SCIENCE CHINA
Information Sciences

• REVIEW •

January 2022, Vol. 65 111301:1–111301:21
https://doi.org/10.1007/s11432-021-3354-7

# Quantum-safe cryptography: crossroads of coding theory and cryptography

Jiabo WANG[1], Ling LIU[2], Shanxiang LYU[3], Zheng WANG[4], Mengfan ZHENG[5], Fuchun LIN[5], Zhao CHEN[1], Liuguo YIN[1], Xiaofu WU[6] & Cong LING[5*]

[1] *Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing* 100084*, China;*
[2] *Department of Software Engineering, Shenzhen University, Shenzhen* 518000*, China;*
[3] *College of Cyber Security, Jinan University, Guangzhou* 510632*, China;*
[4] *School of Information Science and Engineering, Southeast University, Nanjing* 211189*, China;*
[5] *Department of Electrical and Electronic Engineering, Imperial College London, London* SW7 2AZ*, UK;*
[6] *National Engineering Research Center of Communications and Networking,*
*Nanjing University of Posts and Telecommunications, Nanjing* 210023*, China*

Received 1 September 2021/Accepted 26 October 2021/Published online 21 December 2021

**Abstract** We present an overview of quantum-safe cryptography (QSC) with a focus on post-quantum cryptography (PQC) and information-theoretic security. From a cryptographic point of view, lattice and code-based schemes are among the most promising PQC solutions. Both approaches are based on the hardness of decoding problems of linear codes with different metrics. From an information-theoretic point of view, lattices and linear codes can be constructed to achieve certain secrecy quantities for wiretap channels as is intrinsically classical- and quantum-safe. Historically, coding theory and cryptography are intimately connected since Shannon's pioneering studies but have somehow diverged later. QSC offers an opportunity to rebuild the synergy of the two areas, hopefully leading to further development beyond the NIST PQC standardization process. In this paper, we provide a survey of lattice and code designs that are believed to be quantum-safe in the area of cryptography or coding theory. The interplay and similarities between the two areas are discussed. We also conclude our understandings and prospects of future research after NIST PQC standardisation.

**Keywords** post-quantum cryptography, lattice-based cryptography, code-based cryptogarphy, information-theoretic security, lattice reduction, lattice codes, linear codes

**Citation** Wang J B, Liu L, Lyu S X, et al. Quantum-safe cryptography: crossroads of coding theory and cryptography. Sci China Inf Sci, 2022, 65(1): 111301, https://doi.org/10.1007/s11432-021-3354-7

## 1 Introduction

Cryptography is a subject area of long history, playing a crucial role in human society ever since the dawn of civilization. Claude Shannon's 1945 paper "A mathematical theory of cryptography" is widely regarded as the beginning of modern cryptography, as a new science. In 1949, Shannon published another landmark paper "A mathematical theory of communication", founding the area of information theory. As noted by Shannon himself, cryptography and information theory are very close together. Cryptography is the science of concealing information, while information theory is primarily concerned with transmitting information. Thus the two subjects can benefit from each other. In fact, the insights Shannon obtained from the study of cryptography were key to his development of information theory.

The two subjects share many common concepts and formulations, and it seemed they were to live happily ever after — until the rise of public-key cryptography. In 1976, Diffie and Hellman published a famous paper "New directions in cryptography" in *IEEE Transactions on Information Theory*. Ironically, this paper triggered the decline of the influence of information theory on cryptography, whereas cryptography as a stand-alone subject has grown tremendously in the past few decades. This is because

[*] Corresponding author (email: c.ling@imperial.ac.uk)

modern cryptography is mostly based on computational security, rather than information-theoretic security conceived by Shannon, even though the latter still plays a (relatively minor) role in cryptography. One could say that information theory and cryptography are now separate, if not divorced. Fortunately, quantum-safe cryptography offers an opportunity to bring them back together.

Current public-key cryptographic schemes based on integer factoring and discrete logarithm would collapse under quantum computing attacks using Shor's algorithm. This is a serious concern to our modern data-driven society, which has been scrutinized by governments, companies, and research institutions. Standardisation bodies such as the National Institute of Standards and Technology (NIST), European Telecommunications Standards Institute (ETSI), and International Organization for Standardization (ISO) are in the process of developing standards of post-quantum cryptography. In particular, the NIST process of post-quantum cryptography has attracted broad attention and research throughout the world. Among the prospective methods which are expected to be standardized for post-quantum cryptography, lattice and code-based cryptography emerge as the most promising approaches.

Historically, there is a close connection between code and lattice-based cryptography, since both can be viewed as linear codes. The first code-based public-key cryptosystem is the well-known McEliece cryptosystem, published in 1978. The main drawback of McEliece-type cryptosystems is the large public key size. Lattice-based cryptography (LBC) can be viewed as a renaissance of code-based cryptography, using the Euclidean metric rather than the Hamming metric. In fact, developments in lattice and code-based cryptography often inspire each other.

At a high level, both code and lattice-based cryptography are often based on a fundamental problem of coding theory, i.e., maximum-likelihood decoding: upon receiving vector

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e},$$

where $\boldsymbol{A}$ is a (public) matrix, $\boldsymbol{x}$ is a vector and $\boldsymbol{e}$ is noise, the problem is to recover $\boldsymbol{x}$. This is also known as learning with errors (LWE) in lattice-based cryptography. Since maximum-likelihood decoding is computationally hard for random $\boldsymbol{A}$, $\boldsymbol{y}$ is pseudorandom and can be used to hide a message $\boldsymbol{m}$. Therefore, the ciphertext is given by

$$\boldsymbol{y}' = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e} + \mathrm{Enc}(\boldsymbol{m}),$$

where $\mathrm{Enc}(\cdot)$ represents an error-correcting code to cope with noise, and informally, $\boldsymbol{x}$ can be thought of as a secret key. A legitimate user with access to the key will be able to decode the message $\boldsymbol{m}$ after subtracting out $\boldsymbol{A}\boldsymbol{x}$ from $\boldsymbol{y}'$, but an adversary without access to the key cannot distinguish the ciphertext $\boldsymbol{y}'$ from a random string. Note the two-fold role of noise here: it is added to hide the message, but it also causes decoding errors, prompting the usage of error correction coding. Another popular problem in code and lattice-based cryptography is syndrome decoding: upon receiving vector $\boldsymbol{y}$, computing the syndrome

$$\boldsymbol{s} = \boldsymbol{A}\boldsymbol{y},$$

the problem is to find a low-weight vector $\boldsymbol{x}$ satisfying the above syndrome equation, i.e., $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{s}$. This is also known as the inhomogeneous short integer solution (ISIS) problem in lattice-based cryptography. Clearly, LWE and ISIS are inspired by their counterparts in coding theory.

In this paper, we present a coherent view of cryptography and coding theory in the context of quantum-safe cryptography, and the bridge is built by lattices and codes. Quantum-safe cryptographic primitives and channel coding schemes are reviewed and standard designs are described. Especially for post-quantum cryptography (PQC) which has developed rapidly in the last decade, we give our understanding of the interdisciplinary "genes" of cryptography (especially the lattice and code-based primitives) and coding theory. We also conclude what can be projected into future research from the reintegration of the two areas.

**Roadmap.** This rest of the paper is organized as follows. In Section 2, an overview of the NIST standardization process is given. In Section 3, public-key encryption/key encapsulation mechanisms and signature schemes based on hard lattice problems are introduced; a key technique of lattice cryptography, lattice Gaussian sampling, is also reviewed. Section 4 presents a review of code-based cryptography. An interplay between lattices and codes can be found in Section 5. Section 6 presents information-theoretic security which will also remain secure in the post-quantum age and Section 7 summarizes the paper.

**Table 1** NIST third round finalists and alternate candidates

| Thrid-round finalists | Public-key encryption/KEM | Digital signatures |
|---|---|---|
| Lattice-based | CRYSTALS-KYBER<br>NTRU, SABER | CRYSTALS-DILITHIUM<br>FALCON |
| Code-based | Classic McEliece | – |
| Others | – | Rainbow |
| Alternate candidates | Public-key encryption/KEM | Digital signatures |
| Lattice-based | FrodoKEM, NTRU Prime | – |
| Code-based | BIKE, HQC | – |
| Others | SIKE | GeMSS, Picnic, SPHINCS+ |

## 2 Overview of NIST standardization

It is complicated and controversial to answer when a full-fledged quantum computer will be available because current quantum machines operating on a few dozen quantum bits are far from doing anything dazzling. As the world's tech giants (Intel, Google, IBM, etc.) are continuing hitting new milestones in the field of quantum computing and some governments are supporting the research strategically and financially, the reality is that the quantum revolution is happening right now and we must stay ahead of the curve in light of the approaching quantum era. One area of urging importance is the migration to post-quantum cryptography. In the past few years, industry and standard organizations have started their own activities in this field.

In December 2016 the NIST (USA) announced a call for proposals for quantum-resistant algorithms including public-key encryption (PKE)/key encapsulation mechanism (KEM) and digital signatures. While symmetric cryptography (e.g., advanced encryption standard (AES), secure Hash algorithm 2 (SHA-2)) is regarded as quantum-safe as the impact of quantum attack can be effectively mitigated by increasing the key size, the severe impact is mainly on asymmetric cryptography like RSA (Rivest-Shamir-Adleman), Diffie-Hellman, and elliptic curve. The focus of NIST PQC standardization is on public-key solutions including digital signatures and public-key encryption/key establishment.

Cryptographic researchers and practitioners from over 25 countries actively contributed to NIST's standardization process. By December of 2017, NIST received 69 valid submissions for the first round evaluation. These submissions use techniques from a number of different mathematical families, including lattices, error-correcting codes, multivariate equations, hash functions, elliptic curves, and others. In January 2019 the field was cut to 26 candidates for the second round, and in July 2020 the program proceeded to the third round (and apparently final) announcing a short list of 15 candidate proposals with 7 finalists and another 8 as optional candidates. Identified by their underlying hard problems, 7 schemes are built on lattice cryptography, 3 on code-based cryptography, 2 on multivariate methods, 1 on hash functions, and 1 on isogenies of elliptic curves[1]. As shown in Table 1, the vast majority of these successful candidates (10 out of 15) are based on lattices (7) and codes (3).

Making decisions on which proposal to adopt and standardize is sophisticated when every proposal has its pros and cons. For every submission, NIST investigates and assesses its security strength from both theoretical and practical aspects. In the context of public-key encryption/key encapsulation mechanism, NIST intends to standardize schemes that can achieve IND-CCA2 (indistinguishability under adaptive chosen ciphertext attack) security. But if ephemeral-only PKE/KEM is considered, security guarantee under chosen plaintext attack (i.e., IND-CPA security) suffices. In the case of digital signature schemes, proposals should enable existentially unforgeability with respect to adaptive chosen message attacks (i.e., EUF-CMA security). NIST categorizes 5 security levels from I to V each of which defines a security threshold. To break a proposed primitive of a certain security level is supposed to consume at least comparable computational resource to an existing NIST standard in symmetric cryptography. The computational resource is deemed to be a variety of different metrics (e.g., the number of classical elementary operations, quantum circuit size, etc.). Additional security factors NIST will consider include resistance to side-channel attacks, perfect forward secrecy, and resistance to multi-key attacks.

While the security of a PQC proposal is obviously the factor referees care about the most, NIST's competition also scopes out their complexity and compatibility. Tradeoffs have to be made between

[1] Picnic is a signature scheme in none of the above categories because it does not rely on number theoretic or structured hardness assumptions. Instead, it is designed using zero-knowledge proof and symmetric primitives.

| Parameters are $m, n, q$; error distribution $\chi$ on $\mathbb{Z}^m$ | | |
|---|---|---|
| **Alice** | | **Bob** |
| $\boldsymbol{A} \overset{\$}{\leftarrow} \mathbb{Z}_q^{m\times n}$ | | |
| $\boldsymbol{s} \overset{\$}{\leftarrow} \mathbb{Z}_q^{n}$ | | |
| $\boldsymbol{e} \leftarrow \chi$ | | $\boldsymbol{x} \overset{\$}{\leftarrow} \{0,1\}^m$ |
| $\boldsymbol{b} := \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e} \mod q$ | $\xrightarrow{(\boldsymbol{A},\boldsymbol{b})}$ | One bit message $\mu \in \{0,1\}$ |
| | | $\boldsymbol{c}_1 := \boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} \mod q$ |
| $\boldsymbol{y} := c_2 - \langle \boldsymbol{c}_1, \boldsymbol{s}\rangle$<br>$= \mu\lfloor \frac{q}{2} \rfloor + \boldsymbol{e}^{\mathrm{T}}\boldsymbol{x}$ | $\xleftarrow{(\boldsymbol{c}_1,c_2)}$ | $c_2 := \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} + \mu\lfloor \frac{q}{2} \rfloor \mod q$ |

**Figure 1** An example of LWE-based public-key encryption.

security, performance, and complexity after a comprehensive study of every submission. As indicated in NIST's document on submission requirement and evaluation criteria, NIST's PQC standardization is exactly but should not be treated as a competition. It is admitted that different schemes may work with different scenarios and different platforms. Rather than drawing the conclusion one scheme is better than another, NIST's competition aims to encourage discussion and evaluation on proposals which forms the decision for NIST's standardization.

## 3 Lattice-based cryptography

### 3.1 PKE/KEM

#### 3.1.1 *An overview*

A large family of lattice-based PKE/KEM is constructed with the LWE problem [1] and its variants including ring learning with errors (ring-LWE or RLWE), module-LWE (MLWE), and learning with rounding (LWR).

**Definition 1** (LWE sample and LWE distribution)**.** An LWE sample is defined as $(\boldsymbol{A}, \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e} \mod q)$ where the $\boldsymbol{A}$ is uniform in $\mathbb{Z}_q^{m\times n}$, the secret $\boldsymbol{s}$ is uniform in $\mathbb{Z}_q^n$ and the error term $\boldsymbol{e}$ is drawn from some distribution $\psi$ over $\mathbb{Z}^n$. The LWE sample satisfies an LWE distribution $A_{\boldsymbol{s},\psi}$.

Observing arbitrarily many $(\boldsymbol{A}, \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e})$, the search-LWE problem is to find the secret vector $\boldsymbol{s}$ while the decision-LWE is to distinguish $A_{\boldsymbol{s},\psi}$ from a uniform distribution with non-negligible advantage. The hardness of LWE problems relies on the worst-case approximate shortest independent vectors problem (SIVP$_\gamma$) and the decisional approximate shortest vector problem (GapSVP$_\gamma$). If we look into LWE from a coding perspective, an LWE sample $(\boldsymbol{A}, \boldsymbol{b} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e})$ can be viewed as a random lattice code with some additive errors. The search version of LWE problem is to find $\boldsymbol{s}$ observing $(\boldsymbol{A}, \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e})$. The vector $\boldsymbol{b}$ is relatively close to a point in the LWE lattice defined as

$$\mathcal{L} = \{\boldsymbol{A}\boldsymbol{s} : \boldsymbol{s} \in \mathbb{Z}_q^n\} + q\mathbb{Z}^m.$$

So the search-LWE problem can be viewed as bounded distance decoding (BDD) of lattice $\mathcal{L}$ in the average case.

A typical LWE-based public-key encryption is given in Figure 1. As the decryption step on Alice's side, the residue error is small enough by choosing appropriate $\psi$ so that the plaintext can be recovered almost surely. The computational complexity of encrypting a single bit in a typical LWE-based PKE requires $O(n^2)$ scalar operations and the sizes of the public key and the ciphertext are $O(n^2)$ and $O(n)$, respectively. The security of LWE-based PKE schemes (e.g., Frodo and Lizard) relies upon the worst-case hardness of the approximate SIVP. These schemes provide good resilience against quantum attacks but suffer from large key sizes and high computational complexity.

The first step towards facilitating LWE using structured lattice is the ring learning with errors problem, as described in Definition 2. Observing arbitrarily many RLWE samples, the search RLWE problem is to recover $s$ and its decision version is to distinguish an RLWE distribution from a uniform one.

Parameters are $R, n, q$; error distribution $\chi$ on $R_q := R/qR$

| **Alice** | | **Bob** |
|---|---|---|
| $a \overset{\$}{\leftarrow} R_q$ | | |
| $s, e \leftarrow \chi$ | | $t, e', e'' \leftarrow \chi$ |
| $b := a \cdot s + e \in R_q$ | $\xrightarrow{(a,b)}$ | $n$-bit secret $\mu \in \{0,1\}^n$ |
| | | $c_1 := a \cdot t + e' \in R_q$ |
| | | shared secret= hash($\mu$) |
| $\boldsymbol{y} := c_2 - c_1 \cdot s$ | $\xleftarrow{(c_1,c_2)}$ | $c_2 := b \cdot t + e'' + \mu\lfloor \frac{q}{2} \rfloor \in R_q$ |
| $= \mu\lfloor \frac{q}{2} \rfloor + e \cdot t - e' \cdot s + e''$ | | |
| shared secret= hash($\mu$) | | |

**Figure 2** An example of RLWE-based public-key KEM.

The cyclotomic ring admits quite fast and elegant ring operations using an FFT-like technique, i.e., number theoretic transform (NTT), and consequentially enhances the efficiency and compactness in the application.

**Definition 2** (RLWE sample and RLWE distribution)**.** Define a cyclotomic ring $R := \frac{\mathbb{Z}[x]}{1+x^n}$ and denote by $R_q = R/qR$ a quotient ring with an integer prime modulus $q \equiv 1 \mod 2n$. Let $\cdot$ be polynomial multiplication. For a secret $s \in R_q$ and some error distribution $\psi$ over $R$, we derive a RLWE sample $(a, a \cdot s + e) \in R_q \times R_q$ by drawing $a$ from $R_q$ uniformly at random and drawing $e$ from $\psi$. A RLWE sample follows the RLWE distribution $A_{s,\psi}$.

Compared with Definition 2, a more rigorous definition of RLWE is given in [2] elaborating a RLWE sample as $(a,\ a \cdot s + e \mod qR^\vee)$ where $a$ is uniformly drawn from the ring of integers $R$ modulo $q$, $s$ is drawn from the dual ring $R_q^\vee$ and $e$ adheres to some distribution $\psi$ over $K_\mathbb{R}$ ($K$ is the cyclotomic number field associated with the ring $R$ and $K_\mathbb{R}$ is the tensor product of $K$ and $\mathbb{R}$). Compact and efficient public-key cryptosystems based on RLWE are designed in [3] though they are a bit far from real-world protocols due to the choice of $\psi$ and the usages of fractional ideal $R^\vee$ (unless in the 2-power cyclotomic setting where $R^\vee = \frac{1}{n}R$ is a scaling of $R$). Peikert took one more step to develop these cryptosystems into "drop-in" components as Internet protocols [4]. In his work, the $R^\vee$ is translated to $R$ cautiously without distorting the canonical geometry and a reconciliation method is employed for key agreement. A consensus technique similar to reconciliation can be found in another PQC candidate key consensus from lattice (KCL). Other NIST candidates, e.g., NewHope, HILA5, and lattice-based cryptography (LAC), adopt error-correcting codes for key agreement/encryption.

A RLWE-based KEM scheme is given in Figure 2 where the hash($\cdot$) function yields an $n$-bit shared secret given an $n$-bit secret $\mu$ as input. To encrypt an $n$-bit message or to share an $n$-bit secret, the public key and ciphertext size is as large as $O(n)$. Moreover, the polynomial multiplication $(a \cdot s)$ in the ring setting can be performed in $O(n \log n)$ scalar operations using NTT. Although none of the RLWE-based PKE/KEMs are selected as NIST's third round finalists in the fierce competition, we cannot deny their great potential in future research and standardization due to their attractive features especially the efficiency.

### 3.1.2 *Error correction coding*

Since we expect to give some flavor of coding in PQC, it is worthwhile revisiting the error correction problem in lattice based PKE/KEM. We found error-correction codes in some PKE/KEM proposals, e.g., repetition codes in NewHope, BCH codes in LAC, and another polynomial codes named XEf in Round5. This "unusual design", remarked in NIST's report, is employed to fix the decryption failures as a result of a minor residue error term after decryption. Some attacks using the decryption failure rate (DFR) put some schemes under threat because the residue error term might be related to some ciphertexts and even some specific secret terms. Once the underlying relation is learned by attackers, the security will be impacted. There are already attacks of this type. For example, the "failure boosting" and "directional failure boosting", proposed by D'Anvers et al. [5,6] can be used to find some ciphertexts which are more likely to trigger decryption failures. They also verified these attacks on some basic versions of ring/module-LWE/LWR PKE/KEMs with comparable parameterization to NIST candidates, e.g., NTRUEncrypt, KYBER, SABER. The security is impacted assuming unlimited decryption queries are allowed. Another attack proposed by Guo et al. [7] also exploits the relation between some "weak" ciphertexts and some secret keys of certain Hamming weight pattern.

Besides the error-correction codes used in NIST submissions, some academic researches [8] can be found to apply modern error correcting codes and its soft-decision decoding to lattice cryptography, e.g., low-density parity-check (LDPC) codes. Normally, soft-decision decoding algorithms give much better error-correcting capability, and to the best of our knowledge, they assume independent channel models. It is common in communication systems whereas in RLWE-based cryptography it is not the case. As in Figure 2, the residue error term $e{\cdot}t{-}e'{\cdot}s{+}e''$ has correlated coordinates due to polynomial multiplications over $R_q$. As in Figure 2, the ring structure introduces dependency between the coordinates of the residue error. It becomes obvious if we present the residue error term $e \cdot t - e' \cdot s + e''$ in vector format, i.e.,

$$\begin{pmatrix} e & -e_{n-1} & \cdots & -e_1 \\ e_1 & e_0 & \cdots & -e_2 \\ \vdots & \vdots & \ddots & \vdots \\ e_{n-1} & e_{n-2} & \cdots & e_0 \end{pmatrix} \boldsymbol{t} - \begin{pmatrix} s_0 & -s_{n-1} & \cdots & -s_1 \\ s_1 & s_0 & \cdots & -s_2 \\ \vdots & \vdots & \ddots & \vdots \\ s_{n-1} & s_{n-2} & \cdots & s_0 \end{pmatrix} \boldsymbol{e}' + \boldsymbol{e}''.$$

In [8], this dependency is assumed to be negligible. However, the impact of the dependency is analyzed by D'Anvers et al. [9] showing that the DFR will be underestimated (the security will be overestimated consequently) when error-correcting codes are used. They also proposed a relaxed independence assumption by which they gave a new DFR estimation for LAC. A drawback of this relaxed assumption is that it works with schemes like LAC which draws secret terms and error terms from centered binomial over $\{-1, 0, 1\}$. In the case when a genuine discrete Gaussian is used or when the binomial distribution is wide, their method becomes computationally infeasible.

In [10], polar codes are employed for error correction in RLWE-based PKE. They address the error dependency issue using canonical embedding. To be specific, the polynomial multiplication of the ring elements is converted to coordinate-wise multiplication of vectors under canonical bedding, leaving the residue error term with identically independent coordinates. Moreover, some knowledge about the residue error is actually known by the decoder in RLWE-based PKE. By analogy with channel coding, they name this knowledge as channel state information (CSI) which can be exploited to improve decoding performance. Compared with the BCH code and LDPC code, polar code is more friendly to constant-time implementation because both encoding and decoding can be realized by a butterfly circuit with quasi-linear complexity $O(N \log N)$.

### 3.2 Signatures

In the early stage of lattice-based cryptography, signature schemes following the Goldreich-Goldwasser-Halevi (GGH) strategy suffer certain vulnerability that each signature leaks information on the signer's secret key (i.e., the signer's secret lattice basis), and this property has been exploited in Nguyen and Regev's attack of "learning the parallelotope". Specifically, the shape of the parallelotope defined by the secret basis can be learned by using a number of signature samples to perform blind source separation. The NIST digital signature schemes based on LBC fall into two categories: (a) Gentry-Peikert-Vaikuntanathan (GPV) framework with trapdoor sampling and (b) signatures using Fiat-Shamir transform.

Gentry et al. [11] first constructed a type of "trapdoor" cryptographic tools based on the (inhomogeneous) short integer solution (SIS) which can be reduced to the standard worst-case lattice problems. Such trapdoor construction immediately finds its applications in cryptography including digital signatures. Before giving an instantiation of GPV signature using SIS, we first review the SIS as described in Definition 3.

**Definition 3** (SIS)**.** Given a uniformly random matrix $\boldsymbol{A} \in \mathbb{Z}_q^{n\times m}$, find a nonzero integer vector $\boldsymbol{z} \in \mathbb{Z}^m$ of norm $\|\boldsymbol{z}\| \leqslant \beta$ such that

$$f_{\boldsymbol{A}}(\boldsymbol{z}) := \boldsymbol{A}\boldsymbol{z} = \boldsymbol{0} \in \mathbb{Z}_q^n.$$

For $m = \mathrm{poly}(n)$, $\beta > 0$ and $q \geqslant \beta \cdot \mathrm{poly}(n)$, solving the above SIS problem is as hard as $\mathrm{GapSVP}_\gamma$ and $\mathrm{SIVP}_\gamma$. Assuming the hardness of SIS, function $f_{\boldsymbol{A}}(\boldsymbol{z})$ is collision resistant (or one-way) because the equation $f_{\boldsymbol{A}}(\boldsymbol{z}') = f_{\boldsymbol{A}}(\boldsymbol{z}'')$ for distinct $\boldsymbol{z}'$ and $\boldsymbol{z}''$ implies a solution to SIS.

| Parameters are $m, n, q, s$; $\eta(\mathcal{L})$: smoothing parameter of $\mathcal{L}^{\perp}(\boldsymbol{A})$ | | |
|---|---|---|
| **Signer** | | **Verifier** |
| Public key: $\boldsymbol{A} \xleftarrow{\$} \mathbb{Z}_q^{n\times m}$. | | |
| Secret key: $\boldsymbol{S} \in \mathbb{Z}^{m\times m}$ be a good basis of $\mathcal{L}^{\perp}(\boldsymbol{A})$ | | |
| s.t. $\boldsymbol{AS} = \boldsymbol{0} \mod q$. | | |
| Message: $\boldsymbol{m}$ | | |
| $\boldsymbol{u} := \text{hash}(\boldsymbol{m})$ | | |
| Signature: $\boldsymbol{y} \leftarrow D_{\mathcal{L}_{\boldsymbol{u}}^{\perp}(\boldsymbol{A}),s}$ for $s \geqslant \eta(\mathcal{L})$ | $\xrightarrow{(\boldsymbol{A},\boldsymbol{y})}$ | If $\boldsymbol{Ay} = \boldsymbol{u}$ and $\|\boldsymbol{y}\| \leqslant s\sqrt{m}$, then accept; |
| s.t. $\|\boldsymbol{y}\| \leqslant s\sqrt{m}$. | | otherwise, invalid signature. |

**Figure 3** A GPV instantiation using SIS.

Then we define a $q$-array $m$-dimensional integer lattice and its coset as

$$\mathcal{L}^{\perp}(\boldsymbol{A}) := \left\{\boldsymbol{z} \in \mathbb{Z}^m : \boldsymbol{Az} = \boldsymbol{0} \in \mathbb{Z}_q^n\right\},$$

and

$$\mathcal{L}_{\boldsymbol{u}}^{\perp}(\boldsymbol{A}) := \left\{\boldsymbol{z} \in \mathbb{Z}^m : \boldsymbol{Az} = \boldsymbol{u} \in \mathbb{Z}_q^n\right\}, \text{ respectively.}$$

A concrete instantiation of GPV signature using $q$-array integer lattice and preimage sampling is given in Figure 3.

- At key generation step, a signer has a uniformly random public matrix $\boldsymbol{A} \in \mathbb{Z}_q^{n\times m}$ and a secret matrix $\boldsymbol{S} \in \mathbb{Z}_q^{m\times m}$ such that $\boldsymbol{AS} = \boldsymbol{0} \mod q$. The secret $\boldsymbol{S}$ is said to be a good basis of a lattice $\mathcal{L}^{\perp}(\boldsymbol{A})$ with small coefficients and good orthogonality. A specific way to construct $\boldsymbol{A}$, $\boldsymbol{S}$ called "Gadget" can be found in [12].
- The signer hash the message $\boldsymbol{m}$ into a vector $\boldsymbol{u} \in \mathbb{Z}_q^n$. A not necessarily short solution $\boldsymbol{x}$ to $\boldsymbol{Ax} = \boldsymbol{u}$ can be derived using Gaussian elimination. Using the good basis $\boldsymbol{S}$, the signer derives the signature $\boldsymbol{y}$ by sampling from a lattice Gaussian distribution (LGD) $D_{\mathcal{L}_{\boldsymbol{u}}^{\perp}(\boldsymbol{A}),s}$ over a coset $\mathcal{L}_{\boldsymbol{u}}^{\perp}(\boldsymbol{A})$ of $\mathcal{L}^{\perp}(\boldsymbol{A})$ with standard deviation $s$ larger than the smoothing parameter $\eta(\mathcal{L})$. The signer accepts the signature $\boldsymbol{y}$ if $\|\boldsymbol{y}\| \leqslant s\sqrt{m}$.
- At the verifier's side, a signature $\boldsymbol{y}$ is considered to be valid if $\boldsymbol{Ay} = \boldsymbol{u}$ and $\|\boldsymbol{y}\| \leqslant s\sqrt{m}$. Otherwise, $\boldsymbol{y}$ is invalid.

The above hash-and-sign signature is unforgeable because an adversary who only knows $\boldsymbol{A}$ and $\boldsymbol{u}$ cannot recover a valid $\boldsymbol{y}$ assuming the hardness of SIS. It is notable that in the above GPV framework, a trapdoor sampler that can yield a smaller standard deviation enables stronger security while Klein's and Peikert's algorithms do not work with arbitrarily small standard deviations. Another lattice sampling algorithm using Monte Carlo Markov Chain supports arbitrary standard deviations, therefore, enjoys higher security at the cost of longer running time [13, 14].

Falcon adapts NTRU lattice to the GPV framework while another third-round signature candidate CRYSTALS-DILITHIUM employs the Fiat-Shamir transform and rejection sampling. Its security relies on the hardness of MLWE and module-SIS. Digital signatures using Fiat-Shamir's technique was first proposed by Lyubashevsky [15] and the rejection sampling was later adapted and improved in [16] and the signature scheme BLISS [17]. As a high-level description, the signer has a secret key $\boldsymbol{S} \in \mathbb{Z}_q^{m\times n}$ with small coefficients and a public key consists of $\boldsymbol{A} \in \mathbb{Z}_q^{n\times m}$ and the other matrix $\boldsymbol{T} = \boldsymbol{AS}$. The signer firstly picks a short vector $\boldsymbol{y}$ from a normal distribution $\mathcal{N}(0, \sigma^2\boldsymbol{I}^{m\times m})$, calculates $\boldsymbol{c} = \boldsymbol{Ay}$ and then computes $\boldsymbol{z} = \boldsymbol{Sc} + \boldsymbol{y}$ which adheres to $\mathcal{N}(\boldsymbol{Sc}, \sigma^2\boldsymbol{I}^{m\times m})$. A rejection sampling is applied to shape the distribution of $\boldsymbol{z}$ by a shift of $-\boldsymbol{Sc}$. In this manner, the signature is statistically independent of the secret $\boldsymbol{S}$, therefore, leaks almost no information about $\boldsymbol{S}$. The singer outputs $(\boldsymbol{z}, \boldsymbol{c})$ as the signature and the verification is done by checking if the norm of $\boldsymbol{z}$ is short and if $\boldsymbol{Ay} = \boldsymbol{Az} - \boldsymbol{Tc}$.

**Remark 1.** For some lattice-based cryptosystems (and some code-based ones in the sequel), we observe some links between encryption and channel coding, signature and source coding, identity-based encryption (IBE) and joint source-channel coding.

- Taking LWE for example, encryption is to map the plaintext to a codeword of a random linear code disrupted by the noise of a certain distribution. The search-LWE problem is similar to the maximum-likelihood decoding problem in the Euclidean metric. Another example is that the ISIS problem can be

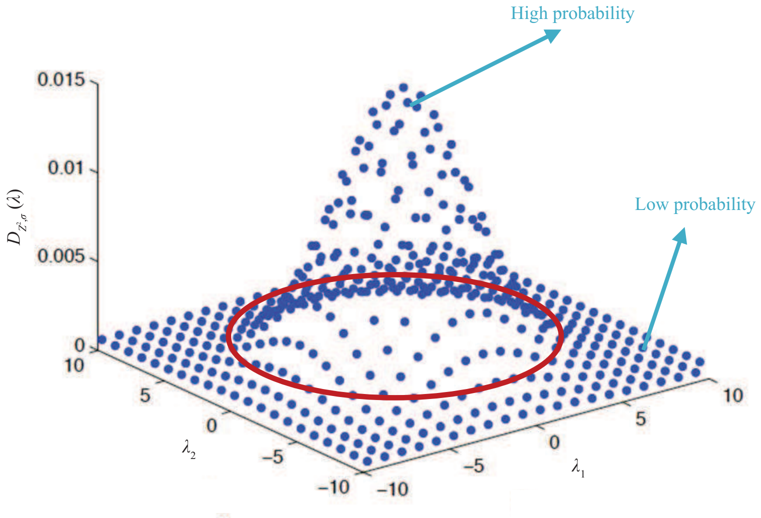

**Figure 4** (Color online) Lattice Gaussian distribution.

seen as syndrome decoding. Similar links can be found in code-based encryption schemes in Hamming or rank metric in the sequel.

- For lattice-based digital signatures, the signature is either drawn from a certain distribution (e.g., discrete Gaussian over a lattice coset) or randomized to hide the secret (e.g., Lyubashevsky's scheme without trapdoor). We can consider it as producing an unbiased distribution while source coding is on the contrary to remove the randomness of a certain unbiased distribution. This relation can be reflected by the Knuth-Yao sampling and Huffman source coding techniques both of which employ a binary tree but for opposite purposes.
- In some sense, we can found a link between IBE from GPV [11] and joint source-channel coding. The GPV IBE can be seen as the "dual" version of Regev's LWE cryptosystem. The master public key of the authority is a matrix $\boldsymbol{A}$ and its trapdoor is the master secret key. The public key of a user is a concatenation of $\boldsymbol{A}$ and a hash $\boldsymbol{u}$ of its unique identity (e.g., email address). The user secret key $\boldsymbol{x}$ is a preimage of $\boldsymbol{u} = \boldsymbol{A}\boldsymbol{x}$. As introduced earlier in this section, vector $\boldsymbol{x}$ is derived by trapdoor sampling from distribution $D_{\mathcal{L}_{\boldsymbol{u}}^{\perp}(\boldsymbol{A}),s}$ which can be seen as the "reverse process" of source coding. What follows is the standard encryption and decryption of an LWE-based scheme as is analogous to the channel coding problem.

### 3.3 Lattice Gaussian sampling

**Definition 4** (LGD)**.** Let $\Lambda$ be an $n$-dimensional lattice in $\mathbb{R}^n$ and let $\rho_{\boldsymbol{c},s}(\boldsymbol{x}) = \exp(-\pi\|\boldsymbol{x}-\boldsymbol{c}\|^2/s^2)$ be a spherical Gaussian function centered at $\boldsymbol{x} = \boldsymbol{c}$ with Gaussian parameter $s$. The Gaussian distribution over a lattice $\Lambda$ is defined as

$$\forall \boldsymbol{x} \in \Lambda,\ D_{\Lambda,\boldsymbol{c},s}(\boldsymbol{x}) = \rho_{\boldsymbol{c},s}(\boldsymbol{x})/\rho_{\boldsymbol{c},s}(\Lambda),$$

where $\rho_{\boldsymbol{c},s}(\Lambda) = \sum_{z\in\Lambda} \rho_{\boldsymbol{c},s}(\boldsymbol{z})$. We can use $D_{\Lambda,s}(\boldsymbol{x})$ for abbreviation if $\boldsymbol{c} = 0$.

Lattice Gaussian sampling (LGS) has become a fundamental tool in LBC. Typically, the task of LGS is to draw vectors from LGD, which is a discretized Gaussian distribution over a lattice. In LGD, every lattice point in the lattice is assigned with a probability, and the lattice point closer to the center of the distribution naturally corresponds to a larger sampling probability. Intuitively, the probabilities of lattice points are controlled by the standard deviation of a LGD, where a small standard deviation accounts for a sharp distribution and vice-versa. An example of a LGD over a $\mathbb{Z}^2$-lattice is illustrated in Figure 4.

LGD is frequently used in lattice cryptosystems especially lattice signatures. A special case of LGS is integer Gaussian sampling which is deemed as the fundamental building block of LBC. However, inappropriate LGS sometimes accounts for some efficiency and security issues. Many lattice-based proposals managed to avoid the lattice/integer Gaussian sampling operations, e.g., CRYSTALS-DILITHIUM uses uniform distribution instead, and SABER and KYBER use binomial distribution instead. Nonetheless, LGS, especially integer Gaussian sampling, cannot be completely avoided.

Well-studied LGS algorithms include Klein's sampler which is a randomized version of Babai's nearest plane algorithm and Peikert's parallelizable sampler [18]. Falcon, a NIST' third round finalist, employs a fast Fourier orthogonalization method [19] to construct a trapdoor sampler for NTRU lattices which combines the quality of Klein's sampler and the efficiency of Peikert's. A "Gadget"-based trapdoor was devised in [12] which enables efficient trapdoor sampling using "Gadget".

Another line of research focuses on integer Gaussian sampling though it is indeed an aged topic. Traditional solutions include the cumulative distribution table (CDT) sampler, the Knuth-Yao sampler, and the discrete Ziggurat sampler. Recent researches [20,21] suggested an integer Gaussian sampler works in a two-phase manner: a base sampler (e.g., Knuth-Yao, CDT) produces and stores plenty of samples from an integer Gaussian distribution of small parameters; the target distribution is shaped with the stored samples by a constant-time convolution/expansion method.

**Remark 2.** From a coding perspective, Knuth-Yao sampler employs a binary searching tree analogous to the Huffman coding tree where a parent node splits into two child nodes according to the occurrence frequency. The difference between the two is that the binary tree constructed by Knuth-Yao is for distribution generation while that for Huffman coding is for lossless data compression. This implies that sampling can be viewed as a reverse of data compression and related work about sampling from the integers using polar source coding was proposed in [22]. It is featured with asymptotically information-theoretic optimality: to produce the desired distribution with the optimal randomness.

## 4 Code-based cryptography

Code-based cryptography employs coding theory and the hard decoding problems to build primitives such as encryption schemes, one-way functions, digital signatures, key exchange[2]. Generally speaking, code-based KEM requires more bandwidth and offers slower key generation in comparison with lattice-based ones. Driven by the NIST process, code-based cryptography has overcome the limitation of the classic McEliece cryptosystem and efficiency enhancements have been made to reduce the key size and to accelerate the key generation using structured codes (e.g., cyclic codes). Moreover, a specialty of code-based cryptography is the strong confidence in security. While lattice-based cryptography emerged in the last decade, code-based cryptography originated from the scheme proposed by McEliece in 1978 [23] and its "dual" version was proposed by Niederreiter in 1986 [24]. Efforts have been put into the cryptanalysis of McElice cryptosystems with little if any significant and practical outcomes for over 40 years.

The idea of McEliece's scheme is to use an apparently random generator matrix of an error-correcting code (a random binary Goppa code) as a public key to encrypt a message, where $t$ random bit errors are also added to it. The code is chosen to be able to correct up to $t$ errors. Thus, legitimate users who know a fast decoding algorithm for the code as a private key can recover the plaintext. The security of McEliece's scheme relies on the following two computational assumptions:

- Decoding a random linear code in the fixed-error model is hard on average;
- The generator matrix (public key) is hard to distinguish from a random matrix.

The first problem is proved to be nondeterministic polynomial time (NP)-complete [25] and is believed to be hard on average, while the second problem is more open. The McEliece encryption scheme is described as follows in Figure 5.

Among NIST third-round finalists, classic McEliece is the only code-based scheme KEM. It is the dual version of the original McEliece with moderate efficiency improvement but no security compromise. The major changes include the migration to Niederreiter's dual variant and some refinements of parameters in order to keep up with the increased computing power. Due to its early invention, classic McEliece is the most researched one among all NIST candidates and thus is better understood compared with other schemes. Furthermore, classic McEliece enables efficient and straightforward conversion of one-way-CPA PKE into IND-CCA2 KEM. Besides, it can be configured to match all five NIST security levels. The main drawback is that it also has extremely large public key sizes, ranging from 250 KB for NIST security level 1 up to 1.3 MB for NIST security level 5.

[2] Linear codes are also used to construct secret sharing scheme (SSS), e.g., Massey's scheme. Moreover, there is a natural correspondence between linear codes and linear secure secret sharing which is the central building block for information-theoretically secure cryptographic primitives such as multiparty computation (MPC). The details of SSS and MPC are beyond the scope of this work.

Parameters: $k, n, t$; $\mathcal{S}_t^n(\mathbb{F}_2) := \{\boldsymbol{x} \in \mathbb{F}_2^n : \ \mathrm{wt}(\boldsymbol{x}) = t\}$; $\mathrm{wt}(\cdot)$: Hamming weight.

| **Alice** | | **Bob** |
|---|---|---|
| $\boldsymbol{G} \in \mathbb{F}_2^{k\times n}$: generator matrix of code $\mathcal{C}$ | | |
| $\boldsymbol{S} \in \mathbb{F}_2^{k\times k}$: a non-singular matrix | | |
| $\boldsymbol{P} \in \mathbb{F}_2^{n\times n}$: a permutation matrix | | |
| $\boldsymbol{A} = \boldsymbol{SGP}$ | $\xrightarrow{\boldsymbol{A}}$ | $\boldsymbol{m} \in \mathbb{F}_2^k$ |
| | | $\boldsymbol{e} \overset{\$}{\leftarrow} \mathcal{S}_t^n(\mathbb{F}_2)$ |
| $\boldsymbol{m}' = \mathcal{C}.\mathrm{Decode}(\boldsymbol{y}\boldsymbol{P}^{\mathrm{T}})$ | $\xleftarrow{(\boldsymbol{y})}$ | $\boldsymbol{y} = \boldsymbol{mA} + \boldsymbol{e}$ |
| $\boldsymbol{m} = \boldsymbol{m}'\boldsymbol{S}^{-1}$ | | |

**Figure 5** Illustration of McEliece public-key encryption scheme.

### 4.1 Quasi-cyclic code-based public-key encryption/KEM

The main reason for the extremely large public key size of classic McEliece is that one has to store the whole generator matrix of the code. To reduce the key size, one approach is to choose a proper code family such that the generator matrix can be expressed more compactly. In [26], it was proposed to use quasi-cyclic codes. A code spanned by a block-circulant matrix is called quasi-cyclic. A circulant matrix is a square matrix in which all row vectors are composed of the same elements and each row vector is rotated one element to the right relative to the preceding row vector. A block-circulant matrix is formed of blocks of circulant square matrices. The generator matrix of a quasi-cyclic code is completely defined by its first row, and thus, the public key size of quasi-cyclic code-based cryptographic schemes can be greatly reduced. Similar to the original McEliece scheme, the security proof of the quasi-cyclic code-based McEliece scheme relies on the following two computational assumptions:

- Generic decoding of a random quasi-cyclic code is hard;
- The generator matrix (public key) is hard to distinguish from a random block-circulant matrix.

Since quasi-cyclic codes are more structured, the code family must be chosen carefully, otherwise, some types of attacks are possible.

Recently, quasi-cyclic moderate density parity-check (QC-MDPC) codes [27] have attracted research interests in cryptography society. MDPC codes are variants of the well-known LDPC codes with moderately sparse parity-check matrices. Specifically, the rows of the parity-check matrix of an MDPC code have length $n$ and Hamming weight of order $\sqrt{n}$. The denser parity-check matrices deteriorate the error-correcting performance. However, in cryptography, we are not interested in correcting as many errors as possible, but only a number that is enough for the schemes. The most important reason for considering moderately sparse parity-check matrices is to avoid certain types of attacks. When using LDPC codes, low weight parity-check rows can be seen as dual codewords. If one searches for low-weight dual codewords to build a sparse parity-check matrix, which is for sure easily decodable, the scheme may be effectively attacked. It deserves to mention that Guo et al. [28] also presented an attack against the QC-MDPC McEliece encryption scheme.

Among NIST third-round candidates, two quasi-cyclic code-based schemes are selected as alternates, i.e., Hamming quasi-cyclic (HQC) and bit flipping key encapsulation (BIKE). HQC is based on BCH code, while BIKE is based on QC-MDPC codes. Both schemes have parameters that target NIST levels 1, 3, and 5 with much smaller public key sizes compared with classic McEliece. The public-key encryption algorithm of HQC is illustrated in Figure 6, where matrix $\boldsymbol{G}$ is public known, the public key is $(\boldsymbol{h}, \boldsymbol{s})$ and the secret key is $(\boldsymbol{x}, \boldsymbol{y})$.

### 4.2 Rank metric-based public-key encryption/KEM

According to the underlying hard problems, lattice-based and code-based cryptography can be interpreted as distance-based cryptography. The former is a Euclidean-based one while the latter is Hamming-based. Different types of metrics have different properties, resulting in their respective advantages and drawbacks. Rank metric codes are a special type of linear error-correcting codes that use the rank metric instead of the Hamming metric. Loo-Keng Hua first introduced rank metric in 1951 and Delsarte introduced the metric, rank distance, and constructed the optimal matrix codes in bilinear representation. In 1985,

Parameters: $k, n, w$; $\mathcal{S}_w^n(\mathbb{F}_2) := \{\boldsymbol{x} \in \mathbb{F}_2^n : \ \mathrm{wt}(\boldsymbol{x}) = w\}$; $\mathrm{wt}(\cdot)$: Hamming weight;
$\boldsymbol{a} \odot \boldsymbol{b}$: vector $\boldsymbol{a}$ multiplied by the circulant matrix generated by vector $\boldsymbol{b}$.

| **Alice** | | **Bob** |
|---|---|---|
| $\boldsymbol{G} \in \mathbb{F}_2^{k \times n}$: generator matrix of code $\mathcal{C}$ | | |
| $\boldsymbol{x}, \boldsymbol{y} \xleftarrow{\$} \mathcal{S}_w^n(\mathbb{F}_2)$ | | |
| $\boldsymbol{h} \xleftarrow{\$} \mathbb{F}_2^n$, $\boldsymbol{s} = \boldsymbol{x} + \boldsymbol{h} \odot \boldsymbol{y}$ | $\xrightarrow{(\boldsymbol{G},\boldsymbol{h},\boldsymbol{s})}$ | $\boldsymbol{m} \in \mathbb{F}_2^k$ |
| | | $\boldsymbol{r}_1, \boldsymbol{r}_2 \xleftarrow{\$} \mathcal{S}_w^n(\mathbb{F}_2)$, $\boldsymbol{e} \xleftarrow{\$} \mathcal{S}_w^n(\mathbb{F}_2)$ |
| | | $\boldsymbol{u} = \boldsymbol{r}_1 + \boldsymbol{h} \odot \boldsymbol{r}_2$ |
| $\boldsymbol{m} = \mathcal{C}.\mathrm{Decode}(\boldsymbol{v} - \boldsymbol{u} \odot \boldsymbol{y})$ | $\xleftarrow{(\boldsymbol{u},\boldsymbol{v})}$ | $\boldsymbol{v} = \boldsymbol{m}\boldsymbol{G} + \boldsymbol{s} \odot \boldsymbol{r}_2 + \boldsymbol{e}$ |

**Figure 6** Illustration of HQC PKE.

Parameters: $n$, $k$, $m$, $w$, and $P \in \mathbb{F}_q[X]$; $\mathcal{S}_w^n(\mathbb{F}_{q^m}) = \{\boldsymbol{x} \in \mathbb{F}_{q^m}^n : \ \mathrm{rank}(\boldsymbol{x}) = w\}$; $\mathrm{rank}(\cdot)$: rank weight;
$\boldsymbol{a} \cdot \boldsymbol{b} := (\sum_{i=0}^n a_i X^i)(\sum_{i=0}^n b_i X^i)$.

| **Alice** | | **Bob** |
|---|---|---|
| $\boldsymbol{G} \in \mathbb{F}_{q^m}^{k \times n}$: generator matrix of code $\mathcal{C}$ | | |
| $(\boldsymbol{x}, \boldsymbol{y}) \xleftarrow{\$} \mathcal{S}_w^{2n}(\mathbb{F}_{q^m})$ | | |
| $\boldsymbol{h} \xleftarrow{\$} \mathbb{F}_{q^m}^n$, $\boldsymbol{s} = \boldsymbol{x} + \boldsymbol{h} \cdot \boldsymbol{y} \mod P$ | $\xrightarrow{(\boldsymbol{g},\boldsymbol{h},\boldsymbol{s})}$ | |
| | | $(\boldsymbol{r}_1, \boldsymbol{r}_2, \boldsymbol{r}_3) \xleftarrow{\$} \mathcal{S}_w^{3n}(\mathbb{F}_{q^m})$ |
| | | $\boldsymbol{u} = \boldsymbol{r}_1 + \boldsymbol{h} \cdot \boldsymbol{r}_2 \mod P$ |
| $\boldsymbol{m} \leftarrow \mathcal{C}.\mathrm{Decode}(\boldsymbol{v} - \boldsymbol{u} \cdot \boldsymbol{y})$ | $\xleftarrow{(\boldsymbol{u},\boldsymbol{v})}$ | $\boldsymbol{v} = \boldsymbol{m}\boldsymbol{G} + \boldsymbol{s} \cdot \boldsymbol{r}_2 + \boldsymbol{r}_3 \mod P$ |

**Figure 7** Illustration of RQC PKE.

Gabidulin proposed vector-form rank codes as well as efficient encoding and decoding algorithms. Since then, rank codes have been used in many applications, such as communication as well as cryptography.

Let $\boldsymbol{u} = \{u_1, \ldots, u_n\}$, $\boldsymbol{v} = \{v_1, \ldots, v_n\} \in \mathbb{F}_{q^m}^n$ be two length-$n$ vectors in the vector space of $\mathbb{F}_{q^m}^n$, where $q$ is a prime, $\mathbb{F}_{q^m}$ is the finite field with $q^m$ elements and $m$ is a positive integer. The rank weight $\mathrm{rank}(\boldsymbol{u})$ of $\boldsymbol{u}$ is defined as the dimension of the $\mathbb{F}_q$-subspace generated by $\{u_1, \ldots, u_n\}$. The rank distance $\mathrm{rd}(\boldsymbol{u}, \boldsymbol{v})$ between $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined as $\mathrm{rank}(\boldsymbol{u} - \boldsymbol{v})$. A $[n, k]$ rank code $\mathcal{C}$ over $\mathbb{F}_{q^m}$ is a subspace of dimension $k$ of $\mathbb{F}_{q^m}^n$ embedded with rank metric. When using rank codes in cryptography, the main problem is the generalization of the generic decoding problem with Hamming distance in the case of rank distance. Specifically, let $\boldsymbol{H}$ be an $((n-k) \times n)$ matrix over $\mathbb{F}_{q^m}$ with $k \leqslant n$, $s \in \mathbb{F}_{q^m}^n$ and $r$ an integer. The problem is to find $x$ such that $\mathrm{rank}(x) = r$ and $Hx^t = s$. This problem has been proved to be NP-hard with a randomized reduction [29].

Rank quasi-cyclic (RQC), a NIST 2nd-round candidate, is based on Gabidulin codes, which has an even shorter public key size compared with HQC and BIKE. Unlike McEliece whose security relies on the hardness of syndrome decoding problem and Goppa code distinguishing problem, the security of RQC relies solely on the decisional version of syndrome decoding of quasi-cyclic codes. Moreover, the family of Gabidulin codes has zero probability of decoding failure when using deterministic decoding. The main drawback of RQC is that the security analysis under algebraic attacks deserves more time to mature [30]. Besides, the decoding complexity of Gabidulin codes, approximately $O(n^2)$, is much higher than its Hamming metric-based counterparts. The public-key encryption algorithm of RQC is illustrated in Figure 7 where $\boldsymbol{G}$ is the generator matrix of a Gabidulin code $\mathcal{C}$ and can be represented by the first row $\boldsymbol{g}$ vector. The public key is $(\boldsymbol{g}, \boldsymbol{h}, \boldsymbol{s})$ and the secret key is $(\boldsymbol{x}, \boldsymbol{y})$.

### 4.3 Code-based signatures

While code-based cryptography provides promising PKE/KEM solutions to be standardized in NIST competition, code-based signatures are not in an advantageous position. Attempts to devise code-based signature schemes to meet the efficiency and security requirements were unsuccessful and none of the existing code-based signatures in the literature is now under consideration for standardisation in the NIST competition. Similar to lattice signatures, code-based signatures fall into two categories: hash-and-sign paradigm using trapdoor and Fiat-Shamir paradigm.

By the first approach, one hashes the message to a syndrome and finds its preimage as the signature subject to some limitations (e.g., Hamming distance, rank distance). Roughly speaking, a one-way function is devised based on the hardness of syndrome decoding of random linear codes. The preimage can be derived using the trapdoor but is hard to recover from the public key and signature due to the one-wayness. Typical signatures of the first type include CFS [31], WAVE [32], and RankSign [33]. CFS exploits the decoding capability of high rate Goppa codes for which a non-negligible fraction of the syndromes can be decoded to the nearest codeword. However, the CFS is unpractical in terms of efficiency and security. One drawback of CFS is that the performance scales poorly with the security. To be specific, achieving 128 bits of classical security requires a public key of several gigabytes and a signature generation of several seconds. Besides, for high rate Goppa codes, the public key was found to be distinguishable from random matrix [34]. Similar distinguisher also exists for the rank-based signature RankSign [35]. The hidden structure as the trapdoor has its pros and cons. On the one hand, it enables efficient preimage calculation, but on the other hand, there might exist unknown structural attacks to recover the hidden structure and break the scheme.

The other signature paradigm using Fiat-Shamir transform and zero-knowledge identification protocol circumvents the above problem. It does not employ any hidden structure because the public key is exactly the parity-check matrix of the underlying linear code and no syndrome decoding is involved. Lattice signature schemes of this type proposed by Lyubashevsky are introduced in Subsection 3.2 where the signature is randomized in Euclidean metric to hide the secret key. When this approach is adapted to code-based schemes, randomizing the signatures in Hamming and rank metric is nontrivial because one has to consider the whole codeword rather than independent coordinates of a lattice vector. Persichetti [36] proposed a one-time signature. However, this scheme is afterward shown to suffer from statistical attacks and the secret key can be recovered using a single signature [37]. Other signatures of this type include NIST 1st-round candidate RaCoSS [38] and Durandal [39], whereas secret information leakage is a common issue to solve.

## 5 The interplay between lattices and codes

In addition to the aforementioned KEM, coding theory has a lot to contribute to lattice cryptography. For instance, Regev's LWE problem [1], as an average-case hard problem of lattices, can be equivalently presented as the problem of decoding random linear codes. Thus the complexity (security) reduction from one side implies the security guarantees from the other side. In addition, the lattice-reduction-aided cryptanalysis technique also benefits a lot by refining the algorithm from the perspective of structured codes. This section inspects these two fronts with details.

### 5.1 On designing hard problems: LWE and random codes

Lattice problems have become popular in quantum resistant cryptographic schemes. The classic computational problems on lattices are the following.

- Shortest vector problem (SVP). Given a lattice $\Lambda$, find the shortest nonzero vector in $\Lambda$.
- Closest vector problem (CVP). Given a lattice $\Lambda$ and a query point $\boldsymbol{y}$, find the closest vector to $\boldsymbol{y}$ in $\Lambda$.
- SIVP. Given a lattice $\Lambda$, find $n$ linearly independent vectors such that the length of the longest vector is minimized.
- BDD. Given a lattice $\Lambda$ and a query point $\boldsymbol{y}$, if the query point is not too far from the lattice, solve the CVP problem.

The hardness of these lattice problems can be analyzed by using complexity theory while these problems have been proved computationally intractable [40]. In the early ages of lattice-based cryptography, cryptographic schemes are built directly from these hard problems and migrated to their approximate versions afterwards. There exists a tight connection between the hardness of a problem and the security of a cryptographic scheme. Nevertheless, modern lattice-based cryptographic schemes are mostly designed from average-case hard problems, rather than the above worst-case hard problems. On one hand, worst-case hardness means that the worst-case instance of the problem is hard to solve. On the other hand, average-case hardness means that given any random instances of the problem, it is computationally hard to solve. While in the early ages many have believed that NP-hard problems would have been the perfect basis to build cryptographic schemes, not all NP-hard problems can achieve average case hardness. Thus it is crucial to design average-case hard problems which also feature proofs of worst-case hardness.

#### 5.1.1 *LWE*

The randomly constructed LWE is a popular average-case problem. As introduced in Section 3, the search-version LWE is to find $\boldsymbol{s} \in \mathbb{Z}_q^n$ given $(\boldsymbol{A}, \boldsymbol{b} = \boldsymbol{A}\boldsymbol{s} + \boldsymbol{e} \mod q)$, where $\boldsymbol{e} \in \mathbb{Z}^m$ is the noise.

#### 5.1.2 *Construction A*

In coding theory, construction A is a method for generating a lattice by "lifting" a linear code to the Euclidean space. Let $\mathcal{C} = C[n,k] \subseteq \mathbb{Z}_q^n$ be a linear code with dimension $k$ and length $n$, where $q$ is a prime number. A lattice $\Lambda$ constructed from the code $\mathcal{C}$ based on construction A is defined by

$$\Lambda = q\mathbb{Z}^n + \phi(\mathcal{C}), \tag{1}$$

where $\phi : \mathbb{Z}_p^n \to \mathbb{R}^n$ is the embedding function which maps a vector in $\mathbb{Z}_q^n$ to its real-valued version. Many properties of construction A lattices can be related to the properties of their underlying codes.

LWE can be interpreted as the decoding over construction A lattices. As the public basis $\boldsymbol{A}$ defines a random code $\mathcal{C}$, $\boldsymbol{A}\boldsymbol{s}$ represents the codeword generated by message $\boldsymbol{s}$. Thus the problem of recovering $\boldsymbol{s}$ from the noisy observation $\boldsymbol{b} \in \mathbb{Z}_q^m$ is in essence the decoding of a random linear code.

### 5.2 On the decoding of hard problems: lattice reduction and algebraic codes

Both algebraic and non-algebraic lattice problems are presumed hard classically and quantumly. For the sake of evaluating the bit-level security of cryptographic schemes based on LWE, SIS, ring-LWE, ring-SIS, module-LWE, and module-SIS, a major approach is to employ lattice reduction-based cryptanalysis [41]. Lattice reduction is to find a basis with short and nearly orthogonal vectors when given a basis as input. Its applications include not only cryptanalysis but also information theory (e.g., designing the network coding coefficients in compute-and-forward [42]) and wireless communications (e.g., lattice-reduction-aided MIMO detection/precoding [43]).

This subsection will review the lattice reduction attacks in analyzing the actual complexity of non-algebraic and algebraic lattice problems, in which the algebraic lattice reduction is examined from the perspective of algebraic codes. Specifically, an essential question in the LBC community is whether the algebraic lattices induced by ring-LWE, ring-SIS, module-LWE, and module-SIS, may downgrade the security level of the associated cryptographic schemes. While the cryptanalytic state of the art about lattice reduction is to unfold the ring/module to a large degree, the algebraic lattice coding theory shows that small polynomial speed-ups can always be guaranteed.

#### 5.2.1 *Lattice reduction attacks*

Lattice reduction has been well investigated for conventional $\mathbb{Z}$-lattices. Some of the reduction algorithms include: the celebrated Lenstra-Lenstra-Lovász (LLL) [44] and its variants [45–47], block-Korkine-Zolotarev (BKZ) [48], Korkine-Zolotarev (KZ) [49], and Minkowski [50].

The LLL algorithm is polynomial time, but the first vector of the output basis is an approximation of the shortest vector of the lattice with an exponential approximation bound. Block-wise generalisations of LLL include the BKZ, whose complete analysis has been open until very recently [51], Schnorr's algorithm, the transference algorithm by Gama et al., and Gama-Nguyen's slide algorithm (see [41] for a survey containing the analysis of the latter three). With the help of a given subroutine for solving SVP in lattices of dimension at most the block size, block-wise generalisations achieve better provable approximation bounds than LLL. Moreover, the choice of the block size offers flexibility for fine-tuning the algorithms catering to different application settings. When the block size $\beta$ is logarithmic in the dimension $n$ of the lattice, these algorithms remain polynomial-time and the approximation bounds remain exponential. For applications in cryptography, the block size is set to be linear in the dimension

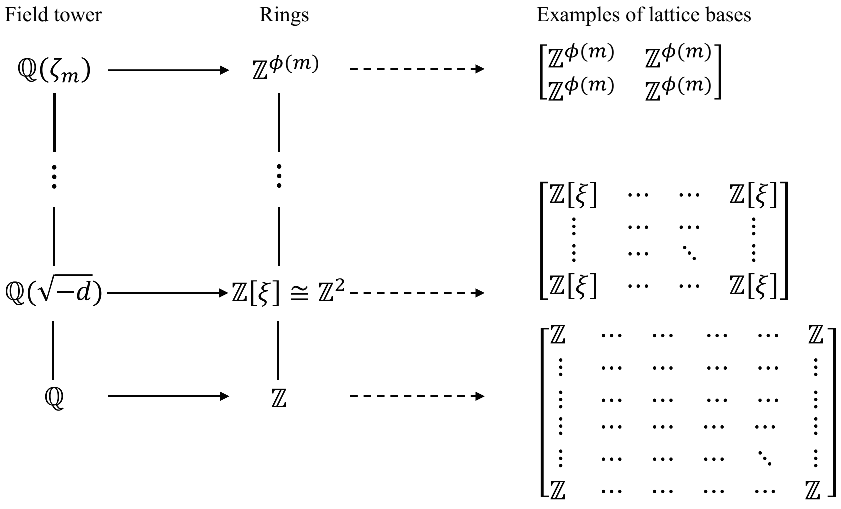
Field tower
Rings
Examples of lattice bases
$\mathbb{Q}(\zeta_m)$
$\mathbb{Z}^{\phi(m)}$
$\begin{bmatrix}\mathbb{Z}^{\phi(m)} & \mathbb{Z}^{\phi(m)}\\ \mathbb{Z}^{\phi(m)} & \mathbb{Z}^{\phi(m)}\end{bmatrix}$
$\mathbb{Q}(\sqrt{-d})$
$\mathbb{Z}[\xi] \cong \mathbb{Z}^2$
$\begin{bmatrix}\mathbb{Z}[\xi] & \cdots & \cdots & \mathbb{Z}[\xi]\\ \vdots & \cdots & \cdots & \vdots\\ \vdots & \cdots & \ddots & \vdots\\ \mathbb{Z}[\xi] & \cdots & \cdots & \mathbb{Z}[\xi]\end{bmatrix}$
$\mathbb{Q}$
$\mathbb{Z}$
$\begin{bmatrix}\mathbb{Z} & \cdots & \cdots & \cdots & \cdots & \mathbb{Z}\\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots\\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots\\ \vdots & \cdots & \cdots & \cdots & \cdots & \vdots\\ \vdots & \cdots & \cdots & \cdots & \ddots & \vdots\\ \mathbb{Z} & \cdots & \cdots & \cdots & \cdots & \mathbb{Z}\end{bmatrix}$

**Figure 8** The tower decomposition of fields.

of the lattice, in which case these algorithms are only known to be the exponential time while the approximation bounds are polynomial in the dimension of the lattice. Gama-Nguyen's slide algorithm has recently been revisited [52] with the following improved approximation bound:

$$\gamma = \gamma' \cdot \left(\gamma' \sqrt{\beta}\right)^{\frac{2(n-\beta)}{\beta-1}},$$

where $\gamma'$ denotes the approximation bound of the given subroutine that solves SVP of lattices of dimension at most $\beta$.

### 5.2.2 *Algebraic variants*

To achieve lower storage and communication costs, many lattice-based cryptographic schemes rely on using algebraic lattices, e.g., RingLWE, RingSIS. These types of lattices can be classified as $\mathcal{O}_{\mathbb{K}}$-lattices, also referred to as $\mathcal{O}_{\mathbb{K}}$-modules, where $\mathcal{O}_{\mathbb{K}}$ denotes the ring of integers of a number field $\mathbb{K}$.

Based on the tower decomposition of fields, as shown in Figure 8, the $\mathcal{O}_{\mathbb{K}}$-lattices can be unfolded to modules of different degrees. By transforming the $\mathcal{O}_{\mathbb{K}}$-modules to $\mathbb{Z}$-modules, conventional lattice reduction algorithms can be readily applied, but this treatment has failed to capture/employ the structures by the algebraic lattices.

The second type of reduction algorithms is to transform the $\mathcal{O}_{\mathbb{K}}$-modules to $\mathbb{Z}[\xi]$-modules [53,54], where $\mathbb{Z}[\xi] \cong \mathbb{Z}^2$ refers to the ring of integers of imaginary quadratic fields. These algorithms were actually motivated by the requirements in coding theory and communications. Specifically, if signal constellations and codes carved from $\mathbb{Z}[\xi]$ are used in the side of transmission [55–57], then the decoding tasks in the side of receivers require performing algebraic lattice reduction over $\mathbb{Z}[\xi]$-lattices. It has been shown that the $\mathbb{Z}[\xi]$-lattice reduction algorithms are approximately 2 times faster than their $\mathbb{Z}$-lattice counterparts.

More advanced treatment is to transform the $\mathcal{O}_{\mathbb{K}}$-modules to high-degree modules and design LLL/BKZ reduction algorithms directly for these modules. This line of works include: Napias's work [58] extends LLL to lattices defined by Euclidean rings, Fieker and Pohst's work [59] of LLL over Dedekind domains, and Kim and Lee's work [60] of LLL for arbitrary Euclidean domains.

This line of investigations have accumulated into the breakthrough studies on module lattice reductions [61,62], where the lattices are $R$-modules $\mathcal{M}$ with a general extension ring $R$ of $\mathbb{Z}$ (usually the ring of integers of an extension field of $\mathbb{Q}$). These results are developed with their cryptographic applications in mind and are naturally comparable with the block-wise generalisations of LLL (in particular, when the block size is linear in the dimension of the lattice). For module lattices the block-size $\beta$ is always a multiple of the degree $d = [R : \mathbb{Z}]$ for the obvious reason that a rank one sub-module (also called an ideal) of $\mathcal{M}$ is corresponding to a $d$-dimensional $\mathbb{Z}$-lattice. The reduction uses an SVP solver for sub-modules

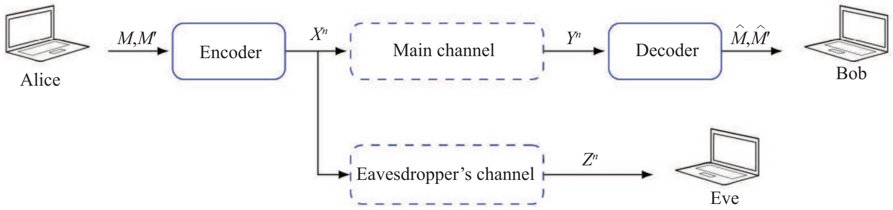

**Figure 9** (Color online) Illustration of the wiretap channel.

of rank $\beta/d$ ($\beta/d \geqslant 2$) to solve an SVP in an $R$-module of rank $n/d$. The approximation bound achieved is

$$\gamma = (\gamma')^2 d \cdot \left(\gamma' \sqrt{\beta d}\right)^{\frac{2(n-\beta)}{\beta-d}},$$

where $\gamma'$ denotes the approximation bound of the given subroutine that solves SVP of $R$-module lattices of rank at most $\beta/d$. This shows that SVP solvers for $R$-module lattices of dimension $\beta$ are almost as good as SVP solvers for general lattices in $R$-module lattice reduction.

## 6 Information-theoretic security

In this section, we present information-theoretic security to establish secure networks and communication systems. Since information-theoretic security does not rely on computation hardness at all, it is also quantum safe. The starting point of information-theoretic security is generally attributed to Shannon's work on the so-called perfect secrecy. However, compared with the more mainstream encryption algorithms, information-theoretic security was regarded as no more than a beautiful, yet unpractical, theoretical construct, because such security is based on the characteristics of the communication channel instead of mathematical operations that are assumed to be hard to compute. With the development of capacity-approaching codes such as turbo codes, LDPC codes, and polar codes, information theorists are now aware of more practical methods to resist the unpredictable disturbance over the communication channel, gradually removing the shadow of cryptography. The potential of information-theoretic security to strengthen the security of the physical layer becomes increasingly clear.

It is important to indicate the principle differences between classical cryptography and information-theoretic security, which may help one select the appropriate method in practice. In general, the security of classical cryptography such as public-key cryptography is based on the conjecture that several one-way functions are hard to invert, meaning that the attackers cannot break the cryptographic system with efficient algorithms and limited computational resource. The security-based on computational hardness cannot be persistently guaranteed from a mathematical perspective, as the computing power continues to increase at a very fast pace. A typical example is that some quantum algorithms that are able to solve the prime factorization problem are now within reach. In addition, when it comes to compare the strengths of different cipher systems, there are no precise metrics available. Whereas for information-theoretic security, no computational restrictions are forced on the eavesdropper and very precise metrics such as the information leakage to the eavesdropper can be evaluated to measure its strength of security. Moreover, the system architecture for information-theoretic security is basically compatible to the one for communication, making it possible to provide an additional layer of security to the existing communication networks without any changes to their infrastructure.

In the following content, we will focus on the wiretap channel model and briefly show how lattices and linear codes manage to achieve the secrecy capacity inherently associated with such a model under an information-theoretic secrecy requirement. The design of secure channel codes for information-theoretic security dates back to Wyner's study on the wiretap channel model [63]. As shown in Figure 9, a wiretap channel is a broadcast channel where one of the receivers is legitimate and the other is treated as an adversary. The channel between the transmitter (Alice) and the legitimate receiver (Bob) is called the main channel, and the one between Alice and the eavesdropper (Eve) is called the wiretapper's channel. For transmission, both the confidential message $\mathsf{M}$ and the auxiliary message $\mathsf{M}'$ are encoded to the codeword $\mathsf{X}^n$. The outputs of the main channel and the wiretapper's channel are given by $\mathsf{Y}^n$ and $\mathsf{Z}^n$, respectively. The reliability requirement says that $\mathsf{M}$ should be recovered correctly on Bob's side, and the security requirement is related to the measure of information leakage. When the average information

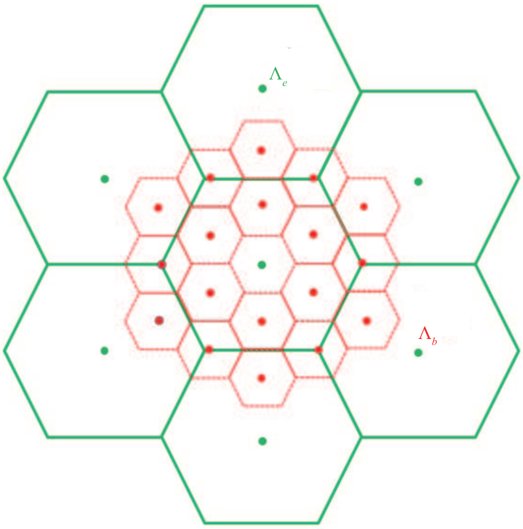

**Figure 10** (Color online) An example: nested lattices.

leakage is asymptotically vanishing, i.e., $\frac{1}{n}I(\mathsf{M};\mathsf{Z}^n) \to 0$, the notion is usually called the weak secrecy condition [63]. A stronger condition directly requires the information leakage $I(\mathsf{M};\mathsf{Z}^n)$ to be vanishing as $n$ increases, which is called the strong secrecy condition [64]. The secrecy capacity is defined as the maximum rate of $\mathsf{M}$ when the reliability requirement and the security requirement are both satisfied. It is also interesting that the secrecy capacity remains unchanged when the weak secrecy condition is replaced by the strong secrecy condition [64].

When both the main channel and the wiretapper's channel are symmetric, and the latter one is degraded with respect to the former one, the secrecy capacity is given by $C_b - C_e$, where $C_b$ and $C_e$ denote the capacities of the main channel and the wiretapper's channel, respectively. In this case, it turns out that the design of secure channel codes is highly related to that of capacity-achieving codes. One may construct two capacity-achieving codes $\mathcal{C}_b$ and $\mathcal{C}_e$ for the main channel and the wiretapper's channel, respectively. To fulfill the reliability and the security conditions, $\mathcal{C}_e$ conveys all auxiliary message $\mathsf{M}'$ and the confidential message $\mathsf{M}$ is encoded to the quotient group $\mathcal{C}_b/\mathcal{C}_e$. Since both $\mathcal{C}_b$ and $\mathcal{C}_e$ are capacity-achieving, their rates approach $C_b$ and $C_e$, respectively. Therefore, the secrecy capacity $C_b - C_e$ can be achieved.

## 6.1 Lattices

Particularly, for the Gaussian wiretap channel, where both the main channel and the wiretapper's channel are AWGN (additive white Gaussian noise) channels, we can construct two capacity-achieving lattice codes accordingly for the two channels. More explicitly, we construct two nested lattices $\Lambda_b$ and $\Lambda_e$ for Bob and Eve respectively, where $\Lambda_b$ is AWGN-good [65] and $\Lambda_e$ is secrecy-good [66]. To satisfy the power constraint, we then assign a Gaussian distribution to $\Lambda_b$ and $\Lambda_e$ simultaneously. As can be seen from Figure 10, when the two lattices are nested, i.e., $\Lambda_e$ is a sub-lattice of $\Lambda_b$, it is convenient to shape the two lattices simultaneously because a Gaussian distribution defined over a sub-lattice is still a LGD. It is worth noting that the lattice-based secure coding scheme can even achieve the so-called semantic security [66], namely the scheme is secure for an arbitrarily distributed confidential message. A polar-lattice based implementation of the scheme in [66] can be found in [67].

### 6.1.1 *AWGN-good lattices*

The AWGN-goodness of a lattice can be regarded as the maximum efficiency of the lattice volume when resisting the Gaussian noise. Suppose $\boldsymbol{W}$ is an $n$-dimensional zero-mean white-Gaussian vector. The probability density function of $\boldsymbol{W}$ depends only on the Euclidean norm $\|\boldsymbol{W}\|$. By the law of large numbers (LLN), the normalized squared norm $\frac{1}{n}\|\boldsymbol{W}\|^2$ converges in probability to the Gaussian noise variance $\sigma^2$, which means that a zero-mean white-Gaussian vector tends to be uniformly distributed over a spherical shell of radius $\sqrt{n\sigma^2}$. For a high dimensional sphere, most of its volume is distributed on its surface and the probability of escaping such sphere for $\boldsymbol{W}$ vanishes. Therefore, to guarantee a vanishing error probability, the volume of a lattice is only required to be barely larger than the volume of an $n$-dimensional spherical shell of radius $\sqrt{n\sigma^2}$, which is approximated by $(2\pi e\sigma^2)^{\frac{n}{2}}$. Let $V(\Lambda)$ denote the volume of a lattice $\Lambda$, the AWGN-goodness establishes a lower bound on the normalized volume of $\Lambda$ as

$$\frac{V(\Lambda)^{\frac{2}{n}}}{\sigma^2} > 2\pi e, \tag{2}$$

for any desirable error probability $P_e \to 0$.

Loeliger has proven the existence of AWGN-goodness lattices in [68], using the random linear codes in $\mathbb{F}_q$. This construction method which directly lifts a $q$-ary non-binary linear code to the Euclidean space is named as construction A as is introduced in Section 5. The decoding of such construction A lattices is highly related to that of the embedded $q$-ary non-binary codes, which is relatively more complex than decoding binary linear codes. For this reason, a more practical method of constructing AWGN-good lattices attributes to Forney et al. [69], which utilizes a multi-level structure and a series of nested binary linear codes. This method is referred to as construction D. Forney et al. showed that the construction D lattices are AWGN-good, or equivalently sphere-bound-achieving when the underlying binary codes are capacity-achieving for each level. Following this line, we can use capacity-achieving codes such as polar codes [70] and LDPC codes [71] to construct such construction D lattices.

#### 6.1.2 *Secrecy-good lattices*

Unlike the mentioned AWGN-good lattices, which are constructed to establish reliable transmissions, the so-called secrecy-good lattices [66] are designed for preventing information leakage to adversarial parties in a communication system. For secrecy-good lattices, the inequality (2) is reversed to make it difficult to correctly decode the confidential message. This can be done by forcing the rate of binary codes slightly exceeding the channel capacity at each level in the construction D, and making the resulted lattice relatively denser such that $V(\Lambda)/\sigma^2$ becomes smaller than $2\pi e$. According to Shannon's channel coding theory, it is impossible to recover the correct lattice point from the received signal since the transmission rate is above the channel capacity. A more rigorous proof in [66] shows that secrecy-good lattices are capable of achieving strong secrecy.

### 6.2 Codes

#### 6.2.1 *LDPC codes for BEC wiretap channels*

LDPC codes are famous for their capacity-approaching performance on many communication channels. However, LDPC codes have been used to build wiretap codes only with limited success. When the main channel $V$ is noiseless and the wiretapper's channel $W$ is a binary erasure channel (BEC), LDPC codes for this BEC wiretap channel were presented in [72, 73]. Especially in [73], the authors generalized the link between capacity-approaching codes and weak secrecy capacity. The use of capacity-achieving codes for the wiretapper's channel is a sufficient condition for weak secrecy. This view point provided a clear code construction method for secure communication across arbitrary wiretap channels. Then, they used this idea to construct the first secrecy capacity-achieving LDPC codes for a wiretap channel with a noiseless $V$ and a BEC $W$ under the belief propagation decoding (BP) in terms of weak secrecy. Later, Ref. [74] proved that the same construction can be used to guarantee strong secrecy at lower rates. A similar construction based on two-edge-type LDPC codes was proposed in [75] for the BEC wiretap channel ($V$ is no longer noiseless). Unfortunately, general LDPC codes do not have the capacity-achieving property for binary memoryless symmetric channels (BMSCs) other than BECs. Therefore, the coset coding scheme using general LDPC codes cannot achieve the secrecy capacity when the wiretapper's channel is not a BEC.

In this case, spatially coupled LDPC (SC-LDPC) codes, which are provable to achieve the capacity of general BMSCs, provide us with a promising approach. In [76], a coset coding scheme based on regular two-edge-type SC-LDPC codes is proposed for a BEC wiretap channel, where the main channel is also a BEC. It is shown that the whole rate equivocation region of such BEC wiretap channel can be achieved by using this scheme under weak secrecy condition. Since SC-LDPC codes are universally capacity-achieving, it is also conjectured that this construction is optimal for the class of wiretap channel where the main channel and wiretapper's channel are BMSCs and the wiretapper's channel is physically degraded with respect to the main channel.

### 6.2.2 *Polar codes for degraded wiretap channels*

Compared with LDPC codes, polar codes provide a more powerful approach to design wiretap codes, since they are capacity-achieving for general BMS channels, not just for BECs. Recently there has been a lot of interest in the design of wiretap codes based on polar codes. For example, polar codes are employed to build secure schemes for the degraded wiretap setting with BMS channels in [77], but only weak secrecy is guaranteed. In [78], it was shown that, with a minor modification of the original design, polar codes achieve strong secrecy (and also semantic security) for degraded wiretap channels. Unfortunately, they could not guarantee the reliability of the main channel in the non-degraded case. In [79], a multi-block polar coding scheme was proposed to solve this reliability problem under the condition that the number of blocks is sufficiently large. In the meantime, a similar multi-block coding scheme was discussed in [80]. Their polar coding scheme also achieves the secrecy capacity under strong secrecy condition and guarantees reliability for the legitimate receiver. However, Ref. [80] only proved the existence of this coding scheme, and thus it might be computationally hard to find the explicit structure.

Now we briefly introduce the idea of polar wiretap coding. The construction of polar codes consists of an information set $\mathcal{I}$, which carries message bits to be transmitted, and a frozen set $\mathcal{F}$, which is fixed with some known value. When a channel $W$ is degraded with respect to another channel $V$, their information sets, $\mathcal{I}_W$ and $\mathcal{I}_V$, satisfy $\mathcal{I}_W \subseteq \mathcal{I}_V$ [81, Lemma 1.8]. Then, the indices $1, \ldots, n$ can be divided into 3 sets: $\mathcal{I}_V^c$, $\mathcal{I}_W$, and $\mathcal{I}_V \setminus \mathcal{I}_W$. The wiretap coding scheme is to assign these three sets with frozen bits which are known to both Bob and Eve prior to transmission, random bits in order to confuse the eavesdropper, and message bits, respectively. Due to the capacity-achieving property of polar codes, the secrecy capacity is achieved as $\lim_{n\to\infty} \frac{|\mathcal{I}_V \setminus \mathcal{I}_W|}{n} = C(V) - C(W)$. The reliability for Bob is also guaranteed by the standard polar decoding. The weak secrecy can be proved using Fano's inequality.

For strong secrecy and non-degraded wiretap channels, the above wiretap coding scheme needs to be modified. In these cases, the inclusion relation of information sets in the degraded wiretap channel does not hold, and we have to partition the index set $\{1, \ldots, n\}$ into four sets, one of which is unreliable for Bob but reliable for Eve. This becomes problematic since bits in this set need to be known by Bob but kept secret from Eve. Fortunately, this problem can be solved by a multi-block technique proposed in [79] to achieve reliability and strong secrecy simultaneously. The idea is to allocate some reliable and secure bits in the current block for the bits in the problematic set of the next block. Details of this scheme can be found in [79].

### 6.2.3 *Practical instances*

The above coding techniques to achieve secrecy capacity for wiretap channels assume infinite block length. In the context of real-world design, engineers prefer bit error rate (BER) as a measure of security, i.e., the legitimate recipient on the main channel should have overwhelmingly smaller BER than the eavesdropper on the wiretap channel to achieve secure transmission. The security gap, as a practical metric, is defined by

$$S_{\mathrm{sg}} = \mathrm{SNR}_{\mathrm{th}}(P_e^B) - \mathrm{SNR}_{\mathrm{th}}(P_e^E),$$

where $\mathrm{SNR}_{\mathrm{th}}(P_e^B)$ is the signal-to-noise ratio (SNR) threshold for Bob to reliably recover the message at a BER of $P_e^B$ and $\mathrm{SNR}_{\mathrm{th}}(P_e^E)$ is the threshold for Eve to operate at a BER (approximate to 0.5) not able to extract enough information about the message from the received signal. In other words, the security gap defines the minimum SNR advantage the main channel has over the wiretap channel to achieve reliable and secure transmission.

Considering a typical BER-SNR curve illustrating the decoding performance of certain codes, a sharper falloff will render a smaller security gap. Theorists and practitioners have contributed to reducing the security gap. An exploration towards that direction is the punctured LDPC codes in [82] for the Gaussian wiretap channel where the punctured bits are used to transmit the secret information. As a result, the BER of the eavesdropper deteriorates faster towards 0.5 as SNR decreases than it does without puncturing. The security gap required to reach a certain level of security can be further reduced with scrambling [83].

When the wiretap channel has a quality the same as or even better than the main channel, a feedback mechanism, automatic repeat-request (ARQ) protocol, was used to achieve reliability and security [83]. Another secure transmission method using ARQ can be found in [84] where a code-hopping scheme [85] is employed for encoding with a single-use parity-check matrix and ARQ is used for synchronization.

## 7 Open problems

The NIST standardization process marks the beginning of a paradigm shift to PQC, not the end. We offer a perspective into the future of this exciting area.

- What will post-NIST PQC look like, i.e., PQC beyond the NIST process? Traditionally, cryptography is also known as secure communications. Can we merge communication and cryptography as envisioned by Shannon himself? As cryptography plays an increasingly important role in secure computation, can we develop a unified theory of communications, computation, and security? Compared to information-theoretic security, information theory is of little use for computational security; can algorithmic information theory make a difference? In data science and artificial intelligence, the privacy and security of data become a central concern; thus PQC will find broad applications in these areas.
- In PQC, cryptosystems based on linear codes (e.g., LWE) seem to be the most successful. The analogy of methods (e.g., construction A, ideal lattices) used in coding theory and cryptography is especially striking; PQC would greatly benefit from the synergy of the two fields. Moving further, why do we restrict ourselves to linear codes? Linear codes are arguably the simplest algebraic structure from a mathematical point of view. What about nonlinear codes and more sophisticated algebraic structures such as algebraic geometry?
- Decoding random linear codes, regardless of the metric, is believed to be hard even for quantum computers. But the supporting evidence is not enough — people are merely unaware of devastating quantum attacks to codes and lattices nowadays. More research on the quantum hardness of coding problems is required.
- More interaction between coding theory and cryptography is needed. As discussed in Remark 1 encryption is analogous to channel coding, some lattice signatures can be deemed as the "reverse process" of source coding, and we also find a link between IBE and joint source-channel coding. Lattice and code-based cryptosystems, especially code-based ones, require powerful error-correction codes with extremely low decoding error rates and high decoding speed. Intricate channel models of these cryptosystems invite new methods to design good codes for error correction, as opposed to the standard i.i.d. model.
- So far, quantum computing is destructive to cryptography. Can we use quantum computers constructively in cryptography? That is, can we use the computational power of quantum computers to design new cryptosystems that are also resilient to quantum attacks?
- Even classically, the security of PQC is not well understood. Problems of algebraic lattices, such as lattice reduction, deserve a thorough investigation. Similar techniques for codes are needed as well.
- Crypto conferences are apparently dominated by a small number of countries. The community should become more open and diverse, not only in terms of people but also in terms of disciplines. PQC is a chance to attract people from different countries and different disciplines.

**Acknowledgements** This work was supported by U.K. Engineering and Physical Sciences Research Council (Grant No. EP/S021-043/1) and National Natural Science Foundation of China (Grant Nos. 61871257, 61902149).

**References**

1 Regev O. On lattices, learning with errors, random linear codes, and cryptography. In: Proceedings of the 37th Annual ACM Symposium on Theory of Computing, New York, 2005. 84–93

2 Lyubashevsky V, Peikert C, Regev O. On ideal lattices and learning with errors over rings. J ACM, 2013, 60: 43

3 Lyubashevsky V, Peikert C, Regev O. A toolkit for ring-lwe cryptography. In: Proceedings of Annual International Conference on the Theory and Applications of Cryptographic Techniques, 2013. 35–54

4 Peikert C. Lattice cryptography for the Internet. In: Proceedings of International Workshop on Post-quantum Cryptography, 2014. 197–219

5 D'Anvers J P, Guo Q, Johansson T, et al. Decryption failure attacks on ind-cca secure lattice-based schemes. In: Proceedings of IACR International Workshop on Public Key Cryptography, 2019. 565–598

6 D'Anvers J P, Rossi M, Virdia F. (One) failure is not an option: bootstrapping the search for failures in lattice-based encryption schemes. In: Proceedings of Annual International Conference on the Theory and Applications of Cryptographic Techniques, 2020. 3–33

7 Guo Q, Johansson T, Yang J. A novel CCA attack using decryption errors against LAC. In: Proceedings of International Conference on the Theory and Application of Cryptology and Information Security, 2019. 82–111

8 Fritzmann T, Pöppelmann T, Sepulveda J. Analysis of error-correcting codes for lattice-based key exchange. In: Proceedings of International Conference on Selected Areas in Cryptography, 2019. 369–390

9 D'Anvers J P, Vercauteren F, Verbauwhede I. The impact of error dependencies on ring/mod-LWE/LWR based schemes. In: Proceedings of International Conference on Post-Quantum Cryptography, 2019. 103–115

10 Wang J B, Ling C. Polar coding for ring-LWE-based public key encryption. 2021. https://eprint.iacr.org/2021/619.pdf

11 Gentry C, Peikert C, Vaikuntanathan V. Trapdoors for hard lattices and new cryptographic constructions. In: Proceedings of the 40th Annual ACM Symposium on Theory of Computing, 2008. 197–206

12 Micciancio D, Peikert C. Trapdoors for lattices: simpler, tighter, faster, smaller. In: Proceedings of Annual International Conference on the Theory and Applications of Cryptographic Techniques, 2012. 700–718

13 Wang Z, Ling C. On the geometric ergodicity of metropolis-hastings algorithms for lattice gaussian sampling. IEEE Trans Inform Theor, 2018, 64: 738–751

14 Wang Z, Ling C. Lattice Gaussian sampling by Markov Chain Monte Carlo: bounded distance decoding and trapdoor sampling. IEEE Trans Inform Theor, 2019, 65: 3630–3645

15 Lyubashevsky V. Fiat-Shamir with aborts: applications to lattice and factoring-based signatures. In: Proceedings of International Conference on the Theory and Application of Cryptology and Information Security, 2009. 598–616

16 Lyubashevsky V. Lattice signatures without trapdoors. In: Proceedings of Annual International Conference on the Theory and Applications of Cryptographic Techniques, 2012. 738–755

17 Ducas L, Durmus A, Lepoint T, et al. Lattice signatures and bimodal gaussians. In: Proceedings of Annual Cryptology Conference, 2013. 40–56

18 Peikert C. An efficient and parallel Gaussian sampler for lattices. In: Proceedings of Annual Cryptology Conference, 2010. 80–97

19 Ducas L, Prest T. Fast fourier orthogonalization. In: Proceedings of ACM on International Symposium on Symbolic and Algebraic Computation, New York, 2016. 191–198

20 Micciancio D, Walter M. Gaussian sampling over the integers: efficient, generic, constant-time. In: Proceedings of Annual International Cryptology Conference, 2017. 455–485

21 Zhao R K, Steinfeld R, Sakzad A. FACCT: fast, compact, and constant-time discrete Gaussian sampler over integers. IEEE Trans Comput, 2020, 69: 126–137

22 Wang J B, Ling C. Polar sampler: discrete gaussian sampling over the integers using polar codes. 2019. https://eprint.iacr.org/2019/674.pdf

23 McEliece R J. A Public-key Cryptosystem Based on Algebraic Coding Theory. The Deep Space Network Progress Report, 1978. 114–116

24 Niederreiter H. Knapsack-type cryptosystems and algebraic coding theory. Prob Control Inf Theory, 1986, 15: 159–166

25 Berlekamp E, McEliece R, van Tilborg H. On the inherent intractability of certain coding problems (Corresp.). IEEE Trans Inform Theor, 1978, 24: 384–386

26 Gaborit P. Shorter keys for code based cryptography. In: Proceedings of International Workshop on Coding and Cryptography (WCC), 2005. 81–91

27 Misoczki R, Tillich J P, Sendrier N, et al. MDPC-mceliece: new mceliece variants from moderate density parity-check codes. In: Proceedings of IEEE International Symposium on Information Theory, 2013. 2069–2073

28 Guo Q, Johansson T, Stankovski P. A key recovery attack on MDPC with CCA security using decoding errors. In: Proceedings of International Conference on the Theory and Application of Cryptology and Information Security, 2016. 789–815

29 Gaborit P, Zémor G. On the hardness of the decoding and the minimum distance problems for rank codes. IEEE Trans Inform Theor, 2016, 62: 7245–7252

30 Bardet M, Bros M, Cabarcas D, et al. Improvements of algebraic attacks for solving the rank decoding and minrank problems. In: Proceedings of International Conference on the Theory and Application of Cryptology and Information Security, 2020. 507–536

31 Courtois N T, Finiasz M, Sendrier N. How to achieve a mceliece-based digital signature scheme. In: Proceedings of International Conference on the Theory and Application of Cryptology and Information Security, 2001. 157–174

32 Debris-Alazard T, Sendrier N, Tillich J P. Wave: a new family of trapdoor one-way preimage sampleable functions based on codes. In: Proceedings of International Conference on the Theory and Application of Cryptology and Information Security, 2019. 21–51

33 Gaborit P, Ruatta O, Schrek J, et al. New results for rank-based cryptography. In: Proceedings of International Conference on Cryptology in Africa, 2014. 1–12

34 Faugère J C, Gauthier-Umaña V, Otmani A, et al. A distinguisher for high rate mceliece cryptosystems. In: Proceedings of IEEE Information Theory Workshop, 2011. 282–286

35 Debris-Alazard T, Tillich J P. Two attacks on rank metric code-based schemes: ranksign and an IBE scheme. In: Proceedings of International Conference on the Theory and Application of Cryptology and Information Security, 2018. 62–92

36 Persichetti E. Efficient one-time signatures from quasi-cyclic codes: a full treatment. Cryptography, 2018, 2: 30

37 Deneuville J C, Gaborit P. Cryptanalysis of a code-based one-time signature. Des Codes Cryptogr, 2020, 88: 1857–1866

38 Fukushima K, Sarathi Roy P, Xu R, et al. Racoss: random code-based signature scheme. 2017. https://csrc.nist.gov/Projects/post-quantum-cryptography/Round-1-Submissions

39 Aragon N, Blazy O, Gaborit P, et al. Durandal: a rank metric based signature scheme. In: Proceedings of Annual International Conference on the Theory and Applications of Cryptographic Techniques, 2019. 728–758

40 Micciancio D, Goldwasser S. Complexity of Lattice Problems. Berlin: Springer, 2002

41 Nguyen P Q, Vallée B. The LLL Algorithm: Survey and Applications. Belin: Springer, 2010

42 Feng C, Silva D, Kschischang F R. An algebraic approach to physical-layer network coding. IEEE Trans Inform Theor, 2013, 59: 7576–7596

43 Wubben D, Seethaler D, Jalden J, et al. Lattice reduction. IEEE Signal Process Mag, 2011, 28: 70–91

44 Lenstra A K, Lenstra Jr H W, Lovász L. Factoring polynomials with rational coefficients. Math Ann, 1982, 261: 515–534

45 Schnorr C P, Euchner M. Lattice basis reduction: improved practical algorithms and solving subset sum problems. Math Program, 1994, 66: 181–199

46 Chang X W, Yang X, Zhou T. MLAMBDA: a modified LAMBDA method for integer least-squares estimation. J Geodesy, 2005, 79: 552–565

47 Lyu S, Ling C. Boosted KZ and LLL algorithms. IEEE Trans Signal Process, 2017, 65: 4784–4796

48 Chen Y M, Nguyen P Q. BKZ 2.0: better lattice security estimates. In: Proceedings of International Conference on the

Theory and Application of Cryptology and Information Security, 2011. 1–20

49 Lagarias J C, Lenstra Jr H W, Schnorr C P. Korkin-Zolotarev bases and successive minima of a lattice and its reciprocal lattice. Combinatorica, 1990, 10: 333–348

50 Minkowski H. Diskontinuitätsbereich für arithmetische Äquivalenz. Journal für die reine und angewandte Mathematik, 1905, 129: 220–224

51 Li J W, Nguyen P Q. A complete analysis of the BKZ lattice reduction algorithm. 2020. https://eprint.iacr.org/2020/1237.pdf

52 Aggarwal D, Li J W, Nguyen P Q, et al. Slide reduction, revisited — filling the gaps in SVP approximation. In: Proceedings of Annual International Cryptology Conference, 201

53 Lyu S, Porter C, Ling C. Lattice reduction over imaginary quadratic fields. IEEE Trans Signal Process, 2020, 68: 6380–6393

54 Gan Y H, Ling C, Mow W H. Complex lattice reduction algorithm for low-complexity full-diversity MIMO detection. IEEE Trans Signal Process, 2009, 57: 2701–2710

55 Tunali N E, Huang Y C, Boutros J J, et al. Lattices over eisenstein integers for compute-and-forward. IEEE Trans Inform Theor, 2015, 61: 5306–5321

56 Huang Y C, Narayanan K R, Wang P C. Lattices over algebraic integers with an application to compute-and-forward. IEEE Trans Inform Theor, 2018, 64: 6863–6877

57 Stern S, Fischer R F H. Lattice-reduction-aided precoding for coded modulation over algebraic signal constellations. In: Proceedings of the 20th International ITG Workshop on Smart Antennas, Munich, 2016. 1–8

58 Napias H. A generalization of the LLL-algorithm over euclidean rings or orders. J de Théorie des Nombres de Bordeaux, 1996, 8: 387–396

59 Fieker C, Pohst M. On lattices over number fields. In: Proceedings of Algorithmic Number Theory, Second International Symposium, 1996. 133–139

60 Kim T, Lee C. Lattice reductions over Euclidean rings with applications to cryptanalysis. In: Proceedings of IMA International Conference on Cryptography and Coding, 2017. 371–391

61 Lee C, Pellet-Mary A, Stehlé D, et al. An LLL algorithm for module lattices. In: Proceedings of International Conference on the Theory and Application of Cryptology and Information Security, 2019. 59–90

62 Mukherjee T, Stephens-Davidowitz N. Lattice reduction for modules, or how to reduce moduleSVP to moduleSVP. In: Proceedings of Annual International Cryptology Conference, 2020. 213–242

63 Wyner A D. The wire-tap channel. Bell Syst Technical J, 1975, 54: 1355–1387

64 Csiszár I. Almost independence and secrecy capacity. Probl Inform Transm, 1996, 32: 48–57

65 Erez U, Zamir R. Achieving 1/2 log (1+SNR) on the AWGN channel with lattice encoding and decoding. IEEE Trans Inform Theor, 2004, 50: 2293–2314

66 Ling C, Luzzi L, Belfiore J C, et al. Semantically secure lattice codes for the Gaussian wiretap channel. IEEE Trans Inform Theor, 2014, 60: 6399–6416

67 Liu L, Yan Y F, Ling C. Achieving secrecy capacity of the gaussian wiretap channel with polar lattices. IEEE Trans Inform Theor, 2018, 64: 1647–1665

68 Loeliger H A. Averaging bounds for lattices and linear codes. IEEE Trans Inform Theor, 1997, 43: 1767–1773

69 Forney G D, Trott M D, Chung S Y. Sphere-bound-achieving coset codes and multilevel coset codes. IEEE Trans Inform Theor, 2000, 46: 820–850

70 Arikan E. Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. IEEE Trans Inform Theor, 2009, 55: 3051–3073

71 Sadeghi M R, Banihashemi A H, Panario D. Low-density parity-check lattices: construction and decoding analysis. IEEE Trans Inform Theor, 2006, 52: 4481–4495

72 Suresh A T, Subramanian A, Thangaraj A, et al. Strong secrecy for erasure wiretap channels. In: Proceedings of Information Theory Workshop (ITW), 2010. 1–5

73 Thangaraj A, Dihidar S, Calderbank A R, et al. Applications of LDPC codes to the wiretap channel. IEEE Trans Inform Theor, 2007, 53: 2933–2945

74 Bloch M, Barros J. Physical-Layer Security: From Information Theory to Security Engineering. Cambridge: Cambridge University Press, 2011

75 Rathi V, Andersson M, Thobaben R, et al. Performance analysis and design of two edge-type LDPC codes for the BEC wiretap channel. IEEE Trans Inform Theor, 2013, 59: 1048–1064

76 Rathi V, Urbanke R, Andersson M, et al. Rate-equivocation optimal spatially coupled LDPC codes for the BEC wiretap channel. In: Proceedings of Information Theory Proceedings (ISIT), 2011. 2393–2397

77 Andersson M, Rathi V, Thobaben R, et al. Nested polar codes for wiretap and relay channels. IEEE Commun Lett, 2010, 14: 752–754

78 Mahdavifar H, Vardy A. Achieving the secrecy capacity of wiretap channels using polar codes. IEEE Trans Inform Theor, 2011, 57: 6428–6443

79 Sasoglu E, Vardy A. A new polar coding scheme for strong security on wiretap channels. In: Proceedings of Information Theory Proceedings (ISIT), 2013. 1117–1121

80 Renes J M, Renner R, Sutter D. Efficient one-way secret-key agreement and private channel coding via polarization. In: Proceedings of International Conference on the Theory and Application of Cryptology and Information Security, 2013. 194–213

81 Hussami N, Korada S B, Urbanke R. Polar codes for channel and source coding. 2009. ArXiv:0901.2370

82 Klinc D, Ha J, McLaughlin S W, et al. LDPC codes for the Gaussian wiretap channel. IEEE Trans Inform Forensic Secur, 2011, 6: 532–540

83 Baldi M, Bianchi M, Chiaraluce F. Coding with scrambling, concatenation, and HARQ for the AWGN wire-tap channel: a security gap analysis. IEEE Trans Inform Forensic Secur, 2012, 7: 883–894

84 Yin L G, Hao W T. Code-hopping based transmission scheme for wireless physical-layer security. Wirel Commun Mobile Comput, 2018, 2018: 1–12

85 Chen Z, Yin L G, Pei Y K, et al. CodeHop: physical layer error correction and encryption with LDPC-based code hopping. Sci China Inf Sci, 2016, 59: 102309