

# Achieving Secrecy Capacity of the Gaussian Wiretap Channel with Polar Lattices

Ling Liu, Yanfei Yan, and Cong Ling *Member, IEEE*

## Abstract

In this work, an explicit scheme of wiretap coding based on polar lattices is proposed to achieve the secrecy capacity of the additive white Gaussian noise (AWGN) wiretap channel. Firstly, polar lattices are used to construct secrecy-good lattices for the mod- $\Lambda_s$  Gaussian wiretap channel. Then we propose an explicit shaping scheme to remove this mod- $\Lambda_s$  front end and extend polar lattices to the genuine Gaussian wiretap channel. The shaping technique is based on the lattice Gaussian distribution, which leads to a binary asymmetric channel at each level for the multilevel lattice codes. By employing the asymmetric polar coding technique, we construct an AWGN-good lattice and a secrecy-good lattice with optimal shaping simultaneously. As a result, the encoding complexity for the sender and the decoding complexity for the legitimate receiver are both  $O(N \log N \log(\log N))$ . The proposed scheme is proven to be semantically secure.

## I. INTRODUCTION

Wyner [1] introduced the wiretap channel model and showed that both reliability and confidentiality could be attained by coding without any key bits if the channel between the sender and the eavesdropper (wiretapper's channel  $W$ ) is degraded with respect to the channel between the sender and the legitimate receiver (main channel  $V$ ). The goal of wiretap coding is to design a coding scheme that makes it possible to communicate both reliably and securely between the sender and the legitimate receiver. Reliability is measured by the decoding error probability for the legitimate user, namely  $\lim_{N \rightarrow \infty} \Pr\{\hat{M} \neq M\} = 0$ , where  $N$  is the length of transmitted codeword,  $M$  is the confidential message and  $\hat{M}$  is its estimate. Secrecy is measured by the mutual information between  $M$  and the signal received by the eavesdropper  $Z^{[N]}$ . In this work, we will follow the strong secrecy condition proposed by Csiszár [2], i.e.,  $\lim_{N \rightarrow \infty} I(M; Z^{[N]}) = 0$ , which is more widely accepted than the weak secrecy criterion  $\lim_{N \rightarrow \infty} \frac{1}{N} I(M; Z^{[N]}) = 0$ . In simple terms, the secrecy capacity is defined as the maximum achievable rate under both the reliability and strong secrecy conditions. When  $W$  and  $V$  are both symmetric, and  $W$  is degraded with respect to  $V$ , the secrecy capacity is given by  $C(V) - C(W)$  [3], where  $C(\cdot)$  denotes the channel capacity.

This work was supported in part by FP7 project PHYLAWS (EU FP7-ICT 317562) and in part by the China Scholarship Council. This work was presented at the IEEE Int. Symp. Inform. Theory (ISIT), Honolulu, USA, 2014 and the IEEE Inform. Theory Workshop, Jerusalem, ISRAEL, 2015.

Ling Liu, Yanfei Yan and Cong Ling are with the Department of Electrical and Electronic Engineering, Imperial College London, London, UK (e-mails: l.liu12@imperial.ac.uk, y.yan10@imperial.ac.uk, cling@ieee.org).

In the study of strong secrecy, plaintext messages are often assumed to be random and uniformly distributed. From a cryptographic point of view, it is crucial that the security does not rely on the distribution of the message. This issue can be resolved by using the standard notion of *semantic security* [4] which means that, asymptotically, it is impossible to estimate any function of the message better than to guess it without accessing  $Z^{[N]}$  at all. The relation between strong secrecy and semantic security was recently revealed in [5], [6], namely, semantic security is equivalent to achieving strong secrecy for all distributions  $p_M$  of the plaintext messages:

$$\lim_{N \rightarrow \infty} \max_{p_M} I(M; Z^{[N]}) = 0. \quad (1)$$

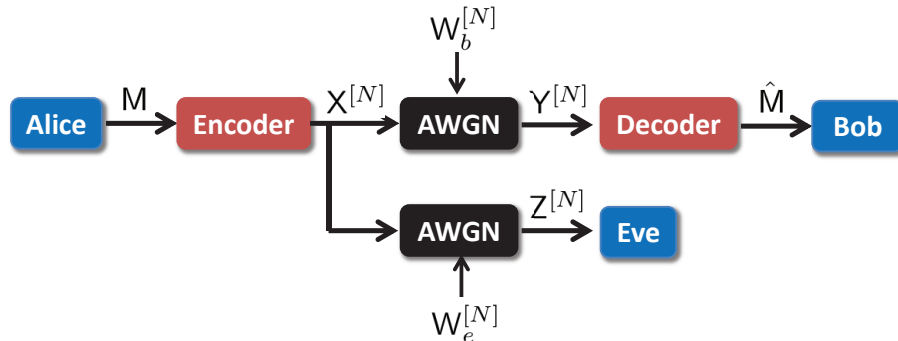


Fig. 1. The Gaussian wiretap channel.

In this work, we construct lattice codes for the Gaussian wiretap channel (GWC) which is shown in Fig. 1. The confidential message  $M$  drawn from the message set  $\mathcal{M}$  is encoded by the sender (Alice) into an  $N$ -dimensional codeword  $X^{[N]}$ . The outputs  $Y^{[N]}$  and  $Z^{[N]}$  received by the legitimate receiver (Bob) and the eavesdropper (Eve) are respectively given by

$$\begin{cases} Y^{[N]} = X^{[N]} + W_b^{[N]} \\ Z^{[N]} = X^{[N]} + W_e^{[N]}, \end{cases}$$

where  $W_b^{[N]}$  and  $W_e^{[N]}$  are  $N$ -dimensional Gaussian noise vectors with zero mean and variance  $\sigma_b^2$ ,  $\sigma_e^2$  respectively. The channel input  $X^{[N]}$  satisfies the power constraint  $P_s$ , i.e.,

$$\frac{1}{N} E[\|X^{[N]}\|^2] \leq P_s.$$

Polar codes [7] have shown their great potential in solving the wiretap coding problem. The polar coding scheme proposed in [8], combined with the block Markov coding technique [9], was proved to achieve the strong secrecy capacity when  $W$  and  $V$  are both binary-input symmetric channels, and  $W$  is degraded with respect to  $V$ . More recently, polar wiretap coding has been extended to general wiretap channels (not necessarily degraded or symmetric) in [10] and [11]. For continuous channels such as the GWC, there also has been notable progress in wiretap lattice coding. On the theoretical aspect, the existence of lattice codes achieving the secrecy capacity to within  $\frac{1}{2}$  nat under the strong secrecy as well as semantic security criterion was demonstrated in [6]. On the practical aspect, wiretap lattice codes were proposed in [12] and [13] to maximize the eavesdropper's decoding error probability.

### A. Our contribution

Polar lattices, the counterpart of polar codes in the Euclidean space, have already been proved to be additive white Gaussian noise (AWGN)-good [14] and further to achieve the AWGN channel capacity with lattice Gaussian shaping [15]<sup>1</sup>. Motivated by [8], we will propose polar lattices to achieve both strong secrecy and reliability over the mod- $\Lambda_s$  GWC. Conceptually, this polar lattice structure can be regarded as a secrecy-good lattice  $\Lambda_e$  nested within an AWGN-good lattice  $\Lambda_b$  ( $\Lambda_e \subset \Lambda_b$ ). Further, we will propose a Gaussian shaping scheme over  $\Lambda_b$  and  $\Lambda_e$ , using the multilevel asymmetric polar coding technique. As a result, we will accomplish the design of an explicit lattice coding scheme which achieves the secrecy capacity of the GWC with semantic security.

- The first technical contribution of this paper is the explicit construction of secrecy-good polar lattices for the mod- $\Lambda_s$  GWC and the proof of their secrecy capacity-achieving. This is an extension of the binary symmetric wiretap coding [8] to the multilevel coding scenario, and can also be considered as the construction of secrecy-good polar lattices for the GWC without the power constraint. The construction for the mod- $\Lambda_s$  GWC provides considerable insight into wiretap coding for the genuine GWC, without deviating to the technicality of Gaussian shaping. This work is also of independent interest to other problems of information theoretic security, e.g., secret key generation from Gaussian sources [19].
- Our second contribution is the Gaussian shaping applied to the secrecy-good polar lattice, which follows the technique of [6], [15]. The resultant coding scheme is proved to achieve the secrecy capacity of the GWC. It is worth mentioning that our proposed coding scheme is not only a practical implementation of the secure random lattice coding in [6], but also an improvement in the sense that we successfully remove the constant  $\frac{1}{2}$ -nat gap to the secrecy capacity<sup>2</sup>.

### B. Comparison with the extractor-based approach

Invertible randomness extractors were introduced into wiretap coding in [5], [20], [21]. The key idea is that an extractor is used to convert a capacity-achieving code with rate close to  $C(V)$  for the main channel into a wiretap code with the rate close to  $C(V) - C(W)$ . Later, this coding scheme was extended to the GWC in [22]. Besides, channel resolvability [23] was proposed as a tool for wiretap codes. An interesting connection between the resolvability and the extractor was revealed in [24].

The proposed approach and the one based on invertible extractors have their respective advantages. The extractor-based approach is modular, i.e., the error-correction code and extractor are realized separately; it is possible to harness the results of invertible extractors in literature. The advantage of our lattice-based scheme is that the wiretap code designed for Eve is nested within the capacity-achieving code designed for Bob, which represents an integrated approach. More importantly, lattice codes are attractive for emerging applications in network information theory

<sup>1</sup>Please refer to [16]–[18] for other methods of achieving the AWGN channel capacity.

<sup>2</sup>The  $\frac{1}{2}$ -nat gap in [6] was due to a requirement on the volume-to-noise ratio of the secrecy-good lattice. In this paper, we employ mutual information, rather than via the flatness factor, to directly bound information leakage, thereby removing that requirement of the secrecy-good lattice.

thanks to their useful structures [16], [25]; thus the proposed scheme may fit better with this landscape when security is a concern [26].

### C. Outline of the paper

The paper is organized as follows: Section II presents some preliminaries of lattice codes. The binary polar codes and multilevel lattice structure [27] are briefly reviewed in Section III, where the original polar wiretap coding scheme in [8] is slightly modified to be compatible to the following shaping operation. In Section IV, we construct secrecy-good polar lattices for the mod- $\Lambda_s$  GWC. In Section V, we show how to implement the discrete Gaussian shaping over the polar lattice to remove the mod- $\Lambda_s$  front end, using the polar coding technique for asymmetric channels. Then we prove that our wiretap lattice coding achieves the secrecy capacity with shaping. Finally, we discuss the relationship between the lattice constructions with and without shaping in Section VI.

### D. Notation

All random variables (RVs) will be denoted by capital letters. Let  $P_X$  denote the probability distribution of a RV  $X$  taking values  $x$  in a set  $\mathcal{X}$  and let  $H(X)$  denote its entropy. For multilevel coding, we denote by  $X_\ell$  a RV  $X$  at level  $\ell$ . The  $i$ -th realization of  $X_\ell$  is denoted by  $x_\ell^i$ . We also use the notation  $x_\ell^{i:j}$  as a shorthand for a vector  $(x_\ell^i, \dots, x_\ell^j)$ , which is a realization of RVs  $X_\ell^{i:j} = (X_\ell^i, \dots, X_\ell^j)$ . Similarly,  $x_{\ell:j}^i$  will denote the realization of the  $i$ -th RVs from level  $\ell$  to level  $j$ , i.e., of  $X_{\ell:j}^i = (X_\ell^i, \dots, X_j^i)$ . For a set  $\mathcal{I}$ ,  $\mathcal{I}^c$  denotes its complement set, and  $|\mathcal{I}|$  represents its cardinality. For an integer  $N$ ,  $[N]$  will be used to denote the set of all integers from 1 to  $N$ . A binary memoryless asymmetric (BMA) channel and a binary memoryless symmetric (BMS) channel will be denoted by  $W$  and  $\widetilde{W}$ , respectively. Following the notation of [7], we denote  $N$  independent uses of channel  $W$  by  $W^N$ . By channel combining and splitting, we get the combined channel  $W_N$  and the  $i$ -th subchannel  $W_N^{(i)}$ . Specifically, for a channel  $W_\ell$  at level  $\ell$ ,  $W_\ell^N$ ,  $W_{\ell,N}$  and  $W_\ell^{(i,N)}$  are used to denote its  $N$  independent expansion, the combined channel and the  $i$ -th subchannel after polarization. An indicator function is represented by  $\mathbb{1}(\cdot)$ . Throughout this paper, we use the binary logarithm, denoted by  $\log$ , and information is measured in bits.

## II. PRELIMINARIES OF LATTICE CODES

### A. Definitions

A lattice is a discrete subgroup of  $\mathbb{R}^n$  which can be described by

$$\Lambda = \{\lambda = Bx : x \in \mathbb{Z}^n\},$$

where  $B$  is an  $n$ -by- $n$  lattice generator matrix and we always assume that it has full rank in this paper.

For a vector  $x \in \mathbb{R}^n$ , the nearest-neighbor quantizer associated with  $\Lambda$  is  $Q_\Lambda(x) = \arg \min_{\lambda \in \Lambda} \|\lambda - x\|$ . We define the modulo lattice operation by  $x \bmod \Lambda \triangleq x - Q_\Lambda(x)$ . The Voronoi region of  $\Lambda$ , defined by  $\mathcal{V}(\Lambda) = \{x : Q_\Lambda(x) = 0\}$ , specifies the nearest-neighbor decoding region. The Voronoi cell is one example of fundamental region of the lattice. A measurable set  $\mathcal{R}(\Lambda) \subset \mathbb{R}^n$  is a fundamental region of the lattice  $\Lambda$  if  $\cup_{\lambda \in \Lambda} (\mathcal{R}(\Lambda) + \lambda) = \mathbb{R}^n$  and if

$(\mathcal{R}(\Lambda) + \lambda) \cap (\mathcal{R}(\Lambda) + \lambda')$  has measure 0 for any  $\lambda \neq \lambda'$  in  $\Lambda$ . The volume of a fundamental region is equal to that of the Voronoi region  $\mathcal{V}(\Lambda)$ , which is given by  $\text{Vol}(\Lambda) = |\det(B)|$ .

The theta series of  $\Lambda$  (see, e.g., [28, p.70]) is defined as

$$\Theta_{\Lambda}(\tau) = \sum_{\lambda \in \Lambda} e^{-\pi\tau\|\lambda\|^2}, \quad \tau > 0.$$

In this paper, the reliability condition for Bob is measured by the block error probability  $P_e(\Lambda, \sigma^2)$  of lattice decoding. It is the probability  $\Pr\{x \notin \mathcal{V}(\Lambda)\}$  that an  $n$ -dimensional independent and identically distributed (i.i.d.) Gaussian noise vector  $x$  with zero mean and variance  $\sigma^2$  per dimension falls outside the Voronoi region  $\mathcal{V}(\Lambda)$ . For an  $n$ -dimensional lattice  $\Lambda$ , define the volume-to-noise ratio (VNR) of  $\Lambda$  by

$$\gamma_{\Lambda}(\sigma) \triangleq \frac{\text{Vol}(\Lambda)^{\frac{2}{n}}}{\sigma^2}.$$

Then we introduce the notion of lattices which are good for the AWGN channel without power constraint.

*Definition 1 (AWGN-good lattices):* A sequence of lattices  $\Lambda_b$  of increasing dimension  $n$  is AWGN-good if, for any fixed  $P_e(\Lambda_b, \sigma^2) \in (0, 1)$ ,  $\lim_{n \rightarrow \infty} \gamma_{\Lambda_b}(\sigma) = 2\pi e$ , and if, for any fixed VNR greater than  $2\pi e$ ,

$$\lim_{n \rightarrow \infty} P_e(\Lambda_b, \sigma^2) = 0.$$

It is worth mentioning here that we do not insist on exponentially vanishing error probabilities, unlike Poltyrev's original treatment of good lattices for coding over the AWGN channel [29]. This is because a sub-exponential or polynomial decay of the error probability is often good enough.

Next, we introduce the notion of secrecy-good lattices. For this purpose, we need the capacity  $C(\Lambda_e, \sigma^2)$  of the mod- $\Lambda_e$  channel, which will be defined in (9).

*Definition 2 (Secrecy-good lattices):* A sequence of lattices  $\Lambda_e$  of increasing dimension  $n$  is secrecy-good if, for any fixed VNR of  $\Lambda_e$  smaller than  $2\pi e$ , the channel capacity  $C(\Lambda_e, \sigma^2)$  vanishes:

$$\lim_{n \rightarrow \infty} C(\Lambda_e, \sigma^2) = 0.$$

Note that this definition is different from that in [6], which is based on the flatness factor associated with the lattice Gaussian distribution. We will show that this definition is also sufficient to guarantee vanishing information leakage (see Remark 3).

### B. Flatness factor and lattice Gaussian distribution

For  $\sigma > 0$  and  $c \in \mathbb{R}^n$ , the Gaussian distribution of mean  $c$  and variance  $\sigma^2$  is defined as

$$f_{\sigma, c}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{\|x-c\|^2}{2\sigma^2}},$$

for all  $x \in \mathbb{R}^n$ . For convenience, let  $f_{\sigma}(x) = f_{\sigma, 0}(x)$ .

Given lattice  $\Lambda$ , we define the  $\Lambda$ -periodic function

$$f_{\sigma, \Lambda}(x) = \sum_{\lambda \in \Lambda} f_{\sigma, \lambda}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \sum_{\lambda \in \Lambda} e^{-\frac{\|x-\lambda\|^2}{2\sigma^2}},$$

for  $x \in \mathbb{R}^n$ .

The flatness factor is defined for a lattice  $\Lambda$  as [6]

$$\epsilon_\Lambda(\sigma) \triangleq \max_{x \in \mathcal{R}(\Lambda)} |\text{Vol}(\Lambda) f_{\sigma, \Lambda}(x) - 1|.$$

It can be interpreted as the maximum variation of  $f_{\sigma, \Lambda}(x)$  from the uniform distribution over  $\mathcal{R}(\Lambda)$ . The flatness factor can be calculated using the theta series [6]:

$$\epsilon_\Lambda(\sigma) = \left( \frac{\gamma_\Lambda(\sigma)}{2\pi} \right)^{\frac{n}{2}} \Theta_\Lambda \left( \frac{1}{2\pi\sigma^2} \right) - 1.$$

We define the *discrete Gaussian distribution* over  $\Lambda$  centered at  $c \in \mathbb{R}^n$  as the following discrete distribution taking values in  $\lambda \in \Lambda$ :

$$D_{\Lambda, \sigma, c}(\lambda) = \frac{f_{\sigma, c}(\lambda)}{f_{\sigma, c}(\Lambda)}, \quad \forall \lambda \in \Lambda,$$

where  $f_{\sigma, c}(\Lambda) \triangleq \sum_{\lambda \in \Lambda} f_{\sigma, c}(\lambda) = f_{\sigma, \Lambda}(c)$ . Again for convenience, we write  $D_{\Lambda, \sigma} = D_{\Lambda, \sigma, 0}$ .

It is also useful to define the discrete Gaussian distribution over a coset of  $\Lambda$ , i.e., the shifted lattice  $\Lambda - c$ :

$$D_{\Lambda - c, \sigma}(\lambda - c) = \frac{f_{\sigma}(\lambda - c)}{f_{\sigma, c}(\Lambda)}, \quad \forall \lambda \in \Lambda.$$

Note the relation  $D_{\Lambda - c, \sigma}(\lambda - c) = D_{\Lambda, \sigma, c}(\lambda)$ , namely, they are a shifted version of each other.

Each component of a lattice point sampled from  $D_{\Lambda - c, \sigma}$  has an average power always less than  $\sigma^2$  by the following lemma.

*Lemma 1 (Average power of lattice Gaussian [30, Lemma 1]):* Let  $x = (x_1, x_2, \dots, x_n)^T \sim D_{\Lambda - c, \sigma}$ . Then, for each  $1 \leq i \leq n$ ,

$$E[x_i^2] \leq \sigma^2. \quad (2)$$

If the flatness factor is negligible, the discrete Gaussian distribution over a lattice preserves the capacity of the AWGN channel.

*Theorem 1 (Mutual information of discrete Gaussian distribution [30, Th. 2]):* Consider an AWGN channel  $Y = X + E$  where the input constellation  $X$  has a discrete Gaussian distribution  $D_{\Lambda - c, \sigma_s}$  for arbitrary  $c \in \mathbb{R}^n$ , and where the variance of the noise  $E$  is  $\sigma^2$ . Let the average signal power be  $P_s$  so that  $\text{SNR} = P_s/\sigma^2$ , and let  $\tilde{\sigma} \triangleq \frac{\sigma_s \sigma}{\sqrt{\sigma_s^2 + \sigma^2}}$ . Then, if  $\epsilon = \epsilon_\Lambda(\tilde{\sigma}) < \frac{1}{2}$  and  $\frac{\pi \epsilon_t}{1 - \epsilon_t} \leq \epsilon$  where

$$\epsilon_t \triangleq \begin{cases} \epsilon_\Lambda \left( \sigma_s / \sqrt{\frac{\pi}{\pi - t}} \right), & t \geq 1/e \\ (t^{-4} + 1) \epsilon_\Lambda \left( \sigma_s / \sqrt{\frac{\pi}{\pi - t}} \right), & 0 < t < 1/e \end{cases}$$

the discrete Gaussian constellation results in mutual information

$$I_D \geq \frac{1}{2} \log(1 + \text{SNR}) - \frac{5\epsilon}{n} \quad (3)$$

per channel use. Moreover, the difference between  $P_s$  and  $\sigma_s^2$  is bounded by

$$|P_s - \sigma_s^2| \leq \frac{2\pi \epsilon_t}{n(1 - \epsilon)} \sigma_s^2.$$

A lattice  $\Lambda$  or its coset  $\Lambda - c$  with a discrete Gaussian distribution is referred to as a *good constellation* for the AWGN channel if  $\epsilon_\Lambda(\tilde{\sigma})$  is negligible [30]. It is further proved in [30] that the channel capacity is achieved with

Gaussian shaping over an AWGN-good lattice and minimum mean square error (MMSE) lattice decoding. Following Theorem 1, it has been shown in [15] that an AWGN-good polar lattice shaped according to the discrete Gaussian distribution achieves the AWGN channel capacity with sub-exponentially vanishing error probability, which means that an explicit polar lattice satisfying the power constraint and the reliability condition for Bob is already in hand. Therefore, the next section will focus on the construction of the secrecy-good polar lattice.

### III. POLAR CODES AND POLAR LATTICES

#### A. Polar codes: brief review

We firstly recall some basics of polar codes. Let  $\widetilde{W}$  be a BMS channel with uniformly distributed input  $X \in \mathcal{X} = \{0, 1\}$  and output  $Y \in \mathcal{Y}$ . The input distribution and transition probability of  $\widetilde{W}$  are denoted by  $P_X$  and  $P_{Y|X}$  respectively. Let  $X^{[N]}$  and  $Y^{[N]}$  be the input and output vector of  $N$  independent uses of  $\widetilde{W}$ . Let  $N = 2^m$  be the block length of polar codes for some integer  $m \geq 1$ . The channel polarization is based on the  $N$ -by- $N$  transform  $U^{[N]} = X^{[N]}G_N$ , where  $G_N = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{\otimes m}$  is the generator matrix and  $\otimes$  denotes the Kronecker product. Then we get an  $N$ -dimensional combined channel  $\widetilde{W}_N$  from  $U^{[N]}$  to  $Y^{[N]}$ . For each  $i \in [N]$ , given the previous bits  $U^{1:i-1}$ , the channel  $\widetilde{W}_N^{(i)}$  seen by each bit  $U^i$  is called the  $i$ -th subchannel channel after the channel splitting process [7], and the transition probability of  $\widetilde{W}_N^{(i)}$  is given by

$$\widetilde{W}_N^{(i)}(y^{[N]}, u^{1:i-1}|u^i) = \sum_{u^{i+1:N} \in \mathcal{X}^{N-i}} \frac{1}{2^{N-1}} \widetilde{W}_N(y^{[N]}|u^{[N]}),$$

where  $u^{[N]}$  and  $y^{[N]}$  are the realizations of  $U^{[N]}$  and  $Y^{[N]}$ , respectively. Arıkan proved that  $\widetilde{W}_N^{(i)}$  is also a BMS channel and it becomes either an almost error-free channel or a completely useless channel as  $N$  grows. According to [7], the goodness of a BMS channel can be estimated by its associate Bhattacharyya parameter, which is defined as follows.

*Definition 3 (Bhattacharyya parameter of BMS channels):* Let  $\widetilde{W}$  be a BMS channel with transition probability  $P_{Y|X}$ , the symmetric Bhattacharyya parameter  $\widetilde{Z} \in [0, 1]$  is defined as

$$\widetilde{Z}(\widetilde{W}) \triangleq \sum_y \sqrt{P_{Y|X}(y|0)P_{Y|X}(y|1)}.$$

**Remark 1.** Although polar codes were originally proposed for binary-input discrete memoryless channels [7], their extension to continuous channels, such as the binary-input AWGN channel, was given in [31]. To construct polar codes efficiently, the authors proposed smart channel degrading and upgrading merging algorithms to quantize continuous channels into their discrete versions. Fortunately, the quantization accuracy can be made arbitrarily small by increasing the quantization level. For this reason, we still use the summation form of Bhattacharyya parameters for continuous channels in this work, which also makes the notations consistent with the literature on polar codes.

It was further shown in [32], [33] that for any  $0 < \beta < \frac{1}{2}$ ,

$$\begin{aligned} \lim_{m \rightarrow \infty} \frac{1}{N} \left| \{i : \widetilde{Z}(\widetilde{W}_N^{(i)}) < 2^{-N^\beta}\} \right| &= I(\widetilde{W}) \\ \lim_{m \rightarrow \infty} \frac{1}{N} \left| \{i : \widetilde{Z}(\widetilde{W}_N^{(i)}) > 1 - 2^{-N^\beta}\} \right| &= 1 - I(\widetilde{W}), \end{aligned}$$

which means the proportion of such roughly error-free subchannels (with negligible Bhattacharyya parameters) approaches the channel capacity  $I(\widetilde{W})$ . The set of the indices of all those almost error-free subchannels is usually called the information set  $\mathcal{I}$  and its complementary is called the frozen set  $\mathcal{F}$ . Consequently, the construction of capacity-achieving polar codes is simply to identify the indices in the information set  $\mathcal{I}$ . However, for a general BMS channel other than binary erasure channel, the complexity of the exact computation for  $\widetilde{Z}(\widetilde{W}_N^{(i)})$  appears to be exponential in the block length  $N$ . An efficient estimation method for  $\widetilde{Z}(\widetilde{W}_N^{(i)})$  was proposed in [31], using the idea of channel upgrading and degrading. It was shown that with a sufficient number of quantization levels, the approximation error is negligible even if  $\widetilde{W}$  has continuous output, and the involved computational complexity is acceptable.

In [7], a bit-wise decoding method called successive cancellation (SC) decoding was proposed to show that polar codes are able to achieve channel capacity with vanishing error probability. This decoding method has complexity  $O(N \log N)$ , and the error probability is given by  $P_e^{SC} \leq \sum_{i \in \mathcal{I}} \widetilde{Z}(\widetilde{W}_N^{(i)})$ .

### B. Polar codes for the binary symmetric wiretap channel

Now we revisit the construction of polar codes for the binary symmetric wiretap channel. We use  $\widetilde{V}$  and  $\widetilde{W}$  to denote the symmetric main channel between Alice and Bob and the symmetric wiretap channel between Alice and Eve, respectively. Both  $\widetilde{V}$  and  $\widetilde{W}$  have binary input  $\mathsf{X}$  and  $\widetilde{W}$  is degraded with respect to  $\widetilde{V}$ . Let  $\mathsf{Y}$  and  $\mathsf{Z}$  denote the output of  $\widetilde{V}$  and  $\widetilde{W}$ . After the channel combination and splitting of  $N$  independent uses of the  $\widetilde{V}$  and  $\widetilde{W}$  by the polarization transform  $U^{[N]} = \mathsf{X}^{[N]} G_N$ , we define the sets of reliability-good indices for Bob and information-poor indices for Eve as

$$\begin{aligned} \mathcal{G}(\widetilde{V}) &= \{i : \widetilde{Z}(\widetilde{V}_N^{(i)}) \leq 2^{-N^\beta}\}, \\ \mathcal{N}(\widetilde{W}) &= \{i : \widetilde{Z}(\widetilde{W}_N^{(i)}) \geq 1 - 2^{-N^\beta}\}, \end{aligned} \quad (4)$$

where  $0 < \beta < 0.5$  and  $\widetilde{V}_N^{(i)}$  ( $\widetilde{W}_N^{(i)}$ ) is the  $i$ -th subchannel of the main channel (wiretapper's channel) after polarization transform.

Note that in the seminal paper [8] of polar wiretap coding, the information-poor set  $\mathcal{N}(\widetilde{W})$  was defined as  $\{i : I(\widetilde{W}^{(i,N)}) \leq 2^{-N^\beta}\}$ . In contrast, our criterion here is based on the Bhattacharyya parameter<sup>3</sup>. This slight modification will bring us much convenience when lattice shaping is involved in Sect. V. The following lemma shows that the modified criterion is similar to the original one in the sense that the mutual information of the subchannels with indices in  $\mathcal{N}(\widetilde{W})$  can still be bounded in the same form.

*Lemma 2:* Let  $\widetilde{W}_N^{(i)}$  be the  $i$ -th subchannel after the polarization transform on independent  $N$  uses of a BMS channel  $\widetilde{W}$ . For any  $0 < \beta < \frac{1}{2}$  and  $\delta > 0$ , if  $\widetilde{Z}(\widetilde{W}_N^{(i)}) \geq 1 - 2^{-N^\beta}$ , the mutual information of the  $i$ -th subchannel can be upper-bounded as

$$I(\widetilde{W}_N^{(i)}) \leq 2^{-N^{\beta'}},$$

where  $\beta(1 - \delta) \leq \beta' \leq \beta$  when  $N$  is sufficiently large.

<sup>3</sup>This idea has already been used in [8] to prove that polar wiretap coding scheme is secrecy capacity-achieving.



*Proof.* When  $\widetilde{W}$  is symmetric,  $\widetilde{W}_N^{(i)}$  is symmetric as well. By [7, Proposition 1], we have

$$\begin{aligned} I(\widetilde{W}_N^{(i)}) &\leq \sqrt{1 - \widetilde{Z}(\widetilde{W}_N^{(i)})^2} \\ &\leq \sqrt{2 \cdot 2^{-N^\beta}} \\ &= 2^{-N^{\beta'}}, \end{aligned}$$

where  $\beta' < \beta$ . Moreover, for sufficiently large  $N$ ,  $\beta'$  can be made arbitrarily close to  $\beta$ , i.e.,  $\beta(1 - \delta) \leq \beta'$  for any  $\delta > 0$ .  $\square$

Since the mutual information of subchannels in  $\mathcal{N}(\widetilde{W})$  can be upper-bounded in the same form, it is not difficult to understand that strong secrecy can be achieved using the index partition proposed in [8]. Similarly, we divide the index set  $[N]$  into the following four sets:

$$\begin{aligned} \mathcal{A} &= \mathcal{G}(\widetilde{V}) \cap \mathcal{N}(\widetilde{W}), \quad \mathcal{B} = \mathcal{G}(\widetilde{V}) \cap \mathcal{N}(\widetilde{W})^c \\ \mathcal{C} &= \mathcal{G}(\widetilde{V})^c \cap \mathcal{N}(\widetilde{W}), \quad \mathcal{D} = \mathcal{G}(\widetilde{V})^c \cap \mathcal{N}(\widetilde{W})^c. \end{aligned} \tag{5}$$

Clearly,  $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C} \cup \mathcal{D} = [N]$ . Then we assign set  $\mathcal{A}$  with message bits  $M$ , set  $\mathcal{B}$  with uniformly random bits  $R_b$ , set  $\mathcal{C}$  with frozen bits  $F$  which are known to both Bob and Eve prior to transmission, and set  $\mathcal{D}$  with uniformly random bits  $R_d$ . The next lemma shows that this assignment achieves strong secrecy. We note that this proof is similar to that in [8], [9] and it is given in [34, Appendix A].

*Lemma 3:* According to the partitions of the index set shown in (5), if we assign the four sets as follows

$$\begin{aligned} \mathcal{A} &\leftarrow M, \quad \mathcal{B} \leftarrow R_b, \\ \mathcal{C} &\leftarrow F, \quad \mathcal{D} \leftarrow R_d, \end{aligned} \tag{6}$$

the information leakage  $I(M; Z^{[N]})$  can be upper-bounded as

$$I(M; Z^{[N]}) \leq N \cdot 2^{-N^{\beta'}}, 0 < \beta' < 0.5. \tag{7}$$

We can also observe that the proportion of the problematic set  $\mathcal{D}$  is arbitrarily small when  $N$  is sufficiently large. This is because set  $\mathcal{D}$  is a subset of the unpolarized set  $\{i : 2^{-N^\beta} < \widetilde{Z}(\widetilde{V}_N^{(i)}) < 1 - 2^{-N^\beta}\}$ . As has been shown in [8], the reliability condition cannot be fulfilled with SC decoding due to the existence of  $\mathcal{D}$ . Fortunately, we can use the Markov block coding technique proposed in [9] to achieve reliability and strong secrecy simultaneously. More details of this Markov block coding technique will be discussed in Section IV-B and Section V-D.

With regard to the secrecy rate, we show that the modified polar coding scheme can also achieve the secrecy capacity.

*Lemma 4:* Let  $C(\widetilde{V})$  and  $C(\widetilde{W})$  denote the channel capacity of the main channel  $\widetilde{V}$  and wiretap channel  $\widetilde{W}$  respectively. Since  $\widetilde{W}$  is degraded with respect to  $\widetilde{V}$ , the secrecy capacity, which is given by  $C(\widetilde{V}) - C(\widetilde{W})$ , is achievable using the modified wiretap coding scheme, i.e.,

$$\lim_{N \rightarrow \infty} |\mathcal{G}(\widetilde{V}) \cap \mathcal{N}(\widetilde{W})|/N = C(\widetilde{V}) - C(\widetilde{W}).$$

*Proof.* See [34, Appendix B].  $\square$

### C. From polar codes to polar lattices

A sublattice  $\Lambda' \subset \Lambda$  induces a partition (denoted by  $\Lambda/\Lambda'$ ) of  $\Lambda$  into equivalence classes modulo  $\Lambda'$ . The order of the partition is denoted by  $|\Lambda/\Lambda'|$ , which is equal to the number of cosets. If  $|\Lambda/\Lambda'| = 2$ , we call this a binary partition. Let  $\Lambda/\Lambda_1/\dots/\Lambda_{r-1}/\Lambda'$  for  $r \geq 1$  be an  $n$ -dimensional self-similar lattice partition chain<sup>4</sup>. For each partition  $\Lambda_{\ell-1}/\Lambda_\ell$  ( $1 \leq \ell \leq r$  with convention  $\Lambda_0 = \Lambda$  and  $\Lambda_r = \Lambda'$ ) a code  $C_\ell$  over  $\Lambda_{\ell-1}/\Lambda_\ell$  selects a sequence of representatives  $a_\ell$  for the cosets of  $\Lambda_\ell$ . Consequently, if each partition is binary, the code  $C_\ell$  is a binary code.

Polar lattices are constructed by ‘‘Construction D’’ [28, p.232] [27] using a set of nested polar codes  $C_1 \subseteq C_2 \dots \subseteq C_r$ . Suppose  $C_\ell$  has block length  $N$  and  $k_\ell$  information bits for  $1 \leq \ell \leq r$ . Choose a basis  $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_N$  from the polar generator matrix  $G_N$  such that  $\mathbf{g}_1, \dots, \mathbf{g}_{k_\ell}$  span  $C_\ell$ . When the dimension  $n = 1$ , we choose the partition chain  $\mathbb{Z}/2\mathbb{Z}/\dots/2^r\mathbb{Z}$ , then the lattice  $L$  admits the form [27]

$$L = \left\{ \sum_{\ell=1}^r 2^{\ell-1} \sum_{i=1}^{k_\ell} u_\ell^i \mathbf{g}_i + 2^r \mathbb{Z}^N \mid u_\ell^i \in \{0, 1\} \right\}, \quad (8)$$

where the addition is carried out in  $\mathbb{R}^N$ . The fundamental volume of a lattice obtained from this construction is given by

$$\text{Vol}(L) = 2^{-NR_C} \cdot \text{Vol}(\Lambda_r)^N,$$

where  $R_C = \sum_{\ell=1}^r R_\ell = \frac{1}{N} \sum_{\ell=1}^r k_\ell$  denotes the sum rate of component codes. In this paper, we limit ourselves to the one-dimensional binary lattice partition chain and binary polar codes for simplicity.

## IV. SECRECY-GOOD POLAR LATTICES FOR THE MOD- $\Lambda_s$ GWC

Before considering the Gaussian wiretap channel, we will tackle a simpler problem of constructing secrecy-good polar lattices over the mod- $\Lambda_s$  GWC shown in Fig. 2. The difference between the mod- $\Lambda_s$  GWC and the genuine GWC is the mod- $\Lambda_s$  operation on the received signal of Bob and Eve. We will assume uniform input messages until we discuss semantic security in the end of this section.

### A. Strong secrecy

With some abuse of notation, the outputs  $\mathbf{Y}^{[N]}$  and  $\mathbf{Z}^{[N]}$  at Bob and Eve’s ends respectively become

$$\begin{cases} \mathbf{Y}^{[N]} = \left[ \mathbf{X}^{[N]} + \mathbf{W}_b^{[N]} \right] \bmod \Lambda_s, \\ \mathbf{Z}^{[N]} = \left[ \mathbf{X}^{[N]} + \mathbf{W}_e^{[N]} \right] \bmod \Lambda_s. \end{cases}$$

The idea of wiretap lattice coding over the mod- $\Lambda_s$  GWC [6] can be explained as follows. Let  $\Lambda_b$  and  $\Lambda_e$  be the AWGN-good lattice and secrecy-good lattice designed for Bob and Eve accordingly. Let  $\Lambda_s \subset \Lambda_e \subset \Lambda_b$  be a nested chain of  $N$ -dimensional lattices in  $\mathbb{R}^N$ , where  $\Lambda_s$  is the shaping lattice. Note that the shaping lattice  $\Lambda_s$  here is employed primarily for the convenience of designing the secrecy-good lattice and secondarily for satisfying the power constraint. Consider a one-to-one mapping:  $\mathcal{M} \rightarrow \Lambda_b/\Lambda_e$  which associates each message  $m \in \mathcal{M}$  to a coset

<sup>4</sup>By saying self-similar, we mean that  $\Lambda_\ell = T^\ell \Lambda$  for all  $\ell$ , with  $T = \alpha V$  for some scale factor  $\alpha > 1$  and orthogonal matrix  $V$ . For example,  $\mathbb{Z}/2\mathbb{Z}/\dots/2^r\mathbb{Z}$  is a one-dimensional self-similar partition chain.

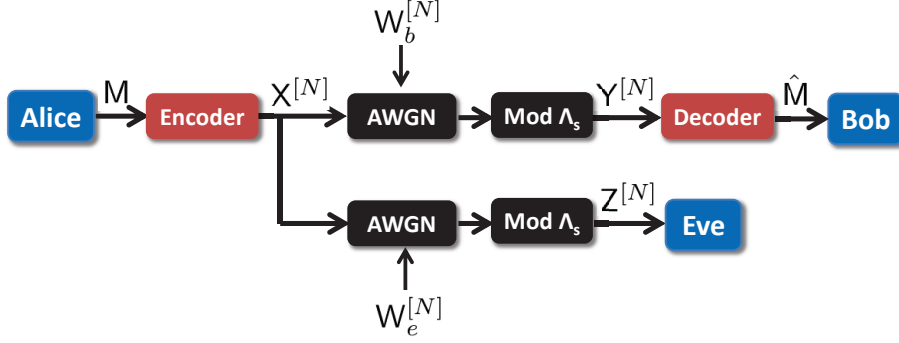


Fig. 2. The mod- $\Lambda_s$  Gaussian wiretap channel.

$\tilde{\lambda}_m \in \Lambda_b/\Lambda_e$ . Alice selects a lattice point  $\lambda \in \Lambda_e \cap \mathcal{V}(\Lambda_s)$  uniformly at random and transmits  $X^{[N]} = \lambda + \lambda_m$ , where  $\lambda_m$  is the coset representative of  $\tilde{\lambda}_m$  in  $\mathcal{V}(\Lambda_e)$ . This scheme has been proved to achieve both reliability and semantic security in [6] by random lattice codes. We will make it explicit by constructing polar lattice codes in this section.

Let  $\Lambda_b$  and  $\Lambda_e$  be constructed from a binary partition chain  $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda_r$ , and assume  $\Lambda_s \subset \Lambda_r^N$  such that  $\Lambda_s \subset \Lambda_r^N \subset \Lambda_e \subset \Lambda_b^5$ . Also, denote by  $X_{1:r}^{[N]}$  the bits encoding  $\Lambda^N/\Lambda_r^N$ , which include all information bits for message  $M$  as a subset. We have that  $[X^{[N]} + W_e^{[N]}] \bmod \Lambda_r^N$  is a sufficient statistic for  $X_{1:r}^{[N]}$ . This can be seen from [27, Lemma 8], rewritten as follows:

*Lemma 5 (Sufficiency of mod- $\Lambda$  output [27]):* For a partition chain  $\Lambda/\Lambda'$  ( $\Lambda' \subset \Lambda$ ), let the input of an AWGN channel be  $X = A + B$ , where  $A \in \mathcal{R}(\Lambda)$  is a random variable, and  $B$  is uniformly distributed in  $\Lambda \cap \mathcal{R}(\Lambda')$ . Reduce the output  $Y$  first to  $Y' = Y \bmod \Lambda'$  and then to  $Y'' = Y' \bmod \Lambda$ . Then the mod- $\Lambda$  map is information-lossless, namely  $I(A; Y') = I(A; Y'')$ , which means that the output  $Y'' = Y' \bmod \Lambda$  of mod- $\Lambda$  map is a sufficient statistic for  $A$ .

In our context, we identify  $\Lambda$  with  $\Lambda_r^N$  and  $\Lambda'$  with  $\Lambda_s$ , respectively. Since the bits encoding  $\Lambda_r^N/\Lambda_s$  are uniformly distributed<sup>6</sup>, the mod- $\Lambda_r^N$  operation is information-lossless in the sense that

$$I(X_{1:r}^{[N]}; Z^{[N]}) = I(X_{1:r}^{[N]}; [X^{[N]} + W_e^{[N]}] \bmod \Lambda_r^N).$$

As far as mutual information  $I(X_{1:r}^{[N]}; Z^{[N]})$  is concerned, we can use the mod- $\Lambda_r^N$  operator instead of the mod- $\Lambda_s$  operator here. Under this condition, we use the multilevel lattice structure introduced in [27] to decompose the mod- $\Lambda_s$  channel into a series of BMS channels according to the partition chain  $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda_r$ . Therefore, the afore-mentioned polar coding technique for BMS channels can be employed. Moreover, the channel resulted from the lattice partition chain can be proved to be equivalent to that based on the chain rule of mutual information

<sup>5</sup>This is always possible with sufficient power, since the power constraint is not our primary concern in this section. We can scale  $\Lambda_s$  as large as possible to make  $\Lambda_s \subset \Lambda_r^N$ .

<sup>6</sup>In fact, all bits encoding  $\Lambda_e/\Lambda_s$  are uniformly distributed in wiretap coding.

(See (11)). Following this channel equivalence, we can construct an AWGN-good lattice  $\Lambda_b$  and a secrecy-good lattice  $\Lambda_e$ , using the wiretap coding technique (4) at each partition level.

A mod- $\Lambda$  channel is a Gaussian channel with a modulo- $\Lambda$  operator in the front end [27], [35]. The capacity of the mod- $\Lambda$  channel is [27]

$$C(\Lambda, \sigma^2) = \log(\text{Vol}(\Lambda)) - h(\Lambda, \sigma^2), \quad (9)$$

where  $h(\Lambda, \sigma^2)$  is the differential entropy of the  $\Lambda$ -aliased noise over  $\mathcal{R}(\Lambda)$ :

$$h(\Lambda, \sigma^2) = - \int_{\mathcal{R}(\Lambda)} f_{\sigma, \Lambda}(t) \log f_{\sigma, \Lambda}(t) dt.$$

The differential entropy reaches its maximum  $\log(\text{Vol}(\Lambda))$  by the uniform distribution over  $\mathcal{R}(\Lambda)$ . The  $\Lambda_{\ell-1}/\Lambda_\ell$  channel is defined as a mod- $\Lambda_\ell$  channel whose input is drawn from  $\Lambda_{\ell-1} \cap \mathcal{R}(\Lambda_\ell)$ . It is known that the  $\Lambda_{\ell-1}/\Lambda_\ell$  channel is symmetric<sup>7</sup>, and the optimum input distribution is uniform [27]. Furthermore, the  $\Lambda_{\ell-1}/\Lambda_\ell$  channel is binary if  $|\Lambda_{\ell-1}/\Lambda_\ell| = 2$ . The capacity of the  $\Lambda_{\ell-1}/\Lambda_\ell$  channel for Gaussian noise of variance  $\sigma^2$  is given by [27]

$$\begin{aligned} C(\Lambda_{\ell-1}/\Lambda_\ell, \sigma^2) &= C(\Lambda_\ell, \sigma^2) - C(\Lambda_{\ell-1}, \sigma^2) \\ &= h(\Lambda_{\ell-1}, \sigma^2) - h(\Lambda_\ell, \sigma^2) + \log(\text{Vol}(\Lambda_\ell)/\text{Vol}(\Lambda_{\ell-1})). \end{aligned}$$

The decomposition into a set of  $\Lambda_{\ell-1}/\Lambda_\ell$  channels is used in [27] to construct AWGN-good lattices. Take the partition chain  $\mathbb{Z}/2\mathbb{Z}/\dots/2^r\mathbb{Z}$  as an example. Given uniform input  $X_{1:r}$ , let  $\mathcal{K}_\ell$  denote the coset indexed by  $x_{1:\ell}$ , i.e.,  $\mathcal{K}_\ell = x_1 + \dots + 2^{\ell-1}x_\ell + 2^\ell\mathbb{Z}$ . Given that  $X_{1:\ell-1} = x_{1:\ell-1}$ , the conditional probability distribution function (PDF) of this channel with binary input  $X_\ell$  and output  $\bar{Z} = Z \bmod \Lambda_\ell$  is

$$f_{\bar{Z}|X_\ell}(\bar{z}|x_\ell) = \frac{1}{\sqrt{2\pi}\sigma_e} \sum_{a \in \mathcal{K}_\ell(x_{1:\ell})} \exp\left(-\frac{1}{2\sigma_e^2}\|\bar{z} - a\|^2\right). \quad (10)$$

Since the previous input bits  $x_{1:\ell-1}$  cause a shift on  $\mathcal{K}_\ell$  and will be removed by the multistage decoder at level  $\ell$ , the code can be designed according to the channel transition probability (10) with  $x_{1:\ell-1} = 0$ . Following the notation of [27], we use  $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$  and  $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$  to denote the  $\Lambda_{\ell-1}/\Lambda_\ell$  channel for Bob and Eve respectively. The  $\Lambda_{\ell-1}/\Lambda_\ell$  channel can also be used to construct secrecy-good lattices. In order to bound the information leakage of the wiretapper's channel, we firstly express  $I(X_{1:r}; Z)$  according to the chain rule of mutual information as

$$I(X_{1:r}; Z) = I(X_1; Z) + I(X_2; Z|X_1) + \dots + I(X_r; Z|X_{1:r-1}). \quad (11)$$

This equation still holds if  $Z$  denotes the noisy signal after the mod- $\Lambda_r$  operation, namely,  $Z = [X+W_e] \bmod \Lambda_r$ . We will adopt this notation in the rest of this subsection. We refer to the  $\ell$ -th channel associated with mutual information  $I(X_\ell; Z|X_{1:\ell-1})$  as the equivalent channel denoted by  $W'(X_\ell; Z|X_{1:\ell-1})$ , which is defined as the channel from  $X_\ell$  to  $Z$  given the previous  $X_{1:\ell-1}$ . Then the transition probability distribution of  $W'(X_\ell; Z|X_{1:\ell-1})$  is [27, Lemma 6]

$$\begin{aligned} f_{Z|X_\ell}(z|x_\ell) &= \frac{1}{\Pr(\mathcal{K}_\ell(x_{1:\ell}))} \sum_{a \in \mathcal{K}_\ell(x_{1:\ell})} \Pr(a) f_Z(z|a) \\ &= \frac{1}{|\Lambda_\ell/\Lambda_r|} \frac{1}{\sqrt{2\pi}\sigma_e} \sum_{a \in \mathcal{K}_\ell(x_{1:\ell})} \exp\left(-\frac{1}{2\sigma_e^2}\|z - a\|^2\right), \quad z \in \mathcal{V}(\Lambda_r). \end{aligned} \quad (12)$$

<sup>7</sup>This is ‘‘regular’’ in the sense of Delsarte and Piret and symmetric in the sense of Gallager [27].

From (10) and (12), we can observe that the channel output likelihood ratio (LR) of the  $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$  channel is equal to that of the  $\ell$ -th equivalent channel  $W'(X_\ell; Z|X_{1:\ell-1})$ . Then we have the following channel equivalence lemma.

*Lemma 6:* Consider a lattice  $L$  constructed by a binary lattice partition chain  $\Lambda/\Lambda_1/\dots/\Lambda_{r-1}/\Lambda_r$ . Constructing a polar code for the  $\ell$ -th equivalent binary-input channel  $W'(X_\ell; Z|X_{1:\ell-1})$  defined by the chain rule (11) is equivalent to constructing a polar code for the  $\Lambda_{\ell-1}/\Lambda_\ell$  channel  $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ , i.e., the mutual information and Bhattacharyya parameters of the subchannels resulted from  $W'(X_\ell; Z|X_{1:\ell-1})$  are equivalent to that of the subchannels resulted from  $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ , respectively.

*Proof.* See Appendix C. □

Note that another proof based on direct calculation of the mutual information and Bhattacharyya parameters of the subchannels can be found in [36].

**Remark 2.** Observe that if we define  $V'(X_\ell; Y|X_{1:\ell-1})$  as the equivalent channel according to the chain rule expansion of  $I(X; Y)$  for the main channel, the same result can be obtained between  $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$  and  $V'(X_\ell; Y|X_{1:\ell-1})$ . Moreover, this lemma also holds without the mod- $\Lambda_s$  front-end, i.e., without power constraint. The construction of AWGN-good polar lattices was given in [15], where nested polar codes were constructed based on a set of  $\Lambda_{\ell-1}/\Lambda_\ell$  channels. We note that the  $\Lambda_{\ell-1}/\Lambda_\ell$  channel is degraded with respect to the  $\Lambda_\ell/\Lambda_{\ell+1}$  channel [15, Lemma 3].

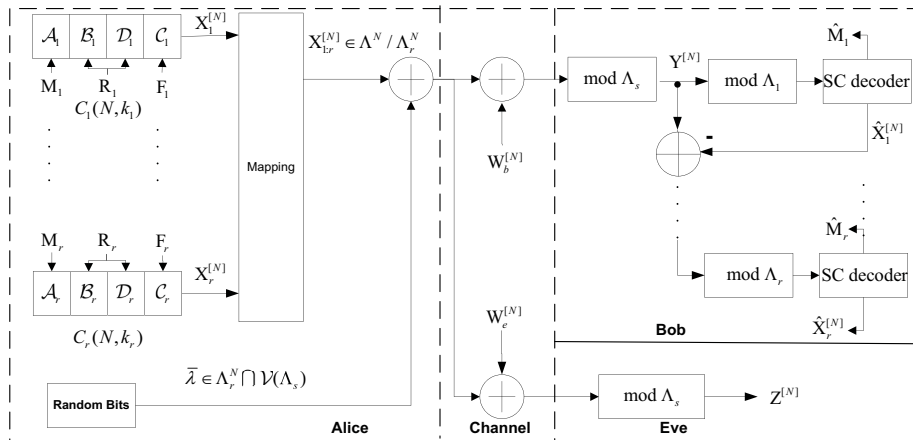


Fig. 3. The multilevel lattice coding system over the mod- $\Lambda_s$  Gaussian wiretap channel.

Now we are ready to introduce the polar lattice construction for the mod- $\Lambda_s$  GWC shown in Fig. 3. A polar lattice  $L$  is constructed by a series of nested polar codes  $C_1(N, k_1) \subseteq C_2(N, k_2) \subseteq \dots \subseteq C_r(N, k_r)$  and a binary lattice partition chain  $\Lambda/\Lambda_1/\dots/\Lambda_r$ . The block length of polar codes is  $N$ . Alice splits the message  $M$  into  $M_1, \dots, M_r$ . We follow the same rule (6) to assign bits in the component polar codes to achieve strong secrecy. Note that  $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$  is degraded with respect to  $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$  for  $1 \leq \ell \leq r$  because  $\sigma_b^2 \leq \sigma_e^2$ . Treating

$V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$  and  $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$  as the main channel and wiretapper's channel at each level and using the partition rule (5), we can get four sets  $\mathcal{A}_\ell, \mathcal{B}_\ell, \mathcal{C}_\ell$  and  $\mathcal{D}_\ell$ . Similarly, we assign the bits as follows

$$\begin{aligned} \mathcal{A}_\ell &\leftarrow M_\ell, & \mathcal{B}_\ell &\leftarrow R_\ell^b, \\ \mathcal{C}_\ell &\leftarrow F_\ell, & \mathcal{D}_\ell &\leftarrow R_\ell^d \end{aligned} \quad (13)$$

for each level  $\ell$ , where  $M_\ell, F_\ell$  and  $R_\ell^b$  ( $R_\ell^d$ ) represent message bits, frozen bits (could be set as all zeros) and uniformly random bits for set  $\mathcal{B}_\ell$  ( $\mathcal{D}_\ell$ ) at level  $\ell$ . Since the  $\Lambda_{\ell-1}/\Lambda_\ell$  channel is degraded with respect to the  $\Lambda_\ell/\Lambda_{\ell+1}$  channel. According to [33, Lemma 4.7], when a BMS channel  $\tilde{W}$  is degraded with respect to a BMS channel  $\tilde{V}$ , the Bhattacharyya parameters of the subchannels satisfy  $\tilde{Z}(\tilde{W}_N^{(i)}) \geq \tilde{Z}(\tilde{V}_N^{(i)})$ . Thus, it is easy to obtain that  $\mathcal{C}_\ell \supseteq \mathcal{C}_{\ell+1}$ , which means  $\mathcal{A}_\ell \cup \mathcal{B}_\ell \cup \mathcal{D}_\ell \subseteq \mathcal{A}_{\ell+1} \cup \mathcal{B}_{\ell+1} \cup \mathcal{D}_{\ell+1}$ . This construction is clearly a lattice construction as polar codes constructed for each level are nested. We skip the proof of nested polar codes here. A similar proof can be found in [14] and [15].

As a result, the above multilevel construction yields an AWGN-good lattice  $\Lambda_b$  and a secrecy-good lattice  $\Lambda_e$  simultaneously. More precisely,  $\Lambda_b$  is constructed from a set of nested polar codes  $C_1(N, |\mathcal{A}_1| + |\mathcal{B}_1| + |\mathcal{D}_1|) \subseteq \dots \subseteq C_r(N, |\mathcal{A}_r| + |\mathcal{B}_r| + |\mathcal{D}_r|)$ , while  $\Lambda_e$  is constructed from a set of nested polar codes  $C_1(N, |\mathcal{B}_1| + |\mathcal{D}_1|) \subseteq \dots \subseteq C_r(N, |\mathcal{B}_r| + |\mathcal{D}_r|)$  and with the same lattice partition chain. Note that the random bits in set  $\mathcal{D}_\ell$  should be shared to Bob to guarantee the AWGN-goodness of  $\Lambda_b$ . More details will be given in the next subsection. It is clear that  $\Lambda_e \subset \Lambda_b$ . Thus, our proposed coding scheme instantiates the coset coding scheme introduced in [6], where the confidential message is mapped to the coset  $\tilde{\lambda}_m \in \Lambda_b/\Lambda_e$ . However, unlike the work of [6], our scheme does not require an asymptotically vanishing flatness factor, since the upper-bound of the information leakage can be calculated directly. The flatness factor will show up with the lattice Gaussian shaping in the next section.

By using the above assignments and Lemma 3, we have

$$I(M_\ell F_\ell; Z_\ell^{[N]}) \leq N2^{-N^{\beta'}}, \quad (14)$$

where  $Z_\ell^{[N]} = Z^{[N]} \bmod \Lambda_\ell$  is the output of the  $\Lambda_{\ell-1}/\Lambda_\ell$  channel for Eve. In other words, the employed polar code for the channel  $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$  can guarantee that the mutual information between the input message and the output is upper bounded by  $N2^{-N^{\beta'}}$ .

We assume uniform  $M_\ell$  and  $F_\ell$  such that  $X_\ell$  is uniformly distributed at each level. We will remove this restriction to the uniform distribution in Proposition 1. According to Lemma 6, the constructed polar code can also guarantee the same upper-bound on the mutual information between the input message and the output of the channel  $W'(X_\ell; Z|X_{1:\ell-1})$ , as shown in the following inequality ( $X_\ell$  is independent of the previous  $X_{1:\ell-1}$ ):

$$I(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}) \leq N2^{-N^{\beta'}}.$$

Recall that  $Z^{[N]}$  is the signal received by Eve after the mod- $\Lambda_r$  operation. Let  $F$  denote the combination of  $F_1, F_2, \dots, F_r$ . From the chain rule of mutual information, we obtain

$$\begin{aligned} &I(MF; Z^{[N]}) \\ &= \sum_{\ell=1}^r I(Z^{[N]}; M_\ell F_\ell | M_{1:\ell-1} F_{1:\ell-1}) \end{aligned}$$

$$\begin{aligned}
&= \sum_{\ell=1}^r H(\mathbf{M}_\ell \mathbf{F}_\ell | \mathbf{M}_{1:\ell-1} \mathbf{F}_{1:\ell-1}) - H(\mathbf{M}_\ell \mathbf{F}_\ell | \mathbf{Z}^{[N]}, \mathbf{M}_{1:\ell-1} \mathbf{F}_{1:\ell-1}) \\
&\leq \sum_{\ell=1}^r H(\mathbf{M}_\ell \mathbf{F}_\ell) - H(\mathbf{M}_\ell \mathbf{F}_\ell | \mathbf{Z}^{[N]}, \mathbf{M}_{1:\ell-1} \mathbf{F}_{1:\ell-1}) \\
&= \sum_{\ell=1}^r I(\mathbf{M}_\ell \mathbf{F}_\ell; \mathbf{Z}^{[N]}, \mathbf{M}_{1:\ell-1} \mathbf{F}_{1:\ell-1}) \\
&\leq \sum_{\ell=1}^r I(\mathbf{M}_\ell \mathbf{F}_\ell; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) \leq rN2^{-N^{\beta'}},
\end{aligned} \tag{15}$$

where the second inequality holds because  $I(\mathbf{M}_\ell \mathbf{F}_\ell; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) = I(\mathbf{M}_\ell \mathbf{F}_\ell; \mathbf{Z}^{[N]}, \mathbf{U}_{1:\ell-1}^{[N]})$  and adding more variables will not decrease the mutual information. Since  $\lim_{N \rightarrow \infty} I(\mathbf{M}\mathbf{F}; \mathbf{Z}^{[N]}) = 0$ , strong secrecy is achieved.

### B. Achieving secrecy capacity

In the original polar coding scheme for the binary wiretap channel [8], how to assign set  $\mathcal{D}$  is a problem. Assigning frozen bits to  $\mathcal{D}$  guarantees reliability but only achieves weak secrecy, whereas assigning random bits to  $\mathcal{D}$  guarantees strong secrecy but may violate the reliability requirement because  $\mathcal{D}$  may be nonempty. In order to ensure strong secrecy,  $\mathcal{D}$  is assigned with random bits ( $\mathcal{D} \leftarrow \mathbf{R}$ ), which makes this scheme failed to accomplish the theoretical reliability. In simple words, to satisfy the strong secrecy and reliability conditions simultaneously, the bits corresponding to  $\mathcal{D}$  must be kept frozen to Bob but uniformly random to Eve. For any  $\ell$ -th level channel  $V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$  at Bob's end, if set  $\mathcal{D}_\ell$  is fed with random bits, the probability of error is upper-bounded by the sum of the Bhattacharyya parameters  $\tilde{Z}(V_N^{(j)}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2))$  of subchannels that are not frozen to zero [7]. For each bit-channel index  $j$  and  $\beta < 0.5$ , we have

$$j \in \mathcal{G}(V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)) \cup \mathcal{D}_\ell.$$

By the definition (4), the sum of  $\tilde{Z}(V_N^{(j)}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2))$  over the set  $\mathcal{G}(V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2))$  is bounded by  $2^{-N^\beta}$ , therefore the error probability of the  $\ell$ -th level channel under the SC decoding, denoted by  $P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$ , can be upper-bounded by [7]

$$P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2) \leq N2^{-N^\beta} + \sum_{j \in \mathcal{D}_\ell} \tilde{Z}(V_N^{(j)}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)).$$

Since multistage decoding is utilized, by the union bound, the final decoding error probability for Bob is bounded as

$$\Pr\{\widehat{\mathbf{M}} \neq \mathbf{M}\} \leq \sum_{i=1}^r P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2).$$

Unfortunately, a bound on the sum  $\sum_{j \in \mathcal{D}_\ell} \tilde{Z}(V_N^{(j)}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2))$  is unavailable, making the proof of reliability out of reach. There is numerical evidence of low probabilities of error nonetheless. The proportion of  $\mathcal{D}_\ell$  vanishes as  $N \rightarrow \infty$  [8, Prop. 22]. In fact, numerical examples in [8, Sect. VI-F] showed that  $\mathcal{D}_\ell = \emptyset$  in most cases of interest. In any case, Bob can run some exhaustive search or form a small list of paths for those unreliable indexes.

The reliability problem was recently solved in [9], where a new scheme dividing the information message into several blocks was proposed. For a specific block,  $\mathcal{D}_\ell$  is still assigned with random bits and transmitted in advance

in the set  $\mathcal{A}_\ell$  of the previous block. This scheme involves negligible rate loss and finally realizes reliability and strong security simultaneously. In this case, if the reliability of each partition channel can be achieved, i.e., for any  $\ell$ -th level partition  $\Lambda_{\ell-1}/\Lambda_\ell$ ,  $P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)$  vanishes as  $N \rightarrow \infty$ , then the total decoding error probability for Bob can be made arbitrarily small. Consequently, based on this new scheme of assigning the problematic set, the error probability on level  $\ell$  can be upper-bounded by

$$P_e^{SC}(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2) \leq \epsilon_{N'}^\ell + k_\ell \cdot O(2^{-N'^\beta}), \quad (16)$$

where  $k_\ell$  is the number of information blocks on the  $\ell$ -th level,  $N'$  is the length of each block which satisfies  $N' \times k_\ell = N$  and  $\epsilon_{N'}^\ell$  is caused by the first separate block consisting of the initial bits in  $\mathcal{D}_\ell$  at the  $\ell$ -th level. Since  $|\mathcal{D}_\ell|$  is extremely small comparing to the block length  $N$ , the decoding failure probability for the first block can be made arbitrarily small when  $N$  is sufficiently large. Meanwhile, by the analysis in [15], when  $h(\Lambda, \sigma_b^2) \rightarrow \log(V(\Lambda))$ ,  $h(\Lambda_r, \sigma_b^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_b^2)$ , and  $R_C \rightarrow C(\Lambda/\Lambda_r, \sigma_b^2)$ , we have  $\gamma_{\Lambda_b}(\sigma_b) \rightarrow 2\pi e$ . Therefore,  $\Lambda_b$  is an AWGN-good lattice<sup>8</sup>.

Note that the rate loss incurred by repeatedly transmitted bits in  $\mathcal{D}_\ell$  is negligible because of its small size. Specifically, the actual secrecy rate in the  $\ell$ -th level is given by  $\frac{k_\ell}{k_\ell+1}[C(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2) - C(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)]$ . Clearly, this rate can be made close to the secrecy capacity by choosing sufficiently large  $k_\ell$  as well.

*Theorem 2 (Achieving secrecy capacity of the mod- $\Lambda_s$  GWC):* Consider a sequence of multi-level polar lattices  $L(N)$  of increasing dimensions  $N$ . Let  $L(N)$  be constructed according to (13) with the binary lattice partition chain  $\Lambda/\Lambda_1/\dots/\Lambda_r$  and  $r$  binary nested polar codes where  $r = O(\log N)$ . Scale the lattice partition chain to satisfy the following conditions:

- (i)  $\epsilon_\Lambda(\sigma_b) \rightarrow 0$ ,
- (ii)  $\epsilon_e = \frac{1}{2} \log(2\pi e \sigma_e^2) - h(\Lambda_r, \sigma_e^2) \rightarrow 0$ .

Given  $\sigma_e^2 > \sigma_b^2$ , the secrecy capacity  $\frac{1}{2} \log \frac{\sigma_e^2}{\sigma_b^2}$  of the mod- $\Lambda_s$  Gaussian wiretap channel is achievable by using the polar lattices  $L(N)$ , i.e., for any rate  $R < \frac{1}{2} \log \frac{\sigma_e^2}{\sigma_b^2}$ , there exists a sufficiently large  $N$  such that the realized rate  $R(N)$  of  $L(N)$  satisfies  $R(N) > R$ .

<sup>8</sup>More precisely, to make  $\Lambda_b$  AWGN-good, we need  $P_e(\Lambda_b, \sigma_b^2) \rightarrow 0$  by definition. By [15, Theorem 2],  $P_e(\Lambda_b, \sigma_b^2) \leq rN2^{-N^\beta} + N \cdot P_e(\Lambda_r, \sigma_b^2)$ . According to the analysis in Remark 6,  $r = O(\log N)$  is sufficient to guarantee  $P_e(\Lambda_r, \sigma_b^2) = e^{-\Omega(N)}$ , meaning that a sub-exponentially vanishing  $P_e(\Lambda_b, \sigma_b^2)$  can be achieved.



*Proof.* By Lemma 4 and (13),

$$\begin{aligned}
\lim_{N \rightarrow \infty} R(N) &= \sum_{\ell=1}^r \lim_{N \rightarrow \infty} \frac{|\mathcal{A}_\ell|}{N} \\
&= \sum_{\ell=1}^r C(V_\ell) - C(W_\ell) \\
&= \sum_{\ell=1}^r C(V(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_b^2)) - C(W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)) \\
&= C(V(\Lambda/\Lambda_r, \sigma_b^2)) - C(W(\Lambda/\Lambda_r, \sigma_e^2)) \\
&= C(\Lambda_r, \sigma_b^2) - C(\Lambda, \sigma_b^2) - C(\Lambda_r, \sigma_e^2) + C(\Lambda, \sigma_e^2) \\
&= h(\Lambda_r, \sigma_e^2) - h(\Lambda_r, \sigma_b^2) + h(\Lambda, \sigma_b^2) - h(\Lambda, \sigma_e^2) \\
&= \frac{1}{2} \log \frac{\sigma_e^2}{\sigma_b^2} - (\epsilon_e - \epsilon_b) - \epsilon_1,
\end{aligned} \tag{17}$$

where

$$\begin{cases} \epsilon_1 = C(\Lambda, \sigma_b^2) - C(\Lambda, \sigma_e^2) = h(\Lambda, \sigma_e^2) - h(\Lambda, \sigma_b^2) \geq 0, \\ \epsilon_b = h(\sigma_b^2) - h(\Lambda_r, \sigma_b^2) = \frac{1}{2} \log(2\pi e \sigma_b^2) - h(\Lambda_r, \sigma_b^2) \geq 0, \\ \epsilon_e = h(\sigma_e^2) - h(\Lambda_r, \sigma_e^2) = \frac{1}{2} \log(2\pi e \sigma_e^2) - h(\Lambda_r, \sigma_e^2) \geq 0 \end{cases}$$

and  $\epsilon_e - \epsilon_b \geq 0$ .

By scaling  $\Lambda$ , we can have  $h(\Lambda, \sigma_b^2) \rightarrow \log(\text{Vol}(\Lambda))$ . Since  $\sigma_e^2 > \sigma_b^2$ , we also have  $h(\Lambda, \sigma_e^2) \rightarrow \log(\text{Vol}(\Lambda))$ . More precisely, by [15, Lemma 1],  $\epsilon_1$  can be upper-bounded by the flatness factor as

$$\epsilon_1 \leq C(\Lambda, \sigma_b^2) \leq \log(e) \cdot \epsilon_\Lambda(\sigma_b).$$

Then, according to [6, Corollary 1], we can make  $\epsilon_\Lambda(\sigma_b) \rightarrow 0$  by scaling  $\Lambda$ .

The number of levels is set such that  $h(\Lambda_r, \sigma_e^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_e^2)$ . By [15, Theorem 2],  $r = O(\log N)$  is sufficient to guarantee  $P_e(\Lambda_r, \sigma_b^2) = e^{-\Omega(N)}$ , meaning that the volume  $\text{Vol}(\Lambda_r)$  is sufficiently large such that  $h(\Lambda_r, \sigma_e^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_e^2)$  as  $N \rightarrow \infty$ . Again, since  $\sigma_e^2 > \sigma_b^2$ , we immediately have  $h(\Lambda_r, \sigma_b^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_e^2)$ , and  $\epsilon_e - \epsilon_b \rightarrow 0$ . Therefore by scaling  $\Lambda$  and adjusting  $r$ , the secrecy rate can get arbitrarily close to  $\frac{1}{2} \log \frac{\sigma_e^2}{\sigma_b^2}$ .  $\square$

**Remark 3.** The constructed lattice  $\Lambda_e$  is secrecy-good in the sense of Definition 2. Recall that  $\Lambda_e$  is constructed from the partition chain  $\Lambda/\cdots/\Lambda_r$ , which gives us the  $N$ -dimensional partition chain  $\Lambda^N/\Lambda_e/\Lambda_r^N$ . Then,

$$\begin{aligned}
C(\Lambda_e, \sigma_e^2) &= C(\Lambda^N, \sigma_e^2) + C(\Lambda^N/\Lambda_e, \sigma_e^2) \\
&= C(\Lambda^N, \sigma_e^2) + I(\text{MF}; \mathbf{Z}^{[N]}) \\
&\leq \log(e) \cdot \epsilon_{\Lambda^N}(\sigma_e) + I(\text{MF}; \mathbf{Z}^{[N]}) \\
&\leq \log(e) \cdot ([1 + \epsilon_\Lambda(\sigma_e)]^N - 1) + I(\text{MF}; \mathbf{Z}^{[N]}),
\end{aligned}$$

where we use [6, Corollary 1] and [30, Lemma 3] in the last two inequalities, respectively.

Since  $r = O(\log N)$ , the top lattice  $\Lambda$  can be scaled down so that  $\epsilon_\Lambda(\sigma_e)$  vanishes as fast as  $O(2^{-\sqrt{N}})$  by [37, Proposition 2]. When  $N \rightarrow \infty$ , we have

$$C(\Lambda_e, \sigma_e^2) \leq N \log(e) \cdot \epsilon_\Lambda(\sigma_e) + I(\text{MF}; \mathbf{Z}^{[N]}) + O(2^{-\sqrt{N}}).$$

Recalling (15), we immediately have  $C(\Lambda_e, \sigma_e^2) \rightarrow 0$ .

Meanwhile, following the analysis of [15], we can show that the VNR  $\gamma_{\Lambda_e}(\sigma_e^2) \rightarrow 2\pi e$  from below. More precisely, the logarithmic VNR of  $\Lambda_e$  satisfies

$$\log\left(\frac{\gamma_L(\sigma)}{2\pi e}\right) = 2(\epsilon_{e1} - \epsilon_{e2} - \epsilon_{e3})$$

where

$$\begin{cases} \epsilon_{e1} = C(\Lambda, \sigma_e^2) \\ \epsilon_{e2} = \frac{1}{2} \log 2\pi e \sigma_e^2 - h(\Lambda_r, \sigma_e^2) \\ \epsilon_{e3} = \sum_{\ell=1}^r R_\ell - C(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2). \end{cases} \quad (18)$$

We note that,  $\epsilon_{e1} \leq C(\Lambda, \sigma_b^2) \rightarrow 0$ ,  $\epsilon_{e2} \rightarrow 0$  (condition (ii) in Theorem 2), and  $\epsilon_{e3}$  is the total extra rate of component codes to guarantee security. Since  $R_\ell = |\mathcal{R}_\ell|/N = (|\mathcal{B}_r| + |\mathcal{D}_r|)/N \rightarrow C(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ , we also have  $\epsilon_{e3} \rightarrow 0$ .

Let  $U_{\mathcal{R}(\Lambda_e)}$  denote the uniform distribution over a fundamental region  $\mathcal{R}(\Lambda_e)$ . Note that condition  $C(\Lambda_e, \sigma_e^2) \rightarrow 0$  implies the following statements, which all state that the distribution  $f_{\sigma_e, \Lambda_e}$  of the mod- $\Lambda_e$  Gaussian noise converges to the uniform distribution:

- 1) Differential entropy  $h(\Lambda_e, \sigma_e^2) \rightarrow \log(\text{Vol}(\Lambda_e))$ ;
- 2) Kullback-Leibler divergence  $\mathbb{D}(f_{\sigma_e, \Lambda_e} \| U_{\mathcal{R}(\Lambda_e)}) \rightarrow 0$ ;
- 3) Variational distance  $\mathbb{V}(f_{\sigma_e, \Lambda_e}, U_{\mathcal{R}(\Lambda_e)}) \rightarrow 0$

where 1) is by definition, 2) from the relation between mutual information and Kullback-Leibler divergence<sup>9</sup>, and 3) by Pinsker's inequality.

**Remark 4.** The secrecy capacity of the mod- $\Lambda_s$  Gaussian wiretap channel per use is given by

$$C_s = \frac{1}{N} C(\Lambda_s, \sigma_b^2) - \frac{1}{N} C(\Lambda_s, \sigma_e^2) = \frac{1}{N} h(\Lambda_s, \sigma_e^2) - \frac{1}{N} h(\Lambda_s, \sigma_b^2)$$

since the wiretapper's channel is degraded with respect to the main channel. Because  $h(\Lambda_r, \sigma_e^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_e^2)$  and  $\Lambda_s \subset \Lambda_r^N$ , we have  $\frac{1}{N} h(\Lambda_s, \sigma_e^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_e^2)$  and  $\frac{1}{N} h(\Lambda_s, \sigma_b^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_b^2)$ . Hence  $C_s \rightarrow \frac{1}{2} \log \frac{\sigma_e^2}{\sigma_b^2}$ . It also equals the secrecy capacity of the Gaussian wiretap channel when the signal power goes to infinity. It is noteworthy that we successfully remove the  $\frac{1}{2}$ -nat gap in the achievable secrecy rate derived in [6] which is caused by the limitation of the  $L^\infty$  distance associated with the flatness factor.

**Remark 5.** The mild conditions (i) and (ii) stated in the theorem are easy to meet, by scaling top lattice  $\Lambda$  and choosing the number of levels  $r$  appropriately. Consider an example for  $\sigma_e^2 = 4$  and  $\sigma_b^2 = 1$ . We choose  $r = 3$

<sup>9</sup>In fact, it is easy to show that  $\mathbb{D}(f_{\sigma_e, \Lambda_e} \| U_{\mathcal{R}(\Lambda_e)}) = \log(\text{Vol}(\Lambda_e)) - h(\Lambda_e, \sigma_e^2) = C(\Lambda_e, \sigma_e^2)$ , thanks to the symmetry of the mod- $\Lambda_e$  channel.

levels and a partition chain  $\mathbb{Z}/2\mathbb{Z}/4\mathbb{Z}$  with scaling factor 2.5. The difference between the achievable rate computed from (17) and the upper bound  $\frac{1}{2} \log \frac{\sigma_e^2}{\sigma_b^2}$  on secrecy capacity is about 0.05.

**Remark 6.** From conditions (i) and (ii), we can see that the construction for secrecy-good lattices requires more levels than the construction of AWGN-good lattices.  $\epsilon_1$  can be made arbitrarily small by scaling down  $\Lambda$  such that both  $h(\Lambda, \sigma_e^2)$  and  $h(\Lambda, \sigma_b^2)$  are sufficiently close to  $\log(\text{Vol}(\Lambda))$ . For polar lattices for AWGN-goodness [14], we only need  $h(\Lambda_{r'}, \sigma_b^2) \approx \frac{1}{2} \log(2\pi e \sigma_b^2)$  for some  $r' < r$ . Since  $\epsilon_b < \epsilon_e$ ,  $\Lambda_{r'}$  may be not enough for the wiretapper's channel. Therefore, more levels are needed in the wiretap coding context. To satisfy the condition  $h(\Lambda_r, \sigma_e^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_e^2)$ , it is sufficient to guarantee that  $P_e(\Lambda_r, \sigma_e^2) \rightarrow 0$  by [27, Theorem 13]. When one-dimensional binary partition  $\mathbb{Z}/2\mathbb{Z}/4\mathbb{Z}/\dots$  is used, we have  $P_e(\Lambda_r, \sigma_e^2) \leq Q(\frac{2^r}{2\sigma_e}) \leq e^{-\frac{2^{2r}}{8\sigma_e^2}}$ , where  $Q(\cdot)$  is the Q-function. Letting  $r = O(\log N)$ , the error probability vanishes as  $P_e(\Lambda_r, \sigma_e^2) = e^{-\Omega(N)}$ , which implies that  $h(\Lambda_r, \sigma_e^2) \rightarrow \frac{1}{2} \log(2\pi e \sigma_e^2)$  as  $N \rightarrow \infty$ . We also note that when lattice Gaussian shaping is considered in Sect. V, the probability of selecting a lattice point from  $\Lambda_r$  decays exponentially as  $r$  increases. The requirement is relaxed to  $r = O(\log \log(N))$  to achieve the secrecy capacity.

### C. Semantic security

So far we have assumed that the message is uniformly distributed. In fact, this assumption is not needed because of the symmetry of the  $\Lambda_b/\Lambda_e$  channel [27]. It is well known that the error probability of polar codes in a symmetric channel is independent of the transmitted message [7]; thus the input distribution does not matter for reliability. Moreover, the foregoing security analysis also implies *semantic security*, i.e., (15) holds for arbitrarily distributed M and F. This  $\Lambda_b/\Lambda_e$  channel can be seen as the counterpart in lattice coding of the randomness-induced channel defined in [8].

*Proposition 1:* Semantic security holds for the polar lattice construction for the mod- $\Lambda_s$  GWC shown in Fig. 3, i.e.,

$$I(\text{MF}; \mathbf{Z}^{[N]}) \leq rN2^{-N^{\beta'}}$$

for arbitrarily distributed M and F.

*Proof.* Since MF is drawn from  $\mathcal{R}(\Lambda_e)$  and the random bits are drawn from  $\Lambda_e \cap \mathcal{R}(\Lambda_s)$ , by Lemma 5, the mod- $\Lambda_e$  map is information lossless and its output is a sufficient statistic for MF. Therefore, the channel between MF and the eavesdropper can be viewed as a  $\Lambda_b/\Lambda_e$  channel. Because the  $\Lambda_b/\Lambda_e$  channel is symmetric, the maximum mutual information is achieved by the uniform input. Consequently, the mutual information corresponding to other input distributions can also be upper-bounded by  $rN2^{-N^{\beta'}}$  as in (15), and we can also freeze the bits F.  $\square$

## V. ACHIEVING SECRECY CAPACITY WITH DISCRETE GAUSSIAN SHAPING

In this section, we apply Gaussian shaping on the AWGN-good and secrecy-good polar lattices. The idea of lattice Gaussian shaping was proposed in [30] and then implemented in [15] to construct capacity-achieving polar lattices. For wiretap coding, the discrete Gaussian distribution can also be utilized to satisfy the power constraint. In

simple terms, after obtaining the AWGN-good lattice  $\Lambda_b$  and the secrecy-good lattice  $\Lambda_e$ , Alice maps each message  $m$  to a coset  $\tilde{\lambda}_m \in \Lambda_b/\Lambda_e$  as mentioned in Sect. IV. However, instead of the mod- $\Lambda_s$  operation, Alice samples the encoded signal  $X^N$  from  $D_{\Lambda_e+\lambda_m, \sigma_s}$ , where  $\lambda_m$  is the coset representative of  $\tilde{\lambda}_m$  and  $\sigma_s^2$  is arbitrarily close to the signal power  $P_s$  (see [6] for more details). Again, we assume uniform messages until we prove semantic security in the end of this section.

The construction of polar lattices with Gaussian shaping is reviewed in Sect. V-A. With Gaussian shaping, we propose a new partition of the index set for the genuine GWC in Sect. V-B. Strong secrecy is proved in Sect. V-C, and reliability is then discussed in Sect. V-D. Extension to semantical security is given in Sect. V-E. Moreover, we will show that this shaping operation does not hurt the secrecy rate and that the secrecy capacity can be achieved.

#### A. Gaussian shaping over polar lattices

In this subsection, we introduce the lattice shaping technique for polar lattices. The idea is to select the lattice points according to a carefully chosen lattice Gaussian distribution, which makes a non-uniform input distribution for each partition channel. As shown in [15], the shaping scheme is based on the technique of polar codes for asymmetric channels. For the paper to be self-contained, a brief review will be presented in this subsection. A more detailed account of Gaussian shaping can be found in [15].

Similarly to the polar coding on symmetric channels, the Bhattacharyya parameter for a binary memoryless asymmetric (BMA) channel is defined as follows.

*Definition 4 (Bhattacharyya parameter for BMA channel):* Let  $W$  be a BMA channel with input  $X \in \mathcal{X} = \{0, 1\}$  and output  $Y \in \mathcal{Y}$ . The input distribution and channel transition probability is denoted by  $P_X$  and  $P_{Y|X}$  respectively. The Bhattacharyya parameter  $Z$  for  $W$  is defined as

$$\begin{aligned} Z(X|Y) &= 2 \sum_y P_Y(y) \sqrt{P_{X|Y}(0|y)P_{X|Y}(1|y)} \\ &= 2 \sum_y \sqrt{P_{X,Y}(0, y)P_{X,Y}(1, y)}. \end{aligned}$$

The following lemma, which will be useful for the forthcoming new partition scheme, shows that by adding observable at the output of  $W$ ,  $Z$  will not increase.

*Lemma 7 (Conditioning reduces Bhattacharyya parameter  $Z$  [15]):* Let  $(X, Y, Y') \sim P_{X,Y,Y'}$ ,  $X \in \mathcal{X} = \{0, 1\}$ ,  $Y \in \mathcal{Y}$ ,  $Y' \in \mathcal{Y}'$ , we have

$$Z(X|Y, Y') \leq Z(X|Y).$$

When  $X$  is uniformly distributed, the Bhattacharyya parameter of BMA channels coincides with that of BMS channels defined in Definition 3. Moreover, the calculation of  $Z$  can be converted to the calculation of the Bhattacharyya parameter  $\tilde{Z}$  for a related BMS channel. The following lemma is implicitly considered in [38] and then explicitly expressed in [15]. We show it here for completeness.

*Lemma 8 (From Asymmetric to Symmetric channel [15]):* Let  $W$  be a binary input asymmetric channel with input  $X \in \mathcal{X} = \{0, 1\}$  and  $Y \in \mathcal{Y}$ . We define a new channel  $\tilde{W}$  corresponding to  $W$  which has input  $\tilde{X} \in \mathcal{X} = \{0, 1\}$

and output  $\tilde{Y} \in \mathcal{Y} \times \mathcal{X}$ . The relationship between  $\tilde{W}$  and  $W$  is shown in Fig. 4. The input of  $\tilde{W}$  is uniformly distributed, i.e.,  $P_{\tilde{X}}(\tilde{x} = 0) = P_{\tilde{X}}(\tilde{x} = 1) = \frac{1}{2}$ , and the output of  $\tilde{W}$  is given by  $(Y, X \oplus \tilde{X})$ , where  $\oplus$  denotes the bitwise XOR operation. Then,  $\tilde{W}$  is a binary symmetric channel in the sense that  $P_{\tilde{Y}|\tilde{X}}(y, x \oplus \tilde{x}|\tilde{x}) = P_{Y,X}(y, x)$ .

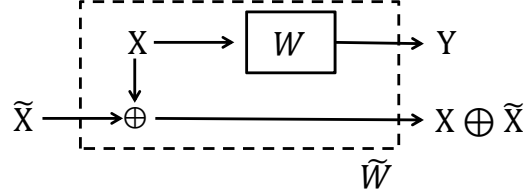


Fig. 4. The relationship between  $\tilde{W}$  and  $W$ .

The following lemma describes how to construct a polar code for a BMA channel  $W$  from that for the associated BMS channel  $\tilde{W}$ .

*Lemma 9 (The equivalence between symmetric and asymmetric Bhattacharyya parameters [38]):* For a BMA channel  $W$  with input  $X \sim P_X$ , let  $\tilde{W}$  be its symmetrized channel constructed according to Lemma 8. Suppose  $X^{[N]}$  and  $Y^{[N]}$  be the input and output vectors of  $W^N$ , and let  $\tilde{X}^{[N]}$  and  $\tilde{Y}^{[N]} = (X^{[N]} \oplus \tilde{X}^{[N]}, Y^{[N]})$  be the input and output vectors of  $\tilde{W}^N$ , where  $\tilde{X}$  is uniform. Consider polarized random variables  $U^{[N]} = X^{[N]} G_N$  and  $\tilde{U}^{[N]} = \tilde{X}^{[N]} G_N$ , and denote by  $W_N$  and  $\tilde{W}_N$  the combining channel of  $N$  uses of  $W$  and  $\tilde{W}$ , respectively. The Bhattacharyya parameter for each subchannel of  $W_N$  is equal to that of each subchannel of  $\tilde{W}_N$ , i.e.,

$$Z(U^i | U^{1:i-1}, Y^{[N]}) = \tilde{Z}(\tilde{U}^i | \tilde{U}^{1:i-1}, X^{[N]} \oplus \tilde{X}^{[N]}, Y^{[N]}).$$

To obtain the desired input distribution of  $P_X$  for  $W$ , the indices with very small  $Z(U^i | U^{1:i-1})$  should be removed from the information set of the symmetric channel. Following [15], the resultant subset is referred to as the information set  $\mathcal{I}$  for the asymmetric channel  $W$ . For the remaining part  $\mathcal{I}^c$ , we further find out that there are some bits which can be made independent of the information bits and uniformly distributed. The purpose of extracting such bits is for the interest of our lattice construction. We name the set that includes those independent frozen bits as the independent frozen set  $\mathcal{F}$ , and the remaining frozen bits are determined by the bits in  $\mathcal{F} \cup \mathcal{I}$ . We name the set of all those deterministic bits as the shaping set  $\mathcal{S}$ . The three sets are formally defined as follows:

$$\begin{cases} \text{the independent frozen set: } \mathcal{F} = \{i \in [N] : Z(U^i | U^{1:i-1}, Y^{[N]}) \geq 1 - 2^{-N^\beta}\} \\ \text{the information set: } \mathcal{I} = \{i \in [N] : Z(U^i | U^{1:i-1}, Y^{[N]}) \leq 2^{-N^\beta} \text{ and } Z(U^i | U^{1:i-1}) \geq 1 - 2^{-N^\beta}\} \\ \text{the shaping set: } \mathcal{S} = (\mathcal{F} \cup \mathcal{I})^c. \end{cases} \quad (19)$$

To identify these three sets, one can use Lemma 9 to calculate  $Z(U^i | U^{1:i-1}, Y^{[N]}, X^{[N]})$  using the known constructing techniques for symmetric polar codes [31] [39]. We note that  $Z(U^i | U^{1:i-1})$  can be computed in a similar way, by constructing a symmetric channel between  $\tilde{X}$  and  $X \oplus \tilde{X}$ . Besides the construction, the decoding process for the asymmetric polar codes can also be converted to the decoding for the symmetric polar codes.

The polar coding scheme according to (19), which can be viewed as an extension of the scheme proposed in [38], has been proved to be capacity-achieving in [15]. Moreover, it can be extended to the construction of multilevel asymmetric polar codes.

Let us describe the encoding strategy for the channel of the  $\ell$ -th ( $\ell \leq r$ ) level  $W_\ell$  with the channel transition probability  $P_{Y|X_\ell, X_{1:\ell-1}}(y|x_\ell, x_{1:\ell-1})$  as follows.

- Encoding: Before sending the codeword  $x_\ell^{[N]} = u_\ell^{[N]} G_N$ , the index set  $[N]$  are divided into three parts: the independent frozen set  $\mathcal{F}_\ell$ , information set  $\mathcal{I}_\ell$ , and shaping set  $\mathcal{S}_\ell$ , which are defined as follows:

$$\begin{cases} \mathcal{F}_\ell = \left\{ i \in [N] : Z\left(\mathsf{U}_\ell^i | \mathsf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{Y}^{[N]}\right) \geq 1 - 2^{-N^\beta} \right\} \\ \mathcal{I}_\ell = \left\{ i \in [N] : Z\left(\mathsf{U}_\ell^i | \mathsf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{Y}^{[N]}\right) \leq 2^{-N^\beta} \text{ and } Z\left(\mathsf{U}_\ell^i | \mathsf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}\right) \geq 1 - 2^{-N^\beta} \right\} \\ \mathcal{S}_\ell = (\mathcal{F}_\ell \cup \mathcal{I}_\ell)^c. \end{cases}$$

The encoder first places uniformly distributed information bits in  $\mathcal{I}_\ell$ . Then the frozen set  $\mathcal{F}_\ell$  is filled with a uniform random sequence which is shared between the encoder and the decoder. The bits in  $\mathcal{S}_\ell$  are generated by a random mapping  $\Phi_{\mathcal{S}_\ell}$ , which yields the following distribution:

$$u_\ell^i = \begin{cases} 0 & \text{with probability } P_{\mathsf{U}_\ell^i | \mathsf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}}(0 | u_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}), \\ 1 & \text{with probability } P_{\mathsf{U}_\ell^i | \mathsf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}}(1 | u_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}). \end{cases} \quad (20)$$

*Theorem 3 (Construction of multilevel polar codes [15]):* Consider a polar code with the above encoding strategy. Then, any message rate arbitrarily close to  $I(X_\ell; Y | X_{1:\ell-1})$  is achievable using SC decoding<sup>10</sup> and the expectation of the decoding error probability over the randomized mappings satisfies  $E_{\Phi_{\mathcal{S}_\ell}}[P_e(\phi_{\mathcal{S}_\ell})] = O(2^{-N^{\beta'}})$  for any  $\beta' < \beta < 0.5$ .

Now let us pick a suitable input distribution  $P_{X_{1:r}}$  to implement the shaping. As shown in Theorem 1, the mutual information between the discrete Gaussian lattice distribution  $D_{\Lambda, \sigma_s}$  and the output of the AWGN channel approaches  $\frac{1}{2} \log(1 + \text{SNR})$  as the flatness factor  $\epsilon_\Lambda(\tilde{\sigma}) \rightarrow 0$ . Therefore, we use the lattice Gaussian distribution  $P_X \sim D_{\Lambda, \sigma_s}$  as the constellation, which gives us  $\lim_{r \rightarrow \infty} P_{X_{1:r}} = P_X \sim D_{\Lambda, \sigma_s}$ . By [15, Lemma 5], when  $N \rightarrow \infty$ , the mutual information  $I(X_r; Y | X_{1:r-1})$  at the bottom level goes to 0 if  $r = O(\log \log N)$ , and using the first  $r$  levels would involve a capacity loss  $\sum_{\ell > r} I(X_\ell; Y | X_{1:\ell-1}) \leq O(\frac{1}{N})$ .

From the chain rule of mutual information,

$$I(X_{1:r}; Y) = \sum_{\ell=1}^r I(X_\ell; Y | X_{1:\ell-1}),$$

we have  $r$  binary-input channels and the  $\ell$ -th channel according to  $I(X_\ell; Y | X_{1:\ell-1})$  is generally asymmetric with the input distribution  $P_{X_\ell | X_{1:\ell-1}}$  ( $1 \leq \ell \leq r$ ). Then we can construct the polar code for the asymmetric channel at each level according to Lemma 8. As a result, the  $\ell$ -th symmetrized channel is equivalent to the MMSE-scaled  $\Lambda_{\ell-1}/\Lambda_\ell$  channel in the sense of channel polarization. (See [15] for more details.)

<sup>10</sup>It is possible to derandomize the mapping  $\Phi_{\mathcal{S}_\ell}$  for the purpose of achieving capacity alone. However, it is tricky to handle the random mapping in order to achieve the secrecy capacity: it requires either to share a secret random mapping or to use the Markov block coding technique (see Sect. V-D).

Therefore, when power constraint is taken into consideration, the multilevel polar codes before shaping are constructed according to the symmetric channel  $V(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}_b^2)$  and  $W(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}_e^2)$ , where  $\tilde{\sigma}_b^2 = \left(\frac{\sigma_s \sigma_h}{\sqrt{\sigma_s^2 + \sigma_b^2}}\right)^2$  and  $\tilde{\sigma}_e^2 = \left(\frac{\sigma_s \sigma_e}{\sqrt{\sigma_s^2 + \sigma_e^2}}\right)^2$  are the MMSE-scaled noise variance of the main channel and of the wiretapper's channel, respectively. This is similar to the mod- $\Lambda_s$  GWC scenario mentioned in the previous section. The difference is that  $\sigma_b^2$  and  $\sigma_e^2$  are replaced by  $\tilde{\sigma}_b^2$  and  $\tilde{\sigma}_e^2$  accordingly. As a result, we can still obtain an AWGN-good lattice  $\Lambda_b$  and a secrecy-good lattice  $\Lambda_e$  by treating  $V(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}_b^2)$  and  $W(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}_e^2)$  as the main channel and wiretapper's channel at each level.

### B. Three-dimensional partition

When lattice Gaussian shaping is performed over the AWGN-good lattice  $\Lambda_b$  and the secrecy-good lattice  $\Lambda_e$  simultaneously, we have a new shaping induced partition. The polar coding scheme for the mod- $\Lambda_s$  wiretap channel given in Sect. IV needs to be modified. Now we consider the partition of the index set  $[N]$  with shaping involved. According to the analysis of asymmetric polar codes, we have to eliminate those indices with small  $Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]})$  from the information set of the symmetric channels. Therefore, Alice cannot send message on those subchannels with  $Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta}$ . Note that this part is the same for  $\tilde{V}_\ell$  and  $\tilde{W}_\ell$ , because it only depends on the shaping distribution. At each level, the index set which is used for shaping is given as

$$\mathcal{S}_\ell \triangleq \left\{ i \in [N] : Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta} \right\},$$

and the index set which is not for shaping is denoted by  $\mathcal{S}_\ell^c$ . Recall that for the index set  $[N]$ , we already have two partition criteria, i.e., reliability-good and information-bad (see (4)). We rewrite the reliability-good index set  $\mathcal{G}_\ell$  and information-poor index set  $\mathcal{N}_\ell$  at level  $\ell$  as

$$\begin{aligned} \mathcal{G}_\ell &\triangleq \left\{ i \in [N] : Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{Y}^{[N]}) \leq 2^{-N^\beta} \right\}, \\ \mathcal{N}_\ell &\triangleq \left\{ i \in [N] : Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{Z}^{[N]}) \geq 1 - 2^{-N^\beta} \right\}. \end{aligned} \quad (21)$$

Note that  $\mathcal{G}_\ell$  and  $\mathcal{N}_\ell$  are defined by the asymmetric Bhattacharyya parameters. Nevertheless, by Lemma 9 and the channel equivalence, we have  $\mathcal{G}_\ell = \mathcal{G}(\tilde{V}_\ell)$  and  $\mathcal{N}_\ell = \mathcal{N}(\tilde{W}_\ell)$  as defined in (4), where  $\tilde{V}_\ell$  and  $\tilde{W}_\ell$  are the respective symmetric channels or the MMSE-scaled  $\Lambda_{\ell-1}/\Lambda_\ell$  channels for Bob and Eve at level  $\ell$ . The four sets  $\mathcal{A}_\ell$ ,  $\mathcal{B}_\ell$ ,  $\mathcal{C}_\ell$ , and  $\mathcal{D}_\ell$  are defined in the same fashion as (5), with  $\mathcal{G}_\ell$  and  $\mathcal{N}_\ell$  replacing  $\mathcal{G}(\tilde{V}_\ell)$  and  $\mathcal{N}(\tilde{W}_\ell)$ , respectively. Now the whole index set  $[N]$  is divided like a cube in three directions, which is shown in Fig. 5.

Clearly, we have eight blocks:

$$\begin{aligned} \mathcal{A}_\ell^S &= \mathcal{A}_\ell \cap \mathcal{S}_\ell, \quad \mathcal{A}_\ell^{S^c} = \mathcal{A}_\ell \cap \mathcal{S}_\ell^c \\ \mathcal{B}_\ell^S &= \mathcal{B}_\ell \cap \mathcal{S}_\ell, \quad \mathcal{B}_\ell^{S^c} = \mathcal{B}_\ell \cap \mathcal{S}_\ell^c \\ \mathcal{C}_\ell^S &= \mathcal{C}_\ell \cap \mathcal{S}_\ell, \quad \mathcal{C}_\ell^{S^c} = \mathcal{C}_\ell \cap \mathcal{S}_\ell^c \\ \mathcal{D}_\ell^S &= \mathcal{D}_\ell \cap \mathcal{S}_\ell, \quad \mathcal{D}_\ell^{S^c} = \mathcal{D}_\ell \cap \mathcal{S}_\ell^c \end{aligned} \quad (22)$$

By Lemma 7, we observe that  $\mathcal{A}_\ell^S = \mathcal{C}_\ell^S = \emptyset$ ,  $\mathcal{A}_\ell^{S^c} = \mathcal{A}_\ell$ , and  $\mathcal{C}_\ell^{S^c} = \mathcal{C}_\ell$ . The shaping set  $\mathcal{S}_\ell$  is divided into two sets  $\mathcal{B}_\ell^S$  and  $\mathcal{D}_\ell^S$ . The bits in  $\mathcal{S}_\ell$  are determined by the bits in  $\mathcal{S}_\ell^c$  according to the mapping. Similarly,  $\mathcal{S}_\ell^c$  is divided

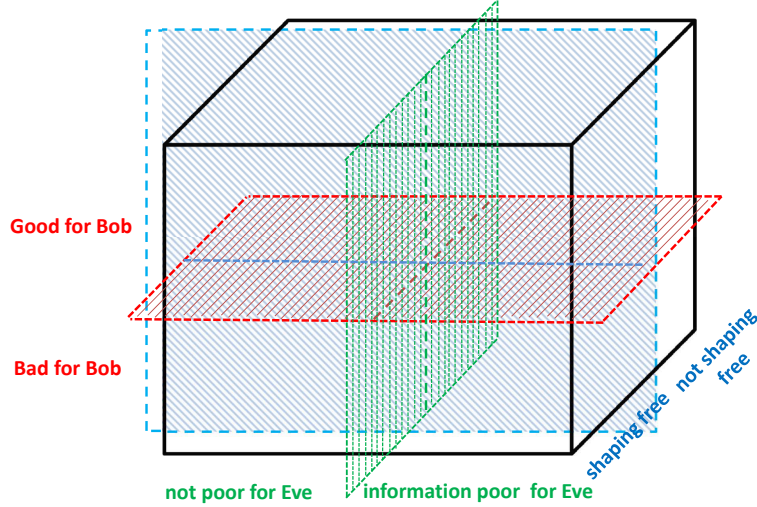


Fig. 5. Partitions of the index set  $[N]$  with shaping.

into the four sets  $\mathcal{A}_\ell^{S^c} = \mathcal{A}_\ell$ ,  $\mathcal{B}_\ell^{S^c}$ ,  $\mathcal{C}_\ell^{S^c} = \mathcal{C}_\ell$ , and  $\mathcal{D}_\ell^{S^c}$ . Note that for wiretap coding, the frozen set becomes  $\mathcal{C}_\ell^{S^c}$ , which is slightly different from the frozen set for channel coding. To satisfy the reliability condition, the frozen set  $\mathcal{C}_\ell^{S^c}$  and the problematic set  $\mathcal{D}_\ell^{S^c}$  cannot be set uniformly random any more. Recall that only the independent frozen set  $\mathcal{F}_\ell$  at each level, which is defined as  $\{i \in [N] : Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{Y}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) \geq 1 - 2^{-N^\beta}\}$ , can be set uniformly random (which are already shared between Alice and Bob), and the bits in the unpolarized frozen set  $\bar{\mathcal{F}}_\ell$ , defined as  $\{i \in [N] : 2^{-N^\beta} < Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{Y}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta}\}$ , should be determined according to the mapping. Moreover, we can observe that  $\mathcal{F}_\ell \subset \mathcal{C}_\ell^{S^c}$  and  $\mathcal{D}_\ell^{S^c} \subset \mathcal{D}_\ell \subset \bar{\mathcal{F}}_\ell$ . Here we make the bits in  $\mathcal{F}_\ell$  uniformly random and the bits in  $\mathcal{C}_\ell^{S^c} \setminus \mathcal{F}_\ell$  and  $\mathcal{D}_\ell^{S^c}$  determined by the mapping. Therefore, from now on, we adjust the definition of the shaping bits as:

$$\mathcal{S}_\ell \triangleq \left\{ i \in [N] : Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta} \text{ or } 2^{-N^\beta} < Z(\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{Y}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta} \right\}, \quad (23)$$

which is essentially equivalent to the definition of the shaping set given in Theorem 3.

To sum up, at level  $\ell$ , we assign the sets  $\mathcal{A}_\ell^{S^c}$ ,  $\mathcal{B}_\ell^{S^c}$ , and  $\mathcal{F}_\ell$  with message bits  $M_\ell$ , uniformly random bits  $R_\ell$ , and uniform frozen bits  $F_\ell$ , respectively. The rest bits  $S_\ell$  (in  $\mathcal{S}_\ell$ ) will be fed with random bits according to  $P_{\mathbf{U}_\ell^i | \mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}}$ . Clearly, this shaping operation will make the input distribution arbitrarily close to  $P_{\mathbf{X}_\ell | \mathbf{X}_{1:\ell-1}}$ , for  $\beta$  fixed and  $N$  tending to infinity. In this case, we can obtain the equality between the Bhattacharyya parameter of asymmetric setting and symmetric setting (see Lemma 9). This provides us a convenient way to prove the strong secrecy of the wiretap coding scheme with shaping because we have already proved the strong secrecy of a symmetric wiretap coding scheme using the Bhattacharyya parameter of the symmetric setting. A detailed proof will be presented in the following subsection. Before this, we show that the shaping will not change the message rate.

*Lemma 10:* For the symmetrized main channel  $\tilde{V}_\ell$  and wiretapper's channel  $\tilde{W}_\ell$ , consider the reliability-good indices set  $\mathcal{G}_\ell$  and information-bad indices set  $\mathcal{N}_\ell$  defined as in (21). By eliminating the shaping set  $\mathcal{S}_\ell$  from the



original message set defined in (5), we get the new message set  $\mathcal{A}_\ell^{S^c} = \mathcal{G}_\ell \cap \mathcal{N}_\ell \cap \mathcal{S}_\ell^c$ . The proportion of  $|\mathcal{A}_\ell^{S^c}|$  equals to that of  $|\mathcal{A}_\ell|$ , and the message rate after shaping can still be arbitrarily close to  $\frac{1}{2} \log \frac{\tilde{\sigma}_\ell^2}{\sigma_\ell^2}$ .

*Proof.* By Theorem 2, when shaping is not involved, the message rate can be made arbitrarily close to  $\frac{1}{2} \log \frac{\tilde{\sigma}_\ell^2}{\sigma_\ell^2}$ . By the new definition (23) of  $\mathcal{S}_\ell$ , we still have  $\mathcal{A}_\ell^S = \emptyset$ , which means the shaping operation will not affect the message rate.  $\square$

### C. Strong secrecy

In this subsection, we prove that strong secrecy can still be achieved when shaping is involved. To this end, we introduce a new induced channel from Eve's perspective and prove that the information leakage over this channel vanishes at each level. Then, strong secrecy is proved by using the chain rule of mutual information as in (15).

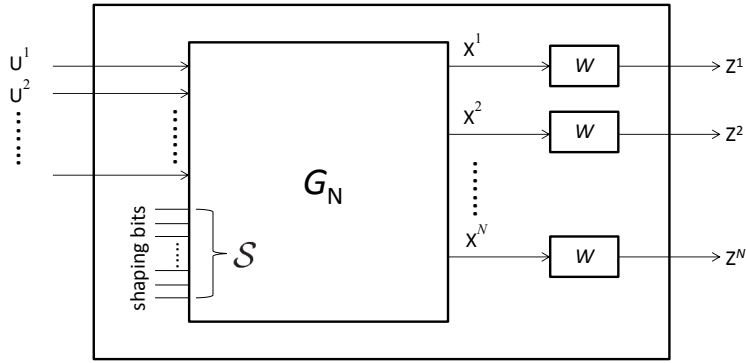


Fig. 6. Block diagram of the shaping-induced channel  $\mathcal{Q}_N(W, \mathcal{S})$ .

In [8], an induced channel is defined in order to prove strong secrecy. Here we call it the randomness-induced channel because it is induced by feeding the subchannels in the sets  $\mathcal{B}_\ell$  and  $\mathcal{D}_\ell$  with uniformly random bits. However, when shaping is involved, the set  $\mathcal{B}_\ell$  and  $\mathcal{D}_\ell$  are no longer fed with uniformly random bits. In fact, some subchannels (covered by the shaping mapping) should be fed with bits according to a random mapping. We define the channel induced by the shaping bits as the shaping-induced channel.

*Definition 5 (Shaping-induced channel):* The shaping-induced channel  $\mathcal{Q}_N(W, \mathcal{S})$  is defined in terms of  $N$  uses of an asymmetric channel  $W$ , and a shaping subset  $\mathcal{S}$  of  $[N]$  of size  $|\mathcal{S}|$ . The input alphabet of  $\mathcal{Q}_N(W, \mathcal{S})$  is  $\{0, 1\}^{N-|\mathcal{S}|}$  and the bits in  $\mathcal{S}$  are determined by the input bits according to a random shaping  $\Phi_{\mathcal{S}}$ . A block diagram of the shaping induced channel is shown in Fig. 6.

Based on the shaping-induced channel, we define a new induced channel, which is caused by feeding a part of the input bits of the shaping-induced channel with uniformly random bits.

*Definition 6 (New induced channel):* Based on a shaping induced channel  $\mathcal{Q}_N(W, \mathcal{S})$ , the new induced channel  $\mathcal{Q}_N(W, \mathcal{S}, \mathcal{R})$  is specified in terms of a randomness subset  $\mathcal{R}$  of size  $|\mathcal{R}|$ . The randomness is introduced into the input set of the shaping-induced channel. The input alphabet of  $\mathcal{Q}_N(W, \mathcal{S}, \mathcal{R})$  is  $\{0, 1\}^{N-|\mathcal{S}|-|\mathcal{R}|}$  and the bits in  $\mathcal{R}$  are uniformly and independently random. A block diagram of the new induced channel is shown in Fig. 7.

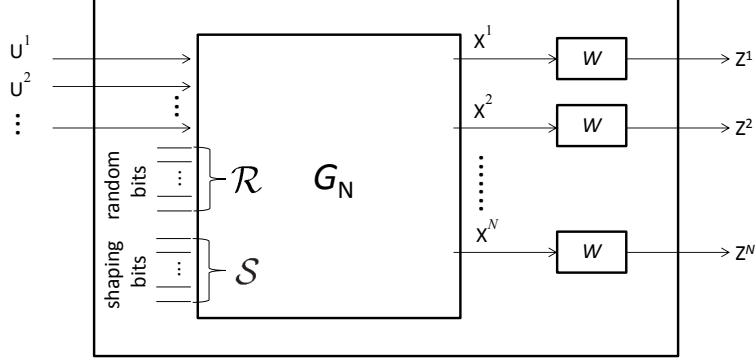


Fig. 7. Block diagram of the new induced channel  $\mathcal{Q}_N(W, \mathcal{S}, \mathcal{R})$ .

The new induced channel is a combination of the shaping-induced channel and randomness-induced channel. This is different from the definition given in [8] because the bits in  $\mathcal{S}$  are neither independent to the message bits nor uniformly distributed. As long as the input bits of the new induced channel are uniform and the shaping bits are chosen according to the random mapping, the new induced channel can still generate  $2^N$  possible realizations  $x_\ell^{[N]}$  of  $\mathcal{X}_\ell^{[N]}$  as  $N$  goes to infinity, and those  $x_\ell^{[N]}$  can be viewed as the output of  $N$  i.i.d binary sources with input distribution  $P_{\mathcal{X}_\ell|\mathcal{X}_{1:\ell-1}}$ . These are exactly the conditions required by Lemma 9. Specifically, we have  $Z(\mathbf{U}_\ell^i|\mathbf{U}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{Z}^{[N]}) = \tilde{Z}(\tilde{\mathbf{U}}_\ell^i|\tilde{\mathbf{U}}_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}, \mathbf{X}_\ell^{[N]} \oplus \tilde{\mathbf{X}}_\ell^{[N]}, \mathbf{Z}^{[N]})$ . In simple words, this equation holds when  $x_\ell^{[N]}$  and  $x_\ell^{[N]} \oplus \tilde{x}_\ell^{[N]}$  are all selected from  $\{0, 1\}^N$  according to their respective distributions. Then we can exploit the relation between the asymmetric channel and the corresponding symmetric channel to bound the mutual information of the asymmetric channel. Therefore, we have to stick to the input distribution (uniform) of our new induced channel and also the distribution of the random mapping. This is similar to the setting of the randomness induced channel in [8], where the input distribution and the randomness distribution are both set to be uniform. In [8], the randomness-induced channel is further proved to be symmetric; then any other input distribution can also achieve strong secrecy and the symmetry finally results in semantic security. In this work, however, we do not have a proof of the symmetry of the new induced channel. For this reason, we assume for now that the message bits are uniform distributed. To prove semantic security, we will show that the information leakage of the symmetrized version of the new induced channel is vanishing in Sect. V-E.

*Lemma 11:* Let  $M_\ell$  be the uniformly distributed message bits and  $F_\ell$  be the independent frozen bits at the input of the channel at the  $\ell$ -th level. When shaping bits  $S_\ell$  are selected according to the random mapping  $\Phi_{S_\ell}$ <sup>11</sup> and  $N$  is sufficiently large, the mutual information can be upper-bounded as

$$I(M_\ell F_\ell; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) \leq O(N^2 2^{-N^{\beta'}}).$$

<sup>11</sup>We will further show that the number of shaping bits  $S_\ell$  covered by random mapping can be significantly reduced in Sect. V-E. Then, to achieve reliability,  $S_\ell$  can be shared between Alice and Bob, or we can use the Markov block coding technique to hide  $S_\ell$  with negligible rate loss.

*Proof.* We firstly assume that  $U_\ell^i$  is selected according to the distribution  $P_{U_\ell^i|U_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}}$  for all  $i \in [N]$ , i.e.,

$$u_\ell^i = \begin{cases} 0 & \text{with probability } P_{U_\ell^i|U_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}}(0|u_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}), \\ 1 & \text{with probability } P_{U_\ell^i|U_\ell^{1:i-1}, \mathbf{X}_{1:\ell-1}^{[N]}}(1|u_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}). \end{cases} \quad (24)$$

for all  $i \in [N]$ . In this case, the input distribution  $P_{\mathbf{X}_\ell|\mathbf{X}_{1:\ell-1}}$  at each level is exactly the optimal input distribution obtained from the lattice Gaussian distribution. The mutual information between  $M_\ell F_\ell$  and  $(\mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]})$  in this case is denoted by  $I_P(M_\ell F_\ell; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]})$ .

For the shaping induced channel  $\mathcal{Q}_N(W_\ell, \mathcal{S}_\ell, \mathcal{R}_\ell)$  ( $\mathcal{R}_\ell$  is  $\mathcal{B}_\ell^{\mathcal{S}^c}$  according to the above analysis), we write the indices of the input bits  $(\mathcal{S}_\ell \cup \mathcal{R}_\ell)^c = [N] \setminus (\mathcal{S}_\ell \cup \mathcal{R}_\ell)$  as  $\{i_1, i_2, \dots, i_{N-s_\ell-r_\ell}\}$ , where  $|\mathcal{R}_\ell| = r_\ell$  and  $|\mathcal{S}_\ell| = s_\ell$ , and assume that  $i_1 < i_2 < \dots < i_{N-s_\ell-r_\ell}$ . We have

$$\begin{aligned} I_P(M_\ell F_\ell; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) &= I_P(U_\ell^{(\mathcal{S}_\ell \cup \mathcal{R}_\ell)^c}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) \\ &= I_P(U_\ell^{i_1}, U_\ell^{i_2}, \dots, U_\ell^{i_{N-s_\ell-r_\ell}}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}) \\ &= \sum_{j=1}^{N-s_\ell-r_\ell} I_P(U_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]} | U_\ell^{i_1}, U_\ell^{i_2}, \dots, U_\ell^{i_{j-1}}) \\ &= \sum_{j=1}^{N-s_\ell-r_\ell} I_P(U_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, U_\ell^{i_1}, U_\ell^{i_2}, \dots, U_\ell^{i_{j-1}}) \\ &\stackrel{(a)}{\leq} \sum_{j=1}^{N-s_\ell-r_\ell} I_P(U_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, U_\ell^1, U_\ell^2, \dots, U_\ell^{i_{j-1}}), \end{aligned}$$

where (a) holds because adding more variables will not decrease the mutual information.

Then the above mutual information can be bounded by the mutual information of the symmetric channel plus an infinitesimal term as follows:

$$\begin{aligned} &\sum_{j=1}^{N-s_\ell-r_\ell} I_P(U_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, U_\ell^{1:i_j-1}) \\ &\stackrel{(a)}{\leq} \sum_{j=1}^{N-s_\ell-r_\ell} I(\tilde{U}_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \tilde{\mathbf{X}}_\ell^{[N]} \oplus \mathbf{X}_\ell^{[N]}, \tilde{U}_\ell^{1:i_j-1}) + H(\tilde{U}_\ell^{i_j} | \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \tilde{\mathbf{X}}_\ell^{[N]} \oplus \mathbf{X}_\ell^{[N]}, \tilde{U}_\ell^{1:i_j-1}) \\ &\quad - \sum_{j=1}^{N-s_\ell-r_\ell} H(U_\ell^{i_j} | \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, U_\ell^{1:i_j-1}) \\ &\stackrel{(b)}{\leq} \sum_{j=1}^{N-s_\ell-r_\ell} I(\tilde{U}_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \tilde{\mathbf{X}}_\ell^{[N]} \oplus \mathbf{X}_\ell^{[N]}, \tilde{U}_\ell^{1:i_j-1}) \\ &\quad + \sum_{j=1}^{N-s_\ell-r_\ell} Z(U_\ell^{i_j} | \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, U_\ell^{1:i_j-1}) - (Z(U_\ell^{i_j} | \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, U_\ell^{1:i_j-1}))^2 \\ &\stackrel{(c)}{\leq} \sum_{j=1}^{N-s_\ell-r_\ell} I(\tilde{U}_\ell^{i_j}; \mathbf{Z}^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \tilde{\mathbf{X}}_\ell^{[N]} \oplus \mathbf{X}_\ell^{[N]}, \tilde{U}_\ell^{1:i_j-1}) + N2^{-N^\beta} \\ &\stackrel{(d)}{\leq} N2^{-N^{\beta'}} + N2^{-N^\beta} \\ &\leq 2N2^{-N^{\beta'}} \end{aligned}$$

for  $0 < \beta' < \beta < 0.5$ . Inequalities (a)-(d) follow from

- (a) uniformly distributed  $\tilde{U}_\ell^{i_j}$ ,
- (b) [40, Proposition 2] which gives  $H(X|Y) - H(X|Y, Z) \leq Z(X|Y) - (Z(X|Y, Z))^2$  and Lemma 9,
- (c) our coding scheme guaranteeing that  $Z(U_\ell^{i_j} | Z^{[N]}, X_{1:\ell-1}^{[N]}, U_\ell^{1:i_j-1})$  is greater than  $1 - 2^{-N^\beta}$  for the frozen bits and information bits,
- (d) Lemma 2.

For wiretap coding, the message  $M_\ell$ , frozen bits  $F_\ell$  and random bits  $R_\ell$  are all uniformly random, and the shaping bits  $S_\ell$  are determined by  $S_\ell^c$  according to  $\Phi_{S_\ell}$ . Let  $Q_{U_\ell^{[N]}, X_{1:\ell-1}^{[N]}, Z^{[N]}}$  denote the joint distribution of  $(U_\ell^{[N]}, X_{1:\ell-1}^{[N]}, Z^{[N]})$  resulted from uniformly distributed  $M_\ell F_\ell R_\ell$  and  $S_\ell$  according to  $\Phi_{S_\ell}$ . By the proofs of [15, Th. 5] and [15, Th. 6], the total variation distance can be bounded as

$$\left\| Q_{U_\ell^{[N]}, X_{1:\ell-1}^{[N]}, Z^{[N]}} - P_{U_\ell^{[N]}, X_{1:\ell-1}^{[N]}, Z^{[N]}} \right\| \leq N2^{-N^{\beta'}} \quad (25)$$

for sufficiently large  $N$ .

By [41, Proposition 5], the mutual information  $I(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]})$  due to  $Q_{U_\ell^{[N]}, X_{1:\ell-1}^{[N]}, Z^{[N]}}$  satisfies

$$\begin{aligned} \left| I(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}) - I_P(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}) \right| &\leq 7N2^{-N^{\beta'}} \log 2^N + h_2(N2^{-N^{\beta'}}) + h_2(4N2^{-N^{\beta'}}) \\ &= O(N^2 2^{-N^{\beta'}}), \end{aligned}$$

where  $h_2(\cdot)$  denotes the binary entropy function. □

Finally, strong secrecy (for uniform message bits) can be proved in the same fashion as shown in (15) as:

$$I(\text{MF}; Z^{[N]}) \leq \sum_{\ell=1}^r I(M_\ell F_\ell; Z^{[N]}, X_{1:\ell-1}^{[N]}) = O(rN^2 2^{-N^{\beta'}}). \quad (26)$$

Therefore we conclude that the whole shaping scheme is secure in the sense that the mutual information leakage between  $M$  and  $Z^{[N]}$  vanishes with the block length  $N$ .

#### D. Reliability

The reliability analysis in Sect. IV-B holds for the wiretap coding without shaping. When shaping is involved, the problematic set  $\mathcal{D}_\ell$  at each level is included in the shaping set  $\mathcal{S}_\ell$  and hence determined by the random mapping  $\Phi_{S_\ell}$ . In this subsection, we propose two decoders to achieve reliability for the shaping case. The first one requires a private link between Alice and Bob to share a vanishing fraction of the random mapping  $\Phi_{S_\ell}$  and the second one uses the Markov block coding technique [9] without sharing the random mapping.

**Decoder 1:** If  $\Phi_{S_\ell}$  is secretly shared between Alice and Bob (we will show in a moment that only a vanishing fraction of  $\Phi_{S_\ell}$  needs to be shared), the bits in  $\mathcal{D}_\ell$  can be recovered by Bob simply by the shared mapping but not requiring the Markov block coding technique. By Theorem 3, the reliability at each level can be guaranteed by uniformly distributed independent frozen bits and a random mapping  $\Phi_{S_\ell}$  according to  $P_{U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}}$  at each level. The decoding rule is given as follows.

- Decoding: The decoder receives  $y^{[N]}$  and estimates  $\hat{u}_\ell^{[N]}$  based on the previously recovered  $x_{1:\ell-1}^{[N]}$  according to the rule

$$\hat{u}_\ell^i = \begin{cases} u_\ell^i, & \text{if } i \in \mathcal{F}_\ell \\ \phi_i(\hat{u}_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}), & \text{if } i \in \mathcal{S}_\ell \\ \operatorname{argmax}_u P_{U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}, Y^{[N]}}(u | \hat{u}_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}, y^{[N]}), & \text{if } i \in \mathcal{I}_\ell \end{cases}.$$

Note that probability  $P_{U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}, Y^{[N]}}(u | \hat{u}_\ell^{1:i-1}, x_{1:\ell-1}^{[N]}, y^{[N]})$  can be calculated by the SC decoding algorithm efficiently, treating  $Y$  and  $X_{1:\ell-1}$  (already decoded by the SC decoder at previous levels) as the outputs of the asymmetric channel. As a result, the expectation of the decoding error probability over the randomized mappings satisfies  $E_{\Phi_{\mathcal{S}_\ell}}[P_e(\phi_{\mathcal{S}_\ell})] = O(2^{-N^{\beta'}})$  for any  $\beta' < \beta < 0.5$ .

Consequently, by the multilevel decoding and union bound, the expectation of the block error probability of our wiretap coding scheme is vanishing as  $N \rightarrow \infty$ . However, this result is based on the assumption that the mapping  $\Phi_{\mathcal{S}_\ell}$  is only shared between Alice and Bob. To share this mapping, we can let Alice and Bob have access to the same source of randomness, which may be achieved by a private link between Alice and Bob. Fortunately, the rate of this private link can be made vanishing since the proportion of the shaping bits covered by the mapping  $\Phi_{\mathcal{S}_\ell}$  can be significantly reduced.

Recall that the shaping set  $\mathcal{S}_\ell$  is defined by

$$\mathcal{S}_\ell \triangleq \left\{ i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta} \text{ or } 2^{-N^\beta} < Z(U_\ell^i | U_\ell^{1:i-1}, Y^{[N]}, X_{1:\ell-1}^{[N]}) < 1 - 2^{-N^\beta} \right\}. \quad (27)$$

It has been shown in [37, Th. 2] and [42, Th. 15] that the shaping bits in the subset  $\{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}) \leq 2^{-N^\beta}\}$  can be recovered according to the rule

$$u_\ell^i = \operatorname{argmax}_u P_{U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{1:N}}(u | u_\ell^{1:i-1}, x_{1:\ell-1}^{1:N}) \quad \text{if } Z(U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}) \leq 2^{-N^\beta},$$

instead of mapping. This modification has negligible impact on strong secrecy. Let us explain it briefly. For the shaping bits in  $\mathcal{S}_\ell$  with  $Z(U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}) \leq 2^{-N^\beta}$ , we also have  $H(U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}) \leq 2^{-N^\beta}$ . This means that  $U_\ell^i$  in  $\mathcal{S}_\ell$  is almost determined by  $U_\ell^{1:i-1}$  and  $X_{1:\ell-1}^{[N]}$  when  $N$  is sufficiently large. The probability  $P_{U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{1:N}}(u | u_\ell^{1:i-1}, x_{1:\ell-1}^{1:N})$  for those bits can be arbitrarily close to either 0 or 1. Therefore, replacing the random rounding rule with the MAP decision rule for those bits will yield another vanishing term  $N2^{-N^{\beta'}}$  on the right hand side of the upper bound of the total variation distance as shown in (25), which results in negligible difference on the information leakage when  $N$  grows large. Moreover, since  $Z(U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{[N]}) \leq 2^{-N^\beta}$  for these shaping bits, using the MAP decision rule will also yield an additional vanishing term  $N2^{-N^{\beta'}}$  on the upper bound of the decoding error probability for Bob. As a result, the deterministic mapping has only to cover the unpolarized set

$$d\mathcal{S}_\ell = \left\{ i \in [N] : 2^{-N^\beta} < Z(U_\ell^i | U_\ell^{1:i-1}, X_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta} \text{ or } 2^{-N^\beta} < Z(U_\ell^i | U_\ell^{1:i-1}, Y^{1:N}, X_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta} \right\},$$

whose proportion  $\frac{|d\mathcal{S}_\ell|}{N} \rightarrow 0$  as  $N \rightarrow \infty$ .

**Remark 7.** By the channel equivalence, when  $\Phi_{S_\ell}$  is shared to Bob, the decoding of  $\Lambda_b$  is equivalent to the MMSE lattice decoding proposed in [6] for random lattice codes. When instantiated with a polar lattice, we use multistage lattice decoding. More explicitly, by [15, Lemma 7], the SC decoding of the asymmetric channel can be converted to the SC decoding of its symmetrized channel, which is equivalent to the MMSE-scaled partition channel in the lattice Gaussian shaping case [15, Lemma 9].

**Decoder 2:** Alternatively, one can also use the block Markov coding technique [9] to achieve reliability without sharing  $\Phi_{S_\ell}$ . As shown in Fig. 8, the message at  $\ell$ -th level is divided into  $k_\ell$  blocks. Denote by  $\Delta S_\ell$  the bits in unpolarized set  $dS_\ell$ . The shaping bits  $S_\ell$  for each block is further divided into unpolarized bits  $\Delta S_\ell$  and polarized shaping bits  $S_\ell \setminus \Delta S_\ell$ . As mentioned above, only  $\Delta S_\ell$  needs to be covered by mapping and its proportion is vanishing. We can sacrifice some message bits to convey  $\Delta S_\ell$  for the next block without involving significant rate loss. These wasted message bits are denoted by  $E_\ell$ . For encoding, we start with the last block (Block  $k_\ell$ ). Given  $F_\ell$ ,  $M_\ell$  (no  $E_\ell$  for the last block) and  $R_\ell$ , we can obtain  $\Delta S_\ell$  according to  $\Phi_{S_\ell}$ . Then we copy  $\Delta S_\ell$  of the last block to the bits  $E_\ell$  of its previous block and do encoding to get the  $\Delta S_\ell$  of block  $k_\ell - 1$ . This process ends until we get the  $\Delta S_\ell$  of the first block. This scheme is similar to the one we discussed in Sect. IV-B. To achieve reliability, we need a secure code with vanishing rate to convey the bits  $\Delta S_\ell$  of the first block to Bob. See [43] for an example of such codes. To guarantee an insignificant rate loss,  $k_\ell$  is required to be sufficiently large. We may set  $k_\ell = O(N^\alpha)$  for some  $\alpha > 0$ .

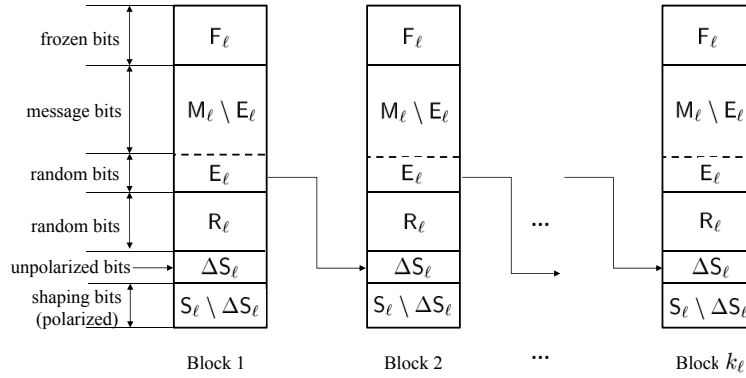


Fig. 8. Markov block coding scheme without sharing the secret mapping.

Now we present the main theorem of the paper.

*Theorem 4 (Achieving secrecy capacity of the GWC):* Consider a multilevel lattice code constructed from polar codes based on asymmetric channels and lattice Gaussian shaping  $D_{\Lambda, \sigma_s}$ . Given  $\sigma_e^2 > \sigma_b^2$ , let  $\epsilon_\Lambda(\tilde{\sigma}_e)$  be negligible and set the number of levels  $r = O(\log \log N)$  for  $N \rightarrow \infty$ . Then all strong secrecy rates  $R$  satisfying  $R < \frac{1}{2} \log \left( \frac{1 + \text{SNR}_b}{1 + \text{SNR}_e} \right)$  are achievable for the Gaussian wiretap channel, where  $\text{SNR}_b$  and  $\text{SNR}_e$  denote the SNR of the main channel and wiretapper's channel, respectively.

*Proof.* The reliability condition and the strong secrecy condition are satisfied by Theorem 3 and Lemma 11,

respectively. It remains to illustrate that the secrecy rate approaches the secrecy capacity. For some  $\epsilon' \rightarrow 0$ , we have

$$\begin{aligned}
\lim_{N \rightarrow \infty} R &= \sum_{\ell=1}^r \lim_{N \rightarrow \infty} \frac{|\mathcal{A}_\ell^{\mathcal{S}^c}|}{N} \\
&= \sum_{\ell=1}^r I(\mathbf{X}_\ell; \mathbf{Y} | \mathbf{X}_1, \dots, \mathbf{X}_{\ell-1}) - I(\mathbf{X}_\ell; \mathbf{Z} | \mathbf{X}_1, \dots, \mathbf{X}_{\ell-1}) \\
&\stackrel{(a)}{=} \frac{1}{2} \log \left( \frac{\tilde{\sigma}_e^2}{\tilde{\sigma}_b^2} \right) - \epsilon' \\
&\stackrel{(b)}{\geq} \frac{1}{2} \log \left( \frac{1 + \text{SNR}_b}{1 + \text{SNR}_e} \right) - \epsilon',
\end{aligned} \tag{28}$$

where (a) is due to Lemma 10, and (b) is because the signal power  $P_s \leq \sigma_s^2$  [30, Lemma 1]<sup>12</sup>, respectively.  $\square$

### E. Semantic security

In this subsection, we extend strong secrecy of the constructed polar lattices to semantic security, namely the resulted strong secrecy does not rely on the distribution of the message. We take the level-1 wiretapper's channel  $W_1$  as an example. Our goal is to show that the maximum mutual information between  $M_1 F_1$  and  $\mathbf{Z}^{[N]}$  is vanishing for any input distribution as  $N \rightarrow \infty$ . Unlike the symmetric randomness induced channel introduced in [8], the new induced channel is generally asymmetric with transition probability

$$Q(z|v) = \frac{1}{2^{r_1}} \sum_{\Phi_{\mathcal{S}_1}} P(\Phi_{\mathcal{S}_1}) \sum_{e \in \{0,1\}^{r_1}} W_1^N(z|(v, e, \Phi_{\mathcal{S}_1}(v, e))G_N),$$

where  $\Phi_{\mathcal{S}_1}(v, e)$  represents the shaping bits determined by  $v$  (the frozen bits and message bits together) and  $e$  (the random bits) according to the random mapping  $\Phi_{\mathcal{S}_1}$ . It is difficult to find the optimal input distribution to maximize the mutual information for the new induced channel.

To prove the semantic security, we investigate the relationship between the  $i$ -th subchannel of  $W_{1,N}$  and the  $i$ -th subchannel of its symmetrized version  $\widetilde{W}_{1,N}$ , which are denoted by  $W_1^{(i,N)}$  and  $\widetilde{W}_1^{(i,N)}$ , respectively. According to Lemma 8, the asymmetric wiretap channel  $W_1 : \mathbf{X}_1 \rightarrow \mathbf{Z}$  is symmetrized to channel  $\widetilde{W}_1 : \widetilde{\mathbf{X}}_1 \rightarrow (\mathbf{Z}, \widetilde{\mathbf{X}}_1 \oplus \mathbf{X}_1)$ . After the  $N$ -by- $N$  polarization transform, we obtain  $W_1^{(i,N)} : \mathbf{U}_1^i \rightarrow (\mathbf{U}_1^{1:i-1}, \mathbf{Z}^{[N]})$  and  $\widetilde{W}_1^{(i,N)} : \widetilde{\mathbf{U}}_1^i \rightarrow (\widetilde{\mathbf{U}}_1^{1:i-1}, \widetilde{\mathbf{X}}_1^{[N]} \oplus \mathbf{X}_1^{[N]}, \mathbf{Z}^{[N]})$ . The next lemma shows that if we symmetrize  $W_1^{(i,N)}$  directly, i.e., construct a symmetric channel  $\widetilde{W}_1^{(i,N)} : \widetilde{\mathbf{U}}_1^i \rightarrow (\mathbf{U}_1^{1:i-1}, \mathbf{Z}^{[N]}, \widetilde{\mathbf{U}}_1^i \oplus \mathbf{U}_1^i)$  in the sense of Lemma 8,  $\widetilde{W}_1^{(i,N)}$  is degraded with respect to  $\widetilde{W}_1^{(i,N)}$ .

*Lemma 12:* The symmetrized channel  $\widetilde{W}_1^{(i,N)}$  derived directly from  $W_1^{(i,N)}$  is degraded with respect to the  $i$ -th subchannel  $\widetilde{W}_1^{(i,N)}$  of  $\widetilde{W}_1$ .

*Proof.* According to the proof of [38, Theorem 2], we have the relationship

$$\widetilde{W}_1^{(i,N)}(\widetilde{u}_1^{1:i-1}, \widetilde{x}_1^{[N]} \oplus x_1^{[N]}, z^{[N]} | \widetilde{u}_1^i) = 2^{-N+1} P_{\mathbf{U}_1^{1:i}, \mathbf{Z}^{[N]}}(u_1^{1:i}, z^{[N]}).$$

Letting  $\widetilde{x}_1^{[N]} \oplus x_1^{[N]} = 0^{[N]}$ , the equation becomes  $\widetilde{W}_1^{(i,N)}(u_1^{1:i-1}, 0^{[N]}, z^{[N]} | u_1^i) = 2^{-N+1} P_{\mathbf{U}_1^{1:i}, \mathbf{Z}^{[N]}}(u_1^{1:i}, z^{[N]})$ , which has already been addressed in [38]. However, for a fixed  $x_1^{[N]}$  and  $\widetilde{u}_1^i = u_1^i$ , since  $G_N$  is full rank, there are  $2^{N-1}$  choices of  $\widetilde{x}_1^{[N]}$  remaining, which means that there exists  $2^{N-1}$  outputs symbols of  $\widetilde{W}_1^{(i,N)}$  having the same

<sup>12</sup>Of course,  $R$  cannot exceed the secrecy capacity, so this inequality implies that  $P_s \rightarrow \sigma_s^2$ .

transition probability  $2^{-N+1}P_{U_1^{1:i}, Z^{[N]}}(u_1^{1:i}, z^{[N]})$ . Suppose a middle channel which maps all these output symbols to one single symbol, which is with transition probability  $P_{U_1^{1:i}, Z^{[N]}}(u_1^{1:i}, z^{[N]})$ . The same operation can be done for  $\tilde{u}_1^i = u_1^i \oplus 1$ , making another symbol with transition probability  $P_{U_1^{1:i}, Z^{[N]}}(u_1^{1:i}, z^{[N]})$  corresponding to the input  $u_1^i \oplus 1$ . This is a channel degradation process, and the degraded channel is symmetric.

Then we show that the symmetrized channel  $\widetilde{W_1^{(i,N)}}$  is equivalent to the degraded channel mentioned above. By Lemma 8, the channel transition probability of  $\widetilde{W_1^{(i,N)}}$  is

$$\widetilde{W_1^{(i,N)}}(u_1^{1:i-1}, \tilde{u}_1^i \oplus u_1^i, z^{[N]} | \tilde{u}_1^i) = P_{U_1^{1:i}, Z^{[N]}}(u_1^{1:i}, z^{[N]}),$$

which is equal to the transition probability of the degraded channel discussed in the previous paragraph. Therefore,  $\widetilde{W_1^{(i,N)}}$  is degraded with respect to  $\widetilde{W_1^{(i,N)}}$ .  $\square$

**Remark 8.** In fact, a stronger relationship that  $\widetilde{W_1^{(i,N)}}$  is equivalent to  $\widetilde{W_1^{(i,N)}}$  can be proved. This is because that the output symbols combined in the channel degradation process have the same LR. An evidence of this result can be found in [38, Equation (36)], where  $\tilde{Z}(\widetilde{W_1^{(i,N)}}) = Z(U_1^i | U_1^{1:i-1}, Z^{[N]}) = \tilde{Z}(\widetilde{W_1^{(i,N)}})$ . Nevertheless, the degradation relationship is sufficient for this work. Notice that Lemma 12 can be generalized to high level  $\ell$ , with outputs  $Z^{[N]}$  replaced by  $(Z^{[N]}, X_{1:\ell-1}^{[N]})$ .

Illuminated by Lemma 12, we can also symmetrize the new induced channel at level  $\ell$  and show that it is degraded with respect to the randomness-induced channel constructed from  $\widetilde{W}_\ell$ . For simplicity, letting  $\ell = 1$ , the new induced channel at level 1 is  $\mathcal{Q}_N(W_1, \mathcal{S}_1, \mathcal{R}_1) : U_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c} \rightarrow Z^{[N]}$ , which is symmetrized to  $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1) : \tilde{U}_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c} \rightarrow (Z^{[N]}, \tilde{U}_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c} \oplus U_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c})$  in the same fashion as in Lemma 8. Recall that the randomness-induced channel of  $\widetilde{W}_1$  defined in [8] can be denoted as  $\mathcal{Q}_N(\widetilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1) : \tilde{U}_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c} \rightarrow (Z^{[N]}, \tilde{X}_1^{[N]} \oplus X_1^{[N]})$ . Note that for the randomness-induced channel  $\mathcal{Q}_N(\widetilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1)$ , set  $\mathcal{R}_1 \cup \mathcal{S}_1$  is fed with uniformly random bits, which is different from the shaping-induced channel.

*Lemma 13:* For an asymmetric channel  $W_1 : X_1 \rightarrow Z$  and its symmetrized channel  $\widetilde{W}_1 : \tilde{X}_1 \rightarrow (Z, \tilde{X}_1 \oplus X_1)$ , the symmetrized version of the new induced channel  $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1)$  is degraded with respect to the randomness-induced channel  $\mathcal{Q}_N(\widetilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1)$ .

*Proof.* The proof is similar to that of Lemma 12. For a fixed realization  $x_1^{[N]}$  and input  $\tilde{u}_1^{(\mathcal{S}_1 \cup \mathcal{R}_1)^c}$ , there are  $2^{|\mathcal{S}_1 \cup \mathcal{R}_1|}$  choice of  $\tilde{x}_1^{[N]}$  remaining. Since  $z^{[N]}$  is only dependent on  $x_1^{[N]}$ , we can build a middle channel which merges the  $2^{|\mathcal{S}_1 \cup \mathcal{R}_1|}$  output symbols of  $\mathcal{Q}_N(\widetilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1)$  to one output symbol of  $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1)$ , which means that  $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1)$  is degraded with respect to  $\mathcal{Q}_N(\widetilde{W}_1, \mathcal{R}_1 \cup \mathcal{S}_1)$ . Again, this result can be generalized to higher levels.  $\square$

Finally, we are ready to prove the semantic security of our wiretap coding scheme. For brevity, let  $M_\ell F_\ell$  and  $\tilde{M}_\ell \tilde{F}_\ell$  denote  $U_\ell^{(\mathcal{S}_\ell \cup \mathcal{R}_\ell)^c}$  and  $\tilde{U}_\ell^{(\mathcal{S}_\ell \cup \mathcal{R}_\ell)^c}$ , respectively. Recall that  $M$  is divided into  $M_1, \dots, M_r$  at each level. We express  $MF$  and  $\tilde{M}\tilde{F}$  as the collection of message and frozen bits on all levels of the new induced channel and the symmetric randomness-induced channel, respectively. We also define  $\tilde{M}\tilde{F} \oplus MF$  as the operation  $\tilde{M}_\ell \tilde{F}_\ell \oplus M_\ell F_\ell$  from level 1 to level  $r$ .



*Theorem 5 (Semantic security):* For arbitrarily distributed message  $M$ , the information leakage  $I(M; Z^{[N]})$  of the proposed wiretap lattice code is upper-bounded as

$$I(M; Z^{[N]}) \leq I(\tilde{M}\tilde{F}; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF) \leq rN2^{-N^{\beta'}},$$

where  $I(\tilde{M}\tilde{F}; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF)$  is the capacity of the symmetrized channel derived from the non-binary channel  $MF \rightarrow Z^{[N]}$ <sup>13</sup>.

*Proof.* By [8, Proposition 16], the channel capacity of the randomness-induced channel  $\mathcal{Q}_N(\tilde{W}_1, \mathcal{S}_1, \mathcal{R}_1)$  is upper-bounded by  $N2^{-N^{\beta'}}$  when partition rule (4) is used. By channel degradation, the channel capacity of the symmetrized new induced channel  $\tilde{\mathcal{Q}}_N(W_1, \mathcal{S}_1, \mathcal{R}_1)$  can also be upper-bounded by  $N2^{-N^{\beta'}}$ . Since this result can be generalized to higher level  $\ell$  ( $\ell \geq 1$ ), we obtain  $C(\tilde{\mathcal{Q}}_N(W_\ell, \mathcal{S}_\ell, \mathcal{R}_\ell)) \leq N2^{-N^{\beta'}}$ , which means  $I(\tilde{M}_\ell \tilde{F}_\ell; Z^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \tilde{M}_\ell \tilde{F}_\ell \oplus M_\ell F_\ell) \leq N2^{-N^{\beta'}}$ . Similarly to (15), we have

$$\begin{aligned} & I(\tilde{M}\tilde{F}; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF) \\ &= \sum_{\ell=1}^r I(\tilde{M}_\ell \tilde{F}_\ell; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF | \tilde{M}_{1:\ell-1} \tilde{F}_{1:\ell-1}) \\ &= \sum_{\ell=1}^r H(\tilde{M}_\ell \tilde{F}_\ell | \tilde{M}_{1:\ell-1} \tilde{F}_{1:\ell-1}) - H(\tilde{M}_\ell \tilde{F}_\ell | Z^{[N]}, \tilde{M}\tilde{F} \oplus MF, \tilde{M}_{1:\ell-1} \tilde{F}_{1:\ell-1}) \\ &\leq \sum_{\ell=1}^r H(\tilde{M}_\ell \tilde{F}_\ell) - H(\tilde{M}_\ell \tilde{F}_\ell | Z^{[N]}, \tilde{M}\tilde{F} \oplus MF, \tilde{M}_{1:\ell-1} \tilde{F}_{1:\ell-1}) \\ &= \sum_{\ell=1}^r I(\tilde{M}_\ell \tilde{F}_\ell; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF, \tilde{M}_{1:\ell-1} \tilde{F}_{1:\ell-1}) \\ &\stackrel{(a)}{=} \sum_{\ell=1}^r I(\tilde{M}_\ell \tilde{F}_\ell; Z^{[N]}, M_{1:\ell-1} F_{1:\ell-1}, \tilde{M}_\ell \tilde{F}_\ell \oplus M_\ell F_\ell) \\ &\stackrel{(b)}{\leq} \sum_{\ell=1}^r I(\tilde{M}_\ell \tilde{F}_\ell; Z^{[N]}, \mathbf{X}_{1:\ell-1}^{[N]}, \tilde{M}_\ell \tilde{F}_\ell \oplus M_\ell F_\ell) \\ &\leq rN2^{-N^{\beta'}}, \end{aligned}$$

where equality (a) holds because  $Z^{[N]}$  is determined by MFR and  $\tilde{M}_\ell \tilde{F}_\ell$  is independent of  $\tilde{M}_{\ell+1:r} \tilde{F}_{\ell+1:r} \oplus M_{\ell+1:r} F_{\ell+1:r}$ , and inequality (b) holds because adding more variables will not decrease the mutual information.

Therefore, we have

$$\begin{aligned} I(M; Z^{[N]}) &\leq I(MF; Z^{[N]}) \\ &\stackrel{(a)}{\leq} H(\tilde{M}\tilde{F} \oplus MF) - H(MF) + I(MF; Z^{[N]}) \\ &\stackrel{(b)}{=} I(\tilde{M}\tilde{F}; Z^{[N]}, \tilde{M}\tilde{F} \oplus MF) \\ &\leq rN2^{-N^{\beta'}}, \end{aligned}$$

where the equality in (a) holds iff  $MF$  is also uniform, and (b) is due to the chain rule.  $\square$

<sup>13</sup>The symmetrization of a non-binary channel is similar to that of a binary channel as shown in Lemma 8. When  $X$  and  $\tilde{X}$  are both non-binary,  $X \oplus \tilde{X}$  denotes the result of the exclusive or (xor) operation of the binary expressions of  $X$  and  $\tilde{X}$ .

## VI. DISCUSSION

We would like to elucidate our coding scheme for the Gaussian wiretap channel in terms of the lattice structure. In Sect. IV, we constructed the AWGN-good lattice  $\Lambda_b$  and the secrecy-good lattice  $\Lambda_e$  without considering the power constraint. When the power constraint is taken into consideration, the lattice Gaussian shaping was implemented in Sect. V.  $\Lambda_b$  and  $\Lambda_e$  were then constructed according to the MMSE-scaled main channel and wiretapper's channel, respectively. We note that these two lattices themselves are generated only if the independent frozen bits on all levels are 0s. Since the independent frozen set of the polar codes at each level is filled with random bits, we actually obtain a coset  $\Lambda_b + \chi$  of  $\Lambda_b$  and a coset  $\Lambda_e + \chi$  of  $\Lambda_e$  simultaneously, where  $\chi$  is a uniformly distributed shift. This is because we are unable to fix the independent frozen bits  $F_\ell$  in our scheme (due to the lack of the proof that the shaping-induced channel is symmetric). By using the lattice Gaussian  $D_{\Lambda, \sigma_s}$  as our constellation in each lattice dimension, we would obtain  $D_{\Lambda^N, \sigma_s}$  without coding. Since  $\Lambda_e + \chi \subset \Lambda_b + \chi \subset \Lambda^N$ , we actually implemented the lattice Gaussian shaping over both  $\Lambda_b + \chi$  and  $\Lambda_e + \chi$ . To summarize, Alice firstly assigns each message  $m \in \mathcal{M}$  to a coset  $\tilde{\lambda}_m \in \Lambda_b/\Lambda_e$ , then randomly sends a point in the coset  $\Lambda_e + \chi + \lambda_m$  ( $\lambda_m$  is the coset leader of  $\tilde{\lambda}_m$ ) according to the distribution  $D_{\Lambda_e + \chi + \lambda_m, \sigma_s}$ . This scheme is consistent with the theoretical model proposed in [6].

On the mod- $\Lambda_s$  wiretap channel, semantic security was obtained for free due to the channel symmetry. On the power-constrained wiretap channel, a symmetrized new induced channel from  $\tilde{\text{MF}}$  to  $(\mathbf{Z}^{[N]}, \tilde{\text{MF}} \oplus \text{MF})$  was constructed to upper-bound the information leakage. This channel is directly derived from the new induced channel from MF to  $\mathbf{Z}^{[N]}$ . According to Lemma 12, this symmetrized new induced channel is degraded with respect to the symmetric randomness-induced channel from  $\tilde{\text{MF}}$  to  $(\mathbf{Z}^{[N]}, \tilde{\mathbf{X}}_{1:r}^{[N]} \oplus \mathbf{X}_{1:r}^{[N]})$ . Moreover, when  $\tilde{\text{F}}$  is frozen, the randomness-induced channel from  $\tilde{\text{M}}$  to  $(\mathbf{Z}^{[N]}, \tilde{\mathbf{X}}_{1:r}^{[N]} \oplus \mathbf{X}_{1:r}^{[N]})$  corresponds to the  $\Lambda_b/\Lambda_e$  channel given in Sect. IV (with MMSE scaling).

### APPENDIX A

#### PROOF OF LEMMA 3

*Proof.* It is sufficient to show  $I(\text{MF}; \mathbf{Z}^{[N]}) \leq N \cdot 2^{-N^{\beta'}}$  since  $I(\text{M}; \mathbf{Z}^{[N]}) \leq I(\text{MF}; \mathbf{Z}^{[N]})$ . As has been shown in [8], the induced channel  $\text{MF} \rightarrow \mathbf{Z}^{[N]}$  is symmetric when  $\mathcal{B}$  and  $\mathcal{D}$  are fed with random bits  $\text{R}$ . For a symmetric channel, the maximum mutual information is achieved by uniform input distribution. Let  $\tilde{\text{U}}^{\mathcal{A}}$  and  $\tilde{\text{U}}^{\mathcal{C}}$  denote independent and uniform versions of  $\text{M}$  and  $\text{F}$  and  $\tilde{\mathbf{Z}}^{[N]}$  be the corresponding channel output. Assuming  $i_1 < i_2 < \dots < i_{|\mathcal{A} \cup \mathcal{C}|}$  are the indices in  $\mathcal{A} \cup \mathcal{C}$ ,

$$\begin{aligned}
 I(\text{MF}; \mathbf{Z}^{[N]}) &\leq I(\tilde{\text{U}}^{\mathcal{A}} \tilde{\text{U}}^{\mathcal{C}}; \tilde{\mathbf{Z}}^{[N]}) \\
 &= \sum_{j=1}^{|\mathcal{A} \cup \mathcal{C}|} I(\tilde{\text{U}}^{i_j}; \tilde{\mathbf{Z}}^{[N]} | \tilde{\text{U}}^{i_1}, \dots, \tilde{\text{U}}^{i_{j-1}}) \\
 &= \sum_{j=1}^{|\mathcal{A} \cup \mathcal{C}|} I(\tilde{\text{U}}^{i_j}; \tilde{\mathbf{Z}}^{[N]}, \tilde{\text{U}}^{i_1}, \dots, \tilde{\text{U}}^{i_{j-1}}) \\
 &\leq \sum_{j=1}^{|\mathcal{A} \cup \mathcal{C}|} I(\tilde{\text{U}}^{i_j}; \tilde{\mathbf{Z}}^{[N]}, \tilde{\text{U}}^{1:i_{j-1}})
 \end{aligned}$$

$$= \sum_{j=1}^{|\mathcal{A} \cup \mathcal{C}|} I(\widetilde{W}_N^{(i_j)}) \leq N \cdot 2^{-N^{\beta'}}.$$

□

## APPENDIX B

## PROOF OF LEMMA 4

*Proof.* According to the definitions of  $\mathcal{G}(\widetilde{V})$  and  $\mathcal{N}(\widetilde{W})$  presented in (4),

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{|\mathcal{G}(\widetilde{V})|}{N} &= \lim_{N \rightarrow \infty} \frac{1}{N} |\{i : \widetilde{Z}(\widetilde{V}_N^{(i)}) \leq 2^{-N^\beta}\}| = C(\widetilde{V}), \\ \lim_{N \rightarrow \infty} \frac{|\mathcal{N}(\widetilde{W})|}{N} &= \lim_{N \rightarrow \infty} \frac{1}{N} |\{i : \widetilde{Z}(\widetilde{W}_N^{(i)}) \geq 1 - 2^{-N^\beta}\}| = 1 - C(\widetilde{W}). \end{aligned}$$

Here we define another two sets  $\bar{\mathcal{G}}(\widetilde{V})$  and  $\bar{\mathcal{N}}(\widetilde{W})$  as

$$\begin{aligned} \bar{\mathcal{G}}(\widetilde{V}) &= \{i : \widetilde{Z}(\widetilde{V}_N^{(i)}) \geq 1 - 2^{-N^\beta}\}, \\ \bar{\mathcal{N}}(\widetilde{W}) &= \{i : \widetilde{Z}(\widetilde{W}_N^{(i)}) \leq 2^{-N^\beta}\}. \end{aligned}$$

Similarly, we have  $\lim_{N \rightarrow \infty} \frac{|\bar{\mathcal{G}}(\widetilde{V})|}{N} = 1 - C(\widetilde{V})$  and  $\lim_{N \rightarrow \infty} \frac{|\bar{\mathcal{N}}(\widetilde{W})|}{N} = C(\widetilde{W})$ . Since  $\widetilde{W}$  is stochastically degraded with respect to  $\widetilde{V}$ ,  $\bar{\mathcal{G}}(\widetilde{V})$  and  $\bar{\mathcal{N}}(\widetilde{W})$  are disjoint with each other [33], then we have

$$\lim_{N \rightarrow \infty} \frac{|\bar{\mathcal{G}}(\widetilde{V}) \cup \bar{\mathcal{N}}(\widetilde{W})|}{N} = 1 - C(\widetilde{V}) + C(\widetilde{W}).$$

By the property of polarization, the proportion of the unpolarized part is vanishing as  $N$  goes to infinity, i.e.,

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{|\mathcal{G}(\widetilde{V}) \cup \bar{\mathcal{G}}(\widetilde{V})|}{N} &= 1, \\ \lim_{N \rightarrow \infty} \frac{|\mathcal{N}(\widetilde{W}) \cup \bar{\mathcal{N}}(\widetilde{W})|}{N} &= 1, \end{aligned}$$

Finally, we have

$$\lim_{N \rightarrow \infty} \frac{|\mathcal{G}(\widetilde{V}) \cap \mathcal{N}(\widetilde{W})|}{N} = 1 - \lim_{N \rightarrow \infty} \frac{|\bar{\mathcal{G}}(\widetilde{V}) \cup \bar{\mathcal{N}}(\widetilde{W})|}{N} = C(\widetilde{V}) - C(\widetilde{W}).$$

□

## APPENDIX C

## PROOF OF LEMMA 6

*Proof.* It is sufficient to demonstrate that channel  $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$  is degraded with respect to  $W'(X_\ell; Z|X_{1:\ell-1})$  and  $W'(X_\ell; Z|X_{1:\ell-1})$  is degraded with respect to  $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$  as well. To see this, we firstly construct a middle channel  $\widehat{W}$  from  $Z \in \mathcal{V}(\Lambda_r)$  to  $\bar{Z} \in \mathcal{V}(\Lambda_\ell)$ . For a specific realization  $\bar{z}$  of  $\bar{Z}$ , this  $\widehat{W}$  maps  $\bar{z} + [\Lambda_\ell/\Lambda_r]$  to  $\bar{z}$  with probability 1, where  $[\Lambda_\ell/\Lambda_r]$  represents the set of the coset leaders of the partition  $\Lambda_\ell/\Lambda_r$ . Then we obtain channel  $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$  by concatenating  $W'(X_\ell; Z|X_{1:\ell-1})$  and  $\widehat{W}$ , which means  $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$  is degraded to  $W'(X_\ell; Z|X_{1:\ell-1})$ . Similarly, we can also construct a middle channel  $\check{W}$  from  $\bar{Z}$  to  $Z$ . For a specific realization  $\bar{z}$  of  $\bar{Z}$ , this  $\check{W}$  maps  $\bar{z}$  to  $\bar{z} + [\Lambda_\ell/\Lambda_r]$  with probability  $\frac{1}{|\Lambda_\ell/\Lambda_r|}$ , where  $|\Lambda_\ell/\Lambda_r|$  is the order of this partition. This means that  $W'(X_\ell; Z|X_{1:\ell-1})$  is also degraded to  $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$ .

By channel degradation and [31, Lemma 1], letting channel  $W$  and  $W'$  denote  $W(\Lambda_{\ell-1}/\Lambda_\ell, \sigma_e^2)$  and  $W'(X_\ell; Z|X_{1:\ell-1})$  for short, we have

$$\begin{aligned}\tilde{Z}(W_N^{(i)}) &\leq \tilde{Z}(W'_N^{(i)}) \text{ and } \tilde{Z}(W_N^{(i)}) \geq \tilde{Z}(W'_N^{(i)}), \\ I(W_N^{(i)}) &\leq I(W'_N^{(i)}) \text{ and } I(W_N^{(i)}) \geq I(W'_N^{(i)}),\end{aligned}$$

meaning that  $\tilde{Z}(W_N^{(i)}) = \tilde{Z}(W'_N^{(i)})$  and  $I(W_N^{(i)}) = I(W'_N^{(i)})$ .  $\square$

## REFERENCES

- [1] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.
- [2] I. Csiszár, "Almost independence and secrecy capacity," *Probl. of Inform. Transmission*, vol. 32, pp. 48–57, 1996.
- [3] S. Leung-Yan-Cheong, "On a special class of wiretap channels," *IEEE Trans. Inf. Theory*, vol. 23, no. 5, pp. 625–627, Sep. 1977.
- [4] S. Goldwasser and S. Micali, "Probabilistic encryption," *J. Comput. Syst. Sci.*, vol. 28, no. 2, pp. 270–299, 1984.
- [5] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Proc. CRYPTO 2012*, ser. Lecture Notes in Computer Science, vol. 7417. Springer-Verlag, 2012, pp. 294–311.
- [6] C. Ling, L. Luzzi, J. Belfiore, and D. Stehle, "Semantically secure lattice codes for the Gaussian wiretap channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6399–6416, Oct. 2014.
- [7] E. Arkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.
- [8] H. Mahdaviifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6428–6443, Oct. 2011.
- [9] E. Şaşıoğlu and A. Vardy, "A new polar coding scheme for strong security on wiretap channels," in *Proc. 2013 IEEE Int. Symp. Inform. Theory*, Istanbul, Turkey, July 2013, pp. 1117–1121.
- [10] T. C. Gulcu and A. Barg, "Achieving secrecy capacity of the wiretap channel and broadcast channel with a confidential component," *IEEE Transactions on Information Theory*, vol. 63, no. 2, pp. 1311–1324, Feb 2017.
- [11] Y.-P. Wei and S. Ulukus, "Polar coding for the general wiretap channel," in *Proc. 2015 IEEE Inform. Theory Workshop*, Jerusalem, Israel, April 2015, pp. 1–5.
- [12] F. Oggier, P. Solé, and J. C. Belfiore, "Lattice codes for the wiretap Gaussian channel: Construction and analysis," *IEEE Transactions on Information Theory*, vol. 62, no. 10, pp. 5690–5708, Oct 2016.
- [13] A. Ernvall-Hytonen and C. Hollanti, "On the eavesdropper's correct decision in Gaussian and fading wiretap channels using lattice codes," in *Proc. 2011 IEEE Inform. Theory Workshop*, Paraty, Brazil, Oct. 2011, pp. 210–214.
- [14] Y. Yan, C. Ling, and X. Wu, "Polar lattices: Where Arkan meets Forney," in *Proc. 2013 IEEE Int. Symp. Inform. Theory*, Istanbul, Turkey, July 2013, pp. 1292–1296.
- [15] Y. Yan, L. Liu, C. Ling, and X. Wu, "Construction of capacity-achieving lattice codes: Polar lattices," Nov. 2014. [Online]. Available: <http://arxiv.org/abs/1411.0187>
- [16] R. Zamir, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*. Cambridge, UK: Cambridge University Press, 2014.
- [17] E. Abbe and A. Barron, "Polar coding schemes for the AWGN channel," in *Proc. 2011 IEEE Int. Symp. Inform. Theory*, St. Petersburg, Russia, July 2011.
- [18] A. Joseph and A. Barron, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2541–2557, May 2012.
- [19] C. Ling, L. Luzzi, and M. Bloch, "Secret key generation from Gaussian sources using lattice hashing," in *Proc. 2013 IEEE Int. Symp. Inform. Theory*, Istanbul, Turkey, July 2013, pp. 2621–2625.
- [20] M. Hayashi and R. Matsumoto, "Construction of wiretap codes from ordinary channel codes," in *Proc. 2010 IEEE Int. Symp. Inform. Theory*, Austin, USA, June 2010, pp. 2538–2542.
- [21] M. Cheraghchi, F. Didier, and A. Shokrollahi, "Invertible extractors and wiretap protocols," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1254–1274, Feb 2012.

- [22] H. Tyagi and A. Vardy, "Explicit capacity-achieving coding scheme for the Gaussian wiretap channel," in *Proc. 2014 IEEE Int. Symp. Inform. Theory*, Honolulu, USA, June 2014, pp. 956–960.
- [23] M. Bloch and J. Laneman, "Strong secrecy from channel resolvability," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8077–8098, Dec. 2013.
- [24] R. A. Chou, M. R. Bloch, and J. Kliewer, "Low-complexity channel resolvability codes for the symmetric multiple-access channel," in *Proc. 2014 IEEE Inform. Theory Workshop*, Hobart, Australia, Nov. 2014, pp. 466–470.
- [25] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6463–6486, Oct. 2011.
- [26] Y. Liang, H. Vincent, and S. Shamai, "Information theoretic security," in *Found. Trends Commun. Inf. Theory*. Norwell, MA, USA: Now Publishers, 2009.
- [27] G. D. Forney Jr., M. Trott, and S.-Y. Chung, "Sphere-bound-achieving coset codes and multilevel coset codes," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 820–850, May 2000.
- [28] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups*. New York: Springer, 1993.
- [29] G. Poltyrev, "On coding without restrictions for the AWGN channel," *IEEE Trans. Inf. Theory*, vol. 40, pp. 409–417, Mar. 1994.
- [30] C. Ling and J. Belfiore, "Achieving AWGN channel capacity with lattice Gaussian coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5918–5929, Oct. 2014.
- [31] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.
- [32] E. Arkan and I. Telatar, "On the rate of channel polarization," in *Proc. 2009 IEEE Int. Symp. Inform. Theory*. Seoul, South Korea: IEEE, June 2009, pp. 1493–1495.
- [33] S. B. Korada, "Polar codes for channel and source coding," Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2009.
- [34] L. Liu, Y. Yan, and C. Ling, "Achieving secrecy capacity of the Gaussian wiretap channel with polar lattices," March 2015. [Online]. Available: <https://arxiv.org/abs/1503.02313>
- [35] R. Fischer, "The modulo-lattice channel: The key feature in precoding schemes," *Int. J. Electron. Commun. (AEÜ)*, vol. 59, no. 4, pp. 244–253, June 2005.
- [36] Y. Yan, L. Liu, and C. Ling, "Polar lattices for strong secrecy over the mod- $\Lambda$  Gaussian wiretap channel," in *Proc. 2014 IEEE Int. Symp. Inform. Theory*, Honolulu, USA, June 2014, pp. 961–965.
- [37] L. Liu and C. Ling, "Polar lattices for lossy compression," Jan. 2015. [Online]. Available: <http://arxiv.org/abs/1501.05683>
- [38] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7829–7838, Dec. 2013.
- [39] R. Mori and T. Tanaka, "Performance of polar codes with the construction using density evolution," *IEEE Commun. Lett.*, vol. 13, no. 7, pp. 519–521, July 2009.
- [40] E. Arkan, "Source polarization," in *Proc. 2010 IEEE Int. Symp. Inform. Theory*, Austin, USA, June 2010, pp. 899–903.
- [41] M. Mondelli, S. H. Hassani, and R. Urbanke, "How to achieve the capacity of asymmetric channels," Sep. 2014. [Online]. Available: <http://arxiv.org/abs/1103.4086>
- [42] E. E. Gad, Y. Li, J. Kliewer, M. Langberg, A. Jiang, and J. Bruck, "Asymmetric error correction and flash-memory rewriting using polar codes," Oct. 2014. [Online]. Available: <http://arxiv.org/abs/1410.3542>
- [43] M. R. Bloch, "Covert communication over noisy channels: A resolvability perspective," Mar. 2015. [Online]. Available: <http://arxiv.org/abs/1503.08778>