Unified Dual Attack Analyses: Covariance-Based Score Distribution Prediction for LWE

Yechen Li[1] and Qunxiong Zheng[1(✉)]

Information Engineering University, Zhengzhou, 450001, Henan, China
qunxiong_zheng@163.com

Abstract. The Learning with Errors (LWE) problem underpins many post-quantum cryptosystems, including the NIST-selected CRYSTALS-KYBER and CRYSTALS-DILITHIUM. Recent dual attacks have demonstrated remarkable effectiveness against concrete LWE-based schemes, with some claims suggesting that the security of CRYSTALS-KYBER may be reduced below the NIST threshold. However, the analysis of the score distribution for the correct guess in dual attacks has consistently relied on a flawed independence assumption, leading to variance estimates that are far smaller than the true score variance. This issue has been highlighted in several studies. For instance, Bashiri and Wiemers (JMC 2025) proposed an estimate of the variance, yet our experiments reveal that their approach performs poorly in medium-to-high dimensions. On the other hand, many works have characterized the success probability of dual attacks, but these are either based on the BDD problem or limited to specific attack types, lacking a unified analytical framework for dual attacks on LWE. These theoretical gaps motivate us to develop a unified estimation of the expectation and variance of the score in dual attacks.

In this paper, we propose a unified predictive model for the expectation and variance of the score, covering three types of dual attacks: the original dual attack, the dual attack with modulus switching, and the dual attack with decoding (Crypto 2025). Our key observation is that the cosine of the angle between different short vectors follows a normal distribution, which we use to estimate the covariance between individual scores. By decomposing the score expression into a combination of simple distributions, we obtain estimates for the expectation and variance of individual scores, and combine these with the covariance to derive closed-form estimates for the total variance. Experiments show that our estimates achieve high accuracy and outperform previous work in medium-to-high dimensions. We also extend the prediction method of Ducas and Pulles (JOC 2026) for the score of incorrect guesses to a more general setting and, together with our predictive model, provide an extended characterization of the tail behavior.

Keywords: Lattice-based Cryptography · Learning with Errors · Dual Attack · Covariance.

1 Introduction

1.1 Background

Recent years have seen sustained interest in Learning With Errors (LWE), introduced by Regev [27], due to its central role in lattice-based cryptography. In particular, the security

of many post-quantum schemes relies on the hardness of LWE, notably CRYSTALS-KYBER [29] and CRYSTALS-DILITHIUM [23], which were standardized by NIST. The LWE problem is parameterized by m, n, q and distributions χ_s, χ_e over $\mathbb{Z}_q$. Given $\mathbf{A} \leftarrow \mathbb{Z}_q^{m\times n}$ and $\mathbf{b} = \mathbf{As} + \mathbf{e} \bmod q$, where secret vector $\mathbf{s} \leftarrow \chi_s^n$ and error vector $\mathbf{e} \leftarrow \chi_e^m$, the task is to recover a valid short $\mathbf{s}$ or $\mathbf{e}$. Equivalently, LWE can be viewed as a Bounded Distance Decoding (BDD) problem over the corresponding q-ary lattice

$$\Lambda_q(\mathbf{A}) \triangleq \left\{\mathbf{Ax} \bmod q \mid \mathbf{x} \in \mathbb{Z}_q^n\right\}.$$

Dual attacks. The central role of LWE in lattice-based cryptography has motivated extensive research on it, leading to several classes of attacks, including primal attacks [18,19], dual attacks [10,17,24,25], algebraic attacks [7], and machine learning attacks [22]. Among these, dual attack [10] has emerged as one of the most effective approaches in practice. Dual attack was first proposed by Micciancio and Regev [25], with the underlying idea tracing back to parity-check codes [15]. Its goal is to use short vectors in the dual lattice to distinguish points that are close to the lattice from those that are far away. Concretely, given a q-ary lattice $\Lambda_q(\mathbf{A})$ with $\mathbf{A} \in \mathbb{Z}_q^{m\times n}$ and a vector $\mathbf{v} \in \mathbb{Z}_q^m$, the attack finds a short vector $\mathbf{w}$ in the dual lattice $\Lambda_q(\mathbf{A})^\vee = \frac{1}{q}\Lambda_q^\perp(\mathbf{A})$ and checks the inner product $\langle \mathbf{w}, \mathbf{v}\rangle$, where

$$\Lambda_q^\perp(\mathbf{A}) \triangleq \left\{\mathbf{x} \in \mathbb{Z}_q^m \mid \mathbf{A}^\top\mathbf{x} \equiv \mathbf{0} \bmod q\right\}.$$

When $\mathbf{v}$ is very close to the lattice (e.g., an LWE instance), the inner product $\langle \mathbf{w}, \mathbf{v}\rangle$ is close to an integer. In contrast, if $\mathbf{v}$ is a random vector, $\langle \mathbf{w}, \mathbf{v}\rangle$ mod 1 is uniformly distributed over the discrete set $\{0, 1/q, \ldots, (q-1)/q\}$. In Eurocrypt 2017, this foundational idea was developed by Albrecht [2]. For an LWE instance $(\mathbf{A} \in \mathbb{Z}_q^{m\times n}, \mathbf{b} = \mathbf{As}+\mathbf{e} \bmod q)$, the author considered finding short vectors $\mathbf{w} = (\mathbf{x}^\top, \mathbf{y}^\top)^\top$ in the lattice

$$\Lambda = \left\{(\mathbf{x}^\top, \mathbf{y}^\top)^\top \in \mathbb{Z}_q^m \times \mathbb{Z}_q^n \mid \mathbf{A}^\top\mathbf{x} \equiv \mathbf{y} \bmod q\right\}.$$

Then the inner product becomes $\langle \mathbf{x}, \mathbf{b}\rangle = \langle \mathbf{x}, \mathbf{e}\rangle + \langle \mathbf{y}, \mathbf{s}\rangle \bmod q$, which is biased toward 0 since $\mathbf{s}$ and $\mathbf{e}$ are short. Furthermore, Guo and Johansson [17] proposed splitting the matrix $\mathbf{A} \in \mathbb{Z}_q^{m\times n}$ into three blocks and using the Fast Fourier Transform (FFT) to accelerate the process of guessing the secret vector. In detail, the matrix $\mathbf{A} \in \mathbb{Z}_q^{m\times n}$ is partitioned as $[\mathbf{A}_{\text{lat}}, \mathbf{A}_{\text{enu}}, \mathbf{A}_{\text{fft}}]$ with

$$\mathbf{A}_{\text{lat}} \in \mathbb{Z}_q^{m\times n_{\text{lat}}}, \quad \mathbf{A}_{\text{enu}} \in \mathbb{Z}_q^{m\times n_{\text{enu}}}, \quad \mathbf{A}_{\text{fft}} \in \mathbb{Z}_q^{m\times n_{\text{fft}}},$$

where $n_{\text{lat}} + n_{\text{enu}} + n_{\text{fft}} = n$. Correspondingly, the secret vector is decomposed as $\mathbf{s}^\top = (\mathbf{s}_{\text{lat}}^\top, \mathbf{s}_{\text{enu}}^\top, \mathbf{s}_{\text{fft}}^\top)$. The lattice

$$\Lambda_{\text{lat}} = \left\{(\mathbf{x}^\top, \mathbf{y}^\top)^\top \in \mathbb{Z}_q^m \times \mathbb{Z}_q^{n_{\text{lat}}} \mid \mathbf{A}_{\text{lat}}^\top\mathbf{x} \equiv \mathbf{y} \bmod q\right\}$$

which is exactly of the form discussed earlier, is then considered. Next, apply lattice basis reduction (e.g., the BKZ algorithm [11]) followed by a sieve algorithm [3] to this lattice to yield a list $\mathcal{W}$ of short vectors. Recalling that $\mathbf{b} = \mathbf{A}_{\text{lat}}\mathbf{s}_{\text{lat}} + \mathbf{A}_{\text{enu}}\mathbf{s}_{\text{enu}} + \mathbf{A}_{\text{fft}}\mathbf{s}_{\text{fft}} + \mathbf{e} \bmod q$, for each vector $\mathbf{w}^\top = (\mathbf{x}^\top, \mathbf{y}^\top) \in \mathbb{Z}_q^m \times \mathbb{Z}_q^{n_{\text{lat}}}$ in $\mathcal{W}$, one obtains

$$\langle \mathbf{x}, \mathbf{b} - \mathbf{A}_{\text{enu}}\mathbf{s}_{\text{enu}} - \mathbf{A}_{\text{fft}}\mathbf{s}_{\text{fft}}\rangle = \langle \mathbf{x}, \mathbf{e}\rangle + \langle \mathbf{x}, \mathbf{A}_{\text{lat}}\mathbf{s}_{\text{lat}}\rangle = \langle \mathbf{x}, \mathbf{e}\rangle + \langle \mathbf{y}, \mathbf{s}_{\text{lat}}\rangle \bmod q. \quad (1)$$

Because both $(\mathbf{x}^\top, \mathbf{y}^\top)$ and $(\mathbf{e}^\top, \mathbf{s}_{\mathrm{lat}}^\top)$ are short, inner products on the right-hand side are biased toward zero when the guess for $(\mathbf{s}_{\mathrm{enu}}, \mathbf{s}_{\mathrm{fft}})$ is correct. Conversely, under an incorrect guess $(\widetilde{\mathbf{s}_{\mathrm{enu}}}, \widetilde{\mathbf{s}_{\mathrm{fft}}}) \neq (\mathbf{s}_{\mathrm{enu}}, \mathbf{s}_{\mathrm{fft}})$, the quantity $\langle \mathbf{x}, \mathbf{b} - \mathbf{A}_{\mathrm{enu}}\widetilde{\mathbf{s}_{\mathrm{enu}}} - \mathbf{A}_{\mathrm{fft}}\widetilde{\mathbf{s}_{\mathrm{fft}}} \rangle$ is uniformly distributed over $\mathbb{Z}_q$. A natural score that highlights the shortness of the guess is

$$s_{\mathbf{w}} \triangleq \cos\left(\frac{2\pi}{q}\langle \mathbf{x}, \mathbf{b} - \mathbf{A}_{\mathrm{enu}}\widetilde{\mathbf{s}_{\mathrm{enu}}} - \mathbf{A}_{\mathrm{fft}}\widetilde{\mathbf{s}_{\mathrm{fft}}} \rangle\right).$$

If a guess $\mathbf{b} - \mathbf{A}_{\mathrm{enu}}\widetilde{\mathbf{s}_{\mathrm{enu}}} - \mathbf{A}_{\mathrm{fft}}\widetilde{\mathbf{s}_{\mathrm{fft}}}$ is correct, then $s_{\mathbf{w}}$ should be close to 1; however, the converse need not hold. To increase confidence in the reliability of this score, one naturally considers the total score over many vectors in $\mathcal{W}$. Accordingly, the following score function, as used in [10], is considered:

$$F_{\mathcal{W}}(\widetilde{\mathbf{s}_{\mathrm{enu}}}, \widetilde{\mathbf{s}_{\mathrm{fft}}}) = \sum_{(\mathbf{x}^\top, \mathbf{y}^\top) \in \mathcal{W}} \cos\left(\frac{2\pi}{q}\left(\langle \mathbf{x}, \mathbf{b} - \mathbf{A}_{\mathrm{enu}}\widetilde{\mathbf{s}_{\mathrm{enu}}} \rangle - \langle \mathbf{A}_{\mathrm{fft}}^\top \mathbf{x}, \widetilde{\mathbf{s}_{\mathrm{fft}}} \rangle\right)\right).$$

This function is referred to as the simple distinguisher f_{simple} by Laarhoven and Walter [21], and resembles the Aharonov-Regev [1] distinguisher closely. When $(\widetilde{\mathbf{s}_{\mathrm{enu}}}, \widetilde{\mathbf{s}_{\mathrm{fft}}}) = (\mathbf{s}_{\mathrm{enu}}, \mathbf{s}_{\mathrm{fft}})$, each term of $F_{\mathcal{W}}(\widetilde{\mathbf{s}_{\mathrm{enu}}}, \widetilde{\mathbf{s}_{\mathrm{fft}}})$ tends to be close to 1, and hence the overall score is significantly larger than 0. In contrast, for an incorrect guess $(\widetilde{\mathbf{s}_{\mathrm{enu}}}, \widetilde{\mathbf{s}_{\mathrm{fft}}}) \neq (\mathbf{s}_{\mathrm{enu}}, \mathbf{s}_{\mathrm{fft}})$, the expectation of $F_{\mathcal{W}}(\widetilde{\mathbf{s}_{\mathrm{enu}}}, \widetilde{\mathbf{s}_{\mathrm{fft}}})$ is close to 0. This statistical gap makes it possible to distinguish the correct guess from incorrect ones.

Since n_{enu} is typically small in practice, exhaustive search over $\widetilde{\mathbf{s}_{\mathrm{enu}}}$ is feasible. For each fixed $\widetilde{\mathbf{s}_{\mathrm{enu}}}$, FFT is then used to evaluate the score simultaneously for all $\mathbf{s}_{\mathrm{fft}} \in \mathbb{Z}_q^{n_{\mathrm{fft}}}$, after which the maximum score is taken as the score associated with the guess $\widetilde{\mathbf{s}_{\mathrm{enu}}}$. Accordingly, the score function corresponding to $\widetilde{\mathbf{s}_{\mathrm{enu}}}$ as

$$E_{\mathcal{W}}(\widetilde{\mathbf{s}_{\mathrm{enu}}}) \triangleq \max_{\mathbf{z} \in \mathbb{Z}_q^{n_{\mathrm{fft}}}} F_{\mathcal{W}}(\widetilde{\mathbf{s}_{\mathrm{enu}}}, \mathbf{z})$$

is considered. If $E_{\mathcal{W}}(\widetilde{\mathbf{s}_{\mathrm{enu}}}) > T$, where T is a prescribed threshold, then $\widetilde{\mathbf{s}_{\mathrm{enu}}}$ is regarded as the correct guess. Specifically, we refer to this approach as the *original dual attack* in this paper. It differs slightly from the method of Guo and Johansson [17], which we do not discuss further in our work.

The efficiency of computing $E_{\mathcal{W}}(\widetilde{\mathbf{s}_{\mathrm{enu}}})$ crucially relies on the FFT, which calculate the scores for all possible candidates simultaneously. However, when the modulus q is large, as in CRYSTALS-KYBER where $q = 3329$, the cost of the FFT becomes prohibitive. To address this issue, MATZOV [24] introduced a modulus switching technique, which replaces the large modulus q with a much smaller integer p. Under this approach, the MATZOV attack considers the score function

$$F_{\mathcal{W}}^{(\mathrm{ms})}(\widetilde{\mathbf{s}_{\mathrm{enu}}}, \widetilde{\mathbf{s}_{\mathrm{fft}}}) = \sum_{(\mathbf{x}^\top, \mathbf{y}^\top) \in \mathcal{W}} \cos\left(\frac{2\pi}{p}\left(\frac{p}{q}\langle \mathbf{x}, \mathbf{b} - \mathbf{A}_{\mathrm{enu}}\widetilde{\mathbf{s}_{\mathrm{enu}}} \rangle - \langle \left\lfloor \frac{p}{q}\mathbf{A}_{\mathrm{fft}}^\top \mathbf{x} \right\rceil, \widetilde{\mathbf{s}_{\mathrm{fft}}} \rangle\right)\right),$$

where $\lfloor \mathbf{x} \rceil$ denotes component-wise rounding to the nearest integer. For an incorrect guess, the expectation of $F_{\mathcal{W}}^{(\mathrm{ms})}(\widetilde{\mathbf{s}_{\mathrm{enu}}}, \widetilde{\mathbf{s}_{\mathrm{fft}}})$ remains close to 0. For the correct guess, it

follows from Eq. (1) that

$$\begin{aligned}&\frac{1}{p}\left(\frac{p}{q}\langle \mathbf{x}, \mathbf{b}-\mathbf{A}_{\text{enu}}\mathbf{s}_{\text{enu}}\rangle - \langle \lfloor \frac{p}{q}\mathbf{A}_{\text{fft}}^{\top}\mathbf{x}\rceil, \mathbf{s}_{\text{fft}}\rangle\right)\\ &= \frac{1}{q}\left(\langle \mathbf{x}, \mathbf{e}\rangle + \langle \mathbf{y}, \mathbf{s}_{\text{lat}}\rangle\right) + \frac{1}{p}\langle \left\{\frac{p}{q}\mathbf{A}_{\text{fft}}^{\top}\mathbf{x}\right\}, \mathbf{s}_{\text{fft}}\rangle \bmod 1,\end{aligned} \tag{2}$$

where $\{\mathbf{x}\} = \mathbf{x} - \lfloor \mathbf{x} \rceil$. Therefore, after modulus switching, an additional perturbation $\frac{2\pi}{p}\langle \left\{\frac{p}{q}\mathbf{A}_{\text{fft}}^{\top}\mathbf{x}\right\}, \mathbf{s}_{\text{fft}}\rangle$ appears within the cosine of each term. Since both $\mathbf{s}_{\text{fft}}$ and $\left\{\frac{p}{q}\mathbf{A}_{\text{fft}}^{\top}\mathbf{x}\right\}$ are short vectors, the inner product $\langle \left\{\frac{p}{q}\mathbf{A}_{\text{fft}}^{\top}\mathbf{x}\right\}, \mathbf{s}_{\text{fft}}\rangle$ is close to 0. However, if p is small, the factor $1/p$ amplifies this inner product, which reduces the score for the correct guess compared to that of the original dual attack. The amount of reduction is determined by the chosen modulus p. The corresponding score function is then expressed as

$$E_{\mathcal{W}}^{(\text{ms})}(\widetilde{\mathbf{s}_{\text{enu}}}) \triangleq \max_{\mathbf{z}\in\mathbb{Z}_p^{n_{\text{fft}}}} F_{\mathcal{W}}^{(\text{ms})}(\widetilde{\mathbf{s}_{\text{enu}}}, \mathbf{z}).$$

In this way, the FFT is performed over $\mathbb{Z}_p^{n_{\text{fft}}}$ instead of $\mathbb{Z}_q^{n_{\text{fft}}}$, thereby significantly reducing the FFT complexity, at the expense of lowering the score of the correct guess.

In Crypto 2025, Carrier et al. [10] proposed a dual attack based on lossy source coding. They observed that the perturbation introduced by coding is smaller than that caused by modulus switching, yielding a higher score for the correct guess. Specifically, for $\mathbf{w}^{\top} = (\mathbf{x}^{\top}, \mathbf{y}^{\top})$, let $\mathbf{G}\mathbf{u}_{\text{lsc}}$ denote the list-decoding output for $\mathbf{A}_{\text{fft}}^{\top}\mathbf{x}$ in the code $C(\mathbf{G})$. The concrete definitions of the list decoder and $\mathbf{G}$ are given in [10] and are not needed here. Then the corresponding score function takes the following form

$$\begin{aligned}F_{\mathcal{W}}^{(\text{lsc})}(\widetilde{\mathbf{s}_{\text{enu}}}, \widetilde{\mathbf{s}_{\text{fft}}}) &= \sum_{(\mathbf{x}^{\top},\mathbf{y}^{\top})\in\mathcal{W}} \cos\left(\frac{2\pi}{q}\left(\langle \mathbf{x}, \mathbf{b}-\mathbf{A}_{\text{enu}}\widetilde{\mathbf{s}_{\text{enu}}}\rangle - \langle \mathbf{G}\mathbf{u}_{\text{lsc}}, \widetilde{\mathbf{s}_{\text{fft}}}\rangle\right)\right)\\ &= \sum_{(\mathbf{x}^{\top},\mathbf{y}^{\top})\in\mathcal{W}} \cos\left(\frac{2\pi}{q}\left(\langle \mathbf{x}, \mathbf{b}-\mathbf{A}_{\text{enu}}\widetilde{\mathbf{s}_{\text{enu}}}\rangle - \langle \mathbf{u}_{\text{lsc}}, \mathbf{G}^{\top}\widetilde{\mathbf{s}_{\text{fft}}}\rangle\right)\right).\end{aligned}$$

For the correct guess, Eq. (1) implies that

$$\begin{aligned}&\frac{1}{q}\left(\langle \mathbf{x}, \mathbf{b}-\mathbf{A}_{\text{enu}}\mathbf{s}_{\text{enu}}\rangle - \langle \mathbf{G}\mathbf{u}_{\text{lsc}}, \mathbf{s}_{\text{fft}}\rangle\right)\\ &= \frac{1}{q}\left(\langle \mathbf{x}, \mathbf{e}\rangle + \langle \mathbf{y}, \mathbf{s}_{\text{lat}}\rangle + \langle \mathbf{e}_{\text{lsc}}, \mathbf{s}_{\text{fft}}\rangle\right) \bmod 1,\end{aligned} \tag{3}$$

where $\mathbf{e}_{\text{lsc}} \triangleq \mathbf{A}_{\text{fft}}^{\top}\mathbf{x} - \mathbf{G}\mathbf{u}_{\text{lsc}}$ denotes the decoding error vector. Since both $\mathbf{e}_{\text{lsc}}$ and $\mathbf{s}_{\text{fft}}$ are short, the right-hand side of Eq. (3) remains biased toward 0, and hence each term of $F_{\mathcal{W}}^{(\text{lsc})}(\mathbf{s}_{\text{enu}}, \mathbf{s}_{\text{fft}})$ tends to be close to 1. Moreover, since $\langle \mathbf{G}\mathbf{u}_{\text{lsc}}, \mathbf{s}_{\text{fft}}\rangle = \langle \mathbf{u}_{\text{lsc}}, \mathbf{G}^{\top}\mathbf{s}_{\text{fft}}\rangle$, when $\mathbf{s}_{\text{fft}}$ ranges over $\mathbb{Z}_q^{n_{\text{fft}}}$, the vector $\mathbf{G}^{\top}\mathbf{s}_{\text{fft}}$ ranges over $\mathbb{Z}_q^{k_{\text{fft}}}$. Therefore, it suffices in practice to enumerate over $\mathbb{Z}_q^{k_{\text{fft}}}$ only, and the corresponding score function is defined as

$$E_{\mathcal{W}}^{(\text{lsc})}(\widetilde{\mathbf{s}_{\text{enu}}}) \triangleq \max_{\mathbf{z}\in\mathbb{Z}_q^{k_{\text{fft}}}} \sum_{(\mathbf{x}^{\top},\mathbf{y}^{\top})\in\mathcal{W}} \cos\left(\frac{2\pi}{q}\left(\langle \mathbf{x}, \mathbf{b}-\mathbf{A}_{\text{enu}}\widetilde{\mathbf{s}_{\text{enu}}}\rangle - \langle \mathbf{u}_{\text{lsc}}, \mathbf{z}\rangle\right)\right).$$

Hence, the FFT dimension is reduced from n_{fft} to k_{fft}, again at the cost of lowering the score for the correct guess and introducing an additional decoding cost. In this paper, these two types of dual attacks are referred to as the *dual attack with modulus switching* and the *dual attack with decoding*, respectively.

Analysis of the Effectiveness of Dual Attacks. As dual attacks have been refined, research has increasingly focused on analyzing their practical efficiency. Laarhoven and Walter [21] conducted an initial evaluation of dual attack and gave three types of distinguisher. They also gave estimating the expectation and variance of the distinguisher and evaluating the effectiveness of dual attacks. However, this variance estimation relied on the independence assumption, which posits that the total score variance equals the sum of the variances of individual scores, implicitly assuming that scores corresponding to each short vector are mutually independent. Subsequently, MATZOV [24] adopted the same independence assumption to estimate the variance and success probability in their analysis of the dual attack with modulus switching.

In Crypto 2023, Ducas and Pulles [12] generalized the sieve-based dual attack to arbitrary BDD instances and revealed fundamental flaws in the underlying independence assumption. They showed that this assumption leads to contradictions with known statistical impossibility results in certain regimes. Through extensive experiments, they further demonstrated that the predicted score distributions for both uniform and BDD targets are inaccurate: the tail of the uniform score exhibits a waterfall-floor phenomenon, while the variance of the BDD score is severely underestimated. Consequently, the success probabilities claimed for dual attacks with FFT are likely overestimated, and previous analyses are deemed unreliable. Bashiri and Wiemers [30] confirmed, via covariance analysis, that the independence assumption does not hold, and estimated the variance using a chi-squared distribution. On the other hand, Ducas and Pulles [13] proposed a framework based on a weakened independence assumption combined with conditional probability integration. They derived the cumulative distribution function of the score for both correct and incorrect guesses in the dual attack on the BDD instance, and their experiments confirmed that this framework yields highly accurate predictions. In Eurocrypt 2024, Pouly and Shen [26] established a provable dual attack without relying on any statistical assumptions and clearly identified a set of parameters for which their attack works. More recently, Carrier et al. [10] replaced modulus switching with lossy source coding, achieving superior performance compared to the MATZOV attack. Their method sets the threshold T to the expected score for correct guesses, thereby avoiding variance considerations and enabling a complexity evaluation that does not rely on the independence assumption.

1.2 Our Contribution

A central question in dual attack analysis is how to bring together the diverse methods and inconsistent analytical frameworks for LWE-based problems, so that different dual attacks can be compared and understood within a common perspective. This paper addresses this question by proposing a unified predictive model with improved accuracy, whose validity is supported by our experiments. We further extend the analytical

framework of Ducas and Pulles [13], offering broader research tools and theoretical foundations for dual attack analysis on LWE. Our specific contributions are as follows.

A more accurate variance estimate based on covariance. To the best of our knowledge, the work in [30] first analyzed the independence assumption in dual attacks from a covariance perspective, and provided estimates for both covariance and variance using chi-squared distributions. Specifically, for two short vectors $\mathbf{w}_1$ and $\mathbf{w}_2$, the covariance between their corresponding scores depends on the distributions of $\|\mathbf{w}_1 \pm \mathbf{w}_2\|^2$ and $\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2$. They assumed that the components of $\mathbf{w}_1$ and $\mathbf{w}_2$ are independent and identically distributed (i.i.d.) normal with mean zero and variance τ_0. Consequently, the components of $\mathbf{w}_1 + \mathbf{w}_2$ are normal with mean zero and variance $2\tau_0$, and the squared lengths follow chi-squared distributions. Based on this, they estimated $\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2$ and $\|\mathbf{w}_1 \pm \mathbf{w}_2\|^2$. However, for the medium-to-high dimensional regimes typical in dual attacks, our experiments show that this estimation is inaccurate, likely because the component-wise assumption on short vectors is too crude in higher dimensions. In contrast, our model focuses directly on the geometric relationship between short vectors, specifically the distribution of $\cos\theta$ in the identity $\|\mathbf{w}_1 \pm \mathbf{w}_2\|^2 = \|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 \pm 2\|\mathbf{w}_1\|\|\mathbf{w}_2\| \cos\theta$. Experimental observations indicate that the cosine of the angle between short vectors follows a normal distribution, which we use to estimate the above quantities. Our experiments show that, in medium-to-high dimensions, our model significantly outperforms the estimate of [30] and achieves high accuracy, because the geometric relationship between short vectors becomes more pronounced in higher dimensions.

A unified predictive model with intuitive derivation and closed-form estimates. Several works have predicted the score distribution for the correct guess. Ducas and Pulles [13] derived the cumulative distribution function for the original dual attack, directly characterizing the score distribution. However, their analysis was developed for general BDD instances rather than LWE instances, and the underlying lattice structures differ: one uses a dual lattice, while the other uses the lattice defined in Eq. (4). Consequently, their model cannot be directly applied to the problem considered here. In Crypto 2025, Carrier et al. [10] estimated the expected score via Poisson summation and Fourier analysis, but their derivation is tailored to their proposed attack. Despite these differences, the score functions of different dual attacks share a similar form, which motivates us to develop a unified analytical model directly from the score expression. By decomposing the score into a combination of simple distributions, we estimate its expectation and variance using only elementary probability theory. The resulting closed-form estimates are intuitive and concise, clearly revealing how each attack parameter influences the score distribution and thereby deepening our understanding of dual attacks. Moreover, our model serves as a simple predictive tool that can be directly used in the complexity analysis of dual attacks.

Extending the Ducas–Pulles framework with a refined tail-behavior analysis. The analysis of dual attacks inevitably involves characterizing the tail behavior, which corresponds to the probability that an incorrect guess is mistaken for a correct one. Ducas and

Pulles [12] first identified this phenomenon, showing that for sufficiently large thresholds T, the probability that an incorrect guess exceeds T is higher than predicted by the theoretical distribution, resulting in a floor in the survival function. Their analysis, however, was developed for BDD instances and implicitly assumed that the sieve runs on the full lattice. Carrier et al. [10] later related tail behavior to the sieve dimension β_{sieve} in their decoding-based attack, but a systematic treatment for LWE-based dual attacks with sieving on a sublattice is still lacking. Our work extends the Ducas–Pulles framework to LWE-based dual attacks and refines it for the practical case where the sieve operates on a sublattice, i.e., $\beta_{\text{sieve}} < d$. Within this extended framework, we characterize the tail behavior based on our predictive model and apply it to the dual attack with decoding.

1.3 Organization

Notation, basic concepts, and lemmas are introduced in Section 2. Our predictive model is presented in Section 3: the cosine distribution of short vectors is analyzed in Section 3.1, and its application to three dual attacks is carried out in Sections 3.2–3.4. Validation of the model via comparison with experiments and application to an existing attack is given in Section 4. An extended tail behavior analysis is provided in Section 5. Finally, conclusions and open problems are discussed in Section 6.

2 Preliminaries

We denote vectors by bold lowercase letters and matrices by bold uppercase letters (e.g., $\mathbf{v}$ and $\mathbf{M}$), and write $\mathbf{v}^\top$ or $\mathbf{M}^\top$ for their transpose. Unless otherwise specified, all vectors are regarded as column vectors. The concatenation of two row vectors $\mathbf{x}^\top$ and $\mathbf{y}^\top$ is denoted by $(\mathbf{x}^\top, \mathbf{y}^\top)$ and the inner product of $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\langle \mathbf{x}, \mathbf{y} \rangle$. For a binary string $\mathbf{x} \in \{0,1\}^*$ and a bit $b \in \{0,1\}$, we use $b||\mathbf{x}$ to denote the prefix concatenation of b to $\mathbf{x}$. The exponential function is denoted by e^x or $\exp(x)$, with the latter being preferred when the expression becomes cumbersome.

2.1 Probability and Distributions

The expectation and variance of a random variable X are denoted by $\mathbb{E}[X]$ and $\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, respectively. Its standard deviation is $\sigma_X = \sqrt{\mathbb{V}[X]}$. For two random variables X and Y, their covariance is denoted by $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$, and the variance of $X + Y$ is $\mathbb{V}[X+Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\mathrm{Cov}(X,Y)$. For a distribution $\mathcal{D}$, the *cumulative distribution function* (CDF) at $x \in \mathbb{R}$ is $\Pr_{X \leftarrow \mathcal{D}}[X \le x]$, while the *probability density function* (PDF) at $x \in \mathbb{R}$ is the derivative of $\Pr_{X \leftarrow \mathcal{D}}[X \le x]$ with respect to x. The *survival function* (SF) of $\mathcal{D}$ is $\Pr_{X \leftarrow \mathcal{D}}[X \ge x] = 1 - \Pr_{X \leftarrow \mathcal{D}}[X < x]$. If $\Pr_{X \leftarrow \mathcal{D}}[X \le x] = \Pr_{X \leftarrow \mathcal{D}}[X \ge -x]$ holds for all $x \in \mathbb{R}$, then the distribution $\mathcal{D}$ is called *symmetric*.

Let $\mathcal{N}(\mu, \sigma^2)$ denote the normal distribution with mean μ and variance σ^2, and let $\mathcal{U}(\mathcal{X})$ denote the uniform distribution over a finite set $\mathcal{X}$. It is clear that both $\mathcal{N}(0, \sigma^2)$ and $\mathcal{U}([-a, a])$ are symmetric for any $\sigma, a > 0$. Another important symmetric distribution in lattice-based cryptography is the centered binomial distribution, which is formally defined as follows.

Definition 1 (Centered Binomial Distribution). *The centered binomial distribution $\mathcal{B}_\eta$ of parameter $\eta \in \mathbb{N}$ is defined as $\mathcal{B}_\eta \sim \sum_{i=1}^{\eta}(X_i - Y_i)$, where X_i and Y_i are i.i.d. uniform random variables over $\{0,1\}$. In particular, for random $X \sim \mathcal{B}_\eta$, $\Pr(X = j) = \frac{1}{2^{2\eta}}\binom{2\eta}{\eta+j}$ for all $j \in [-\eta, \eta]$.*

It is clear that $\mathcal{B}_\eta$ is a symmetric distribution with mean 0 and standard deviation $\sigma = \sqrt{\eta/2}$. The following lemmas summarize several elementary properties of symmetric distributions that will be used throughout the paper. The proofs of these lemmas are deferred to the appendix A.

Lemma 1. *If $X \sim \mathcal{U}([-\frac{1}{2}, \frac{1}{2}])$, then for any constant $t \in \mathbb{R}$, $\mathbb{E}[\cos(tX)] = 2\sin(\frac{t}{2})/t$.*

Lemma 2. *Let X and Y be independent random variables, both following symmetric distributions. Then, $\mathbb{E}[\cos(X+Y)] = \mathbb{E}[\cos(X)] \cdot \mathbb{E}[\cos(Y)]$.*

Lemma 3. *Let $X, Y \sim \mathcal{U}([-\frac{1}{2}, \frac{1}{2}])$ be independent random variables. Then, for any constant $t \in \mathbb{R}$, $\mathbb{E}[\cos(t(X+Y)] = 2(1-\cos(t))/t^2$.*

Lemma 4. *Let $X \sim \mathcal{N}(\mu, \sigma^2)$. For any constant $t \in \mathbb{R}$, $\mathbb{E}\left[e^{tX}\right] = \exp(t\mu + \frac{t^2\sigma^2}{2})$.*

2.2 Lattice Background

Given a full column-rank matrix $\mathbf{B} \in \mathbb{R}^{m\times d}$, the lattice generated by $\mathbf{B}$, denoted by $\Lambda(\mathbf{B})$, is the set of all integer linear combinations of the column vectors of $\mathbf{B}$, namely, $\Lambda(\mathbf{B}) = \left\{\mathbf{Bx} \mid \mathbf{x} \in \mathbb{Z}^d\right\}$. The column vectors $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_d$ of $\mathbf{B}$ are called a *basis* of the lattice $\Lambda(\mathbf{B})$, and the integer d is called its rank. For simplicity, we refer to $\mathbf{B}$ itself as the basis of $\Lambda(\mathbf{B})$. In particular, when the basis matrix is square ($m = d$) and nonsingular, $\Lambda(\mathbf{B})$ is called a *full-rank lattice*. The determinant of $\Lambda(\mathbf{B})$ is defined as $\det(\Lambda(\mathbf{B})) = \sqrt{\det(\mathbf{B}^\top\mathbf{B})}$. Noted that the determinant is independent of basis choice. The *first successive minimum* of a lattice Λ denoted by $\lambda_1(\Lambda)$, is defined as the Euclidean norm of the shortest non-zero vector in the lattice, i.e., $\lambda_1(\Lambda) \triangleq \min_{\mathbf{v}\in\Lambda\setminus\{\mathbf{0}\}} \|\mathbf{v}\|$. For $k \in \{1, \ldots, d\}$, denote by π_k the orthogonal projection onto the orthogonal complement $\mathrm{span}(\mathbf{b}_1, \ldots, \mathbf{b}_{k-1})^\perp$. A *projected sublattice* $\Lambda(\mathbf{B}_{[k,\ell]})$ is generated by the basis $\mathbf{B}_{[k,\ell]} \triangleq (\pi_k(\mathbf{b}_k), \ldots, \pi_k(\mathbf{b}_\ell))$ for $1 \le k < \ell \le d$, and we denote by $\Lambda(\mathbf{B}_{(k,\ell)})$ the sublattice generated by $\mathbf{b}_k, \ldots, \mathbf{b}_\ell$, where $\mathbf{B}_{(k,\ell)} \triangleq (\mathbf{b}_k, \ldots, \mathbf{b}_\ell)$. The dual of a lattice Λ is defined as $\Lambda^\vee = \{\mathbf{w} \in \mathrm{span}(\Lambda) \mid \langle \mathbf{w}, \mathbf{v}\rangle \in \mathbb{Z} \text{ for all } \mathbf{v} \in \Lambda\}$. Let $\mathbf{B}$ be a basis of Λ, then $\mathbf{D} = \mathbf{B} \cdot (\mathbf{B}^\top\mathbf{B})^{-1}$ is a basis of $\Lambda^\vee$. If Λ is full-rank, then $\mathbf{D} = (\mathbf{B}^\top)^{-1}$. For all $k \in \{1, \ldots, d\}$, the vectors $(\mathbf{b}_d, \ldots, \mathbf{b}_k)$ generate a lattice that is dual to the projected sublattice generated by $\mathbf{D}_{[k,d]} = (\pi_k(\mathbf{d}_k), \ldots, \pi_k(\mathbf{d}_d))$. We now give the definition that is central to the lattice-based problem considered in this paper.

Definition 2 (LWE Instance). *Let n and m be positive integers, χ_e a probability distribution on $\mathbb{Z}_q$ and $\mathbf{s} \in \mathbb{Z}_q^n$ a secret vector. A pair*

$$(\mathbf{A}, \mathbf{b} = \mathbf{As} + \mathbf{e} \bmod q) \in \mathbb{Z}_q^{m\times n} \times \mathbb{Z}_q^m$$

is called an LWE instance, where $\mathbf{A}$ is chosen uniformly at random and $\mathbf{e} \leftarrow \chi_e^m$.

Many modern LWE-based cryptosystems adopt the standard setting where both secret and error vectors are short, with coordinates taken from a small interval. For the dual attack, we assume they follow the same distribution $\chi_s = \chi_e$ (zero mean, small σ_e). The following two assumptions are commonly used in lattice security evaluation.

Assumption 1 (Geometric Series Assumption (GSA) [28]). *Let Λ be a full-rank lattice of dimension d, and let $2 \le \beta \le d$ be some integer. Denote by $\{\mathbf{b}_i^*\}_{i=1}^d$ the Gram-Schmidt orthogonalization of the basis of Λ after running $BKZ_{d,\beta}$, which means the BKZ algorithm for lattices of dimension d and block-size β. Assuming that $\beta \gg 200$ and $k < d - 2\beta$, then the first k vectors $\mathbf{b}_1^*, \mathbf{b}_2^*, \ldots, \mathbf{b}_k^*$ follow the Geometric Series Assumption. That is,*

$$\delta(\beta) \triangleq \frac{\|\mathbf{b}_i^*\|}{\|\mathbf{b}_{i+1}^*\|} \approx \left(\frac{\beta}{2\pi e} \cdot (\pi\beta)^{1/\beta}\right)^{\frac{1}{\beta-1}} \text{ for } 1 \le i < k.$$

Assumption 2 (Gaussian Heuristic (GH)). *Let Λ be a random full-rank lattice of dimension d and determinant $\det(\Lambda)$. Let S be a centrally symmetric convex set. Then with high probability,*

$$\#(\Lambda \cap S) \approx \mathrm{Vol}(S)/\det(\Lambda).$$

In particular, the length of its shortest non-zero vector

$$\lambda_1(\Lambda) \approx \sqrt{\frac{d}{2\pi e} \cdot (\pi d)^{1/d}} \cdot \det(\Lambda)^{1/d},$$

and for every $c > 1$,

$$\#\{v \in \Lambda \mid \|v\| \le c \cdot \lambda_1(\Lambda)\} \approx c^d.$$

In this paper, we adopt the approximation $\lambda_1(\Lambda) \approx \sqrt{\frac{d}{2\pi e}} \cdot \det(\Lambda)^{1/d}$, noting that $(\pi d)^{1/d} \approx 1$ for large d. Although this heuristic is not rigorous in full generality, it has proven remarkably accurate in practice [16] and is routinely used in the analysis of lattice basis reduction and sieving algorithms.

Short Vectors Sampling. For an LWE instance $(\mathbf{A}, \mathbf{b} = \mathbf{As} + \mathbf{e} \bmod q)$, the dual attack samples short vectors by applying the BKZ algorithm [11] and a lattice sieve [3] to the constructed lattice

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{A}_{\text{lat}}^\top & q\mathbf{I}_{n_{\text{lat}}} \end{bmatrix}, \tag{4}$$

where $\mathbf{A}_{\text{lat}}$ is an $m \times n_{\text{lat}}$ submatrix of $\mathbf{A}$. The sieve algorithm outputs a list $\mathcal{W}$ of short vectors $\mathbf{w} \in \mathbb{Z}_q^{m+n_{\text{lat}}}$, whose length can be estimated using the block size β_{bkz} in BKZ and the sieve dimension β_{sieve} by GSA and GH.

Lemma 5. *([24, Lemma 4.2]) Let Λ be a d-dimensional full-rank lattice. Then, applying a β_{sieve}-dimensional sieve algorithm to a β_{bkz}-reduced basis outputs at least $N = (\frac{4}{3})^{\beta_{\text{sieve}}/2}$ vectors of expected length ℓ given by*

$$\ell \triangleq \det(\Lambda)^{\frac{1}{d}} \cdot \sqrt{\frac{4}{3}} \cdot \delta\left(\beta_{\text{sieve}}\right)^{\beta_{\text{sieve}}-1} \cdot \delta\left(\beta_{\text{bkz}}\right)^{d-\beta_{\text{sieve}}},$$

where $\delta(\beta) = \left(\frac{\beta}{2\pi e}(\pi\beta)^{\frac{1}{\beta}}\right)^{\frac{1}{2(\beta-1)}}$ is the root-Hermite factor.

3 Expectation And Variance Estimation of Dual Attacks

In this section, we present our method for estimating the expectation and variance of the correct-guess score in dual attacks on LWE. We first introduce two assumptions derived from experimental observations, and provide theoretical justifications for their reasonableness. Based on these assumptions, we analyze the score distribution for the original dual attack, and then extend the analysis to dual attacks that incorporate either modulus switching or decoding.

3.1 Cosine Distribution for Pairwise Short Vectors

In this subsection, we model the distribution of the cosine of the angle between short dual vectors as a normal distribution. To highlight the significance of this distribution, we first recall the original dual attack and its corresponding analysis. For a correct guess and a short vector $\mathbf{w}^\top = (\mathbf{x}^\top, \mathbf{y}^\top)$, Eq. (1) gives

$$\cos\left(\frac{2\pi}{q}\langle \mathbf{x}, \mathbf{b} - \mathbf{A}_{\text{enu}}\mathbf{s}_{\text{enu}} - \mathbf{A}_{\text{fft}}\mathbf{s}_{\text{fft}}\rangle\right) = \cos\left(\frac{2\pi}{q}\langle \mathbf{w}, \mathbf{e}'\rangle\right),$$

where $\mathbf{e}' = (\mathbf{e}^\top, \mathbf{s}_{\text{lat}}^\top)^\top$. We define the individual score for a correct guess as

$$f_{\mathbf{w}}(\mathbf{e}') \triangleq \cos\left(\frac{2\pi}{q}\langle \mathbf{w}, \mathbf{e}'\rangle\right).$$

For convenience, we write $f_{\mathbf{w}}$ in place of $f_{\mathbf{w}}(\mathbf{e}')$, with the understanding that its expectation and variance are taken with respect to the distribution of $\mathbf{e}'$. To estimate the distribution of $f_{\mathbf{w}}$, we make use of the following lemma, which is analogous to Lemma 6 of [21]. Its proof is adapted from that of [21] and is deferred to Appendix A.

Lemma 6. *If $\mathbf{e}'$ follows the distribution $\chi_e^{m+n_{\text{lat}}}$ with zero mean and standard deviation σ_e, then*

$$\mathbb{E}[f_{\mathbf{w}}] = \rho(\|\mathbf{w}\|) \text{ and } \mathbb{V}[f_{\mathbf{w}}] = \frac{1}{2} - \rho(\|\mathbf{w}\|)^2 + \frac{1}{2}\rho(\|\mathbf{w}\|)^4,$$

where $\rho(\|\mathbf{w}\|) = \exp(-2\pi^2 \frac{\|\mathbf{w}\|^2\sigma_e^2}{q^2})$.

Before proceeding further, we recall the independence assumption.

Assumption 3 (Independence Assumption). *For any fixed set of dual vectors $\mathcal{W}$, the random variables $\{f_{\mathbf{w}}\}_{\mathbf{w}\in\mathcal{W}}$ are assumed to be mutually independent. Consequently,*

$$\mathbb{V}\left[\sum_{\mathbf{w}\in\mathcal{W}} f_{\mathbf{w}}\right] = \sum_{\mathbf{w}\in\mathcal{W}} \mathbb{V}[f_{\mathbf{w}}].$$

The analysis of dual attack in [21,24] implicitly relies on this assumption for the correct guess. However, the true variance of the total score is substantially larger than this estimate, mainly because the covariance terms arising from different short vectors cannot be neglected. For $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W}$, the covariance between the corresponding score terms is given by

$$\mathrm{Cov}(f_{\mathbf{w}_1}, f_{\mathbf{w}_2}) = \mathbb{E}[f_{\mathbf{w}_1} f_{\mathbf{w}_2}] - \mathbb{E}[f_{\mathbf{w}_1}]\mathbb{E}[f_{\mathbf{w}_2}].$$

By Lemma 6 and product-to-sum identity, we obtain

$$\begin{aligned}
\mathbb{E}[f_{\mathbf{w}_1} f_{\mathbf{w}_2}] &= \mathbb{E}[\cos\left(\frac{2\pi}{q}\langle \mathbf{w}_1, \mathbf{e}'\rangle\right) \cdot \cos\left(\frac{2\pi}{q}\langle \mathbf{w}_2, \mathbf{e}'\rangle\right)] \\
&= \mathbb{E}[\frac{1}{2}\cos\left(\frac{2\pi}{q}\langle \mathbf{w}_1 + \mathbf{w}_2, \mathbf{e}'\rangle\right) + \frac{1}{2}\cos\left(\frac{2\pi}{q}\langle \mathbf{w}_1 - \mathbf{w}_2, \mathbf{e}'\rangle\right)] \\
&= \frac{1}{2}\mathbb{E}[f_{\mathbf{w}_1+\mathbf{w}_2}] + \frac{1}{2}\mathbb{E}[f_{\mathbf{w}_1-\mathbf{w}_2}] \\
&= \frac{1}{2}\rho\left(\|\mathbf{w}_1 + \mathbf{w}_2\|\right) + \frac{1}{2}\rho\left(\|\mathbf{w}_1 - \mathbf{w}_2\|\right).
\end{aligned} \tag{5}$$

It follows that estimating the expectation of the covariance requires an evaluation of $\|\mathbf{w}_1 \pm \mathbf{w}_2\|^2$. Since $\|\mathbf{w}_1 \pm \mathbf{w}_2\|^2 = \|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 \pm 2\|\mathbf{w}_1\|\|\mathbf{w}_2\| \cos\theta$, if the distributions of $\|\mathbf{w}_1\|$ and $\|\mathbf{w}_2\|$ are known, then only the distribution of $\cos\theta$ is needed to obtain an estimate of covariance. Regarding the distribution of $\cos\theta$, we are inspired by the following experimental results, which are presented in Fig. 1. As can be seen from the figure, $\cos\theta$ can be approximated by a normal distribution with mean zero. The remaining question is how to estimate the variance σ_θ^2 of this distribution. As justified below, we set $\sigma_\theta^2 = 1/\beta_{\mathrm{sieve}}$, where β_{sieve} denotes the dimension of the sieve algorithm.

We first assume that the directions of the vectors in $\mathcal{W}$ are uniformly distributed in the ambient space $\mathbb{R}^d$, as in [13]. Equivalently, after normalization, these vectors can be regarded as uniformly distributed on the unit sphere $\mathcal{S}^{d-1}$. For any two vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{S}^{d-1}$, let θ denote the angle between them. Since the angle is invariant under orthogonal transformations, we may rotate the coordinate axes and assume without loss of generality that $\mathbf{v}_1 = (1, 0, \ldots, 0)$. Then $\cos\theta = \langle \mathbf{v}_1, \mathbf{v}_2\rangle = v_{2,1}$, where $v_{2,1}$ is the first component of $\mathbf{v}_2$. Hence, it suffices to analyze the distribution of a single component of $\mathbf{v}_2$. For a vector $\mathbf{v}$ uniformly distributed on the unit sphere $\mathcal{S}^{d-1}$, symmetry gives $\mathbb{E}[v_i] = 0$ for all i. Moreover,

$$\mathbb{E}\left[\|\boldsymbol{v}\|^2\right] = \mathbb{E}\left[\sum_{i=1}^{d} v_i^2\right] = \sum_{i=1}^{d} \mathbb{E}\left[v_i^2\right] = 1,$$

which implies that $\mathbb{E}\left[v_i^2\right] \approx \frac{1}{d}$. Therefore, the variance of each component is approximately

$$\mathbb{V}[v_i] = \mathbb{E}\left[v_i^2\right] - (\mathbb{E}[v_i])^2 \approx \frac{1}{d}.$$

Since vectors in $\mathcal{W}$ are generated by sieving in a β_{sieve}-dimensional subspace, their normalized directions are more appropriately modeled as uniformly distributed on

$\mathcal{S}^{\beta_{\text{sieve}}-1}$ rather than on the full-dimensional unit sphere. Consequently, $\sigma_\theta^2 = \mathbb{V}[\cos\theta] = \mathbb{V}[v_i]$ is taken to be $1/\beta_{\text{sieve}}$.

We further observe that, for typical sieving algorithms, the cosine of the angle between short output vectors is concentrated in the interval $\left[-\frac{1}{2}, \frac{1}{2}\right]$. To assess whether this normal distribution assumption holds for the list $\mathcal{W}$, we consider the condition $3\sigma_\theta \leq \frac{1}{2}$, which yields $\beta_{\text{sieve}} \geq 36$. This suggests that the approximation becomes reliable once the sieve dimension is sufficiently large. Since our experiments employ $\beta_{\text{sieve}} \geq 80$, this assumption is well justified.

Similarly, if the decoding error vector $\mathbf{e}_{\text{lsc}}$ is assumed to have a uniformly random direction in $\mathbb{Z}_q^{n_{\text{fft}}}$, then the cosine of the angle between the vectors $\mathbf{e}_{\text{lsc}}$ associated with $\mathcal{W}$ can be modeled as following a normal distribution $\mathcal{N}(0, \sigma_{\text{lsc}}^2)$ as well, with $\sigma_{\text{lsc}}^2 = 1/n_{\text{fft}}$. We provide experimental evidence supporting this approximation, as shown in Fig. 1. To demonstrate that our data are not coincidental, additional data and figures are provided as auxiliary supporting materials.

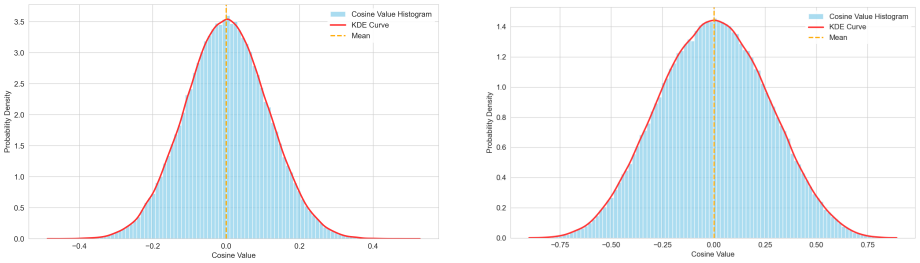

Fig. 1. Kernel density estimation (KDE) of cosine values. Left: between different short vectors from a sieve output (dimension 80, mean $\approx$ 0, variance $\approx$ 0.013, 2^{18} pairs). Right: between different decoding error vectors when $n_{\text{fft}} = 15$ (mean $\approx$ 0, variance $\approx$ 0.068, 2^{18} pairs).

Based on the above discussion, we make the following two assumptions.

Assumption 4. *Let $\mathcal{W}$ be the set of short vectors obtained by applying lattice basis reduction followed by sieving in dimension β_{sieve}. For any $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W}$, the cosine of the angle between $\mathbf{w}_1$ and $\mathbf{w}_2$ is assumed to follow a normal distribution $\mathcal{N}(0, \sigma_\theta^2)$ with $\sigma_\theta^2 = 1/\beta_{\text{sieve}}$.*

Assumption 5. *For any $\mathbf{w}_1 = (\mathbf{x}_1^\top, \mathbf{y}_1^\top)^\top, \mathbf{w}_2 = (\mathbf{x}_2^\top, \mathbf{y}_2^\top)^\top \in \mathcal{W}$, let*

$$\mathbf{e}_{\text{lsc},1} = \mathbf{A}_{\text{fft}}^\top \mathbf{x}_1 - \mathbf{G}\mathbf{u}_{\text{lsc},1}, \quad \mathbf{e}_{\text{lsc},2} = \mathbf{A}_{\text{fft}}^\top \mathbf{x}_2 - \mathbf{G}\mathbf{u}_{\text{lsc},2}.$$

The cosine of the angle between $\mathbf{e}_{\text{lsc},1}$ and $\mathbf{e}_{\text{lsc},2}$ is assumed to follow a normal distribution $\mathcal{N}(0, \sigma_{\text{lsc}}^2)$ with $\sigma_{\text{lsc}}^2 = 1/n_{\text{fft}}$.

3.2 Estimate of the Original Dual Attack

Recalling Eq. (5), it follows from Assumption 4 and Lemma 4 that

$$\begin{aligned}\mathbb{E}[f_{\mathbf{w}_1} f_{\mathbf{w}_2}] &= \frac{1}{2}\rho\left(\|\mathbf{w}_1 + \mathbf{w}_2\|\right) + \frac{1}{2}\rho\left(\|\mathbf{w}_1 - \mathbf{w}_2\|\right) \\ &= \frac{1}{2}\rho\left(\|\mathbf{w}_1\|\right)\rho\left(\|\mathbf{w}_2\|\right)\left(\exp(2\Delta\|\mathbf{w}_1\|\|\mathbf{w}_2\|\cos\theta) + \exp(-2\Delta\|\mathbf{w}_1\|\|\mathbf{w}_2\|\cos\theta)\right) \\ &= \rho\left(\|\mathbf{w}_1\|\right)\rho\left(\|\mathbf{w}_2\|\right)\exp(2\Delta^2\|\mathbf{w}_1\|^2\|\mathbf{w}_2\|^2\sigma_\theta^2),\end{aligned}$$

where $\Delta = -2\pi^2\sigma_e^2/q^2$. In the following, the inner expectation is taken over the random error $\mathbf{e}'$, while the outer expectation is taken over dual vectors $\mathbf{w}_i \in \mathcal{W}$. Assuming that all vectors in $\mathcal{W}$ share the same expected length ℓ, we obtain

$$\mathbb{E}_{\mathbf{w}_i \in \mathcal{W}}\left[\mathbb{E}[f_{\mathbf{w}_1} f_{\mathbf{w}_2}]\right] = \rho(\ell)^2 \cdot \exp(2\Delta^2\ell^4\sigma_\theta^2).$$

Therefore, from the identity $\mathrm{Cov}(f_{\mathbf{w}_1}, f_{\mathbf{w}_2}) = \mathbb{E}[f_{\mathbf{w}_1} f_{\mathbf{w}_2}] - \mathbb{E}[f_{\mathbf{w}_1}]\mathbb{E}[f_{\mathbf{w}_2}]$, we obtain

$$\mathbb{E}\big[\mathrm{Cov}(f_{\mathbf{w}_1}, f_{\mathbf{w}_2})\big] = \rho(\ell)^2\big(\exp(2\Delta^2\ell^4\sigma_\theta^2) - 1\big) > 0.$$

Combining this with the variance decomposition formula

$$\mathbb{V}\left[\sum_{\mathbf{w}\in\mathcal{W}} f_{\mathbf{w}}\right] = \sum_{\mathbf{w}\in\mathcal{W}} \mathbb{V}\left[f_{\mathbf{w}}\right] + \sum_{\mathbf{w}_i \neq \mathbf{w}_j} \mathrm{Cov}(f_{\mathbf{w}_i}, f_{\mathbf{w}_j})$$

yields the following estimate for the variance of the total score.

Theorem 1. *Assume that the vectors in $\mathcal{W}$ have expected length ℓ, and that for any distinct $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W}$, the cosine of the angle between them follows $\mathcal{N}(0, 1/\beta_{\mathrm{sieve}})$. Let $\Delta = -2\pi^2\sigma_e^2/q^2$ and define $\rho(t) = \exp(\Delta \cdot t^2)$. Then, for the correct guess $(\mathbf{s}_{\mathrm{enu}}, \mathbf{s}_{\mathrm{fft}})$, the expectation and variance of the score function can be estimated as follows:*

$$\begin{aligned}\mathbb{E}\big[F_{\mathcal{W}}(\mathbf{s}_{\mathrm{enu}}, \mathbf{s}_{\mathrm{fft}})\big] &\approx \sum_{\mathbf{w}\in\mathcal{W}} \rho(\|\mathbf{w}\|) \\ \mathbb{V}\big[F_{\mathcal{W}}(\mathbf{s}_{\mathrm{enu}}, \mathbf{s}_{\mathrm{fft}})\big] &\approx \sum_{\mathbf{w}\in\mathcal{W}} \left(\frac{1}{2} - \rho(\|\mathbf{w}\|)^2 + \frac{1}{2}\rho(\|\mathbf{w}\|)^4\right) \\ &\quad + |\mathcal{W}|(|\mathcal{W}| - 1)\,\rho(\ell)^2\left(\exp\left(2\Delta^2\ell^4/\beta_{\mathrm{sieve}}\right) - 1\right).\end{aligned}$$

3.3 Estimate of Dual Attack with Modulus Switching

We now estimate the expectation and variance of the score for the correct guess in the dual attack with modulus switching. From Eq. (2), we define the individual score as

$$f_{\mathbf{w}}^{(\mathrm{ms})} \triangleq \cos\left(\frac{2\pi}{q}\langle\mathbf{w}, \mathbf{e}'\rangle + \frac{2\pi}{p}\left\langle\left\{\frac{p}{q}\mathbf{A}_{\mathrm{fft}}^{\top}\mathbf{x}\right\}, \mathbf{s}_{\mathrm{fft}}\right\rangle\right). \tag{6}$$

To estimate the covariance of $f_{\mathbf{w}}^{(\mathrm{ms})}$, we decompose Eq. (6) into two parts and introduce the following assumptions.

Assumption 6. *For* $\mathbf{w}^\top = (\mathbf{x}^\top, \mathbf{y}^\top) \in \mathcal{W}$*, the vector* $\mathbf{w}$ *is independent of* $\left\{\frac{p}{q}\mathbf{A}_{\text{fft}}^\top \mathbf{x}\right\}$*. Consequently,* $\langle \mathbf{w}, \mathbf{e}' \rangle$ *is independent of* $\left\langle \left\{\frac{p}{q}\mathbf{A}_{\text{fft}}^\top \mathbf{x}\right\}, \mathbf{s}_{\text{fft}} \right\rangle$.

Assumption 7. *The components of* $\left\{\frac{p}{q}\mathbf{A}_{\text{fft}}^\top \mathbf{x}\right\}$ *are mutually independent and each follows the uniform distribution over* $\left[-\frac{1}{2}, \frac{1}{2}\right]$.

These two assumptions are natural. Assumption 6 holds intuitively because the modulus-switched vector is essentially uncorrelated with the original vector $\mathbf{x}$, making it independent of $\mathbf{w}$. Assumption 7 can be verified experimentally. Under these assumptions and by Lemma 2, we can derive

$$\mathbb{E}[f_{\mathbf{w}}^{(\text{ms})}] = \mathbb{E}_1 \cdot \mathbb{E}_2,$$

where

$$\mathbb{E}_1 = \mathbb{E}\left[\cos(\frac{2\pi}{q}\langle \mathbf{w}, \mathbf{e}' \rangle)\right], \mathbb{E}_2 = \mathbb{E}\left[\cos\left(\frac{2\pi}{p}\left\langle \left\{\frac{p}{q}\mathbf{A}_{\text{fft}}^\top \mathbf{x}\right\}, \mathbf{s}_{\text{fft}} \right\rangle\right)\right].$$

The term $\mathbb{E}_1$ has been discussed earlier, so we now focus on estimating $\mathbb{E}_2$. Since the components of $\mathbf{s}_{\text{fft}}$ are mutually independent and, by Assumption 7, the components of $\left\{\frac{p}{q}\mathbf{A}_{\text{fft}}^\top \mathbf{x}\right\}$ are also mutually independent, the products

$$\left\{\frac{p}{q}\mathbf{A}_{\text{fft}}^\top \mathbf{x}\right\}_i \cdot (\mathbf{s}_{\text{fft}})_i \qquad (i = 1, 2, \ldots, n_{\text{fft}})$$

are mutually independent as well, where $\left\{\frac{p}{q}\mathbf{A}_{\text{fft}}^\top \mathbf{x}\right\}_i$ and $(\mathbf{s}_{\text{fft}})_i$ denote the i-th components of the respective vectors. If χ_e is discrete, then using conditional probability together with Lemma 1 gives

$$\mathbb{E}\left[\cos\left(\frac{2\pi}{p} \cdot \left\{\frac{p}{q}\mathbf{A}_{\text{fft}}^\top \mathbf{x}\right\}_i \cdot (\mathbf{s}_{\text{fft}})_i\right)\right] = \sum_{t\in\mathbb{Z}} \Pr((\mathbf{s}_{\text{fft}})_i = t) \cdot \frac{2\sin\left(\frac{\pi t}{p}\right)}{\frac{2\pi t}{p}}.$$

Note that the above summation is finite in many cryptographic schemes. If χ_e is a continuous distribution, the same expectation can be evaluated by replacing the sum with a conditional probability integral. Here we typically consider $\chi_e = B_\eta$. Applying Lemma 2 again yields

$$\mathbb{E}_2 = \mathbb{E}\left[\cos\left(\frac{2\pi}{p} \cdot \left\{\frac{p}{q}\mathbf{A}_{\text{fft}}^\top \mathbf{x}\right\}_i \cdot (\mathbf{s}_{\text{fft}})_i\right)\right]^{n_{\text{fft}}} = \left(\sum_{t\in\mathbb{Z}} \Pr(\chi_e = t) \cdot \frac{p\sin(\frac{\pi t}{p})}{\pi t}\right)^{n_{\text{fft}}},$$

and hence

$$\mathbb{E}[f_{\mathbf{w}}^{(\text{ms})}] = \mathbb{E}_1 \cdot \mathbb{E}_2 = \rho(\|\mathbf{w}\|) \cdot \left(\sum_{t\in\mathbb{Z}} \Pr(\chi_e = t) \cdot \frac{p\sin(\frac{\pi t}{p})}{\pi t}\right)^{n_{\text{fft}}}. \tag{7}$$

We next estimate the variance of $f_{\mathbf{w}}^{(\mathrm{ms})}$. Observe that

$$\begin{aligned}\mathbb{E}\left[(f_{\mathbf{w}}^{(\mathrm{ms})})^2\right] &= \mathbb{E}\left[\frac{1}{2}+\frac{1}{2}\cos\left(\frac{4\pi}{q}\langle \mathbf{w}, \mathbf{e}'\rangle + \frac{4\pi}{p}\left\langle\left\{\frac{p}{q}\mathbf{A}_{\mathrm{fft}}^{\top}\mathbf{x}\right\}, \mathbf{s}_{\mathrm{fft}}\right\rangle\right)\right] \\ &= \frac{1}{2}+\frac{1}{2}\mathbb{E}\left[\cos\left(\frac{2\pi}{q}\langle 2\mathbf{w}, \mathbf{e}'\rangle\right)\right]\cdot\mathbb{E}\left[\cos\left(\frac{4\pi}{p}\left\langle\left\{\frac{p}{q}\mathbf{A}_{\mathrm{fft}}^{\top}\mathbf{x}\right\}, \mathbf{s}_{\mathrm{fft}}\right\rangle\right)\right] \\ &= \frac{1}{2}+\frac{1}{2}\rho(\|\mathbf{w}\|)^4\cdot\left(\sum_{t\in\mathbb{Z}}\Pr(\chi_e=t)\cdot\frac{p\sin(\frac{2\pi t}{p})}{2\pi t}\right)^{n_{\mathrm{fft}}},\end{aligned} \tag{8}$$

and so

$$\mathbb{V}\left[f_{\mathbf{w}}^{(\mathrm{ms})}\right] = \mathbb{E}\left[\left(f_{\mathbf{w}}^{(\mathrm{ms})}\right)^2\right] - \mathbb{E}\left[f_{\mathbf{w}}^{(\mathrm{ms})}\right]^2,$$

which yields the variance of an individual score term under modulus switching.

We now turn to the covariance between $f_{\mathbf{w}_1}^{(\mathrm{ms})}$ and $f_{\mathbf{w}_2}^{(\mathrm{ms})}$ for $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W}$. By the product-to-sum identity,

$$\mathbb{E}\left[f_{\mathbf{w}_1}^{(\mathrm{ms})} f_{\mathbf{w}_2}^{(\mathrm{ms})}\right] = \frac{1}{2}\mathbb{E}\left[f_{+}^{(\mathrm{ms})}\right] + \frac{1}{2}\mathbb{E}\left[f_{-}^{(\mathrm{ms})}\right],$$

where $f_{\pm}^{(\mathrm{ms})} = \cos\left(\frac{2\pi}{q}\langle \mathbf{w}_1 \pm \mathbf{w}_2, \mathbf{e}'\rangle + \frac{2\pi}{p}\left\langle\left\{\frac{p}{q}\mathbf{A}_{\mathrm{fft}}^{\top}\mathbf{x}_1\right\} \pm \left\{\frac{p}{q}\mathbf{A}_{\mathrm{fft}}^{\top}\mathbf{x}_2\right\}, \mathbf{s}_{\mathrm{fft}}\right\rangle\right)$. Treating $\left\{\frac{p}{q}\mathbf{A}_{\mathrm{fft}}^{\top}\mathbf{x}_1\right\} \pm \left\{\frac{p}{q}\mathbf{A}_{\mathrm{fft}}^{\top}\mathbf{x}_2\right\}$ as the sum or difference of two independent random vectors following $\mathcal{U}([-\frac{1}{2}, \frac{1}{2}])^{n_{\mathrm{fft}}}$, Lemma 3 implies that

$$\begin{aligned}\mathbb{E}\left[f_{\pm}^{(\mathrm{ms})}\right] &= \mathbb{E}\left[\cos\left(\frac{2\pi}{q}\langle \mathbf{w}_1 \pm \mathbf{w}_2, \mathbf{e}'\rangle\right)\right] \\ &\quad\cdot\mathbb{E}\left[\cos\left(\frac{2\pi}{p}\left\langle\left\{\frac{p}{q}\mathbf{A}_{\mathrm{fft}}^{\top}\mathbf{x}_1\right\} \pm \left\{\frac{p}{q}\mathbf{A}_{\mathrm{fft}}^{\top}\mathbf{x}_2\right\}, \mathbf{s}_{\mathrm{fft}}\right\rangle\right)\right] \\ &= \rho(\|\mathbf{w}_1\|)\rho(\|\mathbf{w}_2\|)\exp\left(2\Delta^2\|\mathbf{w}_1\|^2\|\mathbf{w}_2\|^2\sigma_\theta^2\right) \\ &\quad\cdot\left(\sum_{t\in\mathbb{Z}}\Pr(\chi_e=t)\cdot\frac{p^2\left(1-\cos\left(\frac{2\pi t}{p}\right)\right)}{2\pi^2 t^2}\right)^{n_{\mathrm{fft}}}\end{aligned}$$

Under the same assumptions as in Theorem 1, we obtain

$$\mathbb{E}\left[f_{\mathbf{w}_1}^{(\mathrm{ms})} f_{\mathbf{w}_2}^{(\mathrm{ms})}\right] = \rho(\ell)^2\cdot\exp(2\Delta^2\ell^4/\beta_{\mathrm{sieve}})\cdot\left(\sum_{t\in\mathbb{Z}}\Pr(\chi_e=t)\cdot\frac{p^2(1-\cos(\frac{2\pi t}{p}))}{2\pi^2 t^2}\right)^{n_{\mathrm{fft}}} \tag{9}$$

where ℓ, Δ, and β_{sieve} are defined as in Theorem 1, and for simplicity we denote $\mathbb{E}[f_{\mathbf{w}_1}^{(\mathrm{ms})} f_{\mathbf{w}_2}^{(\mathrm{ms})}] = \mathbb{E}_{\mathbf{w}\in\mathcal{W}}[\mathbb{E}[f_{\mathbf{w}_1}^{(\mathrm{ms})} f_{\mathbf{w}_2}^{(\mathrm{ms})}]]$. We then obtain the following theorem.

Theorem 2. *Assume that the vectors in $\mathcal{W}$ have expected length ℓ, and that for any distinct $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W}$, the cosine of the angle between them follows $\mathcal{N}(0, 1/\beta_{\mathrm{sieve}})$. Let*

$\Delta = -2\pi^2\sigma_e^2/q^2$ *and define* $\rho(t) = \exp(\Delta \cdot t^2)$. *Then, for the correct guess* $(\mathbf{s}_{\text{enu}}, \mathbf{s}_{\text{fft}})$, *the expectation and variance of the score function of dual attack with modulus switching can be estimated as follows:*

$$\mathbb{E}\left[F_{\mathcal{W}}^{(\text{ms})}(\mathbf{s}_{\text{enu}}, \mathbf{s}_{\text{fft}})\right] \approx \sum_{\mathbf{w}\in\mathcal{W}} \mathbb{E}\left[f_{\mathbf{w}}^{(\text{ms})}\right],$$

$$\mathbb{V}\left[F_{\mathcal{W}}^{(\text{ms})}(\mathbf{s}_{\text{enu}}, \mathbf{s}_{\text{fft}})\right] \approx \sum_{\mathbf{w}\in\mathcal{W}} \left(\mathbb{E}\left[(f_{\mathbf{w}}^{(\text{ms})})^2\right] - \mathbb{E}\left[f_{\mathbf{w}}^{(\text{ms})}\right]^2\right) + |\mathcal{W}|(|\mathcal{W}|-1)\left(\mathbb{E}\left[f_{\mathbf{w}_1}^{(\text{ms})} f_{\mathbf{w}_2}^{(\text{ms})})\right] - \mathbb{E}\left[f_{\mathbf{w}}^{(\text{ms})}\right]^2\right),$$

where $\mathbb{E}[f_{\mathbf{w}}^{(\text{ms})}], \mathbb{E}\left[(f_{\mathbf{w}}^{(\text{ms})})^2\right]$ *and* $\mathbb{E}\left[f_{\mathbf{w}_1}^{(\text{ms})} f_{\mathbf{w}_2}^{(\text{ms})}\right]$ *are given in Eq. (7), (8) and (9), respectively.*

3.4 Estimate of Dual Attack with Decoding

We next consider the dual attack with decoding. Following the same approach as above, we define

$$f_{\mathbf{w}}^{(\text{lsc})} = \cos\left(\frac{2\pi}{q}\left(\langle \mathbf{w}, \mathbf{e}'\rangle + \langle \mathbf{e}_{\text{lsc}}, \mathbf{s}_{\text{fft}}\rangle\right)\right)$$

according to Eq. (3). Analogous to the previous analysis, we again decompose the score function into two parts and assume their independence. Specifically, we make the following assumption, which is similar in spirit to that in [10].

Assumption 8. *For* $\mathbf{w}^\top = (\mathbf{x}^\top, \mathbf{y}^\top) \in \mathcal{W}$, *we assume that* $\mathbf{w}$ *and* $\mathbf{e}_{\text{lsc}} = \mathbf{G}\mathbf{u}_{\text{lsc}} - \mathbf{A}_{\text{fft}}^\top\mathbf{x}$ *are independent, where* $\mathbf{G}\mathbf{u}_{\text{lsc}}$ *is the codeword in* $C(\mathbf{G})$ *corresponding to* $\mathbf{A}_{\text{fft}}^\top\mathbf{x}$. *Consequently,* $\langle \mathbf{w}, \mathbf{e}'\rangle$ *and* $\langle \mathbf{e}_{\text{lsc}}, \mathbf{s}_{\text{fft}}\rangle$ *are independent.*

Since

$$f_{\mathbf{w}}^{(\text{lsc})} = \cos\left(\frac{2\pi}{q}\left(\langle \mathbf{w}, \mathbf{e}'\rangle + \langle \mathbf{e}_{\text{lsc}}, \mathbf{s}_{\text{fft}}\rangle\right)\right) = \cos\left(\frac{2\pi}{q}\langle \mathbf{w}', \mathbf{s}'\rangle\right),$$

where $\mathbf{w}' = (\mathbf{w}^\top, \mathbf{e}_{\text{lsc}}^\top)^\top$ and $\mathbf{s}' = (\mathbf{e}'^\top, \mathbf{s}_{\text{fft}}^\top)^\top$, we apply the same derivation as in Lemma 6 together with the above assumption to obtain

$$\mathbb{E}\left[f_{\mathbf{w}}^{(\text{lsc})}\right] = \exp\left(-2\pi^2\sigma_e^2\frac{\|\mathbf{w}\|^2 + \|\mathbf{e}_{\text{lsc}}\|^2}{q^2}\right) = \rho(\|\mathbf{w}\|)\cdot\rho(\|\mathbf{e}_{\text{lsc}}\|).$$

Similarly, the variance of $f_{\mathbf{w}}^{(\text{lsc})}$ is given by

$$\mathbb{V}\left[f_{\mathbf{w}}^{(\text{lsc})}\right] = \frac{1}{2} - \left(\rho(\|\mathbf{w}\|)\rho(\|\mathbf{e}_{\text{lsc}}\|)\right)^2 + \frac{1}{2}\left(\rho(\|\mathbf{w}\|)\rho(\|\mathbf{e}_{\text{lsc}}\|)\right)^4.$$

We next estimate the covariance between $f_{\mathbf{w}_1}^{(\text{lsc})}$ and $f_{\mathbf{w}_2}^{(\text{lsc})}$. First,

$$\mathbb{E}\left[f_{\mathbf{w}_1}^{(\text{lsc})} f_{\mathbf{w}_2}^{(\text{lsc})}\right] = \frac{1}{2}\mathbb{E}\left[f_{+}^{(\text{lsc})}\right] + \frac{1}{2}\mathbb{E}\left[f_{-}^{(\text{lsc})}\right],$$

where $f_{\pm}^{(\mathrm{lsc})} = \cos\left(\frac{2\pi}{q}\left(\langle \mathbf{w}_1 \pm \mathbf{w}_2, \mathbf{e}'\rangle + \langle \mathbf{e}_{\mathrm{lsc},1} \pm \mathbf{e}_{\mathrm{lsc},2}, \mathbf{s}_{\mathrm{fft}}\rangle\right)\right)$. By Assumption 8,

$$\mathbb{E}\left[f_{\pm}^{(\mathrm{lsc})}\right] = \mathbb{E}\left[\cos\left(\frac{2\pi}{q}\langle \mathbf{w}_1 \pm \mathbf{w}_2, \mathbf{e}'\rangle\right)\right] \cdot \mathbb{E}\left[\cos\left(\frac{2\pi}{q}\langle \mathbf{e}_{\mathrm{lsc},1} \pm \mathbf{e}_{\mathrm{lsc},2}, \mathbf{s}_{\mathrm{fft}}\rangle\right)\right].$$

The expectation of the first term is already known. For the second term, by Assumption 5 and Lemma 5, we have

$$\mathbb{E}\left[\cos\left(\frac{2\pi}{q}\langle \mathbf{e}_{\mathrm{lsc},1} \pm \mathbf{e}_{\mathrm{lsc},2}, \mathbf{s}_{\mathrm{fft}}\rangle\right)\right] = \rho(\|\mathbf{e}_{\mathrm{lsc},1}\|)\rho(\|\mathbf{e}_{\mathrm{lsc},2}\|)\cdot\exp\left(2\Delta^2\|\mathbf{e}_{\mathrm{lsc},1}\|^2\|\mathbf{e}_{\mathrm{lsc},2}\|^2\sigma_{\mathrm{lsc}}^2\right).$$

Assuming an average length for $\mathbf{e}_{\mathrm{lsc}}$, we obtain the following theorem.

Theorem 3. *Assume that the vectors in $\mathcal{W}$ have expected length ℓ, and that $\mathbf{e}_{\mathrm{lsc}}$ has expected length ℓ_{lsc}. Moreover, for any distinct $\mathbf{w}_1, \mathbf{w}_2 \in \mathcal{W}$, the cosine of the angle between them follows $\mathcal{N}(0, 1/\beta_{\mathrm{sieve}})$, and the cosine of the angle between their corresponding decoding error vectors $\mathbf{e}_{\mathrm{lsc},1}$ and $\mathbf{e}_{\mathrm{lsc},2}$ follows $\mathcal{N}(0, 1/n_{\mathrm{fft}})$. Let $\Delta = -2\pi^2\sigma_e^2/q^2$ and define $\rho(t) = \exp(\Delta \cdot t^2)$. Then, for the correct guess $(\mathbf{s}_{\mathrm{enu}}, \mathbf{s}_{\mathrm{fft}})$, the expectation and variance of the score function of the dual attack with decoding can be estimated as follows:*

$$\mathbb{E}\left[F_{\mathcal{W}}^{(\mathrm{lsc})}(\mathbf{s}_{\mathrm{enu}}, \mathbf{s}_{\mathrm{fft}})\right] \approx \sum_{\mathbf{w}\in\mathcal{W}} \rho(\|\mathbf{w}\|)\rho(\ell_{\mathrm{lsc}}),$$

$$\begin{aligned}\mathbb{V}\left[F_{\mathcal{W}}^{(\mathrm{lsc})}(\mathbf{s}_{\mathrm{enu}}, \mathbf{s}_{\mathrm{fft}})\right] \approx \sum_{\mathbf{w}\in\mathcal{W}} &\left(\frac{1}{2} - (\rho(\|\mathbf{w}\|)\cdot\rho(\|\mathbf{e}_{\mathrm{lsc}}\|))^2 + \frac{1}{2}(\rho(\|\mathbf{w}\|)\cdot\rho(\|\mathbf{e}_{\mathrm{lsc}}\|))^4\right) \\ &+ |\mathcal{W}|\,(|\mathcal{W}|-1)\,\rho(\ell)^2\rho(\ell_{\mathrm{lsc}})^2\left(\exp\left(2\Delta^2(\ell^4/\beta_{\mathrm{sieve}} + \ell_{\mathrm{lsc}}^4/n_{\mathrm{fft}})\right) - 1\right).\end{aligned}$$

It is worth emphasizing that the model in [13] is fundamentally different from ours. In the score model of [13], the norm of the target vector $\mathbf{t}$ is fixed, while dual vectors $\mathbf{w} \in \mathcal{W}$ are treated as variables of individual scores. Under that framework, vectors $\mathbf{w}$ are assumed to be sampled independently, and thus the individual score terms $f_{\mathbf{w}}$ are mutually independent. Consequently, the expectation and variance of the total score can be computed under the independence assumption.

By contrast, our model treats the target vector $\mathbf{t}$ — specifically its norm $\|\mathbf{t}\|$ — as the random variable, while conditioning on a fixed set of dual vectors $\mathbf{w} \in \mathcal{W}$. Hence, the independence assumption no longer holds. Based on assumptions about the distribution of the cosine of the angle between short vectors, we derive explicit estimates for the expectation and variance of the score. These two moments determine the global shape of the score distribution, although they do not capture finer local features such as the tail behavior for incorrect guesses; this issue will be addressed in later sections. Notably, a key advantage of our approach is that it remains largely insensitive to the specifics of the LWE instance: adapting to a different instance simply requires updating the distribution χ_e of the secret and error vectors, and hence its variance σ_e^2.

4 Experimental Results

In this section, we present detailed experimental results to validate the accuracy of our theoretical estimates, by comparing them with both real experimental data and the results reported in [10].

4.1 Experimental Results of Three Types of Dual Attacks

The vector list $\mathcal{W}$ is generated by applying the progressive BKZ algorithm from BLASter [14] with block size $\beta_{\text{bkz}} = 60$ to the lattice basis constructed in Eq. (4), followed by a progressive sieving with $\beta_{\text{sieve}} = 80$ in G6K [3] (in all sieve experiments, the sieve radius is fixed to $\sqrt{4/3}$ and the sieve ratio is set to 0.5). The mean length of the vectors in $\mathcal{W}$ is estimated using Lemma 5, and the size of $\mathcal{W}$ is fixed at 24860, which follows from the sieve settings (i.e., $|\mathcal{W}| = \frac{1}{2} f_{\text{sat}} r_{\text{sat}}^{\beta_{\text{sieve}}}$ with $r_{\text{sat}} = \sqrt{4/3}$, $f_{\text{sat}} = 0.5$, and $\beta_{\text{sieve}} = 80$). For the attack with decoding, we set ℓ_{lsc} to the empirical mean length of $\mathbf{e}_{\text{lsc}}$ measured in our experiments. Although [10] provides an asymptotic estimate for ℓ_{lsc}, this estimate is noticeably smaller than the observed value in the low-dimensional parameter regimes. The value of σ_θ is chosen according to the rule described in Section 3.1. Since n_{fft} is relatively small in our experiments, the theoretical estimate of σ_{lsc} is less accurate in this setting; we therefore use the experimentally observed value of σ_{lsc} when evaluating the dual attack with decoding. For simplicity, we fix $n_{\text{enu}} = 5$ throughout all experiments. We conduct comparative evaluations under four parameter settings: $(n_{\text{lat}}, n_{\text{fft}}) = (40, 5), (50, 5), (60, 5), (70, 5)$. The corresponding lattice dimension is $m + n_{\text{lat}}$, which equals $90, 110, 130, 150$, respectively. For each setting, we test three attack variants: the original dual attack, the dual attack with modulus switching, and the dual attack with decoding. Specifically, for the original dual attack, we also compare the predictions produced by our model against the estimates obtained from the model in [30]. Tables 1–3 report the experimental results for the three attack variants, respectively. Namely, Table 1 corresponds to the original dual attack, Table 2 presents the results for the modulus-switching variant, and Table 3 gives the results for the attack with decoding.

The experimental results clearly demonstrate that our model achieves substantially higher prediction accuracy than the method in [30], thereby supporting the validity of our assumptions. Moreover, the empirical distribution plots indicate that the score distribution corresponding to the correct guess is well approximated by a normal distribution in low-dimensional experiments. The experiments in [13] report extreme asymmetry for the correct-guess score in the dual attack with modulus switching for some p, contradicting our observations. We believe that this discrepancy can be explained as follows: when p is chosen too small, the expected score becomes rather low, yet the score is bounded below by zero. The combination of a low expectation and a non-negative lower bound inevitably skews the distribution, leading to the pronounced asymmetry observed in their experiments. In contrast, for the parameter regimes we consider, the expected score is sufficiently large that the lower bound does not cause a noticeable distortion, and the distribution remains approximately symmetric. As illustrated in Fig. 2 , treating the score distribution as a normal distribution is a reasonable approximation.

Table 1. Comparison of experimental and predicted mean and standard deviation of the correct-guess score in the original dual attack.

Dimension $(n_{\text{lat}}, n_{\text{fft}})$	(40,5)	(50,5)	(60,5)	(70,5)
predicted mean	24112.68	23445.90	22326.02	20402.28
measured mean	24111.22	23447.62	22313.94	20401.95
ratio meas./pred.	1.00	1.01	1.00	1.01
predicted std. dev.	115.93	209.86	367.19	619.10
measured std. dev.	107.87	198.67	360.51	608.91
ratio meas./pred.	0.93	0.95	0.98	0.98
predicted std. dev. in [30]	109.25	178.91	287.89	451.70
ratio meas./pred.	0.99	1.11	1.25	1.35

Table 2. Comparison of experimental and predicted mean and standard deviation of the correct-guess score in the dual attack with modulus switching ($p = 5$).

Dimension $(n_{\text{lat}}, n_{\text{fft}})$	(40,5)	(50,5)	(60,5)	(70,5)
predicted mean	14938.50	14525.41	13831.62	12674.79
measured mean	15002.02	14511.65	13740.67	12639.81
ratio meas./pred.	1.00	1.00	0.99	0.99
predicted std. dev.	4151.55	4044.94	3859.30	3601.86
measured std. dev.	4073.32	3999.62	3907.26	3488.97
ratio meas./pred.	0.98	0.99	1.01	0.97

Table 3. Comparison of experimental and predicted mean and standard deviation of the correct-guess score in the dual attack with decoding ($k_{\text{fft}} = 2$).

Dimension $(n_{\text{lat}}, n_{\text{fft}})$	(40,5)	(50,5)	(60,5)	(70,5)
predicted mean	23437.52	22848.76	21841.24	19892.10
measured mean	23290.87	22621.00	21472.27	19700.60
ratio meas./pred.	1.02	0.99	0.98	0.99
predicted std. dev.	594.65	646.74	708.19	795.38
measured std. dev.	620.50	636.55	737.91	810.10
ratio meas./pred.	1.04	0.98	1.04	1.02

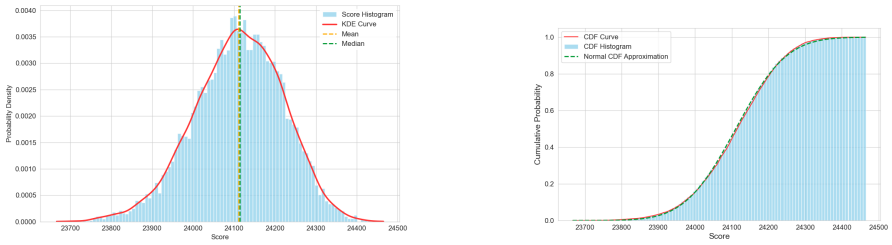

Fig. 2. Experimental results for the correct guess under the original dual attack with parameters $(n_{\text{lat}}, n_{\text{fft}}) = (40, 5)$. Left: KDE of the score distribution. Right: CDF together with its normal approximation using the exact mean and standard deviation.

We further observe that the term $\exp\left(2\Delta^2\ell^4\sigma_\theta^2\right)$ can become extremely large once the ratio $\Delta\ell^2\sigma_\theta$ exceeds 1. Consequently, for large ℓ, our predictive expression shows that the expected score for a correct guess decreases while its variance increases, significantly reducing the discriminative power between correct and incorrect guesses. This observation further implies that the dual attack becomes ineffective when the norm ℓ of the sampled dual vectors is excessively large.

4.2 A More Concrete Analysis of the MATZOV Attack

Building on the covariance-based variance estimate, we revisit the MATZOV attack without invoking the independence assumption. We first validate our model in a high-dimensional regime characterized by long sampled dual vectors, where the factor $\exp(2\Delta^2\ell^4\sigma_\theta^2)$ becomes large and the condition $\Delta\ell^2\sigma_\theta > 1$ is met. We set $n_{\text{lat}} = 160$, $n_{\text{fft}} = n_{\text{enu}} = 5$, $m = 170$, $p = 5$, $\beta_{\text{bkz}} = 60$, and $\beta_{\text{sieve}} = 80$, without restricting the number of output vectors in sieving. This yields $N = 89494$ and an average vector length of $\ell = 2852.31$. The measured mean and standard deviation of $F_{\mathcal{W}}^{(\text{ms})}(\mathbf{s}_{\text{enu}}, \mathbf{s}_{\text{fft}})$ are $\mu_{\text{ip}} = -2.04$ and $\sigma_{\text{ip}} = 212.20$, respectively, whereas our model predicts $\mu_{\text{pre}} = 0.00001$ and $\sigma_{\text{pre}} = 211.53$. The predicted standard deviation is very close to the experimental value, even for such a large ℓ. This provides a partial validation of our estimate in large-scale settings and also shows that when ℓ is too large, the standard deviation of the score can be far larger than its expectation.

We now follow the method of MATZOV [24] and give a more concrete statistical analysis. Let T be the threshold, $|\mathcal{W}| = N$, and let ϵ_1 and ϵ_2 denote the target probabilities that the correct-guess score falls below T and that incorrect-guess scores exceed T, respectively. Let $\mathbb{E}^{(\text{ms})}$ and $\text{Cov}^{(\text{ms})}$ denote the expectation of $f_{\mathbf{w}_1}^{(\text{ms})}$ and the pairwise covariance of $f_{\mathbf{w}_i}^{(\text{ms})}$. For the correct guess, we approximate the total score by a normal distribution with mean $N\mathbb{E}^{(\text{ms})}$ and variance $N^2\text{Cov}^{(\text{ms})}$, while for an incorrect guess we use a normal distribution with mean 0 and variance $N/2$.

For the correct-guess score to fall below T with probability ϵ_1, we require

$$\Pr[F_W^{(\text{ms})} < T] = 1 - \Phi\left(\frac{N\mathbb{E}^{(\text{ms})} - T}{N\sqrt{\text{Cov}^{(\text{ms})}}}\right) \leq \epsilon_1,$$

i.e.

$$\frac{N\mathbb{E}^{(\mathrm{ms})} - T}{N\sqrt{\mathrm{Cov}^{(\mathrm{ms})}}} \geq \Phi^{-1}(1-\epsilon_1) = \psi(\epsilon_1),$$

where $\Phi(\cdot)$ denotes the CDF of the standard normal distribution and $\psi(x) = \Phi^{-1}(1-x)$, so

$$\frac{T}{N} \leq \mathbb{E}^{(\mathrm{ms})} - \psi(\epsilon_1)\sqrt{\mathrm{Cov}^{(\mathrm{ms})}}.$$

Similarly, for an incorrect guess to exceed T with probability $\frac{\epsilon_2}{Rp^{n_{\mathrm{fft}}}}$, where R denotes the number of guesses, we require

$$\Pr[F_W^{(\mathrm{ms})} \geq T] = 1 - \Phi\left(\frac{T}{\sqrt{N/2}}\right) \leq \frac{\epsilon_2}{Rp^{n_{\mathrm{fft}}}},$$

i.e.,

$$\frac{T}{\sqrt{N}} \geq \frac{\psi(\epsilon_2/(Rp^{n_{\mathrm{fft}}}))}{\sqrt{2}}. \tag{10}$$

However, in the high-dimensional regime we have $\sqrt{\mathrm{Cov}^{(\mathrm{ms})}} \gg \mathbb{E}^{(\mathrm{ms})}$. The condition

$$\mathbb{E}^{(\mathrm{ms})} - \psi(\epsilon_1)\sqrt{\mathrm{Cov}^{(\mathrm{ms})}} > 0$$

can then hold only if $\psi(\epsilon_1) \leq 0$, which requires $\epsilon_1 \geq 1/2$. To satisfy this requirement, the optimal choice of the threshold is $T = N\mathbb{E}^{(\mathrm{ms})}$, which yields $\epsilon_1 \approx 0.5$, and is analogous to the strategy of Carrier et al. [10]. By constraining ϵ_2 through Eq. (10), where ϵ_2 is typically set to 0.1, one can obtain an estimate of N, which is then used to evaluate the MATZOV attack. The detailed time complexity evaluation and parameters are deferred to the appendix B, while the resulting complexity estimates are summarized in Table 4. Compared with MATZOV, our evaluation yields a lower time complexity and a higher success probability of approximately 0.4 versus 0.25. This improvement stems from our more refined estimation, which avoids an overly conservative choice of N and the need for re-randomization and multiple sampling, and from our recognition that the optimal threshold is $T = N\mathbb{E}^{(\mathrm{ms})}$.

4.3 Comparison of Our Predictive Model with [10]

We now apply our model to the attack of Carrier et al. [10]. Based on the analysis in the previous subsection, we consider $T = N\mathbb{E}[f_{\mathbf{w}}^{(\mathrm{lsc})}]$ to be a reasonable threshold, while the corresponding time complexity has already been established by Carrier et al. We therefore only report the results of our model under their recommended parameters for KYBER. As shown in Table 5, our results are almost identical to theirs, which further confirms the correctness of our model. Moreover, compared with the analysis of Carrier et al., our model is simpler in form and does not require high-precision numerical integration.

Table 4. Comparison of time complexity (in $\log_2$) between our model and MATZOV under the CC model for KYBER and DILITHIUM.

Scheme	This work			MATZOV (CC)
	C0	CN	CC	
KYBER-512	120.30	135.08	139.33	143.8
KYBER-768	176.71	190.51	195.00	200.5
KYBER-1024	242.69	255.43	260.41	266.0
DILITHIUM2	126.29	144.05	148.19	153.4
DILITHIUM3	183.16	199.30	204.24	210.5
DILITHIUM5	247.67	261.98	268.18	273.3

Table 5. Comparison of predicted result between our model and [10].

Parameters	KYBER-512	KYBER-768	KYBER-1024
(m, n)	(488,512)	(667,768)	(920,1024)
$(\beta_{\text{bkz}}, \beta_{\text{sieve}})$	(393,393)	(588,588)	(815,815)
$(n_{\text{enu}}, n_{\text{fft}}, n_{\text{lat}})$	(1,38,473)	(6,69,693)	(9,100,915)
$(\ell, \ell_{\text{lsc}})$	(2974.57,975.17)	(4401.02,1741.80)	(5190.00,2134.08)
$\log_2(T_{\text{pred}}/N)$ in [10]	-37.73	-57.33	-80.66
$\log_2(T_{\text{pred}}/N)$ in ours	-37.77	-57.57	-80.91
$\log_2(\sigma_{\text{pred}}/N)$ in ours	-35.50	-54.04	-75.90

5 An Extended Perspective on Tail Behavior

In this section, we extend the approach of Ducas and Pulles [13] to predict the score distribution of incorrect guesses in the dual attack. Recall that their analysis was formulated in the context of BDD. Although they discussed the relationship between β_{sieve} and the contradictory regime, their score prediction did not explicitly involve β_{sieve}; instead, it effectively corresponded to the case where the sieve runs on the full lattice, which was indeed the setting of their experiments. Consequently, their predictions matched their experimental data well. In practical dual attacks, however, due to complexity constraints, the sieve dimension is typically chosen to be strictly smaller than the full lattice dimension. This motivates us to extend the prediction framework to the case $\beta_{\text{sieve}} < n$. As we show, this generalization is far from trivial: the lattice structure underlying LWE-based dual attacks is not a dual lattice in the strict sense, and moreover the dual of a sublattice is not a sublattice of the dual. We provide rigorous theoretical derivations and present experimental comparisons for validation.

5.1 Ducas and Pulles' Perspective on Tail Behavior

We begin by reviewing the work of Ducas and Pulles. Consider a BDD instance ($\mathbf{B} \in \mathbb{Z}^{n\times n}, \mathbf{t} = \mathbf{v} + \mathbf{e}$), where $\mathbf{v} \in \Lambda(\mathbf{B})$ and $\mathbf{e}$ is a short vector satisfying $\|\mathbf{e}\| < \lambda_1(\Lambda(\mathbf{B}))/2$.

Ducas and Pulles performed lattice reduction on the dual lattice $\Lambda(\mathbf{B})^\vee$ and then obtained a list $\mathcal{W}$ of short dual vectors by running a sieve algorithm. For any $\mathbf{w} \in \mathcal{W}$, one has $\langle \mathbf{w}, \mathbf{t} \rangle \equiv \langle \mathbf{w}, \mathbf{e} \rangle \bmod 1$. They then considered the score function defined as

$$F^{\mathrm{bdd}}_{\mathcal{W}}(\mathbf{t}) = \sum_{\mathbf{w} \in \mathcal{W}} \cos\bigl(2\pi \langle \mathbf{w}, \mathbf{t} \rangle\bigr).$$

On the one hand, they argued that any target $\mathbf{t}$ can be reduced modulo the lattice so that its norm satisfies $\|\mathbf{t}\| \le \mathrm{GH}(\Lambda(\mathbf{B}))$. Since for any lattice point $\mathbf{v}_1$ one has $F^{\mathrm{bdd}}_{\mathcal{W}}(\mathbf{t}) = F^{\mathrm{bdd}}_{\mathcal{W}}(\mathbf{t} + \mathbf{v}_1)$, they considered the norm of the reduced target (which lies inside the Voronoi cell [20]) as the length r of the target, with $r \le \mathrm{GH}(\Lambda(\mathbf{B}))$. For a target $\mathbf{t}$ of length r, they derived estimates for the expectation $\mathbb{E}(r)$ and variance $\mathbb{V}(r)$ of the corresponding score, and assumed that the score follows a normal distribution depending only on r. Let F_r denote the score of a target $\mathbf{t}$ of length r. Then, for $\mathbf{t} \leftarrow \mathcal{D}$ where $\mathcal{D}$ is some distribution, they obtained the following cumulative distribution function:

$$\Pr_{\mathbf{t} \leftarrow \mathcal{D}}\left(F^{\mathrm{bdd}}_{\mathcal{W}}(\mathbf{t}) < x\right) = \sum_r \Pr(F_r < x) \cdot \Pr_{\mathbf{t} \leftarrow \mathcal{D}}(\|\mathbf{t}\| = r) = \int \Phi\left(\frac{x - \mathbb{E}(r)}{\sqrt{\mathbb{V}(r)}}\right) f_{\mathcal{D}}(r)\,\mathrm{d}r.$$

And consequently, the survival function:

$$\Pr_{\mathbf{t} \leftarrow \mathcal{D}}\left(F^{\mathrm{bdd}}_{\mathcal{W}}(\mathbf{t}) \ge x\right) = 1 - \int \Phi\left(\frac{x - \mathbb{E}(r)}{\sqrt{\mathbb{V}(r)}}\right) f_{\mathcal{D}}(r)\,\mathrm{d}r, \tag{11}$$

where $\Phi(\cdot)$ denotes the CDF of the standard normal distribution, and $f_{\mathcal{D}}(r)$ is the PDF of the length $r = \|\mathbf{t}\|$ of the target vector $\mathbf{t} \leftarrow \mathcal{D}$.

On the other hand, their experiments revealed that for large thresholds T, the probability that the score of an incorrect guess exceeds T decays significantly more slowly than a normal distribution would predict; we refer to this phenomenon as the tail behavior. Ducas and Pulles explained this observation by noting that an incorrect guess can be viewed as a vector uniformly distributed in $\mathbb{R}^n/\Lambda(\mathbf{B})$. When such a uniform vector happens to lie particularly close to the lattice, its associated score becomes anomalously high. To predict the score distribution, they therefore analyzed the distribution of the distance $d(\mathbf{t}, \Lambda(\mathbf{B}))$ of a uniform target to the lattice. After further analysis and approximation, they adopted a spherical average approximation for the Voronoi cell [20]: namely, on average over a random lattice, the distance $d(\mathbf{t}, \Lambda(\mathbf{B}))$ follows the same distribution as the norm of a uniformly random point in a ball of radius $\mathrm{GH}(\Lambda(\mathbf{B}))$. In other words,

$$\Pr_{\mathbf{t} \leftarrow U(\mathbb{R}^n/\Lambda)}\bigl(d(\mathbf{t}, \Lambda) \le r\bigr) = \frac{r^n}{\mathrm{GH}(\Lambda(\mathbf{B}))^n},$$

and they directly used $r = d(\mathbf{t}, \Lambda(\mathbf{B}))$ as the length of the incorrect target to compute the expectation $\mathbb{E}(r)$ and variance $\mathbb{V}(r)$. In this case, the corresponding probability density function is given by $f_{U(\mathbb{R}^n/\Lambda(\mathbf{B}))}(r) = nr^{n-1}/\mathrm{GH}(\Lambda(\mathbf{B}))^n$. Based on the above discussion and Eq. (11), they then obtained the CDF of the score distribution for incorrect guesses.

5.2 Generalization to LWE-based Dual Attacks: Two Subtle Considerations

Based on the above ideas, we extend the framework to estimate the score distribution of incorrect guesses in the LWE-based dual attack. In doing so, two simple yet subtle issues deserve particular attention: (i) why one needs to consider the distance to $\Lambda(\mathbf{B})$, and (ii) why the dimension of the spherical average distribution is n.

First, the distance to $\Lambda(\mathbf{B})$ matters because the dual vectors are taken from $\Lambda(\mathbf{B})^\vee$, and when computing $\mathbb{E}(r)$ and $\mathbb{V}(r)$, the length r refers to the norm of the target $\mathbf{t}'$ after being reduced by a lattice vector. Second, the dimension of the spherical average distribution is n precisely because we are considering the distance to the full-rank lattice $\Lambda(\mathbf{B})$, whose rank is n.

With these observations in mind, we investigate how to choose an appropriate model for $f(r)$ when considering incorrect guesses in the LWE-based dual attack. A key step is to determine which lattice we should be working with. Recall that the dual vectors are obtained from a lattice Λ whose basis is given in Eq. (4). Following the same logic as above, we consider the dual lattice $\Lambda^\vee$. In fact, since

$$\frac{1}{q}\langle \mathbf{x}, \mathbf{A}\mathbf{s}_{\text{lat}} + \mathbf{e}\rangle = \frac{1}{q}\langle \mathbf{x}, \mathbf{e}\rangle + \langle \mathbf{y}, \mathbf{s}_{\text{lat}}\rangle = \frac{1}{q}\langle \mathbf{w}, (\mathbf{e}^\top, \mathbf{s}_{\text{lat}}^\top)^\top\rangle \bmod 1,$$

it suffices to consider $q\Lambda^\vee$.

However, in practice, $\mathcal{W}$ is not taken from Λ but rather from a β_{sieve}-dimensional sublattice of Λ, which we denote by Λ_2. Therefore, the lattice we should focus on is not $q \cdot \Lambda^\vee$, but rather $q \cdot \Lambda_2^\vee$. Let $m' = m + n_{\text{lat}}$, $\beta_{\text{sieve}} = \beta$, and let $(\mathbf{b}_1, \ldots, \mathbf{b}_{m'})$ be a basis of Λ obtained after running lattice basis reduction, where we reorder the basis vectors so that the sublattice Λ_2 corresponds to $(\mathbf{b}_{m'-\beta+1}, \ldots, \mathbf{b}_{m'})$. Let $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_{m'})$ be the dual basis of $(\mathbf{b}_1, \ldots, \mathbf{b}_{m'})$. By the theory of dual lattices,

$$\Lambda_2^\vee = \Lambda\big(\mathbf{D}_{[m'-\beta+1,\, m']}\big).$$

The detailed justification of the above two issues, the construction of the relevant lattice, and the choice of the distance are deferred to Appendix C.

In summary, we should consider the distance from $\mathbf{t}$ to the β_{sieve}-dimensional sublattice

$$q \cdot \Lambda(\mathbf{D}_{(m'-\beta_{\text{sieve}}+1,\, m')}).$$

Let its determinant be $\det_1$. Then its Gaussian heuristic is

$$\text{gh}_1 = \sqrt{\frac{\beta_{\text{sieve}}}{2\pi e}} \cdot \det_1^{1/\beta_{\text{sieve}}},$$

and the corresponding PDF $f(r)$ is given by

$$f(r) = \frac{\beta_{\text{sieve}}\, r^{\beta_{\text{sieve}}-1}}{\text{gh}_1^{\beta_{\text{sieve}}}}, \qquad r \in (0, \text{gh}_1). \tag{12}$$

As a verification, we further predict the score probability of incorrect guesses for the dual attack with decoding. For $\mathbf{h} \in \mathbb{Z}_q^{k_{\text{fft}}}$, there exists $\mathbf{s}'_{\text{fft}} \in \mathbb{Z}_q^{k_{\text{fft}}}$ satisfying

$$\langle \mathbf{u}, \mathbf{h}\rangle = \langle \mathbf{u}, \mathbf{G}^\top \mathbf{s}'_{\text{fft}}\rangle = \langle \mathbf{G}\mathbf{u}, \mathbf{s}'_{\text{fft}}\rangle = \langle \mathbf{A}_{\text{fft}}^\top \mathbf{x} - \mathbf{e}_{\text{lsc}}, \mathbf{s}'_{\text{fft}}\rangle = \langle \mathbf{x}, \mathbf{A}_{\text{fft}}\mathbf{s}'_{\text{fft}}\rangle - \langle \mathbf{e}_{\text{lsc}}, \mathbf{s}'_{\text{fft}}\rangle,$$

and then

$$\langle \mathbf{x}, \mathbf{A}_{\text{lat}}\mathbf{s}_{\text{lat}} + \mathbf{A}_{\text{enu}}\widetilde{\mathbf{s}_{\text{enu}}} + \mathbf{e}\rangle - \langle \mathbf{u}, \mathbf{h}\rangle = \langle \mathbf{x}, \mathbf{A}_{\text{lat}}\mathbf{s}_{\text{lat}} + \mathbf{A}_{\text{enu}}\widetilde{\mathbf{s}_{\text{enu}}} + \mathbf{A}_{\text{fft}}\mathbf{s}'_{\text{fft}} + \mathbf{e}\rangle + \langle \mathbf{e}_{\text{lsc}}, \mathbf{s}'_{\text{fft}}\rangle.$$

To minimize the quantity, it suffices to consider the shortest vector in the coset

$$\mathcal{S} = \{\mathbf{G}^{\perp}\mathbf{x} + \mathbf{z}_{\mathbf{h}} \mid \mathbf{x} \in \mathbb{Z}_q^{n_{\text{fft}}}\},$$

where $\mathbf{h} = \mathbf{G}^{\top}\mathbf{s}'_{\text{fft}} \bmod q$, the matrix $\mathbf{G}^{\perp}$ satisfies $\mathbf{G}^{\top}\mathbf{G}^{\perp} \equiv \mathbf{0} \bmod q$, and $\mathbf{z}_{\mathbf{h}}$ is any vector satisfying $\mathbf{G}^{\top}\mathbf{z}_{\mathbf{h}} \equiv \mathbf{h} \bmod q$. If $\mathbf{G}^{\perp}$ is regarded as a lattice basis, then the norm of the shortest vector in $\mathcal{S}$ equals the distance from $\mathbf{z}_{\mathbf{h}}$ to the lattice $\Lambda(\mathbf{G}^{\perp})$ with determinant $q^{k_{\text{fft}}}$. Since the term $\langle \mathbf{e}_{\text{lsc}}, \mathbf{s}'_{\text{fft}}\rangle$ is not dominant, we can fix this distance as

$$\text{gh}_2 = \sqrt{\frac{n_{\text{fft}}}{2\pi e}} \cdot \det(\Lambda(\mathbf{G}^{\perp}))^{\frac{1}{n_{\text{fft}}}} = \sqrt{\frac{n_{\text{fft}}}{2\pi e}} \cdot q^{\frac{k_{\text{fft}}}{n_{\text{fft}}}}.$$

For the example in [10, Section 4], the parameters are $m = 40$, $n_{\text{lat}} = 42$, $n_{\text{enu}} = 0$, $q = 241$, $n_{\text{fft}} = 8$, $k_{\text{fft}} = 3$. In this case, denote

$$\rho_1(r) \triangleq \exp\left(-2\pi^2 \frac{\ell^2 r^2}{(m + n_{\text{lat}})q^2}\right), \quad \rho_2(r) \triangleq \exp\left(-2\pi^2 \frac{\ell_{\text{lsc}}^2 r^2}{n_{\text{fft}} q^2}\right),$$

then we let

$$\begin{aligned}
\mathbb{E}(r) &= |\mathcal{W}|\rho_1(r)\rho_2(\text{gh}_2), \\
\mathbb{V}(r) &= |\mathcal{W}|\left(\frac{1}{2} - (\rho_1(r)\cdot\rho_2(\text{gh}_2))^2 + \frac{1}{2}(\rho_1(r)\cdot\rho_2(\text{gh}_2))^4\right) \\
&\quad + |\mathcal{W}|\left(|\mathcal{W}| - 1\right)\rho_1(r)^2\rho_2(\text{gh}_2)^2\left(\exp\left(\frac{8\pi^4}{q^4}\left(\frac{\ell^4 r^4 \sigma_\theta^2}{(m + n_{\text{lat}})^2} + \frac{\ell_{\text{lsc}}^4 \text{gh}_2^4 \sigma_{\text{lsc}}^2}{n_{\text{fft}}^2}\right)\right) - 1\right).
\end{aligned}$$

With the above discussion, we substitute the relevant expressions and Eq. (12) into Eq. (11) to obtain an estimate of the survival function of the score distribution for incorrect guesses.

In our experiments, we set

$$\det_1 = q^{\frac{m\cdot\beta_{\text{sieve}}}{m+n_{\text{lat}}}},$$

which is an average-sense estimate and is smaller than the true determinant of the sublattice. As shown in Fig. 3, the curve of our model lies above the experimental data precisely because we have underestimated both $\det_1$ and gh_1. Nevertheless, the two curves exhibit nearly identical trends, which suggests that if a more accurate estimate of $\det_1$ were available, our curve would coincide with the experimental results.

6 Conclusion and Future Work

This work presents a unified evaluation model for dual attacks on LWE. The model provides closed-form estimates for the expectation and variance of the score for correct

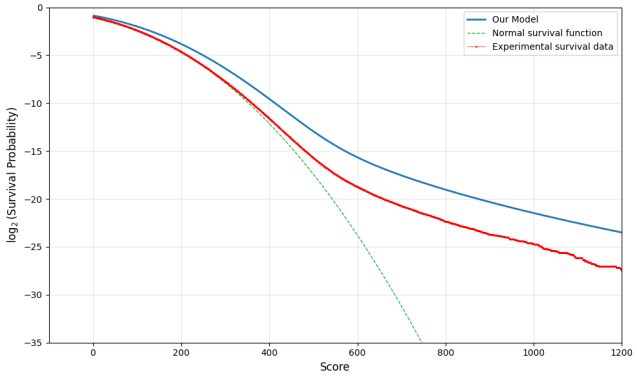

Fig. 3. Comparison of tail floor behavior between our model prediction and experimental data.

guesses, relying on two main ideas: (i) the cosine of the angle between short vectors follows a normal distribution, capturing geometric structure in covariance estimation; (ii) exploiting independence between vector components and the FFT perturbation, the expectation decomposes into a product of simple terms. To our knowledge, this is the first unified framework covering three types of dual attacks, revealing how each parameter affects the score distribution. The model achieves high predictive accuracy and can be used directly for complexity evaluation. Our experiments reveal clear performance differences among the three variants under identical settings. We further extend the tail-behavior characterization of Ducas and Pulles via conditional probability integration and subspace projection, demonstrated on the dual attack with decoding but naturally extendable to other LWE-based dual attacks. Several open issues remain for future work. (i) Ducas and Pulles [13] fix the norm of the target vector $\mathbf{t}$ and assume independence

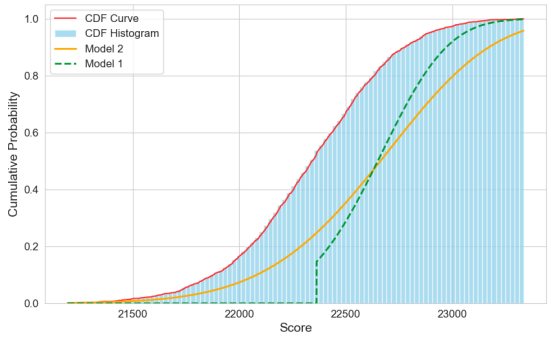

Fig. 4. Comparison of the predicted CDFs of Model 1 and Model 2 for the original dual attack ($(n_{\text{lat}}, n_{\text{fft}}) = (60, 5)$).

among individual scores under this conditioning. Our model, in contrast, provides expectation and variance estimates for individual correct-guess scores in dual attacks. This raises two natural questions. First, if we adopt the independence assumption of [13] and plug our estimates into the conditional probability integral, can we obtain the CDF of the total score? Second, if we directly use our variance estimates (which reflect dependence) in the same integral, does that give a correct CDF for the correct guess? We have compared both approaches on the original dual attack, and the results are illustrated in Fig. 4. Neither reproduces the true distribution faithfully. Under the independence assumption (Model 1), the variance is too small, leading to an underestimate of the probability of low scores. Under the dependence-sensitive model (Model 2), the variance is too large, leading to an underestimate of high scores. How to balance these two to recover the true distribution remains an open problem. (ii) Our prediction model for the tail behavior relies on accurate estimates of the determinant of the corresponding sublattice. In practice, one may use the determinant of its projected sublattice, but note that this is generally smaller than the true determinant of the original sublattice. Developing more precise methods for estimating this determinant would effectively improve the predictive accuracy of our model, and we leave this as an important direction for future work.

In summary, this paper advances the understanding of dual attacks, presents a unified analytical framework encompassing three types of dual attacks, and supplies precise estimates of the expectation and variance of the score for correct guesses. We hope that the perspective developed in this work provides a useful step toward a more accurate and systematic understanding of dual attacks, and contributes to the concrete security evaluation of lattice-based post-quantum cryptosystems.

References

1. Aharonov, D., Regev, O.: Lattice problems in np conp. J. ACM **52**(5), 749–765 (Sep 2005). https://doi.org/10.1145/1089023.1089025, https://doi.org/10.1145/1089023.1089025
2. Albrecht, M.R.: On dual lattice attacks against small-secret lwe and parameter choices in helib and seal. In: Coron, J.S., Nielsen, J.B. (eds.) Advances in Cryptology – EUROCRYPT 2017. pp. 103–129. Springer International Publishing, Cham (2017). https://doi.org/10.1007/978-3-319-56614-6_4
3. Albrecht, M.R., Ducas, L., Herold, G., Kirshanova, E., Postlethwaite, E.W., Stevens, M.: The general sieve kernel and new records in lattice reduction. In: Ishai, Y., Rijmen, V. (eds.) Advances in Cryptology – EUROCRYPT 2019. pp. 717–746. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-17656-3_25
4. Albrecht, M.R., Gheorghiu, V., Postlethwaite, E.W., Schanck, J.M.: Estimating quantum speedups for lattice sieves. In: Moriai, S., Wang, H. (eds.) Advances in Cryptology – ASIACRYPT 2020. pp. 583–613. Springer International Publishing, Cham (2020)
5. Albrecht, M.R., Shen, Y.: Quantum augmented dual attack. Cryptology ePrint Archive, Paper 2022/656 (2022), https://eprint.iacr.org/2022/656
6. Alkim, E., Ducas, L., Pöppelmann, T., Schwabe, P.: Post-quantum key Exchange—A new hope. In: 25th USENIX Security Symposium (USENIX Security 16). pp. 327–343. USENIX Association, Austin, TX (Aug 2016), https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/alkim

7. Arora, S., Ge, R.: New algorithms for learning in presence of errors. In: Aceto, L., Henzinger, M., Sgall, J. (eds.) Automata, Languages and Programming. pp. 403–415. Springer Berlin Heidelberg, Berlin, Heidelberg (2011), https://doi.org/10.1007/978-3-642-22006-7_34
8. Bashiri, K., Glaser, T., May, A., Nowakowski, J.: Super-quadratic quantum speed-ups and guessing many likely keys. In: Daemen, J., Thomé, E. (eds.) Advances in Cryptology – EUROCRYPT 2026. pp. 457–480. Springer Nature Switzerland, Cham (2026)
9. Becker, A., Ducas, L., Gama, N., Laarhoven, T.: New directions in nearest neighbor searching with applications to lattice sieving. In: Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms. p. 10–24. SODA '16, Society for Industrial and Applied Mathematics, USA (2016)
10. Carrier, K., Meyer-Hilfiger, C., Shen, Y., Tillich, J.P.: Assessing the impact of a variant of matzov's dual attack on kyber. In: Tauman Kalai, Y., Kamara, S.F. (eds.) Advances in Cryptology – CRYPTO 2025. pp. 444–476. Springer Nature Switzerland, Cham (2025), https://doi.org/10.1007/978-3-032-01855-7_15
11. Chen, Y., Nguyen, P.Q.: Bkz 2.0: Better lattice security estimates. In: Lee, D.H., Wang, X. (eds.) Advances in Cryptology – ASIACRYPT 2011. pp. 1–20. Springer Berlin Heidelberg, Berlin, Heidelberg (2011). https://doi.org/10.1007/978-3-642-25385-0_1
12. Ducas, L., Pulles, L.N.: Does the dual-sieve attack on learning with errors even work? In: Handschuh, H., Lysyanskaya, A. (eds.) Advances in Cryptology – CRYPTO 2023. pp. 37–69. Springer Nature Switzerland, Cham (2023), https://doi.org/10.1007/978-3-031-38548-3_2
13. Ducas, L., Pulles, L.N.: Accurate score prediction for dual-sieve attacks: L. ducas, ln pulles. Journal of Cryptology **39**(1), 8 (2026), https://doi.org/10.1007/s00145-025-09560-7
14. Ducas, L., Pulles, L.N., Stevens, M.: Towards a modern lll implementation. In: Hanaoka, G., Yang, B.Y. (eds.) Advances in Cryptology – ASIACRYPT 2025. pp. 65–99. Springer Nature Singapore, Singapore (2026). https://doi.org/10.1007/978-981-95-5099-9_3
15. Gallager, R.: Low-density parity-check codes. IRE Transactions on Information Theory **8**(1), 21–28 (1962). https://doi.org/10.1109/TIT.1962.1057683
16. Gama, N., Nguyen, P.Q., Regev, O.: Lattice enumeration using extreme pruning. In: Gilbert, H. (ed.) Advances in Cryptology – EUROCRYPT 2010. pp. 257–278. Springer Berlin Heidelberg, Berlin, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13190-5_13
17. Guo, Q., Johansson, T.: Faster dual lattice attacks for solving lwe with applications to crystals. In: Tibouchi, M., Wang, H. (eds.) Advances in Cryptology – ASIACRYPT 2021. pp. 33–62. Springer International Publishing, Cham (2021), https://doi.org/10.1007/978-3-030-92068-5_2
18. Kannan, R.: Improved algorithms for integer programming and related lattice problems. In: Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing. p. 193–206. STOC '83, Association for Computing Machinery, New York, NY, USA (1983). https://doi.org/10.1145/800061.808749, https://doi.org/10.1145/800061.808749
19. Karenin, A., Kirshanova, E., Nowakowski, J., May, A.: Fast slicer for batch-cvp: Making lattice hybrid attacks practical. In: Hanaoka, G., Yang, B.Y. (eds.) Advances in Cryptology – ASIACRYPT 2025. pp. 100–132. Springer Nature Singapore, Singapore (2026), https://doi.org/10.1007/978-981-95-5099-9_4
20. Laarhoven, T.: Approximate voronoi cells for lattices, revisited. Journal of Mathematical Cryptology **15**(1), 60–71 (2021). https://doi.org/doi:10.1515/jmc-2020-0074, https://doi.org/10.1515/jmc-2020-0074
21. Laarhoven, T., Walter, M.: Dual lattice attacks for closest vector problems (with preprocessing). In: Paterson, K.G. (ed.) Topics in Cryptology – CT-RSA 2021. pp. 478–

502. Springer International Publishing, Cham (2021), https://doi.org/10.1007/978-3-030-75539-3_20
22. Li, C., Wenger, E., Allen-Zhu, Z., Charton, F., Lauter, K.E.: Salsa verde: a machine learning attack on lwe with sparse small secrets. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (eds.) Advances in Neural Information Processing Systems. vol. 36, pp. 53343–53361. Curran Associates, Inc. (2023), https://proceedings.neurips.cc/paper_files/paper/2023/file/a75db7d2ee1e4bee8fb819979b0a6cad-Paper-Conference.pdf
23. Lyubashevsky, V., et al.: CRYSTALS-DILITHIUM. Technical report, National Institute of Standards and Technology (2022), https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022
24. MATZOV, Q.: Report on the security of lwe: Improved dual lattice attack (2022)
25. Micciancio, D., Regev, O.: Lattice-based Cryptography, pp. 147–191. Springer Berlin Heidelberg, Berlin, Heidelberg (2009). https://doi.org/10.1007/978-3-540-88702-7_5, https://doi.org/10.1007/978-3-540-88702-7_5
26. Pouly, A., Shen, Y.: Provable dual attacks on learning with errors. In: Joye, M., Leander, G. (eds.) Advances in Cryptology – EUROCRYPT 2024. pp. 256–285. Springer Nature Switzerland, Cham (2024), https://doi.org/10.1007/978-3-031-58754-2_10
27. Regev, O.: On lattices, learning with errors, random linear codes, and cryptography. J. ACM **56**(6) (Sep 2009). https://doi.org/10.1145/1568318.1568324, https://doi.org/10.1145/1568318.1568324
28. Schnorr, C.P.: Lattice reduction by random sampling and birthday methods. In: Alt, H., Habib, M. (eds.) STACS 2003. pp. 145–156. Springer Berlin Heidelberg, Berlin, Heidelberg (2003). https://doi.org/10.1007/3-540-36494-3_14
29. Schwabe, P., et al.: CRYSTALS-KYBER. Technical report, National Institute of Standards and Technology (2022), https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022
30. Wiemers, A., Bashiri, K.: On the independence heuristic in the dual attack. Cryptology ePrint Archive (2023). https://doi.org/10.1515/jmc-2024-0028

A Proofs of Lemmas

Proof of Lemma 1.

Proof. By the definition of expectation,

$$\mathbb{E}\left[\cos(tX)\right] = \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos(tX)\, dX = \frac{2\sin\left(\frac{t}{2}\right)}{t}.$$

This completes the proof.

Proof of Lemma 2.

Proof. Using the cosine angle addition formula, we have

$$\mathbb{E}\left[\cos(X+Y)\right] = \mathbb{E}\left[\cos(X)\cos(Y)\right] - \mathbb{E}\left[\sin(X)\sin(Y)\right].$$

Since X and Y are independent, it follows that

$$\begin{aligned}\mathbb{E}\left[\cos(X)\cos(Y)\right] &= \mathbb{E}\left[\cos(X)\right]\mathbb{E}\left[\cos(Y)\right],\\ \mathbb{E}\left[\sin(X)\sin(Y)\right] &= \mathbb{E}\left[\sin(X)\right]\mathbb{E}\left[\sin(Y)\right].\end{aligned}$$

Moreover, the symmetry of the distributions of X and Y implies that

$$\mathbb{E}\left[\sin(X)\right] = \mathbb{E}\left[\sin(Y)\right] = 0.$$

Substituting these equalities gives the desired result.

Proof of Lemma 3.

Proof. By Lemmas 1 and 2,

$$\begin{aligned}\mathbb{E}\left[\cos(t(X+Y))\right] &= \mathbb{E}\left[\cos(tX)\right]\cdot\mathbb{E}\left[\cos(tY)\right]\\ &= \left(\frac{2\sin(\frac{t}{2})}{t}\right)^2 = \frac{2(1-\cos(t))}{t^2}.\end{aligned}$$

Proof of Lemma 4.

Proof. By the definition of expectation,

$$\mathbb{E}\left[e^{tX}\right] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,\mathrm{d}x,$$

where $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ is the probability density function of $X \sim \mathcal{N}(\mu, \sigma^2)$. Substituting $f_X(x)$ into the integral gives

$$\mathbb{E}\left[e^{tX}\right] = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} \exp\left(tx - \frac{(x-\mu)^2}{2\sigma^2}\right)\mathrm{d}x.$$

Completing the square in the exponent, we obtain

$$tx - \frac{(x-\mu)^2}{2\sigma^2} = -\frac{\left(x-(\mu+t\sigma^2)\right)^2}{2\sigma^2} + t\mu + \frac{t^2\sigma^2}{2}.$$

Therefore,

$$\mathbb{E}\left[e^{tX}\right] = \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right) \cdot \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{\left(x-(\mu+t\sigma^2)\right)^2}{2\sigma^2}\right) \mathrm{d}x.$$

The integral is equal to 1, since it is the integral of the density function of a normal random variable with mean $\mu + t\sigma^2$ and variance σ^2. Hence,

$$\mathbb{E}\left[e^{tX}\right] = \exp\left(t\mu + \frac{t^2\sigma^2}{2}\right).$$

This completes the proof.

Proof of Lemma 6.

Proof. Following the proof in Section 5 of [21], we have

$$\langle \mathbf{w}, \mathbf{e}' \rangle \sim \mathcal{N}\Big(0, \frac{\|\mathbf{w}\|^2 \|\mathbf{e}'\|^2}{m + n_{\text{lat}}}\Big).$$

We approximate $\|\mathbf{e}'\|^2/(m+n_{\text{lat}}) \approx \sigma_e^2$. Then

$$\frac{1}{q}\langle \mathbf{w}, \mathbf{e}' \rangle \bmod 1 \sim \mathcal{N}\Big(0, \frac{\|\mathbf{w}\|^2 \sigma_e^2}{q^2}\Big) \bmod 1.$$

By Lemma 5 of [21], for a Gaussian random variable $Z \sim \mathcal{N}(0, \sigma^2)$ and $X = [Z \bmod 1] \in [-1/2, 1/2)$, we have

$$\mathbb{E}[\cos(2\pi X)] = \rho, \qquad \mathbb{V}[\cos(2\pi X)] = \frac{1}{2} - \rho^2 + O(\rho^4),$$

where $\rho = \exp(-2\pi^2\sigma^2)$. Setting $\rho(\|\mathbf{w}\|) = \exp\big(-2\pi^2 \frac{\|\mathbf{w}\|^2\sigma_e^2}{q^2}\big)$ and substituting $\frac{\langle \mathbf{w}, \mathbf{e}' \rangle}{q}$ as Z, we obtain

$$\mathbb{E}[f_{\mathbf{w}}] = \rho(\|\mathbf{w}\|), \qquad \mathbb{V}[f_{\mathbf{w}}] = \frac{1}{2} - \rho(\|\mathbf{w}\|)^2 + O\big(\rho(\|\mathbf{w}\|)^4\big).$$

To see the variance more explicitly, note that

$$\mathbb{V}[\cos(2\pi X)] = \mathbb{E}[\cos^2(2\pi X)] - \mathbb{E}[\cos(2\pi X)]^2 = \mathbb{E}\big[\tfrac{1}{2} + \tfrac{1}{2}\cos(4\pi X)\big] - \rho^2.$$

Approximating $2X \sim \mathcal{N}(0, 4\sigma^2) \bmod 1$, we have $\mathbb{E}[\cos(4\pi X)] \approx \rho^4$. Inserting this gives the approximate expression $\mathbb{V}[f_{\mathbf{w}}] \approx \frac{1}{2} - \rho(\|\mathbf{w}\|)^2 + \frac{1}{2}\rho(\|\mathbf{w}\|)^4$.

B Security Evaluation of KYBER and DILITHIUM

Based on the analysis in Section 4.2, we evaluate the security of KYBER and DILITHIUM. The time complexity model largely follows Theorem 7.6 of MATZOV [24], namely

$$T_{\text{total}} = T_{\text{sample}} + R\,(T_{\text{fft}} + T_{\text{table}})\,,$$

where T_{sample} is the time complexity of lattice reduction and sieving, determined by β_{bkz} and β_{sieve}. Similar to Carrier et al. [10], we consider three cost models for lattice reduction: C0, CC, and CN. They differ in how the cost of lattice reduction is estimated, and we adopt the same conventions as in [5].

- **C0.** The "Core-SVP" model [6], where a single SVP call is assumed to suffice and all lower-order terms are ignored.
- **CC.** A classical circuit model [4,24], giving the most detailed estimates currently available, though memory access is not accounted for. It is termed "list_decoding-classical" in [4].
- **CN.** A query model included for completeness, termed "list_decoding-naive_classical" in [4].

All three models employ [9] as the sieving oracle. R denotes the number of guesses. According to [8], using $2^{n_{\text{enu}}H(\chi_e)}$ guesses recovers the correct solution with probability greater than $1/2$, where $H(\cdot)$ is the Shannon entropy. This matches the parameters used in MATZOV. To further increase the probability of finding the correct solution, we set $R = 2^{n_{\text{enu}}(H(\chi_e)+1)}$.

For the FFT step, we follow Carrier et al. [10] and set

$$T_{\text{fft}} = \left(C_{\text{add}}N_{\text{add}}^{\text{FFT}}(p) + C_{\text{mul}}N_{\text{mul}}^{\text{FFT}}(p)\right) \cdot n_{\text{fft}} \cdot p^{n_{k_{\text{fft}}+1}},$$

where $N_{\text{add}}^{\text{FFT}}(p)$ and $N_{\text{mul}}^{\text{FFT}}(p)$ are the numbers of additions and multiplications required by an FFT over $\mathbb{Z}_p$, and the costs of one addition and one multiplication are $C_{\text{add}} = 160$ and $C_{\text{mul}} = 1024$.

The time complexity of updating the FFT table for each guess is denoted by T_{table}. Following MATZOV [24], we set

$$T_{\text{table}} = 4 \cdot N \cdot C_{\text{add}},$$

where N is the number of short vectors required, which can be computed from Eq. (10). We usually fix the error probability ϵ_2 to 0.1. Since the computed N is usually smaller than the number of vectors output by the sieve, re-randomization and multiple sampling as in MATZOV [24] are not required. We evaluate the time complexity of the MATZOV attack for the KYBER and DILITHIUM parameter sets, with p ranging from 3 to 14. The values of $N_{\text{add}}^{\text{FFT}}(p)$ and $N_{\text{mul}}^{\text{FFT}}(p)$ are given in Table 6. The complete parameters and complexity estimates for KYBER and DILITHIUM are summarized in Table 7. The computation of $N_{\text{add}}^{\text{FFT}}(p)$ and $N_{\text{mul}}^{\text{FFT}}(p)$ is performed using FFTW[1], in the same way as in Carrier et al. [10].

[1] The software is available at `https://www.fftw.org/`. Its usage follows the same procedure as the open-source code of [10]; the numerical values may vary slightly across different machines.

Table 6. FFT complexity coefficients for p from 3 to 14.

p	$N_{\text{add}}^{\text{FFT}}$	$N_{\text{mul}}^{\text{FFT}}$	p	$N_{\text{add}}^{\text{FFT}}$	$N_{\text{mul}}^{\text{FFT}}$
3	12	4	9	80	40
4	16	0	10	84	24
5	30	12	11	150	100
6	36	8	12	96	16
7	60	36	13	176	68
8	52	4	14	148	60

C Details of the Tail-Behavior Analysis

We provide the detailed derivation omitted in Section 5. First, the distance to $\Lambda(\mathbf{B})$ matters because the dual vectors are taken from $\Lambda(\mathbf{B})^\vee$, and when computing $\mathbb{E}(r)$ and $\mathbb{V}(r)$, the length r refers to the norm of the target $\mathbf{t}'$ after being reduced by a lattice vector. Second, the dimension of the spherical average distribution is n precisely because we are considering the distance to the full-rank lattice $\Lambda(\mathbf{B})$, whose rank is n.

Following the same logic as above, we consider the dual lattice $\Lambda^\vee$. In fact, since

$$\frac{1}{q}\langle \mathbf{x}, \mathbf{A}\mathbf{s}_{\text{lat}} + \mathbf{e}\rangle = \frac{1}{q}\langle \mathbf{x}, \mathbf{e}\rangle + \langle \mathbf{y}, \mathbf{s}_{\text{lat}}\rangle = \frac{1}{q}\langle \mathbf{w}, (\mathbf{e}^\top, \mathbf{s}_{\text{lat}}^\top)^\top\rangle \bmod 1,$$

it suffices to consider $q\Lambda^\vee$. Let Λ_1 be the lattice generated by the following basis matrix:

$$\begin{bmatrix} q\mathbf{I}_m & -\mathbf{A}_{\text{lat}} \\ \mathbf{0} & \mathbf{I}_{n_{\text{lat}}} \end{bmatrix}.$$

One can observe that $\Lambda_1 = q \cdot \Lambda^\vee$. In particular,

$$\begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{A}_{\text{lat}}^\top & q\mathbf{I}_{n_{\text{lat}}} \end{bmatrix}^\top \cdot \begin{bmatrix} q\mathbf{I}_m & -\mathbf{A}_{\text{lat}} \\ \mathbf{0} & \mathbf{I}_{n_{\text{lat}}} \end{bmatrix} = \begin{bmatrix} q\mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & q\mathbf{I}_{n_{\text{lat}}} \end{bmatrix}.$$

Hence, it seems that we ought to consider the distance of the target to Λ_1.

There is an alternative interpretation of Λ_1. Consider an incorrect guess

$$\mathbf{t} = \mathbf{A}_{\text{lat}}\mathbf{s}_{\text{lat}} + \mathbf{A}_{\text{enu}}\widetilde{\mathbf{s}_{\text{enu}}} + \mathbf{A}_{\text{fft}}\widetilde{\mathbf{s}_{\text{fft}}} + \mathbf{e}.$$

There exists $\hat{\mathbf{s}}_{\text{lat}}$ such that

$$\mathbf{t} = \mathbf{A}_{\text{lat}}\hat{\mathbf{s}}_{\text{lat}} + \hat{\mathbf{e}},$$

where $\hat{\mathbf{e}}$ is a relatively short vector, yielding

$$\langle \mathbf{x}, \mathbf{t}\rangle = \langle \mathbf{x}, \hat{\mathbf{e}}\rangle + \langle \mathbf{y}, \hat{\mathbf{s}}_{\text{lat}}\rangle \equiv \langle \mathbf{w}, (\hat{\mathbf{e}}^\top, \hat{\mathbf{s}}_{\text{lat}}^\top)^\top\rangle \bmod 1.$$

Here $(\hat{\mathbf{e}}^\top, \hat{\mathbf{s}}_{\text{lat}}^\top)^\top$ can be viewed as the distance vector from $(\mathbf{t}^\top, \mathbf{0}^\top)^\top$ to the lattice Λ_1. In particular, since a vector $(\mathbf{x}_1^\top, \mathbf{y}_1^\top)^\top$ in Λ_1 satisfies $\mathbf{x}_1 \equiv -\mathbf{A}_{\text{lat}}\mathbf{y}_1 \bmod q$, there exists a lattice point

$$\left((-\mathbf{A}_{\text{lat}}\hat{\mathbf{s}}_{\text{lat}})^\top, \hat{\mathbf{s}}_{\text{lat}}^\top\right)^\top$$

Table 7. Complete parameters and complexity estimates for KYBER and DILITHIUM.

Scheme	Model	Optimal p	$\log_2(\text{comp})$	β_{bkz}	β_{sieve}	m	n_{enu}	n_{fft}
KYBER-512	C0	4	120.30	407	407	514	15	33
KYBER-512	CN	4	135.08	387	390	444	15	40
KYBER-512	CC	5	139.33	382	386	474	15	36
KYBER-768	C0	4	176.71	601	601	679	26	51
KYBER-768	CN	4	190.51	586	579	629	26	58
KYBER-768	CC	4	195.00	581	574	659	28	59
KYBER-1024	C0	4	242.69	824	824	856	36	75
KYBER-1024	CN	4	255.43	819	800	826	36	80
KYBER-1024	CC	4	260.41	814	795	836	38	81
DILITHIUM2	C0	6	126.29	429	429	1052	17	24
DILITHIUM2	CN	6	144.05	414	421	947	17	31
DILITHIUM2	CC	6	148.19	409	416	992	17	34
DILITHIUM3	C0	11	183.16	623	623	1370	18	38
DILITHIUM3	CN	12	199.30	613	609	1300	18	38
DILITHIUM3	CC	13	204.24	608	605	1345	18	38
DILITHIUM5	C0	6	247.67	842	842	1714	32	59
DILITHIUM5	CN	6	261.98	837	822	1694	32	64
DILITHIUM5	CC	6	268.18	837	822	1589	32	64

such that

$$(\hat{\mathbf{e}}^\top, \hat{\mathbf{s}}_{\text{lat}}^\top)^\top = (\mathbf{t}^\top, \mathbf{0}^\top)^\top + \left((-\mathbf{A}_{\text{lat}}\hat{\mathbf{s}}_{\text{lat}})^\top, \hat{\mathbf{s}}_{\text{lat}}^\top\right)^\top.$$

This also suggests that we should consider the distance from $(\mathbf{t}^\top, \mathbf{0}^\top)^\top$ to the lattice Λ_1.

However, in practice, $\mathcal{W}$ is not taken from $\Lambda^\vee$ but rather from a β_{sieve}-dimensional sublattice of $\Lambda^\vee$, which we denote by Λ_2. Therefore, the lattice we should focus on is not $q \cdot \Lambda^\vee$, but rather $q \cdot \Lambda_2^\vee$. Let $m' = m + n_{\text{lat}}$, $\beta_{\text{sieve}} = \beta$, and let $(\mathbf{b}_1, \ldots, \mathbf{b}_{m'})$ be a basis of Λ obtained after running lattice basis reduction, where we reorder the basis vectors so that the sublattice Λ_2 corresponds to $(\mathbf{b}_{m'-\beta+1}, \ldots, \mathbf{b}_{m'})$. Let $\mathbf{D} = (\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_{m'})$ be the dual basis of $(\mathbf{b}_1, \ldots, \mathbf{b}_{m'})$. By the theory of dual lattices, the vectors $(\mathbf{b}_{m'}, \ldots, \mathbf{b}_{m'-\beta+1})$ and

$$\left(\pi_{m'-\beta+1}(\mathbf{d}_{m'-\beta+1}), \ldots, \pi_{m'-\beta+1}(\mathbf{d}_{m'})\right) = \mathbf{D}_{[m'-\beta+1,\, m']}$$

form a pair of dual bases. Consequently,

$$\Lambda_2^\vee = \Lambda\left(\mathbf{D}_{[m'-\beta+1,\, m']}\right).$$

Observe that the length r in $\mathbb{E}(r)$ and $\mathbb{V}(r)$ refers to the norm of the vector obtained by reducing the target $\mathbf{t}$ by a lattice point in Λ_1. To clarify this, we expand $\mathbf{t}$ in the orthogonal basis $(q\mathbf{d}_1^*, q\mathbf{d}_2^*, \ldots, q\mathbf{d}_{m'}^*)$ (note that $q\mathbf{D}$ is also a basis of Λ_1):

$$\mathbf{t} = a_1 q\mathbf{d}_1^* + a_2 q\mathbf{d}_2^* + \cdots + a_{m'} q\mathbf{d}_{m'}^*,$$

where $a_i \in \mathbb{R}$. If we consider the distance vector $\mathbf{dis}_1$ from $\mathbf{t}$ to $q\Lambda_2^\vee$, then

$$\mathbf{dis}_1 = a_1 q\mathbf{d}_1^* + a_2 q\mathbf{d}_2^* + \cdots + a_{m'-\beta} q\mathbf{d}_{m'-\beta}^*,$$

and

$$\mathbf{t} - \mathbf{dis}_1 = a_{m'-\beta+1} q\mathbf{d}_{m'-\beta+1}^* + \cdots + a_{m'} q\mathbf{d}_{m'}^*,$$

which with very high probability does not belong to Λ_1. Hence, $\|\mathbf{dis}_1\|$ cannot be used as r in $\mathbb{E}(r)$ and $\mathbb{V}(r)$.

On the other hand, suppose we consider the distance vector $\mathbf{dis}_2$ from $\mathbf{t}$ to

$$q\Lambda((\mathbf{d}_{m'-\beta+1}, \ldots, \mathbf{d}_{m'})) = q\Lambda(\mathbf{D}_{(m'-\beta+1, m')}).$$

Then we can write

$$\begin{aligned}
\mathbf{t} &= c_1 q\mathbf{d}_1 + c_2 q\mathbf{d}_2 + \cdots + c_{m'} q\mathbf{d}_{m'} + \mathbf{e}_1, \\
\mathbf{dis}_2 &= c_1 q\mathbf{d}_1 + c_2 q\mathbf{d}_2 + \cdots + c_{m'-\beta} q\mathbf{d}_{m'-\beta} + \mathbf{e}_1, \\
\mathbf{t} - \mathbf{dis}_2 &= c_{m'-\beta+1} q\mathbf{d}_{m'-\beta+1} + \cdots + c_{m'} q\mathbf{d}_{m'} \in \Lambda_1,
\end{aligned}$$

where $c_i \in \mathbb{Z}$, $\mathbf{e}_1 \notin \Lambda_1$. Since $\mathbf{t} - \mathbf{dis}_2 \in \Lambda_1$, $\|\mathbf{dis}_2\|$ can be legitimately substituted into $\mathbb{E}(r)$ and $\mathbb{V}(r)$. Therefore, the dimension of the spherical average distribution coincides with the sublattice dimension, namely β_{sieve}.

In summary, we should consider the distance from $\mathbf{t}$ to the β_{sieve}-dimensional sublattice

$$q \cdot \Lambda(\mathbf{D}_{(m'-\beta_{\text{sieve}}+1,\, m')}).$$

Let its determinant be $\det_1$. Then its Gaussian heuristic is

$$\text{gh}_1 = \sqrt{\frac{\beta_{\text{sieve}}}{2\pi e}} \cdot \det_1^{1/\beta_{\text{sieve}}},$$

and the corresponding PDF $f(r)$ is given by Eq. (12) in the main text.

As a verification, we further predict the score probability of incorrect guesses for the dual attack with decoding. For $\mathbf{h} \in \mathbb{Z}_q^{k_{\text{fft}}}$, there exists $\mathbf{s}'_{\text{fft}} \in \mathbb{Z}_q^{k_{\text{fft}}}$ satisfying

$$\langle \mathbf{u}, \mathbf{h} \rangle = \langle \mathbf{u}, \mathbf{G}^\top \mathbf{s}'_{\text{fft}} \rangle = \langle \mathbf{Gu}, \mathbf{s}'_{\text{fft}} \rangle = \langle \mathbf{A}_{\text{fft}}^\top \mathbf{x} - \mathbf{e}_{\text{lsc}}, \mathbf{s}'_{\text{fft}} \rangle = \langle \mathbf{x}, \mathbf{A}_{\text{fft}} \mathbf{s}'_{\text{fft}} \rangle - \langle \mathbf{e}_{\text{lsc}}, \mathbf{s}'_{\text{fft}} \rangle,$$

and then

$$\langle \mathbf{x}, \mathbf{A}_{\text{lat}} \mathbf{s}_{\text{lat}} + \mathbf{A}_{\text{enu}} \widetilde{\mathbf{s}_{\text{enu}}} + \mathbf{e} \rangle - \langle \mathbf{u}, \mathbf{h} \rangle = \langle \mathbf{x}, \mathbf{A}_{\text{lat}} \mathbf{s}_{\text{lat}} + \mathbf{A}_{\text{enu}} \widetilde{\mathbf{s}_{\text{enu}}} + \mathbf{A}_{\text{fft}} \mathbf{s}'_{\text{fft}} + \mathbf{e} \rangle + \langle \mathbf{e}_{\text{lsc}}, \mathbf{s}'_{\text{fft}} \rangle.$$

To minimize the quantity, it suffices to consider the shortest vector in the coset

$$\mathcal{S} = \{ \mathbf{G}^\perp \mathbf{x} + \mathbf{z}_{\mathbf{h}} \mid \mathbf{x} \in \mathbb{Z}_q^{n_{\text{fft}}} \},$$

where $\mathbf{h} = \mathbf{G}^\top \mathbf{s}'_{\text{fft}} \bmod q$, the matrix $\mathbf{G}^\perp$ satisfies $\mathbf{G}^\top \mathbf{G}^\perp \equiv \mathbf{0} \bmod q$, and $\mathbf{z}_{\mathbf{h}}$ is any vector satisfying $\mathbf{G}^\top \mathbf{z}_{\mathbf{h}} \equiv \mathbf{h} \bmod q$. If $\mathbf{G}^\perp$ is regarded as a lattice basis, then the norm of the shortest vector in $\mathcal{S}$ equals the distance from $\mathbf{z}_{\mathbf{h}}$ to the lattice $\Lambda(\mathbf{G}^\perp)$ with determinant $q^{k_{\text{fft}}}$. Since the term $\langle \mathbf{e}_{\text{lsc}}, \mathbf{s}'_{\text{fft}} \rangle$ is not dominant, we can fix this distance as

$$\text{gh}_2 = \sqrt{\frac{n_{\text{fft}}}{2\pi e}} \cdot \det(\Lambda(\mathbf{G}^\perp))^{\frac{1}{n_{\text{fft}}}} = \sqrt{\frac{n_{\text{fft}}}{2\pi e}} \cdot q^{\frac{k_{\text{fft}}}{n_{\text{fft}}}}.$$

This completes the detailed derivation used in Section 5.