Polar Coding for Ring-LWE-Based Public Key Encryption

Jiabo Wang*, Cong Ling†, *Member, IEEE*

* Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China, Email:wangjiabo.2013@tsinghua.org.cn

† Deparment of Electrical and Electronic Engineering, Imperial College, London SW7 2AZ, UK, Email:c.ling@imperial.ac.uk

Abstract

Cryptographic constructions based on *ring learning with errors* (RLWE) have emerged as one of the front runners for the standardization of post quantum public key cryptography. As the standardization process continues, optimizing specific parts of proposed schemes becomes a worthwhile endeavor. In this work we focus on using error correcting codes to alleviate a natural trade-off present in most schemes; namely, we would like a wider error distribution to increase security, but a wider error distribution comes at the cost of an increased probability of decryption error. The motivation of this work is to improve the security level of RLWE-based public key encryption (PKE) while keeping the target decryption failure rate (DFR) achievable using error-correcting codes. Specifically, we explore how to implement a family member of error correcting codes, known as polar codes, in RLWE-based PKE schemes in order to effectively lower the DFR. The dependency existing in the additive noise term is handled by mapping every error term (e.g., e, t, s, e_1, e_2) under canonical embedding to the space H where a product in the number field K gives rise to a coordinate-wise product in H. An attempt has been made to make the modulation constellation (message basis) fit in with the canonical basis. Furthermore, we exploit the actuality of some error terms known by the decoder to further lower the DFR. Using our method, the DFR is expected to be as low as 2^{-298} for code rate 0.25, $n = 1024, q = 12289$ and binomial parameter $k = 8$ as is exactly the setting of the post-quantum scheme NewHope; DFR is 2^{-156} for code rate 0.25, $n = 1024, q = 12289, k = 16$. This new DFR margin enables us to improve the security level by 9.4% compared with NewHope. Moreover, polar encoding and decoding have quasi-linear complexity $O(N \log N)$ and they can be implemented in constant time.

Index Terms

Ring LWE; Polar Codes; Public Key Encryption; Error Dependency; Canonical Embedding

I. Introduction

A. *Error-Correcting for Ring-LWE-Based Public Key Encryption*

As the world's top tech companies and research labs compete in the race to build a quantum computer, real world public key cryptography, such as digital signature schemes, public key encryption, and key exchange protocols, must be made quantum resistant. The RLWE problem was introduced in [1] in 2010, expanding on the classical version of the problem (i.e., LWE) introduced by Regev in [2]. Since then, cryptography based on the RLWE problem has become one of the most attractive post quantum candidates. Its quantum security relies on the worst-case approximate SIVP (shortest independent vector problem) on ideal lattices and it gives better efficiency compared with plain LWE because of the ring structure. Many of the prominent submissions to the NIST's call for proposals [3], for example NewHope [4] and LAC [5], are based on RLWE cryptography. Though neither of the two advanced to NIST's third round after a tough decision was made between a module-LWE scheme, Kyber, and the RLWE-based ones, academic and industrial study on RLWE cryptography and their applications never stops. One topic of pressing importance is to refine such schemes for better efficiency and security. In this work, we focus on the issue of error correcting for RLWE-based public key encryption.

Among the RLWE-based public key exchange protocols, there are essentially two major approaches to the problem of sharing a ephemeral secret key that can be used to protect further communication; the reconciliation approach of [6] and the encryption approach of [7]. In the former approach, both parties agree on a shared value from some pseudorandom signals with the help of a robust extractor. It is the latter on which our work is focused, where the protocol resembles the compact RLWE public key encryption scheme first described in [8]. Taking NewHope for example, the binary secret to be shared is encoded with repetition codes, mapped to $\{0, \lfloor q/2 \rfloor\}^n$ and then added to a sequence of error terms jointly produced by both parties. There will be a residue noise term after the decryption operation. Upon getting the decrypted secret, the decoder then sum up the symbols corresponding to the repeated secret bits and infer if the secret should be 0 or 1 according to a threshold. Taking the telecommunication system as an analogy, this is exactly a hard-decision decoding process. Similar ideas can be found in [9], where encoding is done byte-wise rather than bit-wise. Unsurprisingly, this error-correcting scheme cannot give the optimal decoding performance. Decreasing the DFR is of major importance for several reasons. Firstly, if we seek CCA security of the above cryptosystem using the classical Fujisaki-Okamoto transform [10], the NIST standardization targets at a failure rate lower than 2^{-128}. Secondly, more capable error correction can save the bandwidth by allowing a longer ephemeral key to be shared without increasing the length of cipher text or compromising the DFR. Finally, more capable error correction allows larger error term

e, t, s, e_1, e_2, increasing the hardness of the underlying RLWE problem and therefore the security of the cryptosystem.

To improve the error correction and security of RLWE-based PKE, some researchers have exploited the goodness of multi-dimensional lattices. For example, Leech lattice which gives the densest sphere packing of it dimension is used in [11] to encode bit strings as 24-dimensional symbols. The corresponding decoding problem is solved by CVP algorithm. An alternative way to reduce DFR is to apply error-correcting codes (ECC), a method that appears in LAC and is remarked to be distinguishable in NIST report [12]. In [13], Fritzmann et al. considered how much the RLWE-based PKE protocol, NewHope Simple, can profit from BCH codes, LDPC codes, and a hybrid of the two regarding the DFR. They achieve a DFR of 2^{-140} using these codes, but it is unclear whether the decoding algorithms are constant time. In an independent line of work, Saarinen designed a linear block code called XEf and implemented it in Hila5 which is a RLWR (ring learning with rounding)-based PKE scheme [14][1]. This method is able to share 256 bits of message and additional 240 bits of redundancy at DFR below 2^{-128}. The decoding algorithm runs in constant time, which provides resilience to side channel attacks.

How to deal with the dependency existing in the residue noise term of RLWE-based PKE is closely related to the DFR estimation no matter error-correcting codes are applied or not [13], [15], [16]. Fritzmann et al. gave upper bounds on DFR using their error-correcting codes assuming that the residue noise can be seen as independent [13]. They claimed to improve the security level of NewHope Simple by 31.76% for $n = 1024,\ q = 12289$ targeting at DFR=2^{-140}. D'Anvers et al. assume the residue noise is independent conditional on its norm. This method can be applied to calculate the DFR for LAC where the error term is ternary but impractical for wider error distribution and true discrete Gaussian errors [15]. Song et al. formulated the NewHope round 2 as a digital communication system and solved a part of the dependency. They claimed to improve the security by 7.2% ($n = 1024,\ q = 12289$) targeting at a DFR of 2^{-140} as well [16].

B. Contribution

The contribution of this paper is as follows.

1) We are the first to formulate the RLWE-based PKE as an i.i.d. fading channel with channel state information (CSI) available to the receiver without any "independence" assumptions.
    a) Given the residue noise term $e \cdot t - s \cdot e_1 + e_2$, we are the first, without any assumptions, to completely unfasten the dependency between the coefficients by mapping every noise

[1] Hila5 has been withdrawn from the NIST Post Quantum Cryptography standardization process, but the XEf code appears in new entry called Round5.

term to canonical basis where the polynomial multiplication is converted to coordinate-wise multiplication of two vectors[2].

b) We formulate the RLWE-based PKE model as an i.i.d. fading channel with CSI known to the receiver while the transmitter only knows the channel distribution information (CDI). We derive this novel formulation by exploiting the actuality that Alice, at the decryption stage, knows some partial information of the residue noise, i.e., e and s sampled by herself at key generation stage.

c) Taking telecommunication system as an analogy, mapping a single bit 0 or 1 of the plain text to a symbol on the constellation $\{0, \lfloor q/2 \rceil\}$ is called modulation. To make the modulation scheme fit in with the i.i.d. fading channel in canonical basis, we proposed a new modulation scheme at the cost of error tolerance.

2) We give the explicit construction of polar codes for RLWE-based PKE channel model. The encoding and decoding routines allow quasi-linear (i.e., $O(N \log N)$) and constant-time implementations. Experimental results and theoretical estimation of DFR are also given. Specifically, we derive a new DFR of 2^{-298} for $q = 12289, n = 1024, r = 2$ ($r = \sqrt{k/2}$) and code rate=0.25; we derive a new DFR of 2^{-156} for $r = 2.83$ ($k = 16$) and code rate=0.25. The DFR margin enables us to improve the security by 9.4% while keeping the target DFR of 2^{-140} (2^{-128} as well) achievable.

C. *Roadmap*

This paper is organized as follows. A review of the necessary algebraic number theory, some basics of fading channels and polar codes can be found in Section II. In Section III we explain how to formulate a typical RLWE-based PKE scheme as an i.i.d. fading channel. How to handle the dependency in canonical basis is also demonstrated. Section IV gives a high-level description of RLWE-based PKE with the proposed polar coding scheme. Section V gives the explicit construction of polar codes for RLWE channel. Section VI analyzes the DFR theoretically and experimentally when polar decoding (SC decoding) is applied. Section VII discusses the security improvement derived by the new DFR margin. Section VIII concludes this paper.

[2] A concurrent work also used canonical embedding to analyse the statistical framework of RLWE-based cryptography [17], [18]; the novel contribution of our work is to derive an i.i.d. channel model under canonical embedding and to use the CSI for error-correction.

II. Preliminaries

A. Algebraic Number Theory

We review the necessary concepts from algebraic number theory required for our discussion of ring-LWE. In particular, we will relate many of our definitions to power-of-two cyclotomic fields, which are popular in modern cryptography.

A number field K is a finite degree field extension of the rationals $\mathbb{Q}$. Any number field K may be defined by adjoining some element $\alpha \in \mathbb{C}$ and setting $K = \mathbb{Q}[\alpha]$. Let α have minimal polynomial $f(X) \in \mathbb{Q}[X]$. Then, the degree of K over $\mathbb{Q}$ is precisely the degree n of $f(X)$, and there exists a canonical isomorphism $K \cong \dfrac{\mathbb{Q}[X]}{f(X)}$ defined by sending α to X. Because $f(\alpha) = 0$, K can be seen as a vector space over $\mathbb{Q}$ endowed with a basis $\{1, \alpha, ..., \alpha^{n-1}\}$ known as the power basis. In this paper we are interested in power-of-two cyclotomic fields, where the m^{th} cyclotomic field for 2-power m is defined by $K \cong \dfrac{\mathbb{Q}[X]}{x^n + 1}$, $n = \phi(m) = m/2$. Equivalently, we define $K = \mathbb{Q}[\zeta_m]$, where ζ_m is the m^{th} root of unity which has minimum polynomial $x^n + 1$.

A number field K of degree n permits n ring embeddings $\sigma_i : K \to \mathbb{C}, i = 1, ..., n$, which correspond to n distinct injective ring homomorphisms mapping α to the other roots of its minimum polynomial $f(X)$. These embeddings are split into the s_1 real embeddings, with images in $\mathbb{R}$, and s_2 complex conjugate pairs of complex embeddings, ordered such that $\sigma_{s_1+j} = \overline{\sigma_{s_1+s_2+j}}$. Hence we have $n = s_1 + 2s_2$. Now we define the canonical embedding $\sigma : K \to \mathbb{R}^{s_1} \times \mathbb{C}^{2s_2}$ as the map

$$\sigma(x) = (\sigma_1(x), ..., \sigma_n(x)).$$

More precisely, we say that this is a map from K into the space

$$H = \{(x_1, ..., x_n) \in \mathbb{R}^{s_1} \times \mathbb{C}^{2s_2} | x_{s_1+j} = \overline{x_{s_1+s_2+j}},\ \forall 1 \leq j \leq s_2\} \subset \mathbb{C}^n$$

and observe that H is isomorphic to $\mathbb{R}^n$ as a complex conjugate pair space, which can be seen by considering the orthonormal basis $\{\mathbf{h}_1, ..., \mathbf{h}_n\}$ of H, where $\mathbf{h}_i = \mathbf{e}_i, 1 \leq i \leq s_1$ and $\mathbf{h}_i = \frac{1}{\sqrt{2}}(\mathbf{e}_i + \mathbf{e}_{i+s_2}), \mathbf{h}_{i+s_2} = \frac{\sqrt{-1}}{\sqrt{2}}(\mathbf{e}_i - \mathbf{e}_{i+s_2}), s_1 < i \leq s_1 + s_2$, and $\mathbf{e}_i \in \mathbb{C}^n$ be a vector with 1 in its i-th coordinate and 0s in other positions. This isomorphism allows us to define ℓ_p norms on K by setting $|x|_p = \|\sigma(x)\|_p$ for $x \in K$. We also remark that under the embedding σ multiplication in K maps to coordinatewise multiplication in H.

B. Sample Ring-LWE Public Key Encryption Scheme

For concreteness, we give an example of a public key scheme based on ring-LWE. This scheme is well known in the literature, and was first described in [8]. Many ring-LWE schemes and protocols, including

NewHope and LAC, closely resemble this scheme. The scheme is paramaterized by an integer modulus q, dimension n, and error distribution χ over R_q. We will take the example of NewHope and view R_q as $\frac{\mathbb{Z}_q[X]}{x^n+1}$ and define sampling from χ to be sampling each coefficient of the polynomial s from the discrete Gaussian over $\mathbb{Z}$. The scheme proceeds as follows.

- Alice samples a secret key $s \leftarrow \chi$ and publishes as a public key an ring-LWE sample $(a, b) = (a, a \cdot s + e) \in R_q \times R_q$, where a is uniformly random and $e \leftarrow \chi$.
- Bob encrypts a message $m \in \{0, 1\}^n$ as $(c_1, c_2) = (a \cdot t + e_1, b \cdot t + e_2 + \lfloor \frac{q}{2} \rfloor \cdot m)$, where e_1, e_2, t are sampled independently from χ.
- Alice decrypts using s by computing $d := c_2 - c_1 \cdot s = \lfloor \frac{q}{2} \rfloor \cdot m + e \cdot t - s \cdot e_1 + e_2$.

Alice then recovers the message m by decoding: if the i^{th} coordinate of d is closer to 0 than $\lfloor q/2 \rfloor$, Alice assumes the i^{th} coordinate of m was 0, otherwise she assumes it was 1. We observe a few key facts about this scheme that we will need for our work. Firstly, although its formal security proof may be found in [8], the main idea is that b, c_1 and c_2 leak no information because they are ring-LWE samples, which are assumed to be pseudorandom by the hardness of the ring-LWE decision problem. However, one could alternate the encoding term $\lfloor \frac{q}{2} \rfloor \cdot m$ without affecting security, as long as the encoding is independent of the actualization of the variables s, e, e_1, e_2, t. We will use this fact implicitly while constructing polar codes in the sequel. Secondly, we observe that Alice knows the actualization of s and e, and so may use these for decoding. Finally, although we assumed m is a binary message of length n, typically one encodes a shorter message into n coordinates by introducing redundancy for error correction.

C. Fading Channel

In wireless communications, a fading channel arises due to a time-varying attenuation of signal quality caused by either the propagation environment or by movement of the transmitter/receiver. We consider a discrete-time fading channel model W

$$y_i = h_i x_i + z_i, \quad i = 1, \cdots, N,$$

where h_i is the channel gain, z_i is additive white Gaussian noise (AWGN) and N is the frame length. Denote by T_c the *coherence interval* of a fading channel W. In the context of a fading channel with memory, the channel gain h_i is believed to be a constant within one coherence interval and varies independently as the next coherence interval approaches. The realization of h_i is called *channel state information* (CSI) and the distribution of h_i is called *channel distribution information* (CDI). When designing a telecommunication system, we prefer i.i.d. fading channels where h_i are independent. There are a few methods to deal with the correlation. Let $m = T_c > 1$ and $N/m = n$. Since a fading channel with coherence interval T_c can be seen as m parallel sub-channels, *bit-interleaved coded modulation* (BICM) technique can be used to deal with the correlation between sub-channels [19], [20]. Another solution is to use multilevel codes [21], [22], [23] to design a coded modulation scheme with signal points in an $m-$dimensional signal space. In [24], a properly chosen lattice partition chain $\Lambda_1/\cdots/\Lambda_{l-1}/\Lambda_l$ is employed to design multi-level polar codes to achieve fading channel capacity. In this case, the dimension m of Λ_1 is properly chosen such that the channel gain h_i is assumed to be a constant amid the whole transmission of m symbols, i.e. $T_c = m$. A component code $\mathcal{C}_i$ at the ith level of the partition chain is designed in order to achieve the capacity of a Λ_i/Λ_{i+1} fading channel. The component codes are combined by *construction D* giving rise to a lattice which is good for the fading channel. For more information of the multi-level construction and the Λ_i/Λ_{i+1} channel, see [24] and [23]. We give an example of a mod $\mathbb{Z}$ channel and a $\mathbb{Z}/2\mathbb{Z}$ channel as follows and the fading version will be given in Section III.

Example 1: A mod $\mathbb{Z}$ channel is an *additive white Gaussian noise* (AWGN) channel with input restricted to $a \in \mathcal{V}(\mathbb{Z})$ where $\mathcal{V}(\mathbb{Z})$ is the *fundamental region* [3] of $\mathbb{Z}$. At the receiver's end, there is a mod $\mathcal{V}(\mathbb{Z})$ operation giving the equivalent channel output as

$$y = a + n \text{ mod } \mathbb{Z} = (a + n') \text{ mod } \mathbb{Z},$$

where n is the AWGN noise and $n' = n$ mod $\mathbb{Z}$.

Example 2: A $\mathbb{Z}/2\mathbb{Z}$ channel is an AWGN channel with input restricted to $r \in (\mathbb{Z} + a) \cap \mathcal{V}(2\mathbb{Z})$ for some offset $a \in \mathbb{R}$. At the receiver's end, the equivalent channel output is

$$y = r + n \text{ mod } 2\mathbb{Z} = r + n' \text{ mod } 2\mathbb{Z}.$$

It can be viewed as a mod $2\mathbb{Z}$ channel with input restricted to a set of elements of $\mathbb{Z} + a$ that fall in $\mathcal{V}(2\mathbb{Z})$.

In the special case of $T_c = 1$, channel W is referred to as an identically independently distributed (i.i.d.) fading channel. The design and performance of error-correcting codes for i.i.d. fading channels with/without CSI is well studied [25], [26], [27], [28], [29]. In [24], Liu etc., proposed a polar coding scheme for i.i.d. fading channels to achieve the Ergodic capacity. Unlike previous work of [27] in which CSI is given to both ends of communication, in Liu's scheme CSI is only known to the receiver which is more ordinary and frequent in practice.

[3] A *fundamental region* of a lattice Λ is a region that includes one and only one point from each coset of Λ in $\mathbb{R}^n$. Algebraically, $\mathcal{V}(\Lambda)$ is a set of coset representatives for all the cosets of Λ in $\mathbb{R}^n$, e.g., we can define $\mathcal{V}(\mathbb{Z})$ to be $[0, 1)$ but not necessarily the *fundamental Voronoi cell* $[-0.5, 0.5)$.

D. Polar Codes for BDMS Channels

Polar codes, introduced by Arıkan in [30], are linear block codes of length $N = 2^n$ for a positive integer n that achieves the capacity of any binary discrete memoryless symmetric (BDMS) channels asymptotically[4]. We firstly recall some basics of the polar coding for a BDMS channel. Given a BDMS channel W, there are two commonly used metrics in information theory to measure the quality of W: the mutual information[5] and the reliability.

Definition 1 (Mutual information of BDMS channels): The mutual information $I(W) \in [0, 1]$ of a BDMS channel $W : \mathcal{X} \to \mathcal{Y}$ is the maximum rate at which information can be successfully transmitted from the transmitter to the receiver.

$$I(W) \triangleq \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} \frac{1}{2} W(y|x) \log \frac{W(y|x)}{\frac{1}{2}W(y|0) + \frac{1}{2}W(y|1)},$$

In here, we use the definition of symmetric mutual information assuming uniform channel input which is also the capacity of the BDMS channel. We use the notations $I(W)$ and $I(Y; X)$ interchangeably to denote the mutual information of W.

Definition 2 (Bhattacharyya parameter of BDMS channels): The Bhattacharyya parameter $Z(W) \in [0, 1]$ is a measure of channel reliability for a BDMS channel W.

$$Z(W) \triangleq \sum_{y \in \mathcal{Y}} \sqrt{W(y|0)W(y|1)},$$

A small $Z(W)$ indicates a more reliable channel while a large $Z(W)$ implies a channel with more inference.

The capacity-achieving nature of polar codes arises from the so-called channel polarization phenomenon as a result of recursive applications of Arıkan's transform to identical Ws and their synthesized derivatives. The overall recursive transform can be done in a channel combining phase and a channel splitting phase. In the channel combining phase, a linear transformation defined as $X^{1:N} = U^{1:N} G_N$ is performed on a vector $U^{1:N} \in \mathcal{X}^{1:N}$ over $GF(2)$, where $G_N = B_N \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}^{\otimes n}$. B_N is a permutation matrix: if $U'^{1:N} = U^{1:N} B_N$ and $n = \log_2 N$, the $i' = ((b_n, \cdots, b_2, b_1)_2 + 1)$-th coordinate of $U'^{1:N}$ is the $i = ((b_1, b_2, \cdots, b_n)_2 + 1)$-th coordinate of $U^{1:N}$. By taking $X^{1:N}$ as the raw input of W, one derives a combined channel $W_N : \mathcal{X}^{1:N} \to \mathcal{Y}^{1:N}$ with a transition probability of

$$W_N(y^{1:N}|u^{1:N}) = \prod_{i \in \{1, \cdots, N\}} W(y^{(i)}|x^{(i)} = (u^{1:N} G_N)_i), \tag{1}$$

[4] In fact, the generalizations of polar codes are extended to arbitrary block length and a large class of channels.

[5] The maximum mutual information over all possible channel input distributions is the channel capacity.

where $(\cdot)_i$ denotes i-the coordinate. Since G_N induces a one-to-one mapping between $U^{1:N}$ and $X^{1:N}$, the mutual information of W_N is

$$I(W_N) = I((Y^{1:N}; U^{1:N})) = NI(W). \tag{2}$$

In the channel splitting phase, W_N is further split back into N synthesized channels $W_N^{(i)} : \mathcal{X} \to \mathcal{Y}^N \times \mathcal{X}^{i-1}$ whose transition probability is defined by

$$W_N^{(i)}(y^{1:N}, u^{1:i-1}|u^{(i)}) = \sum_{U^{i+1:N} \in \mathcal{X}^{N-i}} \frac{1}{2^{N-1}} W_N(Y^{1:N}|U^{1:N}). \tag{3}$$

It is proved in [30] that Arıkan's transform preserves the mutual information in the sense that

$$I(W_N) = NI(W) = \sum_{i \in \{1,\cdots,N\}} I(W_N^{(i)}).$$

More importantly, the quality of the synthesized channels polarizes asymptotically as the recursion proceeds.

Theorem 1 (Channel polarization of mutual information[30]): For any BDMS channel W, the synthesized channels $W_N^{(i)}$ polarize in the sense that, for any fixed $\delta \in (0, 1)$, as N goes to infinity through powers of two, the fraction of indices $i \in \{1, \cdots, N\}$ for which $I(W_N^{(i)}) \in (1 - \delta, 1]$ goes to $I(W)$ and the fraction for which $I(W_N^{(i)}) \in [0, \delta)$ goes to $1 - I(W)$.

The channel polarization theorem from above can also be stated in the metric of Bhattacharyya parameter by replacing $I(W_N^{(i)})$ by $Z(W_N^{(i)})$.

For any desired transmission rate $R < I(W)$, we can partition $\{1, \cdots, N\}$ into a subset $\mathcal{A}$ and its complement $\mathcal{A}^{\mathcal{C}}$ such that (i) $|\mathcal{A}| = \lfloor NR \rfloor$ and (ii) for any $i \in \mathcal{A}$ and $j \in \mathcal{A}^{\mathcal{C}}$, $Z(W_N^{(i)}) \leq Z(W_N^{(j)})$. Given the "best" $\lfloor NR \rfloor$ channels indexed by $\mathcal{A}$, one can construct polar codes following the encoding rule:

$$X^{1:N} = U_{\mathcal{A}} G_N(\mathcal{A}) \oplus U_{\mathcal{A}^c} G_N(\mathcal{A}^{\mathcal{C}}),$$

where $U_{\mathcal{A}}$ is the binary message vector of length $\lfloor NR \rfloor$ and $U_{\mathcal{A}^c}$ is some pre-determined information known to both encoder and decoder , e.g., $U_{\mathcal{A}^c} = \mathbf{0}$. In this manner, the useful information is transmitted via the most reliable synthesized channels. A question may arise on how to efficiently calculate $Z(W_N^{(i)})$. A brief review can be found in Section II-E and Section V-C but a detailed description of these methods are beyond the scope of this work.

The successive cancellation (SC) decoder is the initial decoding algorithm for polar codes. It gives an estimation of $u^{(i)}$, the i-th coordinate of $U^{1:N}$, in the natural order of i. Given a polar code initialized by code length N, information set $\mathcal{A}$, and frozen bits $U_{\mathcal{A}^c}$, one can derive the recovered message $\bar{u}^{(i)}$ of $u^{(i)}$ in sequential order of index i according to the decoding rule specified as

$$\bar{u}^{(i)} = \begin{cases} u^{(i)} & i \in \mathcal{A}^{\mathcal{C}}, \\ 0 & L_N^{(i)}(y^{1:N}, \bar{u}^{1:i-1}) \geq 1 \text{ and } i \in \mathcal{A}, \\ 1 & \text{otherwise,} \end{cases}$$

where $\bar{u}^{1:i-1}$ is the estimation of $u^{1:i-1}$ recovered before $\bar{u}^{(i)}$ and $L_N^{(i)}(y^{1:N}, \bar{u}^{1:i-1})$ is the likelihood ratio function defined as

$$L_N^{(i)}(y^{1:N}, \bar{u}^{1:i-1}) = \frac{W_N^{(i)}(y^{1:N}, \bar{u}^{1:i-1}|u^{(i)} = 0)}{W_N^{(i)}(y^{1:N}, \bar{u}^{1:i-1}|u^{(i)} = 1)}.$$

Denote by P_e the averaged probability of block decoding error. As a result of polar encoding and SC decoding, it is proved in [30] that P_e is upper bounded as follows.

Theorem 2 (Decoding Performance [30]): For any BDMS channel W and any choices of parameter $(N, R, \mathcal{A})$,

$$P_e \leq \sum_{i \in \mathcal{A}} Z(W_N^{(i)}).$$

E. Channel Degradation and Upgradation

The construction of polar codes can be addressed if all the Bhattacharyya parameters of synthesized channels can be efficiently calculated. In [30], an efficient solution to compute $Z(W_N^{(i)})$ for binary erasure channels (BEC) is given while it is suggested to use Monte-Carlo method to deal with more general BDMS channels. R. Mori and T. Tanaka made an attempt to solve this problem for arbitrary symmetric binary-input memoryless (BMS) channels using density evolution [31], [32], [33] of belief propagation (BP) decoding. However, they also mentioned that it was unclear how to handle the computational efficiency when the block length N was large and the requirement for precision was high. In [34], a quantization method was proposed to construct a degraded or upgraded approximation of a general BMS channel. If the degraded or upgraded relation exists, one can approximate $Z(W_N^{(i)})$ efficiently.

Definition 3 (Degraded and Upgraded Channel, [34]): A channel $\mathcal{Q} : \mathcal{X} \to \mathcal{Z}$ is (stochastically) degraded with respect to a channel $\mathcal{W} : \mathcal{X} \to \mathcal{Y}$ if there exists a channel $\mathcal{P} : \mathcal{Y} \to \mathcal{Z}$ such that

$$\mathcal{Q}(z|x) = \sum_{y \in \mathcal{Y}} \mathcal{W}(y|x)\mathcal{P}(z|y)$$

for all $z \in \mathcal{Z}$ and $x \in \mathcal{X}$. We denote by $\mathcal{Q} \preceq \mathcal{W}$ the relation that $\mathcal{Q}$ is degraded with respect to $\mathcal{W}$. Conversely, we denote by $\mathcal{Q}' \succeq \mathcal{W}$ the relation that $\mathcal{Q}'$ is upgraded with respect to $\mathcal{W}$ if there exists a channel $\mathcal{Q}' : \mathcal{X} \to \mathcal{Z}'$ and a channel $\mathcal{P} : \mathcal{Z}' \to \mathcal{Y}$ such that

$$\mathcal{W}(y|x) = \sum_{z' \in \mathcal{Z}'} \mathcal{Q}'(z'|x)\mathcal{P}(y|z')$$

for $y \in \mathcal{Y}$ and $x \in \mathcal{X}$.

Moreover, the synthesized channels of $\mathcal{Q}, \mathcal{W}, \mathcal{Q}'$ under Arıkan's transform also fulfill the channel degradation and upgradation relation.

Lemma 1 (restatement of Lemma 4.7 in [35]): Given BMS channels $\mathcal{W}$, $\mathcal{Q}$, and $\mathcal{Q}'$, we denote by $\mathcal{W}_N^{(i)}$, $\mathcal{Q}_N^{(i)}$ and $\mathcal{Q}'^{(i)}_N$ for $i \in [1, N]$ the synthesized channels obtained by Arıkan transformation. If $\mathcal{Q}' \succeq \mathcal{W} \succeq \mathcal{Q}$ for all i, then $\mathcal{Q}'^{(i)}_N \succeq \mathcal{W}_N^{(i)} \succeq \mathcal{Q}_N^{(i)}$.

If the channel degradation or upgradation relation is set up, their channel capacity, reliability and error probability will be related as follows.

Lemma 2 ([34]): Let $\mathcal{W}$ be a BMS channel and suppose there exists the other channel $\mathcal{Q}$ such that $\mathcal{Q} \preceq \mathcal{W}$. Then

$$C(\mathrm{Q}) \leq C(\mathcal{W}),$$

$$Z(\mathrm{Q}) \geq Z(\mathcal{W}),$$

$$P_e(\mathrm{Q}) \geq P_e(\mathcal{W}).$$

The inequality will reverse if we replace "degraded" by "upgraded".

III. RLWE Channel Model

A. RLWE Channel Model in Canonical Basis

Definition 4: The real multivariate normal distribution has density function

$$f(x; \mu, \Sigma) = \frac{e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}}{\sqrt{|2\pi\Sigma|}} \quad x \in \mathbb{R}^n$$

where $|\cdot|$ denotes the determinant, $\mu = \mathbb{E}[X] \in \mathbb{R}^n$, $\Sigma = \mathbb{E}\left[(X-\mu)(X-\mu)^T\right]$; we write $X \sim \mathcal{N}(\mu, \Sigma)$. One natural generalization would be to consider $Z \sim \mathcal{NC}(\mu, \Gamma)$, the complex multivariate normal, with density function

$$f(z; \mu, \Gamma) = \frac{e^{-(z-\mu)^* \Gamma^{-1}(z-\mu)}}{|\pi\Gamma|} \quad z \in \mathbb{C}^n,$$

where z^* denotes the Hermitian transpose of the vector z, μ is the (possibly complex) mean and $\Gamma = \mathbb{E}[(Z-\mu)(Z-\mu)^*]$.

Let ζ be the $2n^{\text{th}}$ root of unity for a 2-power n. The canonical embedding $\sigma : K \to \mathbb{C}^n$ maps the $2n^{th}$ cyclotomic number field K to the space H which is endowed with an orthogonal basis defined by the columns of the matrix

$$B = \begin{pmatrix} 1 & \zeta & \zeta^2 & \cdots & \zeta^{n-1} \\ 1 & \zeta^3 & \zeta^6 & \cdots & \zeta^{3(n-1)} \\ 1 & \zeta^5 & \zeta^{10} & \cdots & \zeta^{5(n-1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \zeta^{(2n-1)} & \zeta^{(2n-1)2} & \cdots & \zeta^{(2n-1)(n-1))} \end{pmatrix}_{n\times n} .$$

Recall the definition of canonical embedding in Section II-A, i.e.,

$$\sigma(x) = (\sigma_1(x), ..., \sigma_n(x)),$$

where the conjugate pairs are $\sigma_i(x)$ and $\sigma_{i+n/2}(x)$. We observe that we have re-ordered the embedding in the basis of B where $\sigma_i(x)$ is the complex conjugate of $\sigma_{n-i+1}(x)$ for $1 \leq i \leq n/2$. This can be interpreted as a permutation of coordinates and it does not change the homomorphism relationship. Moreover, we can also represent $\mathbb{C}^n$ with $\mathbb{R}^n$ by mapping the first $n/2$ embeddings to their real and complex parts, i.e., we rewrite the canonical embedding as $\sigma' : K \to \mathbb{R}^n$

$$\sigma'(x) = (\Re[\sigma_1(x)], \Im[\sigma_1(x)], ..., \Re[\sigma_{n/2}(x)], \Im[\sigma_{n/2}(x)]).$$

The corresponding basis $\tilde{B}$ of $\sigma'(x)$ is

$$\tilde{B} = \begin{pmatrix} 1 & \Re[\zeta] & \Re[\zeta^2] & \cdots & \Re[\zeta^{n-1}] \\ 0 & \Im[\zeta] & \Im[\zeta^2] & \cdots & \Im[\zeta^{n-1}] \\ 1 & \Re[\zeta^3] & \Re[\zeta^6] & \cdots & \Re[\zeta^{3(n-1)}] \\ 0 & \Im[\zeta^3] & \Im[\zeta^6] & \cdots & \Im[\zeta^{3(n-1)}] \\ 1 & \Re[\zeta^5] & \Re[\zeta^{10}] & \cdots & \Re[\zeta^{5(n-1)}] \\ 0 & \Im[\zeta^5] & \Im[\zeta^{10}] & \cdots & \Im[\zeta^{5(n-1)}] \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \Re[\zeta^{(n-1)}] & \Re[\zeta^{(n-1)2}] & \cdots & \Re[\zeta^{(n-1)(n-1)}] \\ 0 & \Im[\zeta^{(n-1)}] & \Im[\zeta^{(n-1)2}] & \cdots & \Im[\zeta^{(n-1)(n-1)}] \end{pmatrix}_{n\times n} .$$

Note that both B and $\tilde{B}$ are orthogonal matrices. The determinant of B is $\sqrt{n}^n$ while that of $\tilde{B}$ is $(\sqrt{n/2})^n$.

We already given an RLWE-based PKE instance in Section II-B. Now we consider the problem of decoding the message m from the polynomial

$$y = \lfloor \frac{q}{2} \rfloor \cdot m + e \cdot t - s \cdot e_1 + e_2 \mod R_q, \tag{4}$$

where $e \cdot t$ and $s \cdot e_1$ are convolution of polynomials in $\mathbb{Z}[x]/(1 + x^n)$. To visualize the decoding step of RLWE-based PKE as a channel decoding problem, we rewrite formula (4) in vector form as

$$\mathbf{y} = \lfloor \frac{q}{2} \rfloor \cdot \mathbf{m} + \mathbf{E} \cdot \mathbf{t} - \mathbf{S} \cdot \mathbf{e}_1 + \mathbf{e}_2 \mod R_q, \quad \mathbf{e}, \mathbf{t}, \mathbf{e}_1, \mathbf{e}_2 \leftarrow \mathcal{N}(0, r^2\mathbb{I}), \tag{5}$$

where $\mathbf{m}, \mathbf{t}, \mathbf{e}_2, \mathbf{e}_2$ are vectors, $\mathbf{E}$ is a negacyclic matrix with the first column to be the coefficients of the polynomial e and $\mathbf{S}$ is also a negacyclic matrix defined in the same manner. Note that the coefficients of e, t, s, e_1, e_2 should be drawn from discrete Gaussian $D_{\mathbb{Z},r}$. We use continuous normal distribution $\mathcal{N}(0, r^2)$ instead to simplify the distribution analysis of the noise term. Formula (5) can be viewed as a channel model where $\mathbf{y}$ is the channel output vector, $\mathbf{m}$ is the codewords to be modulated as $\lfloor q/2 \rfloor \cdot \mathbf{m}$ and $\mathbf{E} \cdot \mathbf{t} - \mathbf{S} \cdot \mathbf{e}_1 + \mathbf{e}_2$ is the channel noise. In this work, we refer to formula (5) as a *RLWE channel.*

Theorem 3 (Diagonalization of negacyclic matrix,[36]): Let $N(x)$ be an $n \times n$ negacyclic matrix whose first column is x. Then

$$N(x) = G^* \text{diag}(\lambda_1(x), \lambda_2(x), \cdots, \lambda_n(x)) G,$$

where the element $G_{p,q} = \frac{1}{\sqrt{n}} w^{(2p-1)(q-1)}$, w is the $2n$th root of unity, G^* is the conjugate transpose of G and $\lambda_j(x) = \sum_{k=1}^{n} x_k w^{(2j-1)(k-1)}$.

According to Theorem 3 this RLWE channel can be formulated as

$$B\mathbf{y} = B\lfloor \frac{q}{2} \rfloor \mathbf{m} + BB^{-1}\text{diag}(\mathbf{e})B\mathbf{t} - BB^{-1}\text{diag}(\mathbf{s})B\mathbf{e}_1 + B\mathbf{e}_2, \mod BR_q \tag{6}$$

which is equivalent to the polynomial description under canonical embedding $\sigma : K \to \mathbb{C}^n$ as

$$\sigma(y) = \sigma(\lfloor \frac{q}{2} \rfloor m) + \sigma(e)\sigma(t) - \sigma(s)\sigma(e_1) + \sigma(e_2) \qquad \mod BR_q. \tag{7}$$

If we chop off the bottom half of the vectors in equation (6) (i.e., the conjugates) and separate the real part from the imaginary part, we derive the RLWE channel as

$$\underbrace{\begin{pmatrix} \Re[B_1\mathbf{y}] \\ \Im[B_1\mathbf{y}] \\ \Re[B_2\mathbf{y}] \\ \Im[B_2\mathbf{y}] \\ \vdots \\ \Re[B_{\frac{n}{2}}\mathbf{y}] \\ \Im[B_{\frac{n}{2}}\mathbf{y}] \end{pmatrix}}_{Y} = \underbrace{\begin{pmatrix} \Re[B_1\lfloor\frac{q}{2}\rfloor\mathbf{m}] \\ \Im[B_1\lfloor\frac{q}{2}\rfloor\mathbf{m}] \\ \Re[B_2\lfloor\frac{q}{2}\rfloor\mathbf{m}] \\ \Im[B_2\lfloor\frac{q}{2}\rfloor\mathbf{m}] \\ \vdots \\ \Re[B_{\frac{n}{2}}\lfloor\frac{q}{2}\rfloor\mathbf{m}] \\ \Im[B_{\frac{n}{2}}\lfloor\frac{q}{2}\rfloor\mathbf{m}] \end{pmatrix}}_{X} + \underbrace{\begin{pmatrix} \Re[\sigma_1(e)\sigma_1(t) - \sigma_1(s)\sigma_1(e_1) + \sigma_1(e_2)] \\ \Im[\sigma_1(e)\sigma_1(t) - \sigma_1(s)\sigma_1(e_1) + \sigma_1(e_2)] \\ \Re[\sigma_2(e)\sigma_2(t) - \sigma_2(s)\sigma_2(e_1) + \sigma_2(e_2)] \\ \Im[\sigma_2(e)\sigma_2(t) - \sigma_2(s)\sigma_2(e_1) + \sigma_2(e_2)] \\ \vdots \\ \Re[\sigma_{\frac{n}{2}}(e)\sigma_{\frac{n}{2}}(t) - \sigma_{\frac{n}{2}}(s)\sigma_{\frac{n}{2}}(e_1) + \sigma_{\frac{n}{2}}(e_2)] \\ \Im[\sigma_{\frac{n}{2}}(e)\sigma_{\frac{n}{2}}(t) - \sigma_{\frac{n}{2}}(s)\sigma_{\frac{n}{2}}(e_1) + \sigma_{\frac{n}{2}}(e_2)] \end{pmatrix}}_{N} \mathrm{mod}\tilde{B}R_q, \tag{8}$$

where B_j represents the j^{th} row of B, $Y = \tilde{B}\mathbf{y} = \sigma'(y)$, and $X = \tilde{B}\lfloor\frac{q}{2}\rfloor\mathbf{m} = \sigma'(\lfloor\frac{q}{2}\rfloor m)$. To see the how the noise term N is distributed, we rewrite formula (8) for all the odd indices $i = 1, 3, 5, \cdots, n/2-1$ as

$$\begin{aligned} \begin{bmatrix} \sigma'_i(y) \\ \sigma'_{i+1}(y) \end{bmatrix} &= \begin{bmatrix} \sigma'_i(\lfloor\frac{q}{2}\rfloor m) \\ \sigma'_{i+1}(\lfloor\frac{q}{4}\rfloor m) \end{bmatrix} + \begin{bmatrix} \sigma'_i(e) & -\sigma'_{i+1}(e) \\ \sigma'_{i+1}(e) & \sigma'_i(e) \end{bmatrix} \begin{bmatrix} \sigma'_i(t) \\ \sigma'_{i+1}(t) \end{bmatrix} \\ &\quad - \begin{bmatrix} \sigma'_i(s) & -\sigma'_{i+1}(s) \\ \sigma'_{i+1}(s) & \sigma'_i(s) \end{bmatrix} \begin{bmatrix} \sigma'_i(e_1) \\ \sigma'_{i+1}(e_1) \end{bmatrix} + \begin{bmatrix} \sigma'_i(e_2) \\ \sigma'_{i+1}(e_2) \end{bmatrix}, \end{aligned} \tag{9}$$

where $\tilde{B}_i(\cdot) = \sigma'_i(\cdot)$ and $\tilde{B}_{i+1}(\cdot) = \sigma'_{i+1}(\cdot)$. Under embedding $\sigma : K \to \mathbb{C}^n$, the spherical normal distributed vectors, **e** and **t**, are mapped to complex spherical normal vectors, $\sigma(e), \sigma(t) \sim \mathcal{NC}(0, nr^2\mathbb{I})$. As for the embedding $\sigma' : K \to \mathbb{R}^n$, the spherical normal $\mathcal{N}(0, r^2\mathbb{I})$ is transformed to a new spherical normal with distribution $\mathcal{N}(0, nr^2/2\mathbb{I})$. Since $\mathbf{e}, \mathbf{t}$ are coordinate-wise i.i.d. their embedding $\sigma(e)$, $\sigma(t)$, $\sigma'(e)$, $\sigma'(t)$ are coordinate-wise independent as well. We observe from formula (9) that every odd-indexed coordinate and the next even-indexed coordinate are somehow correlated because they share the same $\sigma'_i(e), \sigma'_{i+1}(e)$, $\sigma'_i(t), \sigma'_{i+1}(t)$, $\sigma'_i(s), \sigma'_{i+1}(s)$ and $\sigma'_i(e_1), \sigma'_{i+1}(e_1)$ although $\sigma'_i(e_2), \sigma'_{i+1}(e_2)$ are independent.

To further refine the RLWE channel model, we can rewrite formula (8) and (9) as

$$\tilde{B}\mathbf{y} = \tilde{B}\lfloor\frac{q}{2}\rfloor\mathbf{m} + \mathbf{N}, \quad \bmod \tilde{B}R_q \tag{10}$$

where for $i = 1, 2, \cdots, n$, $N_i = H_i * Z_i$, $Z_i \leftarrow \mathcal{N}(0, \frac{nr^2}{2})$, and

$$H_i = \sqrt{\sigma^*_{\lceil i/2\rceil}(e)\sigma_{\lceil i/2\rceil}(e) + \sigma^*_{\lceil i/2\rceil}(s)\sigma_{\lceil i/2\rceil}(s) + 1}.$$

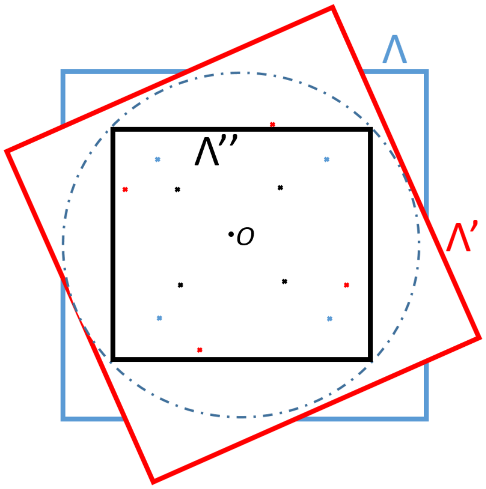


Fig. 1. Switch of constellation diagram.

Because of the correlation between every two coordinates, H_i and H_j are independent for two different indices i, j as long as $\lceil i/2 \rceil \neq \lceil j/2 \rceil$; otherwise $H_i = H_j$. Similarly, Z_i and Z_j are correlated if $\lceil i/2 \rceil = \lceil j/2 \rceil$; otherwise they are independent.

Unlike in NewHope and other RLWE-based encryption schemes where the plain text is encoded and decoded in the polynomial basis, we will carry out encoding and decoding in canonical basis. Observe that the channel given by formula (10) is almost a *fading channel* with coherence interval $T_c = 2$ coordinates except that the symbols to be transmitted after modulation, i.e., $\tilde{B}\lfloor \frac{q}{2} \rfloor \mathbf{m}$, are not coordinate-wise independent. In next subsection, we will adjust the modulation scheme such that a tailored constellation diagram can fit in with the *fading channel*.

B. A Tailored Constellation Diagram

The RLWE channel in formula (5) can be interpreted as n parallel $\mathbb{Z}/2\mathbb{Z}$ channels where the message $\mathbf{m} \in \{0,1\}^n$ is mapped to symbols on the *constellation diagram* $\{0, \lfloor \frac{q}{2} \rfloor\}^n$. The mod R_q operation defines a valid *constellation space* as an $n-$dimensional cube Λ with vertices $\{0, q\}^n$. To ease the description of how we design a new constellation diagram in canonical basis, we make a modification to the modulation scheme in formula (5): the message $\mathbf{m} \in \{-1, 1\}^n$ is mapped onto the constellation diagram $\{\pm \lfloor \frac{q}{4} \rfloor\}^n$ and the valid constellation space is a cube Λ with vertices $\{\pm \lfloor \frac{q}{2} \rfloor\}^n$. This modification will preserve the capacity of the $\mathbb{Z}/2\mathbb{Z}$ channel because they are statistically equivalent if we ignore geometrical approximation caused by the round-off operation $\lfloor \cdot \rceil$.

According to formula (10), after applying the canonical embedding, the constellation diagram turns into $\tilde{B}\{\pm\lfloor\frac{q}{4}\rfloor\}^n$. Similarly, we can obtain the new constellation space $\Lambda' = \tilde{B}\Lambda$ by rotating Λ and scaling it up by a factor of $\sqrt{n/2}$.

As discussed in last subsection, the coherence interval T_c of the residue noise equals to 2 coordinates while the constellation symbol $\tilde{B}\lfloor\frac{q}{4}\rfloor\mathbf{m}$ has memory throughout n coordinates. In a communication system, the interleaving technique can be used to alleviate the correlation of the source by permuting symbols of different code blocks. Unfortunately, interleaving is impractical in the RLWE channel because there is only one code block of length n. At the cost of distance between the constellation symbols, we tailor the constellation space Λ' to fit in with the *fading channel* with correlation T_c.

Essentially, we are looking for a new modulation scheme meeting two conditions: (a) we desire the symbols after modulation (or the modulated message) to be coordinate-wise i.i.d.; in other words, we expect a valid constellation diagram inside the space Λ' such that for coordinate-wise i.i.d. message $\mathbf{m}$, the modulated message is coordinate-wise i.i.d. as well; (b) the new modulation scheme gives us a $\mathbb{Z}/2\mathbb{Z}$ channel. Conceptually, the maximal n-dimensional cube Λ'' enclosed in Λ' and parallel to Λ is our target constellation space. In this case, the symbols to be transmitted can be easily made to be binary and i.i.d. if we divide the cube Λ'' equally into 2^n small cubes and select all the centers of the small cubes to be the constellation diagram. However, looking for such a Λ'' in practice is not tractable when the dimension n is large and we are unclear about in what direction and by what degree the cube Λ' is rotated with respect to Λ. Instead, we compromise and use the cube Λ'' which is parallel to Λ and is enclosed in the maximal ball inscribed in Λ'. In this manner, we can make sure there always exists such a constellation space Λ'' and it is straightforward to calculate its size. Figure 1 illustrates this idea in the 2-dimensional case. If the side length of Λ is q, the side of Λ' turns out to have length $q\sqrt{n/2}$, and the side of Λ'' will be $q/\sqrt{2}$. Observe that $\Lambda' = \sqrt{2}\tilde{B}\Lambda''$.

C. Tailored RLWE Channel Model in Canonical Basis

Given the tailored constellation space Λ'' and its corresponding constellation diagram, we now have a tailored RLWE channel model in the canonical basis:

$$\mathbf{y} = \lfloor\frac{q}{2}\rfloor\frac{1}{\sqrt{2}}\mathbf{m} + \mathbf{N}, \quad \bmod \Lambda'', \tag{11}$$

where $\mathbf{m} \in \{0,1\}^n$, $N_i = H_i * Z_i$ and $Z_i \leftarrow \mathcal{N}(0, nr^2/2)$ for $1 \le i \le n$. As discussed in formula (10), H_i and H_j are independent for two different indices i, j as long as $\lceil i/2\rceil \neq \lceil j/2\rceil$; otherwise $H_i = H_j$.

$$H_i = \sqrt{\sigma^*_{\lceil i/2\rceil}(e)\sigma_{\lceil i/2\rceil}(e) + \sigma^*_{\lceil i/2\rceil}(s)\sigma_{\lceil i/2\rceil}(s) + 1},$$

where $\sigma_{\lceil i/2\rceil}(e), \sigma_{\lceil i/2\rceil}(s) \leftarrow \mathcal{NC}(0, nr^2)$. Similarly, Z_i and Z_j are independent if $\lceil i/2\rceil \neq \lceil j/2\rceil$ otherwise they are correlated.

We observe that the tailored channel model in formula (11) can be seen as a fading channel where H_i is the channel gain and Z_i is the additive noise. A family of fading channels (e.g., i.i.d. fading, block fading, compound fading) are well studied in existing work of [25], [27], [26], [24], [37] and explicit constructions of error-correcting codes are given. In this work, since H_i and Z_i have the same coherence interval of two coordinates, our strategy is to divide the n parallel channels into two groups of i.i.d. channels and we construct two parallel polar codes of equal block length $n/2$ for the two $\mathbb{Z}/2\mathbb{Z}$ fading channels. Note that in this work we use parameters similar to NewHope, e.g., $q = 12289, n = 1024$, $r \in \{1, 2, 6, 9\}$ where the values of r correspond to the 'Short' and 'Tall' parameters in [38].

Denote by L and L' two one-dimensional lattices $\lfloor\frac{q}{2}\rfloor\frac{1}{\sqrt{2}}\mathbb{Z}$ and $q\frac{1}{\sqrt{2}}\mathbb{Z}$ respectively. The above channel model can also be written as a fading L/L' channel, i.e.,

$$Y_i = \lfloor\frac{q}{2}\rfloor\frac{1}{\sqrt{2}}m_i + H_i * Z_i, \quad \bmod q\frac{1}{\sqrt{2}}\mathbb{Z},\ i = 1, \cdots n, \tag{12}$$

where $m_i \in \{0, 1\}$ and the channel input X is restricted to the discrete alphabet $\mathcal{X} = L \cap \mathcal{R}(L') = \{0, \lfloor\frac{q}{2}\rfloor\frac{1}{\sqrt{2}}\}$. Since Alice knows exactly what e and s is, she knows both the distribution and realization of the channel gain H_i. At the transmitter's end, Bob only knows the distribution of H_i. Both of them know the distribution of Z_i. How to achieve the *ergodic capacity* of such an i.i.d. fading channel using polar codes is well studied in [24] and we are about to adapt their strategy to our tailored RLWE channel model. A diagram of a fading L/L' channel with CSI available to the decoder is shown in Figure 2. Denote by $W : X \to (\tilde{Y}, H)$ the fading L/L' channel with CSI available to the decoder. The transition

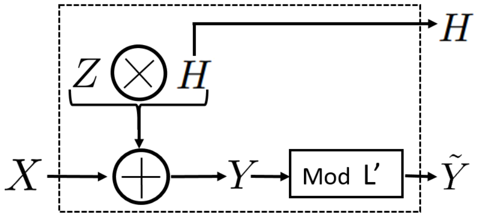


Fig. 2. A block diagram of a fading L/L' channel.

probability of W is

$$\begin{aligned} P_{\tilde{Y},H|X}(\tilde{y}, h|x) &= P_{Y,H|X}(y = \tilde{y} + L', h|x)\frac{d\tilde{y}}{dy} \\ &= P_H(h)P_{Y|H,X}(y = \tilde{y} + L'|h, x) \\ &= P_H(h)\sum_{\lambda\in L'}\frac{1}{\sqrt{2\pi}h\sigma}\exp\left\{-\frac{(\tilde{y} + \lambda - x)^2}{2\sigma^2h^2}\right\}, \end{aligned} \tag{13}$$

where $\sigma = \sqrt{\frac{n}{2}}r$. The distribution of H is

$$P_H(h) = \frac{1}{2\sigma^4}h(h^2-1)\exp\left\{-\frac{h^2-1}{2\sigma^2}\right\} = \frac{2h(h^2-1)}{n^2r^4}e^{-\frac{(h^2-1)}{nr^2}},\ h > 1.$$

The pdf of H in terms of various choices of parameter r is illustrated in Figure 3.

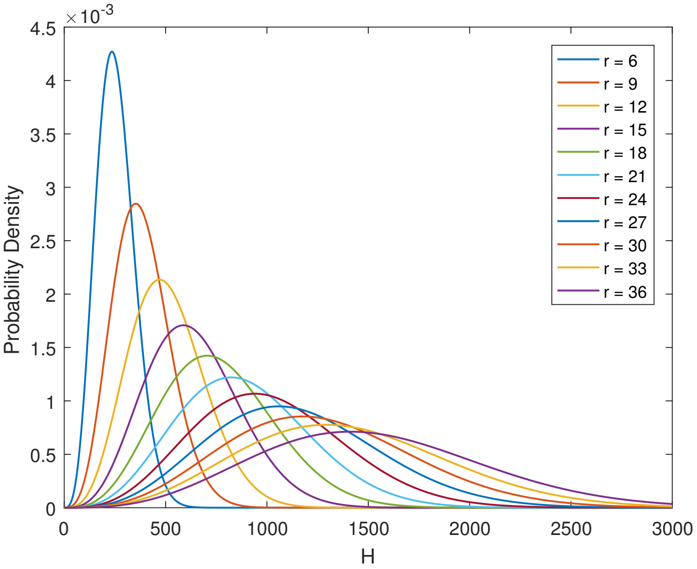


Fig. 3. Probability density function of fading coefficient H.

As discussed in [39] and [24], the capacity of the fading L/L' channel is given by

$$\begin{aligned} C(L/L',\sigma^2) &= E_H\left[C(L',(h\sigma)^2)\right] - E_H\left[C(L,(h\sigma)^2)\right] \\ &= E_H\left[\mathfrak{h}(L,(h\sigma)^2)\right] - E_H\left[\mathfrak{h}(L',(h\sigma)^2)\right] + \log|L/L'|, \end{aligned}$$

where $E_H[\cdot]$ denotes the expectation over the fading coefficient, $\mathfrak{h}(L,\sigma^2)$ and $\mathfrak{h}(L',\sigma^2)$ are differential entropies of mod-L and mod-L' channels respectively, and $|L/L'|$ is the order of the partition L/L'. Specifically, $\mathfrak{h}(L,\sigma^2)$ is given by

$$\mathfrak{h}(L,(h\sigma)^2) = -\int_{\mathcal{R}(L)} f_{L,(h\sigma)^2}(n')\log f_{L,(h\sigma)^2}(n')dn',$$

$$f_{L,(h\sigma)^2}(n') = \sum_{\lambda\in L} g_{(h\sigma)^2}(n'+\lambda),\ n'\in\mathcal{R}(L),$$

where $\mathcal{R}$ is a fundamental region of lattice L, $g_{(h\sigma)^2}(n)$ is the density function of $n \leftarrow \mathcal{N}(0,h^2\sigma^2)$. We refer to $f_{L,(h\sigma)^2}$ as L-aliased Gaussian density function or L-periodic Gaussian density function which is defined by summing up a set of copies of a Gaussian density function centered at the every lattice point of L. The value of an L-aliased Gaussian variable n' is restricted to any fundamental region of

L such that the integral of its density function over $\mathcal{R}(L)$ is obviously 1. See Figure 4 for the Ergodic capacity of the fading L/L' channel $W : X \to (\tilde{Y}, H)$ with respect to different choices of r. In a communication system, the *signal-to-noise ratio* (SNR) is a measure of how much useful information can be transmitted from information source to recipient. It is defined as the ratio of the signal strength over the noise strength[6].

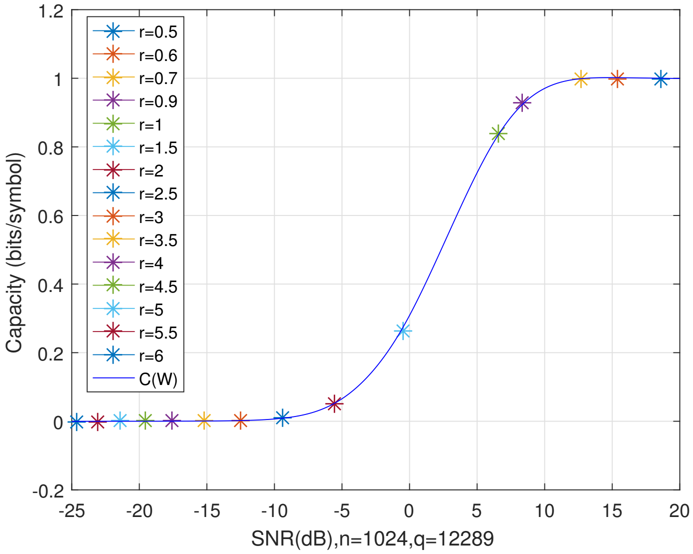


Fig. 4. Capacity of RLWE channel vs SNR given $X \in \{0, \lfloor \frac{q}{2} \rfloor \frac{1}{\sqrt{2}}\}$, $n = 1024$, $q = 12289$.

Observe that the fading L/L' channel W is a BMS channel in the sense that the equation $P_{\tilde{Y},H|X}(\tilde{y}, h|x = 0) = P_{\tilde{Y},H|X}(\pi(\tilde{y}, h)|x = \lfloor \frac{q}{2} \rfloor \frac{1}{\sqrt{2}})$ holds for a permutation $\pi(\tilde{y}, h) = ([\lfloor \frac{q}{2} \rfloor \frac{1}{\sqrt{2}} - \tilde{y}] \text{ mod-}q\frac{1}{\sqrt{2}}\mathbb{Z}, h)$ over the outputs $(\tilde{y}, h)$. Thus, we can achieve the capacity of the W with an equiprobable input distribution, i.e. X is uniformly random over $\mathcal{X}$.

IV. Description of the Encryption Scheme

Table I gives a high-level description of the RLWE-based PKE scheme using polar codes which are customized for our tailored RLWE channel model in canonical basis. The functions *PolarEnc*$(\cdot)$ and *PolarDec*$(\cdot)$ are encoding and decoding algorithms of polar codes which will be explicitly introduced in the sequel.

Remark 1: Unlike most RLWE encryption schemes (e.g., NewHope and LAC) where the error distribution χ is defined over R_q, we use the definition of χ when the Ideal Learning With Errors problem

[6] Different r induces different SNR. The calculation of SNR with respect to the fading L/L' channel is given in Appendix A

was initially proposed by D. Stehlé, R. Steinfeld et al. in [40] where χ is defined on $\mathbb{R}/[0,q)$. Moreover, in the formal definition of ring-LWE hard problems when it introduced in [8], [41], the error distribution is also continuous over the field tensor product $K \otimes_{\mathbb{Q}} \mathbb{R}$.

Remark 2: A plaintext $\mathbf{m}$ is uniquely mapped to a symbol $\lfloor \frac{q}{2} \rfloor \frac{1}{\sqrt{2}}\textit{PolarEnc}(\mathbf{m})$ on the constellation diagram in canonical basis and then it is switched to polynomial basis as vector $\mathbf{v}$. Note that $\mathbf{v} \in (\mathbb{R}/[0,q))^n$ but not in R_q. We see it reasonable since χ is also real and continuous.

Parameters are n, q; error distribution χ on $(\mathbb{R}/[0,q))^n$		
Alice (Server)		**Bob (Client)**
$\mathbf{a} \leftarrow R_q$		
$\mathbf{s}, \mathbf{e} \leftarrow \chi$		$\mathbf{t}, \mathbf{e}_1, \mathbf{e}_2 \leftarrow \chi$
$\mathbf{b} = \mathbf{a} \cdot \mathbf{s} + \mathbf{e} \mod R_q$	$\xrightarrow{(\mathbf{b},\mathbf{a})}$	
		$\mathbf{c}_1 = \mathbf{a} \cdot \mathbf{t} + \mathbf{e}_1 \mod R_q$
		$\mathbf{v} = \tilde{B}^{-1} \lfloor \frac{q}{2} \rfloor \frac{1}{\sqrt{2}}\textit{PolarEnc}(\mathbf{m})$
		$\mathbf{c}_2 = \mathbf{b} \cdot \mathbf{t} + \mathbf{e}_2 + \mathbf{v} \mod R_q$
$\mathbf{y} = \mathbf{c}_2 - \mathbf{c}_1 \cdot \mathbf{s} \mod R_q$	$\xleftarrow{(\mathbf{c}_1,\mathbf{c}_2)}$	
$\tilde{B}\mathbf{y} = \tilde{B}\mathbf{v} + \mathbf{N} \mod \tilde{B}R_q$		
$\mathbf{m} = \textit{PolarDec}(\tilde{B}\mathbf{y})$		

TABLE I

A RLWE PUBLIC KEY ENCRYPTION SCHEME WITH POLAR CODING AND DECODING.

One may notice in Table I that Alice finally derives a mod-$\tilde{B}R_q$ channel (or equivalently a mod-Λ' channel) as in Figure 1 rather than the mod-Λ'' in formula (11) (or equivalently the mod-L' channel in (12)). Questions arise whether the tailored RLWE channel model in formula (11) makes sense and how it will behave if we construct a polar code for the mod-Λ'' channel when we actually have a mod-Λ' channel. Lemma 3 illustrates the channel degradation relation between the above two different but related channels.

Lemma 3: (Channel Degradation Relation Between RLWE Channel and Its Tailored Variant) Let Λ' be the constellation space and let Λ'' be its tailored variant as in Figure 1. Given the tailored RLWE channel model as in formula (11) with CSI H_i known to the decoder as in Figure 2, the fading L^n/Λ'' channel is degraded with respect to the fading L^n/Λ' channel.

Proof 1: Denote by W' the fading L^n/Λ' channel $y' = x + h * z \mod \Lambda'$ where $y' \in \mathcal{R}(\Lambda')$, $x \in L^n \cap \mathcal{R}(\Lambda')$ is the channel input, h is the channel gain and z is the Gaussian noise. In the same fashion, we define the fading L^n/Λ'' channel W'' as $y'' = x + h * z \mod \Lambda''$ where $y'' \in \mathcal{R}(\Lambda'')$, $x \in L^n \cap \mathcal{R}(\Lambda'')$.

As formula (13) indicates, the L/L' fading channel with CSI known to the receiver in formula (12) can be viewed as an independent combination of channel gain H and a L/L' Gaussian channel. Therefore, with no loss of generality, we can view the channel gain h as a constant. We can rewrite channel W' as $W' : y' = x + z' \mod \Lambda'$ where $z' \sim \mathcal{N}(0, h^2\sigma^2)$ and rewrite W'' as $W'' : y'' = x + z' \mod \Lambda''$ where $z' \sim \mathcal{N}(0, h^2\sigma^2)$. The channel transition probability of W' is

$$\begin{aligned} W'(y'|x) &= \sum_{\lambda' \in \Lambda'} g_{(h\sigma)^2}(y' - x + \lambda'), && y' \in \mathcal{R}(\Lambda') \\ &= \sum_{\lambda' \in \Lambda'} g_{(h\sigma)^2}(n' + \lambda'), && n' \in \mathcal{R}(\Lambda') \qquad (14) \end{aligned}$$

where $g_{(h\sigma)^2}$ represents the density function of $\mathcal{N}(0, (h\sigma)^2\mathbb{I})$. The channel transition probability of W'' is

$$\begin{aligned} W''(y''|x) &= \sum_{\lambda'' \in \Lambda''} g_{(h\sigma)^2}(y'' - x + \lambda''), && y'' \in \mathcal{R}(\Lambda'') \\ &= \sum_{\lambda'' \in \Lambda''} g_{(h\sigma)^2}(n'' + \lambda''), && n'' \in \mathcal{R}(\Lambda'') \\ &\overset{(a)}{=} \sum_{\lambda' \in \Lambda'} g_{(h\sigma)^2}(n' \frac{\tilde{B}^{-1}}{\sqrt{2}} + \lambda' \frac{\tilde{B}^{-1}}{\sqrt{2}}), && n' \in \mathcal{R}(\Lambda') \\ &= \sum_{\lambda' \in \Lambda'} g_{(h\sigma)^2}\left(\frac{\tilde{B}^{-1}}{\sqrt{2}}(n' + \lambda')\right), && n' \in \mathcal{R}(\Lambda') \\ &= \sum_{\lambda' \in \Lambda'} g_{(h\sigma\sqrt{2}\tilde{B})^2}(n' + \lambda'), && n' \in \mathcal{R}(\Lambda') \qquad (15) \end{aligned}$$

where the equality (a) is due to the relation $\Lambda' = \sqrt{2}\tilde{B}\Lambda''$, $\lambda' = \sqrt{2}\tilde{B}\lambda''$, and $n' \in \mathcal{R}(\Lambda')$, $n'' \in \mathcal{R}(\Lambda'')$. We observe from equation (15) that channel W'' is statistically equivalent to $W'' : y'' = x + z'' \mod \Lambda'$ where $z'' \sim \mathcal{N}(0, (h\sigma\sqrt{2}\tilde{B})^2)$. Since the transition probabilities in equation (14) and equation (15) are two Λ'-aliased Gaussian distributions featured with variances $(h\sigma)^2 < (h\sigma\sqrt{2}\tilde{B})^2$, we can prove W'' is degraded with respect to W' by introducing an intermediate L^n/Λ' channel W''' with additive Gaussian noise $z''' \sim \mathcal{N}(0, (h\sigma\sqrt{2}\tilde{B})^2 - (h\sigma)^2)$ such that W'' is a concatenation of W' and W''', i.e.,

$$\begin{aligned} (x + z'') \mod \Lambda' &= (x + z' + z''') \mod \Lambda' \\ &= \big((x + z') \mod \Lambda'\big) + z''' \mod \Lambda'. \end{aligned}$$

The above concatenation satisfies the definition of channel degradation (Definition 3).

Give the channel degradation relation between the fading L^n/Λ' channel W' and the fading L^n/Λ'' channel W'', we can guarantee that the polar codes constructed for W'' also fit in with W'. How to explicitly construct polar codes will be shown in next section.

V. Polar Coding for the Tailored RLWE Channel

As discussed in Section II-D, we need a BDMS channel before we can adapt the polar coding methods, including calculating the Bhattacharyya parameters of the synthesized channels, defining the information indices and frozen indices, encoding and SC decoding. We have already proved the fading L/L' channel $W : X \to (\tilde{Y}, H)$ as in formula (12) is a BMS channel in Section III-C. Since we assume the channel gain H and Gaussian noise Z to be continuous and so is the channel output, we need to discretize the channel output $H, \tilde{Y}$ before constructing polar codes. An elegant channel quantization scheme was proposed in [24] where the two output H and $\tilde{Y}$ are quantized independently with reasonable loss of channel capacity. Basically, the channel gain H is quantized into a series of discrete values with uniform occurrence probability. As for the output $\tilde{Y}$, we will decompose the L/L' channel into multiple BDMS channels such that the overall channel capacity almost preserves with negligible loss.

A. *Quantization of the Fading Coefficient*

As discussed in previous sections, the fading L/L' channel with CSI available to the decoder is statistically equivalent to an independent combination of the fading coefficient H and an L/L' channel with additive Gaussian noise of variance $(h\sigma)^2$. Therefore, we can firstly quantize H then the L/L' channel. Define an ascending sequence $\{\alpha_i\}$ in the following form

$$\alpha_1 = 1, \alpha_2, \cdots, \alpha_m, \alpha_{m+1} = +\infty,$$

so that for $1 \leq i \leq m$ we have

$$\int_{\alpha_i}^{\alpha_{i+1}} P_H(h)dh = \frac{1}{m}.$$

We take the centroid with respect to the interval (α_i, α_{i+1}) as the quantized alphabet $\mathcal{H}_q = \{h_i\}$ for $i = 1, \cdots, m$ where h_i is calculated as follows.

$$h_i = \int_{\alpha_i}^{\alpha_{i+1}} mhP_H(h)dh.$$

B. *Degrading Transform Quantization*

As in Figure 2 we viewed the tailored RLWE channel as an i.i.d. fading channel. For such a channel, polar codes are constructed in [24] to achieve the ergodic capacity $C(W)$ as long as the receiver knows the CSI and the transmitter knows the CDI. Given N ($N = 2^m, m \in \mathbb{Z}$) i.i.d. tailored RLWE channels $W : X \to (\tilde{Y}, H)$, we define the channel input as $X^{1:N} = U^{1:N} G_N$ where $U^{1:N} \in \{0, 1\}^{1:N}$ and G_N is the generator matrix[7]. We obtain N synthesized channels $W_N^{(i)} : U^{(i)} \to (U^{1:i-1}, \tilde{Y}^{1:N}, H^{1:N})$ for

[7] We are using the notation from coding theory where capital N indicates the block length of a channel; it is equal to the degree n of the $2n$-th cyclotomic polynomial $x^n + 1$.

$1 \leq i \leq N$ by performing channel combining and channel splitting. The Bhattacharyya parameter for W is given by

$$Z(W) \triangleq \sum_{\tilde{y},h} \sqrt{P_{\tilde{Y},H|X}(\tilde{y},h|0) P_{\tilde{Y},H|X}(\tilde{y},h|1)}.$$

To compute $Z(W_N^{(i)})$ efficiently, we employ the degrading transform proposed in [34] to quantize a BMS channel W with continuous output alphabet into a degraded and approximated channel W_Q with finite output alphabet size. Intuitively, the finer the quantized output alphabet is, the better W_Q approximates W. Since we have already quantized H as h_i for $i = 1, \cdots, m$, we can consider h_i as a constant and quantize the L/L' channel $W_{h_i} : \tilde{Y} \leftarrow X, h_i$ for each h_i.

We define the likelihood ratio (LR) of a channel W as

$$\lambda(\tilde{y}, h_i) := \frac{W_{\tilde{Y}|X,h_i}(\tilde{y}|0, h_i)}{W_{\tilde{Y}|X,h_i}(\tilde{y}|\lfloor\frac{q}{2}\rfloor\frac{1}{\sqrt{2}}, h_i)}, \tag{16}$$

where the transition probability $W_{\tilde{Y}|X,h_i}$ is

$$W_{\tilde{Y}|h_i,X}(\tilde{y}|0, h_i) = f_{L',0,h_i^2\sigma^2}(\tilde{y}) = \sum_{\lambda \in L'} g_{0,(h_i\sigma)^2}(\tilde{y} + \lambda),$$

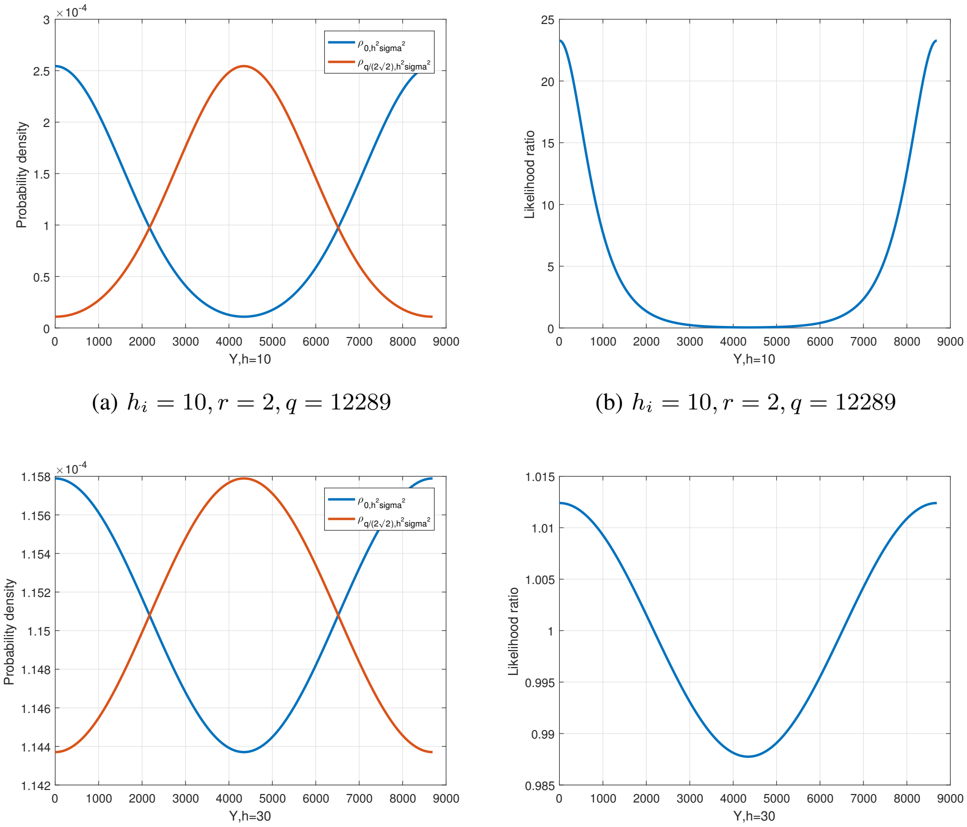
$$W_{\tilde{Y}|h_i,X}(\tilde{y}|\lfloor\frac{q}{2}\rfloor\frac{1}{\sqrt{2}}, h_i) = f_{L',\lfloor\frac{q}{2}\rfloor\frac{1}{\sqrt{2}},h_i^2\sigma^2}(\tilde{y}) = \sum_{\lambda \in L'} g_{\lfloor\frac{q}{2}\rfloor\frac{1}{\sqrt{2}},(h_i\sigma)^2}(\tilde{y} + \lambda).$$

Figure 5 depicts $W_{\tilde{Y}|X,h_i}$ and $\lambda(\tilde{y}, h_i)$ by giving some examples when $q = 12289, r = 2$ and $h_i = 10, 20$. We can see it in Figure 5 that the channel $W_{h_i} : \tilde{Y} \leftarrow X, h_i$ is BMS with $\tilde{Y}$ continuously located over the interval $[0, q/\sqrt{2})$. There exists a permutation function $\pi(\tilde{y}) = (\lfloor\frac{q}{2}\rfloor\frac{1}{\sqrt{2}} - \tilde{y}) \mod q/\sqrt{2}$ such that $W(\tilde{y}|0, h_i) = W(\pi(\tilde{y})|\lfloor\frac{q}{2}\rfloor\frac{1}{\sqrt{2}}, h_i)$. Intuitively, the BMS channel W_{h_i} can be decomposed into infinite binary symmetric channel (BSC) channels $W_c : \tilde{Y}_c \leftarrow X, h_i$ where the output is $\tilde{Y}_c \in \{y_c, \pi(y_c)\}$ for continuous $y_c \in [0, q/\sqrt{2})$, $X \in \{0, \lfloor\frac{q}{2}\rfloor\frac{1}{\sqrt{2}}\}$ and the crossover probability is the corresponding probability density $W(\tilde{y}_c|X, h_i)$. If we focus on the likelihood ratio $\lambda(\tilde{y}_c, h_i) \geq 1$, the crossover probability of BSC channel W_c is $\frac{1}{\lambda(\tilde{y}_c,h_i)+1}$. The capacity of this BSC is

$$C[\lambda(\tilde{y}_c, h_i)] = 1 - \frac{\lambda(\tilde{y}_c, h_i)}{\lambda(\tilde{y}_c, h_i) + 1} \log \frac{\lambda(\tilde{y}_c, h_i) + 1}{\lambda(\tilde{y}_c, h_i)} - \frac{1}{\lambda(\tilde{y}_c, h_i) + 1} \log (\lambda(\tilde{y}_c, h_i) + 1),$$

where $\lambda(\tilde{y}_c, h_i) \geq 1$. Quantitatively, the continuous decomposition of W_{h_i} preserves the channel capacity in the sense that

$$C(W_{h_i}) = \int_{\lambda(\tilde{y},h_i)\geq 1} (W_{\tilde{Y}|X,h_i}(\tilde{y}|0, h_i) + W_{\tilde{Y}|X,h_i}(\tilde{y}|\lfloor\frac{q}{2}\rfloor\frac{1}{\sqrt{2}}, h_i)) C[\lambda(\tilde{y}, h_i)]\, d\tilde{y},$$



(a) $h_i = 10, r = 2, q = 12289$

(b) $h_i = 10, r = 2, q = 12289$

(c) $h_i = 30, r = 2, q = 12289$

(d) $h_i = 30, r = 2, q = 12289$

Fig. 5. The probability density and likelihood ratio of $W : X, h_i \to \tilde{Y}$.

where the integral interval is restricted to $\tilde{y}$ such that $\lambda(\tilde{y}, h_i) \geq 1$. If we ignore the subtle geometrical error introduced by rounding $\lfloor \cdot \rceil$, we can observe a symmetry feature in the graphs in Figure 5 and we find that the valid integral interval is

$$A := [0, \lfloor \frac{q}{2} \rceil \frac{1}{2\sqrt{2}}] \cup [\lfloor \frac{q}{2} \rceil \frac{3}{2\sqrt{2}}, q\frac{1}{\sqrt{2}}].$$

We divide the interval A into ν segments A_j for $j \in [\nu]$ such that

$$\begin{aligned} A_j &= \left\{ \tilde{y} \in A : \frac{j-1}{\nu} \leq C[\lambda(\tilde{y}, h_i)] < \frac{j}{\nu} \right\} \\ &= \left\{ \tilde{y} \in A : \frac{1}{\mathfrak{h}_2^{-1}\left(\frac{\nu-i+1}{\nu}\right)} - 1 \leq \lambda(\tilde{y}, h_i) < \frac{1}{\mathfrak{h}_2^{-1}\left(\frac{\nu-i}{\nu}\right)} - 1 \right\}, \end{aligned}$$

where $\mathfrak{h}_2(\cdot)$ is the binary entropy function. Each A_j corresponds to a BSC channel with crossover probability

$$p_j = \frac{\int_{A_j} W_{\tilde{Y}|X,h_i}(\tilde{y}|\lfloor \frac{q}{2} \rceil \frac{1}{\sqrt{2}}, h_i) d\tilde{y}}{\int_{A_j} W_{\tilde{Y}|X,h_i}(\tilde{y}|\lfloor \frac{q}{2} \rceil \frac{1}{\sqrt{2}}, h_i) d\tilde{y} + \int_{A_j} W_{\tilde{Y}|X,h_i}(\tilde{y}|0, h_i) d\tilde{y}}, \tag{17}$$

where

$$\int_{A_j} W_{\tilde{Y}|X,h_i}(\tilde{y}|0,h_i)d\tilde{y} = \int_{A_j} \sum_{\lambda \in L'} g_{0,(h_i\sigma)^2}(\tilde{y}+\lambda)d\tilde{y}$$

$$\int_{A_j} W_{\tilde{Y}|X,h_i}(\tilde{y}|\lfloor \frac{q}{2} \rfloor \frac{1}{\sqrt{2}}, h_i)d\tilde{y} = \int_{A_j} \sum_{\lambda \in L'} g_{\lfloor \frac{q}{2} \rfloor \frac{1}{\sqrt{2}},(h_i\sigma)^2}(\tilde{y}+\lambda)d\tilde{y}.$$

Since lattice L' is infinite, we can numerically approximate $f_{L',0,h_i^2\sigma^2}(\tilde{y})$, $f_{L',\lfloor \frac{q}{2} \rfloor \frac{1}{\sqrt{2}},h_i^2\sigma^2}(\tilde{y})$ then $\lambda(\tilde{y},h_i)$, A_j and p_j.

If we define z_j and its conjugate $\bar{z}_j$ to be the channel output of the BSC associated with A_j, we will obtain the quantized output alphabet of W_{h_i} as

$$\mathcal{Z} := \{z_1, \bar{z}_1, z_2, \bar{z}_2, \cdots, z_\nu, \bar{z}_\nu\}.$$

If we denote by W_Q the quantized version of the original fading L/L' channel $W : X \to \tilde{Y}, H$, the output alphabet of W_Q is $\mathcal{H}_q \otimes \mathcal{Z} := \{h_i\} \otimes \{z_1, \bar{z}_1, \cdots, z_\nu, \bar{z}_\nu\}$ for $i \in [m]$ and $j \in [\nu]$ where $\otimes$ is a tensor product of two sets.

Lemma 4: The channel $W_Q : X \to Z, H_q$ is degraded with respect to W.

Proof 2: We supply an intermediate channel $W_P : (\tilde{Y}, H) \to (Z, H_q)$ such that

$$W_P(z, h_q|\tilde{y}, h) = \begin{cases} 1, & \text{if } z = z_j, \tilde{y} \in A_j, \text{ and } h_q = h_i, h \in [\alpha_i, \alpha_{i+1}), \\ 1, & \text{if } z = \bar{z}_j, \pi(\tilde{y}) \in A_j, \text{ and } h_q = h_i, h \in [\alpha_i, \alpha_{i+1}), \\ 0, & \text{otherwise.} \end{cases}$$

We can find there exits a channel degradation relation such that

$$W_Q(z, h_q|x) = \int W_{\tilde{Y},H|X}(\tilde{y}, h|x) W_P(z, h_q|\tilde{y}, h) d\tilde{y}\, dh.$$

Corollary 1: Given that $W_Q : X \to Z, H_q$ is degraded with respect to W, the capacity, Bhattacharyya parameter and frame error rate are of the two channels are related as follows.

$$C(W_Q) \le C(W),$$

$$Z(W_Q) \ge Z(W),$$

$$P_e(W_Q) \ge P_e(W).$$

Proof 3: As a corollary of Lemma 2 and 4.

It is also indicated in [34] that the capacity loss introduced by the degrading transform is no greater than $1/\nu$. If we choose large alphabet size m and 2ν, the loss of capacity is negligible and so is $Z(\cdot)$ and $P_e(\cdot)$. We also verified our channel quantization scheme with respect to the channel capacity. As is

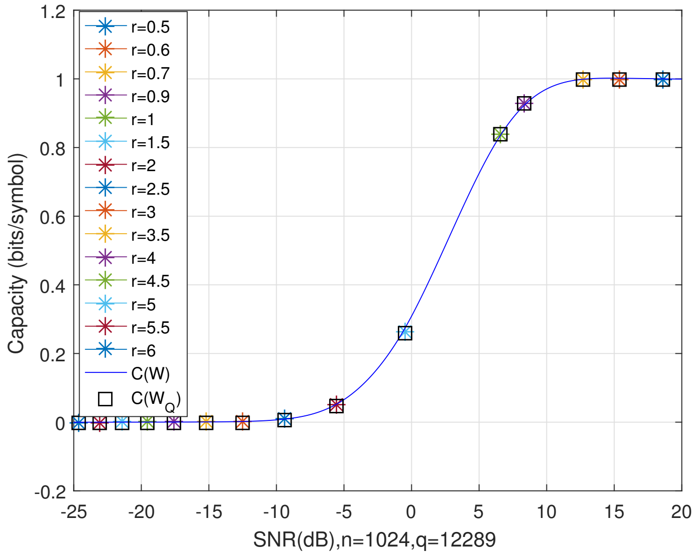


Fig. 6. A comparison between $C(W)$ and $C(W_Q)$ for different r, when $m = 20, \nu$ = 50.

shown in Figure 6, for $m = 20, \nu = 50$ and multiple choices of r, $C(W_Q)$ is comparable to $C(W)$ with only negligible difference.

To summarize, what the degrading transform does is to convert the RLWE channel W with continuous output alphabet into a BDMS channel W_Q with finite output, which can be viewed as a combination of $m \times \nu$ BSC channels. In this way the $Z(W_Q)$ can be calculated according to Definition 2.

C. *Polar Encoding and SC Decoding*

Given the BDMS channel W_Q obtained by channel quantization, we can adapt the polar encoding and decoding method introduced in Section II-D to W_Q. Recall that the output alphabet of W_Q is $m \times 2\nu$. As the channel combining and splitting process continue, the alphabet size of the synthesized channels $W_{QN}^{(i)}$ will increase exponentially as the recursion proceeds. To handle this problem, we employ an approximation method proposed in [42] which can reduce the alphabet size of a BDMS channel with negligible and traceable loss of performance by merging some of the output symbols.

After we finish computing the Bhattacharyya parameters of all the $W_{QN}^{(i)}$, we can define the information set $\mathcal{A}$ and frozen set $\mathcal{A}^c$ for code rate R. We construct the polar codewords as

$$x^{1:N} = u_{\mathcal{A}} G_N(\mathcal{A}) \oplus u_{\mathcal{A}^c} G_N(\mathcal{A}^c).$$

Upon observing the signal $\tilde{y}^{1:N}$ and invoking their knowledge of the CSI $h^{1:N}$, the recipient can apply the successive cancellation (SC) decoder to $\tilde{y}^{1:N}, h^{1:N}$ and give an estimation of $u_{\mathcal{A}}$ according to the decision function

$$\overline{u}^{(i)} = \begin{cases} 0, & \text{if } L_N^{(i)}(\tilde{y}^{1:N}, h^{1:N}, \overline{u}^{1:i-1}) \geq 1 \\ 1, & \text{otherwise} \end{cases},$$

where the likelihood ratio $L_N^{(i)}(\tilde{y}^{1:N}, h^{1:N}, \overline{u}^{1:i-1}) \triangleq \frac{W_N^{(i)}(\tilde{y}^{1:N}, h^{1:N}, \overline{u}^{1:i-1}|0)}{W_N^{(i)}(\tilde{y}^{1:N}, h^{1:N}, \overline{u}^{1:i-1}|1)}$ can be calculated recursively by SC decoding algorithm in [30]. Note that the input of SC decoder $\lambda(\tilde{y}, h)$ is given in formula (16). A block decoding error occurs if $\overline{u}^{1:N} \neq u^{1:N}$; we may interchangeably use block error probability and DFR in this work. The complexity of both polar encoding and SC decoding are $O(N \log N)$. Additionally, the two algorithms both require constant steps of operations for fixed choices of $R, N, \mathcal{A}$, making constant-time implementations possible.

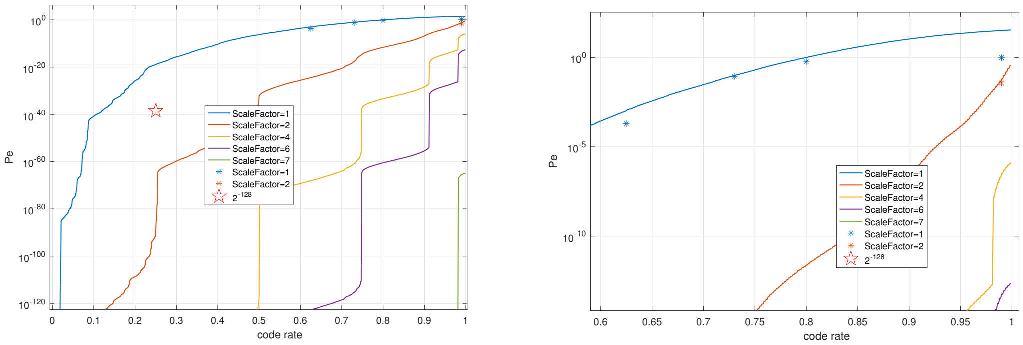
VI. Results: Performance Analysis and Improvement

According to Theorem 2, the block error probability $P_e(N, R, \mathcal{A})$ of SC decoding is upper bounded by the sum of $Z(W_N^{(i)})$. Since $W_Q \preceq W$ and $W_{QN}^{(i)} \preceq W_N^{(i)}$ according to Lemma 1, we derive

$$P_e(N, R, \mathcal{A}) \leq \sum_{i \in \mathcal{A}} Z(W_{NQ}^{(i)}). \tag{18}$$

Recall it in Figure 6 that the capacity of our tailored RLWE channel deteriorates dramatically because we use a tailored and shrunk constellation diagram. As a result, for most choices of r which are believed to be secure in RLWE-based PKE, we cannot obtain a desired DFR lower than 2^{-128} which is used as a benchmark in NIST standardization. As explained in Section III-B, we carefully and conservatively choose a cube Λ'' which is enclosed in the maximal sphere inscribed in Λ'. Almost surely there exists other valid choices of Λ'' lager than the one we choose, though it is not easy at all to figure out the optimal one. Our solution to this harsh problem is to gradually scale Λ'' up by a factor $t \geq 1$ and run simulations for each to justify if the numerical results of P_e coincide with the upper bound in formula (18).

Figure 7 compares the upper bound of decoding probability P_e with our simulation results in the setting of $q = 12289, n = 1024, r = 1$. The solid lines indicate the upper bound of decoding error probability P_e with respect to different code rate R. The solid lines with stars represent our simulation results which, for reasonably small DFR, comply with the upper bound. We aim to achieve $P_e = 2^{-128}$ at code rate $R = 0.25$. Obviously it is unachievable when the scale factor $t = 1$ though our experimental results indicated by the blue stars comply with our estimation indicated by the blue line. We gradually increase t and obtain the corresponding estimation of P_e. We can see that the decoding performance is improved significantly upon a slightly larger t, e.g., P_e is smaller than $10^{-60}(\approx 2^{-200})$ at $R = 0.25$ for $t = 2$. When $t = 2$, the experiment result, the red star, also complies with its corresponding theoretical estimation, i.e., the red solid line. It implies that our estimation of P_e for $t = 2$ is reliable to some extent. Please note that all these experiments target at relatively large P_e which is feasible to verify.



(a) $r = 1$

(b) $r = 1$, enlarged view

Fig. 7. Decoding error probability for RLWE-Based PKE with $q = 12289, n = 1024, r = 1$ and multiple choices of scale factor t: simulation results vs. upper bound.

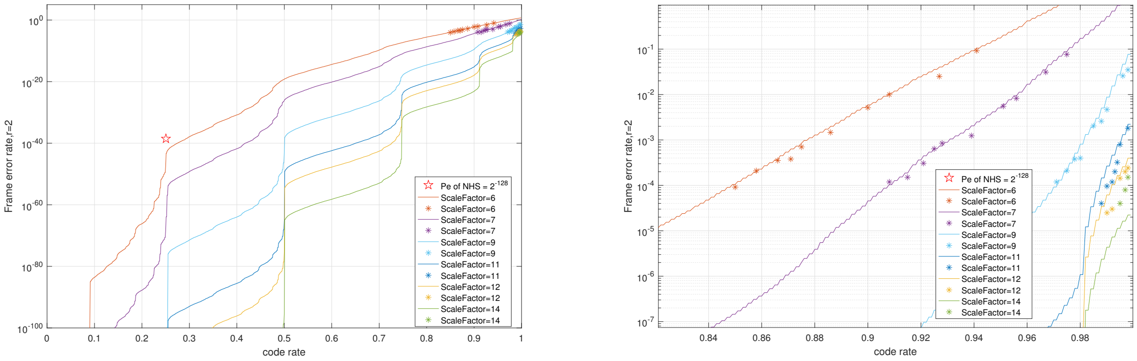
Figure 8 can be interpreted in the same manner as Figure 7. The only different parameter used here is $r = 2$. The solid lines in different colors represent our estimation of P_e and the stars are our simulation results. By making scale factor t as large as 6, the target R and P_e can be achieved. For relatively large P_e shown in the graph, we observed that our simulation results comply with our estimation when $t = 6, 7, 9, 11, 12$. However, when $t = 14$, simulation results are worse than our estimation, implying that the constellation diagram Λ'' is overwhelmingly large and goes beyond the valid domain.

In Figure 9, $r = 2.83$. We can observe that our estimation are effective for $t = 8, 12$ but fails for $t > 12$. We can see that none of our simulation results comply with the estimation in Figure 10. It implies that the scaling method cannot be applied for $r \geq 3.46$.

Remark 3: The error sources for the scaled and tailored RLWE channel model are concluded as follows:

(a) As t increases, the constellation space Λ'' may go beyond Λ' and our model will fail to describe the statistical feature of the real channel;

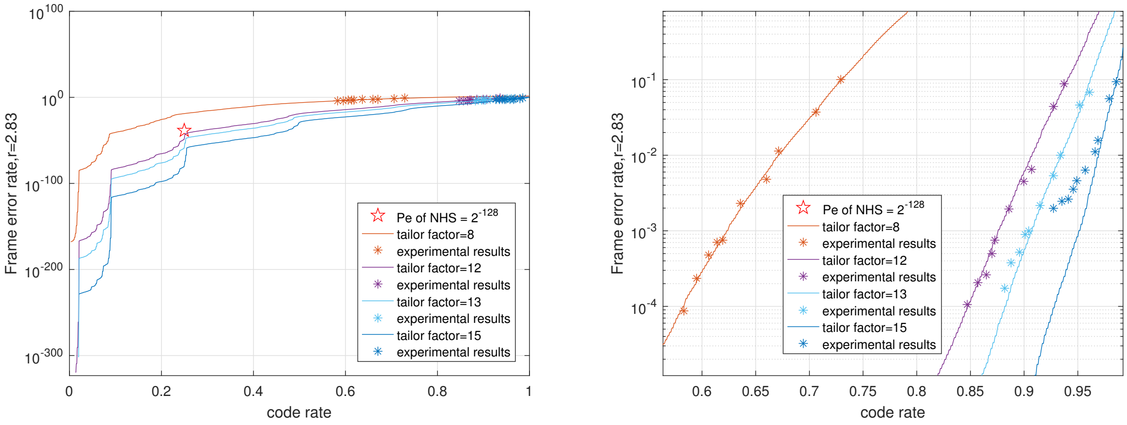
(b) The SC decoder takes $\tilde{B}\mathbf{y}$ to be the channel output of a fading L^n/Λ'' channel while it is actually a fading $L^n/\Lambda'/\Lambda''$ channel according to Table I. This is because Alice firstly performs a mod R_q operation and then calculates $\tilde{B}\mathbf{y}$ (equivalent to L^n/Λ') upon receiving $\mathbf{y}$ from Bob. For small r, the two have quite close distributions but become less likely as r goes larger. This explain why our model fails when $r \geq 3.46$ in Figure 10.



(a) $r = 2$

(b) $r = 2$, enlarged view

Fig. 8. Decoding error probability for RLWE-Based PKE with $q = 12289, n = 1024, r = 2$ and multiple choices of scale factor t: simulation results vs. upper bound.



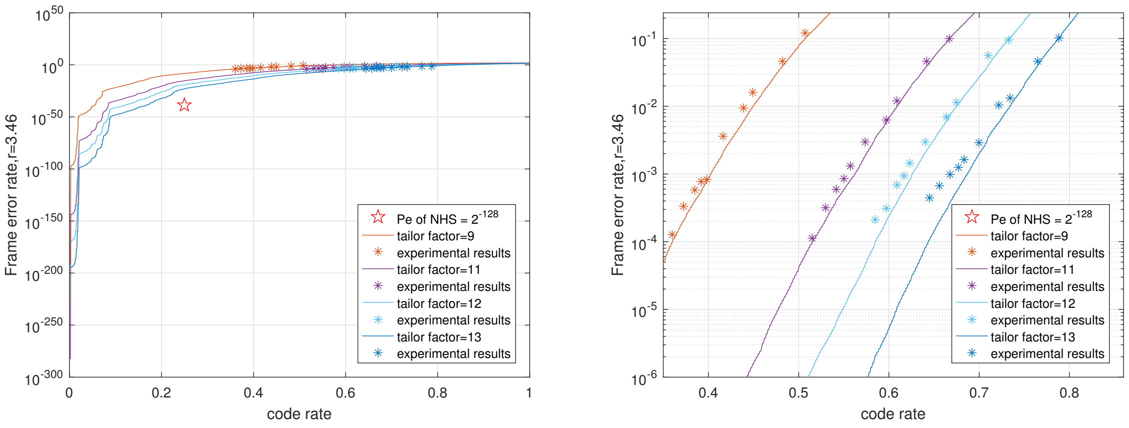
(a) $r = 2.83$

(b) $r = 2.83$, enlarged view

Fig. 9. Decoding error probability for RLWE-Based PKE with $q = 12289, n = 1024, r = 2.83$ and multiple choices of scale factor t: simulation results vs. upper bound.

VII. Discussion: Improving Security Using New DFR

There exists a trade-off relation between DFR and security level of RLWE-based PKE. Basically, larger error term (or large binomial parameter k in NewHope) gives better security but worse DFR. The motivation of this work is to exploit polar codes to give a safer DFR margin such that we can improve



(a) $r = 3.46$

(b) $r = 3.46$, enlarged view

Fig. 10. Decoding error probability for RLWE-Based PKE with $q = 12289, n = 1024, r = 3.46$ and multiple choices of scale factor t: simulation results vs. upper bound.

the security level while achieving the target DFR. In NIST standardization, this target DFR is 2^{-128}. A more conservative target 2^{-140} is used in prior literature [16], [13].

The concrete security analysis of RLWE-based PKE, so far, is based on the hardness of LWE as in [4][8]. Essentially, we will consider (a) a primal attack which consists of constructing and solving unique-SVP given the LWE samples (b) a dual attack which consists of searching for the shortest vector in a dual lattice given LWE samples. Table II illustrates the DFR and security of RLWE-based PKE using our polar coding scheme for different choices of binomial parameter k ($r = \sqrt{k/2}$) and scale factor t. As we discussed in last section, the scale factor of the constellation diagram cannot be larger than 12 for $k = 8$, otherwise the estimation of DFR is no longer valid. We select a more conservative choice $t = 11$ and achieve DFR$= 2^{-298}$ for $n = 1024, q = 12289, k = 8$ using our polar coding scheme as is smaller than the DFR 2^{-216} of NewHope round 2 in the same setting. When $k = 16$, we can achieve a DFR of 2^{-156} while in [7] NewHope gives a DFR of 2^{-61} in the same setting. As discussed in Figure 10, our calculation of DFR for $k \geq 24$ ($r \geq 3.46$) is no longer applicable.

In conclusion, our polar coding scheme allows RLWE-based PKE to achieve the target DFR 2^{-140} (and also 2^{-128}) while improving the security by 9.4% by using larger $k = 16$. The state-of-the-art study of this kind can be classified into two categories. In [13], LDPC codes and BCH codes are used to improve the security level by 31.76% while achieving DFR of 2^{-140}. However, their DFR estimation highly relies on an "independence assumption" and their error-correcting algorithms are not constant-time. In [16],

[8] The security estimator is available at https://github.com/tpoeppelmann/newhope.

Song et al. gave a tighter bound on DFR without amending the coding scheme of NewHope Simple and the security is improved by 7.2%.

TABLE II

Improved security of RLWE-PKE using polar coding for $n = 1024, q = 12289$.

k	t	DFR	cost of primal attacks classical/quantum [bits]	cost of dual attacks classical/quantum [bits]
8	6	2^{-159}	259/235	257/233
8	7	2^{-229}		
8	9	2^{-252}		
8	11	2^{-298}		
16	12	2^{-156}	282/256	281/255
24	N/A			
$k = 14$,[16]		2^{-156}	278/252	276/250
$k = 66$,[13]		2^{-140}	341/309	338/307

VIII. Conclusions

We have presented the first example of a polar coding technique to improve the DFR of RLWE-based PKE which takes the advantage of viewing protocol as a fading channel with CSI known to the decoder. Moreover, switching from polynomial basis to canonical basis unfasten the dependency existing in the residue noise term though a new correlation is introduced to the message term. The constellation space is tailored to construct an i.i.d. fading channel at the cost of decoding performance and a scaling method was employed to compensate the performance loss. Both numerical and theoretical results are given to verify the DFR estimation. The advantages of our method are as follows.

- We derive an i.i.d. channel model of the residue noise term in H space using canonical embeding. The actuality that some knowledge of noise term is known by the decoder is exploited to improve the decoding performance.
- The security is improved by 9.4% while achieving the target DFR of 2^{-140} in the setting of $n = 1024, q = 12289, k = 16$ $(r = 2.83)$. This improvement is better than the benchmark 7.2% in [16]. Though it is not an attractive as the record 31.76% in [13], their results rely on an "independence assumption" that may not hold.
- Polar codes enables constant-time implementation of encoding and decoding while BCH and LDPC codes employed in [13] do not. The quantity of operations required by encoding and decoding is solely determined by block length N and code rate R.

The disadvantages are also given as follows.

- Switching between the two basis by multiplying matrix $\tilde{B}$ and $\tilde{B}^{-1}$ as in Table I increases the complexity of the protocol.
- To give the i.i.d. channel model, we design a tailored constellation diagram which gives shorter code distance than the original modulation $\{0, \lfloor \frac{q}{2} \rfloor\}$. It will hurt the decoding performance though the power of polar coding and the scale factor t can counteract its effects to some extent.

Appendix

SNR is defined as SNR $=$ P$_{\text{signal}}$/P$_{\text{noise}}$ where P$_{\text{signal}}$ and P$_{\text{noise}}$ denote the signal and noise power, respectively.The channel model of RLWE-based PKE in polynomial basis is

$$\lfloor \frac{q}{2} \rfloor \cdot m + e \cdot t - s \cdot e_1 + e_2,$$

where e, t, s, e_1, e_2 are polynomials in $\dfrac{\mathbb{Z}_q[x]}{x^n + 1}$ whose coordinates are independently drawn from a spherical normal $\mathcal{N}(0, \sigma^2 \mathbb{I})$. The multiplication of two polynomials can be interpreted as the convolution of their coordinates, giving rise to n parallel and correlated channels. If we set message m to be $\{-\lfloor \frac{q}{4} \rfloor, \lfloor \frac{q}{4} \rfloor\}^n$, the SNR is roughly $\dfrac{q^2}{16(2n\sigma^4 + \sigma^2)}$.

In the tailored RLWE channel model in formula (11) and (12), polynomials convolutions are transformed to component-wise multiplication in canonical basis. Since we shrink the constellation diagram as described in Section III-B, we set the message m to be $\{-\lfloor \frac{q}{4\sqrt{2}} \rfloor, \lfloor \frac{q}{4\sqrt{2}} \rfloor\}^n$. The channel gain H_i and Gaussian noise Z_i are independent and their distribution are discussed in detail in formula (11). The SNR is $\dfrac{q^2}{32(n^2\sigma^4 + n\sigma^2/2)}$.

Acknowledgment

We thank Charles Grover for providing support on the formulation of RLWE channel from a number theoretic point of view.

References

[1] V. Lyubashevsky, C. Peikert, and O. Regev, "On ideal lattices and learning with errors over rings," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2010, pp. 1–23.

[2] O. Regev, "On lattices, learning with errors, random linear codes, and cryptography," in *Proceedings of the Thirty-seventh Annual ACM Symposium on Theory of Computing*, ser. STOC '05. New York, NY, USA: ACM, 2005, pp. 84–93.

[3] NIST, "Submission requirements and evaluation criteria for the post-quantum cryptography standardization process," 2016.

[4] E. Alkim, L. Ducas, T. Pöppelmann, and P. Schwabe, "Post-quantum key exchange—a new hope," in *25th {USENIX} Security Symposium ({USENIX} Security 16)*, 2016, pp. 327–343.

[5] X. Lu, Y. Liu, Z. Zhang, D. Jia, H. Xue, J. He, B. Li, K. Wang, Z. Liu, and H. Yang, "Lac: Practical ring-lwe based public-key encryption with byte-level modulus." *IACR Cryptol. ePrint Arch.*, vol. 2018, p. 1009, 2018.

[6] J. Ding, X. Xie, and X. Lin, "A simple provably secure key exchange scheme based on the learning with errors problem." *IACR Cryptology EPrint Archive*, vol. 2012, p. 688, 2012.

[7] E. Alkim, L. Ducas, T. Pöppelmann, and P. Schwabe, "Newhope without reconciliation." *IACR Cryptology ePrint Archive*, vol. 2016, p. 1157, 2016.

[8] V. Lyubashevsky, C. Peikert, and O. Regev, "A toolkit for ring-lwe cryptography," in *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 2013, pp. 35–54.

[9] M. Hamburg, "Post-quantum cryptography proposal: Three bears," 2017.

[10] E. Fujisaki and T. Okamoto, "Secure integration of asymmetric and symmetric encryption schemes," in *Annual International Cryptology Conference*. Springer, 1999, pp. 537–554.

[11] A. van Poppelen, "Cryptographic decoding of the leech lattice," *IACR Cryptology ePrint Archive*, vol. 2016, p. 1050, 2016.

[12] NIST, "Status report on the second round of the nist post-quantum cryptography standardization process. https://nvlpubs.nist.gov/nistpubs/ir/2020/nist.ir.8309.pdf," Tech. Rep., 2020.

[13] T. Fritzmann, T. Pöppelmann, and M. J. Sepúlveda, "Analysis of error-correcting codes for lattice-based key exchange," *IACR Cryptology ePrint Archive*, vol. 2018, p. 150, 2018.

[14] M.-J. O. Saarinen, "Hila5: on reliability, reconciliation, and error correction for ring-lwe encryption," in *International Conference on Selected Areas in Cryptography*. Springer, 2017, pp. 192–212.

[15] J.-P. D'Anvers, F. Vercauteren, and I. Verbauwhede, "The impact of error dependencies on ring/mod-lwe/lwr based schemes," in *International Conference on Post-Quantum Cryptography*. Springer, 2019, pp. 103–115.

[16] M. Song, S. Lee, D. Shin, E. Lee, Y. Kim, and J. No, "Analysis of error dependencies on newhope," *IEEE Access*, vol. 8, pp. 45 443–45 456, 2020.

[17] S. Murphy and R. Player, "Discretisation and product distributions in ring-lwe," Cryptology ePrint Archive, Report 2019/596, 2019, https://eprint.iacr.org/2019/596.

[18] ——, "δ-subgaussian random variables in cryptography," in *Information Security and Privacy*, J. Jang-Jaccard and F. Guo, Eds. Cham: Springer International Publishing, 2019, pp. 251–268.

[19] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE transactions on information theory*, vol. 44, no. 3, pp. 927–946, 1998.

[20] A. Martinez, A. G. i Fabregas, G. Caire, and F. M. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2756–2765, 2009.

[21] U. Wachsmann, R. F. Fischer, and J. B. Huber, "Multilevel codes: Theoretical concepts and practical design rules," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1361–1391, 1999.

[22] H. Imai and S. Hirakawa, "A new multilevel coding method using error-correcting codes," *IEEE Transactions on Information Theory*, vol. 23, no. 3, pp. 371–377, 1977.

[23] G. D. Forney, "Coset codes. i. introduction and geometrical classification," *IEEE Transactions on Information Theory*, vol. 34, no. 5, pp. 1123–1151, Sep. 1988.

[24] L. Liu and C. Ling, "Polar Codes and Polar Lattices for Independent Fading Channels." *IEEE Transaction on Communications*, vol. 64, no. 12, pp. 4923–4935, 2016.

[25] E. K. Hall and S. G. Wilson, “Design and analysis of turbo codes on Rayleigh fading channels,” *IEEE journal on selected areas in communications*, vol. 16, no. 2, pp. 160–174, 1998.

[26] P. Trifonov, “Design of polar codes for rayleigh fading channel,” in *2015 International Symposium on Wireless Communication Systems (ISWCS)*, 2015, pp. 331–335.

[27] A. Bravo-Santos, “Polar codes for the Rayleigh fading channel,” *IEEE Communications Letters*, vol. 17, no. 12, pp. 2352–2355, 2013.

[28] S. Liu, Y. Hong, and E. Viterbo, “Polar codes for block fading channels,” in *2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, 2017, pp. 1–6.

[29] M. Zheng, M. Tao, W. Chen, and C. Ling, “Polar coding for block fading channels,” *CoRR*, vol. abs/1701.06111, 2017.

[30] E. Arikan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[31] R. Mori and T. Tanaka, “Performance and construction of polar codes on symmetric binary-input memoryless channels,” in *2009 IEEE International Symposium on Information Theory*, 2009, pp. 1496–1500.

[32] R. Mori and T. Tanaka, “Performance of polar codes with the construction using density evolution,” *IEEE Communications Letters*, vol. 13, no. 7, 2009.

[33] R. Mori, “Properties and construction of polar codes,” Master’s thesis, Kyoto University, 2010.

[34] I. Tal and A. Vardy, “How to construct polar codes,” *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6562–6582, 2013.

[35] S. B. Korada, “Polar codes for channel and source coding,” EPFL, Tech. Rep., 2009.

[36] E. G. Kocer, “Circulant, negacyclic and semicirculant matrices with the modified pell, jacobsthal and jacobsthal-lucas numbers,” *Hacettepe Journal of Mathematics and Statistics*, vol. 36, no. 2, 2007.

[37] A. Campello, L. Liu, and C. Ling, “Multilevel code construction for compound fading channels,” 2017.

[38] E. Crockett and C. Peikert, “Challenges for ring-lwe.” *IACR Cryptol. ePrint Arch.*, vol. 2016, p. 782, 2016.

[39] G. D. Forney, M. D. Trott, and S.-Y. Chung, “Sphere-bound-achieving coset codes and multilevel coset codes,” *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 820–850, May 2000.

[40] D. Stehlé, R. Steinfeld, K. Tanaka, and K. Xagawa, “Efficient public key encryption based on ideal lattices,” in *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 2009, pp. 617–635.

[41] V. Lyubashevsky, C. Peikert, and O. Regev, “On ideal lattices and learning with errors over rings,” *Journal of the ACM (JACM)*, vol. 60, no. 6, pp. 1–35, 2013.

[42] R. Pedarsani, S. H. Hassani, I. Tal, and E. Telatar, “On the construction of polar codes,” in *2011 IEEE International Symposium on Information Theory Proceedings*, 2011, pp. 11–15.