Distinguishing LWE Instances Using Fourier Transform: A Refined Framework and its Applications

Chunhuan Zhao[1], Zhongxiang Zheng[2*], Xiaoyun Wang[1,3], Guangwu Xu[3,4]

[1] Institute for Advanced Study, Tsinghua University, Beijing 100084, China
[2] Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
[3] Key Laboratory of Cryptologic Technology and Information Security, Ministry of Education, Shandong Universtiy, Qingdao 266237, China
[4] Department of Electrical Engineering and Computer Sciences, University of Wisconsin, Milwaukee, WI 53201, USA
[*] Corresponding authors

zhengzx13@tsinghua.org.cn

Abstract. As a fundamental tool in lattice-based cryptosystems, discrete Gaussian samplers play important roles in both efficiency and security of lattice-based schemes. Approximate discrete rounded Gaussian sampler, central binomial sampler and bounded uniform sampler are three types of error samplers that are commonly used in the designs of various schemes. However, known cryptanalytics about error samplers concentrate on their standard deviations and no analysis about distinct structures of distributions have been proposed. In this paper, we address this problem by considering the dual attack for LWE instances and investigating Fourier transforms of these distributions. We introduce the concept of local width which enables us to get a more detailed look of these distributions and the distinguish advantages. We make an analysis of dual attack for different distributions and provide a novel measure model to describe the differences. Within this refined framework, we also propose a novel type of error sampler which can achieve high efficiency, security as well as flexibility.

Key words: discrete Gaussian sampling, lattice, distinguish advantage , LWE, dual attack

1 Introduction

With the rapid developments in quantum algorithms and computations, research in lattice-based cryptography has attracted considerable attention because lattice-based cryptosystems are likely to be effective against quantum computing attacks in the future. Mathematical and computational properties of lattices also provide basis for various advanced schemes, such as digital signatures, identity-based and attribute-based encryption, zero-knowledge proof and fully homomorphic schemes.

The learning with errors (LWE) problem introduced in Regev's work [25] is one of the most popular average-case problems that have been widely studied. Plenty of lattice-based cryptosystems, such as PKE schemes, KEM schemes and KEX schemes [3, 20, 22], are based on LWE problem or its variants such as Ring-LWE [21, 23] and Modula-LWE [8, 17]. In a LWE/RLWE/MLWE-based cryptosystem, the discrete Gaussian sampler works as a basic module which not only influences the efficiency of the whole scheme but also directly affects the decryption failure probability and the securities against known attacks such as primal attack, dual attack, BKW attack as well as algebraic attack [1, 3, 4, 11, 13, 14, 16]. According to results of [25], a LWE scheme which has a discrete Gaussian error sampler with large enough width (standard deviation) enjoys the worst case hardness. However, the scheme is not quite practical because other parameters should also be quite large in order to match the sampler's width. As a result, how to make a good balance between efficiency and security has become a key issue in designing LWE-based cryptosystems. One common way to achieve security with smaller parameters is restricting the number of available samples to avoid BKW attack and algebraic attack where a large number of samples are needed. Under the condition that only a limited number of samples are available, the primal attack and dual attack are usually considered [3, 20, 22].
In practice, three types of error samplers are commonly used in lattice-based schemes, namely rounded discrete Gaussian sampler, central binomial sampler and bounded uniform sampler. The current analysis uses width parameters (which can also be computed with standard deviations by multiplying a constant) to measure the security of these three error samplers, no attack that deals with the structures of error distributions has been ever considered according to [3] In this paper, we study the distributions for sampling errors by means of Fourier analysis. Fourier transform is a powerful tool in analyzing practical distributions. For these distributions, the values of their Fourier transform can be precisely computed in polynomial time. Because the isomorphic property of Fourier transform, these values can be used to reveal full information of the distributions. Therefore, some of the natures of these distributions can be viewed from a different angle. This provides an effective method to measure the differences brought by practical distributions when used in a LWE scheme instead of ideal discrete Gaussian distribution.

In this paper, our contribution can be summarized into three aspects. Firstly, we make use of Fourier transform further by exploring the distinguishing behavior to the components level of the dual lattice vector $\mathbf{v}$. Utilizing this analysis, we are able to study some unique features of an individual sampler. It is shown that the distinguish advantage of ideal discrete Gaussian distribution is related to the length of vectors in the dual lattice and the width s. Differences for approximate rounded Gaussian distribution, central binormal distribution and bounded uniform distribution are displayed. Our results indicate that the differences of the distinguish abilities of vectors with the same length in the dual lattice may be quite large according to the concrete distribution. Secondly, we make an analysis of dual attack for different distributions and provide a new measure model to describe the difference between practical distributions and ideal Gaussian distribution. The results show that the central binomial sampler used in NewHope shares the same property with ideal Gaussian sampler under the measure model while the approximate discrete rounded Gaussian sampler used in Frodo and the bounded uniform sampler used in Saber have gaps compared with the ideal one. Thirdly, we propose a novel type of sampler named mixed sampler which shares good property with ideal Gaussian and central binomial distributions. This sampler outputs a convolution distribution of central binomial distributions and bounded uniform distributions where more flexible choices in sampling widths are allowed, compared to that for the central binomial sampler. Furthermore, by choosing parameters properly, the mixed sampler can also achieve better efficiency and security compared with former samplers.

The rest of the paper is organized as follows. In Section 2, we introduce some background about lattice, discrete Gaussian sampling , LWE problem and dual attack. Our analysis of the distinguish advantage by using Fourier transform and their proofs are presented in Section 3. In Section 4, some applications of the above analysis are described, including a new measure of practical distributions under dual attack and a new sampler. Finally, we give our conclusion in Section 5.

2 Preliminaries

For $x \in R$, let $\lfloor x \rfloor$ be the maximum value among all the integers that are smaller than x, let $\lceil x \rfloor$ be the nearest integer to x.

2.1 Lattice

An m-dimensional *lattice* is a discrete additive subgroup in $\mathbb{R}^m$ which can be represented as the set of linear combination of n linearly independent vectors $\{\mathbf{b}_1, \cdots, \mathbf{b}_n\}$, i.e.

$$\mathcal{L}(\mathbf{B}) = \big\{ \sum_{i=1}^{n} x_i \mathbf{b}_i \mid x_i \in \mathbb{Z}, \forall i \in [1, n] \big\}$$

where $\mathbf{B} = [\mathbf{b}_1, \cdots, \mathbf{b}_n]$ is called a *basis* of $\mathcal{L}$ which is not unique, $n (n \leqslant m)$ is the *rank* of the lattice, a lattice is called full-rank if $m = n$. The determinant of $\mathcal{L}$ is defined as

$$\det(\mathcal{L}) = \sqrt{\det(\mathbf{B}^\top \mathbf{B})}.$$

The quantity $\det(\mathcal{L})$ is invariant regardless of the choice of $\mathbf{B}$. The *dual lattice* $\mathcal{L}^*$ is defined as

$$\mathcal{L}^* = \{\mathbf{w} \in \mathbb{R}^m \mid \forall \mathbf{v} \in \mathcal{L}, \langle \mathbf{w}, \mathbf{v} \rangle \in \mathbb{Z}\}.$$

q-ary lattice As a kind of important lattices in lattice-based cryptography, a *q-ary* lattice refers to the lattice such that $q\mathbb{Z}^n \subseteq \mathcal{L} \subseteq \mathbb{Z}^n$ where q is an integer.

Two types of q-ary lattices frequently used in lattice cryptography are defined as follows with respect to an $n \times m$ matrix $\mathbf{B} \in \mathbb{Z}_q^{n\times m}$,

$$\begin{aligned} \mathcal{L}_q(\mathbf{B}) &= \{\mathbf{y} \in \mathbb{Z}^m \mid \mathbf{y} = \mathbf{B}^\top \mathbf{x} \bmod q, \mathbf{x} \in \mathbb{Z}^n\}, \\ \mathcal{L}_q^\perp(\mathbf{B}) &= \{\mathbf{y} \in \mathbb{Z}^m \mid \mathbf{B}\mathbf{y} = 0 \bmod q\}. \end{aligned}$$

2.2 Gaussian Distribution over Lattices

For $s > 0$, the Gaussian function is defined as

$$\rho_s(\mathbf{y}) = e^{-\pi \|\mathbf{y}\|^2/s^2}$$

for $\mathbf{y} \in \mathbb{R}^m$ where s is called the *width*. When $s = 1$, the subscript is usually omitted for simplicity.

Definition 2.1 (Discrete Gaussian distribution). *For $s > 0$ and $\mathbf{c} \in \mathbb{R}^m$, the discrete Gaussian distribution $D_{\mathcal{L}+\mathbf{c},s}$ over $\mathcal{L} + \mathbf{c}$ is defined as*

$$D_{\mathcal{L}+\mathbf{c},s}(\mathbf{x}) = \frac{\rho_s(\mathbf{x})}{\rho_s(\mathcal{L} + \mathbf{c})}$$

where $\mathbf{x} \in \mathcal{L}+\mathbf{c}$ and $\rho_s(\mathcal{L}+\mathbf{c}) = \sum_{\mathbf{x}\in\mathcal{L}+\mathbf{c}} \rho_s(\mathbf{x})$. We call $\sigma = s/\sqrt{2\pi}$ the standard deviation for $D_{\mathcal{L}+\mathbf{c},s}$.

It is difficult to calculate the sum $\rho_s(\mathcal{L})$ directly, but it is related to the sum of values of a Gaussian function over the dual lattice according to the celebrated Poisson summation formula.

Lemma 2.1 (Poisson summation formula [5]) *For an n-dimensional lattice $\mathcal{L}$, let $s > 0$ and $\mathbf{t} \in \mathbb{R}^n$, the following hold:*

(1) $\rho_s(\mathcal{L}) = \frac{s^n}{\det(\mathcal{L})} \rho_{1/s}(\mathcal{L}^*)$,

(2) $\rho_s(\mathcal{L} + \mathbf{t}) = \frac{s^n}{\det(\mathcal{L})} \sum\limits_{\mathbf{w}\in\mathcal{L}^*} e^{2\pi i \langle \mathbf{w}, \mathbf{t}\rangle} \rho_{1/s}(\mathbf{w})$.

There is a tail bound for the continuous Gaussian distribution and the discrete Gaussian distribution also has a similar property which was first proven by Banaszczyk [5]. The following is a refinement to the bound of Banaszczyk given in [26].

Lemma 2.2 (Tail bound [26]) *For an n-dimensional lattice $\mathcal{L}$ and a vector $\mathbf{t} \in \mathbb{R}^n$, let $s > 0$ and $c \geqslant 1/\sqrt{2\pi}$, we have*

$$\Pr_{X\sim D_{\mathcal{L}+\mathbf{t},s}} [\|X\| > cs\sqrt{n}] \leqslant (2\pi e c^2)^{n/2} e^{-\pi n c^2} \frac{\rho_s(\mathcal{L})}{\rho_s(\mathcal{L} + \mathbf{t})}.$$

2.3 LWE Problem

LWE was proposed by Regev [25] in 2005 and has been widely used in the construction of lattice-based cryptography. We first introduce some definitions in order to describe LWE problems.

Definition 2.2 (LWE distribution). *Let $n \geqslant 1$, $q \geqslant 2$ and χ be an error distribution over $\mathbb{Z}_q$, given a secret vector $\mathbf{s} \in \mathbb{Z}_q^n$, the LWE distribution $L_{\mathbf{s},\chi}$ over $\mathbb{Z}_q^n \times \mathbb{Z}_q$ is sampled by choosing $\mathbf{a} \sim U(\mathbb{Z}_q^n)$ and $e \sim \chi$, and outputting $(\mathbf{a}, b = \langle \mathbf{a}, \mathbf{s} \rangle + e \bmod q)$.*

The LWE problem has a search version and a decision version, which are defined as follows.

Definition 2.3 (Search-LWE). *Given m samples $(\mathbf{a}_i, b_i) \in \mathbb{Z}_q^n \times \mathbb{Z}_q$ that are independently sampled from $L_{\mathbf{s},\chi}$ with a fixed secret $\mathbf{s} \in \mathbb{Z}_q^n$, the goal of search-LWE is to find the secret vector $\mathbf{s}$.*

In the rest of our discussion, we denote $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ to be the matrix formed by m columns $\{\mathbf{a}_i\}_{i=1}^m$ and $\mathbf{b} = (b_1, b_2, \cdots, b_m)^\top \in \mathbb{Z}_q^m$, where $b_i = \langle \mathbf{a}_i, \mathbf{s} \rangle + e_i \bmod q$.

Definition 2.4 (Decision-LWE). *Given m independent samples $(\mathbf{A}, \mathbf{b}) \in \mathbb{Z}_q^{m\times n} \times \mathbb{Z}_q^m$ that follow either the LWE distribution $L_{\mathbf{s},\chi}$ with a fixed secret $\mathbf{s} \in \mathbb{Z}_q^n$ or the uniform distribution, the goal of decision-LWE is to decide which distribution the samples follow.*

To make LWE more practical in cryptography, variants of LWE problems (e.g., Ring-LWE and Modulo-LWE) have been investigated. More details of these variants can be found in [18, 21]. Learning With Rounding (LWR) is another LWE variant with the determined error and defined as follows.

Definition 2.5 (LWR [7]). *For the integer parameters (n, q, p) where $n > 1$ and $q \geqslant p \geqslant 2$, given a secret vector $\mathbf{s} \in \mathbb{Z}_q^n$, the LWR distribution over $\mathbb{Z}_q^n \times \mathbb{Z}_q$ is sampled by choosing $\mathbf{a} \sim U(\mathbb{Z}_q^n)$ and outputting the sample $(\mathbf{a}, \lceil \frac{p}{q} \langle \mathbf{a}, \mathbf{s} \rangle \rfloor)$.*

Since

$$\frac{q}{p} \lceil \frac{p}{q} \langle \mathbf{a}, \mathbf{s} \rangle \rfloor = \langle \mathbf{a}, \mathbf{s} \rangle + e \bmod q$$

where

$$e = \frac{q}{p}(\frac{p}{q} \langle \mathbf{a}, \mathbf{s} \rangle - \lceil \frac{p}{q} \langle \mathbf{a}, \mathbf{s} \rangle \rfloor) \in (-\frac{q}{2p}, \frac{q}{2p}],$$

LWR is often viewed as the LWE problem with the error following the bounded uniform distribution [11]. Accordingly, LWR with structure are also commonly used in practice, such as Modulo-LWR and Ring-LWR.

2.4 Dual attack against decisional-LWE

The dual attack is to find a short vector $\mathbf{w}$ in the lattice $\mathcal{L}_q^{\perp}(\mathbf{A})$ and then make a distinguish. When $(\mathbf{A}, \mathbf{b})$ is a LWE sample, there is a distinguish advantage as $| \langle \mathbf{w}, \mathbf{b} \rangle |$ is small. The cost of obtaining a short vector and its corresponding distinguish advantage decide the whole complexity under dual attack.

As for the estimation of the complexity of obtaining a short vector, BKZ is usually used as it is the best performing algorithm in practical experiments. BKZ algorithm with block size b reduces the lattice basis by making use of SVP oracles in b dimension lattice iteratively. The cost of BKZ running time depends on the numbers of calls of SVP oracles which is known as polynomial [15]. As the polynomial factor is difficult to estimate, the most popular way is to adopt a very conservative approach by considering only one SVP oracle call in the iteration and take the "core-SVP" complexity as the estimation of cost of BKZ [3, 22] . Among various SVP oracle models, heuristic sieving algorithm is often considered in predicting the hardness of high dimensional lattice. Accordingly, the complexity is $2^{0.292b+o(b)}, 2^{0.265b+o(b)}, 2^{0.2075b+o(b)}$ which responds to the complexity of best current classical sieving, quantum sieving and the plausible sieving respectively and the factor in the $o(b)$ is ignored in estimation.

As for the lattice $\mathcal{L}_q^{\perp}(\mathbf{A})$, the length of vector outputted by BKZ algorithm with block size b is estimated as $l = \delta^{m-1} q^{\frac{n}{m}}$ where $\delta = ((\pi b)^{\frac{1}{b}} \frac{b}{2\pi e})^{\frac{1}{2(b-1)}}$. Since the sieving algorithm provides $2^{0.2075b}$ vectors, the whole complexity of dual attack is

$$2^{c_B b} \max\{1, \frac{1}{\epsilon^2} 2^{-0.2075b}\},$$

where $c_B = 0.292$ under classical computation and 0.265 under quantum computation. The distinguish advantage $\epsilon = e^{-\pi \frac{s^2 l^2}{q^2}}$ according to [19].

3 Fourier Transform and Dual Attack for LWE Instances

The duality in Fourier analysis is a fundamental mathematical thought in which a function localized in the time domain can be also viewed to spread out across the frequency domain. It has been shown to be very powerful for lattice theory, for example, the Fourier transform for discrete Gaussian and the corresponding Poisson summation formula (discussed in the previous section) are crucial for the improved transference bounds of lattice by Banaszczyk [5]. In this section, we will discuss the discrete Flourier transform over the Abelian group $\mathbb{Z}_q$ and use it to analyze several probability distributions over $\mathbb{Z}_q$. This enables us to provide a refined framework for dual attacks for some LWE instances.

For a function $f : \mathbb{Z}_q \to \mathbb{C}$, its Fourier transform $\hat{f}$ is given by

$$\hat{f}(k) = \sum_{j=0}^{q-1} e^{-\frac{2\pi i j k}{q}} f(j), \tag{1}$$

for each $k \in \mathbb{Z}_q$. The transform is invertible, so f can be uniquely determined by $\hat{f}$. Among the properties of Fourier transform, the uncertainty principle of of some remanence to our discussion. Let $\text{supp}(g) = \{j \in \mathbb{Z}_q : g(j) \neq 0\}$ be the support of a function g, then the uncertainty principle of Donoho and Stark [10] for discrete Fourier transform over $\mathbb{Z}_q$ states that, for any function $f : \mathbb{Z}_q \to \mathbb{C}$,

$$|\text{supp}(f)||\text{supp}(\hat{f})| \geq q. \tag{2}$$

The essence of the uncertainty principle is saying that $|\text{supp}(f)|$ and $|\text{supp}(\hat{f})|$ cannot be both small. In our later discussion of dual attack, we hope $|\text{supp}(\hat{f})|$ is a big integer for certain non-uniform distribution. As we shall see later, some error distributions for LWE have probability functions f with small $|\text{supp}(f)|$, and the actual $|\text{supp}(\hat{f})|$ is even bigger than the theoretical estimation in (2).

3.1 Distinguishing Advantage for Discrete Gaussian

Dual attack and primal attack are popular methods for solving LWE problems, they are especially effective in the case of have only a limited number of samples. We will focus on the dual attack against LWE.

The aim of dual attack is to solve decision-LWE problem, i.e. to distinguish whether the m independent samples $(\mathbf{A}, \mathbf{b}) \in \mathbb{Z}_q^{n\times m} \times \mathbb{Z}_q^m$ are drawn from LWE distribution or the uniform distribution.

The procedure of making distinction is to choose a (non-zero) vector $\mathbf{v}$ in the q-ary lattice $\mathcal{L}_q^{\perp}(\mathbf{A}) = \{\mathbf{x} \in \mathbb{Z}_q^m \mid \mathbf{A}\mathbf{x} = 0 \bmod q\}$. It can be seen that $\langle \mathbf{v}, \mathbf{X} \rangle$ is uniformly distributed over $\mathbb{Z}_q$ if $\mathbf{X} \sim U(\mathbb{Z}_q^m)$. However, we have $\langle \mathbf{v}, \mathbf{X} \rangle = \langle \mathbf{v}, \mathbf{e} \rangle \bmod q$ when $\mathbf{X} \sim L_{s,\chi}$.

In order to get a numeric distinguish advantage of a distribution over the uniform distribution, we can use the Fourier transform. Write random variable $\mathbf{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$ where components $\mathbf{x}_i$ are sampled from $\mathbb{Z}_q$ independently and have the same probability function $f(x)$. We denote the distribution of $\langle \mathbf{v}, \mathbf{X} \rangle$ as $f_{\langle \mathbf{v}, \mathbf{X} \rangle}$. By using convolution and its transform, we see that

$$\hat{f}_{\langle \mathbf{v}, \mathbf{X} \rangle}(1) = \prod_{j=1}^{n} \hat{f}_{v_j \mathbf{x}_j}(1) = \prod_{j=1}^{n} \hat{f}(v_j) \tag{3}$$

It is obvious that $\hat{f}_{\langle \mathbf{v}, \mathbf{X} \rangle}(1) = 0$ if $\mathbf{X} \sim U(\mathbb{Z}_q^n)$ and $\mathbf{v} \neq 0$. However, $\hat{f}_{\langle \mathbf{v}, \mathbf{X} \rangle}(1) \neq 0$ when $\mathbf{x}_i \sim \chi$ for some other error distribution χ and for a suitable vector $\mathbf{v}$ (i.e., $\hat{f}(v_j) \neq 0$ for all components of $\mathbf{v}$). A positive lower bound of $|\hat{f}_{\langle \mathbf{v}, \mathbf{X} \rangle}(1)|$ can be regarded as a distinguishing advantage.

An ideal error distribution is the discrete Gaussian for $\mathbb{Z}_q$ whose definition is $D_{s,q}(x) = \frac{\sum_{t\in\mathbb{Z}} \rho_s(x+tq)}{\rho_s(\mathbb{Z})}$, where s is called the width (which is $\sqrt{2\pi}$ times the standard deviation). As mentioned in [19], the distinguishing advantage for this case is

$$\epsilon(\|\mathbf{v}\|) = e^{-\pi\frac{s^2\|\mathbf{v}\|^2}{q^2}}. \tag{4}$$

We include a proof of (4) by showing $e^{-\pi\frac{s^2\|\mathbf{v}\|^2}{q^2}} \leq (\widehat{D_{s,q}})_{\langle\mathbf{v},\mathbf{X}\rangle}$. The following result also produces a tighter upper bound, namely $(\widehat{D_{s,q}})_{\langle\mathbf{v},\mathbf{X}\rangle} \leq 2e^{-\pi\frac{s^2\|\mathbf{v}\|^2}{q^2}}$.

Lemma 3.1 *For the variable X sampled from discrete Gaussian distribution $D_{s,q}$, we have*

$$\widehat{D_{s,q}}(k) = \frac{\rho_{\frac{1}{s}}(\mathbb{Z} + \frac{k}{q})}{\rho_{\frac{1}{s}}(\mathbb{Z})}.$$

Moreover, for $k = 1, 2, \cdots, \lfloor q/2 \rfloor$,

$$e^{-\frac{\pi s^2 k^2}{q^2}} \leqslant \widehat{D_{s,q}}(k) \leqslant 2e^{-\frac{\pi s^2 k^2}{q^2}}.$$

Proof. By definition and the Poisson summation formula,

$$\begin{aligned}
\widehat{D_{s,q}}(k) &= \sum_{j=0}^{q-1} e^{-\frac{2\pi ijk}{q}} \frac{\sum_{t\in\mathbb{Z}} \rho_s(j+tq)}{\rho_s(\mathbb{Z})} = \frac{1}{\rho_s(\mathbb{Z})} \sum_{j=0}^{q-1}\sum_{t\in\mathbb{Z}} e^{-\frac{2\pi i(j+tq)k}{q}} \rho_s(j+tq) \\
&= \frac{1}{\rho_s(\mathbb{Z})} \sum_{x\in Z} e^{-2\pi ix\frac{k}{q}} \rho_s(x) = \frac{\rho_{\frac{1}{s}}(\mathbb{Z} + \frac{k}{q})}{\rho_{\frac{1}{s}}(\mathbb{Z})}.
\end{aligned}$$

As for the lower bound of function $\widehat{D_{s,q}}(k))$, we have

$$\begin{aligned}
\rho_{\frac{1}{s}}(\mathbb{Z} + \frac{k}{q}) &= e^{-\frac{\pi s^2 k^2}{q^2}} + e^{-\frac{\pi s^2 (q+k)^2}{q^2}} + e^{-\frac{\pi s^2 (-q+k)^2}{q^2}} + \cdots \\
&= e^{-\frac{\pi s^2 k^2}{q^2}} (1 + e^{-\pi s^2}(e^{-\frac{2\pi s^2 k}{q}} + e^{\frac{2\pi s^2 k}{q}}) + \cdots) \\
&\geq e^{-\frac{\pi s^2 k^2}{q^2}} (1 + 2e^{-\pi s^2} + 2e^{-4\pi s^2} + \cdots) = e^{-\frac{\pi s^2 k^2}{q^2}} \rho_{\frac{1}{s}}(\mathbb{Z}).
\end{aligned}$$

On the other hand, lemma 2.4 of [6] states that for any lattice $\mathcal{L}$ and any vector $\mathbf{u} \in \mathbb{R}^n$, $\sum_{\substack{\mathbf{x}\in\mathcal{L}+\mathbf{u} \\ |x_1|\geq t}} \rho(\mathbf{x}) \leq 2e^{-\pi t^2}\rho(\mathcal{L})$ holds. Let $\mathcal{L} = s\mathbb{Z}$ and $t = u = \frac{sk}{q}$, then $|x| \geq t$ is true for any $x \in \mathcal{L} + u$ since $k \leq \frac{q}{2}$. Therefore,

$$\rho_{\frac{1}{s}}(\mathbb{Z} + \frac{k}{q}) = \sum_{\substack{x\in\mathcal{L}+u \\ |x|\geq t}} \rho(x) \leq 2e^{-\pi t^2}\rho(\mathcal{L}) = 2e^{-\frac{s^2 k^2 \pi}{q^2}} \rho_{\frac{1}{s}}(\mathbb{Z}),$$

which gives an upper bound for $\widehat{D_{s,q}}(k)$.

We should remark that many recent LWE schemes do not use the discrete Gaussian but its alternatives. Approximate discrete Gaussian distribution, central binormal distribution and bounded uniform distribution are treated as discrete Gaussian with corresponding width ($\sqrt{2\pi}$ times the standard deviation). See [2].

(4) indicates that one needs to seek shorter vectors in the dual lattice in order to achieve bigger distinguish advantage. Let $C(\ell)$ denote the cost of obtaining (short) vectors $\mathbf{v}$ with length ℓ, then the whole cost for the dual attack against LWE is $\frac{C(\ell)}{\epsilon^2(\|\mathbf{v}\|)}$. This is based on the Chernoff-Hoeffding argument which implies that $\epsilon^2(\|\mathbf{v}\|)$ many of samples (of $\mathbf{v}$) will increase the advantage close to 1.

Examining (3), it can be seen that that the advantage is not entirely depending on the norm of $\mathbf{v}$. The values $\hat{f}(v_j)$ play more role in this matter. We will also describe an idea to increase the advantage by maximizing $|\hat{f}_{\langle \mathbf{v},\mathbf{X}\rangle}(t)|$ for $t \in \mathbb{Z}_q$ later.

3.2 Distinguishing Advantage for Other Alternatives

The above attack applies for LWE problem with ideal discrete Gaussian distributions, while when it comes to practical LWE cases, ideal samplers are not available due to the limitation of precisions and truncations. To achieve high efficiency, rounded discrete Gaussian distribution, central binomial distribution, and bounded uniform distribution are used to sample errors in some NIST PQC candidates. For example, Frodo is based on LWE with rounded discrete Gaussian sampler, NewHope is based on RLWE with central binormal distribution sampler, Saber is based on LWR problem [3, 12, 22] and CRYSTALS-KYBER is based on MLWE where the error distribution can be seen as the convolution of a central binormal distribution and a bounded uniform distribution [24].

One of the main purposes of this paper is to push the using of Fourier transform further by exploring the distinguishing behavior to the components level of the dual lattice vector $\mathbf{v}$. To this end, we first calculate Fourier transforms for the alternative error distributions over $\mathbb{Z}_q$ explicitly. Let us describe these distributions.

1. Rounded discrete Gaussian distribution with width s: the probability assignment is depending on a fix integer $0 < R \leq \frac{q}{2}$, the probability for $x \in \mathbb{Z}_q$ is $f(x) = \begin{cases} \frac{\Psi_s(x)}{\sum_{j=-R}^{R} \Psi_s(x)} & \text{if } |x| \leq R \\ 0 & \text{otherwise} \end{cases}$, where $\Psi_s(x) = \int_{x-\frac{1}{2}}^{x+\frac{1}{2}} \frac{\rho_s(t)}{s} dt$.
2. The central binormal distribution $B(h)$: for a positive integer$h$, a random variable X is said to be sampled from $B(h)$, if it is the convolution of h (independent) variables X_i over $\{-1, 0, 1\}$ with $Pr[X_i = 1] = Pr[X_i = -1] = 1/4, Pr[X_i = 0] = 1/2$. The width of this distribution is $s = \sqrt{h\pi}$.
3. Bounded uniform distribution: for integers $0 \leq a, b \leq \lfloor \frac{q}{2} \rfloor$, a general bounded uniform distribution is simply the uniform distribution for the set $\{-a, -a + 1, \cdots, 0, 1, \cdots, b\}$. The width of this distribution is $s = \sqrt{\frac{\pi((a+b+1)^2-1)}{6}}$.

The Fourier transforms for these three distributions are summarized in the following result. For the distribution that are consist of the above three distributions (e.g. the error distribution used in CRYSTALS-KYBER), its Fourier transforms can be easily obtained according to the convolution property. Let f be the probability function for the distribution in consideration, then we have

Theorem 3.2 *1. For the rounded discrete Gaussian distribution, we have*

$$\left|\widehat{f}(k)-\widehat{\Psi_s}(k)\right| \le 2e^{-\pi\frac{(R+\frac{1}{2})^2}{s^2}}.$$

2. For the central binormal distribution $B(h)$, we have

$$\hat{f}(k)=\cos^{2h}(\frac{\pi k}{q}).$$

3. For the bounded uniform distribution $U[-a,-a+1,\cdots,0,\cdots,b]$, we have

$$|\hat{f}(k)|^2=\frac{1-\cos\frac{2\pi(a+b+1)k}{q}}{(a+b+1)^2(1-\cos\frac{2\pi k}{q})},$$

for $k=1,2,\cdots,\lfloor q/2\rfloor-1$.

Proof. 1. We will use the following estimation for Gaussian distribution [9]: for $x\ge 0$,

$$2\int_x^{\infty} e^{-\pi t^2}dt \le e^{-\pi x^2}. \tag{5}$$

Let $A=\sum_{j=-R}^{R}\Psi_s(j)=2\int_0^{\frac{2R+1}{2s}} e^{-\pi t^2}dt$. Then $\widehat{f}(k)=\frac{1}{A}\sum_{j=-R}^{R}e^{-\frac{2\pi ijk}{q}}\Psi_s(j)$, and $1-A\le e^{-\pi\frac{(R+\frac{1}{2})^2}{s^2}}$. Now

$$\begin{aligned}\left|\widehat{\Psi_s}(k)-\widehat{f}(k)\right| &\le \left|\widehat{\Psi_s}(k)-A\widehat{f}(k)\right|+\left|\widehat{f}(k)-A\widehat{f}(k)\right| \le \left|\sum_{|j|\ge R+1} e^{-\frac{2\pi ijk}{q}}\Psi_s(j)\right|+(1-A)\\ &\le \sum_{|j|\ge R+1}\Psi_s(j)+(1-A)=2\int_{\frac{2R+1}{2s}}^{\infty} e^{-\pi t^2}dt+(1-A)\le 2e^{-\pi\frac{(R+\frac{1}{2})^2}{s^2}}.\end{aligned}$$

2. For variable $x\sim B(1)$, it is easy to calculate that $\hat{f}_x(k)=\frac{1}{2}+\frac{1}{2}\cos(\frac{2\pi k}{q})=\cos^2(\frac{\pi k}{q})$. So for the variable $X\sim B(h)$, we have

$$\hat{f}(k)=\hat{f}_x^h(k)=\cos^{2h}(\frac{\pi k}{q})$$

according to the convolution property of the Fourier transform.

3. Let $t=\frac{2\pi k}{q}$, then

$$\begin{aligned}\hat{f}(k)&=\frac{1}{a+b+1}\left(\sum_{x=0}^{b}e^{-ixt}+\sum_{x=1}^{a}e^{ixt}\right)=\frac{1}{a+b+1}\left(\frac{e^{-it(b+1)}-1}{e^{-it}-1}+\frac{e^{it(a+1)}-1}{e^{it}-1}-1\right)\\&=\frac{e^{ita}-e^{it(a+1)}+e^{-itb}-e^{-it(b+1)}}{2(a+b+1)(1-\cos t)}.\end{aligned}$$

Therefore

$$\begin{aligned}|\hat{f}(k)|^2 &= \hat{f}(k)\overline{\hat{f}(k)} = \frac{(e^{ita} - e^{it(a+1)} + e^{-itb} - e^{-it(b+1)})(e^{-ita} - e^{-it(a+1)} + e^{itb} - e^{it(b+1)})}{4(a+b+1)^2(1-\cos t)^2} \\ &= \frac{4 - 2e^{it} - 2e^{-it} - 2e^{it(a+b+1)} - 2e^{-it(a+b+1)} + e^{it(a+b)} + e^{-it(a+b)} + e^{it(a+b+2)} + e^{-it(a+b+2)}}{4(a+b+1)^2(1-\cos t)^2} \\ &= \frac{4(1-\cos t) + 2(\cos(a+b)t - 2\cos(a+b+1)t + \cos(a+b+2)t)}{4(a+b+1)^2(1-\cos t)^2} \\ &= \frac{1-\cos(a+b+1)t}{(a+b+1)^2(1-\cos t)}.\end{aligned}$$

The proof is completed.

We would like to emphasis that the contribution of an individual component of the $\mathbf{v}$ in the distinguishing attack. To this end, in order to better compare $\prod_{j=1}^{m}\hat{f}(v_j)$ and $\epsilon(\|\mathbf{v}\|) = e^{-\frac{\pi s^2\|\mathbf{v}\|^2}{q^2}}$, we define the *local width* $s(k)$ as follows

Definition 3.1 (Local Width). *For a given random variable X over $\mathbb{Z}_q$ and its probabilistic function $f(x)$, if $1 \le k \le \lfloor \frac{q}{2} \rfloor$ and $\hat{f}(k) \neq 0$, the local width $s(k)$ is defined to be*

$$s(k) = \frac{q}{k}\sqrt{\frac{-\ln|\hat{f}(k)|}{\pi}}.$$

Let us make some remarks.

- The case of $\hat{f}(k) = 0$ if not of our interest. According to (4), if there is a component v_j of $\mathbf{v}$ such that $\hat{f}(v_j) = 0$, then this $\mathbf{v}$ cannot be used in distinguishing.
- In the definition, we restricted k in between 1 and $\lfloor \frac{q}{2} \rfloor$. It can extended to integers in between $(-\frac{q}{2}, \frac{q}{2}]$ if $|\hat{f}|$ is even function. We also have $s(0) = 0$. Assume that $\hat{f}(v_j) \neq 0$ for all j, then

$$\prod_{j=1}^{m}\left|\hat{f}(v_j)\right| = e^{-\pi\frac{s^2(v_1)v_1^2 + s^2(v_2)v_2^2 + \cdots + s^2(v_m)v_m^2}{q^2}}.$$

  This is close to $\epsilon(\mathbf{v}) = e^{-\frac{\pi s^2\|\mathbf{v}\|^2}{q^2}}$ if all $s(k)$ are close to the given width s.
- It is observed that (in the examples below), for certain distributions over $\mathbb{Z}_q$, there is a large subset $S \subset \mathbb{Z}_q \cap (-\frac{q}{2}, \frac{q}{2}]$ such that $s(k) < s$ for $k \in S$. If we are able to find a vector $\mathbf{v}$ in the dual lattice with $v_j \in S$, then $\prod_{j=1}^{m}\left|\hat{f}(v_j)\right|$ gives us a greater advantage than $\epsilon(\mathbf{v})$. This idea will be developed in the later discussion.

With these Fourier transforms, we can provide more precise analysis of LWE. To this end, some comparisons between widths and local widths of the relevant distributions are presented.

Theorem 3.3 *1. For the central binormal distribution $B(h)$, we have*

$$s(k)^2 \geq s^2 + 2\pi h \left(\frac{k\pi}{q}\right)^2 \left(\frac{1}{12} + \frac{1}{45}(\frac{k\pi}{q})^2 + \frac{17}{2520}(\frac{k\pi}{q})^4\right),$$

for $k = 1, 2, \cdots, \lfloor\frac{q}{2}\rfloor - 1$ where $s = \sqrt{h\pi}$.

2. For the bounded uniform distribution $U[-a, -a+1, \cdots, 0, \cdots, b]$ with $a+b \geq 7$, we have

$$s(k)^2 + \frac{5(16qk - 3q^2)}{8k^2} < s^2,$$

for $k \in \{1, 2, \cdots, \lfloor q/2 \rfloor\}$ with exceptions that $k \geq \frac{3q}{16\pi}$ and for any integer ℓ, $|\frac{dk\pi}{q} - \ell\pi| \geq \frac{\pi}{24}$. Here $s = \sqrt{\pi\frac{(a+b+1)^2-1}{6}}$.

Proof. 1. Now consider the central binormal distribution $B(h)$. By theorem 3.2, $\hat{f}(k) = \cos^{2h}(\frac{\pi k}{q})$. Write $t = \frac{k\pi}{q}$. Then

$$s(k)^2 - s^2 = \frac{-q^2}{k^2}\frac{\ln|\hat{f}(k)|}{\pi} - h\pi = \frac{2\pi h}{t^2}(-\ln\cos t - \frac{t^2}{2}).$$

The result follows by noticing that for any $x \in (0, \frac{\pi}{2})$,

$$\begin{aligned} -\ln\cos x - \frac{x^2}{2} &= \int_0^x \tan\theta d\theta - \frac{x^2}{2} = \int_0^x (\theta + \frac{\theta^3}{3} + \frac{2\theta^5}{15} + \frac{17\theta^7}{315} + \cdots)d\theta - \frac{x^2}{2} \\ &= \frac{x^4}{12} + \frac{x^6}{45} + \frac{17x^8}{2520} + \cdots. \end{aligned}$$

2. For the error distribution $U[-a, -a+1, \cdots, 0, 1, \cdots, b]$, we have $|\hat{f}(k)|^2 = \frac{1-\cos\frac{2\pi(a+b+1)k}{q}}{(a+b+1)^2(1-\cos\frac{2\pi k}{q})}$. Let $t = \frac{2k\pi}{q}$ and $d = a+b+1$.

$$\begin{aligned} s^2 - s(k)^2 &= \pi\frac{(a+b+1)^2 - 1}{6} + \frac{q^2}{k^2}\frac{\ln|\hat{f}(k)|}{\pi} = \frac{4\pi}{t^2}\ln\left(e^{\frac{t^2(d^2-1)}{24}}\sqrt{\frac{1-\cos dt}{d^2(1-\cos t)}}\right) \\ &= \frac{2\pi}{t^2}\ln\left(e^{\frac{t^2(d^2-1)}{12}}\frac{1-\cos dt}{d^2(1-\cos t)}\right). \end{aligned}$$

Under the assumption, we have $d \geq 8, t \geq \frac{3\pi}{8}$ and $|dt - 2\ell\pi| \geq \frac{\pi}{12}$ for any integer ℓ. Write $t = \frac{3\pi}{8} + \varDelta$. Then

$$\begin{aligned} e^{\frac{t^2(d^2-1)}{12}} &= e^{\frac{(d^2-1)(\frac{3\pi}{8})^2}{12}} e^{\frac{(d^2-1)(t+\frac{3\pi}{8})\varDelta}{12}} \geq e^{\frac{(d^2-1)(\frac{3\pi}{8})^2}{12}} e^{\frac{(64-1)(\frac{3\pi}{4})\varDelta}{12}} \\ &> e^{\frac{23(d^2-1)}{200}} e^{12\varDelta}. \end{aligned}$$

On the other hand, since $\sin\varDelta \leq \varDelta$ and $1 - \cos\varDelta \leq \varDelta$,

$$\begin{aligned} \frac{1-\cos dt}{d^2(1-\cos t)} &\geq \frac{1-\cos\frac{\pi}{12}}{d^2(1-\cos\frac{3\pi}{8}\cos\varDelta + \sin\frac{3\pi}{8}\sin\varDelta)} \\ &= \frac{1-\cos\frac{\pi}{12}}{d^2(1-\cos\frac{3\pi}{8} + \cos\frac{3\pi}{8}(1-\cos\varDelta) + \sin\frac{3\pi}{8}\sin\varDelta)} > \frac{1}{20d^2(1+2\varDelta)}. \end{aligned}$$

Thus

$$s^2 - s(k)^2 \geq \frac{2\pi}{t^2} \ln \left(\frac{e^{\frac{23(d^2-1)}{200}}}{20d^2} \frac{e^{2\varDelta}}{1+2\varDelta} e^{10\varDelta} \right) \geq \frac{20\pi\varDelta}{t^2} = \frac{5(16qk-3q^2)}{8k^2}.$$

Next, we calculate and analyze local widths for three lattice-based cryptography schemes. These schemes are Frodo based on LWE with approximate discrete rounded Gaussian sampler, NewHope based on RLWE with central binormal distribution sampler, and Saber based on LWR problem. The experiments show similar behaviors as described in theorem 3.3, but the actual computation results are more precise than the theoretical estimation. The result for comparing width for rounded Gaussian is quite complicated and was not included in theorem 3.3. But our experiment shows that an improved analysis can be made.

Frodo. As one of the candidates in the second round of NIST PQC standardization, Frodo scheme is based on the LWE problem and the noise follows rounded Gaussian distribution. We take one of the two recommended sets of parameters as an example, that is

$$(n, q, \sigma) = (640, 32768, 2.8),$$

the explicit error distribution is given by

Table 1. The Error Distribution in Frodo

standard deviation	0	±1	±2	±3	±4	±5	±6	±7	±8	±9	±10	±11	±12
2.8	$\frac{9288}{65536}$	$\frac{8720}{65536}$	$\frac{7216}{65536}$	$\frac{5264}{65536}$	$\frac{3384}{65536}$	$\frac{1918}{65536}$	$\frac{958}{65536}$	$\frac{422}{65536}$	$\frac{164}{65536}$	$\frac{56}{65536}$	$\frac{17}{65536}$	$\frac{4}{65536}$	$\frac{1}{65536}$

The calculation result of $s(k)$ for the above parameters is shown in Figure 1 where $s_0 = \sigma\sqrt{2\pi} = 7.02$. We see that the value of $s(k)$ gets below s_0 when $k > q/3$ with a decreasing tendency (most of the time). This is suggestive that finding a suitable vector in the dual lattice to utilize local width may result a bigger distinguish advantage.

NewHope. NewHope is based on the hardness of RLWE problem and its error follows central binormal distribution. We use the following recommended set of parameters as an example

$$(n, q, \sigma) = (512, 12289, 2).$$

with the error $e \sim B(8)$. The calculation result of $s(k)$ for the above parameters is shown in Figure 2. It is seen that the width $s_0 = 2\sqrt{2\pi} \approx 5.01$ is the lower bound of $s(k)$ for all $k \leqslant q/2$. When k gets bigger, $s(k) - s_0$ becomes bigger as

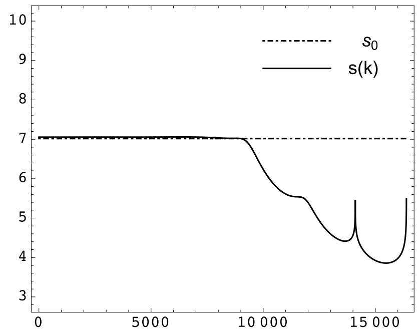

Figure 1. Comparison between local width $s(k)$ and given width s_0 in Frodo.

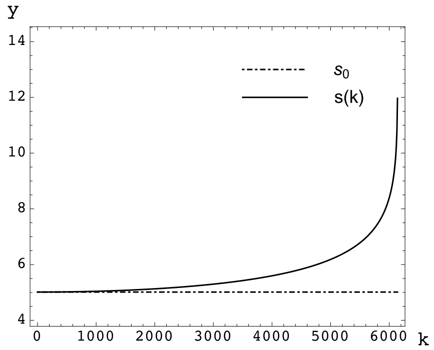

Figure 2. Comparison between local width $s(k)$ and given width s_0 in NewHope.

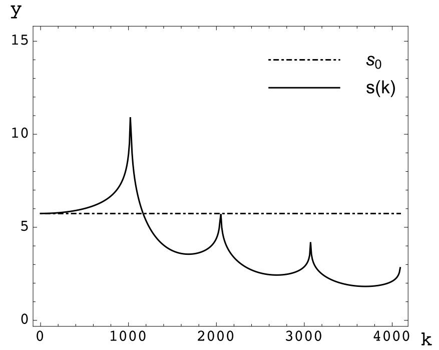

Figure 3. Comparison between local width $s(k)$ and given width s_0 in Saber.

predicted by theorem 3.3. This means that using $e^{-\pi s_0^2\|\mathbf{v}\|^2/q^2}$ as an advantage is quite conservative.

Saber. Saber scheme is based on the hardness of Mod-LWR problem. It is always treated as the normal LWE problem with the bounded uniform distribution as noise sampling. We consider the following parameters

$$(n, q, \sigma) = (768, 8192, 2.29)$$

with the error $e \sim U[-3, -2, -1, 0, 1, 2, 3, 4]$. The calculation result of $s(k)$ for the above parameters is shown in Figure 3 . We see that the value of $s(k)$ is smaller than s_0 when $l \geq \frac{q}{5}$. This is better than the theoretical estimation presented in theorem 3.3. As is shown in the above figure, there is a large gap between $s(k)$ and s_0 when k is bigger than certain value for the bounded uniform distribution.

4 Applications of the distinguish advantage analysis for LWE instances

In the above discussion, the distinguish advantage can be revealed completely by the Fourier transform of error distribution which is shown to be not only related to the length of vectors in the dual lattice, but also relevant to the components of vectors. For the distributions in Frodo and Saber analyzed above, those k with corresponding $s(k)$ smaller than s_0 is quite large (for example, $k \approx q/3$ in Frodo and $k \approx q/5$ in Saber). If we can assume the outputted vectors of lattice reduction algorithm distribute uniformly on the sphere, there is little influence on the estimation for LWE instances under dual attack model as the proportion of short vectors which has a component larger than k is small. However, we can see different properties of the three distributions do exist compared with the ideal Gaussian distribution when looking at the Fourier transform. So a new measure to describe the difference against distinguish attack between practical distributions and ideal Gaussian distribution can be naturally established based on the Fourier transform.

4.1 Analysis of dual attack for different distributions

In this subsection, we make a analysis of dual attack for three practical distributions, known as rounded Gaussian, bounded uniform and central binormal distribution. Let $C(\ell)$ denote the cost of obtaining (short) vectors $\mathbf{v}$ with length ℓ in an n dimensional lattice, the whole complexity of dual attack is the tradeoff between the cost of finding short vectors and the number of vectors needed which is decided by the corresponding distinguish advantage, i.e.

$$\frac{C(\ell)}{\epsilon^2(\ell)}$$

where $\epsilon(\ell) = e^{-\pi s^2 \ell^2 / q^2}$.

As it is shown in Section 3, there are underestimations of distinguish advantage for some vectors for bounded uniform distribution and rounded Gaussian distribution, a natural idea is to use the vectors of bigger advantages to make a distinguish, therefore less number of vectors are needed and it may have an influence on hardness of LWE instances.

Accordingly, let C_ℓ be the set of vectors with length ℓ in the dual lattice, we prefer vectors denoted as $\mathbf{w}$ such that $\epsilon(\mathbf{w})$ is close to

$$\epsilon'(\ell) = \max_{(v_1,\cdots,v_n)\in C_l} |\hat{f}(v_1)|\cdots|\hat{f}(v_n)|.$$

In this case, $\epsilon'(\ell)$ would be much larger than $\epsilon(\ell)$ for some lengths ℓ. It is easy to see that vectors with large components are likely to be chosen. The tradeoff between $\epsilon'(\ell)$ and the cost of obtaining such vectors denoted as $C'(\ell)$ would give the final result. However, little research of investigating the components of lattice vectors has been reported and we pose it as an open question.

Since there are little results of the cost of finding vectors with large components which is an important factor in estimating the hardness of LWE instances, we make further discussion based on the following assumption. We use parameter β to describe the complexity of finding vectors with large components and the assumptions are listed as follows.

Assumption 1. The cost of obtaining a short vector $\mathbf{v}$ with length ℓ such that its distinguish advantage is roughly $\epsilon'(\ell)$ in an n dimension lattice is $C(\ell)\cdot 2^{\beta n}$ where $\beta > 0$.

Remark 4.1. If $\beta = 0$, the assumption means that the extra overheads can be regarded as negligible when compared to the cost of finding short vectors, i.e. the cost is $C(\ell)\cdot poly(n)$.

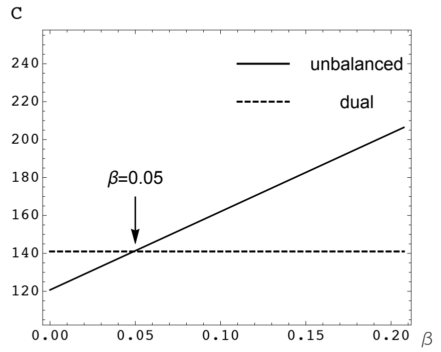

Figure 4. Complexities of parameters in Frodo under dual attack and unbalanced model.

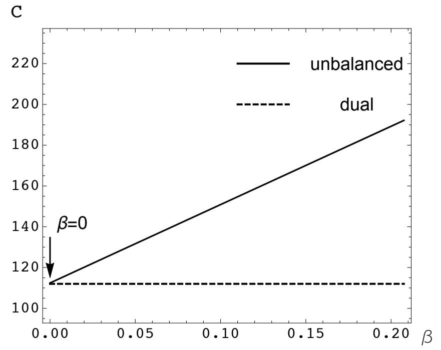

Figure 5. Complexities of parameters in NewHope under dual attack and unbalanced model.

Based on Assumption 1, we can further discuss the differences by using the three practical distributions with parameter β. For simplicity, we call this measure model as "unbalanced" model where the core hardness in estimating BKZ complexity is taken as $2^{(0.292+\beta)b}$ and the distinguish advantage is taken as $\epsilon'(\ell)$. The $\epsilon'(\ell)$ is calculated according to the local width. Here we give analysis of parameters in Frodo, NewHope and Saber under unbalanced model.

We show the relationship between complexity (the log of complexity is shown in the ordinate C) and the parameter β (it is shown in the abscissa β). We remark that the complexity under dual attack is calculated under the model introduced in Section 2.4 where the core hardness in estimating BKZ complexity is taken as $2^{0.292b}$.

From the experiment, it is seen that different distribution performs quite differently. For example, the result of Frodo in Figure 4 shows the complexity

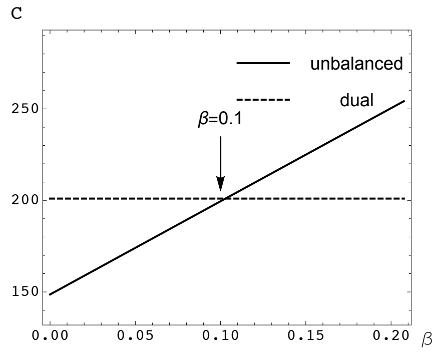

Figure 6. Complexities of parameters in Saber under dual attack and unbalanced model.

varies from 2^{120} to 2^{140} with β varies from 0 to 0.05, the result of Saber in Figure 6 shows the complexity varies from 2^{150} to 2^{200} with β varies from 0 to 0.1, while the complexity of central binormal distribution shown in Figure 5 is always larger than 2^{112} for any $\beta \geqslant 0$. Let β_Z denote the point of intersection where the complexity under unbalanced model equals to that under ideal model. As the complexity increases with β increasing, the less β_Z is, the stronger assumptions the distribution seems to be security against.

The discussion of results.

- It is clear that the complexity of dual attack against ideal Gaussian distribution is always no more than results for any $\beta \geqslant 0$ since the upper bound of distinguish advantage among all vectors is used, that is $\beta_Z(\text{Ideal Gaussian}) = 0$. As for the parameters in NewHope where the error follows binormal distribution, it is shown that $\beta_Z(\text{NewHope}) = 0$.

- We can see that $\beta_Z(\text{Frodo}) = 0.05$ and $\beta_Z(\text{Saber}) = 0.1$ from experiments. It means that the hardness of parameters in Frodo (Saber) would be lower than the present result if the assumption of $\beta < \beta_Z(\text{Frodo}) = 0.05(\beta < \beta_Z(\text{Saber}) = 0.1)$ is true.

It should be noted that the analysis does not mean the distributions used in Frodo and Saber are not secure unless algorithms that satisfy the assumptions can be found. And if algorithms that satisfy the assumptions are proven nonexistent, then it is seen that the two distribution can provide the same security of ideal Gaussian distribution against distinguish attack. The results provide an alternative value to measure the security against distinguish attack provided by practical distributions compared with ideal ones. However, since the existence of those algorithms is unknown and the central binormal distribution share the same result with ideal Gaussian distribution under the strongest assumption, a natural question arises that can we find some other practical distributions that also have this good property?

4.2 A New Sampler

In this section, we propose a new sampler denoted as "Mixed Sampling" that has the same property with central binormal distribution. It is denoted as "mixed distribution" and we firstly present the definition.

Definition 4.1 (Mixed distribution). *Let k_1, k_2 be positive integers, $\{X_1, \cdots, X_{k_1}\}$ is a sequence of independent and identically distributed variables where $X_i \sim B(1)$, $\{Y_1, \cdots, Y_{k_2}\}$ is a sequence of independent and identically distributed variables where $Y_i \sim U[-1, 0, 1]$, the variable X following "mixed distribution" denoted as $\Phi(k_1, k_2)$ is the convolution of $\{X_i\}_{i=1}^{k_1}$ and $\{Y_i\}_{i=1}^{k_2}$, i.e.*

$$X = X_1 + \cdots + X_{k_1} + Y_1 + \cdots + Y_{k_2}.$$

As for the "mixed distribution" $\Phi(k_1, k_2)$, we calculate the expectation and variance.

Lemma 4.1 *Let $X \sim \Phi(k_1, k_2)$, then*

$$E[X] = 0, D[X] = \frac{k_1}{2} + \frac{2k_2}{3}.$$

Proof. Since $X \sim \Phi(k_1, k_2)$, then it could be express of the following form,

$$X = \sum_{i=1}^{k_1}(b_i - b_i') + \sum_{i=1}^{k_2} u_i.$$

where $b_i, b_i' \sim U[0, 1]$ and $Pr[b = 0] = Pr[b = 1] = 1/2$, $u_i \sim U[-1, 0, 1]$ and $Pr[u_i = j] = 1/3$ for $j = -1, 0, 1$. Therefore we have

$$E[X] = 2k_1 E[b_i] + k_2 E[u_i] = 0,$$
$$D[X] = 2k_1 D[b_i] + k_2 D[u_i] = \frac{k_1}{2} + \frac{2k_2}{3}.$$

4.2.1 The lower bound of local width In this subsection, we prove that the mixed distribution with properly chosen parameters provides claimed security under dual attack where the local width takes the original width $s_0 = \sqrt{2\pi}\sigma$ as a lower bound.

Theorem 4.2 *Let q be an integer and the variable $X \sim \Phi(k_1, k_2)$, if $k_1 \geqslant k_2$, for $j = 1, 2, \cdots, \lfloor q/2 \rfloor$, we have*

$$s(j) \geqslant s_0,$$

where $s_0 = \sqrt{2\pi D[X]} = \sqrt{(k_1 + 4k_2/3)\pi}$.

Proof. Let f be the probability function for X, then from theorem 3.2 we see that for $j = 1, 2, \cdots, \lfloor q/2 \rfloor$,

$$|\widehat{f}(j)| = \left(\frac{|1 + 2\cos\frac{2\pi j}{q}|}{3} \right)^{k_2} \cos^{2k_1} \frac{\pi j}{q}.$$

This gives that $s(j) = \frac{q}{j}\sqrt{\frac{-\ln|\hat{f}(j)|}{\pi}}$. Therefore we need to prove

$$\left(\frac{|1 + 2\cos\frac{2\pi j}{q}|}{3} \right)^{k_2} \cos^{2k_1} \frac{\pi j}{q} \leqslant e^{-\pi^2 j^2 \frac{(k_1 + \frac{4k_2}{3})}{q^2}} \tag{6}$$

Write $y = \frac{\pi j}{q}$. Notice that for any $\theta \in [0, \frac{\pi}{2})$, $e^{\theta^2} \cos^2\theta \leq 1$, so

$$\begin{aligned} \left(\frac{|1 + 2\cos y|}{3} \right)^{k_2} \cos^{2k_1} y &= \left(\frac{|1 + 2\cos 2y|}{3} \right)^{k_2} \left(e^{y^2}\cos^2 y\right)^{k_1} e^{-k_1 y^2} \\ &\leq \left(\frac{|1 + 2\cos 2y|}{3} e^{y^2} \cos^2 y \right)^{k_2} e^{-k_1 y^2} \end{aligned}$$

To prove (6), it suffices to show for $y \in [0, \frac{\pi}{2})$,

$$\frac{|1 + 2\cos 2y|}{3} \cos^2 y \leq e^{-\frac{7}{3}y^2}. \tag{7}$$

This inequality is trivial for $y = \frac{\pi}{3}$. For $y \in [0, \frac{\pi}{3}) \cup (\frac{\pi}{3}, \frac{\pi}{2})$, we let

$$h(y) = \frac{7}{3}y^2 + \ln\frac{|1 + 2\cos 2y|}{3} + 2\ln\cos y.$$

The derivative and second derivative of h are

$$h'(y) = \frac{14}{3}y - \frac{4\sin 2y}{1 + 2\cos 2y} - 2\tan y;\ h''(y) = \frac{14}{3} - \frac{16 + 8\cos 2y}{(1 + 2\cos 2y)^2} - \frac{2}{\cos^2 y}.$$

To get (7), we just need to show $h(y) \leq 0$. Two cases need to be considered.

Case I: $y \in [0, \frac{\pi}{3})$. In this case, $h(y) = \frac{7}{3}y^2 + \ln\frac{1+2\cos 2y}{3} + 2\ln\cos y$ and $h(0) = 0$. Notice that $h''(y) = \frac{14}{3} - \frac{12}{(1+2\cos 2y)^2} - \frac{4}{(1+2\cos 2y)} - \frac{2}{\cos^2 y} \leq 0$, as h'' is decreasing on $[0, \frac{\pi}{3})$. Together with the fact that $h'(0) = 0$, this implies that $h' \leq 0$ on $[0, \frac{\pi}{3})$.

Therefore, we must have $h(y) \leq 0$ on $[0, \frac{\pi}{3})$.

Case II: $y \in (\frac{\pi}{3}, \frac{\pi}{2})$. In this case, $h(y) = \frac{7}{3}y^2 + \ln\frac{-1-2\cos 2y}{3} + 2\ln\cos y$. Simple calculation shows $0.231 > h'(1.335) > 0.23$ and $h'(1.35) < -0.49$. Since $h''(y) < \frac{14}{3} - \frac{2}{\cos^2 y} < \frac{14}{3} - \frac{2}{\cos^2\frac{\pi}{3}} < 0$, $h'(y)$ is strictly decreasing on $(\frac{\pi}{3}, \frac{\pi}{2})$. This means that $h'(y)$ has only one zero y_0 on the interval $(\frac{\pi}{3}, \frac{\pi}{2})$, and $1.335 < y_0 < 1.35$. This also implies that $h(y_0)$ is the maximum value of h on $(\frac{\pi}{3}, \frac{\pi}{2})$.

It is noted that $h(1.335) < -0.094$. By Lagrange's mean value theorem, there is a $\xi \in (1.335, y_0)$ such that $h(y_0) = h(1.335) + h'(\xi)(y_0 - 1.335)$. Thus

$$h(y) \leq h(y_0) < h(1.335) + h'(1.335)(1.35 - 1.335) < -0.094 + 0.231 \cdot 0.015 < 0.$$

4.2.2 Mixed Sampling Algorithm and Efficiency Analysis.

Mixed sampling could be implemented by using the central binomial distribution and the uniform distribution on $\{-1, 0, 1\}$ as the underlying modules. The algorithm is listed below.

Mixed sampling algorithm

Input: The parameter (k_1, k_2).

Output: The value of variable following the distribution $\Phi(k_1, k_2)$.

1: Set a random number generated by a 2 bits random source and output an integer a that subjects to a central binomial distribution. For example, when the random number input is 00/01/10/11 respectively, the output is $-1/0/0/1$.

2: Repeat the central binomial distribution sampling k_1 times, then calculate the sum of k_1 values. That is, let the i-th $(i = 1, \cdots, k_1)$ output is a_i, then calculate $A = a_1 + a_2 + \cdots + a_{k_1}$.

3: Take a random number generated by $f = \lceil \log_2(3^{k_2}) \rceil$ bits random source. If the value of the random number in binary is greater than 3^{k_2}, then enter a random number generated by the f bits random source again until the value is not greater than 3^{k_2}. Let the random number be expressed as a k_2 ternary string, then count the number of $0, 2$ which is denoted respectively as b_0, b_2, output $B = b_2 - b_0$;

4: Output the value $S = A + B \bmod q$.

Theorem 4.3. *The mixed sampling algorithm outputs a sample distributed as $\Phi(k_1, k_2)$ correctly and the expectation of bits used to output a sample is*

$$2k_1 + \frac{\lceil k_2 \log_2 3 \rceil 2^{\lceil k_2 \log_2 3 \rceil}}{3^{k_2}}.$$

Proof. Since the variable X could be expressed as

$$X = \sum_{i=1}^{k_1} b_i + \sum_{i=1}^{k_2} u_i,$$

where b_i follows the distribution that

$$Pr[b_i = -1] = Pr[b_i = 1] = \frac{1}{4}, Pr[b_i = 0] = \frac{1}{2},$$

and $u_i \sim U[-1, 0, 1]$. Therefore the sampling of $\sum_{i=1}^{k_1} b_i$ is obtained by repeating the central binormal module k_1 times and the number of bits is $2k_1$. As for the variable $\sum_{i=1}^{k_2} u_i$, let Y be the random value outputted by step 4, we have

$$Pr[Y \leqslant 3^{k_2}] = \frac{3^{k_2}}{2^{\lceil \log_2 3^{k_2} \rceil}}.$$

Therefore the expectation times of step 4 is

$$\frac{2^{\lceil k_2 \log_2 3 \rceil}}{3^{k_2}}.$$

In summary, we can prove the conclusion in the theorem.

Comparison with other samplings. We make a comparison of the variance, sampling bits and the lower bound of local width among different sampling methods, including mixed sampling, approximate discrete rounded Gaussian sampling and central binormal sampling.

As for approximating discrete Gaussian sampling, although there are many different algorithms to approximate DGS, such as Knuth-Yao method and reject sampling, the common feature is that they approximate probabilities by sampling uniformly within a certain integer N whose bits can be denoted as $\log_2 N$. Under the observations from experiments which is shown in Appendix I, we need $N = O(e^{\pi s^2/4})$ to make $s(k) \approx s_0$ for $k < q/2$ which can provide security under the strongest assumption as it is with central binormal sampling, thus the number of random source bits required is approximately

$$\frac{\pi s^2 \log_2 e}{4}.$$

The comparison among different samplings is shown in Table 4.

Table 4. Comparison among different samplings.

sampling algorithm	bits	variance	lower bound of local width
discrete Gaussian sampling	$(\pi s^2 \log_2 e)/4$	$s^2/(2\pi)$	s
mixed sampling	$2k_1 + f'2^{f'}/3^{k_2}$	$k_1/2 + 2k_2/3$	$\sqrt{(k_1 + 4k_2/3)\pi}$
central binormal sampling	$2k$	$k/2$	$\sqrt{k\pi}$

* $f = \lceil \log_2(3^k) \rceil, f' = \lceil \log_2(3^{k_2}) \rceil$.

- In terms of the selection range of sampling variance, it is wide for DGS as we can set arbitrary width. And the sampling width has only a limited number of values to select for the central binomial distribution since the parameter k is a positive integer. The selection of mixed sampling parameters is also discrete and depends on the selection of parameters (k_1, k_2). It is more flexible when compared with central binomial distribution.

- In terms of sampling efficiency, the efficiency of mixed sampling is higher than that of central binomial distribution and approximated discrete Gaussian sampling under appropriate selection and it is shown in Figure 7.

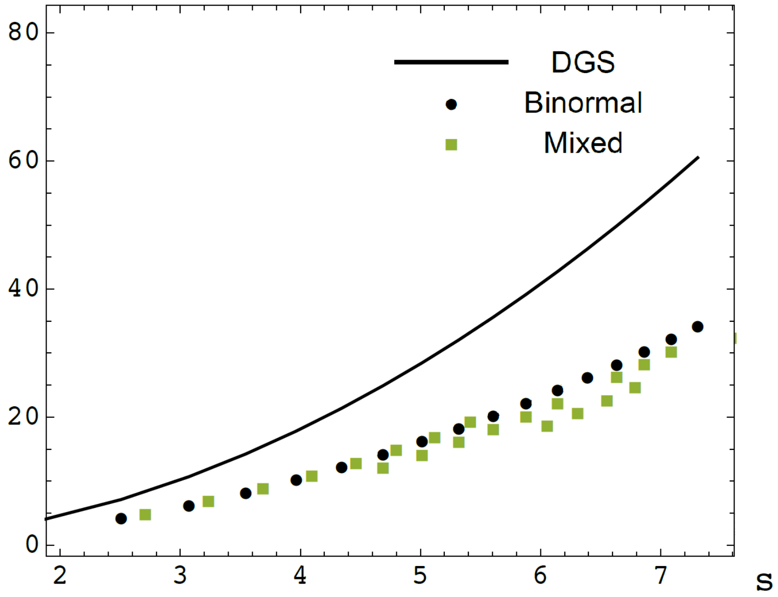

Figure 7. Efficiency comparison of sampling algorithms

- *The abscissa "s" denotes the sampler's width and longitudinal axis is the bits in the sampling algorithm. The square symbol denotes the width and bits under proper choice of parameters (k_1, k_2) in mixed sampling.

- *Let y be the bits of sampling, then

$$y = \begin{cases} 1.1331s^2, & \text{when it is discrete sampling;} \\ 0.6366s^2, & \text{when it is binormal sampling;} \\ 0.4442s^2 + 3.9269s - 23.7935, & \text{when it is mixed sampling.} \end{cases}$$

5 Conclusion

In this paper, we introduce a refined framework on the distinguish advantage of LWE instances by using Fourier transform. We use the proposed framework to analyze the practical parameters used in NIST PQC candidates where the structure of error distribution plays different roles. Furthermore, a novel type of error sampler with higher efficiency, security as well as flexibility is described.

References

1. Albrecht, M.R., Cid, C., Faugère, J., Fitzpatrick, R., Perret, L.: Algebraic algorithms for LWE problems. ACM Comm. Computer Algebra **49**(2), 62 (2015)
2. Albrecht, M.R., Curtis, B.R., Deo, A., Davidson, A., Player, R., Postlethwaite, E.W., Virdia, F., Wunderer, T.: Estimate all the {LWE, NTRU} schemes! In: International Conference on Security and Cryptography for Networks. pp. 351–367. Springer (2018)
3. Alkim, E., Ducas, L., Pöppelmann, T., Schwabe, P.: Post-quantum key exchange - A new hope. In: Holz, T., Savage, S. (eds.) 25th USENIX Security Symposium, USENIX Security 16, Austin, TX, USA, August 10-12, 2016. pp. 327–343. USENIX Association (2016), https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/alkim
4. Arora, S., Ge, R.: New algorithms for learning in presence of errors. In: Automata, Languages and Programming - 38th International Colloquium, ICALP 2011, Zurich, Switzerland, July 4-8, 2011, Proceedings, Part I. pp. 403–415 (2011)
5. Banaszczyk, W.: New bounds in some transference theorems in the geometry of numbers. Mathematische Annalen **296**(1), 625–635 (1993)
6. Banaszczyk, W.: Inequalities for convex bodies and polar reciprocal lattices inr n. Discrete & Computational Geometry **13**(2), 217–231 (1995)
7. Banerjee, A., Peikert, C., Rosen, A.: Pseudorandom functions and lattices. In: Annual International Conference on the Theory and Applications of Cryptographic Techniques. pp. 719–737. Springer (2012)
8. Brakerski, Z., Gentry, C., Vaikuntanathan, V.: (leveled) fully homomorphic encryption without bootstrapping. In: Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8-10, 2012. pp. 309–325 (2012)
9. Chang, S.H., Cosman, P.C., Milstein, L.B.: Chernoff-type bounds for the gaussian error function. IEEE Transactions on Communications **59**(11), 2939–2944 (2011)
10. Donoho, D.L., Stark, P.B.: Uncertainty principles and signal recovery. SIAM Journal on Applied Mathematics **49**(3), 906–931 (1989)
11. Duc, A., Tramer, F., Vaudenay, S.: Better algorithms for lwe and lwr. In: Annual International Conference on the Theory and Applications of Cryptographic Techniques. pp. 173–202. Springer (2015)
12. D'Anvers, J.P., Karmakar, A., Roy, S.S., Vercauteren, F.: Saber: Mod-lwr based kem. Proposal to NIST PQC Standardization (2017)
13. Guo, Q., Johansson, T., Mårtensson, E., Stankovski, P.: Coded-bkw with sieving. In: Advances in Cryptology - ASIACRYPT 2017 - 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part I. pp. 323–346 (2017)
14. Guo, Q., Johansson, T., Stankovski, P.: Coded-bkw: Solving LWE using lattice codes. In: Advances in Cryptology - CRYPTO 2015 - 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part I. pp. 23–42 (2015)
15. Hanrot, G., Pujol, X., Stehlé, D.: Terminating bkz. IACR Cryptology ePrint Archive **2011**, 198 (2011)

16. Kirchner, P., Fouque, P.: An improved BKW algorithm for LWE with applications to cryptography and lattices. In: Advances in Cryptology - CRYPTO 2015 - 35th Annual Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2015, Proceedings, Part I. pp. 43–62 (2015)
17. Langlois, A., Stehlé, D.: Worst-case to average-case reductions for module lattices. Des. Codes Cryptography **75**(3), 565–599 (2015)
18. Langlois, A., Stehlé, D.: Worst-case to average-case reductions for module lattices. Designs, Codes and Cryptography **75**(3), 565–599 (2015)
19. Lindner, R., Peikert, C.: Better key sizes (and attacks) for lwe-based encryption. In: Cryptographers' Track at the RSA Conference. pp. 319–339. Springer (2011)
20. Lu, X., Liu, Y., Zhang, Z., Jia, D., Xue, H., He, J., Li, B.: LAC: practical ring-lwe based public-key encryption with byte-level modulus. IACR Cryptology ePrint Archive **2018**, 1009 (2018), https://eprint.iacr.org/2018/1009
21. Lyubashevsky, V., Peikert, C., Regev, O.: On ideal lattices and learning with errors over rings. In: Advances in Cryptology - EUROCRYPT 2010, 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Monaco / French Riviera, May 30 - June 3, 2010. Proceedings. pp. 1–23 (2010)
22. Naehrig, M., Alkim, E., Bos, J., Ducas, L., Easterbrook, K., LaMacchia, B., Longa, P., Mironov, I., Nikolaenko, V., Peikert, C., Raghunathan, A., Stebila, D.: Frodokem learning with errors key encapsulation. http://frodokem.org (2018)
23. Peikert, C., Regev, O., Stephens-Davidowitz, N.: Pseudorandomness of ring-lwe for any ring and modulus. In: Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017. pp. 461–473 (2017)
24. Peter, S., Roberto, A., Joppe, B., Leo, D., Eike, K., Tancrede, L., Vadim, L., John, M.S., Gregor, S., Damien, S.: Crystals-kyber. https://pq-crystals.org/ (2017)
25. Regev, O.: On lattices, learning with errors, random linear codes, and cryptography. In: Gabow, H.N., Fagin, R. (eds.) Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005. pp. 84–93. ACM (2005). https://doi.org/10.1145/1060590.1060603, https://doi.org/10.1145/1060590.1060603
26. Tian, C., Liu, M., Xu, G.: Measure inequalities and the transference theorem in the geometry of numbers. Proceedings of the American Mathematical Society **142**(1), 47–57 (2014)

A Approximate discrete Gaussian sampling

As for approximating discrete Gaussian sampling, the common way is to approximate probabilities by sampling uniformly within a certain integer N whose bits can be denoted as $\log_2 N$. The question we are concerned with is the relationship between the size of N and the lower bound of the local width. As shown in our analysis, the local width of discrete Gaussian approximates the given width for any $k \leqslant q/2$. Given a small N to approximate a distribution with a large variance, it would have a larger statistical distance with DGS which can be reflected by the gap between local width and the initial given width.
Let $X \sim D_{s,q}$, $p_0, p_1, \cdots, p_{q-1}$ be the corresponding probability when the value takes $0, 1, \cdots, q-1$ module q, if we approximate the distribution by sampling with N points, the practical probability would be shown as

$$p_0 + \epsilon_0, p_1 + \epsilon_1, \cdots, p_{q-1} + \epsilon_{q-1},$$

where $|\epsilon_i| \leqslant \min\{\frac{1}{2N}, e^{-\pi i^2/s^2}\}$ if $p_0, p_1, \cdots, p_{q-1}$ are chosen properly. We have

$$\begin{aligned}\hat{f}(k) &= \sum_{x=0}^{q-1}(p_x + \epsilon_x)e^{-2\pi ikx/q} \\ &= \sum_{x=0}^{q-1} p_x e^{-2\pi ikx/q} + \epsilon_0 + \sum_{x=1}^{q-1} \epsilon_x e^{-2\pi ikx/q}.\end{aligned}$$

As is shown in Theorem **??**, we have

$$e^{-\pi s^2k^2/q^2} \leqslant \sum_{x=0}^{q-1} p_x e^{-2\pi ikx/q} \leqslant 2e^{-\pi s^2k^2/q^2}.$$

Let $\delta = \epsilon_0 + \sum_{x=1}^{q-1} \epsilon_x e^{-2\pi ikx/q}$, there would be the dominant term if $\delta > e^{-\pi s^2k^2/q^2}$ and therefore have an effect on the local width. We construct several experiments to discuss the relationship between N and the range of local width and have the following observations.

- **Observation 1:** When approximating the DGS with $p_0, p_1, \cdots, p_{q-1}$ by using N points, if

  $$N = O(e^{\pi s^2k^2/q^2}),$$

  where $k \in [1, 2, \cdots, \lfloor \frac{q}{2} \rfloor - 1]$, then for $j \in [1, 2, \cdots, k]$,

  $$s(j) \approx s_0.$$

  As for the Observation 1, we design the following experiment. For the discrete Gaussian distribution with $s_0 = 6$ and $q = 8192$, we select

  $$N = e^{\pi s^2/16}, e^{9\pi s^2/100}, e^{4\pi s^2/25}, e^{\pi s^2/4}$$

sampling points respectively for approximate DGS. It is shown in Figure 8 that the value which the local width deviates considerably from the initial width corresponds to

$$0.25q, 0.3q, 0.4q, 0.5q.$$

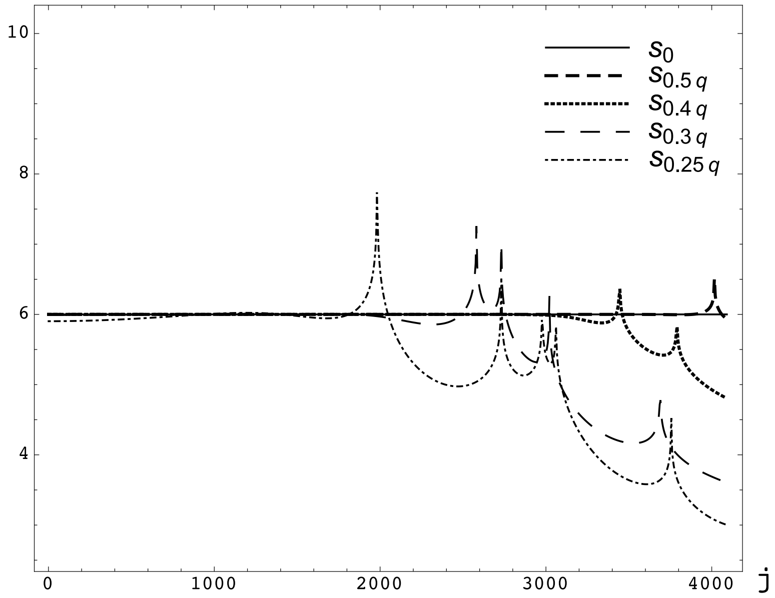

Figure 8. Contrast between local width and given width under different points approximation of N

– **Observation 2:** To ensure that $s(j) \approx s_0$ for $j < q/2$, , we need that

$$N = O(e^{\pi s^2/4}).$$

We construct the following experiment. For Gaussian distribution with different widths

$$s_0 = 3, 4, 5, 6$$

we take

$$N = 10e^{9\pi/4}, e^{4\pi}, e^{25\pi/4}, e^{9\pi}$$

sampling points respectively to reveal the relationship between local width and initial width. The result is shown in Figure 9. The local width begins to deviate from the initial width near $q/2$ and the deviation is different when compared with the given width.

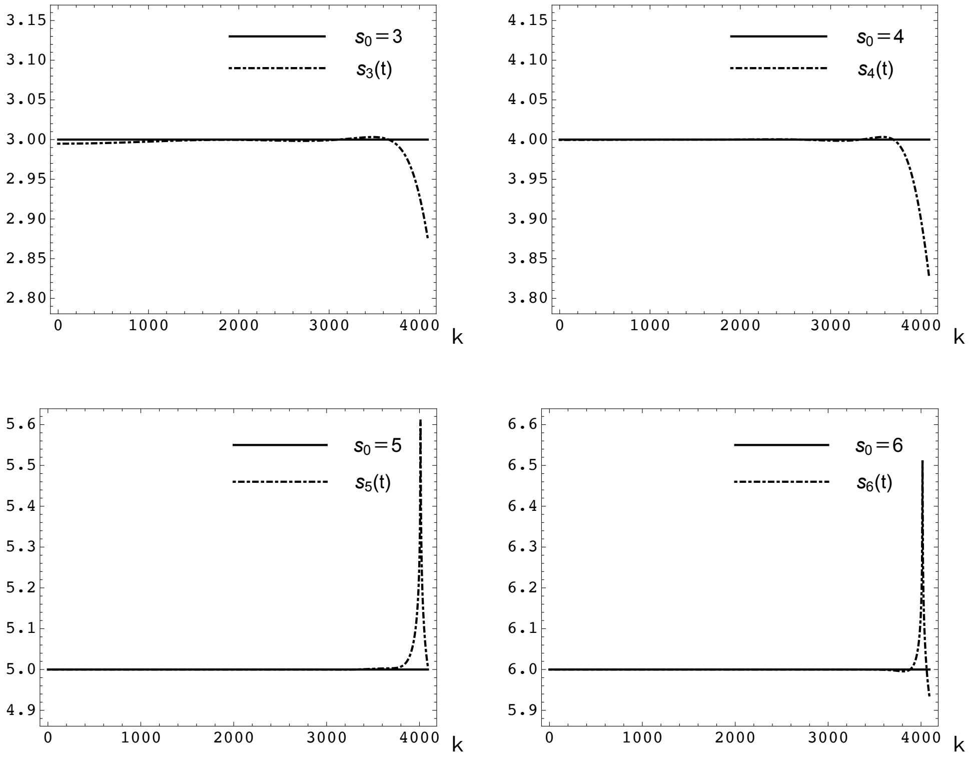

Figure 9. Comparison between the local width of approximating different DGS and given width

In summary, given a fixed N, the range of variance selection is very large, but approximating too large variance with fewer points will cause the local width to fluctuate which may result insecurity risks. In order to ensure that the local width is always near the initial width, at least $e^{\pi s^2/4}$ points are needed to approximate and the number of random source bits required is approximately

$$\frac{\pi s^2 \log_2 e}{4}.$$