Learning with Quantization, Polar Quantizer, and Secure Source Coding

Shanxiang Lyu[1], Ling Liu[2], and Cong Ling[3]

[1] Jinan University, Guangzhou, China
lsx07@jnu.edu.cn
[2] Xidian University, Xi'an, China
liuling@xidian.edu.cn
[3] Imperial College London, London, UK
c.ling@imperial.ac.uk

Abstract. This paper presents a generalization of the Learning With Rounding (LWR) problem, initially introduced by Banerjee, Peikert, and Rosen, by applying the perspective of vector quantization. In LWR, noise is induced by rounding each coordinate to the nearest multiple of a fraction, a process inherently tied to scalar quantization. By considering a new variant termed Learning With Quantization (LWQ), we explore large-dimensional fast-decodable lattices with superior quantization properties, aiming to enhance the compression performance over conventional scalar quantization. We identify polar lattices as exemplary structures, effectively transforming LWQ into a problem akin to Learning With Errors (LWE), where the distribution of quantization noise is statistically close to discrete Gaussian. Furthermore, we develop a novel "quancryption" scheme for secure source coding. Notably, the scheme achieves near-optimal rate-distortion ratios for bounded rational signal sources, and can be implemented efficiently with quasi-linear time complexity. Python code of the polar-lattice quantizer is available at https://github.com/shx-lyu/PolarQuantizer.

Keywords: Lattice-Based Cryptography · Learning with Quantization · Polar Lattice · Ciphertext Compression.

1 Introduction

Recent advancements have firmly established lattice-based cryptography (LBC) as a leading candidate to replace number-theoretic cryptography, particularly in anticipation of the quantum computing era. A shared objective across various LBC domains, including secret-key encryption [25], homomorphic encryption [10], trapdoor-based public key encryption (PKE) [28], reconciliation-based key encapsulation mechanism (KEM) [2], and encryption-based KEM [30], is the compression of ciphertexts.

To encrypt analog or floating-point signal sources, such as gradients in federated learning [31, 32] and audios [24], using lattice-based secret-key encryption

(e.g., via the Micciancio-Schultz encryption framework [25]), a series of sequential steps are typically executed: source quantization, error correction coding (ECC), encryption, ciphertext compression, and ECC decoding. This process involves quantization being employed twice: first, a lattice quantizer to reduce the source rate, and then encryption followed by quantization to reduce the ciphertext rate. However, the computational complexity of a robust lattice quantizer, along with ECC encoding and decoding, can be prohibitive. Additionally, managing public parameters in the separated setting can be cumbersome. To strike the optimal balance between rate and distortion in source coding, a common approach is dithered quantization, which involves applying a uniform dither to the signal before quantization, necessitating the sharing of the dither as public randomness.

1.1 Our Results

We define the Learning With Quantization (LWQ) problem, which serves as the cornerstone of cryptographic functionalities. LWQ extends the concepts of Learning with Errors (LWE) [29] and Learning with Rounding (LWR) [7] by incorporating a general lattice quantizer. When a lattice quantizer effectively quantizes, its error distribution resembles a uniform distribution over a ball-shaped Voronoi cell. Leveraging the sphere-Gaussian equivalence, the per-component marginal of this uniform distribution converges to a Gaussian density. Thus, a well-designed quantizer-enabled LWQ behaves analogously to LWE with Gaussian noise, while a hypercube lattice $\frac{q}{p}\mathbb{Z}^m$ (p divides q) based LWQ corresponds to LWE with uniform noise (i.e., LWR).

Furthermore, we introduce a high-dimensional lattice quantizer based on polar lattices. Our design adheres to several crucial criteria in lattice-based cryptography: the ciphertext resides within a large ring, the quantizer's computational complexity remains manageable, and the quantizer's output does not compromise the system's security. Through the application of polar lattices, we demonstrate that LWQ achieves security levels comparable to LWE with discrete Gaussian noise. This demonstration relies on establishing the close proximity of distributions between LWQ samples $(\mathbf{A}, Q_\Lambda(\mathbf{As}))$ and LWE samples $(\mathbf{A}, Q_\Lambda(\mathbf{As}+\mathbf{e}))$ in terms of Rényi divergence. Consequently, breaking LWQ would imply an advantage for solving LWE. The computational complexity of our proposed polar-lattice quantizer is $O(m \log m)$ for blocklength m.

Lastly, we present a novel approach integrating source and ciphertext quantization into a single process. Unlike conventional methods involving multiple stages of quantization, our scheme applies lattice quantization only once, directly transforming the signal source into ciphertext. The encryption-decryption cycle bypasses error correction mechanisms, such as $x \mapsto \frac{q}{2}x$ or general lattice codes, instead opting for direct lifting of messages to the $\mathbb{Z}_q$ domain for dithered quantization. In this setup, the ciphertext represents a compressed version of the encrypted source message $\mathbf{m}$, with the decrypted message $\hat{\mathbf{m}}$ lying within the Voronoi region of a lattice coset $\Lambda + \mathbf{m}$, denoted as $\mathbf{m} - \hat{\mathbf{m}} \in \mathcal{V}_\Lambda$. Notably, our approach achieves the highest achievable source-to-ciphertext ratio (SCR). For

context, the most favorable reported SCR to date stands at $1 - o(1)$ [25, Table 1], where the vanishing term $o(1)$ requires a large number m of samples.

1.2 Related Work

While sphere packing focuses on achieving the highest sphere packing density [34, 12], the inquiry into optimal lattices for quantization, aiming for the smallest average distortion, appears less mature. The theoretical proof of optimal lattice quantizers has been limited to dimensions up to 3 (*i.e.*, $\mathbb{Z}$, A_2, A_3^*) [8], although efforts to identify good lattice quantizers have resulted in periodic updates of tables for small-dimensional lattices $n \leq 24$ [3, 1].

In source coding, beyond minimizing quantization distortion, achieving unbiased quantization (where the expected quantization error is zero) is often desirable, although challenging due to its dependence on the density of source signals. Another pertinent task is achieving the rate-distortion bound for Gaussian source signals [36]. In this context, dithered quantization has been under development for decades [16, 37], where a (pseudo-)random signal, known as a dither, is introduced to the input signal before quantization. This regulated perturbation has the potential to enhance the statistical characteristics of the quantization error. While obtaining the rate-distortion bound with random lattices seems feasible [35], decoding a high-dimensional random lattice poses challenges, albeit mitigated by the law of large numbers. For a continuous Gaussian source, an explicit construction of polar lattices to achieve the rate-distortion bound has been presented in [22], where the computational complexity of the quantizer is $O(m \log m)$.

Ciphertext compression in LBC is closely tied to lattice quantization, striking a balance between structured lattices for efficiency and random lattices for security. A prevalent compression technique is scalar quantization, also known as modulus switching/modulus reduction. For instance, CKKS homomorphic encryption [10] employs simple modulus reduction to a smaller modulus before computation on ciphertexts at different levels, while CRYSTALS-Kyber [30] utilizes it for ciphertext compression. In contrast, (lattice) vector quantization, rooted in Shannon's rate-distortion theory, consistently outperforms scalar quantization by quantizing vectors rather than individual scalars. Until the advent of polar lattices, it remained unknown whether there exists a fast decodable, high-dimensional quantizer achieving optimal quantization performance.

Scalar quantization has been adapted to define a variant of LWE. Banerjee, Peikert, and Rosen [7] introduced the LWR problem, serving as a derandomized version of LWE. By replacing Gaussian sampling in LWE with deterministic rounding, LWR samples can be generated faster and with less randomness. The hardness of LWR has been established only for restricted settings. Reference [7] demonstrated that if one can distinguish the LWR distribution from uniform distribution with advantage δ, then one can also distinguish LWE with advantage $\delta - O(mBp/q)$, where the error of LWE is assumed to be uniformly distributed over $\{-B, \ldots, B\}$, and the modulus q is exponential. The size of q was reduced by assuming it is a prime in [4], while [9] showed that q can be polynomial when

the given number of LWR samples is bounded. Restrictions on the number of samples were removed in [26], and a lower bound for proving the hardness of LWR with polynomial modulus was provided.

2 Preliminaries

Table 1 summarizes a few important notations in this paper for easy reference.

Symbol	Definition
$\mathbf{x}$	a boldface lower case for vectors
$\mathbf{X}$	a boldface capital for matrices
$x \sim U$	(random variable) x admits a uniform distribution on U
$x \leftarrow \chi$	(sample) x is drawn according to distribution χ
$\mathbb{Z}_q$	set $\{0, 1, ..., q-1\}$
$\mathbb{Z}_q^{n*}$	set of integer vectors in $\mathbb{Z}_q^n$ with $\gcd(s_1, ..., s_n, q) = 1$
X_ℓ	binary representation random variable of X at level ℓ
x_ℓ^i	i-th realization of X_ℓ
$x_\ell^{i:j}$	shorthand for $(x_\ell^i, ..., x_\ell^j)$
$x_{\ell:\jmath}^i$	realization of i-th random variable from level ℓ to level $\jmath$
$[m]$	set of all integers from 1 to m
$X^{\mathcal{I}}$	subvector of $X^{[m]}$ with indices limited in $\mathcal{I} \subseteq [m]$

Table 1. IMPORTANT NOTATIONS

2.1 Lattices and Quantization

A lattice is a discrete subgroup $\Lambda \subseteq \mathbb{R}^n$. The rank of a lattice is the dimension of the subspace of $\mathbb{R}^n$ that it spans. A lattice is called full-rank if its rank equals its dimension.

Definition 1 (Partition Cell). *A partition cell of the lattice Λ is a bounded set $\mathcal{P}_\Lambda$ that satisfies the following properties:*

1. *Covering Property: The union of translates of $\mathcal{P}_\Lambda$ by lattice points covers the entire space $\mathbb{R}^n$, i.e., $\cup_{\mathbf{v}\in\Lambda}(\mathbf{v} + \mathcal{P}_\Lambda) = \mathbb{R}^n$.*
2. *Partitioning Property: For any pair of distinct lattice points $\mathbf{v}$ and $\mathbf{w}$ in Λ, if their corresponding translated partition cells intersect, then $\mathbf{v}$ must equal $\mathbf{w}$.*

For instance, the half-open Voronoi cell $\mathcal{V}_\Lambda$ is a partition cell. This cell encompasses the set of points in $\mathbb{R}^n$ that are closer to a specific lattice point (referred to as the generating lattice point) within Λ than to any other lattice point. Essentially, it defines the region surrounding each generating lattice point where it is the closest lattice point.

A quantization function maps a vector $\mathbf{y} \in \mathbb{R}^n$ to the nearest lattice point in Λ. This is formulated as:

$$Q_\Lambda(\mathbf{y}) = \arg\min_{\lambda \in \Lambda} \|\mathbf{y} - \lambda\| \tag{1}$$

where the implicit selection of a half-open Voronoi cell is crucial, as it allows Q_Λ to consistently choose a single representative when multiple lattice points are equidistant from $\mathbf{y}$.

Definition 2 (Dithered quantizer). *A dithered quantizer over lattice Λ is defined by sampling $\mathbf{g} \leftarrow \mathcal{V}_\Lambda$ and outputting*

$$Q_{\Lambda+\mathbf{g}}(\mathbf{y}) = \mathbf{g} + Q_\Lambda(\mathbf{y} - \mathbf{g}). \tag{2}$$

Definition 3 (Second moment). *The second moment of a lattice is defined as the second moment per dimension of a random variable $\mathbf{u}$ which is uniformly distributed over the fundamental Voronoi cell $\mathcal{V}$:*

$$\tilde{\sigma}^2(\Lambda) = \frac{1}{n}\mathbb{E}\|\mathbf{u}\|^2 = \frac{1}{n}\frac{1}{\det(\Lambda)}\int_{\mathcal{V}} \|\mathbf{x}\|^2 \, d\mathbf{x}$$

where $\mathbb{E}$ denotes expectation, and $\det(\Lambda)$ is the volume of a Voronoi cell.

For a dithered quantizer, $\mathbf{y} - Q_{\Lambda+\mathbf{g}}(\mathbf{y})$ is uniformly distributed over $\mathcal{V}_\Lambda$, so the averaged quantization error of the dithered quantizer can be quantified by $\tilde{\sigma}^2(\Lambda)$: for any distribution of $\mathbf{y}$, with $\mathbf{g} \leftarrow \mathcal{V}_\Lambda$, then

$$\frac{1}{n}\mathbb{E}\left\|\mathbf{y} - Q_{\Lambda+\mathbf{g}}(\mathbf{y})\right\|^2 = \tilde{\sigma}^2(\Lambda). \tag{3}$$

The normalized second moment (NSM), i.e., the second-moment to volume ratio, is defined as

$$G(\Lambda) = \frac{\tilde{\sigma}^2(\Lambda)}{\det^{2/n}(\Lambda)}. \tag{4}$$

The minimum possible value of $G(\Lambda)$ over all lattices in $\mathbb{R}^n$ is denoted by G_m.

Definition 4 (Quantization-good). *A sequence of lattices $\Lambda^{(n)}$ with growing dimension is called good for mean squared error quantization if*

$$\lim_{n\to\infty} G(\Lambda^{(n)}) = \frac{1}{2\pi e}. \tag{5}$$

2.2 LWE, LWR, LWQ

This section reviews the definitions of LWE [29] and LWR [7], and presents our generalization called LWQ. For the sake of theoretical analysis, this work considers $\mathbf{s} \in \mathbb{Z}_q^{n*}$ rather than $\mathbf{s} \in \mathbb{Z}_q^n$. This adaption is minor as the probability of $\mathbf{s} \in \mathbb{Z}_q^{n*}$ is at least $1 - O(1/2^n)$ for $\mathbf{s} \in \mathbb{Z}_q^n$.

Definition 5 (LWE/LWR/LWQ distributions). *Let $n, m, q \in \mathbb{N}$, $p \mid q$ and $p \geq 2$, χ^m be a distribution on $\mathbb{Z}_q^m$, and Λ is an m-dimensional integer lattice satisfying $q\mathbb{Z}^m \subset \Lambda \subset \mathbb{Z}^m$.*

- *Samples from LWE distribution: $\mathbf{A} \leftarrow \mathbb{Z}_q^{m\times n}$, $\mathbf{s} \leftarrow \mathbb{Z}_q^{n*}$, $\mathbf{e} \leftarrow \chi^m$, set $\mathbf{b} = \mathbf{As} + \mathbf{e}$ and output $(\mathbf{A}, \mathbf{b}) \in \mathbb{Z}_q^{m\times n} \times \mathbb{Z}_q^m$.*
- *Samples from LWR distribution: $\mathbf{A} \leftarrow \mathbb{Z}_q^{m\times n}$, $\mathbf{s} \leftarrow \mathbb{Z}_q^{n*}$, set $\mathbf{b} = Q_{\frac{q}{p}\mathbb{Z}^m}(\mathbf{As})$ and output $(\mathbf{A}, \mathbf{b}) \in \mathbb{Z}_q^{m\times n} \times (\mathbb{Z}_q^m \cap \frac{q}{p}\mathbb{Z}^m)$ [4].*
- *Samples from LWQ distribution: $\mathbf{A} \leftarrow \mathbb{Z}_q^{m\times n}$, $\mathbf{s} \leftarrow \mathbb{Z}_q^{n*}$, set $\mathbf{b} = Q_\Lambda(\mathbf{As})$ and output $(\mathbf{A}, \mathbf{b}) \in \mathbb{Z}_q^{m\times n} \times (\mathbb{Z}_q^m \cap \Lambda)$*

Definition 6 (LWE/LWR/LWQ problems).

- *Search Problem: Given $(\mathbf{A}, \mathbf{b})$ from the LWE/LWR/LWQ distribution, the Search problem asks to find $\mathbf{s}$.*
- *Decisional Problem: Given $(\mathbf{A}, \mathbf{b})$ from the LWE/LWR/LWQ distribution, the decisional problem asks to distinguish whether $(\mathbf{A}, \mathbf{b})$ is generated from the LWE/LWR/LWQ distribution or a uniform distribution over $\mathbb{Z}_q^{m\times n} \times \mathbb{Z}_q^m$, $\mathbb{Z}_q^{m\times n} \times (\mathbb{Z}_q^m \cap \frac{q}{p}\mathbb{Z}^m)$, $\mathbb{Z}_q^{m\times n} \times (\mathbb{Z}_q^m \cap \Lambda)$.*

In terms of efficiency, both LWR and LWQ sidestep Gaussian sampling by discarding the $\mathbf{e}$ terms. The advantage of the proposed LWQ over LWR is that, due to the sphere-like effective noise $\mathbf{e}_{\text{ef}}$ ($\mathbf{e}_{\text{ef}} = \mathbf{As} - Q_\Lambda(\mathbf{As})$) of LWQ if enabled by a good quantization lattice, its noise term has a smaller second moment.

In terms of computational hardness, LWQ amounts to LWR by setting the quantization lattice as $\Lambda = \frac{q}{p}\mathbb{Z}^m$. The hardness of LWR is often evaluated by running security reduction from "quantized" LWE with $\mathbf{e}$ admitting bounded uniform errors. The hardness of the LWE problem has been justified for specific error distributions χ, under the assumption of worst-case hardness for certain lattice problems. In particular, this holds true when χ is a discrete Gaussian distribution with an appropriate variance.

3 Polar Lattice for Quantization

Our recent work in [21] has shown that polar lattices are good for quantization ($G(\Lambda^{(n)}) \to \frac{1}{2\pi e}$). In this section, we will adopt this polar quantizer to compress the discrete sources. The technical novelty is to prove the quantization noise converges to a discrete Gaussian distribution. This is key to prove the closeness of the LWQ and LWE distributions, therefore justifying the hardness of LWQ.

Polar lattices are an instance of the well-known "Construction D" [13, p.232] which uses a set of nested polar codes as component codes. Thanks to the nice structure of "Construction D", both the encoding and decoding complexity of polar lattices are quasilinear in the block length (*i.e.*, dimension of the lattice). A construction of polar lattices achieving the Shannon capacity of the Gaussian

[4] $\frac{p}{q}(\mathbb{Z}_q^m \cap \frac{q}{p}\mathbb{Z}^m) = \mathbb{Z}_p^m$, so this definition is the same as the conventional LWR definition.

noise channel was presented in [23]. A follow-up work [22] gave a construction of polar lattices to achieve the rate-distortion bound of source coding for Gaussian sources. Note that the two types of polar lattices constructed in [23, 22] are related but not the same (*i.e.*, one for channel coding and the other for source coding). The multilevel structure of polar lattices enables not only efficient encoding and decoding algorithms, but also a layer-by-layer lattice shaping implementation.

3.1 Polar Codes

As a major breakthrough in coding and information theory, polar coding [5] presents arguably the first explicit construction of codes that are capacity-achieving for any binary-input memoryless symmetric channels (BMSCs). Given a BMSC $W : X \to Y$, a polar code with block length $m = 2^t$ selects K rows from the generator matrix $G_m = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]^{\otimes t}$, where $\otimes$ denotes the Kronecker product. The set containing the K row indices is named as the information set $\mathcal{I}$, and its complement set $\mathcal{F}$ is called the frozen set. The combination of N identical copies of W is denoted by $W_m : X^{[m]} \to Y^{[m]}$. After the polarization transform $U^{[m]} = X^{[m]}G_m$, W_m can be successively split into m binary memoryless symmetric subchannels according to the chain rule of mutual information, denoted by $W_N^{(i)} : U^i \to (Y^{[m]}, U^{1:i-1})$ with $1 \le i \le m$. By channel polarization [5], the fraction of good (roughly error-free) subchannels approaches the capacity C of W as $m \to \infty$. Therefore, to achieve the capacity, the K selected row indices of a polar code are corresponding to these good subchannels, while the rest $m - K$ rows are abandoned as if frozen bits are assigned to those non-good subchannels. The quality of a subchannel is generally identified based on its associated Bhattacharyya parameter.

Definition 7. *Given a BMSC W with transition probability $\mathsf{P}_{Y|X}$, the Bhattacharyya parameter $Z \in [0, 1]$ is defined as*

$$Z(W) = Z(X|Y) \triangleq \sum_y \sqrt{\mathsf{P}_{Y|X}(y|0)\mathsf{P}_{Y|X}(y|1)}. \tag{6}$$

In [6], the rate of channel polarization is characterized in terms of the Bhattacharyya parameter as $\lim_{m\to\infty} \mathsf{Pr}\left(Z(W_m^{(i)}) < 2^{-m^\beta}\right) = C$, for any $0 < \beta < 0.5$. For efficient construction of polar codes, $Z(W_m^{(i)})$ can be evaluated using the methods introduced in [33, 27]. The above mentioned channel splitting process also gives rise to a simple decoding algorithm called Successive Cancellation (SC) decoding [5], which executes the maximum a posteriori (MAP) decoding for each subchannel sequentially from $i = 1$ to m. Consequently, by the union bound, the block error probability of the SC decoding can be upper-bounded by $\sum_{i\in\mathcal{I}} Z(W_m^{(i)})$.

In the context of lossy compression, polar codes have been shown to be able to achieve the rate-distortion bound for binary symmetric sources [17]. To achieve a target distortion, a test channel $W : X \to Y$ is built for the source Y and the

reconstruction X. The polar codes for compression are constructed according to the test channel W, with slight modification on the information set, which is defined as $\mathcal{I} \triangleq \{i \in [m] : Z(W_m^{(i)}) < 1 - 2^{-m^\beta}\}$. By the duality between channel coding and source coding, the SC decoding algorithm for polar channel coding can be transformed to the SC encoding algorithm for polar source coding. Given m i.i.d. sources $Y^{[m]}$, the polarized bits $U^{\mathcal{F}}$ are almost independent to $Y^{[m]}$ since $Z(W_m^{(i)}) \geq 1 - 2^{-m^\beta}$ by definition. The compression of $Y^{[m]}$ is achieved by replacing $U^{\mathcal{F}}$ with irrelevant random bits and saving the relevant bits $U^{\mathcal{I}}$, which can be determined from $Y^{[m]}$ and $U^{\mathcal{F}}$ using the SC encoder.

Polar lattices [22] offer an effective solution for achieving the rate-distortion bound in the context of the i.i.d. Gaussian source. In essence, one constructs a polar lattice for the Gaussian source by utilizing a series of nested polar codes, as introduced by Forney *et al.* [15]. These polar codes compress the Gaussian source vector based on the characteristics of the test channel at each level. Moreover, research [23] indicates that employing a binary lattice partition keeps the number of levels r relatively small ($r = O(\log \log m)$), yet still enables the attainment of the capacity $\frac{1}{2}\log(1 + \mathsf{SNR})$ of the additive white Gaussian noise (AWGN) channel, where SNR represents the signal-to-noise ratio.

The concept of duality between source coding and channel coding allows us to interpret quantization polar lattices as analogous to a channel coding lattice constructed on the test channel. In the scenario of a Gaussian source with variance σ_s^2 and an average distortion Δ, the test channel effectively becomes an AWGN channel with a noise variance of Δ. Consequently, the SNR of this test channel equals $\frac{\sigma_s^2 - \Delta}{\Delta}$, while its capacity is $\frac{1}{2}\log\left(\frac{\sigma_s^2}{\Delta}\right)$. This insight suggests that the rate of the polar lattice quantizer can be finely adjusted to approach $\frac{1}{2}\log\left(\frac{\sigma_s^2}{\Delta}\right)$. Consequently, polar lattices demonstrate the capability to achieve the rate-distortion bound of Gaussian sources by employing discrete Gaussian distribution instead of continuous, offering a notable advancement in compression techniques.

3.2 Polar Lattice: Performance Analysis

In this subsection, we present an explicit construction of polar lattices for the quantization of random integers, which produces Gaussian-like quantization errors. Before that, we need some preliminaries on the lattice structure based on multi-level codes [15]. A sublattice $\Lambda' \subset \Lambda$ induces a partition (denoted by Λ/Λ') of Λ into equivalence groups modulo Λ'. The order of the partition is denoted by $|\Lambda/\Lambda'|$, which is equal to the number of the cosets. If $|\Lambda/\Lambda'| = 2$, we call this a binary partition. Let $\Lambda(\Lambda_0)/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'(\Lambda_r)$ for $r \geq 1$ be an n-dimensional lattice partition chain. If only one level is applied ($r = 1$), the construction is known as "Construction A". If multiple levels are used, the construction is known as "Construction D". For each partition $\Lambda_{\ell-1}/\Lambda_\ell$ ($1 \leq \ell \leq r$) a code C_ℓ over $\Lambda_{\ell-1}/\Lambda_\ell$ selects a sequence of coset representatives a_ℓ in a set A_ℓ of representatives for the cosets of Λ_ℓ. This construction requires a set of nested linear binary codes C_ℓ with block length m and dimension of information bits

k_ℓ, which are represented as $[m, k_\ell]$ codes for $1 \leq \ell \leq r$ and $C_1 \subseteq C_2 \cdot \cdot \subseteq C_r$. Let ψ be the natural embedding of $\mathbb{F}_2^m$ into $\mathbb{Z}^m$, where $\mathbb{F}_2$ is the binary field. Consider $\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_m$ be a basis of $\mathbb{F}_2^m$ such that $\mathbf{g}_1, \cdots \mathbf{g}_{k_\ell}$ span C_ℓ. When $n = 1$, the binary lattice L of Construction D consists of all vectors of the form

$$\sum_{\ell=1}^{r} 2^{\ell-1} \sum_{j=1}^{k_\ell} u_\ell^j \psi(\mathbf{g}_j) + 2^r z, \tag{7}$$

where $u_\ell^j \in \{0, 1\}$, $z \in \mathbb{Z}^m$ and ψ denotes the embedding into $\mathbb{R}^m$.

From the perspective of information theory, to compress the source Y which is uniformly random in $[-2^{r-1}, 2^{r-1})$, we build a test channel $Y = X + E \mod q\mathbb{Z}$ ($q = 2^r$), where $E \sim D_{\mathbb{Z},\sigma}$ is a discrete Gaussian noise random variable and the reconstruction X is uniformly random in $[-2^{r-1}, 2^{r-1})$. Our polar lattice quantizer is constructed on this test channel using the binary partition chain $\mathbb{Z}/2\mathbb{Z}/\cdots/2^r\mathbb{Z}$. We also assume that r is sufficiently large such that the modulo $2^r\mathbb{Z}$ operation is insignificant on E. X can be represented by a bit sequence $X_1, ..., X_\ell, ..., X_r$, where X_ℓ specifies the coset $2^{\ell-1}\mathbb{Z}/2^\ell\mathbb{Z}$. Then, $X_1, ..., X_r$ uniquely describes the cosets of $\mathbb{Z}/2^r\mathbb{Z}$. For the first partition level, the quantizer executes the SC decoding to obtain $X_1^{[m]}$ from $Y^{[m]}$, using the statistic of the first partition channel $\mathsf{P}_{Y|X_1}$. For the second level, $X_2^{[m]}$ is decoded from $Y^{[m]}$ and $X_1^{[m]}$ using $\mathsf{P}_{Y,X_1|X_2}$. This process ends until $X_r^{[m]}$ is decoded. Finally, $X^{[m]}$ is recovered as $X^{[m]} = X_1^{[m]} + 2X_2^{[m]} + \cdots + 2^{r-1}X_r^{[m]} \mod 2^r\mathbb{Z}$. We will show that the distribution of $Y^{[m]} - X^{[m]}$ is close to that of m i.i.d. discrete Gaussian random variables.

Because the quantization noise is represented by $q = 2^r$ integers in $[-2^{r-1}, 2^{r-1})$ but not exactly in $\mathbb{Z}$, the following lemma may be needed to show that a discrete Gaussian distribution with truncated tail behaves similarly to the standard one.

Lemma 1. *Let $E \sim D_{\mathbb{Z},\sigma}$ be a discrete Gaussian random variable, and let $E' = E \mod q\mathbb{Z}$ be the residue in $[-2^{r-1}, 2^{r-1})$. The total variation distance between P_E and $\mathsf{P}_{E'}$ is upper-bounded as follows.*

$$V(\mathsf{P}_E, \mathsf{P}_{E'}) = \frac{1}{2} \sum_{e \in \mathbb{Z}} |\mathsf{P}_E(e) - \mathsf{P}_{E'}(e)| \leq M \cdot \exp\left(-\frac{(2^{r-1}-1)^2}{2\sigma^2}\right), \tag{8}$$

where $M = 2/\big(\frac{1}{\sqrt{2\pi\sigma^2}} \sum_{\lambda \in \mathbb{Z}} \exp\big(-\frac{\lambda^2}{2\sigma^2}\big)\big)$.

Proof.

$$V(\mathsf{P}_E, \mathsf{P}_{E'}) = \min \sum_{L \subset \mathbb{Z}/q\mathbb{Z}} \mathsf{Pr}(L) - \mathsf{Pr}(\text{coset leader of } L) \tag{9}$$

$$= \sum_{\lambda=\{-2^{r-1},\ldots,2^{r-1}-1\}} \mathsf{Pr}(\lambda + q\mathbb{Z}) - \mathsf{Pr}(\lambda) \tag{10}$$

$$= \sum_{\lambda=-2^{r-1}-1}^{-\infty} \frac{\exp(-\frac{\lambda^2}{2\sigma^2})}{\sum_{\lambda' \in \mathbb{Z}}(\exp(-\frac{\lambda'^2}{2\sigma^2}))} + \sum_{\lambda=2^{r-1}}^{\infty} \frac{\exp(-\frac{\lambda^2}{2\sigma^2})}{\sum_{\lambda' \in \mathbb{Z}}(\exp(-\frac{\lambda'^2}{2\sigma^2}))} \tag{11}$$

$$\leq 2 \sum_{\lambda=2^{r-1}}^{\infty} \frac{\exp(-\frac{\lambda^2}{2\sigma^2})}{\sum_{\lambda' \in \mathbb{Z}}(\exp(-\frac{\lambda'^2}{2\sigma^2}))} \tag{12}$$

$$\leq 2 \frac{\int_{2^{r-1}-1}^{\infty} \exp(-\frac{t^2}{2\sigma^2}) dt}{\sum_{\lambda' \in \mathbb{Z}}(\exp(-\frac{\lambda'^2}{2\sigma^2}))} \tag{13}$$

$$= M \cdot Q\left(\frac{2^{r-1}-1}{\sigma}\right) \tag{14}$$

$$\leq M \cdot \exp\left(-\frac{(2^{r-1}-1)^2}{2\sigma^2}\right), \tag{15}$$

where $Q(x) = 1 - \Phi(x)$ is the Q-function of a standard normal distribution, and we use $Q(x) \leq \exp(-\frac{x^2}{2})$ in the last inequality. □

We now analyze the distribution of quantization noise. Let $Y^{[m]}$ denote m samples drawn from **As**. The quantization result or the so-called reconstruction of $Y^{[m]}$ is denoted by $X^{[m]}$, which is also in $\mathbb{Z}_q^m$.

- Consider the first case in which the correlation between $Y^{[m]}$ and $X^{[m]}$ is due to an i.i.d. discrete Gaussian random vector $E^{[m]}$, i.e., $Y^i = X^i + E^i \mod q\mathbb{Z}$ for each $i \in [m]$, and $E^i \sim D_{\mathbb{Z},\sigma}$. The joint distribution between $X^{[m]}$ and $Y^{[m]}$ in this case is denoted by $\mathsf{P}_{X^{[m]},Y^{[m]}}$.
- Consider the second case in which the correlation between $Y^{[m]}$ and $X^{[m]}$ is generated by the polar lattice quantizer, i.e., $X^{[m]} = \lfloor Y^{[m]} \rceil_Q$. The joint distribution between $X^{[m]}$ and $Y^{[m]}$ in this case is denoted by $\mathsf{Q}_{X^{[m]},Y^{[m]}}$.

We will show the total variation distance $V(\mathsf{P}_{X^{[m]},Y^{[m]}}, \mathsf{Q}_{X^{[m]},Y^{[m]}})$ vanishes sub-exponentially in m through a layer-by-layer manner, which is corresponding to the multi-level quantization process of polar lattices. Notice that each $X^i \in \mathbb{Z}_q, i \in [m]$ can be uniquely represented by a binary sequence $X_1^i, ..., X_\ell^i, ..., X_r^i$, and X_ℓ^i determines the coset of the binary partition $2^{\ell-1}\mathbb{Z}/2^\ell\mathbb{Z}$ for $1 \leq \ell \leq r$. Given a source vector $Y^{[m]}$, the (m-dimensional) polar lattice quantizer tries to find the coset leader $X_1^{[m]}$ at the first level; then it decides the coset leader $X_2^{[m]}$ at the second level using both $X_1^{[m]}$ and $Y^{[m]}$; the process keeps going at level ℓ, where $X_\ell^{[m]}$ is decoded from $Y^{[m]}$ and $X_{1:\ell-1}^{[m]}$; the process ends at the final r-th level, where $X_r^{[m]}$ is decoded from $Y^{[m]}$ and $X_{1:r-1}^{[m]}$.

From the perspective of lossy compression in information theory, $\mathsf{P}_{Y|X}$ is called the test channel with input (reconstruction) X and output (source) Y. As can be seen, since $Y = X + E \mod q\mathbb{Z}$, the test channel is a discrete additive white Gaussian noise channel with a modulo $q\mathbb{Z}$ operation at the end. Following the step of Forney et al. [15], the test channel can be partitioned into r $2^{\ell-1}\mathbb{Z}/2^{\ell}\mathbb{Z}$ binary-input channels with $1 \le \ell \le r$, which are called binary partition channels.

In fact, the polar lattice consists of the component polar codes designed for these r partition channels. More explicitly, the first level $\mathbb{Z}/2\mathbb{Z}$ partition channel completely determines the joint distribution $\mathsf{P}_{X_1,Y}$ of X_1 and Y, and $Y \mod 2\mathbb{Z}$ is a sufficient statistic of Y with respect to X_1. The polar code C_1 at the first level is constructed according to the $\mathbb{Z}/2\mathbb{Z}$ channel, which is equivalently described by $W_1 : X_1 \underset{\mathsf{P}_{Y|X_1}}{\longrightarrow} Y$. Let $U_1^{[m]} = X_1^{[m]} G_m$ be the bits after channel polarization at level 1. The information set of C_1 is defined as $\mathcal{I}_1 \triangleq \{i \in [m] : Z(U_1^i | U_1^{1:i-1}, Y^{[m]}) \le 1 - 2^{-m^\beta}\}$ for any $0 < \beta < 0.5$, and the frozen set of C_1 is the complement set $\mathcal{F}_1 \triangleq \mathcal{I}_1^c$. By this definition, the correlation between $U_1^{\mathcal{F}_1}$ and $Y^{[m]}$ is negligible. The polar quantizer assigns uniformly random bits that are independent of $Y^{[m]}$ to $U_1^{\mathcal{F}_1}$, and then determines $U_1^{\mathcal{I}_1}$ from $Y^{[m]}$ and $U_1^{\mathcal{F}_1}$ using the SC encoding algorithm. The reconstruction at level 1 is obtained from the inverse polarization transform $X_1^{[m]} = U_1^{[m]} G_m^{-1} = U_1^{[m]} G_m$.

Lemma 2. *Let $\mathsf{Q}_{U_1^{[m]},Y^{[m]}}$ denote the resulted joint distribution of $U_1^{[m]}$ and $Y^{[m]}$ according to the encoding rules* (17) *and* (18) *at the first partition level. Let $\mathsf{P}_{U_1^{[m]},Y^{[m]}}$ denote the joint distribution directly generated from $\mathsf{P}_{X_1,Y}$, i.e., U_1^i is generated according to the encoding rule* (17) *for all $i \in [m]$. The total variation distance between $\mathsf{P}_{U_1^{[m]},Y^{[m]}}$ and $\mathsf{Q}_{U_1^{[m]},Y^{[m]}}$ is upper-bounded as follows.*

$$V\left(\mathsf{P}_{U_1^{[m]},Y^{[m]}}, \mathsf{Q}_{U_1^{[m]},Y^{[m]}}\right) \le m\sqrt{\ln 2 \cdot 2^{-m^\beta}}. \tag{16}$$

.

$$U_1^i = \begin{cases} 0 \text{ w. p. } P_{U_1^i|U_1^{1:i-1},Y^{[m]}}\left(0|u_1^{1:i-1}, y^{[m]}\right) \\ 1 \text{ w. p. } P_{U_1^i|U_1^{1:i-1},Y^{[m]}}\left(1|u_1^{1:i-1}, y^{[m]}\right) \end{cases} \text{if } i \in \mathcal{I}_1, \tag{17}$$

$$U_1^i = \begin{cases} 0 \text{ w. p. } \dfrac{1}{2} \\ 1 \text{ w. p. } \dfrac{1}{2}. \end{cases} \text{if } i \in \mathcal{F}_1, \tag{18}$$

Proof. See Appendix I.

Remark 1. It seems that the encoding rules (17) and (18) are not deterministic. We note that the randomized forms in (17) and (18) are just for convenience of proof. By the symmetry of the $\mathbb{Z}/2\mathbb{Z}$ channel, it can be shown that any fixed realization $U_1^{\mathcal{F}_1} = u_1^{\mathcal{F}_1}$ causes the same total variation distance [17], meaning that

one can safely choose all-zero frozen bits in practice. Similarly, by the polarization effect, the bit U_1^i for $i \in \mathcal{I}_1$ has conditional entropy $H(U_1^i|U_1^{1:i-1}, Y^{[m]}) \to 0$ almost surely as $m \to \infty$. The rule (17) can be replaced with a deterministic MAP rule.

After finishing the encoding at level 1, the polar lattice quantizer proceeds to level 2 in a similar manner. The $2\mathbb{Z}/4\mathbb{Z}$ partition channel completely determines the joint distribution $\mathsf{P}_{X_2,Y|X_1}$ of X_2 and Y given the previous quantization result X_1, and $Y - X_1 \mod 4\mathbb{Z}$ is a sufficient statistic of Y with respect to X_2. The polar code C_2 at the second level is constructed according to the $2\mathbb{Z}/4\mathbb{Z}$ channel, which is equivalently described by $W_2 : X_2 \underset{\mathsf{P}_{Y,X_1|X_2}}{\longrightarrow} (Y, X_1)$. Let $U_2^{[m]} = X_2^{[m]} G_m$ be the bits after channel polarization at level 2. The information set of C_2 is defined as $\mathcal{I}_2 \triangleq \{i \in [m] : Z(U_2^i|U_2^{1:i-1}, X_1^{[m]}, Y^{[m]}) \le 1 - 2^{-m^\beta}\}$, and the frozen set is defined as $\mathcal{F}_2 \triangleq \mathcal{I}_2^c$.

Lemma 3. *Let $\mathsf{Q}_{U_1^{[m]},U_2^{[m]},Y^{[m]}}$ denote the resulted joint distribution of $U_1^{[m]}$, $U_2^{[m]}$ and $Y^{[m]}$ according to the encoding rules* (17) *and* (18) *at the first partition level, and then rules* (20) *and* (21) *at the second partition level. Let $\mathsf{P}_{U_1^{[m]},U_2^{[m]},Y^{[m]}}$ denote the joint distribution directly generated from $\mathsf{P}_{X_1,X_2,Y}$, i.e., U_1^i and U_2^i are generated according to the encoding rule* (17) *and rule* (20) *for all $i \in [m]$, respectively. The total variation distance between $\mathsf{P}_{U_1^{[m]},U_2^{[m]},Y^{[m]}}$ and $\mathsf{Q}_{U_1^{[m]},U_2^{[m]},Y^{[m]}}$ is upper-bounded as follows.*

$$V\left(\mathsf{P}_{U_1^{[m]},U_2^{[m]},Y^{[m]}}, \mathsf{Q}_{U_1^{[m]},U_2^{[m]},Y^{[m]}}\right) \le 2m\sqrt{\ln 2 \cdot 2^{-m^\beta}}. \tag{19}$$

.

$$U_1^i = \begin{cases} 0 \text{ w. p. } P_{U_2^i|U_2^{1:i-1},X_1^{[m]},Y^{[m]}}\left(0|u_2^{1:i-1}, x_1^{[m]}, y^{[m]}\right) \\ 1 \text{ w. p. } P_{U_2^i|U_2^{1:i-1},X_1^{[m]},Y^{[m]}}\left(1|u_2^{1:i-1}, x_1^{[m]}, y^{[m]}\right) \end{cases} \text{if } i \in \mathcal{I}_2, \tag{20}$$

$$U_2^i = \begin{cases} 0 \text{ w. p. } \dfrac{1}{2} \\ 1 \text{ w. p. } \dfrac{1}{2}. \end{cases} \text{if } i \in \mathcal{F}_2, \tag{21}$$

Proof. Assume an auxiliary joint distribution $\mathsf{Q}'_{U_1^{[m]},U_2^{[m]},Y^{[m]}}$ resulted from using the encoding rule (17) for all U_1^i with $i \in [m]$ at the first partition level, and rules (20) and (21) at the second partition. Clearly, $\mathsf{Q}'_{U_1^{[m]},Y^{[m]}} = \mathsf{P}_{U_1^{[m]},Y^{[m]}}$ and $\mathsf{Q}'_{U_2^{[m]}|U_1^{[m]},Y^{[m]}} = \mathsf{Q}_{U_2^{[m]}|U_1^{[m]},Y^{[m]}}$. By the triangle inequality,

$$\begin{aligned} &V\left(\mathsf{P}_{U_1^{[m]},U_2^{[m]},Y^{[m]}}, \mathsf{Q}_{U_1^{[m]},U_2^{[m]},Y^{[m]}}\right) \\ &\le V\left(\mathsf{P}_{U_1^{[m]},U_2^{[m]},Y^{[m]}}, \mathsf{Q}'_{U_1^{[m]},U_2^{[m]},Y^{[m]}}\right) + V\left(\mathsf{Q}'_{U_1^{[m]},U_2^{[m]},Y^{[m]}}, \mathsf{Q}_{U_1^{[m]},U_2^{[m]},Y^{[m]}}\right), \end{aligned} \tag{22}$$

where the first term in the right hand side can be upper bounded by $m\sqrt{\ln 2 \cdot 2^{-m^\beta}}$ using the same method as in the proof of Lemma 2, and the second term is equal to $V\left(\mathsf{P}_{U_1^{[m]},Y^{[m]}}, \mathsf{Q}_{U_1^{[m]},Y^{[m]}}\right)$. □

After the lattice quantization process with r sequential levels, the joint distribution produced by the lattice quantizer is denoted by $\mathsf{Q}_{U_{1:r}^{[m]},Y^{[m]}}$, and the joint distribution directly generated from m i.i.d. test channels is denoted by $\mathsf{P}_{U_{1:r}^{[m]},Y^{[m]}}$. By induction, we obtain $V\left(\mathsf{P}_{U_{1:r}^{[m]},Y^{[m]}}, \mathsf{Q}_{U_{1:r}^{[m]},Y^{[m]}}\right) \leq rm\sqrt{\ln 2 \cdot 2^{-m^\beta}}$. Combining this result with Lemma 1, we have the following theorem on the distribution of quantization noise.

Theorem 1. *The distribution of quantization noise induced by the polar lattice can be rendered arbitrarily close to a discrete Gaussian distribution in terms of the total variation distance, ensuring that*

$$\begin{aligned} &V\left(\mathsf{P}_{X'^{[m]},Y^{[m]}}, \mathsf{Q}_{X^{[m]},Y^{[m]}}\right) \\ &\quad \leq r \cdot m\sqrt{\ln 2 \cdot 2^{-m^\beta}} + M \cdot m \cdot \exp\left(-\frac{(2^{r-1}-1)^2}{2\sigma^2}\right). \end{aligned} \tag{23}$$

Proof. By the inverse polarization transform $X_\ell^{[m]} = U_\ell^{[m]} G_m$ from $\ell = 1$ to r, we immediately have $V\left(\mathsf{P}_{X^{[m]},Y^{[m]}}, \mathsf{Q}_{X^{[m]},Y^{[m]}}\right) \leq r \cdot m\sqrt{\ln 2 \cdot 2^{-m^\beta}}$.

Recall that the test channel $X \underset{\mathsf{P}_{Y|X}}{\longrightarrow} Y$ is given by $Y = X + E \mod q\mathbb{Z}$, where $E \sim D_{\mathbb{Z},\sigma}$. Suppose now P_Y is fixed, and $\mathsf{P}_{X|Y}$ is replaced with $\mathsf{P}_{X'|Y}$ by removing the modulo $q\mathbb{Z}$ operation, i.e., $X' = Y - E$. The total variation distance $V(\mathsf{P}_{X'^{[m]},Y^{[m]}}, \mathsf{P}_{X^{[m]},Y^{[m]}})$ is equal to $V(\mathsf{P}_{E'^{[m]}}, \mathsf{P}_{E^{[m]}})$ as shown in Lemma 1. By using the telescoping expansion (54) and the triangle inequality again, the proof is completed. □

Fig. 1 shows a comparison between the distribution of quantization noise $Y - X$ achieved by the polar lattice quantizer and the genuine discrete Gaussian distribution $D_{\mathbb{Z},\sigma}$ with parameters $\sigma = 3$, $r = 8$ and $m = 2^{20}$.

Remark 2. We note that the validity of polar lattice structure can be easily guaranteed. Taking the above simulation as an example, when constructing multilevel polar codes along the binary partition chain $\mathbb{Z}/2\mathbb{Z}/\cdots/2^r\mathbb{Z}$ for the additive discrete Gaussian test channel ($\sigma = 3$), the capacities of the partition channels from $\ell = 1$ to r are given by 0, 3.2732×10^{-10}, 0.0056, 0.3933, 0.9690, 1.0000 and 1.0000, respectively. The size of the information set is chosen as $|\mathcal{I}_\ell| = \lceil m \cdot C(W_\ell) \rceil$, where $C(W_\ell)$ denotes the capacity of the ℓ-th partition channel. As a result, the component polar codes are consecutively nested by ensuring $\mathcal{I}_\ell \subseteq \mathcal{I}_{\ell+1}$ for $1 \leq \ell \leq r-1$, and we have an ascertained polar lattice quantizer. Moreover, the constructed polar lattice is roughly sphere-bound achieving, by the capacity-achieving property of polar codes for all partition levels.

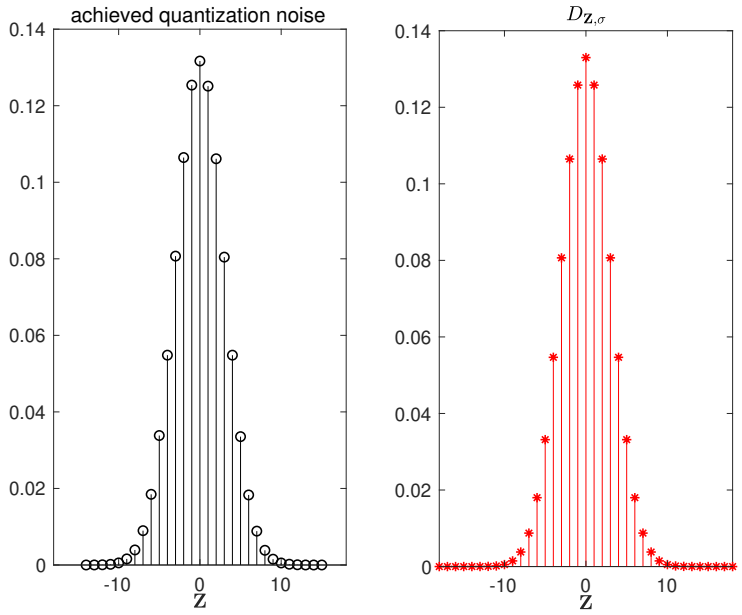

Fig. 1. A comparison between the distribution of quantization noise $Y-X$ and $D_{\mathbb{Z},\sigma_s=3}$.

4 Hardness of LWQ

By extending the LWE distribution to a LWR/LWQ distribution, the loss of security can be quantified by computing the Rényi divergence between the two distributions.

Definition 8 (Rényi divergence). *Rényi divergence of order α between two discrete distributions X and Y is defined as*

$$D_\alpha(X||Y) = \frac{1}{\alpha-1}\ln \sum_{t\in \mathrm{supp}X} X(t)\left(\frac{X(t)}{Y(t)}\right)^{\alpha-1}. \tag{24}$$

For the sake of simplicity, we set $\alpha = 2$ in the following analysis.

Lemma 4. *Let $\mathbf{A} \leftarrow \mathbb{Z}_q^{m\times n}$, $\mathbf{s} \leftarrow \mathbb{Z}_q^{n*}$, $q\mathbb{Z}^m \subset \Lambda \subset \mathbb{Z}^m$. For any $\mathbf{s} \in \mathbb{Z}_q^{n*}$, let $\mathsf{X_s}$ be the distribution of m LWQ samples $(\mathbf{A}, Q_\Lambda(\mathbf{As}))$, and $\mathsf{Y_s}$ be the distribution of m quantized LWE samples $(\mathbf{A}, Q_\Lambda(\mathbf{As}+\mathbf{e}))$. We have*

1. *If $\mathbf{e} \sim \mathcal{V}_{p\Lambda}$ (large noise), $p \geq 2$, $p \in \mathbb{Z}$, then $e^{D_2(\mathsf{X_s}||\mathsf{Y_s})} \leq \frac{1}{p^m}$.*
2. *If $\mathbf{e} \sim \mathcal{V}_{\frac{1}{p}\Lambda}$ (small noise), $p \geq 2$, $p \in \mathbb{Z}$, then $e^{D_2(\mathsf{X_s}||\mathsf{Y_s})} \leq \left(\frac{p-1}{p}\right)^m + \frac{1}{p^m}\times\frac{1}{2^m}$.*

Proof. Since $\mathbf{s} \in \mathbb{Z}^{n*}$, $\mathbf{As}$ admits a uniform distribution in $\mathbb{Z}_q^m$. Let $\chi_\sigma^m \mod q\mathbb{Z}^m$ be the discrete Gaussian distribution after modulo q. Using the definition of Rényi divergence, we have

$$e^{D_2(\mathsf{X}_\mathbf{s}||\mathsf{Y}_\mathbf{s})} = \mathsf{E}_{\mathsf{X}_\mathbf{s}} \frac{\mathsf{Pr}(\mathsf{X}_\mathbf{s} = (\mathbf{A}, Q_\Lambda(\mathbf{As})))}{\mathsf{Pr}(\mathsf{Y}_\mathbf{s} = (\mathbf{A}, Q_\Lambda(\mathbf{As})))} \tag{25}$$

$$= \mathsf{E}_{\mathbf{A}\leftarrow\mathbb{Z}_q^{m\times n}} \frac{1}{\mathsf{Pr}_\mathbf{e}(Q_\Lambda(\mathbf{As}) = Q_\Lambda(\mathbf{As}+\mathbf{e}) \mod q\mathbb{Z}^m)} \tag{26}$$

$$= \mathsf{E}_{\mathbf{u}\leftarrow\mathbb{Z}_q^m} \frac{1}{\mathsf{Pr}_\mathbf{e}(Q_\Lambda(\mathbf{u}) = Q_\Lambda(\mathbf{u}+\mathbf{e}) \mod q\mathbb{Z}^m)} \tag{27}$$

$$= \mathsf{E}_{\mathbf{u}\leftarrow\mathcal{V}_\Lambda} \frac{1}{\mathsf{Pr}_\mathbf{e}(Q_\Lambda(\mathbf{u}+\mathbf{e}) = \mathbf{0} \mod q\mathbb{Z}^m)} \tag{28}$$

$$\leq \mathsf{E}_{\mathbf{u}\sim\mathcal{V}_\Lambda} \frac{1}{\mathsf{Pr}_\mathbf{e}(Q_\Lambda(\mathbf{u}+\mathbf{e}) = \mathbf{0})} \tag{29}$$

If $\mathbf{e} \sim \mathcal{V}_{p\Lambda}$, $\mathsf{Pr}_\mathbf{e}(Q_\Lambda(\mathbf{u}+\mathbf{e}) = \mathbf{0}) = \frac{|\mathcal{V}_\Lambda|}{|\mathcal{V}_{p\Lambda}|} = \frac{1}{p^m}$ for any $\mathbf{u} \leftarrow \mathcal{V}_\Lambda$, thus for Case 1 we have

$$\mathsf{E}_{\mathbf{u}\sim\mathcal{V}_\Lambda} \frac{1}{\mathsf{Pr}_\mathbf{e}(Q_\Lambda(\mathbf{u}+\mathbf{e}) = \mathbf{0})} = \frac{1}{\mathsf{Pr}_\mathbf{e}(Q_\Lambda(\mathbf{e}) = \mathbf{0})} = \frac{1}{p^m}. \tag{30}$$

If $\mathbf{e} \sim \mathcal{V}_{\frac{1}{p}\Lambda}$, this Voronoi region can be partitioned into two parts: the first part that corresponds to $Q_\Lambda(\mathbf{u}+\mathbf{e}) = \mathbf{0}$ has probability $\left(\frac{p-1}{p}\right)^m$ over $\mathbf{u}$, while the second part that corresponds to $\mathsf{Pr}_\mathbf{e}(Q_\Lambda(\mathbf{u}+\mathbf{e}) = \mathbf{0}) \geq \frac{1}{2^m}$ has probability $\frac{1}{p^m}$ over $\mathbf{u}$. Then for Case 2 we have

$$\mathsf{E}_{\mathbf{u}\sim\mathcal{V}_\Lambda} \frac{1}{\mathsf{Pr}_\mathbf{e}(Q_\Lambda(\mathbf{u}+\mathbf{e}) = \mathbf{0})} \leq \left(\frac{p-1}{p}\right)^m + \frac{1}{p^m} \times \frac{1}{2^m}. \tag{31}$$

□

Lemma 5 ([9]). *For any two distributions* X *and* Y*, for any event* E*,*

$$\mathsf{Pr}(\mathsf{Y} \in E) \geq \mathsf{Pr}(\mathsf{X} \in E)^2 / e^{D_2(\mathsf{X}||\mathsf{Y})}. \tag{32}$$

Combining the above lemmas, we arrive at the following theorem. The proof is straightforward and omitted.

Theorem 2. *Let* $\mathbf{A} \leftarrow \mathbb{Z}_q^{m\times n}$, $\mathbf{s} \leftarrow \mathbb{Z}_q^{n*}$, $\mathbb{Z}_q^m \subset \Lambda \subset \mathbb{Z}^m$. *For every algorithm* Learn*, we have*

1. *If* $\mathbf{e} \sim \mathcal{V}_{p\Lambda}$, $p \geq 2$, $p \in \mathbb{Z}$, *then*

$$\mathsf{Pr}_{\mathbf{A},\mathbf{s},\mathbf{e}}\left(\mathsf{Learn}(\mathbf{A}, Q_\Lambda(\mathbf{As}+\mathbf{e})) = \mathbf{s}\right) \geq p^m \mathsf{Pr}_{\mathbf{A},\mathbf{s},\mathbf{e}}\left(\mathsf{Learn}(\mathbf{A}, Q_\Lambda(\mathbf{As})) = \mathbf{s}\right)^2. \tag{33}$$

2. If $\mathbf{e} \sim \mathcal{V}_{\frac{1}{p}\Lambda}$, $p \geq 2$, $p \in \mathbb{Z}$, *then*

$$\begin{aligned}&\mathsf{Pr}_{\mathbf{A},\mathbf{s},\mathbf{e}}\left(\mathsf{Learn}(\mathbf{A}, Q_\Lambda(\mathbf{As}+\mathbf{e})) = \mathbf{s}\right)\\ &\geq 1/\left(\left(\frac{p-1}{p}\right)^m + \frac{1}{p^m} \times \frac{1}{2^m}\right)\mathsf{Pr}_{\mathbf{A},\mathbf{s},\mathbf{e}}\left(\mathsf{Learn}(\mathbf{A}, Q_\Lambda(\mathbf{As})) = \mathbf{s}\right)^2. \quad (34)\end{aligned}$$

Remark 3. For general sub-Gaussian $\mathbf{e}, \mathbf{u}$, we can prove a lower bound for the function

$$f(\mathbf{u}, \mathbf{e}) = \mathsf{E}_{\mathbf{u}} \frac{1}{\mathsf{Pr}_{\mathbf{e}}(Q_\Lambda(\mathbf{u}+\mathbf{e}) = \mathbf{0})}. \quad (35)$$

By Jensen's inequality and the fact that $\frac{1}{\mathsf{Pr}_{\mathbf{e}}(Q_\Lambda(\mathbf{u}+\mathbf{e})=\mathbf{0})}$ is strictly convex for $\mathsf{Pr}_{\mathbf{e}}(Q_\Lambda(\mathbf{u}+\mathbf{e}) = \mathbf{0}) \in \mathbb{R}^+$, we obtain

$$\begin{aligned}f(\mathbf{u}, \mathbf{e}) &\geq \frac{1}{\mathsf{E}_{\mathbf{u}}\mathsf{Pr}_{\mathbf{e}}(Q_\Lambda(\mathbf{u}+\mathbf{e}) = \mathbf{0})} \quad (36)\\ &= \frac{1}{\mathsf{Pr}_{\mathbf{e},\mathbf{u}}(Q_\Lambda(\mathbf{u}+\mathbf{e}) = \mathbf{0})}. \quad (37)\end{aligned}$$

Since both $\mathbf{u}$ and $\mathbf{v}$ are sub-Gaussian variables, the result can be proved by analyzing the error probability of sub-Gaussian.

Remark 4. If $\Lambda = \frac{q}{p}\mathbb{Z}^m$, and $\mathbf{e}$ admits i.i.d. Gaussian of standard deviation σ, with $1 \leq B \leq \frac{q}{2p}$, we have

$$e^{D_2(\mathsf{X}_\mathbf{s}||\mathsf{Y}_\mathbf{s})} \leq \left(\frac{q-2Bp}{q(\Phi(B/\sigma) - \Phi(-B/\sigma))} + \frac{4Bp}{q}\right)^m, \quad (38)$$

where $\Phi(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right)$ is the cumulative distribution function (CDF) of a standard normal distribution. Empirically, this function achieves the smallest value when $B = 2\sigma$.

If $\Lambda = \frac{q}{p}\mathbb{Z}^m$, and $\mathbf{e}$ admits i.i.d. symmetric bounded noise in $\{-B, ..., B\}$, with $1 \leq B \leq \frac{q}{2p}$, we have

$$e^{D_2(\mathsf{X}_\mathbf{s}||\mathsf{Y}_\mathbf{s})} \leq \left(1 + \frac{2Bp}{q}\right)^m. \quad (39)$$

5 The Proposed Secure Source Coding Scheme

We propose a novel approach for joint quantization and encryption of a source vector $\mathbf{m} \in \mathbb{Z}_q^m$. It's worth noting that for a bounded rational source $\tilde{\mathbf{m}} \in \frac{L}{q}\mathbb{Z}^m$ (where $\tilde{\mathbf{m}} \in [-L/2, L/2)^m$), it can be converted to reside in $\mathbb{Z}_q^m$ by setting $\mathbf{m} = \frac{q}{L}\tilde{\mathbf{m}} \in \mathbb{Z}_q^m$.

The quancryption scheme $\mathsf{LWQ}_\Lambda^{m,n,q}$, depicted in Figure 2, is parameterized by the source dimension m, secret dimension n, modulus q, and a quantization

$\mathsf{KGen}(1^n)$	$\mathsf{Enc}_{\mathbf{s}}(\mathbf{m})$	$\mathsf{Dec}_{\mathbf{s}}(\mathbf{A}, \mathbf{b})$
$\mathbf{s} \leftarrow \mathbb{Z}_q^{n*}$	$\mathbf{A} \leftarrow \mathbb{Z}_q^{n \times m}$	Return $\mathbf{b} - \mathbf{A}\mathbf{s}$
Return $\mathbf{s}$	$\mathbf{b} = Q_\Lambda(\mathbf{A}\mathbf{s} + \mathbf{m})$	
	Return $(\mathbf{A}, \mathbf{b})$	

Fig. 2. The quancryption scheme $\mathsf{LWQ}_\Lambda^{m,n,q}$ with $\mathbf{m} \in \mathbb{Z}_q^m$.

lattice Λ, where $q\mathbb{Z}^m \subset \Lambda \subset \mathbb{Z}^m$. It consists of three components: key generation (KGen), encryption (Enc), and decryption (Dec).

In the encryption process, let $\mathbf{A} \leftarrow \mathbb{Z}_q^{m \times n}$ and $\mathbf{s} \leftarrow \mathbb{Z}_q^{n*}$. The ciphertext is generated as follows:

$$\mathsf{Enc}_{\mathbf{s}}(\mathbf{m}) = (\mathbf{A}, \mathbf{b} = Q_\Lambda(\mathbf{A}\mathbf{s} + \mathbf{m})) \in \mathbb{Z}_q^{m \times n} \times (\mathbb{Z}_q^m \cap \Lambda). \tag{40}$$

The decryption process retrieves an estimate $\hat{\mathbf{m}}$ of the original message $\mathbf{m}$:

$$\hat{\mathbf{m}} = \mathbf{b} - \mathbf{A}\mathbf{s} \in \mathbb{Z}_q^m. \tag{41}$$

Unlike traditional secret key encryption, $\hat{\mathbf{m}} \neq \mathbf{m}$ in general. However, the pseudorandom signal $\mathbf{A}\mathbf{s}$ serves as a form of dithering, ensuring that $\mathbf{m} - \hat{\mathbf{m}}$ becomes independent of $\mathbf{m}$. The correctness of quancryption is evaluated based on how closely $\hat{\mathbf{m}}$ approximates $\mathbf{m}$.

Definition 9 (Correctness). *The decryption* $\hat{\mathbf{m}}$ *is considered correct if*

$$Q_{\Lambda + \mathbf{m}}(\hat{\mathbf{m}}) = \mathbf{0}, \tag{42}$$

which implies that $\mathbf{m} - \hat{\mathbf{m}} \in \mathcal{V}_\Lambda$.

The performance of quancryption is assessed using the following metrics:

- Rate of ciphertext:

$$R_C = \frac{1}{m} \log \frac{q^m}{\det \Lambda} \text{ bits/dimension.} \tag{43}$$

- Source to ciphertext ratio:

$$r_{SC} = \frac{\log q^m}{\log q^m - \log \det \Lambda}, \tag{44}$$

  which represents the source rate $\frac{1}{m} \log q^m$ divided by R_C.
- Mean square error (MSE):

$$\text{MSE} = \frac{1}{m} \mathbb{E} \left\| \mathbf{m} - \hat{\mathbf{m}} \right\|^2, \tag{45}$$

  where the expectation is taken over the randomness of $\mathbf{A}$.

Remark 5. If the input message of $\mathsf{LWQ}_{\Lambda}^{m,n,q}$ has already been quantized, i.e., $\mathbf{m} \in \mathbb{Z}_q^m \cap \Lambda$, the decryption algorithm can be modified as $Q_\Lambda(\mathsf{Dec}_{\mathbf{s}}(\cdot))$. In this case, we obtain error-free decryption:

$$Q_\Lambda(\mathsf{Dec}_{\mathbf{s}}(\mathsf{Enc}_{\mathbf{s}}(\mathbf{m}))) = \mathbf{m}. \tag{46}$$

The source to ciphertext ratio of this scheme is $r_{SC} = 1$, which surpasses those in existing literature [25, Table 1]. This improvement is attributed to the use of the lattice code $\mathbb{Z}_q^m \cap \Lambda$ for both error correction and ciphertext compression.

5.1 Cryptographic Properties

We use the security notion of RND-CPA which is better suited to lattice-based primitives, as RND-CPA security implies IND-CPA security [25].

Definition 10 (RND-CPA). *An encryption scheme* $(\mathsf{KGen}, \mathsf{Enc}, \mathsf{Dec})$ *is said to be pseudorandom under chosen plaintext attack if any efficient (probabilistic polynomial-time) adversary* $\mathcal{A}$ *can only achieve at most negligible advantage in the following game, parameterized by a bit* $b \in \{0, 1\}$*:*

1. $\mathsf{sk} \leftarrow \mathsf{KGen}(1^n)$,
2. $b' \leftarrow \mathcal{A}^{O_b(\cdot)}$ *where* $O_b(m)$ *returns either an encryption* $\mathsf{Enc}_{\mathsf{sk}}(m)$ *of the message* m *under the key* sk *if* $b = 0$*, or a sample from a distribution that has support* $\{\mathsf{Enc}_{\mathsf{sk}}(m) \mid \mathsf{sk} \in \mathrm{supp}(\mathsf{KGen}(1^n)), m \in \mathcal{M}\}$ *if* $b = 1$.

The adversary's advantage is defined as $\mathrm{Adv}(\mathcal{A}) = |\Pr(b' = 1|b = 1) - \Pr(b' = 0|b = 1)|$.

Theorem 3. *The quancryption scheme* $\mathsf{LWQ}_{\Lambda}^{m,n,q}$ *is RND-CPA secure if the decisional LWQ problem is hard.*

Proof. In the RND-CPA game of quancryption, the support of is $\mathbb{Z}_q^{m\times n} \times (\mathbb{Z}_q^m \cap \Lambda)$. The hardness of LWQ implies that $\mathcal{A}^{O_b(\mathbf{m}=\mathbf{0})}$ is negligible. For the set of vectors $\mathbf{v}_0, ..., \mathbf{v}_i \in \mathcal{V}_\Lambda \cap \mathbb{Z}^m$, there is a bijection to

$$\mathbf{v}_0 + \mathbf{m} \mod \Lambda, ..., \mathbf{v}_i + \mathbf{m} \mod \Lambda \in \mathcal{V}_\Lambda \cap \mathbb{Z}^m \tag{47}$$

for $\mathbf{m} \in \mathbb{Z}_q^m$. Then the probability distribution of $Q_\Lambda(\mathbf{As})$ is the same as that of $Q_\Lambda(\mathbf{As} + \mathbf{m})$ for $\mathbf{m} \in \mathbb{Z}_q^m$. Thus $\mathcal{A}^{O_b(\mathbf{m})} = \mathcal{A}^{O_b(\mathbf{0})}$ for $\mathbf{m} \in \mathbb{Z}_q^m$, and the RND-CPA security of quancryption can be built upon decisional-LWQ. □

In 2021, Li and Micciancio [18] introduced a model for the passive security of incorrect encryption schemes (IND-CPA^D). In effect, it allows an adversary to decrypt honestly generated ciphertexts, so that a scheme that somehow leaks sensitive information during honest decryption is not seen as secure. In this regard, the proposed quancryption scheme is not IND-CPA^D secure. Nevertheless, a simple solution from the perspective of differential privacy is to add some noises to perturb the decryption result [19].

5.2 Source Coding Properties

Let $\mathcal{P}_\Lambda$ be a general partition cell of Λ, i.e.,

$$\sum_{\lambda\in\Lambda}(\mathcal{P}_\Lambda+\lambda)=\mathbb{R}^m,\ \sum_{\lambda\in\Lambda}(\mathcal{P}_\Lambda\cap\mathbb{Z}^m+\lambda)=\mathbb{Z}^m. \tag{48}$$

It is known that continuous uniform dithers over $\mathcal{P}_\Lambda$ can produce uniform vectors over $\mathcal{V}_\Lambda$. The case of discrete dithers can be proved similarly.

Lemma 6 (Dithered Quantization Error). *If $\mathbf{u}\sim\mathcal{P}_\Lambda\cap\mathbb{Z}^m$, then the dither error $\mathbf{u}+\mathbf{v}-Q_\Lambda(\mathbf{u}+\mathbf{v})$ is uniform over $\mathcal{V}_\Lambda\cap\mathbb{Z}^m$, independent of the source vector $\mathbf{v}\in\mathbb{Z}^m$:*

$$\mathbf{u}+\mathbf{v}-Q_\Lambda(\mathbf{u}+\mathbf{v})\sim\mathcal{V}_\Lambda\cap\mathbb{Z}^m,\quad \textit{for any } \mathbf{v}\in\mathbb{Z}^m.$$

Proof. The function $f(\mathbf{u})=\mathbf{u}-Q_\Lambda(\mathbf{u})\in\mathcal{V}_\Lambda\cap\mathbb{Z}^m$ is a measure-preserving mapping, which implies that for $\mathbf{u}\sim\mathcal{P}_\Lambda\cap\mathbb{Z}^m$, we have $f(\mathbf{u})\sim\mathcal{V}_\Lambda\cap\mathbb{Z}^m$. For any $\mathbf{v}\in\mathbb{Z}^m$, $\mathcal{P}_\Lambda+\mathbf{v}$ also satisfies the general partition cell properties of Eq. (48). Thus $f(\mathbf{u}+\mathbf{v})\sim\mathcal{V}_\Lambda\cap\mathbb{Z}^m$ holds for any $\mathbf{v}\in\mathbb{Z}^m$. □

Theorem 4 (Dithered Quantization Error). *The estimation error of quancryption admits a uniform distribution over the Voronoi region $\mathcal{V}_\Lambda\cap\mathbb{Z}^m$ of the quantization lattice Λ, i.e.,*

$$\mathbf{m}-\hat{\mathbf{m}}\sim\mathcal{V}_\Lambda\cap\mathbb{Z}^m.$$

Proof. In our case, the dither vector $\mathbf{u}=\mathbf{As}$, which admits a uniform distribution over the Voronoi region of the sub-lattice $q\mathbb{Z}^m$, where $q\mathbb{Z}^m\subset\Lambda\subset\mathbb{Z}^m$. So our task is to show that dithers from a partition cell of the sub-lattice can produce uniform error in $\mathcal{V}_\Lambda\cap\mathbb{Z}^m$.

Since $\mathcal{P}_{q\mathbb{Z}^m}=\sum_{\lambda\in q\mathbb{Z}^m\cap\Lambda}(\mathcal{P}_\Lambda+\lambda)$, together with the technique in Lemma 6 showing that each of the $\mathcal{P}_\Lambda+\lambda$ leads to a uniform distribution over $\mathcal{V}_\Lambda\cap\mathbb{Z}^m$, the theorem is proved. □

A direct consequence of the above theorem is that we have

$$\mathrm{MSE}=\frac{1}{m}\mathbb{E}\left\|\mathbf{m}-\hat{\mathbf{m}}\right\|^2\to\tilde{\sigma}^2(\Lambda), \tag{49}$$

where the approximation to $\tilde{\sigma}^2(\Lambda)$ is tight based on the high resolution assumption ([35], here the resolution is q^m).

By substituting $\det(\Lambda)=\left(\tilde{\sigma}^2(\Lambda)/G(\Lambda)\right)^{m/2}$ into the definition of R_C, we have

$$\begin{aligned}R_C&=\log\left(qG(\Lambda)^{1/2}\right)-\log\tilde{\sigma} &(50)\\ &\geq\log\left(q(2\pi e)^{-1/2}\right)-\log\tilde{\sigma}, &(51)\end{aligned}$$

where the equality in Eq. (51) only holds in the asymptotic setting with a quantization-good lattice. This is ciphertext rate to distortion function of quancryption. With chosen q, m and R_C (or $\tilde{\sigma}$), the query of a small $\tilde{\sigma}^2(\Lambda)$ (or R_C) amounts to minimize the NSM $G(\Lambda)$.

The integer lattice $\mathbb{Z}$, checker-board lattice D_4, Gosset lattice E_8, and Leech lattice Λ_{24} have the best reported NSMs in their respective dimensions [1]:

$$0.08333, 0.07660, 0.07168, 0.06577.$$

However, the following lemma shows that the Cartesian product of these lattices can not reach smaller NSM in the large dimensional setting.

Lemma 7. *Assume that $k \mid m$, $k \geq 2$. If Λ is constructed from the m/k-fold Cartesian product of Λ', i.e., $\Lambda = \mathbb{Z}^{m/k} \otimes \Lambda' \subset \mathbb{R}^m$, then the lattices Λ' and Λ have the same NSM, i.e., $G(\Lambda) = G(\Lambda')$.*

Proof. The volume of Λ satisfies

$$\det(\Lambda) = \det(\Lambda')^{m/k}. \tag{52}$$

As $\Lambda = \mathbb{Z}^{m/k} \otimes \Lambda'$, a $\mathbf{x} \in \Lambda$ can be partitioned into m/k independent components $\mathbf{x}_1, ..., \mathbf{x}_{m/k}$, such that $\mathbf{x} = \mathbf{x}_1 + \cdots + \mathbf{x}_{m/k}$. Then we have

$$\frac{1}{m}\mathbb{E}_{\mathbf{x}\sim\mathcal{V}_\Lambda}\|\mathbf{x}\|^2 = \frac{1}{k}\mathbb{E}_{\mathbf{x}_1\sim\mathcal{V}_{\Lambda'}}\|\mathbf{x}_1\|^2. \tag{53}$$

Thus $G(\Lambda) = G(\Lambda')$ can be verified by substituting (52) and (53) into the definition of NSM. □

The proposed secure source coding scheme offers versatile applicability across a spectrum of domains where the confluence of compression and security is paramount. Several notable examples include:

1. **Secure Communication:** In contexts necessitating data transmission over vulnerable channels, such as military communications, IoT (Internet of Things) networks, or telemedicine, secure source quantization may be useful. By compressing data securely prior to transmission, bandwidth requirements are reduced, while concurrently safeguarding the confidentiality of transmitted information.
2. **Privacy-Preserving Machine Learning:** In scenarios demanding the training of machine learning models on sensitive data while upholding privacy, secure source quantization emerges as a salient solution. By securely compressing data before sharing it with third parties or uploading it to the cloud for training, organizations can harness the power of machine learning while fortifying data privacy.
3. **Biometric Data Compression:** Given the sensitivity of biometric data, such as fingerprints, iris scans, or facial recognition data, stringent security measures are imperative. Secure source quantization facilitates the secure compression of biometric data, preserving its confidentiality and integrity against unauthorized access or tampering.

6 Conclusions and Future Work

The paper has explored a novel hardness assumption termed LWQ, similar to the LWR assumption, but is parameterized by an arbitrary lattice Λ (where setting $\Lambda = \frac{q}{p}\mathbb{Z}^m$ recovers LWR). By choosing Λ to be a close-to-optimal lattice quantizer, one can recover a variant of LWR where the noise is Gaussian-like, rather than bounded over an ℓ_∞ ball (which is typical for LWR). This can be used to choose noise of smaller standard deviation at the same security level.

This paper has additionally introduced quancryption, a novel approach that combines source quantization and ciphertext compression into a single process. Utilizing a lattice code $\mathbb{Z}_q^m \cap \Lambda$ as a quantizer, quancryption achieves both tasks in a unified manner, streamlining the overall encryption process. By leveraging the security guarantees offered by LWQ, quancryption facilitates secure dither quantization and achieves a high source-to-ciphertext ratio for lattice-based secret key encryption.

Moving forward, LWQ presents opportunities for adaptation to various cryptography scenarios where LWR serves as a fundamental building block. For instance, we can envision the development of a PKE scheme akin to Lizard [11] or Saber [14] based on LWQ or module-LWQ. Following the structure outlined by Lindner and Peikert [20], such a scheme would incorporate LWE in the key generation phase and LWQ or module-LWQ in the encryption phase. Given the reduced quantization noise inherent in LWQ or module-LWQ, compared to LWR or module-LWR at the same ciphertext rate, we anticipate a lower decryption failure rate for this proposed scheme compared to existing solutions like Lizard and Saber.

7 Appendix I

Proof of Lemma 2.

Proof. Using the telescoping expansion

$$B^{1:n} - A^{1:n} = \sum_{i=1}^{n} (B^i - A^i) A^{1:i-1} B^{i+1:n}, \tag{54}$$

$V\left(\mathsf{P}_{U_1^{[m]},Y^{[m]}},\mathsf{Q}_{U_1^{[m]},Y^{[m]}}\right)$ can be decomposed as

$$
\begin{aligned}
&2V\left(\mathsf{P}_{U_1^{[m]},Y^{[m]}},\mathsf{Q}_{U_1^{[m]},Y^{[m]}}\right)\\
&=\sum_{u_1^{[m]},y^{[m]}}\left|\mathsf{Q}(u_1^{[m]},y^{[m]})-\mathsf{P}(u_1^{[m]},y^{[m]})\right|\\
&=\sum_{u_1^{[m]},y^{[m]}}\left|\sum_i\left(\mathsf{Q}(u_1^i|u_1^{1:i-1},y^{[m]})-\mathsf{P}(u_1^i|u_1^{1:i-1},y^{[m]})\right)\right. \qquad (55)\\
&\qquad\left.\cdot\left(\prod_{j=1}^{i-1}\mathsf{P}(u_1^j|u_1^{1:j-1},y^{[m]})\right)\left(\prod_{j=i+1}^{m}\mathsf{Q}(u_1^j|u_1^{1:j-1},y^{[m]})\right)\mathsf{P}\left(y^{[m]}\right)\right|
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(a)}{\leq}\sum_{i\in\mathcal{F}_1}\sum_{u_1^{[m]},y^{[m]}}\left|\mathsf{Q}(u_1^i|u_1^{1:i-1},y^{[m]})-\mathsf{P}(u_1^i|u_1^{1:i-1},y^{[m]})\right|\left(\prod_{j=1}^{i-1}\mathsf{P}(u_1^j|u_1^{1:j-1},y^{[m]})\right)\\
&\qquad\cdot\left(\prod_{j=i+1}^{m}\mathsf{Q}(u_1^j|u_1^{1:j-1},y^{[m]})\right)\mathsf{P}\left(y^{[m]}\right)\\
&=\sum_{i\in\mathcal{F}_1}\sum_{u_1^{1:i},y^{[m]}}\left|\mathsf{Q}(u_1^i|u_1^{1:i-1},y^{[m]})-\mathsf{P}(u_1^i|u_1^{1:i-1},y^{[m]})\right|\left(\prod_{j=1}^{i-1}\mathsf{P}(u_1^j|u_1^{1:j-1},y^{[m]})\right)\mathsf{P}\left(y^{[m]}\right) \qquad (56)\\
&=\sum_{i\in\mathcal{F}_1}\sum_{u_1^{1:i-1},y^{[m]}}2\mathsf{P}\left(u_1^{1:i-1},y^{[m]}\right)V\left(\mathsf{Q}_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y^{[m]}=y^{[m]}},\mathsf{P}_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y^{[m]}=y^{[m]}}\right)\\
&\overset{(b)}{\leq}\sum_{i\in\mathcal{F}_1}\sum_{u_1^{1:i-1},y^{[m]}}\mathsf{P}\left(u_1^{1:i-1},y^{[m]}\right)\sqrt{2\ln 2D_1\left(\mathsf{P}_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y^{[m]}=y^{[m]}}||\mathsf{Q}_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y^{[m]}=y^{[m]}}\right)}
\end{aligned}
$$

$$
\begin{aligned}
&\overset{(c)}{\leq}\sum_{i\in\mathcal{F}_1}\sqrt{2\ln 2\sum_{u_1^{1:i-1},y^{[m]}}\mathsf{P}\left(u_1^{1:i-1},y^{[m]}\right)D_1\left(\mathsf{P}_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y^{[m]}=y^{[m]}}||\mathsf{Q}_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y^{[m]}=y^{[m]}}\right)}\\
&=\sum_{i\in\mathcal{F}_1}\sqrt{2\ln 2D_1\left(\mathsf{P}_{U_1^i}||\mathsf{Q}_{U_1^i}|U_1^{1:i-1},Y^{[m]}\right)}\\
&\overset{(d)}{=}\sum_{i\in\mathcal{F}_1}\sqrt{2\ln 2\left(1-H(U_1^i|U_1^{1:i-1},Y^{[m]})\right)} \qquad (57)\\
&\overset{(e)}{\leq}\sum_{i\in\mathcal{F}_1}\sqrt{2\ln 2\left(1-Z(U_1^i|U_1^{1:i-1},Y^{[m]})^2\right)}\\
&\overset{(f)}{\leq}m\sqrt{4\ln 2\cdot 2^{-m^\beta}}
\end{aligned}
$$

where $D_1(\cdot||\cdot)$ is the Kullback-Leibler divergence, and the equalities and the inequalities follow from

(a) $\mathsf{Q}\left(u_1^i|u_1^{1:i-1}, y^{[m]}\right) = \mathsf{P}\left(u_1^i|u_1^{1:i-1}, y^{[m]}\right)$ for $i \in \mathcal{I}_1$.
(b) Pinsker's inequality.
(c) Jensen's inequality.
(d) $\mathsf{Q}\left(u_1^i|u_1^{1:i-1}\right) = \frac{1}{2}$ for $i \in \mathcal{F}_1$.
(e) $Z(X|Y)^2 < H(X|Y)$.
(f) Definition of $\mathcal{F}_1$.

□

References

1. Agrell, E., Allen, B.: On the best lattice quantizers. IEEE Transactions on Information Theory **69**(12), 7650–7658 (2023). https://doi.org/10.1109/TIT.2023.3291313
2. Alkim, E., Ducas, L., Pöppelmann, T., Schwabe, P.: Post-quantum key exchange - A new hope. In: Holz, T., Savage, S. (eds.) USENIX Security 2016. pp. 327–343. USENIX Association, Austin, TX, USA (Aug 10–12, 2016)
3. Allen, B., Agrell, E.: The optimal lattice quantizer in nine dimensions. Annalen der Physik **533**(12), 2100259 (2021)
4. Alwen, J., Krenn, S., Pietrzak, K., Wichs, D.: Learning with rounding, revisited - new reduction, properties and applications. In: Canetti, R., Garay, J.A. (eds.) CRYPTO 2013, Part I. LNCS, vol. 8042, pp. 57–74. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 18–22, 2013). https://doi.org/10.1007/978-3-642-40041-4_4
5. Arıkan, E.: Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. IEEE Transactions on Information Theory **55**(7), 3051–3073 (July 2009). https://doi.org/10.1109/TIT.2009.2021379
6. Arıkan, E., Telatar, I.: On the rate of channel polarization. In: Proc. 2009 IEEE Int. Symp. Inform. Theory. pp. 1493–1495. Seoul, South Korea (June 2009)
7. Banerjee, A., Peikert, C., Rosen, A.: Pseudorandom functions and lattices. In: Pointcheval, D., Johansson, T. (eds.) EUROCRYPT 2012. LNCS, vol. 7237, pp. 719–737. Springer, Heidelberg, Germany, Cambridge, UK (Apr 15–19, 2012). https://doi.org/10.1007/978-3-642-29011-4_42
8. Barnes, E., Sloane, N.: The optimal lattice quantizer in three dimensions. SIAM Journal on Algebraic Discrete Methods **4**(1), 30–41 (1983)
9. Bogdanov, A., Guo, S., Masny, D., Richelson, S., Rosen, A.: On the hardness of learning with rounding over small modulus. In: Kushilevitz, E., Malkin, T. (eds.) TCC 2016-A, Part I. LNCS, vol. 9562, pp. 209–224. Springer, Heidelberg, Germany, Tel Aviv, Israel (Jan 10–13, 2016). https://doi.org/10.1007/978-3-662-49096-9_9
10. Cheon, J.H., Kim, A., Kim, M., Song, Y.S.: Homomorphic encryption for arithmetic of approximate numbers. In: Takagi, T., Peyrin, T. (eds.) ASIACRYPT 2017, Part I. LNCS, vol. 10624, pp. 409–437. Springer, Heidelberg, Germany, Hong Kong, China (Dec 3–7, 2017). https://doi.org/10.1007/978-3-319-70694-8_15
11. Cheon, J.H., Kim, D., Lee, J., Song, Y.: Lizard: Cut off the tail! A practical post-quantum public-key encryption from LWE and LWR. In: Catalano, D., De Prisco, R. (eds.) SCN 18. LNCS, vol. 11035, pp. 160–177. Springer, Heidelberg, Germany, Amalfi, Italy (Sep 5–7, 2018). https://doi.org/10.1007/978-3-319-98113-0_9

12. Cohn, H., Kumar, A., Miller, S., Radchenko, D., Viazovska, M.: The sphere packing problem in dimension 24. Annals of Mathematics **185**(3), 1017–1033 (2017). https://doi.org/10.4007/annals.2017.185.3.8
13. Conway, J.H., Sloane, N.J.A.: Sphere Packings, Lattices, and Groups. Springer, New York (1993)
14. D'Anvers, J.P., Karmakar, A., Roy, S.S., Vercauteren, F.: Saber: Module-LWR based key exchange, CPA-secure encryption and CCA-secure KEM. In: Joux, A., Nitaj, A., Rachidi, T. (eds.) AFRICACRYPT 18. LNCS, vol. 10831, pp. 282–305. Springer, Heidelberg, Germany, Marrakesh, Morocco (May 7–9, 2018). https://doi.org/10.1007/978-3-319-89339-6_16
15. Forney, G., Trott, M., Chung, S.Y.: Sphere-bound-achieving coset codes and multilevel coset codes. IEEE Transactions on Information Theory **46**(3), 820–850 (May 2000). https://doi.org/10.1109/18.841165
16. Gray, R.M., Stockham, T.G.: Dithered quantizers. IEEE Transactions on Information Theory **39**(3), 805–812 (1993)
17. Korada, S., Urbanke, R.: Polar codes are optimal for lossy source coding. IEEE Transactions on Information Theory **56**(4), 1751–1768 (April 2010). https://doi.org/10.1109/TIT.2010.2040961
18. Li, B., Micciancio, D.: On the security of homomorphic encryption on approximate numbers. In: Canteaut, A., Standaert, F.X. (eds.) EUROCRYPT 2021, Part I. LNCS, vol. 12696, pp. 648–677. Springer, Heidelberg, Germany, Zagreb, Croatia (Oct 17–21, 2021). https://doi.org/10.1007/978-3-030-77870-5_23
19. Li, B., Micciancio, D., Schultz, M., Sorrell, J.: Securing approximate homomorphic encryption using differential privacy. In: Dodis, Y., Shrimpton, T. (eds.) CRYPTO 2022, Part I. LNCS, vol. 13507, pp. 560–589. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 15–18, 2022). https://doi.org/10.1007/978-3-031-15802-5_20
20. Lindner, R., Peikert, C.: Better key sizes (and attacks) for LWE-based encryption. In: Kiayias, A. (ed.) CT-RSA 2011. LNCS, vol. 6558, pp. 319–339. Springer, Heidelberg, Germany, San Francisco, CA, USA (Feb 14–18, 2011). https://doi.org/10.1007/978-3-642-19074-2_21
21. Liu, L., Lyu, S., Ling, C., Bai, B.: On the quantization goodness of polar lattices. arXiv preprint, arXiv:2405.04051 (2024)
22. Liu, L., Shi, J., Ling, C.: Polar lattices for lossy compression. IEEE Transactions on Information Theory **67**(9), 6140–6163 (2021), https://doi.org/10.1109/TIT.2021.3097965
23. Liu, L., Yan, Y., Ling, C., Wu, X.: Construction of capacity-achieving lattice codes: Polar lattices. IEEE Trans. Commun. **67**(2), 915–928 (Feb 2019)
24. Makhoul, J., Roucos, S., Gish, H.: Vector quantization in speech coding. Proceedings of the IEEE **73**(11), 1551–1588 (1985)
25. Micciancio, D., Schultz, M.: Error correction and ciphertext quantization in lattice cryptography. In: Handschuh, H., Lysyanskaya, A. (eds.) CRYPTO 2023, Part V. LNCS, vol. 14085, pp. 648–681. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 20–24, 2023). https://doi.org/10.1007/978-3-031-38554-4_21
26. Newton, P., Richelson, S.: A lower bound for proving hardness of learning with rounding with polynomial modulus. In: Handschuh, H., Lysyanskaya, A. (eds.) CRYPTO 2023, Part V. LNCS, vol. 14085, pp. 805–835. Springer, Heidelberg, Germany, Santa Barbara, CA, USA (Aug 20–24, 2023). https://doi.org/10.1007/978-3-031-38554-4_26

27. Pedarsani, R., Hassani, S., Tal, I., Telatar, I.: On the construction of polar codes. In: Proc. 2011 IEEE Int. Symp. Inform. Theory. pp. 11–15. St. Petersburg, Russia (July 2011). https://doi.org/10.1109/ISIT.2011.6033724
28. Peikert, C.: Public-key cryptosystems from the worst-case shortest vector problem: extended abstract. In: Mitzenmacher, M. (ed.) 41st ACM STOC. pp. 333–342. ACM Press, Bethesda, MD, USA (May 31 – Jun 2, 2009). https://doi.org/10.1145/1536414.1536461
29. Regev, O.: On lattices, learning with errors, random linear codes, and cryptography. In: Gabow, H.N., Fagin, R. (eds.) 37th ACM STOC. pp. 84–93. ACM Press, Baltimore, MA, USA (May 22–24, 2005). https://doi.org/10.1145/1060590.1060603
30. Schwabe, P., Avanzi, R., Bos, J., Ducas, L., Kiltz, E., Lepoint, T., Lyubashevsky, V., Schanck, J.M., Seiler, G., Stehlé, D., Ding, J.: CRYSTALS-KYBER. Tech. rep., National Institute of Standards and Technology (2022), available at https://csrc.nist.gov/Projects/post-quantum-cryptography/selected-algorithms-2022
31. Shlezinger, N., Chen, M., Eldar, Y.C., Poor, H.V., Cui, S.: Uveqfed: Universal vector quantization for federated learning. IEEE Transactions on Signal Processing **69**, 500–514 (2020)
32. Stevens, T., Skalka, C., Vincent, C., Ring, J., Clark, S., Near, J.P.: Efficient differentially private secure aggregation for federated learning via hardness of learning with errors. In: Butler, K.R.B., Thomas, K. (eds.) USENIX Security 2022. pp. 1379–1395. USENIX Association, Boston, MA, USA (Aug 10–12, 2022)
33. Tal, I., Vardy, A.: How to construct polar codes. IEEE Transactions on Information Theory **59**(10), 6562–6582 (Oct 2013). https://doi.org/10.1109/TIT.2013.2272694
34. Viazovska, M.S.: The sphere packing problem in dimension 8. Annals of Mathematics pp. 991–1015 (2017). https://doi.org/10.4007/annals.2017.185.3.7
35. Zamir, R.: Lattice Coding for Signals and Networks. Cambridge University Press, Cambridge, UK (2014)
36. Zamir, R.: The rate loss in the wyner-ziv problem. IEEE Transactions on Information Theory **42**(6), 2073–2084 (1996). https://doi.org/10.1109/18.556597
37. Zamir, R., Feder, M.: On lattice quantization noise. IEEE Transactions on Information Theory **42**(4), 1152–1159 (1996). https://doi.org/10.1109/18.508838