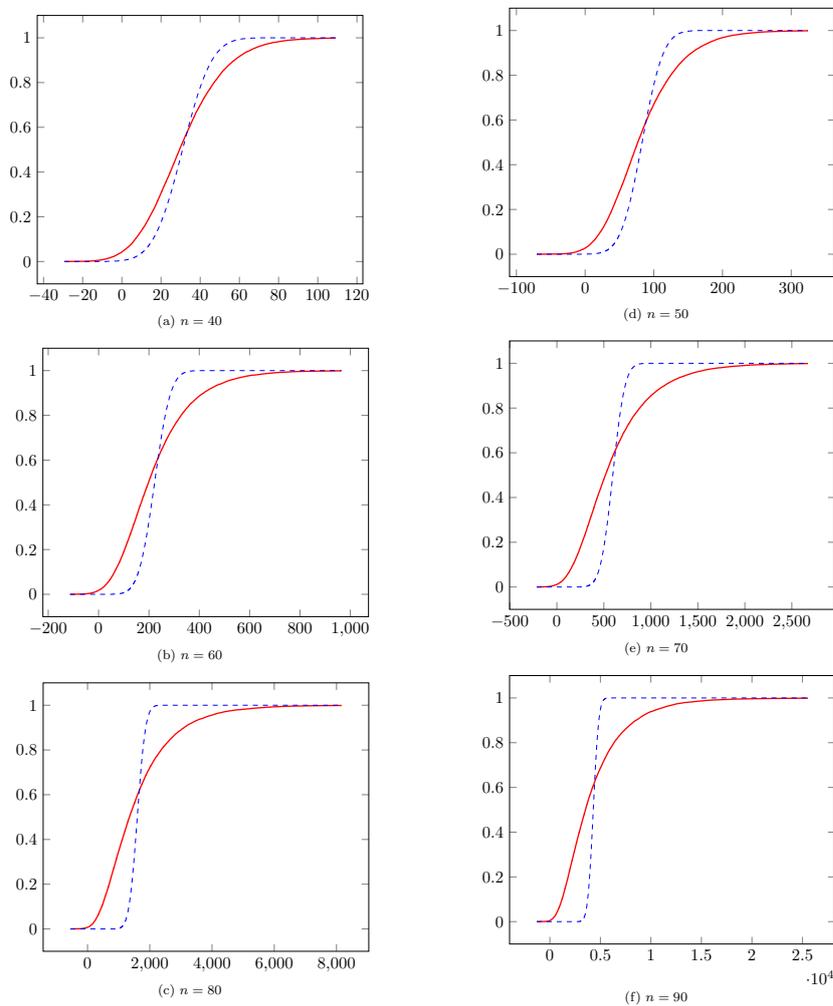
We measured the score distribution for BDD targets sampled from a gaussian of parameter $\sigma = 0.7 \cdot \text{GH}(n) / \sqrt{n}$, over lattices of various dimensions n , and plotted our result in Figure 3. The predicted score distribution is based on the Independence Heuristic, but also takes into account the exact lengths of each dual vector in Equation (8), instead of approximating all the lengths to be equal to $\sqrt{4/3} \cdot \text{GH}(n)$, 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 exponentially 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 Independence Heuristic. And indeed, the prediction is close to the measured average.

A more interesting statistic is the median. According to the standard analysis [LW21, MAT22] reconstructed in Section 2.3, 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 standard analysis [LW21, MAT22], reconstructed in Section 2.3 of the dual attack is completely off when it comes to predicting the score of BDD targets. The distribution is definitely *not* gaussian, nor even symmetric around its average, and its variance is hugely underestimated.



Legend: Cumulative distribution function of the score for gaussian BDD samples, with parameter $\sigma = 0.7 \cdot \text{GH}(n) / \sqrt{n}$. Probability (y -axis) is given on a linear scale. Dashed blue line: prediction from the Heuristic Analysis. Red line: experimental. 2^{15} samples per curve.

Fig. 5. The distribution of scores according to the prediction and the distribution determined experimentally. The experimental data is obtained with `bdd_scores.py` and listed in `data/bdd_scores_nX.csv` of the auxiliary files.

Dimension n	40	50	60	70	80	90
predicted std. dev.	11.77	24.33	50.10	103.0	211.5	434.2
measured std. dev.	20.07	53.89	148.0	412.9	1159	3313
ratio meas./pred.	1.705	2.215	2.954	4.009	5.480	7.630
predicted average	30.97	83.02	222.50	595.11	1598.6	4295.1
measured average	31.10	83.94	223.75	600.91	1611.3	4341.4
ratio meas./pred.	1.00	1.01	1.01	1.01	1.01	1.01
measured average	31.10	83.94	223.75	600.91	1611.3	4341.4
measured median	29.43	76.59	198.21	515.87	1345.2	3521.3
ratio med./avg.	0.95	0.92	0.89	0.87	0.84	0.82

Table 1. Variance of the BDD score distribution

5.4 Distribution of Scores, with Modulus Switching

Lastly, we ran experiments on the score distribution when using modulus switching in [MAT22, Alg. 2]. We compare these experiments to the given bounds: a lower bound for the average [MAT22, Lem. 5.4 and 5.5] and an upper bound on the variance [MAT22, Lem. 5.7]. We also test asymmetry by comparing the average and the median.

p	3	4	5	6	7
std. dev. upper bound	4317	4317	4317	4317	4317
measured	$2.53 \cdot 10^5$	$7.42 \cdot 10^5$	$1.19 \cdot 10^6$	$1.42 \cdot 10^6$	$1.47 \cdot 10^6$
ratio meas./u.b.	58.60	171.9	275.7	328.9	340.5
average lower bound	$8.08 \cdot 10^4$	$4.26 \cdot 10^5$	$1.61 \cdot 10^6$	$3.15 \cdot 10^6$	$4.66 \cdot 10^6$
measured average	$9.46 \cdot 10^4$	$6.81 \cdot 10^5$	$1.89 \cdot 10^6$	$3.36 \cdot 10^6$	$4.79 \cdot 10^6$
ratio meas./l.b.	1.17	1.60	1.17	1.07	1.03
measured average	$9.46 \cdot 10^4$	$6.81 \cdot 10^5$	$1.89 \cdot 10^6$	$3.36 \cdot 10^6$	$4.79 \cdot 10^6$
measured median	$6.84 \cdot 10^3$	$4.32 \cdot 10^5$	$1.57 \cdot 10^6$	$3.06 \cdot 10^6$	$4.54 \cdot 10^6$
ratio med./avg.	0.0723	0.63	0.83	0.91	0.95

Table 2. Predicted vs. measured average, standard deviation and median for scores using modulus switching. Used parameters: $k_{\text{enum}} = 0$, $k_{\text{fft}} = 20$, $k_{\text{lat}} = 45$, $q = 3329$, $\eta = 3$ and 100 samples. A sieve ran in dimension 110 giving 13,393,776 dual vectors.

As in the case of BDD target without modulus switching (Table 1), we note that the standard deviation is significantly underestimated: the given upper

bound is significantly violated. Additionally, the upper bound is constant as a function of p , but we see that the standard deviation increases with p for the values listed in Table 2.

Regarding asymmetry, we see the experimental median is a bit lower than the average for $p \geq 4$, displaying the similar asymmetry than without modulus switching. The case $p = 3$ stands out with an extreme asymmetry: here the median is more than ten times smaller than the average. While the case $p = 3$ might not be of interest in practice, it appears as a good test case for a robust fix of the prediction. One may further wonder whether this phenomenon for $p = 3$ might extend to larger p when changing other parameters.

Conclusion. The prediction for the distribution of scores of BDD targets with modulus switching seems to naturally inherit the issues of the same prediction without modulus switching (i.e. unpredicted asymmetry, underestimated variance), but seems to also raise new ones: an unpredicted growth of variance as p increases, and a very extreme case of asymmetry for certain choices of p .

6 Afterthoughts

As shown in the experiments, the heuristic analysis underlying the dual attack does not match practice. In order to get a precise cost on the dual attack, the analysis has to be fixed, in a way that correctly predicts both distribution of scores, for uniform and for BDD targets.

In this section, we propose a possible explanation of what is happening, while mentioning a similar phenomenon in coding theory. Ultimately, there are some fixes that need to be made to the dual attack and we mention some pitfalls to look out for when fixing the attack.

6.1 A Similar Result from Coding Theory

The *waterfall-floor* phenomenon visible in Figure 4 is something coding theory has encountered before in a rather similar context, namely the error failure probability when decoding Low-Density Parity-Check (LDPC) codes [Ric06, VCN14, ABH⁺22]. We recall that the principle of LDPC decoding is to exploit the low-weight from the parity check matrix, or, said otherwise, the shortness of some vectors of the dual code. The analogy is striking.

6.2 The Origin of Correlation

Formally, the quantities $\langle \mathbf{w}, \mathbf{e} \rangle \bmod 1$ for $\mathbf{w} \in \mathcal{W}$ can only be independent if the set \mathcal{W} is linearly independent. Yet, it appears in Figure 6(a) that the “length” of the linear relation matters when it comes to the impact of such dependencies on the total score $f_{\mathcal{W}}$. Indeed, we see that a long relation $\mathbf{w}_3 = 5\mathbf{w}_1 + 7\mathbf{w}_2$ is not

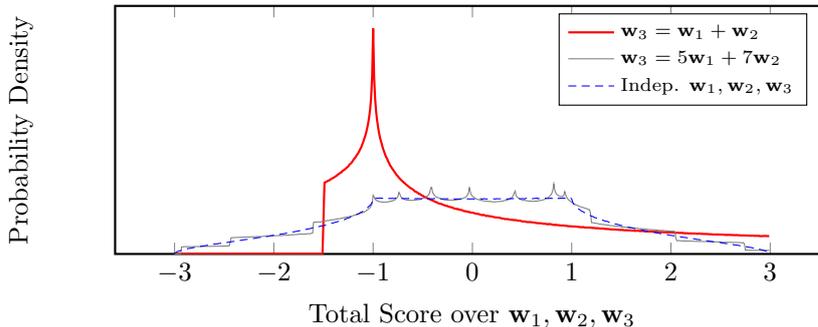


Fig. 6. Distribution of the total score $f_{\mathcal{W}}(\mathbf{t})$ where $|\mathcal{W}| = 3$, for a uniform $\mathbf{t} \bmod A$ when the \mathbf{w} 's are linearly independent versus related. Measured over 2^{25} samples.

that far off from the linearly independent case. On the contrary, a short relation $\mathbf{w}_3 = \mathbf{w}_1 + \mathbf{w}_2$ changes that distribution severely.

Since \mathcal{W} is taken as the output of a sieve [NV08, MV10, BDGL16], i.e. the set of the $(4/3)^{n/2}$ shortest vectors, we get many short relations. Namely, we expect a constant fraction of vectors to belong to a triple of the form $\mathbf{w}_3 = \pm \mathbf{w}_1 \pm \mathbf{w}_2$. For $k \geq 4$, each vector will belong to an exponential amount of such k -tuple $\mathbf{w}_k = \pm \mathbf{w}_1 \pm \mathbf{w}_2 \pm \dots \pm \mathbf{w}_{k-1}$, the precise quantity of such tuples should be analyzable following the theory underlying tuple-sieves [BLS16, HK17].

We clarify that this discussion and experiments are only meant to illustrate where the correlations come from. Fixing the analysis via this angle would require understanding the much harder question of how these correlations compound, and we are doubtful if this would be a tractable route.

6.3 Is the Dual Attack Fixable?

The theoretical analysis of Section 4 and the experiments of Section 5 unequivocally invalidate the standard analysis of the Dual-Sieve attack (with or without FFT) as found in [LW21, GJ21, MAT22, AS22, CST22]. In the context of the Dual-Sieve-FFT attack, as instantiated in [GJ21, MAT22, AS22, CST22], our work point out a presumably large number of false positive, i.e. incorrect answers having a higher score than that of the desired target.

However, if this number of false positives in the current parametrization is reasonable, all might not be lost. Indeed, one may consider the Dual-Sieve-FFT technique as a first filtering stage in an attack with multiple stages; the leftover 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 the leftover problem becomes sufficiently easier to accommodate all these targets, the Dual-Sieve-FFT attack might be salvaged.

Nevertheless, to substantiate such a fix, one must first produce a new convincing model for the score distribution of both BDD and random targets. This should be backed both by theoretical arguments, and by experimental validation.

We insist that pulling the parameters of the attack outside of the contradictory regime is not a convincing way of substantiating a fix. We need a sound analysis that precisely predicts the score behavior, and it should specifically be tested on the regime where the prior analysis made mispredictions. That is, we should have theoretically justified predictions that match with the experimental behavior measured in Figures 3 and 5 and beyond.

In particular, we warn against a flawed argument for a fix, that would consist of constructing the set $\mathcal{W} = \mathcal{W}_1 \cup \mathcal{W}_2$ from two BKZ reductions and sieves instead of just one. One might argue that the dual distinguishing is now happening in a lattice of dimension 2β instead of β , which would push points of Figure 2 outside of the contradictory regime. But considering the experiments of Section 5.2 in Figure 3 we should see that this does not really fix the issue. Each half $f_{\mathcal{W}_i}$ of the score distribution $f_{\mathcal{W}}$ is going to hit its floor, and the sum of those two distributions will also have a floor at essentially the same height. Indeed, it would suffice to hit the floor of one of the functions, to hit the floor of the aggregate.

A more credible approach however is indeed to run two (or a few) BKZ reduction and sieves, obtaining two sets of short dual vectors $\mathcal{W}_1, \mathcal{W}_2$ then considering the aggregate score as the *minimum* of both scores $f' = \min(f_{\mathcal{W}_1}, f_{\mathcal{W}_2})$ rather than its sum. To hit the floor of this new aggregate, a uniform sample 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), that should be much more unlikely than hitting either floor.

Note however that taking such a minimum aggregate of scores might also amplify the issues with low scores for BDD targets observed in Section 5.3. A robust model for all the distributions at hand is therefore still required. Also note that taking the smallest of both scores is conceptually not that far off from the prior fix idea of first filtering with $f_{\mathcal{W}_1}$ and then filtering the survivors again with a new stage of the attack.

Open Question 2 *Convincingly fix the analysis of the Dual-Sieve-FFT attack with robust predictions for both score distributions, matching experiments from Figure 3 and 5 and beyond. Then, consider if one of the suggested fixes above allows to recover a complexity close to that of the original claims [LW21, GJ21, MAT22, AS22, CST22].*

We further invite future work to fix other minor oddities, listed in Appendix A.

References

- ABD⁺19. Roberto Avanzi, Joppe Bos, Léo Ducas, Eike Kiltz, Tancrede Lepoint, Vadim Lyubashevsky, John M Schanck, Peter Schwabe, Gregor Seiler, and Damien Stehlé. Crystals-kyber algorithm specifications and supporting documentation. *NIST PQC Round*, 2(4):1–43, 2019.
- ABH⁺22. Sarah Arpin, Tyler Raven Billingsley, Daniel Rayor Hast, Jun Bo Lau, Ray Perlner, and Angela Robinson. A study of error floor behavior in QC-MDPC codes. In *Post-Quantum Cryptography: 13th International Workshop, PQCrypto 2022, Virtual Event, September 28–30, 2022, Proceedings*, pages 89–103. Springer, 2022.

- AD21. Martin Albrecht and Léo Ducas. Lattice attacks on NTRU and LWE: a history of refinements. *Cryptology ePrint Archive*, 2021.
- ADH⁺19. Martin R. Albrecht, Léo Ducas, Gottfried Herold, Elena Kirshanova, Eamonn W. Postlethwaite, and Marc Stevens. The general sieve kernel and new records in lattice reduction. In Yuval Ishai and Vincent Rijmen, editors, *Advances in Cryptology – EUROCRYPT 2019, Part II*, volume 11477 of *Lecture Notes in Computer Science*, pages 717–746, Darmstadt, Germany, May 19–23, 2019. Springer, Heidelberg, Germany.
- ADPS16. Erdem Alkim, Léo Ducas, Thomas Pöppelmann, and Peter Schwabe. Post-quantum key exchange - A New Hope. In *USENIX security symposium*, volume 2016, 2016.
- AGPS20. Martin R Albrecht, Vlad Gheorghiu, Eamonn W Postlethwaite, and John M Schanck. Estimating quantum speedups for lattice sieves. In *Advances in Cryptology–ASIACRYPT 2020: 26th International Conference on the Theory and Application of Cryptology and Information Security, Daejeon, South Korea, December 7–11, 2020, Proceedings, Part II 26*, pages 583–613. Springer, 2020.
- Alb17. Martin R Albrecht. On dual lattice attacks against small-secret LWE and parameter choices in HELib and SEAL. In *Advances in Cryptology–EUROCRYPT 2017: 36th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Paris, France, April 30–May 4, 2017, Proceedings, Part II*, pages 103–129. Springer, 2017.
- AR04. Dorit Aharonov and Oded Regev. Lattice problems in NP cap coNP. In *45th Annual Symposium on Foundations of Computer Science*, pages 362–371, Rome, Italy, October 17–19, 2004. IEEE Computer Society Press.
- AS64. Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. U.S. Government Printing Office, 1964.
- AS22. Martin R Albrecht and Yixin Shen. Quantum augmented dual attack. *arXiv preprint arXiv:2205.13983*, 2022.
- BDGL16. Anja Becker, Léo Ducas, Nicolas Gama, and Thijs Laarhoven. New directions in nearest neighbor searching with applications to lattice sieving. In Robert Krauthgamer, editor, *27th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 10–24, Arlington, VA, USA, January 10–12, 2016. ACM-SIAM.
- BLS16. Shi Bai, Thijs Laarhoven, and Damien Stehlé. Tuple lattice sieving. *LMS Journal of Computation and Mathematics*, 19(A):146–162, 2016.
- BM23. Alessandro Budroni and Erik Mårtensson. Improved estimation of key enumeration with applications to solving lwe. *Cryptology ePrint Archive*, Paper 2023/139, 2023. <https://eprint.iacr.org/2023/139>.
- CST22. Kevin Carrier, Yixin Shen, and Jean-Pierre Tillich. Faster dual lattice attacks by using coding theory. *Cryptology ePrint Archive*, Paper 2022/1750, 2022.
- DADRT22. Thomas Debris-Alazard, Léo Ducas, Nicolas Resch, and Jean-Pierre Tillich. Smoothing codes and lattices: Systematic study and new bounds. *Cryptology ePrint Archive*, Paper 2022/615, 2022.
- dt23. The FPLLL development team. `fpyll`, a Python wrapper for the `fplll` lattice reduction library, Version: 0.5.9. Available at <https://github.com/fplll/fpylll>, 2023.

- Duc18. Léo Ducas. Shortest vector from lattice sieving: A few dimensions for free. In Jesper Buus Nielsen and Vincent Rijmen, editors, *Advances in Cryptology – EUROCRYPT 2018, Part I*, volume 10820 of *Lecture Notes in Computer Science*, pages 125–145, Tel Aviv, Israel, April 29 – May 3, 2018. Springer, Heidelberg, Germany.
- DV90. Pierre Duhamel and Martin Vetterli. Fast Fourier transforms: a tutorial review and a state of the art. *Signal processing*, 19(4):259–299, 1990.
- EJK20. Thomas Espitau, Antoine Joux, and Natalia Kharchenko. On a hybrid approach to solve small secret LWE. *Cryptology ePrint Archive*, 2020.
- Gal62. Robert Gallager. Low-density parity-check codes. *IRE Transactions on information theory*, 8(1):21–28, 1962.
- GJ21. Qian Guo and Thomas Johansson. Faster dual lattice attacks for solving LWE with applications to CRYSTALS. In Mehdi Tibouchi and Huaxiong Wang, editors, *Advances in Cryptology – ASIACRYPT 2021, Part IV*, volume 13093 of *Lecture Notes in Computer Science*, pages 33–62, Singapore, December 6–10, 2021. Springer, Heidelberg, Germany.
- Hås88. Johan Håstad. Dual vectors and lower bounds for the nearest lattice point problem. *Combinatorica*, 8(1):75–81, 1988.
- HG07. Nick Howgrave-Graham. A hybrid lattice-reduction and meet-in-the-middle attack against ntru. In *Advances in Cryptology-CRYPTO 2007: 27th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2007. Proceedings 27*, pages 150–169. Springer, 2007.
- HK17. Gottfried Herold and Elena Kirshanova. Improved algorithms for the approximate k-list problem in euclidean norm. In *IACR International Workshop on Public Key Cryptography*, pages 16–40. Springer, 2017.
- Jab01. A Al Jabri. A statistical decoding algorithm for general linear block codes. In *IMA International Conference on Cryptography and Coding*, pages 1–8. Springer, 2001.
- LF06. Éric Leveil and Pierre-Alain Fouque. An improved LPN algorithm. In *International conference on security and cryptography for networks*, pages 348–359. Springer, 2006.
- LW21. Thijs Laarhoven and Michael Walter. Dual lattice attacks for closest vector problems (with preprocessing). In Kenneth G. Paterson, editor, *Topics in Cryptology – CT-RSA 2021*, volume 12704 of *Lecture Notes in Computer Science*, pages 478–502, Virtual Event, May 17–20, 2021. Springer, Heidelberg, Germany.
- MAT22. MATZOV. Report on the security of LWE: Improved dual lattice attack, April 2022.
- Mic14. Daniele Micciancio. Lattice algorithms and applications, lecture 3: The dual lattice. Available at <https://cseweb.ucsd.edu/classes/sp14/cse206A-a/lec3.pdf>, 2014.
- MR09. Daniele Micciancio and Oded Regev. Lattice-based cryptography. *Post-quantum cryptography*, pages 147–191, 2009.
- MV10. Daniele Micciancio and Panagiotis Voulgaris. Faster exponential time algorithms for the shortest vector problem. In Moses Charika, editor, *21st Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1468–1480. ACM-SIAM, January 2010.
- NP33. Jerzy Neyman and Egon Sharpe Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(694-706):289–337, 1933.

- NV08. Phong Q Nguyen and Thomas Vidick. Sieve algorithms for the shortest vector problem are practical. *Journal of Mathematical Cryptology*, 2(2):181–207, 2008.
- Ove06. Raphael Overbeck. Statistical decoding revisited. In *Australasian Conference on Information Security and Privacy*, pages 283–294. Springer, 2006.
- Rad68. Charles M Rader. Discrete Fourier transforms when the number of data samples is prime. *Proceedings of the IEEE*, 56(6):1107–1108, 1968.
- Ric06. Gerd Richter. Finding small stopping sets in the Tanner graphs of LDPC codes. In *4th International Symposium on Turbo Codes & Related Topics; 6th International ITG-Conference on Source and Channel Coding*, pages 1–5. VDE, 2006.
- VCN14. Bane Vasić, Shashi Kiran Chilappagari, and Dung Viet Nguyen. Failures and error floors of iterative decoders. In *Channel Coding: Theory, Algorithms, and Applications: Academic Press Library in Mobile and Wireless Communications*, pages 299–341. Elsevier Inc., 2014.

A Minor Oddities from the Literature

A.1 Length of Vectors After BKZ and Sieving

The work of Guo and Johansson [GJ21, Sec. 6.2] seems to assume that after running BKZ of blocksize β_{BKZ} and then a sieve of dimension β_{sieve} of the first block $\mathbf{B}_{[0, \beta_{\text{sieve}}]}$, then, one should obtain $(4/3)^{\beta_{\text{sieve}}/2}$ vectors of length $\sqrt{4/3}$ times the predicted length of \mathbf{b}_1 . The work of MATZOV [MAT22, Lem. 4.2] makes exactly the same assumption.

Such a claim was made in [ADPS16] in the special case $\beta_{\text{BKZ}} = \beta_{\text{sieve}}$ and can be justified using the usual heuristics (including the Geometric Series Assumption, see [AD21] for example). On the contrary, the same heuristic would lead to a different conclusion when $\beta_{\text{BKZ}} \neq \beta_{\text{sieve}}$, as used in [GJ21, MAT22]. Indeed, this should involve the Gaussian Heuristic of the whole block $\mathbf{B}_{[0, \beta_{\text{sieve}}]}$, and only coincide with the predicted length of \mathbf{b}_1 when $\beta_{\text{BKZ}} = \beta_{\text{sieve}}$.

Given that β_{BKZ} is in practice close to β_{sieve} , the induced error on prediction remains most likely mild.

A.2 The So-Called “G6K model” for Dimensions for Free

What is called the “G6K model” in [GJ21, MAT22] is asymptotically incorrect, and was only meant as a local parametrization in the G6K implementation in [ADH⁺19]. The work of [ADH⁺19] explicitly mark these best linear fits as “unreliable for extrapolation”.

Furthermore, the extra “dimensions for free” obtained in [ADH⁺19] compared to the prior work [Duc18] come from a technique called “on the fly lifting”, and are not really for free: they save space, but barely any time [ADH⁺19, App. A]. Furthermore, the practical cheapness of “on the fly lifting” is rather tied to the specific architecture at hand, and it is dubious whether it would transfer so cheaply in the gate count cost model targeted in [GJ21, MAT22] via the software

of [AGPS20]. It is a matter of relative cost between the two main operations: floating points linear algebra, and `xor-popcounts`; modern architectures are heavily focused on the former, but the latter is to be preferred in the gate-count metric. That is, modern processors will process much more gates per cycle when doing linear algebra than when counting bits. And on the fly lifting is essentially floating point linear algebra.

A.3 Geometric Series Assumption vs. BKZ Simulation

We note that the claims of [GJ21, MAT22] are based on modeling basis reduction according to the geometric series assumption (GSA), but it is costed using progressive sieving [ADH⁺19]. It has been pointed out on the NIST PQC-forum⁶ that this leads to noticeable cost underestimation.

While such a simplification might be a reasonable considered in isolation, one should be more careful when comparing concrete claims. In particular the analysis of the primal attack in the KYBER standardization document [ABD⁺19] does use the more accurate simulation method. The cost comparison given in [GJ21, MAT22] are therefore not apple-to-apple.

A.4 Real Scores

Note that we use $\cos(2\pi \langle \mathbf{w}, \mathbf{t} \rangle)$ on line 10 of Algorithm 1, and not the complex character $\chi_{\mathbf{w}}(\mathbf{t})$ because we can always make the set \mathcal{W} symmetric. Indeed we have the identity $2 \cos(2\pi \langle \mathbf{w}, \mathbf{t} \rangle) = \chi_{\mathbf{w}}(\mathbf{t}) + \chi_{-\mathbf{w}}(\mathbf{t})$. It is customary in lattice sieve algorithms [NV08, MV10, ADH⁺19] that a vector and its opposite is only represented once in the database \mathcal{W} . But this is merely a compression trick, and the output should read as implicitly containing both.

In this light, we find ourselves quite confused by the attempt in [MAT22, Sec 5.2] to exploit the phase of the character as part of the score, supposedly inducing a reduction factor of $D_{\text{arg}} \approx \frac{1}{2}$ on the required number of dual vectors.

A.5 Expected Number of Guesses for s_{enum}

In the analysis of MATZOV [MAT22, Thm. 5.1] it seems to be assumed that enumerating possibilities in decreasing order of probability leads to guessing the correct s_{enum} after an average of $2^{H(X)}$ attempts, where $H(X)$ denotes the entropy of the distribution X at hand. There is no justification for this claim, and it appears to be false.

For example, the geometric distribution G_p of parameter p , whose probabilities are in decreasing order by construction, we have a mean of $1/p$, which does not coincide with the exponential of its entropy,

$$H(G_p) = -\frac{(1-p) \log_2(1-p) + p \log_2(p)}{p}.$$

⁶ <https://groups.google.com/a/list.nist.gov/g/pqc-forum/c/Fm4cDfsx65s/m/BZFRc8hiAAAAJ>

The uniform distribution over a set of size s is another counter-example, with an average guessing time of $(s + 1)/2$, but entropy $\log_2 s$.

In the two counter-examples above, the quantity $2^{H(X)}$ is an overestimate for the average guessing time. But it appears to be an underestimate in the case of the KYBER parameters [ABD⁺19]. More specifically, with $k_{\text{enum}} = 17$ and the distribution as in KYBER512, we numerically computed an entropy of 39.66 bits, but found a guessing time of $2^{39.86}$ attempts using a script `guess_time.py`.

This underestimation factor of $2^{0.2}$ may not be that big, but this factor becomes more significant for larger parameter sets. Our script suggests that the gap between these estimations keeps increasing for larger k_{enum} . In particular, for KYBER1024 ($\eta = 2$), their attack cost has an underestimation factor of $2^{2.03}$, which is note-worthy considering the level of detail of the attack costs in [MAT22].

This issue has already been noted by follow-up work [AS22], and specifically studied in [BM23].

A.6 FFT in Guo–Johansson vs. FFT in MATZOV

At last, we would like to comment on the perceived superiority of the variant of MATZOV [MAT22] over that of Guo–Johansson [GJ21]. Indeed, the former claims better complexity, but it is unclear whether this is really due to modulus switching, given other orthogonal differences in both analysis.

In particular, the used enumeration strategies, which is rather an orthogonal question, differ significantly. That of Guo–Johansson is essentially unpruned: it spends equal effort on the most and less likely candidates, and could easily be improved with the enumeration strategy of MATZOV. This could equally explain the gap in performance. Other optimization efforts may differ, as well as the small oddities mentioned in this Appendix. Furthermore, the work of MATZOV [MAT22] leads to a correction in the gate cost estimate for sieving [AGPS20]; the same estimation was used before that correction in [GJ21].

We, however do not have the means to investigate these potential differences: some data are not available, and the underlying scripts have not been made public. And more crucially, this would also divert us from the higher level purpose of this paper.