Secure Compute-and-Forward in a Bidirectional Relay

Shashank Vatedka, *Student Member, IEEE,*
Navin Kashyap, *Senior Member, IEEE,* Andrew Thangaraj, *Senior Member, IEEE*

Abstract

Bidirectional relaying, where a relay helps two user nodes to exchange messages, has been an active area of recent research. In the compute-and-forward strategy for bidirectional relaying, the relay computes a function of the two messages using the naturally-occurring sum of symbols simultaneously transmitted by user nodes in a Gaussian multiple access channel, and the computed function value is forwarded to the user nodes in an ensuing broadcast phase. This paper studies a problem in which the two user nodes wish to exchange messages using an untrusted relay. It is desired to have statistical independence, or perfect secrecy, between the sum of symbols transmitted by the user nodes and the individual messages. A coding scheme using nested lattices is described. It is shown that the scheme achieves perfect secrecy, and an achievable rate is found. The security conditions are relaxed to strong secrecy, and a coding scheme that achieves this is described. The results are then extended to the multi-hop line network and achievable rates under the two security constraints are found. The coding schemes described in this paper are explicit, to the extent that, given a pair of nested lattices that satisfy certain "goodness" properties, we specify the probability distributions for randomization at the encoders to achieve the desired secrecy criteria.

I. Introduction

Consider a network having three nodes, denoted by `A`, `B` and `R`. The nodes `A` and `B`, henceforth called the user nodes, wish to exchange information with each other. However, they are connected only to `R`, and not to each other directly. The node `R` acts as a bidirectional relay between `A` and `B`, and facilitates communication from `A` to `B`, and from `B` to `A` in the reverse direction. All nodes are assumed to operate in half-duplex mode (they cannot transmit and receive simultaneously), and all links between nodes are wireless (unit channel gain) additive white Gaussian noise (AWGN) channels. Bidirectional relaying in such settings has been studied extensively in the recent literature [2], [23], [28], [34], [36].

S. Vatedka and N. Kashyap ({shashank,nkashyap}@ece.iisc.ernet.in) are with the Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India.

A. Thangaraj (andrew@ee.iitm.ac.in) is with the Department of Electrical Engineering, Indian Institute of Technology, Madras, India.

This work was presented in part at ISIT 2012, Cambridge, Mass., USA, and a part has been accepted for presentation at ISIT 2013, Istanbul, Turkey.

We use the compute-and-forward framework proposed in [23], [34] for bidirectional relaying, and we briefly describe a binary version for completeness and clarity. Suppose that A and B possess bits X and Y, respectively. We will assume that X and Y are generated independently and uniformly at random. The goal in bidirectional relaying is to transmit X to B and Y to A through R. To achieve this goal, a compute-and-forward protocol takes place in two phases as shown in Fig. 2: (1) the (Gaussian) multiple access phase or the MAC phase, where the user nodes simultaneously transmit to the relay, and (2) the broadcast phase, where the relay transmits to the user nodes. In the MAC phase, the user nodes A and B independently modulate their bits X and Y into real valued symbols U and V, respectively. The relay receives an instance of a random variable W, which can be modeled as

$$W = U + V + Z, \tag{1}$$

where it is assumed that the links $\mathtt{A} \to \mathtt{R}$ and $\mathtt{B} \to \mathtt{R}$ have unit gain, Z denotes additive white Gaussian noise independent of U and V, and communication is assumed to be synchronized. Using W, the relay computes the XOR of the two message bits, i.e., $X \oplus Y$, and in the broadcast phase, encodes it into a real symbol which is transmitted to the two users over a broadcast channel. Note that A and B can recover Y and X, respectively, from $X \oplus Y$.

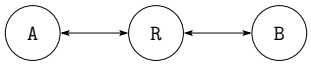

Fig. 1. Bidirectional relay

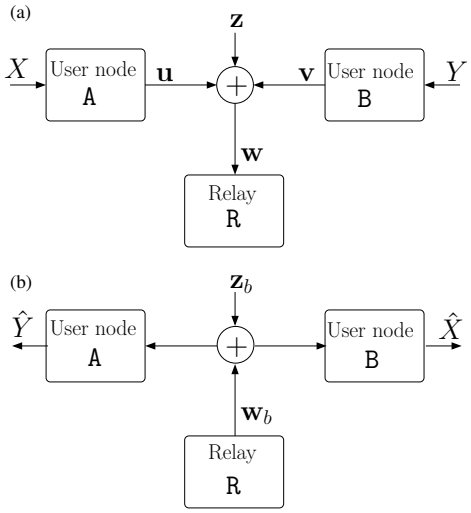

Fig. 2. Bidirectional relaying: (a)MAC phase, (b)Broadcast phase

In a compute-and-forward bidirectional relaying problem such as the above, we study the scenario where an additional secrecy constraint is imposed on the relay R. Specifically, we require that, in the MAC phase, the relay remain ignorant of the individual bits X and Y, while still being able to compute the XOR $X \oplus Y$ reliably. We study the problem under two secrecy constraints. First, we demand that the relay be fully ignorant of the individual bits, i.e., that the random variables $U+V$, X, and Y be pairwise independent. More generally, the user nodes encode the messages X and Y into d-dimensional real vectors $\mathbf{U}$ and $\mathbf{V}$ respectively, and we require $\mathbf{U}+\mathbf{V}$ to be statistically independent of the individual messages. This is referred to as *perfect secrecy*, or *Shannon secrecy* in the literature. The problem of secure bidirectional relaying in presence of an untrusted relay under a perfect secrecy constraint has not been studied prior to this work, and this is a major contribution of this paper. We use a coding scheme that uses a pair of nested lattices $(\Lambda^{(d)}, \Lambda_0^{(d)})$, with $\Lambda_0^{(d)} \subset \Lambda^{(d)}$. The messages are mapped to the cosets of the *coarse lattice* $\Lambda_0^{(d)}$ in the *fine lattice* $\Lambda^{(d)}$. Given a message (say the jth coset, Λ_j) at the user node, the output of the encoder is a random point chosen from that coset according to a distribution p_j. This distribution is obtained by sampling and normalizing over Λ_j, a well-chosen density function f on $\mathbb{R}^d$. We will show that if the characteristic function of f is supported within the fundamental Voronoi region of the Fourier dual of $\Lambda_0^{(d)}$, then it is possible to achieve perfect secrecy. We then study the average transmit power and achievable rates for reliable and secure communication. We will show that a transmission rate of $\frac{1}{2}\log_2 \frac{\mathcal{P}}{\sigma^2} - \log_2 2e$ is achievable with perfect secrecy. Our coding scheme for security is explicit, in that given *any* pair of nested lattices, we precisely specify the distributions p_j that must be used to obtain independence between $\mathbf{U}+\mathbf{V}$ and the individual messages.

We later relax the secrecy constraint, and only demand that the mutual information between $\mathbf{U}+\mathbf{V}$, and the individual messages be arbitrarily small for large block lengths, also referred to as *strong secrecy* [22]. We again use a nested lattice coding scheme, but now the distributions p_j are obtained by sampling and normalizing a Gaussian function, instead of a density having a compactly supported characteristic function. The idea of using probability mass functions (pmfs) obtained by sampling Gaussians was used [20] in the context of the Gaussian wiretap channel, and we will make use of the techniques developed there. Using this scheme, we show that a rate of $\frac{1}{2}\log_2 \frac{\mathcal{P}}{\sigma^2} - \frac{1}{2}\log_2 2e$ is achievable. The rate can be further improved to $\frac{1}{2}\log_2\left(\frac{1}{2}+\frac{\mathcal{P}}{\sigma^2}\right) - \frac{1}{2}\log_2 2e$ using random dithering at the encoders and MMSE equalization at the relay [12], [23].

It is worth emphasizing the basic idea behind the construction of encoders in our coding schemes. Given a pair of nested lattices, the user nodes send points from the fine lattice in the nested lattice pair according to a pmf obtained by sampling a well-chosen density function at the fine lattice points. The choice of the density function determines the level of security that is achievable.

We will restrict our study exclusively to the MAC phase, since there is no security requirement in the broadcast phase and the relay can use a capacity-approaching code to broadcast $X \oplus Y$ to the users.

After discussing secure bidirectional relaying, we show that the principles can be extended to the multi-hop

line network [16] and find achievable transmission rates under the two secrecy constraints.

In prior work, the problem of secure bidirectional relaying in the presence of an untrusted relay was studied by He and Yener in [16], who showed that the mutual information rate, defined to be $\frac{1}{d}\mathcal{I}(X;\mathbf{U}+\mathbf{V}) = \frac{1}{d}\mathcal{I}(Y;\mathbf{U}+\mathbf{V})$ goes to zero for large blocklengths, d. They later studied the problem under a strong secrecy constraint in [17], and gave a scheme based on nested lattice codes and universal hash functions. The achievable rates guaranteed by our strongly secure scheme is slightly lower than that obtained in [17]. However, our scheme is more explicit, in that given a pair of nested lattices that satisfy certain properties [1], we specify exactly what distribution must be used to obtain strong secrecy.

Lattice codes have been proposed for secure communication in different scenarios, particularly the Gaussian wiretap channel (see e.g., [4], [26], [20]). They have also been proposed for use in interference networks [1], and for secret key generation using correlated Gaussian sources [25].

Organization of the paper

We establish some basic notation, and recall some basic lattice definitions in Section II. We describe the secure bidirectional relaying problem in Section III, and then proceed to design coding schemes under the perfect secrecy constraint in Section IV. The main result under the perfect secrecy constraint is given in Theorem 1. We give a randomized encoding scheme for any arbitrary nested lattice code that achieves perfect secrecy in the absence of noise in Section V, then study the effect of additive noise and find achievable transmission rates in Section VI. Thereafter, we study the same problem under a strong secrecy constraint, design coding schemes, and evaluate the performance in Section VII, with the main result summarized in Theorem 17. We extend the results to a multi-hop line network in Section VIII. Most of the technical proofs are given in appendices.

II. Definitions and Notation

We first describe the notation we will use throughout the paper. We denote the set of real numbers by $\mathbb{R}$, and integers by $\mathbb{Z}$. We use the notation $\mathbb{R}^+$ for the set of nonnegative real numbers. The number of elements in a finite set S is denoted by $|S|$. Random vectors are denoted in boldface upper case, e.g., $\mathbf{U}$, and their instances in boldface lower case, as in $\mathbf{u}$. The components of the vectors are denoted in normal font, e.g., $\mathbf{x} = [x_1\, x_2]^T$. Matrices are represented in sans-serif, as in H. The Euclidean (ℓ^2) norm of a column vector $\mathbf{h}$ is denoted by $\|\mathbf{h}\|$. The identity matrix of size $M \times M$ is denoted by I_M. If X is a random variable, then $\mathcal{H}(X)$ denotes the entropy of X, and expectation over the random variable X is denoted by $\mathbb{E}_X(\cdot)$. The probability of an event A is denoted by $\Pr[A]$. For random variables X, Y, the notation $X \perp\!\!\!\perp Y$ means that X and Y are independent. The mutual information between X and Y is denoted by $\mathcal{I}(X;Y)$.

[1] Unfortunately, there are no explicit constructions of lattices that satisfy these properties, but only existence results based on probabilistic arguments.

Let $f(n)$ and $g(n)$ be a sequence of positive real numbers. We say that $g(n) = o(f(n))$ if $g(n)/f(n) \to 0$ as $n \to \infty$. Also, $g(n) = o_n(1)$ if $g(n) \to 0$ as $n \to \infty$. Furthermore, $g(n) = \Omega(f(n))$ if there exists a constant $K > 0$ such that $g(n) > Kf(n)$ for all sufficiently large n, and $g(n) = \mathcal{O}(f(n))$ if there exists a constant $K > 0$ such that $g(n) < Kf(n)$ for all sufficiently large n.

A. Lattices in $\mathbb{R}^d$

We briefly recall some definitions of lattices and their properties. For a more detailed treatment, see e.g., [3], [6].

Let k, d be positive integers with $k \leq d$. Suppose $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are linearly independent column vectors in $\mathbb{R}^d$. Then the set of all integer-linear combinations of the $\mathbf{u}_i$'s, $\Lambda = \{\sum_{i=1}^{k} a_i \mathbf{u}_i : a_i \in \mathbb{Z}, 1 \leq i \leq k\}$, is called a k-dimensional *lattice* in $\mathbb{R}^d$. It is easy to verify that Λ forms an Abelian group under componentwise addition. The collection of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is called a *basis* for the lattice Λ, and it is a standard fact that the basis of a lattice is not unique.

The $k \times d$ matrix $\mathsf{A} := [\mathbf{u}_1 \ \mathbf{u}_2 \ \cdots \mathbf{u}_k]^T$ is called a *generator matrix* of Λ, and we say that the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ generate Λ. We write $\Lambda = \mathsf{A}^T\mathbb{Z}^d := \{\mathsf{A}^T\mathbf{x} : \mathbf{x} \in \mathbb{Z}^d\}$. If Λ is full-rank (i.e., Λ is a d-dimensional lattice in $\mathbb{R}^d$), then the *determinant* of Λ, denoted by $\det\Lambda$ is defined to be $|\det\mathsf{A}|$. It is a standard fact that $\det\Lambda$ does not depend on the generator matrix. Unless mentioned otherwise, we will henceforth consider only full-rank lattices in $\mathbb{R}^d$.

If Λ and Λ_0 are two lattices in $\mathbb{R}^d$ such that $\Lambda_0 \subset \Lambda$, then we say that Λ_0 is a *sublattice* of Λ, or Λ_0 is *nested* within Λ. We call Λ_0 the *coarse lattice* and Λ the *fine lattice*. The number of cosets of Λ_0 in Λ is called the *index* of Λ_0 in Λ, denoted by $|\Lambda/\Lambda_0|$. It is a standard fact that $|\Lambda/\Lambda_0| = \det\Lambda_0/\det\Lambda$ [3, Theorem 5.2].

If A is a generator matrix of a lattice Λ, then $\Lambda^* := \{\mathsf{A}^{-1}\mathbf{z} : \mathbf{z} \in \mathbb{Z}^d\}$ is called the *dual lattice* of Λ. The dual lattice Λ^* is also equal to $\{\mathbf{x} \in \mathbb{R}^d : \sum_{i=1}^{d} x_i y_i \in \mathbb{Z} \text{ for every } \mathbf{y} \in \Lambda\}$[3]. The *Fourier dual* of Λ, denoted $\hat{\Lambda}$, is defined as $2\pi\Lambda^*$.

For any $\mathbf{x} \in \mathbb{R}^d$, we define the nearest neighbour quantizer $Q_\Lambda(\mathbf{x}) := \arg\min_{\lambda\in\Lambda}\|\mathbf{x}-\lambda\|$ to be the function which maps $\mathbf{x}$ to the closest point in Λ. The *fundamental Voronoi region* of Λ is defined as $\mathcal{V}(\Lambda) := \{\mathbf{y} : Q_\Lambda(\mathbf{y}) = \mathbf{0}\}$. The volume of the fundamental Voronoi region, $\mathrm{vol}(\mathcal{V}(\Lambda))$ is equal to $\det\Lambda$ [3], [6].

For any $\mathbf{x} \in \mathbb{R}^d$, we define the modulo-$\Lambda$ operation as $[\mathbf{x}] \bmod \Lambda := \mathbf{x} - Q_\Lambda(\mathbf{x})$. In other words, $[\mathbf{x}] \bmod \Lambda$ gives the quantization error of the nearest neighbour quantizer $Q_\Lambda(\cdot)$. Figure 3 illustrates the $Q_\Lambda(\cdot)$ and the modulo-Λ operations.

The *covering radius* of Λ, denoted by $r_{\mathrm{cov}}(\Lambda)$ is defined as the radius of the smallest closed ball in $\mathbb{R}^d$ centered at $\mathbf{0}$ which contains $\mathcal{V}(\Lambda)$. The *effective radius*, $r_{\mathrm{eff}}(\Lambda)$ is defined as the radius of a ball in $\mathbb{R}^d$ having the same volume as that of $\mathcal{V}(\Lambda)$. The *packing radius* of Λ is the radius of the largest open ball centered at $\mathbf{0}$ which is contained in $\mathcal{V}(\Lambda)$. Clearly, $r_{\mathrm{cov}}(\Lambda) \geq r_{\mathrm{eff}}(\Lambda) \geq r_{\mathrm{pack}}(\Lambda)$. These parameters are illustrated for the hexagonal lattice in Fig. 4.

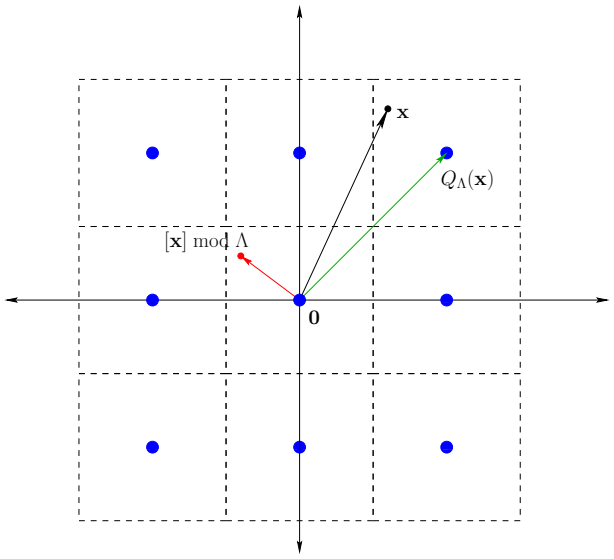

Fig. 3. Illustrating the $Q_\Lambda(.)$ and the $[.] \bmod \Lambda$ operation for the $\mathbb{Z}^2$ lattice.

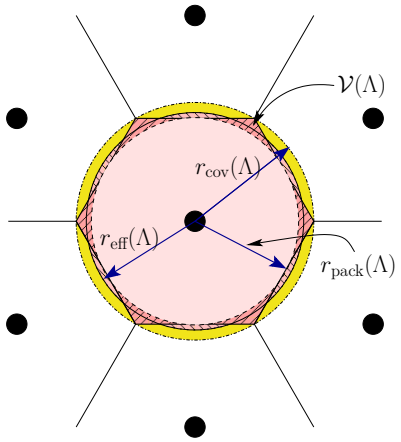

Fig. 4. Illustrating the covering, packing and effective radii of the hexagonal lattice.

The *normalized second moment per dimension* of Λ is defined as

$$\mathcal{G}_\Lambda = \frac{1}{d\left\{\det\Lambda\right\}^{1+2/d}} \int_{\mathcal{V}(\Lambda)} \|\mathbf{y}\|^2 d\mathbf{y}. \tag{2}$$

III. Description of the Problem

The general set-up is as follows: two user nodes, denoted by `A` and `B`, possess messages taking values independently and uniformly in a finite set. For the purposes of computation at the relay, the messages are mapped into random variables X and Y taking values in a finite Abelian group $\mathbb{G}^{(d)}$, where the choice of $\mathbb{G}^{(d)}$ is left to the system designer. The mapping is such that the random variables X and Y remain uniformly distributed over $\mathbb{G}^{(d)}$, and we will see later that this distribution helps in achieving secrecy. The addition

operation in the group $\mathbb{G}^{(d)}$ is denoted $\oplus$. For a given message X at node A, the encoder at A generates a random d-dimensional real vector $\mathbf{U}$. In a similar fashion, for a given Y, the encoder at B generates a random vector $\mathbf{V}$. The user nodes transmit their respective vectors to the relay simultaneously, and at the end of the MAC phase, the relay obtains

$$\mathbf{W} = \mathbf{U} + \mathbf{V} + \mathbf{Z}, \tag{3}$$

where $\mathbf{Z}$ is a Gaussian random vector, with zero mean and covariance matrix $\sigma^2 \mathsf{I}_d$, with I_d being the $d \times d$ identity matrix, and $+$ denotes componentwise real addition. The coding scheme at each user node must ensure that the relay can recover $X \oplus Y$ reliably from $\mathbf{W}$, and one of the following:

- the mutual information between $\mathbf{W}$ and the individual messages, $\mathcal{I}(\mathbf{W}; X)$ and $\mathcal{I}(\mathbf{W}; Y)$ is exactly zero[2] (perfect secrecy)
- $\mathcal{I}(\mathbf{W}; X)$ and $\mathcal{I}(\mathbf{W}; Y)$ can be made arbitrarily small for all sufficiently large d (strong secrecy).

We in fact impose a slightly stronger security criterion than the one mentioned above. Even in the absence of noise, the mutual information between $\mathbf{W} = \mathbf{U}+\mathbf{V}$ and the individual messages must be either zero (perfect secrecy) or can be made arbitrarily small for all sufficiently large d (strong secrecy). Since the additive noise is independent of everything else, $X \to \mathbf{U} + \mathbf{V} \to \mathbf{U} + \mathbf{V} + \mathbf{Z}$ forms a Markov chain, and using the data processing inequality (see e.g., [7]), $\mathcal{I}(X; \mathbf{U}+\mathbf{V}+\mathbf{Z}) \leq \mathcal{I}(X; \mathbf{U}+\mathbf{V})$. Likewise, $\mathcal{I}(Y; \mathbf{U}+\mathbf{V}+\mathbf{Z}) \leq \mathcal{I}(Y; \mathbf{U}+\mathbf{V})$. Therefore, any scheme that achieves perfect (strong) secrecy in the absence of noise will also achieve perfect (strong) secrecy in a noisy channel.

The messages must also be protected from corruption by the additive noise in the multiple access channel. The user nodes need to communicate $X \oplus Y$ to the relay with arbitrarily low probability of error. Since the messages are uniformly distributed over $\mathbb{G}^{(d)}$, $\frac{1}{d}\log_2|\mathbb{G}^{(d)}|$ gives the average number of bits of information sent to the relay by each user node in one channel use in the MAC phase. Our aim will be to ensure secure computation of $X \oplus Y$ at the highest possible rate (which we define to be $\frac{1}{d}\log_2|\mathbb{G}^{(d)}|$) for a given power constraint at the user nodes. To formalize these notions, we have the following definition:

Definition 1. *For a positive integer d, a $(d, M^{(d)})$ code for the MAC phase of the bidirectional relay channel with user nodes A, B and relay R consists of the following:*

1) ***Messages:*** *Nodes A and B possess messages X and Y, respectively, drawn independently and uniformly from a finite Abelian group $\mathbb{G}^{(d)}$ with $M^{(d)} = |\mathbb{G}^{(d)}|$ elements.*
2) ***Codebook:*** *The codebook, denoted by $\mathcal{C}$, is a discrete subset of $\mathbb{R}^d$, not necessarily finite. The elements of $\mathcal{C}$ are called codewords. The codebook consists of all those vectors that are allowed to be transmitted by the user nodes to the relay.*
3) ***Encoder:*** *The encoder is a randomized mapping from $\mathbb{G}^{(d)}$ to $\mathbb{R}^d$, specified by the distribution $p_{\mathbf{U}|X}(\mathbf{u}|x) = Pr[\mathbf{U} = \mathbf{u}|X = x]$ for all $\mathbf{u} \in \mathcal{C}$ and $x \in \mathbb{G}^{(d)}$. The same encoder is used at both nodes, A and B.*

[2]Equivalently, we want $\mathbf{W} \perp\!\!\!\perp X$ and $\mathbf{W} \perp\!\!\!\perp Y$.

At node A*, given a message* $x \in \mathbb{G}^{(d)}$ *as input, the encoder outputs a codeword* $\mathbf{u} \in \mathcal{C}$ *at random, according to* $p_{\mathbf{U}|X}(\mathbf{u}|x)$*. Similarly, at node* B*, with* y *as input, the encoder outputs* $\mathbf{v} \in \mathcal{C}$ *according to* $p_{\mathbf{V}|Y}(\mathbf{v}|y) = p_{\mathbf{U}|X}(\mathbf{v}|y)$*. The encoding of* x *and* y *are done independently. The* rate *of the code is defined to be*

$$R^{(d)} = \frac{\log_2 M^{(d)}}{d}. \tag{4}$$

The code has an average transmit power per dimension *defined as*

$$P^{(d)} = \frac{1}{d}\mathbb{E}\|\mathbf{U}\|^2 = \frac{1}{d}\mathbb{E}\|\mathbf{V}\|^2. \tag{5}$$

4) ***Decoder:*** *The relay* R *receives a vector* $\mathbf{W} \in \mathbb{R}^{(d)}$ *as given in (3). The decoder,* $\mathcal{D}^{(d)} : \mathbb{R}^d \to \mathbb{G}^{(d)}$ *maps the received vector to an element of the set of messages. The average probability of error of the decoder is defined as*

$$\eta^{(d)} := \mathbb{E}\big(Pr[\mathcal{D}^{(d)}(\mathbf{W}) \neq X \oplus Y]\big)$$

where $\mathbb{E}$ *denotes expectation over the messages,* X, Y*, and over the encoders (*$\mathbf{U}, \mathbf{V}$ *given* X, Y*).*

IV. Perfect Secrecy

We first study the case where perfect statistical independence between $\mathbf{U}+\mathbf{V}$ and the individual messages is required, and the relay must be able to reliably compute $X \oplus Y$ (where $\oplus$ denotes addition within $\mathbb{G}^{(d)}$) from the received vector. To summarize, we have the following requirements for secure compute-and-forward:

(S1) $(\mathbf{U}, X) \perp\!\!\!\perp (\mathbf{V}, Y)$.

(S2) $(\mathbf{U} + \mathbf{V}) \perp\!\!\!\perp X$ and $(\mathbf{U} + \mathbf{V}) \perp\!\!\!\perp Y$.

(S3) $\mathbf{U} + \mathbf{V}$ almost surely determines $X \oplus Y$.

If conditions (S1)–(S3) are satisfied, the relay has no means of finding the individual messages. Property (S3) ensures that the relay can decode $X \oplus Y$, which can then be encoded/modulated for further transmission over the broadcast channel. On reception of the broadcast message, since user A (resp. B) knows X (resp. Y), it can recover Y (resp. X).

If the relay only had access to $X \oplus Y$ instead of $\mathbf{U} + \mathbf{V}$, the problem of secure communication would have been trivial due to the uniformity and independence of X and Y. However, the relay receives the real sum of $\mathbf{U}$ and $\mathbf{V}$, which makes the problem harder. For example, suppose that $d = 1$, and $\mathbb{G}^{(1)} = \mathbb{Z}_2$, the group of integers modulo 2. Consider the coding scheme $\mathbf{U} = X$, and $\mathbf{V} = Y$. Then, in the absence of noise, whenever $\mathbf{U} + \mathbf{V} = 0$ or $\mathbf{U} + \mathbf{V} = 2$, the relay can determine both X and Y.

The performance of a coding scheme is generally evaluated in terms of the average transmit power, and the transmission rate. To make these notions formal, we define achievable power-rate pairs as follows.

Definition 2. *A power-rate pair* $(\mathcal{P}, \mathcal{R})$ *is* achievable with perfect secrecy *if, for every* $\delta > 0$*, there exists a sequence of* $(d, M^{(d)})$ *codes such that*

- *conditions (S1)–(S3) are satisfied for all* d*,*

and for all sufficiently large d,

- *the transmission rate, $R^{(d)}$, is greater than $\mathcal{R} - \delta$;*
- *the average transmit power per dimension $P^{(d)}$, is less than $\mathcal{P} + \delta$; and*
- *the average probability of decoding error, $\eta^{(d)}$, is less than δ.*

The objective of the next couple of sections will be to prove the following result.

Theorem 1. *A power-rate pair of*

$$\left(\mathcal{P}, \frac{1}{2}\log_2 \frac{\mathcal{P}}{\sigma^2} - \log_2(2e)\right)$$

is achievable with perfect secrecy in the MAC phase of the bidirectional relay.

V. Perfect secrecy: The Noiseless Setting

To get a clear picture as to how secure communication can be achieved, we first describe the binary case. The messages X and Y are chosen independently and uniformly at random from $\{0, 1\}$, or equivalently, the set of integers modulo-2 ($\mathbb{G} = \mathbb{Z}_2$). They are modulated to U and V respectively, which take values in $\mathbb{R}$. Studying the one-dimensional case will give us the intuition needed to tackle the general case, and we will see that the techniques developed here extend quite naturally to the d-dimensional setting.

We will show that there exist distributions on U and V that permit secure computation defined by properties (S1)–(S3). This is somewhat surprising since we cannot have non-degenerate real valued random variables U, V that satisfy $(U + V) \perp\!\!\!\perp U$ and $(U + V) \perp\!\!\!\perp V$, as shown in the following proposition:

Proposition 2. *Let U and V be independent real-valued random variables, and let $+$ denote addition over $\mathbb{R}$. Then, we have $(U + V) \perp\!\!\!\perp U$ and $(U + V) \perp\!\!\!\perp V$ iff U and V are constant a.s. (i.e., there exist $a, b \in \mathbb{R}$ such that $Pr[U = a] = Pr[V = b] = 1$).*

Proof: The "if" part is trivial, so let us prove the "only if" part. Let $W = U + V$, so that by assumption, U, V and W are pairwise independent. Let φ_U, φ_V and φ_W denote the characteristic functions of U, V and W, respectively. In particular, $\varphi_W = \varphi_U \varphi_V$. From $U = W - V$, we also have that $\varphi_U = \varphi_W \overline{\varphi_V}$, where $\overline{\varphi_V}$ denotes the complex conjugate of φ_V. Putting the two equalities together, we obtain $\varphi_U = \varphi_U |\varphi_V|^2$. To be precise, $\varphi_U(t) = \varphi_U(t)|\varphi_V(t)|^2$ for all $t \in \mathbb{R}$.

Now, characteristic functions are continuous and take the value 1 at $t = 0$. Hence, φ_U is non-zero within the interval $[-\delta, \delta]$ for some $\delta > 0$. Thus, $|\varphi_V(t)| = 1$ for all $t \in [-\delta, \delta]$. By a basic property of characteristic functions (see Lemma 4 of Section XV.1 in [15]), this implies that there exists $b \in \mathbb{R}$ such that $\varphi_V(t) = e^{ibt}$ for all $t \in \mathbb{R}$, thus proving that $V = b$ with probability 1.

A similar argument using $V = W - U$ shows that U is also constant with probability 1. ∎

A. Secure Computation of XOR at the Relay

In this section, X and Y are independent and identically distributed (iid) uniform binary random variables (rvs), and $X \oplus Y$ denotes their modulo-2 sum (XOR). We describe a construction of integer-valued rvs U and V satisfying (S1)–(S3).

1) Conditions on PMFs and Characteristic Functions: We first derive conditions under which integer-valued rvs U and V can satisfy properties (S1)–(S3) stated in Section III. We introduce some notation: for $k \in \mathbb{Z}$, let $p_U(k) = \Pr[U = k]$, $p_V(k) = \Pr[V = k]$, and for $a \in \{0,1\}$, let $p_{U|a}(k) = \Pr[U = k \mid X = a]$, $p_{V|a}(k) = \Pr[V = k \mid Y = a]$. Thus, $p_U = (1/2)(p_{U|0} + p_{U|1})$ and $p_V = (1/2)(p_{V|0} + p_{V|1})$.

Property (S1) is equivalent to requiring that the joint probability mass function (pmf) of (U, V, X, Y) be expressible as

$$p_{UVXY}(k, l, a, b) = (1/2)(1/2)p_{U|a}(k)p_{V|b}(l) \tag{6}$$

for $k, l \in \mathbb{Z}$ and $a, b \in \{0, 1\}$. Without the requirement that $U + V \perp\!\!\!\perp X$ and $U + V \perp\!\!\!\perp Y$, it is trivial to define U and V such that (S3) is satisfied: for example, take $U = X$ and $V = Y$. Property (S3) is satisfied by any U, V such that

$$\begin{aligned} p_{U|0}(k) = p_{V|0}(k) = 0 \quad &\text{for all odd } k \in \mathbb{Z}, \\ p_{U|1}(k) = p_{V|1}(k) = 0 \quad &\text{for all even } k \in \mathbb{Z}. \end{aligned} \tag{7}$$

Finally, we turn our attention to (S2). We want $(U+V) \perp\!\!\!\perp X$ and $(U+V) \perp\!\!\!\perp Y$. Let us define, for $k \in \mathbb{Z}$, $p_{U+V}(k) = \Pr[U + V = k]$, and for $a \in \{0, 1\}$, $p_{U+V|X=a}(k) = \Pr[U + V = k \mid X = a]$ and $p_{U+V|Y=a}(k) = \Pr[U + V = k \mid Y = a]$. Assuming $(U, X) \perp\!\!\!\perp (V, Y)$, we have $p_{U+V} = p_U * p_V$, $p_{U+V|X=a} = p_{U|a} * p_V$, and $p_{U+V|Y=a} = p_U * p_{V|a}$, where $*$ denotes the convolution operation. Thus, when $(U, X) \perp\!\!\!\perp (V, Y)$, (S2) holds iff

$$p_U * p_V = p_{U|a} * p_V = p_U * p_{V|a} \quad \text{for } a \in \{0, 1\}. \tag{8}$$

It helps to view this in the Fourier domain. Let φ_U, φ_V, $\varphi_{U|a}$ etc. denote the respective characteristic functions of the pmfs p_U, p_V, $p_{U|a}$ etc. — for example, $\varphi_{U|a}(t) = \sum_{k \in \mathbb{Z}} p_{U|a}(k)e^{ikt}$. Then, (8) is equivalent to

$$\varphi_U \varphi_V = \varphi_{U|a}\varphi_V = \varphi_U \varphi_{V|a} \quad \text{for } a \in \{0, 1\}. \tag{9}$$

Note that $\varphi_U = (1/2)(\varphi_{U|0} + \varphi_{U|1})$ and $\varphi_V = (1/2)(\varphi_{V|0} + \varphi_{V|1})$. Hence, (9) should be viewed as a requirement on the conditional pmfs $p_{U|a}$ and $p_{V|a}$, $a \in \{0, 1\}$.

In summary, we have the following lemma.

Lemma 3. *Suppose that the conditional pmfs $p_{U|a}$ and $p_{V|a}$, $a \in \{0, 1\}$, satisfy (7) and (9). Then, the rvs U, V, X, Y with joint pmf given by (6) have properties (S1)–(S3).*

The observations made up to this point also allow us to prove the following negative result.[3]

[3]In fact, a stronger negative result can be shown — see Proposition 10.

Proposition 4. *Properties (S1)–(S3) cannot be satisfied by integer-valued rvs U, V that are finitely supported.*

Proof: Suppose that U and V are finitely supported $\mathbb{Z}$-valued rvs. Then, $\varphi_U(t)$ and $\varphi_V(t)$ are finite linear combinations of some exponentials $e^{ik_1 t}, \ldots, e^{ik_n t}$. Equivalently, the real and imaginary parts of φ_U and φ_V are trigonometric polynomials. Thus, either φ_U (resp. φ_V) is identically zero, or it has a discrete set of zeros. The former is impossible as $\varphi_U(0) = \varphi_V(0) = 1$. Now, suppose that (S1) and (S2) are satisfied, which means that (9) must hold. The equality $\varphi_U \varphi_V = \varphi_U \varphi_{V|a}$ in (9) implies that $\varphi_{V|a}(t) = \varphi_V(t)$ for all t such that $\varphi_U(t) \neq 0$. But since $\varphi_U(t)$ has a discrete set of zeros, continuity of characteristic functions in fact implies that $\varphi_{V|a}(t) = \varphi_V(t)$ for all t. An analogous argument shows that $\varphi_{U|a}(t) = \varphi_U(t)$ for all t. Hence, $U \perp\!\!\!\perp X$ and $V \perp\!\!\!\perp Y$. From this, and (S1), we obtain that $U + V \perp\!\!\!\perp X \oplus Y$, thus precluding (S3). ∎

Practical communication systems generally have a maximum power constraint, which means that we would like to have U, V being finitely supported. But from Proposition 4, we see that it is not possible to have finitely supported U, V that permit secure computation of the XOR at the relay. Therefore, in order to ensure secure computation, we will have to relax the criterion to an *average power constraint* on the user nodes. This means that we require finite-variance, integer-valued random variables U, V, with infinite support, that satisfy properties (S1)–(S3), or equivalently, the hypotheses of Lemma 3.

We now give a construction of U, V that satisfy the hypotheses of Lemma 3. We will choose a density function whose characteristic function is finitely supported. The random variables U and V are chosen according to a distribution obtained by sampling and appropriately normalizing this density function. To study this in more detail, we rely upon methods and results from Fourier analysis. The key tool we need is the Poisson summation formula, which we briefly recall here. Our description is based largely on Section XIX.5 in [15].

B. The Poisson Summation Formula

Let ψ be the characteristic function of a real-valued random variable X, such that $\int_{-\infty}^{\infty} |\psi(t)| dt < \infty$. In particular, ψ is continuous and $\psi(0) = 1$. Since ψ is absolutely integrable, the random variable X has a continuous density f. The Poisson summation formula [15, Chapter XIX, equation (5.9)] states that for any $T > 0$ and $s \in \mathbb{R}$, we have for all $\zeta \in \mathbb{R}$,

$$\sum_{n=-\infty}^{\infty} \psi(\zeta + 2n\pi/T)\, e^{-is(2n\pi/T)} = T \sum_{k=-\infty}^{\infty} f(kT + s)\, e^{i(kT+s)\zeta}, \tag{10}$$

provided that the series on the left converges to a continuous function $\Psi(\zeta)$. Note that $\Psi(0) = T \sum_{k=-\infty}^{\infty} f(kT + s)$, which is a non-negative quantity. If $\Psi(0) \neq 0$, then dividing both sides of (10) by $\Psi(0)$ yields the important fact that $\Psi(\zeta)/\Psi(0)$ is the characteristic function of a discrete random variable supported within the set $\{kT + s : k \in \mathbb{Z}\}$, the probability mass at the point $kT + s$ being equal to $f(kT + s)/\sum_{k=-\infty}^{\infty} f(kT + s)$.

A special case of interest is when ψ is compactly supported. Let $T > 0$ be such that $\psi(t) = 0$ whenever $|t| \geq \pi/T$. It is straightforward to see that the series on the left-hand-side of (10) converges to a continuous

function Ψ, and that $\Psi(0) = \psi(0) = 1$. Indeed, the series may be written as $e^{is\zeta}\widetilde{\Psi}(\zeta)$, where

$$\widetilde{\Psi}(\zeta) = \sum_{n=-\infty}^{\infty} \psi(\zeta + 2n\pi/T)\, e^{-is(\zeta+2n\pi/T)}.$$

Note that $\widetilde{\Psi}$ is periodic with period $2\pi/T$. In fact, $\widetilde{\Psi}$ is simply the periodic extension, with period $2\pi/T$, of $\psi(t)e^{-ist}$. Now, ψ (being a characteristic function) is continuous, and hence, so is $\widetilde{\Psi}$. We conclude that $\Psi(\zeta) = e^{is\zeta}\widetilde{\Psi}(\zeta)$ is a continuous function. Furthermore, $\Psi(0) = \psi(0) = 1$, From this, we infer that Ψ is the characteristic function of a discrete rv, as explained above. In fact, by plugging in $\zeta = 0$ in (10) we obtain that $\psi(0) = T\sum_k f(kT+s)$, which shows that $\sum_k f(kT+s) = 1/T$. For future reference, we record this in the form of a proposition.

Proposition 5. *Let ψ be a characteristic function such that $\psi(t) = 0$ whenever $|t| \geq \pi/T$ for some $T > 0$, and let f be the corresponding probability density function. Then, for any $s \in \mathbb{R}$, the function $\Psi : \mathbb{R} \to \mathbb{C}$ defined by*

$$\Psi(\zeta) = \sum_{n=-\infty}^{\infty} \psi(\zeta + 2n\pi/T)\, e^{-is(2n\pi/T)}$$

is the characteristic function of a discrete random variable supported within the set $\{kT + s : k \in \mathbb{Z}\}$. The probability mass at the point $kT + s$ is equal to $Tf(kT + s)$.

It should be noted that compactly supported characteristic functions do indeed exist — see e.g., [15, Section XV.2, Table 1], [10], [31]. We also give an explicit construction after the proof of Theorem 7 in the next subsection.

1) Multi-dimensional Version: We will also need a multi-dimensional version of the Poisson summation formula (see e.g., [32, Chapter VII, Section 2]). It is more natural to express this in terms of lattices. Let Λ be a lattice in $\mathbb{R}^d$. Recall from Section II-A that $\hat{\Lambda}$ denotes the Fourier dual of Λ.

Let $\psi : \mathbb{R}^d \to \mathbb{C}$ now be the characteristic function of an $\mathbb{R}^d$-valued random variable, such that $\int_{\mathbb{R}^d} |\psi(\mathbf{t})| d\mathbf{t} < \infty$. Let the corresponding density function be $f : \mathbb{R}^d \to \mathbb{R}$. The Poisson summation formula in $\mathbb{R}^d$ can be expressed as follows: for any $\mathbf{s} \in \mathbb{R}^d$, we have for all $\boldsymbol{\zeta} \in \mathbb{R}^d$,

$$\sum_{\mathbf{n}\in\hat{\Lambda}} \psi(\boldsymbol{\zeta} + \mathbf{n})\, e^{-i\langle \mathbf{n}, \mathbf{s}\rangle} = (\det \Lambda^{(d)}) \sum_{\mathbf{k}\in\Lambda} f(\mathbf{k} + \mathbf{s})\, e^{i\langle \mathbf{k}+\mathbf{s}, \boldsymbol{\zeta}\rangle}, \tag{11}$$

provided that the series on the left converges to a continuous function $\Psi(\boldsymbol{\zeta})$. It should be pointed out that texts in Fourier analysis typically state the Poisson summation formula for an arbitrary ℓ^1 function f, and would then require that f and ψ decay sufficiently quickly — see e.g., [32, Chapter VII, Corollary 2.6] or [3, Eq. (17.1.2)] — for (11) to hold. However, as argued by Feller in proving (10), in the special case of a non-negative ℓ^1 function f, it is sufficient to assume that the left-hand-side of (11) converges to a continuous function $\Psi(\boldsymbol{\zeta})$.

The following d-dimensional extension of Proposition 5 follows easily from (11).

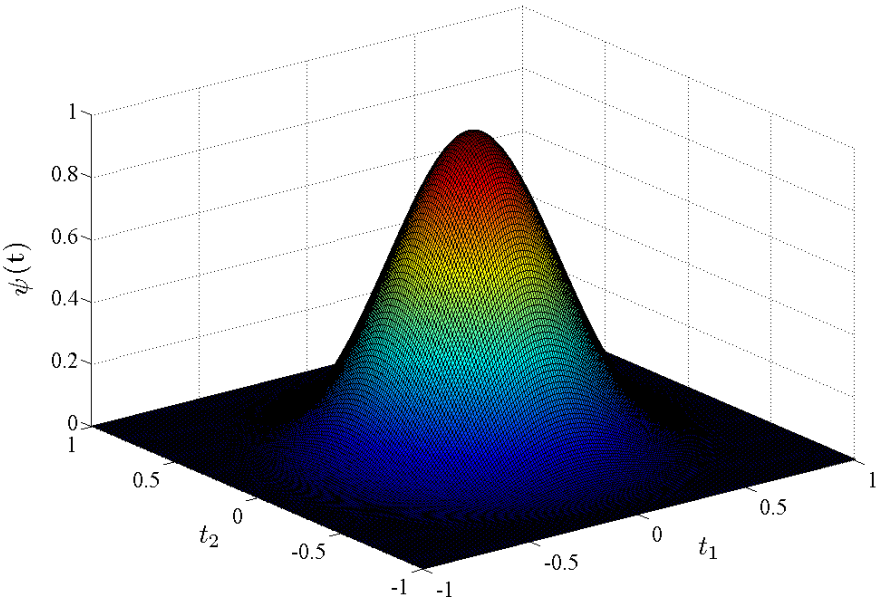

Fig. 5. Example of a characteristic function supported within $\mathcal{V}(2\mathbb{Z}^2)$.

Proposition 6. *Let Λ be a full-rank lattice in $\mathbb{R}^d$. Let $\psi : \mathbb{R}^d \to \mathbb{C}$ be a characteristic function such that $\psi(\mathbf{t}) = 0$ for all $\mathbf{t} \notin \mathcal{V}(\hat{\Lambda})$, and let $f : \mathbb{R}^d \to \mathbb{R}$ be the corresponding probability density function. Then, for any $\mathbf{s} \in \mathbb{R}^d$, the function $\Psi : \mathbb{R}^d \to \mathbb{C}$ defined by*

$$\Psi(\boldsymbol{\zeta}) = \sum_{\mathbf{n} \in \hat{\Lambda}} \psi(\boldsymbol{\zeta} + \mathbf{n})\, e^{-i\langle \mathbf{n}, \mathbf{s} \rangle}$$

is the characteristic function of a random variable supported within the set $\Lambda + \mathbf{s} := \{\mathbf{k} + \mathbf{s} : \mathbf{k} \in \Lambda\}$. The probability mass at the point $\mathbf{k} + \mathbf{s}$ is equal to $(\det \Lambda)\, f(\mathbf{k} + \mathbf{s})$.

C. General Construction

We now describe the construction of integer-valued random variables that satisfy (S1)–(S3). Let ψ be a characteristic function (of a continuous random variable X) with the properties that

(C1) $\psi(t) = 0$ for $|t| \geq \pi/2$, and

(C2) $\psi(t)$ is real and non-negative for all $t \in \mathbb{R}$.[4]

Note that since ψ is real-valued, it must be an even function: $\psi(-t) = \psi(t)$ for all $t \in \mathbb{R}$. Also, $\psi(0) = 1$. Since ψ is integrable over $\mathbb{R}$, by the Fourier inversion formula, the random variable X has a continuous

[4]There is no loss of generality in imposing this requirement. Suppose that a random variable X has characteristic function ψ, which is complex-valued in general. Let X_1, X_2 be iid rvs with the same distribution as X. Then, $X_1 - X_2$ has characteristic function $\psi\bar{\psi} = |\psi|^2$.

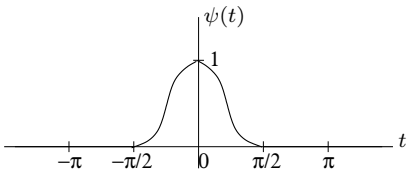

Fig. 6. A generic characteristic function supported on $[-\pi/2, \pi/2]$.

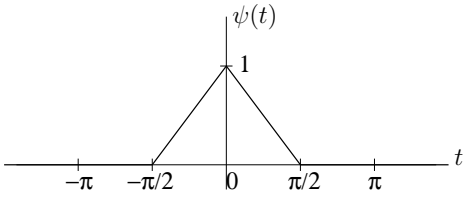

Fig. 7. $\psi(t) = \max\{0, 1 - 2|t|/\pi\}$.

density f. Note that Proposition 5 holds for $T \leq 2$.

A generic such ψ is depicted in Figure 6. As a specific example (see Table 1 of Section XV.2 in [15]), a random variable with density function

$$f(x) = \begin{cases} \frac{1}{4} & \text{if } x = 0 \\ \frac{2}{\pi^2 x^2}(1 - \cos \pi x/2) & \text{if } x \neq 0 \end{cases} \tag{12}$$

has characteristic function

$$\int_{-\infty}^{\infty} f(x)\, e^{itx}\, dx = \max\{0, 1 - 2|t|/\pi\}.$$

Figure 7 shows a plot of this characteristic function.

Reverting to our generic characteristic function ψ, let φ be the periodic function with period 2π that agrees with ψ on $[-\pi, \pi]$, as depicted in Figure 8. Note that $\varphi(\zeta) = \sum_{n=-\infty}^{\infty} \psi(\zeta + 2\pi n)$. Thus, applying Proposition 5 with $T = 1$ and $s = 0$, we find that φ is the characteristic function of an integer-valued random variable, with pmf given by

$$p(k) = f(k) \text{ for all } k \in \mathbb{Z}. \tag{13}$$

Next, for $s = 0, 1$, define φ_s as follows: for $\zeta \in \mathbb{R}$,

$$\varphi_s(\zeta) = \sum_{n=-\infty}^{\infty} \psi(\zeta + n\pi) e^{-isn\pi}.$$

It is easily seen that φ_0 is the periodic extension of ψ with period π, i.e., φ_0 is the periodic function with period π that agrees with ψ on $[-\pi/2, \pi/2]$, as depicted at the top of Figure 9 for a generic ψ shown in Figure 6.

On the other hand, φ_1 is periodic with period 2π: its graph is obtained from that of φ_0 by reflecting about the ζ-axis every second copy of ψ, as depicted at the bottom of Figure 9.

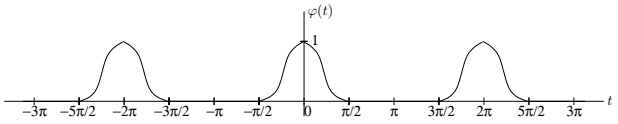

Fig. 8. Period-2π extension of generic ψ from Figure 6.

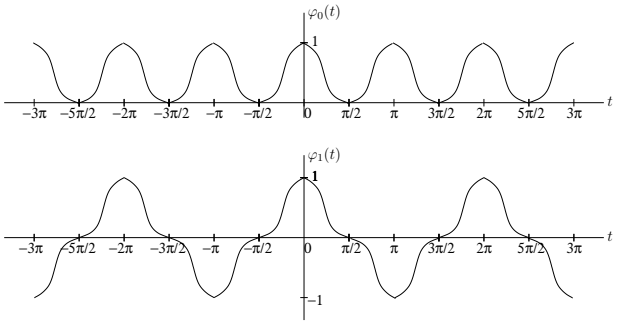

Fig. 9. The periodic functions φ_0 and φ_1 derived from ψ.

Applying Proposition 5 with $T = 2$ and $s \in \{0, 1\}$, we get that φ_0 and φ_1 are characteristic functions of rvs supported within the even and odd integers, respectively. The pmf corresponding to φ_0 is given by

$$p_0(k) = \begin{cases} 2f(k) & \text{if } k \text{ is an even integer} \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

and that corresponding to φ_1 is

$$p_1(k) = \begin{cases} 2f(k) & \text{if } k \text{ is an odd integer} \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

From (13)–(15), we have $p(k) = \frac{1}{2}(p_0(k) + p_1(k))$ for all $k \in \mathbb{Z}$.

Finally, note that since $\varphi_0(t)$ and $\varphi_1(t)$ differ from $\varphi(t)$ only when $\varphi(t) = 0$, we have

$$\varphi^2 = \varphi\varphi_0 = \varphi\varphi_1. \tag{16}$$

We can now prove the following theorem.

Theorem 7. *Let X, Y be iid Bernoulli$(1/2)$ rvs. Suppose that we are given a probability density function $f : \mathbb{R} \to \mathbb{R}^+$ with a non-negative real characteristic function ψ such that $\psi(t) = 0$ for $|t| \geq \pi/2$. Set $p_{U|0} = p_{V|0} = p_0$ and $p_{U|1} = p_{V|1} = p_1$, where p_0 and p_1 are as in (14) and 15). Then, the resulting $\mathbb{Z}$-valued rvs U and V satisfy properties (S1)–(S3). Additionally, the rvs U and V have finite variance iff ψ is twice differentiable, in which case the variance equals $-\psi''(0)$.*

Proof: From the given characteristic function ψ, determine the associated probability density f via Fourier inversion:

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \psi(t) e^{-itx}\, dt.$$

Define the pmfs p_0 and p_1 as in (14) and (15). Set $p_{U|0} = p_{V|0} = p_0$ and $p_{U|1} = p_{V|1} = p_1$. This implies that $p_U = p_V = p$, where p is as defined in (13).

Clearly, (7) holds. To verify (9), note that, by virtue of (16), we have for $a \in \{0, 1\}$,

$$\varphi_U \varphi_V = \varphi^2 = \varphi \varphi_a.$$

But, by construction, $\varphi_U \varphi_{V|a} = \varphi_V \varphi_{U|a} = \varphi \varphi_a$.

Therefore, by Lemma 3, the random variables (U, V, X, Y) with joint pmf given by (6) have the properties (S1)–(S3).

It remains to prove the last statement in the theorem. This follows from the fact [15, pp. 512–513] that a probability distribution F with characteristic function χ has finite variance iff χ is twice differentiable; in this case, $\chi'(0) = i\mu$ and $\chi''(0) = -\mu_2$, where μ and μ_2 are the mean and second moment of F. Thus, the pmf p under consideration has finite variance (to be precise, the distribution specified by p has finite variance) iff φ is twice differentiable. By construction, φ is twice differentiable iff ψ is twice differentiable. In this case, as φ is real, so is $\varphi'(0)$, which implies that p has zero mean. Hence, the variance of p is equal to its second moment, and the final assertion of the theorem follows, since $\varphi''(0) = \psi''(0)$. ■

Based on Theorem 7, secure computation of XOR at the relay works as follows: the nodes `A` and `B` modulate their bits independently to an integer k, with probability $p_0(k)$ (from (14)) if the bit is 0, or with probability $p_1(k)$ (from (15)) if the bit is 1. The probability distributions can be chosen such that the modulated symbols have finite average power. The average transmit power is equal to the variance of the modulated random variable, which is $-\psi''(0)$, and a handle on this can be obtained by choosing ψ carefully. The relay receives the sum of the two integers, which is independent of the individual bits X and Y (of `A` and `B` respectively). However, the XOR of the two bits can be recovered at `R` with probability 1. This is done by simply mapping the received integer W to 1, if W is odd, and 0 if W is even. To gain a better understanding of the construction of the rvs, let us see an example.

As recorded previously (see Table 1 of [15, Section XV.2]), the probability density function f given in (12) has characteristic function

$$\psi(t) = \max\{0, 1 - 2|t|/\pi\}.$$

This function is plotted in Figure 7.

From (13)–(15), we have

$$\begin{aligned} p(k) &= \begin{cases} \frac{1}{4} & \text{if } k = 0 \\ \frac{2}{\pi^2 k^2}(1 - \cos k\pi/2) & \text{if } k \neq 0 \end{cases} \\ &= \begin{cases} \frac{1}{4} & \text{if } k = 0 \\ \frac{2}{\pi^2 k^2} & \text{if } k \text{ is odd} \\ \frac{4}{\pi^2 k^2} & \text{if } k \equiv 2 \pmod 4 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{17}$$

$$p_0(k) = \begin{cases} \frac{1}{2} & \text{if } k = 0 \\ \frac{8}{\pi^2 k^2} & \text{if } k \equiv 2 \pmod 4 \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

and

$$p_1(k) = \begin{cases} \frac{4}{\pi^2 k^2} & \text{if } k \text{ is odd} \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

Note that the pmf p in (17) does not have a finite second moment, and indeed, ψ is not differentiable at 0. However, it is possible to construct compactly supported, twice-differentiable characteristic functions ψ. We give here an explicit construction of such a characteristic function.

Consider the density (from [15, Section XV.2, Table 1])

$$f(x) = \begin{cases} \frac{1}{2\pi} & \text{if } x = 0 \\ \frac{1-\cos x}{\pi x^2} & \text{if } x \neq 0 \end{cases} \tag{20}$$

which has characteristic function

$$\hat{f}(t) = \max\{0, 1 - |t|\} \tag{21}$$

The function $\hat{f}$ has a triangular graph as in Figure 7, except that the base is $[-1, 1]$. In particular, $\hat{f}(t) = 0$ for $|t| \geq 1$.

The function $g = \hat{f} * \hat{f}$, where $*$ denotes convolution, can be explicitly computed to be

$$g(t) = (\hat{f} * \hat{f})(t) = \begin{cases} \frac{1}{2}|t|^3 - t^2 + \frac{2}{3} & \text{if } |t| \leq 1 \\ \frac{1}{6}(2 - |t|)^3 & \text{if } 1 \leq |t| \leq 2 \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

Proposition 8. *The function $h(x) = (3\pi^2/4)\,[f(\pi x/4)]^2$, with f as in (20), is a density function whose characteristic function is given by*

$$\psi(t) = \tfrac{3}{2}\, g(\tfrac{4t}{\pi}),$$

where g is as in (22). The function ψ is non-negative with $\psi(t) = 0$ for $|t| \geq \pi/2$. Furthermore, ψ is twice differentiable, with $\psi''(0) = -48/\pi^2$.

Thus, rvs U and V can be constructed as in Theorem 7 with $\mathrm{var}(U) = \mathrm{var}(V) = 48/\pi^2$.

Proof of Proposition 8. The stated properties of the function ψ can be directly verified from (22). We will show here that h is a density function with characteristic function ψ.

Note first that $\hat{f}$ defined in (21) is also a probability density function — it is non-negative and its integral over $(-\infty, \infty)$ is 1. By Fourier inversion, its characteristic function is $2\pi f$. Therefore, $g = \hat{f} * \hat{f}$ is a density with characteristic function $4\pi^2 f^2$.

Now, f^2 is integrable since $(\hat{f})^2$ is integrable (see corollary to Theorem 3 of Section XV.3 of [15]). Hence, $\tilde{h}(x) = f^2(x)/(\int_{-\infty}^{\infty} f^2(y)\,dy)$ is a probability density function. The integral in the denominator can be explicitly evaluated by means of the Plancherel identity:

$$\int_{-\infty}^{\infty} f^2(y)\,dy = \frac{1}{2\pi}\int_{-\infty}^{\infty} [\hat{f}(t)]^2\,dt = \frac{1}{2\pi}\,g(0) = \frac{1}{3\pi},$$

the last equality following from (22). Thus, $\tilde{h}(x) = 3\pi f^2(x)$.

From the fact that $4\pi^2 f^2$ is the characteristic function of g, it follows by Fourier inversion that $\tilde{h}$ has characteristic function given by $\tilde{\psi}(t) = \frac{3}{2}\,g(t)$. Hence, $h(x) = (\pi/4)\tilde{h}(\pi x/4)$ is a density function with characteristic function $\tilde{\psi}(4t/\pi)$, which is precisely $\psi(t)$. □

Remark 9. *It is even possible to construct compactly supported C^∞ characteristic functions. Constructions of such functions are given in [31]. In fact, [31] constructs compactly supported characteristic functions ψ such that the corresponding density functions f are even functions satisfying $\lim_{x\to\infty} x^m f(x) = 0$ for all $m > 0$. This implies that all the absolute moments $\int_{-\infty}^{\infty} |x|^m f(x)\,dx$ exist, and hence, ψ is a C^∞ function (see [15, p. 512]). If such a characteristic function ψ is used in the construction described in Theorem 7, then the resulting $\mathbb{Z}$-valued rvs U, V will have pmfs $p_U(k), p_V(k)$ whose tails decay faster than any polynomial in k. To be precise, $\lim_{k\to\infty} k^m p_U(k) = \lim_{k\to\infty} k^m p_V(k) = 0$ for any $m > 0$.*

The above remark shows that we can have $\mathbb{Z}$-valued rvs U, V satisfying properties (S1)–(S3), with pmfs decaying faster than any polynomial. However, the rate of decay cannot be much faster than that. Indeed, it is not possible to construct $\mathbb{Z}$-valued rvs with exponentially decaying pmfs that satisfy properties (S1)–(S3). Define a pmf $p(k)$, $k \in \mathbb{Z}$, to be *light-tailed* if there are positive constants C and λ such that $p(k) \leq C\lambda^{-|k|}$ for all sufficiently large $|k|$.

Proposition 10. *Properties (S1)–(S3) cannot be satisfied by integer-valued rvs U, V having light-tailed pmfs.*

Proof.[5] Suppose that U, V are $\mathbb{Z}$-valued rvs satisfying (S1) and (S2). Using $\varphi_U = (1/2)(\varphi_{U|0} + \varphi_{U|1})$ and $\varphi_V = (1/2)(\varphi_{V|0} + \varphi_{V|1})$ in (9), we readily obtain

$$\varphi_{U|0}^2 = \varphi_{U|1}^2 \quad \text{and} \quad \varphi_{V|0}^2 = \varphi_{V|1}^2. \tag{23}$$

If U, V have light-tailed pmfs, then $p_{U|a}$ and $p_{V|a}$, $a \in \{0, 1\}$, must also be light-tailed, since $p_{U|a} \leq 2p_U$ and $p_{V|a} \leq 2p_V$. The key observation is that the characteristic function of a light-tailed pmf is real-analytic, i.e., it has a power series expansion $\sum_{n=0}^{\infty} c_n t^n$, with $c_n \in \mathbb{C}$, that is valid for all $t \in \mathbb{R}$ [21, Chapter 7]. Thus, $\varphi_{U|a}$ and $\varphi_{V|a}$, for $a \in \{0, 1\}$, are real-analytic. It is an easy fact, provable by comparing power series coefficients, that if functions g and h are real-analytic and $g^2 = h^2$, then either $g = h$ or $g = -h$. Applying this to (23), we find that $\varphi_{U|0} = \pm\varphi_{U|1}$, and similarly for V. In fact, since φ_U and φ_V cannot be identically 0, we actually have $\varphi_{U|0} = \varphi_{U|1} = \varphi_U$, and similarly for V. This implies that $U \perp\!\!\!\perp X$ and $V \perp\!\!\!\perp Y$. From this, and (S1), we obtain that $U + V \perp\!\!\!\perp X \oplus Y$, thus precluding (S3). ∎

D. Extension to Finite Abelian Groups

A close look at the modulations in the previous section reveals the following structure: we had a fine lattice $\Lambda = \mathbb{Z}$ and a coarse lattice $\Lambda_0 = 2\mathbb{Z}$, with the quotient group Λ/Λ_0, consisting of the two cosets $2\mathbb{Z}$ and $1 + 2\mathbb{Z}$, making up the probabilistically-chosen modulation alphabet. Given a message $X \in \Lambda/\Lambda_0$, the encoder outputs a random point from the coset X according to a carefully chosen probability distribution. Note that the quotient group in this case is isomorphic to $\mathbb{Z}_2$, and this enables recovery of the XOR of the bits (addition in $\mathbb{Z}_2$) from integer addition of transmitted symbols modulo the coarse lattice. Also, the choice of the probability distribution (from Theorem 7) ensures that the choice of coset at each transmitter is independent of the integer sum at the relay. We shall extend the construction described in the previous subsection to d dimensions, thereby obtaining a scheme that satisfies properties (S1)–(S3).

Now, any finite Abelian group $\mathbb{G}$ can be expressed as the quotient group Λ/Λ_0 for some pair of nested lattices $\Lambda_0 \subseteq \Lambda$. Indeed, any such $\mathbb{G}$ is isomorphic to a direct sum of cyclic groups: $\mathbb{G} \cong \mathbb{Z}_{N_1} \oplus \mathbb{Z}_{N_2} \oplus \cdots \oplus \mathbb{Z}_{N_k}$ for some positive integers $N_1, N_2, \ldots, N_k$ [18, Theorem 2.14.1]. Here, $\mathbb{Z}_{N_j}$ denotes the group of integers modulo-N_j. Taking $\Lambda = \mathbb{Z}^d$ and $\Lambda_0 = \mathsf{A}^T \mathbb{Z}^d$, where A is the diagonal matrix $\mathrm{diag}(N_1, N_2, \ldots, N_k)$, we have $\mathbb{G} \cong \Lambda/\Lambda_0$. So, the finite Abelian group case is equivalent to considering the quotient group, i.e., the group of cosets, of a coarse lattice Λ_0 within a fine lattice Λ. These lattices may be taken to be full-rank lattices in $\mathbb{R}^d$.

As an example, let $N \geq 2$ be an integer, and let $\mathbb{Z}_N = \{0, 1, \ldots, N-1\}$ denote the set of integers modulo N. Let X, Y be iid random variables uniformly distributed over $\mathbb{Z}_N$, and let $X \oplus Y$ now denote their modulo-N sum. Similar to the binary case discussed so far, given a non-negative real characteristic function ψ such that $\psi(t) = 0$ for $|t| \geq \pi/N$, we can construct $\mathbb{Z}$-valued random variables U, V, jointly distributed with

[5]This proof was conveyed to the authors by Manjunath Krishnapur.

X, Y, for which properties (S1)–(S3) hold. In this case, the finite Abelian group can be taken as the group of cosets of the coarse lattice $N\mathbb{Z}$ within the fine lattice $\mathbb{Z}$, which is isomorphic to $\mathbb{Z}_N$.

Let Λ_0 be a sublattice of Λ of index M (i.e., the number of cosets of Λ_0 in Λ is M). List the cosets of Λ_0 in Λ as $\Lambda_0, \Lambda_1, \ldots, \Lambda_{M-1}$, which constitute the quotient group $\mathbb{G} = \Lambda/\Lambda_0$. As before, $\oplus$ denotes addition within $\mathbb{G}$.

Consider rvs X, Y uniformly distributed over $\mathbb{G}$. We wish to construct rvs U, V taking values in Λ, having the properties (S1)–(S3). The following theorem shows that this is possible. Here, $\mathbb{R}^+$ denotes the set of all non-negative real numbers.

Theorem 11. *Suppose that $\psi : \mathbb{R}^d \to \mathbb{R}^+$ is the characteristic function of a probability density function $f : \mathbb{R}^d \to \mathbb{R}^+$, such that $\psi(\mathbf{t}) = 0$ for $\mathbf{t} \notin \mathcal{V}(\hat{\Lambda}_0)$, where $\hat{\Lambda}_0$ is the Fourier dual of Λ_0. For $j = 0, 1, \ldots, M-1$, define the pmf p_j as follows:*

$$p_j(\mathbf{k}) = \begin{cases} |\det \Lambda_0| f(\mathbf{k}) & \text{if } \mathbf{k} \in \Lambda_j \\ 0 & \text{otherwise.} \end{cases} \tag{24}$$

Finally, define a random variable U (resp. V) jointly distributed with X (resp. Y) as follows: if $X = \Lambda_j$ (resp. $Y = \Lambda_j$), U (resp. V) is a random point from Λ_j picked according to the distribution p_j. Then, the resulting Λ-valued rvs U, V satisfy properties (S1)–(S3). Additionally, if ψ is twice differentiable, then $\mathbb{E}\|U\|^2 = \mathbb{E}\|V\|^2 = -\nabla^2\psi(\mathbf{0})$, where $\nabla^2 = \sum_{j=1}^d \partial_j^2$ is the Laplacian operator.

As with Theorem 7 and XOR, the above theorem allows for secure computation at the relay of the group operation $X \oplus Y$. The theorem is proved in a manner completely analogous to Theorem 7, the main difference being that the multi-dimensional Poisson summation formula is used in place of (10). The interested reader is directed to Appendix A for the proof.

Constructing compactly supported twice-differentiable (or even C^∞) characteristic functions $\psi : \mathbb{R}^d \to \mathbb{R}^+$, $d \geq 1$, is straightforward, given our previous constructions of such functions from $\mathbb{R}$ to $\mathbb{R}^+$. Suppose that for $i = 1, 2, \ldots, d$, $\psi_i : \mathbb{R} \to \mathbb{R}^+$ is the characteristic function of a random variable X_i, such that $\psi_i(t) = 0$ for $|t| \geq \lambda_i$, with $\lambda_i > 0$, and $X_1, X_2, \ldots, X_d$ are mutually independent. Then, $\psi(t_1, \ldots, t_d) = \prod_{i=1}^d \psi_i(t_i)$ is the characteristic function of the random vector $\mathbf{X} = (X_1, \ldots, X_d)$. Note that ψ is compactly supported: $\psi(\mathbf{t}) = 0$ for $\mathbf{t} \notin \prod_{i=1}^d (-\lambda_i, \lambda_i)$. Moreover, if the ψ_is are twice-differentiable (or C^∞) for all i, then so is ψ.

Our objective is to design codes (as defined in Definition 1) for secure computation at the relay. With the construction described above, the rate of the code depends on the number of cosets, M, of Λ_0 in Λ. For a given average power constraint, the system designer is usually faced with the task of maximizing the rate. Equivalently, for a given rate, the average transmit power must be kept as small as possible. The transmit power is equal to the second moment of $\mathbf{U}$ (or $\mathbf{V}$). Therefore, while any characteristic function ψ supported within $\mathcal{V}(\hat{\Lambda}_0)$ suffices for the construction of Theorem 11, we must use a ψ for which $-\nabla^2\psi(\mathbf{0})$ is the least among such ψ's. This would yield random variables U and V of least second moment (and hence

least transmit power), and having the desired properties.

It is evident that by simply scaling the nested lattice pair, the average transmit power may be made as small as required. Suppose that the random vectors $\mathbf{U}$ and $\mathbf{V}$, distributed over a fine lattice Λ, have second moment P. Then, for any $\alpha > 0$, the random variables $\mathbf{U}' = \alpha\mathbf{U}$ and $\mathbf{V}' = \alpha\mathbf{V}$, distributed over $\alpha\Lambda := \{\alpha\mathbf{z} : \mathbf{z} \in \Lambda\}$ have second moment $\alpha^2 P$. Choosing a small enough α would suffice to satisfy the power constraint. However, as we will see in the following sections, when we have to deal with the additive noise in the MAC channel, it is not possible to scale down the lattice arbitrarily if the probability of error is to be made small. Also, for a given (fixed) coarse lattice, it turns out that the second moment (which depends solely on the choice of ψ) cannot be made arbitrarily small. Indeed, the following result, adapted from [10], gives a precise and complete answer to the question of how small $-\nabla^2\psi(\mathbf{0})$ can be for a characteristic function ψ supported within a ball of radius ρ in $\mathbb{R}^d$.

Theorem 12 ([10], Theorem 5.1)**.** *Fix a $\rho > 0$. If ψ is a characteristic function of a random variable distributed over $\mathbb{R}^d$ such that $\psi(\mathbf{t}) = 0$ for $\|\mathbf{t}\| \geq \rho$, then*

$$-\nabla^2\psi(\mathbf{0}) \geq \frac{4}{\rho^2}\, j^2_{\frac{d-2}{2}}, \tag{25}$$

with equality iff $\psi(\mathbf{t}) = \widetilde{\psi}(\mathbf{t}/\rho)$ for $\widetilde{\psi} = \omega_d \tilde{*} \omega_d$. Here, $\omega_d(\mathbf{t}) = \gamma_d\, \Omega_d(2\|\mathbf{t}\| j_{\frac{d-2}{2}})$ for $\|\mathbf{t}\| \leq 1/2$ and $\omega_d(\mathbf{t}) = 0$ for $\|\mathbf{t}\| > 1/2$, and

$$\omega_d \tilde{*} \omega_d(\mathbf{t}) = \int \omega_d(\boldsymbol{\tau})\overline{\omega_d(\mathbf{t}+\boldsymbol{\tau})} d\boldsymbol{\tau}$$

denotes the folded-over self convolution of ω_d, with $\overline{\omega_d(\mathbf{t})}$ denoting the complex conjugate of $\omega_d(\mathbf{t})$. Also, j_k denotes the first positive zero of the Bessel function J_k. Furthermore, for $t \in \mathbb{R}$,

$$\Omega_d(t) = \Gamma(d/2)\Big(\frac{2}{t}\Big)^{\frac{d-2}{2}} J_{\frac{d-2}{2}}(t)$$

and

$$\gamma_d^2 = \frac{4 j^{d-2}_{\frac{d-2}{2}}}{\pi^{d/2}\Gamma(d/2) J^2_{\frac{d}{2}}(j_{\frac{d-2}{2}})},$$

where $\Gamma(\cdot)$ denotes the Gamma function. The density f corresponding to the minimum-variance ψ is given by $f(\mathbf{x}) = \rho^d \tilde{f}(\rho\mathbf{x})$, where

$$\tilde{f}(\mathbf{x}) = c_d \left(\frac{\Omega_d(\|\mathbf{x}\|/2)}{j^2_{\frac{d-2}{2}} - (\|\mathbf{x}\|/2)^2} \right)^2, \tag{26}$$

where

$$c_d = \frac{4 j^{d-2}_{\frac{d-2}{2}}}{4^d \pi^{d/2}\Gamma(d/2)}.$$

Remark 13. *Observe that Theorem 11 is true for any nested lattice pair (Λ, Λ_0). As long as $\psi(\mathbf{t})$ is a characteristic function supported within $\mathcal{V}(\Lambda_0)$, we have an encoding scheme that satisfies (S1)–(S3). If we restrict ψ to be supported within a ball of radius ρ, which is contained within $\mathcal{V}(\hat{\Lambda}_0)$, then Theorem 12 gives*

us a suitable candidate for ψ that can be used to obtain perfect secrecy. Since we are interested in minimizing the transmission power, we can choose ρ to be as large as $r_{pack}(\hat{\Lambda}_0)$, where $r_{pack}(\hat{\Lambda}_0)$ denotes the packing radius of $\hat{\Lambda}_0$. Hence, we now have a coding scheme that achieves perfect secrecy for any arbitrary nested lattice pair. This is rather interesting, since earlier work on weak and strong secrecy using lattices [16], [17], [20] invariably required that the nested lattices satisfy certain goodness properties. Therefore, ours is a very general result, and explicit in the sense that once we fix the nested lattice pair, we can exactly specify what distribution should be used to randomize at the encoder to obtain secrecy.

VI. The Gaussian Noise Setting

Given any nested lattice pair, we now have a scheme whereby the relay can compute $X \oplus Y$ from $\mathbf{U}+\mathbf{V}$, but cannot determine X or Y separately. We next consider the scenario where the symbols received by the relay are corrupted by noise, and prove the achievability of the power-rate pairs described in Theorem 1. Recall that in the MAC phase, the relay receives

$$\mathbf{W} = \mathbf{U} + \mathbf{V} + \mathbf{Z},$$

where $\mathbf{Z}$ is zero-mean iid Gaussian noise with variance σ^2. The coding scheme that we use is largely based on the work in [12], [23], and is described below.

A. Coding Scheme for Perfect Secrecy

We now describe the sequence of $(d, M^{(d)})$ (recall Definition 1) codes that achieve perfect secrecy.

Code: A $(\Lambda^{(d)}, \Lambda_0^{(d)})$ *nested lattice code* consists of a pair of full-rank nested lattices $\Lambda_0^{(d)} \subseteq \Lambda^{(d)}$ in $\mathbb{R}^d$. The messages are chosen from the group $\mathbb{G}^{(d)} = \Lambda^{(d)}/\Lambda_0^{(d)}$, whose $M^{(d)} := |\Lambda^{(d)}/\Lambda_0^{(d)}|$ elements are listed as $\Lambda_0, \Lambda_1, \ldots, \Lambda_{M^{(d)}-1}$.

Encoding: We have messages X, Y at nodes $\mathtt{A}, \mathtt{B}$ that are independent rvs, uniformly distributed over $\mathbb{G}^{(d)}$. We first pick a characteristic function ψ supported within $\mathcal{V}(\hat{\Lambda}_0^{(d)})$, as needed in Theorem 11. We impose the restriction that ψ be supported within a ball centered at $\mathbf{0}$ with radius equal to the packing radius, $r_{\text{pack}}(\hat{\Lambda}_0^{(d)})$, of the dual lattice $\hat{\Lambda}_0^{(d)}$. Recall that the packing radius is, by definition, the largest radius of a ball centered at $\mathbf{0}$ that is contained within $\mathcal{V}(\hat{\Lambda}_0^{(d)})$. So, if $\psi(\mathbf{t}) = 0$ for $\|\mathbf{t}\| \geq r_{\text{pack}}(\hat{\Lambda}_0)$, then $\psi(\mathbf{t})$ is certainly supported within $\mathcal{V}(\hat{\Lambda}_0)$. If $X = \Lambda_j$, node $\mathtt{A}$ transmits a random vector $\mathbf{U} \in \Lambda_j$ picked according to the distribution p_j of Theorem 11. Similarly, if $Y = \Lambda_k$, node $\mathtt{B}$ transmits a random vector $\mathbf{V} \in \Lambda_k$ picked according to the distribution p_k. The rate of transmission from $\mathtt{A}$ or $\mathtt{B}$ is $R^{(d)} = \frac{1}{d}\log_2 M^{(d)}$. The average transmit power per dimension at each node is $P^{(d)} = \frac{-\nabla^2\psi(\mathbf{0})}{d}$, as in Theorem 11.

From Theorem 12, we see that an average transmit power per dimension as low as

$$P^{(d)} = \frac{4j^2_{\frac{d-2}{2}}}{d\left(r_{\text{pack}}(\hat{\Lambda}_0^{(d)})\right)^2}, \tag{27}$$

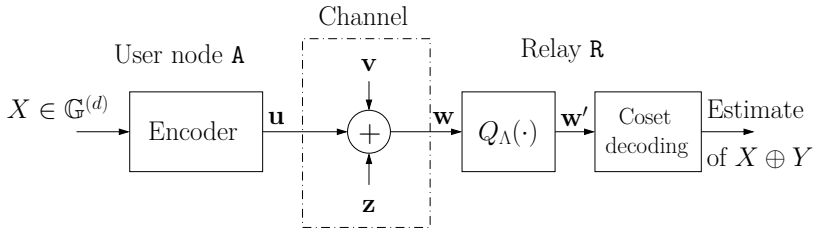

Fig. 10. The operations performed by the user nodes and the relay.

is achievable by a suitable choice of ψ. It was shown in [33] (see also [14]) that the first positive zero of the Bessel function J_k can be written as $j_k = k + bk^{1/3} + \mathcal{O}(k^{-1/3})$, where b is a constant independent of k. Therefore,

$$P^{(d)} = \frac{d}{{r_{\text{pack}}}^2(\hat{\Lambda}_0^{(d)})}(1 + o_d(1)), \tag{28}$$

where $o_d(1) \to 0$ as $d \to \infty$, is achievable by a suitable choice of ψ using Theorem 12.

Decoding: The relay R receives $\mathbf{W} = \mathbf{U} + \mathbf{V} + \mathbf{Z}$, where $\mathbf{Z}$ is a Gaussian noise vector with d independent $\mathcal{N}(0, \sigma^2)$ components, which are all independent of $\mathbf{U}$ and $\mathbf{V}$. The relay estimates $\Lambda_j \oplus \Lambda_k$ to be the coset of $\Lambda_0^{(d)}$ represented by the closest vector to $\mathbf{W}$ in the lattice $\Lambda^{(d)}$, which we denote by $Q_{\Lambda^{(d)}}(\mathbf{W})$.

Security: Since the noise $\mathbf{Z}$ is independent of everything else, Theorem 11 shows that $\mathbf{W}$ is independent of the individual messages X, Y. Hence, even in the noisy setting, perfect security continues to be guaranteed at the relay for any choice of the nested lattice code.

Reliability and achievable rate: Let $\eta^{(d)}$ denote the average probability that $Q_\Lambda(\mathbf{W})$ is different from the coset to which $\mathbf{U} + \mathbf{V}$ belongs. From Definition 2 of achievable power-rate pairs, a pair $(\mathcal{P}, \mathcal{R})$ is achievable if for every $\delta > 0$, there exists a sequence of nested lattice codes $(\Lambda^{(d)}, \Lambda_0^{(d)})$ for which the following hold for sufficiently large d: $R^{(d)} > \mathcal{R} - \delta$, $P^{(d)} < \mathcal{P} + \delta$ and $\eta^{(d)} < \delta$.

For a given nested lattice pair, from Theorem 12, we know the minimum average transmit power per dimension that guarantees perfect secrecy (subject to the condition that the characteristic function is supported within a ball of radius $r_{\text{pack}}(\hat{\Lambda}_0^{(d)})$), and the pmf p_j that achieves the minimum. The choice of the nested lattices determines the error performance of the decoder, and consequently determines reliable transmission rates. For a given transmission power constraint, it is desirable to maximize rate, without compromising on the reliability of computation. The average transmit power, as given by (27), is a function of $r_{\text{pack}}(\hat{\Lambda}_0^{(d)})$ which is a property of the Fourier dual of the coarse lattice. On the other hand, the transmission rate depends on the number of cosets of $\Lambda_0^{(d)}$ in $\Lambda^{(d)}$, which is the number of fine lattice points in the fundamental Voronoi region of the coarse lattice, $\mathcal{V}(\Lambda_0^{(d)})$. In order to maximize the rate, we must pack in as many points of $\Lambda^{(d)}$ as possible in $\mathcal{V}(\Lambda_0^{(d)})$. But having the points of $\Lambda^{(d)}$ too close together adversely affects the error performance of the coding scheme in the presence of noise. A trade-off must be made, and we need to find the maximum rate below which reliable computation of $X \oplus Y$ can be guaranteed. We will restrict

the class of nested lattice pairs to those which give good error performance in the presence of AWGN so that the relay can compute $X \oplus Y$ reliably. To guarantee secure and reliable computation at the relay, we restrict the class of nested lattice pairs $(\Lambda^{(d)}, \Lambda_0^{(d)})$ to those which satisfy the following properties[6]:

(G_1) The sequence of coarse lattices, $\{\Lambda_0^{(d)}\}$, is good for covering and AWGN channel coding.

(G_2) The sequence of dual lattices, $\{\hat{\Lambda}_0^{(d)}\}$, is good for packing.

(G_3) The sequence of fine lattices, $\{\Lambda^{(d)}\}$, is good for AWGN channel coding.

Unlike prior work on nested lattices [1], [12], [23], [25] which required $\{\Lambda_0^{(d)}\}$ and $\{\Lambda^{(d)}\}$ to satisfy properties (G_1) and (G_3) above, we have the additional requirement that the Fourier dual, $\{\Lambda_0^{(d)}\}$ must be good for packing. While the existence of nested lattices that satisfy (G_1) and (G_2) is well established [12], [13], [23], the existence of lattices whose duals simultaneously satisfy these goodness properties seems to be unexplored. In the next section, we will describe a construction of nested lattices based on [12], [23] and show that there are nested lattices that satisfy the above properties.

B. Good Ensembles of Nested Lattices with Good Duals

Let d and k be positive integers with $k \leq d$, and let q be a prime number. Let $\mathbb{Z}_q$ denote the field of integers modulo q.

1) Choose a $k \times d$ matrix G uniformly at random over $\mathbb{Z}_q^{k \times d}$. This is done by choosing each element of G (there are dk of them) uniformly over $\mathbb{Z}_q$, and independently of the other entries. Note that G need not be full-rank. However, the probability that G is full-rank goes to 1 as $(d-k)$ tends to ∞ [13]. The linear code over $\mathbb{Z}_q$ generated by G is denoted by $\mathcal{C}(\mathsf{G})$.
2) Apply Construction A on the code generated above. This is done as follows:
   (c_1) The codebook generated using G is $\mathcal{C}(\mathsf{G}) = \{(\mathsf{G}^T\mathbf{y}) \bmod q : \mathbf{y} \in \mathbb{Z}_q^k\}$.
   (c_2) The codebook is then scaled so that the codeword coordinates are restricted to lie within the d-dimensional unit cube: $\mathcal{C}' = (1/q)\mathcal{C}(\mathsf{G}) = \{(1/q)\mathbf{x} : \mathbf{x} \in \mathcal{C}(\mathsf{G})\}$. The points in the set $\mathcal{C}'$ lie on the vertices of a rectangular grid of side $1/q$ sitting within the unit cube.
   (c_3) The lattice is obtained by repeating these points over the entire space, $\mathbb{R}^d$, i.e., $\tilde{\Lambda}(\mathcal{C}) = \mathcal{C}' + \mathbb{Z}^d := \{\mathbf{c} + \mathbf{x} : \mathbf{c} \in \mathcal{C}', \mathbf{x} \in \mathbb{Z}^d\}$.

Following the terminology used in [13], we will henceforth call the above as the (d, k, q) *ensemble*. Also, from the construction, it is clear that $\mathbb{Z}^d$ is a sublattice of $\tilde{\Lambda}(\mathcal{C})$. More detail regarding Construction-A lattices can be found in [6].

We now describe the construction of the $(\Lambda^{(d)}, \Lambda_0^{(d)})$ nested lattice codes. This is based on [12], [23]. Choose a sequence of coarse lattices $\{\Lambda_0^{(d)}\}$, each $\Lambda_0^{(d)}$ selected uniformly at random from the (d, k, q) ensemble, where k and q may be functions of d chosen beforehand. Let $\mathsf{A}^{(d)}$ be the lattice generator matrix of $\Lambda_0^{(d)}$,

[6]For definitions of lattices good for covering, packing, and AWGN channel coding, the reader is directed to Appendix B.

for $d = 1, 2, ...$ For this choice of $\{\Lambda_0^{(d)}\}$, we construct another ensemble of lattices from which we pick the sequence of fine lattices $\{\Lambda^{(d)}\}$. This consists of two steps:

(f_1) Choose a sequence of lattices, $\{\tilde{\Lambda}_f^{(d)}\}$, with each $\tilde{\Lambda}_f^{(d)}$ coming from the (d, k_1, q_1) ensemble of Construction-A lattices. The parameters k_1 and q_1 could be possibly different from k and q. As mentioned earlier, $\tilde{\Lambda}_f^{(d)}$ contains $\mathbb{Z}^d$ as a sublattice. If the generator matrix of $\tilde{\Lambda}_f^{(d)}$ has full rank, then the number of cosets of $\mathbb{Z}^d$ in $\tilde{\Lambda}_f^{(d)}$ is $q_1^{k_1}$.

(f_2) The lattice $\tilde{\Lambda}_f^{(d)}$ is subjected to a linear transformation by the matrix $(\mathsf{A}^{(d)})^T$, to get $\Lambda^{(d)} = (\mathsf{A}^{(d)})^T \tilde{\Lambda}_f^{(d)} := \{(\mathsf{A}^{(d)})^T \mathbf{y} : \mathbf{y} \in \tilde{\Lambda}_f^{(d)}\}$.

We will call this ensemble of $(\Lambda^{(d)}, \Lambda_0^{(d)})$ pairs as the (d, k, q, k_1, q_1) *ensemble*. The lattice pair can be scaled appropriately so as to satisfy the average power constraint. We have $M^{(d)} = |\Lambda^{(d)}/\Lambda_0^{(d)}| = q_1^{k_1}$ with probability tending to 1 as $d - k$ tends to ∞ [23]. Hence, the rate of the $(\Lambda^{(d)}, \Lambda_0^{(d)})$ code will be

$$R^{(d)} = \frac{k_1}{d} \log_2(q_1). \tag{29}$$

By choosing q_1 and k_1 appropriately, the rate can be varied.

In order to obtain nested lattices from the (d, k, q, k_1, q_1) ensemble that satisfy properties (G_1)–(G_3), we will restrict the choice of k, q, k_1, q_1. We will make sure the values of these parameters can be chosen so that, in spite of the restriction, the transmission rate, as given by (29), can be chosen freely.

1) Choice of k, q, k_1, q_1: Let Λ be a full-rank lattice chosen from the (d, k, q) ensemble. From the construction, we can see that Λ has $\mathbb{Z}^d$ as a sublattice. The nesting ratio of the $(\Lambda, \mathbb{Z}^d)$ nested lattice pair, equal to q^k, is equal to the ratio of the volume of the fundamental Voronoi region of $\mathbb{Z}^d$ to that of Λ. The volume of $\mathcal{V}(\Lambda)$ is in turn equal to the volume of a d-dimensional ball of radius $r_{\text{eff}}(\Lambda)$, and hence,

$$q^k = \frac{\Gamma\left(\frac{d}{2} + 1\right)}{\pi^{d/2} \left(r_{\text{eff}}(\Lambda)\right)^d} \tag{30}$$

Rearranging, we get

$$r_{\text{eff}}(\Lambda) = \left(\frac{\Gamma\left(\frac{d}{2} + 1\right)}{\pi^{d/2} q^k} \right)^{1/d}. \tag{31}$$

We choose

$$k = \beta_0 d, \text{ and } k_1 = \beta_1 d, \tag{32}$$

for some $0 < \beta_0, \beta_1 < 1/2$, and q and q_1 are prime numbers chosen such that

$$\lim_{d \to \infty} \frac{d}{q_1} = 0, \text{ and } r_{\min}^{(0)} < r_{\text{eff}}(\Lambda_0^{(d)}) < 2 r_{\min}^{(0)}, \tag{33}$$

for some $0 < r_{\min}^{(0)} < 1/4$. It is possible to choose primes that satisfy the above conditions. Choosing q_1 to grow faster than d is sufficient to satisfy the first condition. To see that q can be chosen so as to satisfy the second constraint, substitute (30) in (33) ,

$$\frac{\Gamma\left(\frac{d}{2} + 1\right)}{\pi^{d/2} \left(2 r_{\min}^{(0)}\right)^d} < q^k < \frac{\Gamma\left(\frac{d}{2} + 1\right)}{\pi^{d/2} \left(r_{\min}^{(0)}\right)^d}.$$

Since $k = \beta_0 d$, we can rewrite the above inequalities as follows:

$$\left(\frac{\Gamma\left(\frac{d}{2}+1\right)}{\pi^{d/2}\left(2r_{\min}^{(0)}\right)^d}\right)^{\frac{1}{\beta_0 d}} < q < 2^{1/\beta_0}\left(\frac{\Gamma\left(\frac{d}{2}+1\right)}{\pi^{d/2}\left(2r_{\min}^{(0)}\right)^d}\right)^{\frac{1}{\beta_0 d}}.$$

Let α denote the left hand side of the above inequality. From (32), we have $\beta_0 < 1/2$, and hence it is enough to show that there exists a prime q which satisfies $\alpha < q < 4\alpha$. For any $\alpha > 3/2$, there exists an integer n such that $\alpha < n \leq \alpha + 1$, and $2n < 4\alpha$ (since for $\alpha > 3/2$, $2(\alpha+1) < 4\alpha - 1$). By Bertrand's postulate (see e.g., [29]), for every positive integer n, there exists a prime number between n and $2n$, and therefore, we can choose a prime q satisfying (33). We then have the following lemma, which is proved in Appendix C.

Lemma 14. *Let $(\Lambda^{(d)}, \Lambda_0^{(d)})$ be a nested lattice pair chosen uniformly at random from the (d, k, q, k_1, q_1) ensemble, with the parameters k, q, k_1, q_1 chosen so as to satisfy (32) and (33). Then, the probability that $(\Lambda^{(d)}, \Lambda_0^{(d)})$ satisfies (G_1)–(G_3) tends to one as d approaches infinity.*

C. Achievable Rate

We now find achievable transmission rates for reliable and secure computation of $X \oplus Y$ at the relay. This is largely based on [12], [23], [24]. We find sufficient conditions on the transmission rate to ensure that the probability of error in decoding $X \oplus Y$ from $\mathbf{W}$ goes to zero for large block lengths. To study the error performance of the decoder, let us study the relay in more detail. Recall from Section II-A that $Q_{\Lambda^{(d)}}(\cdot)$ denotes the lattice quantizer that maps a d-dimensional real vector to the closest point in $\Lambda^{(d)}$. Similarly, $Q_{\Lambda_0^{(d)}}(\cdot)$ is the lattice quantizer that maps vectors in $\mathbb{R}^d$ to the closest point in $\Lambda_0^{(d)}$. Let $\mathcal{D}(\cdot)$ denote the decoder map as defined in Section VI-A. The quantity $\mathcal{D}(\mathbf{W})$ is the estimate of $X \oplus Y$ made by the relay, and equal to the coset of $\Lambda_0^{(d)}$ to which $Q_{\Lambda^{(d)}}(\mathbf{W})$ belongs. This is the same as the coset represented by $Q_{\Lambda^{(d)}}([\mathbf{W}] \bmod \Lambda_0^{(d)})$.

Each lattice point in $\Lambda^{(d)} \cap \mathcal{V}(\Lambda_0^{(d)})$ is a coset representative for a coset of $\Lambda_0^{(d)}$ in $\Lambda^{(d)}$. This is illustrated in Fig. 11. Suppose that Λ_j and Λ_k are the cosets which represent the messages X and Y respectively. Let $\mathbf{X} = [\mathbf{U}] \bmod \Lambda_0^{(d)}$ and $\mathbf{Y} = [\mathbf{V}] \bmod \Lambda_0^{(d)}$ be the coset representatives of Λ_j and Λ_k respectively. Then, $\Lambda_j \oplus \Lambda_k$ has $[\mathbf{X} + \mathbf{Y}] \bmod \Lambda_0^{(d)}$ as its representative. Therefore, the estimate $\mathcal{D}(\mathbf{W})$ has $\hat{\mathbf{W}} = [Q_{\Lambda^{(d)}}(\mathbf{W})] \bmod \Lambda_0^{(d)}$ as its coset representative. This is equal to $\hat{\mathbf{W}} = [Q_{\Lambda^{(d)}}([\mathbf{W}] \bmod \Lambda_0^{(d)})] \bmod \Lambda_0^{(d)}$. Let $\tilde{\mathbf{W}} = [\mathbf{W}] \bmod \Lambda_0^{(d)}$. Then, $\hat{\mathbf{W}} = [Q_{\Lambda^{(d)}}(\tilde{\mathbf{W}})] \bmod \Lambda_0^{(d)}$. As a consequence of the transmitter-receiver operations, the "effective" channel from $\mathbf{X}, \mathbf{Y}$ to $\tilde{\mathbf{W}}$ can be written as follows [23]:

$$\begin{aligned}
\tilde{\mathbf{W}} &= [\mathbf{U} + \mathbf{V} + \mathbf{Z}] \bmod \Lambda_0^{(d)} \\
&= [\Big([\mathbf{U} + \mathbf{V}] \bmod \Lambda_0^{(d)}\Big) + \mathbf{Z}] \bmod \Lambda_0^{(d)} \\
&= [\Big([\mathbf{X} + \mathbf{Y}] \bmod \Lambda_0^{(d)}\Big) + \mathbf{Z}] \bmod \Lambda_0^{(d)}
\end{aligned}$$

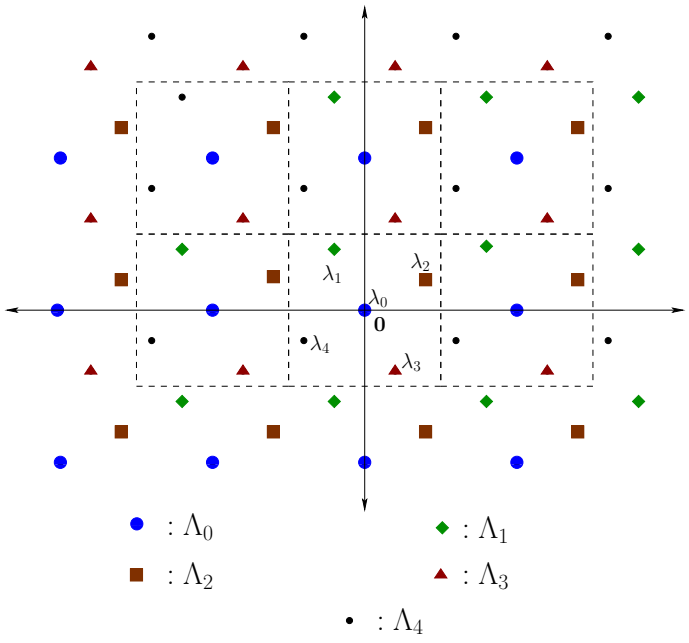

Fig. 11. Different cosets of Λ_0 in Λ in two dimensions. The coset representative of Λ_j within $\mathcal{V}(\Lambda_0)$ is λ_j.

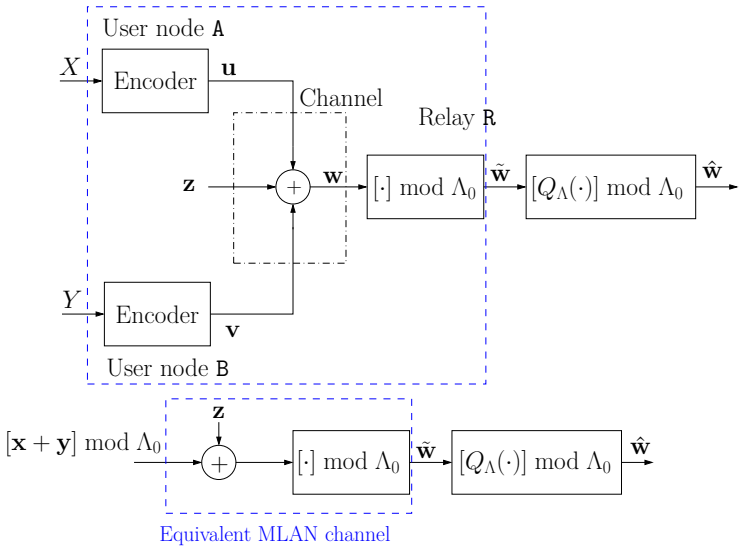

Fig. 12. MAC phase of the bidirectional relay and equivalent MLAN channel representation

A channel of the form $\mathbf{W} = [\mathbf{X}+\mathbf{N}] \bmod \Lambda_0^{(d)}$, where $\mathbf{N}$ denotes the noise vector, is called a $\Lambda_0^{(d)}$-modulo lattice additive noise ($\Lambda_0^{(d)}$-MLAN) channel [12]. The random variable $\tilde{\mathbf{W}}$ behaves like the output of a point-to-point transmission over a $\Lambda_0^{(d)}$-MLAN channel, with the transmitted vector being $[\mathbf{X}+\mathbf{Y}] \bmod \Lambda_0^{(d)}$. Looking from $\tilde{\mathbf{W}}$, the "effective" channel is a $\Lambda_0^{(d)}$-MLAN channel, and the relay has to decode $[\mathbf{X}+\mathbf{Y}]$ mod

$\Lambda_0^{(d)}$ reliably from $\tilde{\mathbf{W}}$. This is illustrated in Fig. 12. We will use the properties of the $\Lambda_0^{(d)}$-MLAN channel to determine achievable rate regions for our coding scheme.

We choose a sequence of nested lattice pairs that satisfy (G_1)–(G_3), with each nested lattice pair coming from a (d, k, q, k_1, q_1) ensemble, where k, q, k_1 and q_1 satisfy (32) and (33). Using the coding scheme of Section VI-A, we can achieve perfect secrecy. From (27), an average transmit power of $\left(4j^2_{\frac{d-2}{2}}/(d{r_{\text{pack}}}^2(\hat{\Lambda}_0^{(d)})\right)$ can be achieved. Since the sequence of nested lattice pairs satisfy the goodness properties (G_1)–(G_3), we have the following result which follows from [12], [23].

Proposition 15. *Let $\mathsf{M} > 0$ be a constant, and $\{\Lambda^{(d)}, \Lambda_0^{(d)}\}$ be a sequence of nested lattice pairs that satisfy (G_1)–(G_3). Then, any rate less than $\mathcal{R} := \frac{1}{2}\log_2\left(\frac{\mathsf{M}}{\sigma^2}\right)$ is achievable with perfect secrecy using the sequence of nested lattice pairs $\left\{\frac{\sqrt{d\mathsf{M}}}{r_{\mathit{eff}}(\Lambda_0^{(d)})}\Lambda^{(d)}, \frac{\sqrt{d\mathsf{M}}}{r_{\mathit{eff}}(\Lambda_0^{(d)})}\Lambda_0^{(d)}\right\}$.*

Proof: See Appendix D. ∎

1) Relating the Power and Rate: From (28), we know that as long as the average transmit power per dimension is less than $\left(d/{r_{\text{pack}}}^2(\hat{\Lambda}_0^{(d)})\right)(1 + o_d(1))$, we can guarantee perfect secrecy at the relay. From Proposition 15, we see that as long as the transmission rate is less than $\frac{1}{2}\log_2({r_{\text{eff}}}^2(\Lambda_0^{(d)})/(d\sigma^2))$, the relay can reliably compute $X \oplus Y$ from $\mathbf{W}$. In order to achieve positive rates, we need $r_{\text{eff}}(\Lambda_0^{(d)})$ to grow at least as fast as $\sqrt{d}$, i.e., $r_{\text{eff}}(\Lambda_0^{(d)}) = \Omega(\sqrt{d})$. Furthermore, to satisfy an average power constraint, we require $r_{\text{pack}}(\hat{\Lambda}_0^{(d)}) = \Omega(\sqrt{d})$. The rate is an increasing function of $r_{\text{eff}}(\Lambda_0^{(d)})$, and the average transmit power per dimension is a decreasing function of $r_{\text{pack}}(\hat{\Lambda}_0^{(d)})$. Since we want to maximize the rate for a given power constraint, we would like both $r_{\text{eff}}(\Lambda_0^{(d)})$ and $r_{\text{pack}}(\hat{\Lambda}_0^{(d)})$ to be as large as possible. However, for any lattice $\Lambda_0^{(d)}$, we have $r_{\text{cov}}(\Lambda_0^{(d)})r_{\text{pack}}(\hat{\Lambda}_0^{(d)}) \leq \pi d$ [3, Theorem 18.3], and since $r_{\text{eff}}(\Lambda_0^{(d)}) \leq r_{\text{pack}}(\Lambda_0^{(d)})$, we get $r_{\text{eff}}(\Lambda_0^{(d)})r_{\text{pack}}(\hat{\Lambda}_0^{(d)}) \leq \pi d$. Hence, to obtain positive rates and at the same time satisfy the power constraint, both $r_{\text{eff}}(\Lambda_0^{(d)})$ and $r_{\text{pack}}(\hat{\Lambda}_0^{(d)})$ must grow roughly as $\sqrt{d}$. Therefore, we seek lattices satisfying properties (G_1)–(G_3), for which the product $r_{\text{eff}}(\Lambda_0^{(d)})r_{\text{pack}}(\hat{\Lambda}_0^{(d)})$ is close to the upper bound of πd.

For a sequence of Construction-A coarse lattices satisfying (G_1) and (G_2), we can find an asymptotic lower bound for $(1/d)r_{\text{eff}}(\Lambda_0^{(d)})r_{\text{pack}}(\hat{\Lambda}_0^{(d)})$,[7] as the following theorem shows.

Lemma 16. *Let $\{\Lambda_0^{(d)}\}$ be a sequence of coarse lattices, with each $\Lambda_0^{(d)}$ chosen from a (d, k, q) ensemble and k, q satisfying (32) and (33). If $\{\Lambda_0^{(d)}\}$ satisfies conditions (G_1)–(G_2), then,*

$$\lim_{d\to\infty} \frac{r_{\mathit{eff}}(\Lambda_0^{(d)})r_{\mathit{pack}}(\hat{\Lambda}_0^{(d)})}{d} \geq \frac{1}{2e}. \tag{34}$$

Proof: See Appendix H. ∎

[7]The product $r_{\text{eff}}(\Lambda_0^{(d)})r_{\text{pack}}(\hat{\Lambda}_0^{(d)})$ is invariant to scaling of $\Lambda_0^{(d)}$. This is because, for a constant $\alpha > 0$, $r_{\text{eff}}(\alpha\Lambda_0^{(d)}) = \alpha r_{\text{eff}}(\Lambda_0^{(d)})$, and if $\Lambda' = \alpha\Lambda_0^{(d)}$, then the Fourier dual of Λ' is $(1/\alpha)\hat{\Lambda}_0^{(d)}$.

D. Proof of Theorem 1

Let us choose $r_{\text{eff}}(\Lambda_0^{(d)}) = \frac{1}{2e}\sqrt{d\mathcal{P}}$, for a constant $\mathcal{P} > 0$. Fix a $\delta > 0$. Using Lemma 16, we see that

$$r_{\text{pack}}(\hat{\Lambda}_0^{(d)}) \geq \frac{d}{2er_{\text{eff}}(\Lambda_0^{(d)})}(1 - o_d(1)) \geq \frac{\sqrt{d}}{\sqrt{\mathcal{P}}}(1 - o_d(1)). \tag{35}$$

From (28), we see that perfect secrecy can be achieved with an average power constraint as low as $P^{(d)} = \left(d/{r_{\text{pack}}}^2(\hat{\Lambda}_0^{(d)})\right)(1 + o_d(1))$. Combining this and (35), perfect secrecy can be achieved with an average transmission power,[8]

$$P^{(d)} < \mathcal{P} + \delta \tag{36}$$

for all sufficiently large d. From Proposition 15, we have seen that the average probability of error can be made to go down to zero as long as

$$R^{(d)} < \mathcal{R} := \frac{1}{2}\log_2 \frac{\mathcal{P}}{(2e)^2\sigma^2}. \tag{37}$$

Therefore, for every $\delta > 0$, we can choose a sequence of nested lattice codes such that for all sufficiently large d, we have $R^{(d)} > \mathcal{R} - \delta$, $P^{(d)} < \mathcal{P} + \delta$ and $\eta^{(d)} < \delta$. Hence, a power-rate pair of

$$\left(\mathcal{P}, \frac{1}{2}\log_2 \frac{\mathcal{P}}{\sigma^2} - \log_2 2e\right)$$

is achievable with perfect secrecy, concluding the proof of Theorem 1. □

VII. Strong Secrecy

Let us quickly summarize the coding scheme used to obtain perfect secrecy. We chose a pair of nested lattices $(\Lambda^{(d)}, \Lambda_0^{(d)})$, where $\Lambda_0^{(d)} \subseteq \Lambda^{(d)}$, and the messages were chosen uniformly over the quotient group, $\Lambda^{(d)}/\Lambda_0^{(d)}$. Given a message (coset) Λ_j, the user node transmits a point from that coset according to a distribution p_j, which is obtained by sampling (and appropriately normalizing) a distribution f over $\mathbb{R}^d$, at precisely those lattice points that belong to the coset Λ_j. As long as the characteristic function corresponding to f is supported within the fundamental Voronoi region of the Fourier dual of $\Lambda_0^{(d)}$, perfect secrecy can be obtained. We imposed an additional constraint, that the characteristic function corresponding to f must be supported within a ball of radius $r_{\text{pack}}(\hat{\Lambda}_0^{(d)})$. Using this, we obtained the minimum average transmit power, and found an achievable rate.

A natural question that arises is what happens if we choose a density f for which the support of the characteristic function goes beyond $\mathcal{V}(\hat{\Lambda}_0^{(d)})$. Suppose that we choose a characteristic function, $\psi(\mathbf{t})$, which is supported within a ball of radius $\rho > r_{\text{pack}}(\hat{\Lambda}_0^{(d)})$, and choose the characteristic function $\phi_{U|X=\mathbf{x}}(\mathbf{t}) = \sum_{\mathbf{n}\in\hat{\Lambda}_0^{(d)}} \psi(\mathbf{t} + \mathbf{n})e^{-i\langle \mathbf{n},\mathbf{x}\rangle}$. Clearly, we cannot expect perfect secrecy, but can we at least obtain strong secrecy? Let us choose ψ to be the characteristic function of the minimum-variance distribution in (26),

[8]Recall that a sequence $f(d)$ is said to be equal to $o_d(1)$ if $f(d) \to 0$ as $d \to \infty$; and hence $(1+o_d(1))/(1-o_d(1))^2 = 1+o_d(1)$.

with the support of ψ chosen to be a ball of radius ρ. Let us choose $\rho = \min\{2r_{\text{eff}}(\hat{\Lambda}_0^{(d)}), 2r_{\text{pack}}(\hat{\Lambda}_0^{(d)})\}$ [9]. Doing so would give us an improved rate of $\frac{1}{2}\log_2 \frac{\mathcal{P}}{\sigma^2} - \log_2 e$. However, for such a coding scheme, we are only able to show that the ℓ^2 norm of the difference between $p_{U+V,X}$ and $p_{U+V}p_X$ goes to zero as $d \to \infty$. Knowing only that the ℓ^2 norm of the difference between $p_{U+V,X}$ and $p_{U+V}p_X$ goes to zero as $d \to \infty$, we cannot conclude whether strong secrecy is obtained. In fact, by itself, the ℓ^2 norm is not a good measure of secrecy. In any case, we will use a different approach to obtaining strong secrecy, and show that an even higher transmission rate of $\frac{1}{2}\log_2 \frac{\mathcal{P}}{\sigma^2} - \frac{1}{2}\log_2 2e$ is achievable.

Instead of using distributions with finitely supported characteristic functions, we will use a sampled Gaussian density for signaling. Such a scheme was used in context of the wiretap channel in [20] with encouraging results, and we will do the same here.

A. The Sampled Gaussian Function

We now introduce some notation that will be used in the sequel. Let Λ be a lattice in $\mathbb{R}^d$. For any $\mathbf{x}, \mathbf{z} \in \mathbb{R}^d$, and any $\kappa > 0$, we define the sampled Gaussian function to be

$$g_{\kappa,\mathbf{x}}(\mathbf{z}) := \frac{1}{(2\pi\kappa^2)^{d/2}} e^{-\frac{\|\mathbf{z}-\mathbf{x}\|^2}{2\kappa^2}}. \tag{38}$$

We also define

$$g_{\kappa,\mathbf{x}}(\Lambda) := \sum_{\lambda\in\Lambda} g_{\kappa,\mathbf{x}}(\lambda). \tag{39}$$

We will use $g_\kappa(\mathbf{z})$ and $g_\kappa(\Lambda)$ to denote $g_{\kappa,\mathbf{0}}(\mathbf{z})$ and $g_{\kappa,\mathbf{0}}(\Lambda)$ respectively.

B. Coding Scheme for Strong Secrecy

Code: Following Section VI-A, we use a $(\Lambda^{(d)}, \Lambda_0^{(d)})$ nested lattice code, with $\Lambda_0^{(d)} \subseteq \Lambda^{(d)}$. As before, the messages are chosen from $\mathbb{G}^{(d)} := \Lambda^{(d)}/\Lambda_0^{(d)}$, and $\oplus$ is the addition operation on $\mathbb{G}^{(d)}$. The $M^{(d)} := |\mathbb{G}^{(d)}|$ cosets of $\Lambda_0^{(d)}$ in $\Lambda^{(d)}$ are denoted by $\Lambda_0, \ldots, \Lambda_{M^{(d)}-1}$.

Encoding: For a coset Λ_j of $\Lambda_0^{(d)}$ in $\Lambda^{(d)}$, let λ_j denote its representative within $\mathcal{V}(\Lambda_0^{(d)})$ (see Fig. 11 for an illustration). Fix a $\kappa > 0$. Corresponding to the message Λ_j, the user node transmits a random lattice point from Λ_j, according to the distribution

$$p_j(\mathbf{u}) = \begin{cases} \frac{g_\kappa(\mathbf{u})}{g_{\kappa,\lambda_j}(\Lambda_0^{(d)})}, & \text{if } \mathbf{u} \in \Lambda_j, \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{40}$$

Decoding: As in the perfect secrecy scenario, the relay uses a lattice decoder. If $\mathbf{W}$ is the vector received in the MAC phase, then the relay estimates $X \oplus Y$ to be the coset represented by the vector in $\Lambda^{(d)}$ closest to $\mathbf{W}$.

[9] If we have $\rho > 2r_{\text{pack}}(\hat{\Lambda}_0^{(d)})$, then $\sum_{\mathbf{n}\in\hat{\Lambda}_0^{(d)}} \psi(\mathbf{t}+\mathbf{n})e^{-i\langle\mathbf{n},\mathbf{x}\rangle}$ would have to be normalized to make this a characteristic function, and this makes analysis more complicated.

Achievable power-rate pair: A power-rate pair of $(\mathcal{P}, \mathcal{R})$ is achievable if for every $\delta > 0$, there exists a sequence of $(\Lambda^{(d)}, \Lambda_0^{(d)})$ nested lattice codes so that for all sufficiently large d,

- the average transmit power per dimension is less than $\mathcal{P} + \delta$:
$$P^{(d)} := \frac{1}{d}\mathbb{E}\|\mathbf{U}\|^2 = \frac{1}{d}\mathbb{E}\|\mathbf{V}\|^2 < \mathcal{P} + \delta;$$
- the transmission rate is greater than $\mathcal{R} - \delta$:
$$R^{(d)} := \frac{1}{d}\log_2 M^{(d)} > \mathcal{R} - \delta;$$
- the average probability of decoding $X \oplus Y$ incorrectly from $\mathbf{W}$ is less than δ; and
- the mutual information between each message and $\mathbf{U} + \mathbf{V}$ is less than δ:
$$\mathcal{I}(X; \mathbf{U} + \mathbf{V}) = \mathcal{I}(Y; \mathbf{U} + \mathbf{V}) < \delta.$$

We will show the following result.

Theorem 17. *For any $\mathcal{P} \geq 4e\sigma^2$, a power-rate pair of*
$$\left(\mathcal{P}, \frac{1}{2}\log_2 \frac{\mathcal{P}}{\sigma^2} - \frac{1}{2}\log_2 2e\right)$$
can be achieved with strong secrecy using the coding scheme of Section VII-B.

C. Strong Secrecy in the Absence of Noise

We will first prove that the scheme described in the previous section achieves strong secrecy. Let us establish some more notation. Let $p_{U+V}(\cdot)$ denote the distribution of $U+V$, and for any $\mathbf{x} \in \Lambda^{(d)} \cap \mathcal{V}(\Lambda_0^{(d)})$, let $p_{U+V|\mathbf{x}}(\cdot)$ denote the distribution of $U+V$ conditioned on the event that X is the coset to which $\mathbf{x}$ belongs. We will show that for every $\mathbf{x}$ in $\Lambda^{(d)} \cap \mathcal{V}(\Lambda_0^{(d)})$ the *variational distance* between p_{U+V} and $p_{U+V|\mathbf{x}}(\cdot)$, defined as
$$\mathbb{V}(p_{U+V}, p_{U+V|\mathbf{x}}) := \sum_{\mathbf{w} \in \Lambda^{(d)}} |p_{U+V}(\mathbf{w}) - p_{U+V|\mathbf{x}}(\mathbf{w})|, \tag{41}$$
goes to zero exponentially in the dimension d. Therefore, the *average variational distance* between the joint pmf of $U+V$ and X, and the product of the marginals,
$$\mathtt{d}_V := \sum_{\mathbf{x} \in \Lambda^{(d)} \cap \mathcal{V}(\Lambda_0^{(d)})} \frac{1}{|\mathbb{G}^{(d)}|}\mathbb{V}(p_{U+V}, p_{U+V|\mathbf{x}}),$$
also goes to zero exponentially in d. We can then use the following lemma, which relates the mutual information and the variational distance.

Lemma 18 ([9], Lemma 1)**.** *For $|\mathbb{G}^{(d)}| \geq 4$, we have*
$$\mathcal{I}(X; \mathbf{U} + \mathbf{V}) \leq \mathtt{d}_V \left(\log_2 |\mathbb{G}^{(d)}| - \log_2 \mathtt{d}_V\right). \tag{42}$$

Since $|\mathbb{G}^{(d)}|$ grows exponentially in d, it is sufficient to have $\mathtt{d}_V$ going to zero as $o(1/d)$ for $\mathcal{I}(X;\mathbf{U}+\mathbf{V})$ to go to zero. The exponential decay of the average variational distance will guarantee that the mutual information also decays exponentially in d. In order to have $\mathtt{d}_V$ going to zero exponentially in d, we will require the coarse and fine lattices to satisfy certain properties. For any lattice Λ in $\mathbb{R}^d$, and any $\theta > 0$, the flatness factor $\epsilon_\Lambda(\theta)$ is defined as [20], [5]

$$\epsilon_\Lambda(\theta) := \frac{\max_{\mathbf{x}\in\mathcal{V}(\Lambda)} \left| \left(\sum_{\lambda\in\Lambda} g_{\theta,\lambda}(\mathbf{x})\right) - (1/\det\Lambda)\right|}{(1/\det\Lambda)}. \tag{43}$$

Following [20], we define a sequence of lattices $\{\Lambda^{(d)}\}$ to be *secrecy-good* if

$$\epsilon_{\Lambda^{(d)}}(\theta) \leq 2^{-\Omega(d)} \text{ for all } \theta \text{ such that } \frac{(\det(\Lambda^{(d)}))^{2/d}}{2\pi\theta^2} < 1.$$

It was shown in [20] that there exist lattices that are secrecy-good and also satisfy all the goodness properties described in Appendix B.

Let us choose κ in (40) to be equal to $\sqrt{\mathcal{P}}$. We can bound the variational distance in terms of the flatness factor of the coarse lattice as follows:

Theorem 19. *If the sequence of nested lattice pairs $\{\Lambda^{(d)}, \Lambda_0^{(d)}\}$ satisfies $\epsilon^{(d)} := \epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}/2}) < 1/2$, then for every $\mathbf{x} \in \Lambda^{(d)} \cap \mathcal{V}(\Lambda_0^{(d)})$, we have*

$$\mathbb{V}(p_{U+V}, p_{U+V|\mathbf{x}}) \leq 216\epsilon^{(d)}. \tag{44}$$

A proof of the above theorem is given in Appendix F.

Furthermore, if the flatness factors of both the coarse and fine lattices go to zero as d goes to infinity, then we can bound the average transmit power per dimension:

Lemma 20. *As $d \to \infty$, if the flatness factors $\epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}/2})$, and $\epsilon_{\Lambda^{(d)}}(\sqrt{\mathcal{P}}/2)$ both converge to zero, then the average transmit power per dimension, $\frac{1}{d}\mathbb{E}\|\mathbf{U}\|^2 = \frac{1}{d}\mathbb{E}\|\mathbf{V}\|^2$, converges to $\mathcal{P}$.*

Proof: See Appendix G. ∎

From Theorem 19 and Lemma 18, we see that strong secrecy can be obtained in the noiseless scenario. Since the noise is independent of everything else, we have strong secrecy in a noisy channel as well. To see why this is the case, observe that $X \to (\mathbf{U}+\mathbf{V}) \to (\mathbf{U}+\mathbf{V}+\mathbf{Z})$ forms a Markov chain. Using the data-processing inequality, we see that $\mathcal{I}(X;\mathbf{U}+\mathbf{V}+\mathbf{Z}) \leq \mathcal{I}(X;\mathbf{U}+\mathbf{V})$, verifying our claim.

D. Achievable Rate and Proof of Theorem 17

We choose our sequence of nested lattices $\{\Lambda^{(d)}, \Lambda_0^{(d)}\}$ so as to satisfy the following properties:

(L1) The sequence of coarse lattices, $\{\Lambda_0^{(d)}\}$, is good for covering, MSE quantization, and AWGN channel coding[10].

[10]For the definitions of lattices good for covering, MSE quantization, and AWGN channel coding, see Appendix B.

(L2) The sequence of coarse lattices, $\{\Lambda_0^{(d)}\}$, is secrecy-good.

(L3) The sequence of fine lattices, $\{\Lambda^{(d)}\}$, is good for AWGN channel coding.

(L4) The sequence of fine lattices, $\{\Lambda^{(d)}\}$, is secrecy-good.

It was shown in [20] that such a sequence of nested lattices indeed exists[11]. Using (L2), to have the flatness factor $\epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}/2}) \to 0$, the coarse lattices must be scaled so that

$$\frac{\left(\det(\Lambda_0^{(d)})\right)^{2/d}}{2\pi(\mathcal{P}/2)} < 1. \tag{45}$$

Let us choose $\left(\det(\Lambda_0^{(d)})\right)^{2/d} = \pi\mathcal{P} - \delta$, for some arbitrary $\delta > 0$, so as to satisfy (45).

In order to have $\epsilon_{\Lambda^{(d)}}(\sqrt{\mathcal{P}}/2) \to 0$ as $d \to \infty$, we require

$$\frac{\left(\det(\Lambda^{(d)})\right)^{2/d}}{2\pi(\mathcal{P}/4)} < 1. \tag{46}$$

Since $|\mathbb{G}^{(d)}|\det(\Lambda^{(d)}) = \det(\Lambda_0^{(d)})$ [3, Theorem 5.2], we can rewrite (46) as follows:

$$\frac{1}{|\mathbb{G}^{(d)}|^{2/d}} \frac{\left(\det(\Lambda_0^{(d)})\right)^{2/d}}{\pi(\mathcal{P}/2)} < 1.$$

But then, we have $\left(\det(\Lambda_0^{(d)})\right)^{2/d} = \pi\mathcal{P} - \delta$, and hence, the above requirement reduces to

$$|\mathbb{G}^{(d)}|^{2/d} > \frac{2}{\left(1 - \frac{\delta}{\pi\mathcal{P}}\right)}.$$

In other words,

$$\frac{1}{d}\log_2|\mathbb{G}^{(d)}| > \frac{1}{2} - \frac{1}{2}\log_2\left(1 - \frac{\delta}{\pi\mathcal{P}}\right). \tag{47}$$

From Lemma 20, we have the average transmit power converging to $\mathcal{P}$ as long as $\epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}/2})$ and $\epsilon_{\Lambda^{(d)}}(\sqrt{\mathcal{P}}/2)$ tend to zero as $d \to \infty$. Our choice of $\left(\det(\Lambda_0^{(d)})\right)^{2/d} = \pi\mathcal{P} - \delta$ ensured that (45) is satisfied and $\epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}/2}) \to 0$. As long as (47) is satisfied, we also have $\epsilon_{\Lambda^{(d)}}(\sqrt{\mathcal{P}}/2) \to 0$, and the average power converges to $\mathcal{P}$. Therefore, we see that requiring the average transmit power to converge to $\mathcal{P}$ places a constraint on the transmission rate.

Since the sequence of fine lattices is good for AWGN channel coding, the probability of error in decoding $X \oplus Y$ from $\mathbf{U} + \mathbf{V} + \mathbf{Z}$ can be made to decay to zero as $d \to \infty$ as long as $\frac{\left(\det(\Lambda^{(d)})\right)^{2/d}}{2\pi e\sigma^2} > 1$, or

$$\frac{1}{|\mathbb{G}^{(d)}|^{(2/d)}} \frac{\left(\det(\Lambda_0^{(d)})\right)^{2/d}}{2\pi e\sigma^2} > 1. \tag{48}$$

Since we have chosen $\det(\Lambda_0^{(d)}) = \pi\mathcal{P} - \delta$, (48) reduces to

$$\frac{1}{d}\log_2|\mathbb{G}^{(d)}| < \frac{1}{2}\log_2\left(\frac{\mathcal{P} - \frac{\delta}{\pi}}{2e\sigma^2}\right).$$

[11] Although it was not explicitly mentioned in [20] that the coarse lattices satisfy property (L2), the construction implicitly guarantees that the requirement is satisfied.

Since δ is arbitrary, we can conclude that as long as

$$\mathcal{R} < \frac{1}{2}\log_2\left(\frac{\mathcal{P}}{\sigma^2}\right) - \frac{1}{2}\log_2 2e, \tag{49}$$

the probability of error can be made to go down to zero as d goes to infinity.

If we choose $\mathcal{P} \geq 4e\sigma^2$, then the right hand side of (49) exceeds that of (47). Therefore, for any $\mathcal{P} \geq 4e\sigma^2$ we see that a power-rate pair of

$$\left(\mathcal{P}, \frac{1}{2}\log_2\left(\frac{\mathcal{P}}{\sigma^2}\right) - \frac{1}{2}\log_2 2e\right)$$

can be achieved with strong secrecy using nested lattice codes, concluding the proof of Theorem 17. □

E. Improving the Rate

The rate can be further improved using random dithering and MMSE equalization at the relay [12], [23]. The user nodes generate dither vectors $\mathbf{D}_1$ (at node A) and $\mathbf{D}_2$ (at node B) independently and uniformly at random from the fundamental Voronoi region $\mathcal{V}(\Lambda_0^{(d)})$. For $\mathbf{D}_1 = \mathbf{d}_1$, and $\mathbf{D_2} = \mathbf{d}_2$, let us define $\tilde{\mathbf{d}}_1 := [\mathbf{x}+\mathbf{d}_1] \bmod \Lambda_0^{(d)}$, and $\tilde{\mathbf{d}}_2 := [\mathbf{y}+\mathbf{d}_2] \bmod \Lambda_0^{(d)}$, where $\mathbf{x}, \mathbf{y} \in \Lambda^{(d)} \cap \mathcal{V}(\Lambda_0^{(d)})$. For a message $X = \Lambda_j \in \Lambda^{(d)}/\Lambda_0^{(d)}$ at node A, having coset representative $\mathbf{x}$ in $\Lambda^{(d)} \cap \mathcal{V}(\Lambda_0^{(d)})$, user node A transmits $\mathbf{u} \in \left(\Lambda_0^{(d)} + \tilde{\mathbf{d}}_1\right) := \{\lambda + \tilde{\mathbf{d}}_1 : \lambda \in \Lambda_0^{(d)}\}$ according to

$$p_j(\mathbf{u}|\mathbf{d}_1) = \begin{cases} \frac{g_{\sqrt{\mathcal{P}}}(\mathbf{u})}{g_{\sqrt{\mathcal{P}},\tilde{\mathbf{d}}_1}(\Lambda_0^{(d)})}, & \text{if } \mathbf{u} \in \Lambda_0^{(d)} + \tilde{\mathbf{d}}_1, \\ \mathbf{0}, & \text{otherwise.} \end{cases}, \tag{50}$$

Node B transmits $\mathbf{v} \in \left(\Lambda_0^{(d)} + \tilde{\mathbf{d}}_2\right)$ in a similar fashion.

A careful study of the proof of Theorem 19 shows that the theorem still holds even if a translate (by a vector $\mathbf{c} \in \mathbb{R}^d$) of the coarse and fine lattices are used for signaling, i.e., $(\Lambda_0^{(d)} + \mathbf{c}, \Lambda^{(d)} + \mathbf{c})$ are used instead of $(\Lambda^{(d)}, \Lambda_0^{(d)})$, and hence adding a dither at the source does not affect security.

Note that $\sum_{\lambda \in \Lambda_0^{(d)}} g_{\theta,\lambda}(\mathbf{x}) = \sum_{\lambda \in \Lambda_0^{(d)}} g_{\theta,\mathbf{x}}(\lambda)$, which is equal to $g_{\theta,\mathbf{x}}(\Lambda_0^{(d)})$. From this, and using the definition of flatness factor, we see that $|\det(\Lambda_0^{(d)}) g_{\theta,\mathbf{x}}(\Lambda_0^{(d)}) - 1| \leq \epsilon_{\Lambda_0^{(d)}}(\theta)$. Rearranging, we get $1 - \epsilon_{\Lambda_0^{(d)}}(\theta) \leq \det\Lambda_0^{(d)} g_{\theta,\mathbf{x}}(\Lambda_0^{(d)}) \leq 1 + \epsilon_{\Lambda_0^{(d)}}(\theta)$. Substituting this in (50), we get

$$\frac{\det\Lambda_0^{(d)} g_{\sqrt{\mathcal{P}}}(\mathbf{u})}{1 + \epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}})} \leq p_j(\mathbf{u}|\mathbf{d}_1) \leq \frac{\det\Lambda_0^{(d)} g_{\sqrt{\mathcal{P}}}(\mathbf{u})}{1 - \epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}})}$$

for $\mathbf{u} \in \Lambda_0^{(d)} + \tilde{\mathbf{d}}_1$, and $p_j(\mathbf{u}|\mathbf{d}_1) = 0$ otherwise. Let f_U denote the density of $\mathbf{U}$ and $\mathbf{V}$. Then, $f_U(\mathbf{u}) = \int_{\mathbf{d}_1 \in \mathcal{V}(\Lambda_0^{(d)})} p_j(\mathbf{u}|\mathbf{d}_1) \frac{1}{\det\Lambda_0^{(d)}}$. Hence,

$$\frac{g_{\sqrt{\mathcal{P}}}(\mathbf{u})}{1 + \epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}})} \leq f_U(\mathbf{u}) \leq \frac{g_{\sqrt{\mathcal{P}}}(\mathbf{u})}{1 - \epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}})},$$

Therefore, $\mathbf{U}$ and $\mathbf{V}$ converge in distribution to a Gaussian vector with zero-mean and variance $\sqrt{\mathcal{P}}$ as long as $\epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}}) \to 0$. From [20, Remark 3], we have $\epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}}) < \epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}/2})$. Therefore, as long as

$\epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}/2}) \to 0$ (which we need anyway, to obtain strong secrecy), $\mathbf{U}$ and $\mathbf{V}$ converge in distribution to the Gaussian random vector.

The relay uses a linear MMSE estimator. It computes $\hat{\mathbf{W}} := [\alpha^*\mathbf{W} - \mathbf{D}_1 - \mathbf{D}_2] \bmod \Lambda_0^{(d)}$, where $\alpha^* := (2\mathcal{P})/(2\mathcal{P}+\sigma^2)$. The estimate of $X \oplus Y$ is the coset represented by the closest vector in $\Lambda_0^{(d)}$ to $\hat{\mathbf{W}}$. The term $\hat{\mathbf{W}}$ can be written as [23]

$$\hat{\mathbf{W}} = \big[[\mathbf{X}+\mathbf{Y}] \bmod \Lambda_0^{(d)} + (1-\alpha^*)(\mathbf{U}+\mathbf{V}) + \alpha^*\mathbf{Z}\big] \bmod \Lambda_0^{(d)}.$$

The density of the effective noise vector $\mathbf{Z}_{\text{eff}} = (1-\alpha^*)(\mathbf{U}+\mathbf{V})+\alpha^*\mathbf{Z}$ can be bounded in terms of a zero-mean Gaussian random vector with covariance matrix $((1-\alpha^*)2\sqrt{\mathcal{P}}+\alpha^*\sigma^2)\mathsf{I}_d$, and the true pdf of $\mathbf{Z}_{\text{eff}}$ converges to the pdf of the Gaussian as $d \to \infty$. The analysis in [12], [23] can be used in this setting, and it can be shown that a power-rate pair of

$$\left(\mathcal{P}, \frac{1}{2}\log_2\left(\frac{1}{2}+\frac{\mathcal{P}}{\sigma^2}\right) - \frac{1}{2}\log_2 2e\right)$$

is achievable with strong secrecy over the bidirectional relay.

Remark 21. *In the perfect secrecy setting, we were not able to show that the technique of random dithering and MMSE equalization can be used to obtain an additional* $1/2$ *in the rate expression. As above, suppose that each user node generates dither vectors, and* $\mathbf{d}_1, \mathbf{d}_2$ *be instances of the randomly generated dither vectors. Let* $\tilde{\mathbf{d}}_1, \tilde{\mathbf{d}}_2$ *be as defined above. Let* $f(\cdot)$ *be a density function on* $\mathbb{R}^d$ *whose characteristic function is compactly supported within* $\mathcal{V}(\hat{\Lambda}_0^{(d)})$*. For a message* $X = \Lambda_j$*, user node* A *transmits* $\mathbf{U} \in \Lambda_0^{(d)} + \tilde{\mathbf{d}}_1$ *according to*

$$p_j(\mathbf{u}|\mathbf{d}_1) = \begin{cases} \det\Lambda_0^{(d)} f(\mathbf{u}), & \text{if } \mathbf{u} \in \Lambda_0^{(d)} + \tilde{\mathbf{d}}_1, \\ \mathbf{0}, & \text{otherwise.} \end{cases}, \tag{51}$$

Similarly, user node B *transmits* $\mathbf{V} \in \Lambda_0^{(d)} + \tilde{\mathbf{d}}_2$ *according to (51). As in the strong-secrecy case, suppose that the relay computes* $\hat{\mathbf{W}} := [\alpha^*\mathbf{W} - \mathbf{D}_1 - \mathbf{D}_2] \bmod \Lambda_0^{(d)}$*, where* $\alpha^* := (2\mathcal{P})/(2\mathcal{P}+\sigma^2)$*. It can be shown that the dithering does not affect security, and* $\mathbf{W}$ *is still independent of the individual messages* X *and* Y*. However, the effective noise vector,* $\mathbf{Z}_{\mathit{eff}} = (1-\alpha^*)(\mathbf{U}+\mathbf{V})+\alpha^*\mathbf{Z}$ *is not Gaussian, since* $\mathbf{U}$ *and* $\mathbf{V}$ *are not Gaussian. In order to find the probability of decoding error, we require an upper bound on the probability that* $\mathbf{Z}_{\mathit{eff}} \notin \mathcal{V}(\Lambda^{(d)})$*, which is not straightforward unlike the Gaussian case. Consequently, one cannot say whether lattice decoding achieve vanishingly small error probabilities in this situation, and it is hard to compute achievable rates.*

Prior work on Strong Secrecy

The strongly secure scheme proposed by He and Yener in [17] also used nested lattice codes as we have done here. They obtain strong secrecy using universal hash functions, and show the existence of a suitable linear hash function that ensures that the mutual information decays exponentially in d. Unlike [17], we have used a sampled Gaussian pmf for randomization at the encoder, and hence, for a given pair of nested lattices, we explicitly specify the distribution used for randomization. Even using our scheme, the mutual

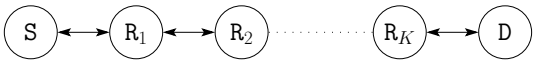

Fig. 13. Multi-hop line network with $K+1$ hops.

information goes down to zero exponentially in d. But unlike [17], which was valid under a maximum power constraint at each node, the codebook we use is unbounded, so our scheme can only satisfy an average power constraint. Also, the achievable rate in the scheme of He and Yener is slightly higher (by $\frac{1}{2}\log_2\frac{e}{2}$ bits per channel use). Our scheme has these two drawbacks but is still attractive because unlike [17], which is only an existence result, we give an explicit randomization technique for security. The scheme in [17] was coupled with an Algebraic Manipulation Detection (AMD) code [8] for Byzantine detection, and it was shown that the probability of a Byzantine attack being undetected could be made to decay to zero exponentially in d. We remark that our coding scheme can also be extended to this scenario, and be used as a replacement for the nested lattice code in [17].

VIII. Multi-hop Line Network

We extend the results of the bidirectional relay setting to the multi-hop line network studied in [16]. In this scenario, a source wants to transmit a message (or more than one message) to a destination via K relay nodes. The structure of a multi-hop line network with $K+1$ hops is shown in Fig. 13. It consists of $K+2$ nodes: a source node, $\mathtt{S}$, a destination node, $\mathtt{D}$, and K relay nodes, $\mathtt{R}_1, \mathtt{R}_2, \ldots, \mathtt{R}_K$. It is assumed that all links are identical AWGN (zero mean, variance σ^2) wireless links. All nodes are half-duplex and can communicate only with their neighbours. Nodes broadcast their messages to their immediate neighbours.

We use the co-operative jamming protocol proposed by He and Yener [16]. The protocol consists of a number of phases, each phase consisting of d channel uses. In any given phase, let $\mathbf{U}_i$ be the d-dimensional vector transmitted by the ith node, and $\mathbf{W}_j$ be the vector received by the jth node, for $i = 0, 1, \ldots, K+1$ and $j = 0, 1, \ldots, K+1$. Here, the source is considered to be the 0th node, and the destination, the $(K+1)$st node. The nodes receive the superposition of the messages sent by their neighbours in that phase. Therefore, for $i = 1, 2, \ldots, K$,

$$\mathbf{W}_i = \mathbf{U}_{i-1} + \mathbf{U}_{i+1} + \mathbf{Z}_i, \tag{52}$$

where $\mathbf{Z}_i$ denotes additive white Gaussian noise (AWGN) with mean zero and variance σ^2. Also,

$$\mathbf{W}_0 = \mathbf{U}_1 + \mathbf{Z}_0, \tag{53}$$

and

$$\mathbf{W}_{K+1} = \mathbf{U}_K + \mathbf{Z}_{K+1}. \tag{54}$$

As earlier, $\mathbf{Z}_0$ and $\mathbf{Z}_{K+1}$ denote AWGN with mean zero and variance σ^2.

The source, S, has to send N messages, $X_1, X_2, \ldots, X_N$, to the destination D across the network. The messages are assumed to be independent and uniformly distributed over the set of all messages.

The relays act as eavesdroppers. It is assumed that the relays do not co-operate with each other, i.e., the information available at a relay is not shared with the other relays. Also, the relays do not tamper with the message in any manner. As remarked by He and Yener in [16], this also takes care of the situation wherein the eavesdropper has access to one of the relays, but it is not known which relay has been compromised. Therefore, the problem is to send the message to the destination via a line network of K *independent, honest but curious* relays. We study this problem mainly under the strong secrecy constraint, but the arguments can easily be extended to the perfect secrecy scenario.

1) The Communication Scheme: We use the scheme proposed by He and Yener [16] for relaying. For clarity and completeness, we will describe the protocol here. Let us choose a sequence of $(\Lambda^{(d)}, \Lambda_0^{(d)})$ nested lattice pairs that satisfy properties (L1)–(L4). The messages at the source are picked from the quotient group $\mathbb{G}^{(d)} := \Lambda^{(d)}/\Lambda_0^{(d)}$. Each node in the network employs the encoding and decoding scheme described in Section VII-B. Let us define some more notation. For any vector $\mathbf{w} \in \mathbb{R}^d$, let $\mathcal{D}(\mathbf{w})$ denote the coset represented by the closest vector to $\mathbf{w}$ in the lattice $\Lambda^{(d)}$, i.e., $\mathcal{D}(\mathbf{w})$ is obtained by passing $\mathbf{w}$ through the lattice decoder of Section VII-B. Also, for any $X \in \mathbb{G}^{(d)}$, having coset representative $\mathbf{X}$ in $\Lambda^{(d)} \cap \mathcal{V}(\Lambda_0^{(d)})$, let $\mathcal{E}(X)$ denote the encoded form of X, i.e., $\mathcal{E}(X)$ is a random variable supported within the coset X, and distributed according to $p_j(\cdot)$ in (40).

- Each relay node i ($i = 1, \ldots, K$) generates a *jamming signal,* J_i, which is chosen uniformly at random from $\mathbb{G}^{(d)}$, and independently of everything else. If the source has to relay N messages to the destination, then the destination generates N independent jamming signals, J_{K+l}, for $l = 1, 2, \ldots, N$.
- The communication takes place in $2N + K$ phases. Each phase consists of d channel uses.
- Let $\mathbf{W}_i[n]$ denote the d-dimensional vector received by the ith node in the nth phase, and let $\mathbf{V}_i[n]$ be the vector transmitted by the ith node in the nth phase.

An average power constraint is imposed at the nodes.

$$\frac{1}{d}\mathbb{E}\|\mathbf{V}_i[n]\|^2 \leq P^{(d)}, \tag{55}$$

for $i = 0, 1, \ldots, K+1$ and $n = 1, 2, \ldots, K+2N$.

We define the rate of the scheme, $R_N^{(d)}$, to be the number of bits of information transmitted per channel use by the source in order to send N messages to the destination. Since it takes $K + 2N$ phases for sending N messages,

$$R_N^{(d)} := \frac{N}{d(K+2N)} \log_2 |\mathbb{G}^{(d)}|. \tag{56}$$

We say that a *power-rate pair* of $(\mathcal{P}, \mathcal{R})$ is *achievable for N-message transmission* with strong secrecy in a multi-hop line network with $K+1$ hops, if for every $\delta > 0$, there exists a sequence of $(\Lambda^{(d)}, \Lambda_0^{(d)})$ nested lattice codes such that for all sufficiently large d, we have

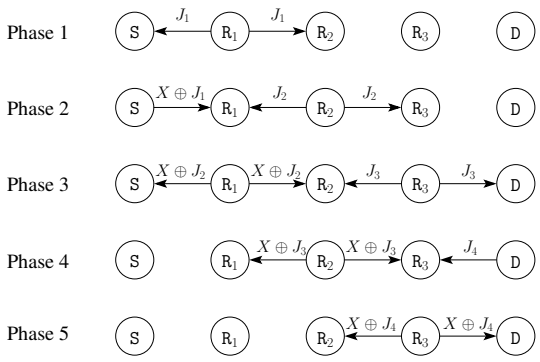

Fig. 14. Secure relaying of a single message in a 4-hop relay network

Phase	Messages available at node at the end of phase				
	S	R_1	R_2	R_3	D
0	X	J_1	J_2	J_3	J_4
1	X, J_1	J_1	J_1, J_2	J_3	J_4
2	X, J_1	J_1, $X \oplus J_2$	J_1, J_2	J_2, J_3	J_4
3	X, J_1	J_1, $X \oplus J_2$	J_1, J_2, $X \oplus J_3$	J_2, J_3	J_3, J_4
4	X, J_1	J_1, $X \oplus J_2$, $X \oplus J_3$	J_1, J_2, $X \oplus J_3$	J_2, J_3, $X \oplus J_4$	J_3, J_4
5	X, J_1	J_1, $X \oplus J_2$, $X \oplus J_3$	J_1, J_2, $X \oplus J_3$	J_2, J_3, $X \oplus J_4$	J_3, J_4, X

TABLE I

MESSAGES AVAILABLE AT VARIOUS NODES AT THE END OF EACH PHASE FOR THE PROTOCOL IN FIG. 14

- $P^{(d)} < \mathcal{P} + \delta$;
- $R_N^{(d)} > \mathcal{R} - \delta$;
- the probability of the destination decoding $X_1, X_2, \ldots, X_N$ incorrectly, $\eta^{(d)}$, is less than δ; and,
- the mutual information between the messages and the variables available at the kth relay, i.e.,

$$\mathcal{I}(X_1, \ldots, X_N; J_k, \mathbf{W}_k[1], \ldots, \mathbf{W}_k[2N+K]) < \delta.$$

We will describe the scheme for secure message relaying in the next section, and find achievable power-rate pairs. Finally, letting the number of messages to go to infinity, we will show that

Theorem 22. *A power-rate pair of*

$$\left(\mathcal{P}, \frac{1}{4}\log_2\left(\frac{\mathcal{P}}{\sigma^2}\right) - \frac{1}{2}\log_2 2e\right)$$

is achievable with perfect secrecy; and for $\mathcal{P} > 4e\sigma^2$, a power-rate pair of

$$\left(\mathcal{P}, \frac{1}{4}\log_2\left(\frac{\mathcal{P}}{\sigma^2}\right) - \frac{1}{4}\log_2 2e\right)$$

is achievable with strong secrecy at the relay nodes in a multi-hop line network with $K+1$ hops.

2) Scheme of He and Yener for Multi-Hop Relaying: We now describe the scheme for secure relaying. The case where S wants to send a single message, X, to the destination is illustrated for a network with three relays in Fig. 14. Only the messages (elements of $\mathbb{G}^{(d)}$) transmitted by each node are indicated in the figure, and it is assumed that actual transmitted vectors are the encoded versions of the messages indicated. The messages available at various nodes at the end of each phase are tabulated in Table I. Suppose that N messages are to be transmitted to the destination. Each relay node generates a jamming signal J_i, chosen uniformly at random from $\mathbb{G}^{(d)}$. The destination node generates N independent jamming signals $J_{K+1}, J_{K+2}, \ldots, J_{K+N}$. Let us use the notation $\oplus_{p=1}^{t} X_p$ to denote $X_1 \oplus X_2 \oplus \cdots \oplus X_t$.

- The source always transmits in phase $2t$, for $t = 1, \ldots, N$.
- The ith node ($i = 1, 2, \ldots, K+1$) transmits in the $(2t+i)$th phase, for $t = 0, 1, \ldots, N$.
- In the ith phase ($i = 1, \ldots, K+1$), the ith node broadcasts the jamming signal J_i (after encoding) to its neighbours. Hence, the signal transmitted by the ith node in the ith phase is

$$\mathbf{V}_i[i] = \mathcal{E}(J_i).$$

- In the $(2t+i)$th phase ($t = 1, 2, \ldots, N$), the ith node sends

$$\mathbf{V}_i[2t+i] = \mathcal{E}\big((\oplus_{p=1}^{t} X_p) \oplus J_{i+t}\big). \tag{57}$$

  This holds for all nodes, $i = 0, 1, \ldots, K+1$. The ith node evaluates $(\oplus_{p=1}^{t} X_p) \oplus J_{i+t}$ by subtracting the message transmitted by it in the $(2t+i-2)$nd phase from the message decoded in the $(2t+i-1)$st phase.

In the first phase, $\mathtt{R}_1$ broadcasts $\mathcal{E}(J_1)$ to the source and $\mathtt{R}_2$, and subsequently, the source can compute $X_1 \oplus J_1$, which it sends in the next phase. In the $(2t-1)$st phase ($t = 1, \ldots, N$), the source receives $\mathcal{E}\big((\oplus_{p=1}^{t-1} X_p) \oplus J_t\big)$ from $\mathtt{R}_1$, and it can compute $(\oplus_{p=1}^{t-1} X_p) \oplus J_t$, and hence $(\oplus_{p=1}^{t} X_p) \oplus J_t$, which is to be sent to $\mathtt{R}_1$ in the next phase.

In the $(i-1)$st phase, the ith relay receives $\mathcal{E}\big(X_1 \oplus J_i\big) + \mathcal{E}(J_{i+1})$, from which it computes $X_1 \oplus J_i \oplus J_{i+1}$, and since $\mathtt{R}_i$ knows J_i, it can find $X_1 \oplus J_{i+1}$, which has to be broadcast in the next phase. In general, in the $(2t+i-1)$st phase, $\mathtt{R}_i$ receives $\mathcal{E}\big((\oplus_{p=1}^{t} X_p) \oplus J_{i+t-1}\big) + \mathcal{E}\big((\oplus_{p=1}^{t-1} X_p) \oplus J_{i+t}\big)$. However, $\mathtt{R}_i$ had sent $\mathcal{E}\big((\oplus_{p=1}^{t-1} X_p) \oplus J_{i+t-1}\big)$ in the $(2t+i-2)$nd phase, and hence it can compute $(\oplus_{p=1}^{t} X_p) \oplus J_{i+t}$, which is to be encoded and sent in the $(2t+i)$th phase.

On similar lines, the destination receives $\mathcal{E}\big(X_1 \oplus J_{K+1}\big)$ in the $(K+2)$nd phase, and transmits $\mathcal{E}\big(X_1 \oplus J_{K+2}\big)$ in the $(K+3)$rd phase, since it knows J_{K+1} which can be subtracted out of $X_1 \oplus J_{K+1}$. In the $(2t+K+1)$st phase, it can compute $(\oplus_{p=1}^{t} X_p) \oplus J_{K+1+t}$ from $\mathcal{E}\big((\oplus_{p=1}^{t} X_p) \oplus J_{K+t}\big)$, which it receives from $\mathtt{R}_K$ in the $(2t+K)$th phase.

3) Secrecy: We now show that using the coding scheme of Section VII-B at each node will guarantee strong secrecy. The argument can be extended similarly for the perfect secrecy case. We will assume for convenience

that all links are noiseless. Since the Gaussian noise is independent of everything else, application of the data processing inequality as in Section VII reveals that secrecy can be obtained even in the noisy case.

Let $\{X_p : p = 1, \ldots, N\}$ denote the set of i.i.d. messages to be sent to the destination. Let us fix a k from $\{1, 2, \ldots, K\}$. We will show that the total information about $\{X_p : p = 1, \ldots, N\}$ obtained by relay k at the end of all relaying operations goes to zero as d goes to infinity.

In the $(2t + k - 1)$st phase, the kth relay receives

$$\mathbf{W}_k[2t + k - 1] = \mathbf{V}_{k-1}[2t + k - 1] + \mathbf{V}_{k+1}[2t + k - 1] \tag{58}$$

$$= \mathcal{E}\Big((\oplus_{p=1}^{t} X_p) \oplus J_{k+t-1}\Big) + \mathcal{E}\Big((\oplus_{p=1}^{t-1} X_p) \oplus J_{k+t}\Big), \tag{59}$$

for $1 \leq t \leq N$. We also have,

$$\mathbf{W}_k[k - 1] = \mathcal{E}(J_{k-1}).$$

For $t = 1, 2, \ldots, N$, let us define

$$\Theta_{k,t} := \{J_k, J_{k-1}, \mathbf{W}_k[2m + k - 1] : 1 \leq m \leq t\} \tag{60}$$

to be the set of all random variables available at the kth relay at the end of the $(2t+k-1)$st phase. We also define $\Theta_{k,0} := \{J_k, J_{k-1}\}$. Note that $\Theta_{k,t-1} \subset \Theta_{k,t}$ for $t = 1, 2, \ldots, N$, and $\Theta_{k,N}$ is the set of all random variables available at the kth relay at the end of all phases. We have to show that $\mathcal{I}(X_1, \ldots, X_N; \Theta_{k,N}) \to 0$ as $d \to \infty$.

Lemma 23. *Let $\epsilon^{(d)} := \epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}/2}) < 1/2$. Then, the total information available at the kth relay node at the end of all relaying phases can be bounded from above as follows:*

$$\mathcal{I}(X_1, \ldots, X_N; \Theta_{k,N}) \leq N\epsilon^{(d)} \left(\log_2 |\mathbb{G}^{(d)}| - \log_2 \epsilon^{(d)}\right), \tag{61}$$

Proof: See Appendix H. ■

Since for our choice of nested lattices, $\epsilon^{(d)} \to 0$ exponentially in d, the mutual information $\mathcal{I}(X_1, \ldots, X_N; \Theta_{k,N})$ also goes to zero exponentially in d, thereby guaranteeing strong secrecy.

4) Achievable Rate and Proof of Theorem 22: Let us fix an arbitrary $\delta > 0$. It was shown in [16, Section IV-C] that using the nested lattice coding scheme of Section VII-B with $(\Lambda^{(d)}, \Lambda_0^{(d)})$ satisfying properties (L1)–(L4), the probability of error, $\eta^{(d)}$ goes to zero as d goes to infinity. Hence, there exists a sequence of lattice pairs such that for sufficiently large d, we can have $\eta^{(d)} < \delta$. Moreover, from Theorem 17, we can have the average transmit power per dimension at each node to be less than $\mathcal{P} + \delta$, and $\frac{1}{d}\log_2 |\mathbb{G}^{(d)}|$ greater than $\frac{1}{2}\log_2 \frac{\mathcal{P}}{\sigma^2} - \frac{1}{2}\log_2 2e - \delta$. Therefore, we can say that a power-rate pair of $\left(\mathcal{P}, \frac{N}{2(K+2N+1)}\left[\log_2\left(\frac{\mathcal{P}}{\sigma^2}\right) - \log_2 2e\right]\right)$ is achievable for the transmission of N messages using this scheme. Letting the number of messages, N, go to infinity, we have the second part of Theorem 22. Analogously, we can show that a power-rate pair of $\left(\mathcal{P}, \frac{N}{2(K+2N+1)}\left[\log_2\left(\frac{\mathcal{P}}{\sigma^2}\right) - 2\log_2 2e\right]\right)$ is achievable for the transmission of N messages with perfect secrecy using the scheme of Section IV, and we have the first part of the theorem. □

IX. Conclusion

We have described two coding schemes for secure bidirectional relaying in presence of an honest-but-curious relay. We saw that using pmfs generated from density functions having compactly supported characteristic functions, one can obtain perfect secrecy. We saw that transmission rates below $\frac{1}{2}\log_2\frac{\mathcal{P}}{\sigma^2} - \log_2 2e$ can be achieved. In order to achieve higher transmission rates, we relaxed the secrecy constraint, and only required that the mutual information between $\mathbf{U}+\mathbf{V}$ and the individual messages go to zero for large block lengths. Using pmfs obtained from sampled Gaussian functions, we could achieve a rate of $\frac{1}{2}\log_2\left(\frac{1}{2}+\frac{\mathcal{P}}{\sigma^2}\right)-\frac{1}{2}\log_2 2e$. These rates are within a constant gap of the best known achievable rate of $\frac{1}{2}\log_2\left(\frac{1}{2}+\frac{\mathcal{P}}{\sigma^2}\right)$ without secrecy constraints [23].

The main theme of this paper was the use of nested lattice codes, and explicit pmfs having infinite support to obtain security. An inherent disadvantage of our scheme is that it is not possible to satisfy a maximum power constraint. One could study the scenario where the support of the distributions we described are truncated, and find the performance of such a scheme, and this is yet to be studied. Two key assumptions made in this paper are that the messages are uniformly distributed, and the channel gains are unity. What happens when these conditions are relaxed is left as future work. In this paper, we only found achievable rates. As remarked in [17], finding a converse result is much harder, and even without any secrecy constraints, a nontrivial outer bound on the capacity of a bidirectional relay is not known.

Appendix A: Proof of Theorem 11

We are given an index-M sublattice Λ_0 of the lattice Λ. Recall from Section II-A that $(\det\Lambda_0)/(\det\Lambda) = M$. Let $\Lambda_0, \Lambda_1, \ldots, \Lambda_{M-1}$ denote the M cosets of Λ_0 in Λ. These constitute the elements of the quotient group $\mathbb{G} = \Lambda/\Lambda_0$.

Suppose that X, Y are iid random variables, each uniformly distributed over $\mathbb{G}$. For each $j \in \{0, 1, \ldots M-1\}$, let p_j be a pmf supported within the coset Λ_j, so that $p_j(\mathbf{k}) = 0$ for $\mathbf{k} \notin \Lambda_j$. We define a random variable U (resp. V) jointly distributed with X (resp. Y) as follows: if $X = \Lambda_j$ (resp. $Y = \Lambda_j$), U (resp. V) is a random point from Λ_j picked according to the distribution p_j. Then, U and V are identically distributed with $p_U = p_V = \frac{1}{M}\sum_{i=0}^{M-1} p_i$. Let φ_U, φ_V and φ_j, $j = 0, 1, \ldots, M-1$, be the characteristic functions corresponding to p_U, p_V and p_j, $j = 0, 1, \ldots, M-1$, respectively. We have the following straightforward generalization of Lemma 3.

Lemma 24. *Suppose that $\varphi_U\varphi_V = \varphi_j\varphi_V = \varphi_U\varphi_j$ for $j = 0, 1, \ldots, M-1$. Then, the random variables (U, V, X, Y) with joint pmf given by*

$$p_{UVXY}(\mathbf{k}, \mathbf{l}, \Lambda_i, \Lambda_j) = (1/M)(1/M)p_i(\mathbf{k})p_j(\mathbf{l}) \quad \text{for } \mathbf{k}, \mathbf{l} \in \Lambda \text{ and } \Lambda_i, \Lambda_j \in \mathbb{G} \tag{62}$$

have properties (S1)–(S3).

We will now construct the characteristic functions φ_j that satisfy the above lemma. Let f be the (continuous) probability density function corresponding to the compactly supported characteristic function ψ in the hypothesis of Theorem 11. The function f can be retrieved from ψ by Fourier inversion:

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \psi(\mathbf{t}) e^{-i\langle \mathbf{t}, \mathbf{x}\rangle}\, d\mathbf{t} \\ &= \frac{1}{(2\pi)^d} \int_{\mathcal{V}(\hat{\Lambda}_0)} \psi(\mathbf{t}) e^{-i\langle \mathbf{t}, \mathbf{x}\rangle}\, d\mathbf{t} \end{aligned} \tag{63}$$

Note that each coset Λ_j can be expressed as $\mathbf{u}_j + \Lambda_0$ for some $\mathbf{u}_j \in \Lambda$. We set

$$\varphi_j(\boldsymbol{\zeta}) = \sum_{\mathbf{n} \in \hat{\Lambda}_0} \psi(\boldsymbol{\zeta} + \mathbf{n})\, e^{-i\langle \mathbf{n},\, \mathbf{u}_j\rangle} \tag{64}$$

for all $\boldsymbol{\zeta} \in \mathbb{R}^d$. Then, by Proposition 6, we have that p_j is supported within Λ_j, and

$$p_j(\mathbf{k}) = (\det \Lambda_0)\, f(\mathbf{k}) \text{ for all } \mathbf{k} \in \Lambda_j. \tag{65}$$

Finally, define

$$\varphi(\boldsymbol{\zeta}) = \sum_{\mathbf{n} \in \hat{\Lambda}} \psi(\boldsymbol{\zeta} + \mathbf{n}) \tag{66}$$

for all $\boldsymbol{\zeta} \in \mathbb{R}^d$.

We make two claims:

(i) $\varphi^2 = \varphi\varphi_j$ for $j = 0, 1, \ldots, M-1$;

(ii) $\varphi = \varphi_U = \varphi_V$.

Given these claims, by Lemma 24, the random variables U, V satisfy the properties (L1)–(L3).

Both claims follow from the fact that $\hat{\Lambda}$ is a sublattice of $\hat{\Lambda}_0$. (If a lattice Γ contains a sublattice Γ_0, then the dual Γ^* is a sublattice of Γ_0^*.) To see (i), we re-write (66) as

$$\varphi(\boldsymbol{\zeta}) = \sum_{\mathbf{n} \in \hat{\Lambda}} \psi(\boldsymbol{\zeta} + \mathbf{n})\, e^{-i\langle \mathbf{n},\, \mathbf{u}_j\rangle} \tag{67}$$

This is possible because, for $\mathbf{n} \in \hat{\Lambda} = 2\pi\Lambda^*$ and $\mathbf{u}_j \in \Lambda$, we have $e^{-i\langle \mathbf{n},\, \mathbf{u}_j\rangle} = 1$. Comparing (64) and (67), and noting that ψ is supported within $\mathcal{V}(\hat{\Lambda}_0)$, it is evident that $\mathrm{supp}(\varphi) := \{\boldsymbol{\zeta} : \varphi(\boldsymbol{\zeta}) \neq 0\}$ is contained in $\mathrm{supp}(\varphi_j) := \{\boldsymbol{\zeta} : \varphi_j(\boldsymbol{\zeta}) \neq 0\}$. Furthermore, for all $\boldsymbol{\zeta} \in \mathrm{supp}(\varphi)$, we have $\varphi(\boldsymbol{\zeta}) = \varphi_j(\boldsymbol{\zeta})$. Claim (i) directly follows from this.

For Claim (ii), we note that $\mathcal{V}(\hat{\Lambda}_0) \subseteq \mathcal{V}(\hat{\Lambda})$, since $\hat{\Lambda}$ is a sublattice of $\hat{\Lambda}_0$. Hence, we can apply Proposition 6 to deduce that φ is the characteristic function of a pmf p supported within Λ, with

$$p(\mathbf{k}) = (\det \Lambda)\, f(\mathbf{k}) \text{ for all } \mathbf{k} \in \Lambda.$$

Thus, from (65) and the fact that $(\det \Lambda_0)/(\det \Lambda) = M$, we see that $p = \frac{1}{M} \sum_{j=0}^{M-1} p_j$. In other words, $p = p_U = p_V$, which proves Claim (ii).

It remains to prove that if ψ is twice differentiable, then $E\|\mathbf{U}\|^2 = -\nabla^2\psi(\mathbf{0})$. Since $\mathbf{U}$ and $\mathbf{V}$ are identically distributed, we would then also have $E\|\mathbf{V}\|^2 = -\nabla^2\psi(\mathbf{0})$. Write $\mathbf{U} = (U_1, \ldots, U_d)$, so that $\|\mathbf{U}\|^2 = U_1^2 +$

$\cdots + U_d^2$. We want to show that $E[U_j^2] = -\frac{\partial^2}{\partial t_j^2}\psi(\mathbf{0})$, for $j = 1, \ldots, d$. For notational simplicity, we show this for $j = 1$. Note that the characteristic function of U_i is given by $\varphi_{U_1}(t_1) = \varphi_U(t_1, 0, \ldots, 0)$. As argued in the proof of Theorem 7, $E[U_1^2] = -\varphi''_{U_1}(0)$. Now, $\varphi''_{U_1}(0) = \frac{\partial^2}{\partial t_1^2}\varphi_U(0, 0, \ldots, 0)$. From (66), we see that $\varphi_U = \psi$ in a small neighbourhood around $\mathbf{0} = (0, 0, \ldots, 0)$. Therefore, $\frac{\partial^2}{\partial t_1^2}\varphi_U(\mathbf{0}) = \frac{\partial^2}{\partial t_1^2}\psi(\mathbf{0})$, and hence, $E[U_1^2] = -\frac{\partial^2}{\partial t_1^2}\psi(\mathbf{0})$, as desired.

This concludes the proof of Theorem 11. □

Appendix B: "Good" Lattice Properties

In this section, we briefly review certain "good" lattice properties, and some results in the literature. This is almost entirely based on [13]. Let $\{\Lambda^{(d)}\}$ be a sequence of lattices, with each $\Lambda^{(d)}$ chosen uniformly at random from a (d, k, q) ensemble described in Section VI-B. It was shown in [13] that with high probability, $\{\Lambda^{(d)}\}$ is simultaneously good for covering, packing, mean-squared error (MSE) quantization, and AWGN channel coding. Let us make these notions more formal.

We say that that the sequence of lattices $\{\Lambda^{(d)}\}$ is *good for covering* if

$$\lim_{d\to\infty} \frac{r_{\text{cov}}(\Lambda^{(d)})}{r_{\text{eff}}(\Lambda^{(d)})} = 1.$$

We say that $\{\Lambda^{(d)}\}$ is *good for packing* if

$$\lim_{d\to\infty} \frac{r_{\text{pack}}(\Lambda^{(d)})}{r_{\text{eff}}(\Lambda^{(d)})} > \frac{1}{2}.$$

Let $\mathcal{G}_{\Lambda^{(d)}}$ denote the normalized second moment per dimension of $\Lambda^{(d)}$, as defined in Section II-A. A sequence of lattices $\{\Lambda^{(d)}\}$ is said to be *good for MSE quantization* if $\mathcal{G}_{\Lambda^{(d)}} \to \frac{1}{2\pi e}$ as $d \to \infty$.

Let $\mathbf{Z}$ be a zero-mean d-dimensional white Gaussian vector having second moment per dimension equal to σ^2. Let

$$\mu := \frac{\text{vol}\big(\mathcal{V}(\Lambda^{(d)})\big)^{2/d}}{\sigma^2}$$

Then we say that $\{\Lambda^{(d)}\}$ is *good for AWGN channel coding* if the probability that $\mathbf{Z}$ lies outside the fundamental Voronoi region of $\Lambda^{(d)}$ is upper bounded by

$$\Pr[\mathbf{Z} \notin \mathcal{V}(\Lambda^{(d)})] \leq e^{-d\left(E_U(\mu) - o_d(1)\right)}$$

for all σ^2 that satisfy $\mu \geq 2\pi e$. Here, $E_U(\cdot)$, called the *Poltyrev exponent* is defined as follows:

$$E_U(\mu) = \begin{cases} \frac{\mu}{16\pi e}, & 8\pi e \leq \mu \\ \frac{1}{2}\ln\frac{\mu}{8\pi}, & 4\pi e \leq \mu \leq 8\pi e \\ \frac{\mu}{4\pi e} - \frac{1}{2}\ln\frac{\mu}{2\pi}, & 2\pi e \leq \mu \leq 4\pi e \end{cases} \tag{68}$$

Suppose that we use a subcollection of points from $\Lambda^{(d)}$ as the codebook for transmission over an AWGN channel. Then, as long as

$$\frac{\text{vol}\big(\mathcal{V}(\Lambda^{(d)})\big)^{2/d}}{\sigma^2} \geq 2\pi e,$$

the probability that a lattice decoder decodes to a lattice point other than the one that was transmitted, decays exponentially in the dimension d, with the exponent given by (68).

It is worth noting that the above "good" properties are invariant to scaling. If $\{\Lambda^{(d)}\}$ is a sequence of lattices that is good for covering, packing, and AWGN channel coding, then for any $\alpha > 0$, $\{\alpha\Lambda^{(d)}\}$ is also good for covering, packing and AWGN channel coding. This is because of the fact that $r_{\text{pack}}(\alpha\Lambda^{(d)}) = \alpha r_{\text{pack}}(\Lambda^{(d)})$, $r_{\text{cov}}(\alpha\Lambda^{(d)}) = \alpha r_{\text{cov}}(\Lambda^{(d)})$, and $r_{\text{eff}}(\alpha\Lambda^{(d)}) = \alpha r_{\text{eff}}(\Lambda^{(d)})$.

Appendix C: Proof of Lemma 14

In proving Lemma 14, we use the following theorem from [13], which says that if the parameters k and q are selected appropriately, then almost all lattices in a (d, k, q) ensemble satisfy the "goodness" properties described in Appendix B.

Theorem 25 ([13], Theorem 5)**.** *Let $0 < r_{min} < \frac{1}{4}$ be chosen arbitrarily. Let $\Lambda^{(d)}$ be a sequence of lattices selected uniformly at random from a (d, k, q) ensemble, such that*

- *$k \leq \beta_1 d$ for some $0 < \beta_1 < 1$, but k grows faster than $\log^2 d$, and*
- *q is chosen so that $r_{\mathit{eff}}(\Lambda^{(d)})$, as given by (31), satisfies $r_{min} < r_{\mathit{eff}}(\Lambda^{(d)}) < 2r_{min}$.*

Then, the sequence of lattices $\Lambda^{(d)}$ is simultaneously good for covering, packing and MSE quantization, with probability approaching 1 as d tends to infinity. If, in addition, we have $\beta_1 < 1/2$, then the sequence of lattices is also simultaneously good for AWGN channel coding with probability tending to 1 as $d \to \infty$.

Therefore, if we choose k and q that satisfy the hypotheses of Lemma 14, then from the above theorem, the probability that a uniformly chosen $\Lambda_0^{(d)}$ satisfies condition (G_1) tends to 1 as $d \to \infty$.

Recall from Section II-A that if A is a generator matrix of a lattice Λ, then the dual lattice of Λ, denoted by Λ^*, is the set of all integer linear combinations of the rows of A^{-1}. It turns out that the dual of a Construction-A lattice is also a Construction-A lattice, as seen from the following.

Proposition 26. *Suppose that G is a $k \times d$ generator matrix of a (d, k) linear code $\mathcal{C}$ over $\mathbb{Z}_q$, q being prime, and G having the form*

$$\mathsf{G} = \begin{bmatrix} \mathsf{I}_k & \mathsf{B} \end{bmatrix},$$

where I_k denotes the $k \times k$ identity matrix. Let $\Lambda(\mathcal{C})$ be the lattice obtained by employing Construction A on the code $\mathcal{C}$. Then, the matrix

$$\mathsf{A} = \frac{1}{q} \begin{bmatrix} \mathsf{I}_k & \mathsf{B} \\ 0 & q\mathsf{I}_{(d-k)} \end{bmatrix} \tag{69}$$

is a generator matrix for the lattice $\Lambda(\mathcal{C})$.

Proof: We want to show that $\mathsf{A}^T\mathbb{Z}^d := \{\mathsf{A}^T\mathbf{y} : \mathbf{y} \in \mathbb{Z}^d\} = \Lambda(\mathcal{C})$. We first prove that $\mathsf{A}^T\mathbb{Z}^d \subseteq \Lambda(\mathcal{C})$. By definition, $\Lambda(\mathcal{C}) = \{\mathbf{x} \in \mathbb{R}^d : (q\mathbf{x}) \bmod q \in \mathcal{C}\}$. Therefore, it is enough to show that $(q\mathsf{A}^T\mathbf{z}) \bmod q \in \mathcal{C}$ for

every $\mathbf{z} \in \mathbb{Z}^d$. Fix a $\mathbf{z} \in \mathbb{Z}^d$. Then,

$$\begin{aligned}
(q\mathsf{A}^T\mathbf{z}) \bmod q &= \left(\begin{bmatrix} \mathsf{I}_k & \mathsf{B} \\ 0 & q\mathsf{I}_{(d-k)} \end{bmatrix}^T [z_1\, z_2 \ldots\, z_d]^T \right) \bmod q \\
&= \Big(\begin{bmatrix} \mathsf{I}_k & \mathsf{B} \end{bmatrix}^T [z_1\, \ldots\, z_k]^T \\
&\quad + \begin{bmatrix} 0 & q\mathsf{I}_{(d-k)} \end{bmatrix}^T [z_{k+1}\, \ldots\, z_d]^T \Big) \bmod q \\
&= \left(\begin{bmatrix} \mathsf{I}_k & \mathsf{B} \end{bmatrix}^T [z_1\, \ldots\, z_k]^T \right) \bmod q \\
&= (\mathsf{G}^T\hat{\mathbf{z}}) \bmod q \in \mathcal{C}. \qquad (70)
\end{aligned}$$

For the converse, define $\mathcal{C}' = \{\frac{1}{q}\mathbf{c} : \mathbf{c} \in \mathcal{C}\}$. Then, $\Lambda(\mathcal{C}) = \mathcal{C}' + \mathbb{Z}^d := \{\mathbf{c} + \mathbf{z} : \mathbf{c} \in \mathcal{C}', \mathbf{z} \in \mathbb{Z}^d\}$. The set $\mathsf{A}^T\mathbb{Z}^d$ forms a group under (componentwise) addition. Hence, it is sufficient to show that $\mathcal{C}' \subseteq \mathsf{A}^T\mathbb{Z}^d$, and $\mathbb{Z}^d \subseteq \mathsf{A}^T\mathbb{Z}^d$. Fix an arbitrary $\mathbf{c} \in \mathcal{C}$. Let $\mathbf{c}' = \frac{1}{q}\mathbf{c}$. By definition, there exists a $\mathbf{x} \in \mathbb{Z}_q^k$ such that

$$\begin{aligned}
\mathbf{c} &= \left(\begin{bmatrix} \mathsf{I}_k & \mathsf{B} \end{bmatrix}^T \mathbf{x} \right) \bmod q \\
&= \begin{bmatrix} \mathbf{x} \\ \mathsf{B}^T\mathbf{x} \end{bmatrix} \bmod q = \begin{bmatrix} \mathbf{x} \\ (\mathsf{B}^T\mathbf{x}) \bmod q \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{x} \\ \mathsf{B}^T\mathbf{x} \end{bmatrix} - q \begin{bmatrix} 0 \\ \mathbf{z}' \end{bmatrix}, \qquad (71)
\end{aligned}$$

for some $\mathbf{z}' \in \mathbb{Z}^{d-k}$. Therefore,

$$\mathbf{c} = \begin{bmatrix} \mathsf{I}_k & \mathsf{B} \\ 0 & q\mathsf{I}_{(d-k)} \end{bmatrix}^T \begin{bmatrix} \mathbf{x} \\ -\mathbf{z}' \end{bmatrix}.$$

Hence, there exists

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} \\ \mathbf{z}' \end{bmatrix} \in \mathbb{Z}^d$$

so that $\mathbf{c}' = \mathsf{A}^T\mathbf{z}$. Therefore, we can say that $\mathcal{C}' \subseteq \mathsf{A}^T\mathbb{Z}^d$. Next, consider $\mathbf{z} \in \mathbb{Z}^d$. Let A^* be as in Lemma 27, and note that $\mathsf{A}^T\mathsf{A}^* = \mathsf{I}_d$, the $d \times d$ identity matrix. Let $\mathbf{z}' = \mathsf{A}^*\mathbf{z} \in \mathbb{Z}^d$. Then, $\mathsf{A}^T\mathbf{z}' = \mathsf{A}^T(\mathsf{A}^*\mathbf{z}) = (\mathsf{A}^T\mathsf{A}^*)\mathbf{z} = \mathbf{z}$. Hence, we can say that for every $\mathbf{z} \in \mathbb{Z}^d$, there exists a $\mathbf{z}' \in \mathbb{Z}^d$ so that $\mathbf{z} = \mathsf{A}^T\mathbf{z}$, and hence $\mathbb{Z}^d \subseteq \mathsf{A}^T\mathbb{Z}^d$, thus concluding the proof. ■

It can be shown in a similar manner that if G has the form

$$\mathsf{G} = \begin{bmatrix} \mathsf{B} & \mathsf{I}_k \end{bmatrix},$$

then,

$$\mathsf{A} = \frac{1}{q} \begin{bmatrix} \mathsf{B} & \mathsf{I}_k \\ q\mathsf{I}_{(d-k)} & 0 \end{bmatrix}$$

is a generator matrix for $\Lambda(\mathcal{C})$.

Lemma 27. *Let $\mathcal{C}$, G, $\Lambda(\mathcal{C})$ be as in Proposition 26. Then, the dual of $\Lambda(\mathcal{C})$, denoted by $\Lambda^*(\mathcal{C})$, has generator matrix*

$$\mathsf{A}^* = \begin{bmatrix} q\mathsf{I}_k & \mathsf{0} \\ -\mathsf{B}^T & \mathsf{I}_{(d-k)} \end{bmatrix}. \tag{72}$$

Therefore, $\Lambda^*(\mathcal{C}) = q\Lambda(\mathcal{C}^\perp)$, where $\mathcal{C}^\perp$ denotes the dual code of $\mathcal{C}$.

Proof: Consider,

$$\begin{aligned} \mathsf{A}(\mathsf{A}^*)^T &= \frac{1}{q}\begin{bmatrix} \mathsf{I}_k & \mathsf{B} \\ \mathsf{0} & q\mathsf{I}_{(d-k)} \end{bmatrix}\begin{bmatrix} q\mathsf{I}_k & -\mathsf{B} \\ \mathsf{0} & \mathsf{I}_{(d-k)} \end{bmatrix} \\ &= \frac{1}{q}\begin{bmatrix} q\mathsf{I}_k & \mathsf{0} \\ \mathsf{0} & q\mathsf{I}_{(d-k)} \end{bmatrix} \\ &= \mathsf{I}_d. \end{aligned}$$

Similarly, $(\mathsf{A}^*)^T\mathsf{A} = \mathsf{I}_d$.

Since a permutation of the rows of a generator matrix of a lattice also yields a valid generator matrix for the same lattice,

$$\mathsf{A}_1^* = \begin{bmatrix} -\mathsf{B}^T & \mathsf{I}_{(d-k)} \\ q\mathsf{I}_k & \mathsf{0} \end{bmatrix}$$

is also a generator matrix for $\Lambda^*(\mathcal{C})$. If $\mathcal{C}^\perp$ denotes the dual code of $\mathcal{C}$, then $\mathcal{C}^\perp$ has a generator matrix [30]

$$\mathsf{G} = \begin{bmatrix} -\mathsf{B}^T & \mathsf{I}_{(d-k)} \end{bmatrix}.$$

Therefore, we can conclude that the dual lattice, $\Lambda^*(\mathcal{C})$ is a q-scaled version of the lattice obtained by applying Construction A to $\mathcal{C}^\perp$, i.e., $\Lambda^*(\mathcal{C}) = q\Lambda(\mathcal{C}^\perp)$. ■

Since, $\Lambda^*(\mathcal{C}) = q\Lambda(\mathcal{C}^\perp)$, if the generator matrix is full-rank, then $\Lambda(\mathcal{C}^\perp)$ belongs to a $(d, d-k, q)$ ensemble. Therefore, from [13], we can say that a randomly picked $\Lambda(\mathcal{C}^\perp)$ is good for packing and covering with probability tending to 1 as $d \to \infty$, as long as $d - k \leq \beta_1 d$ for some $0 < \mu < 1$, and $d - k$ grows faster than $\log^2 d$. From the definitions, we see that the properties of covering and packing goodness are invariant to any scaling of the lattices. Therefore, if $\Lambda(\mathcal{C}^\perp)$ is good for packing and covering, then $q\Lambda(\mathcal{C}^\perp)$, and hence $\Lambda^*(\mathcal{C})$ is also good for packing and covering. Since the probability that $\{\Lambda^{(d)}\}$ is packing- and covering-good tends to one as d tends to ∞, by the union bound, we can argue that if $k = \beta_1 d$ for some $\beta_1 < 1/2$, then a randomly picked sequence of lattices is good for covering, packing and AWGN coding, together with the property that its dual is good for packing and covering with probability going to 1 as $d \to \infty$. Therefore, we can find a sequence of coarse lattices that satisfy (G_1) and (G_2).

It was also shown in [23, Appendix B] (also see [12]) that if the coarse lattices are good for covering and AWGN channel coding, then as long as $d/q_1 \to 0$ as $d \to \infty$, the probability that a uniformly chosen sequence of fine lattices is good for AWGN channel coding tends to 1 as $d \to \infty$. This completes the proof of Lemma 14.

Appendix D: Proof of Proposition 15

We can make a stronger statement than Proposition 15, and characterize the error exponent of the lattice decoder, as the following lemma, adapted from [12], says. As long as the transmission rate is less than $\frac{1}{2}\log_2 \frac{\mathsf{M}}{\sigma^2}$, the error exponent is positive, and hence the probability of error goes to zero as $d \to \infty$. The analysis of the error exponent is interesting in its own right, as it gives insights on the error performance of the decoder when the coarse lattices do not satisfy the goodness properties discussed earlier. The lemma is quite general, as it indicates how well one can do given an arbitrary sequence of coarse lattices.

Lemma 28. *Let $\{\Lambda_0^{(d)}\}$ be a sequence of coarse lattices satisfying $r_{\mathit{eff}}(\Lambda_0^{(d)}) = \sqrt{dM}$ for each d. Then, there exists a sequence of lattices $\{\Lambda^{(d)}\}$, with $\Lambda_0^{(d)} \subseteq \Lambda^{(d)}$ for all d, and $R^{(d)} = \frac{1}{d}\log_2 |\Lambda^{(d)}/\Lambda_0^{(d)}|$, for which the probability of error of decoding $X \oplus Y$ at the relay is upper bounded by*

$$Pr[\mathcal{D}^{(d)}(\mathbf{U}+\mathbf{V}+\mathbf{Z}) \neq x \oplus y | X = x, Y = y] \leq e^{-dE_{\Lambda_0}(R^{(d)})}, \tag{73}$$

where the exponent

$$E_{\Lambda_0}(R^{(d)}) > E_U\left(2\pi e^{1-2(\frac{1}{2}\log_2 \frac{M}{\sigma^2} - R^{(d)} - \delta_1(d))}\right) - \delta_2(d) - \delta_3(d). \tag{74}$$

Here, $E_U(\cdot)$ denotes the Poltyrev exponent given by (68), and $\delta_1(d) \to 0$ as $d \to \infty$. The quantities $\delta_2(d)$ and $\delta_3(d)$, given by (79) and (84), go to zero as $d \to \infty$ provided that the sequence of coarse lattices are good for covering and AWGN channel coding, respectively.

The quantities δ_2, δ_3 quantify how far the coarse lattices are from satisfying (G_1) and show the deviation from the ideal error performance.

Proof: Let us consider a fixed sequence of coarse lattices $\{\Lambda_0^{(d)}\}$. Recall that $M^{(d)}$ denotes the number of distinct messages at each node. Let $R := \frac{1}{d}\log_2 M^{(d)}$ denote the rate of transmission. We omit the superscript (d) in R for convenience. As described in Section VI-C, the decoder at the relay sees an effective $\Lambda_0^{(d)}$-MLAN channel. We study the probability of error for a $\Lambda_0^{(d)}$-MLAN channel with additive Gaussian noise.

Let us pick 2^{dR} codewords uniformly at random over $\mathcal{V}(\Lambda_0^{(d)})$. The messages are mapped to the 2^{dR} codewords. From [12, Appendix B], (also see [11]) the random coding error exponent, $E_{\Lambda_0}^r(R)$, for the $\Lambda_0^{(d)}$-MLAN channel under minimum Euclidean distance decoding is lower bounded as

$$E_{\Lambda_0}^r(R) > \max_{0<\rho\leq 1} \rho\Big[\frac{1}{d}\log_2\left(\mathrm{vol}(\mathcal{V}(\Lambda_0^{(d)}))\right) - \frac{1}{d}h_{\bar{\rho}}(\mathbf{Z}) - R\Big] - \frac{1}{d}\log_2\left(\frac{1}{1-\delta_{\mathbf{z}}(d)}\right), \tag{75}$$

Here, $\bar{\rho} = 1/(1+\rho)$, and $\mathbf{Z}$ denotes the zero-mean Gaussian noise vector with second moment per dimension equal to σ^2, and $\delta_{\mathbf{z}}(d) = \Pr[\mathbf{Z} \notin \mathcal{V}(\Lambda_0^{(d)})]$. The term, $h_{\bar{\rho}}(\mathbf{Z})$ is the Renyi entropy of order $\bar{\rho}$ of the random variable $\mathbf{Z}$ (with density $f_{\mathbf{Z}}$), defined as follows.

$$h_{\bar{\rho}}(\mathbf{Z}) := \frac{\bar{\rho}}{1-\bar{\rho}}\log_2\left(\int_{\mathbf{a}} f_{\mathbf{Z}}(\mathbf{a})^{\bar{\rho}} d\mathbf{a}\right)^{1/\bar{\rho}}. \tag{76}$$

Recall that $r_{\text{eff}}(\Lambda_0^{(d)})$ is the effective radius of $\Lambda_0^{(d)}$. Then, (see e.g., [13, Section II-D])

$$\text{vol}\left(\mathcal{V}(\Lambda_0^{(d)})\right) = \left(\frac{{r_{\text{eff}}}^2(\Lambda_0^{(d)})}{(d+2)\mathcal{G}^{(d)}}\right)^{d/2} = \left(\frac{d\mathsf{M}}{(d+2)\mathcal{G}^{(d)}}\right)^{d/2},$$

where $\mathcal{G}^{(d)}$ is the normalized second moment of the unit ball in d dimensions, and $\mathcal{G}^{(d)} \to \frac{1}{2\pi e}$ as $d \to \infty$. Using this in (76), we get

$$E_{\Lambda_0}^r(R) > \max_{0<\rho\leq 1} \rho\Big[\frac{1}{2}\log_2\left(2\pi e\mathsf{M}\right) - \frac{1}{d}h_{\bar{\rho}}(\mathbf{Z}) - R - \frac{1}{2}\log_2\left(\frac{(d+2)}{d}2\pi e\mathcal{G}^{(d)}\right)\Big] - \frac{1}{d}\log_2\left(\frac{1}{1-\delta_{\mathbf{z}}(d)}\right), \tag{77}$$

Let us define

$$\delta_1(d) := \frac{1}{2}\log_2\left(\frac{(d+2)}{d}2\pi e\mathcal{G}^{(d)}\right), \tag{78}$$

The term $\delta_1(d) \to 0$ as $d \to \infty$. We also define

$$\delta_2(d) = \frac{1}{d}\log_2\left(\frac{1}{1-\delta_{\mathbf{z}}(d)}\right) \tag{79}$$

Since $\delta_{\mathbf{z}}(d)$ denotes the probability that the noise vector $\mathbf{Z}$ is not within $\mathcal{V}(\Lambda_0^{(d)})$, the quantity $\delta_2(d)$ tells us how good the coarse lattices are for AWGN channel coding.

Using (78) and (79), (77) can be written as

$$E_{\Lambda_0}^r(R) > \max_{0<\rho\leq 1} \rho\Big[\frac{1}{2}\log_2\left(2\pi e\mathsf{M}\right) - \frac{1}{d}h_{\bar{\rho}}(\mathbf{Z}) - (R+\delta_1(d))\Big] - \delta_2(d), \tag{80}$$

Define $C := \frac{1}{2}\log_2\left(\frac{\mathsf{M}}{\sigma^2}\right)$. Then, optimizing the above equation with respect to ρ gives us (similar to [12, Appendix A])

$$E_{\Lambda_0}^r(R) > E_U\left(2\pi e^{1+2(C-R-\delta_1(d))}\right) - \delta_2(d), \tag{81}$$

for all transmission rates R that lie between $C-(\log_2 2)/2$ and C. Here, $E_U(\cdot)$ denotes the Poltyrev exponent given by (68). A similar argument also holds for the expurgated error exponent $E_{\Lambda_0}^x(R)$ [12], and

$$E_{\Lambda_0}^x(R) > E_U\left(2\pi e^{1+2(C-R-\delta_1(d))}\right) - \delta_2(d), \tag{82}$$

for R between 0 and $C-(\log_2 2)/2$.

Observe that, in the analysis of the error exponent, the codewords are picked uniformly over $\mathcal{V}(\Lambda_0^{(d)})$. Now suppose that the codewords are instead distributed uniformly over a fine grid within $\mathcal{V}(\Lambda_0^{(d)})$, in particular, over $\left(q_1^{-1}\Lambda_0^{(d)}\right) \cap \mathcal{V}(\Lambda_0^{(d)})$, where q_1 is prime. Our codebook comes from the fine lattice, and these points are distributed uniformly over $\left(q_1^{-1}\Lambda_0^{(d)}\right) \cap \mathcal{V}(\Lambda_0^{(d)})$. This is the case if we pick a sublattice from the (d, k_1, q_1) ensemble described in steps (f_1)–(f_2) as earlier [23, Lemma 3]. Then, this quantization of the input does not affect the error exponent by a significant amount as long as d/q_1 goes to zero as $d \to \infty$. Indeed from [12, Appendix C], we have that

$$E_{\Lambda_0}^r(R) > E_U\left(2\pi e^{1+2(C-R-\delta_1(d))}\right) - \delta_2(d) - \delta_3(d), \tag{83}$$

where

$$\delta_3(d) = \frac{{r_{\rm cov}}^2(\Lambda_0^{(d)})}{q_1\sigma^2}\left(1+\frac{1}{2q_1}\right), \tag{84}$$

with $r_{\rm cov}(\Lambda_0^{(d)})$ being the covering radius of $\Lambda_0^{(d)}$. If the sequence of coarse lattices is good for covering, and $d/q_1 \to 0$ as $d\to\infty$, then $\delta_3(d)\to 0$ as $d\to\infty$. If, in addition, $\Lambda_0^{(d)}$ is good for AWGN channel coding, then, $\delta_2(d)\to 0$ as $d\to\infty$. ■

Appendix E: Proof of Lemma 16

For ease of notation, denote by $r_{\rm eff}$, the effective radius of $\Lambda_0^{(d)}$. The index, d, in $r_{\rm eff}$ has been dropped but it must be understood that this is a function of d. Let $\mathcal{C}^{(d)}$ denote the (d,k) code over $\mathbb{Z}_q$ that is used to generate the coarse lattice. Using (30),

$$\begin{aligned} q^k &= \frac{\Gamma(d/2+1)}{\pi^{d/2}{r_{\rm eff}}^d} \\ &= \sqrt{d\pi}\left(\frac{d}{2\pi e{r_{\rm eff}}^2}\right)^{d/2}(1+o_d(1)), \end{aligned} \tag{85}$$

where the second step uses Stirling's approximation, and $o_d(1)$ is a term that approaches 0 as $d\to\infty$. From (32), $k=\beta_0 d$ for some $0<\beta_0<1/2$. Substituting this in the above, and raising both sides to the power $1/d$, we get

$$q^{\beta_0} = (d\pi)^{\frac{1}{2d}}\left(\frac{d}{2\pi e{r_{\rm eff}}^2}\right)^{1/2}(1+o_d(1))^{1/d} = (d\pi)^{\frac{1}{2d}}\frac{\sqrt{d}}{\sqrt{2\pi e}r_{\rm eff}}(1+o_d(1)). \tag{86}$$

Let $\Lambda_0^{(d)*}$ denote the dual of $\Lambda_0^{(d)}$, and ${r_{\rm eff}}^*$ denote the effective radius of $\Lambda_0^{(d)*}$. Let $\Lambda_0(\mathcal{C}^{(d)\perp})$ be the lattice obtained by applying Construction-A on the dual of $\mathcal{C}^{(d)}$, i.e., on $\mathcal{C}^{(d)\perp}$. As remarked in Appendix C, $\Lambda_0(\mathcal{C}^{(d)\perp})$ comes from a $(d,d-k,q)$ ensemble. From Lemma 27, $\Lambda_0^{(d)*} = q\Lambda_0(\mathcal{C}^{(d)\perp})$. Therefore, $(1/q)\Lambda^{(d)*} = \Lambda(\mathcal{C}^{(d)\perp})$ will satisfy

$$q^{d-k} = \sqrt{d\pi}\left(\frac{d}{2\pi e\left(r_{\rm eff}(\frac{1}{q}\Lambda^{(d)*})\right)^2}\right)^{d/2}(1+o_d(1))$$

where $o_d(1)\to 0$ as $d\to\infty$. But $r_{\rm eff}(\frac{1}{q}\Lambda_0^{(d)*}) = \frac{1}{q}{r_{\rm eff}}^*$, and hence, analogous to (86), we have

$$q^{d(1-\beta_0)} = \sqrt{d\pi}\left(\frac{d}{2\pi e(1/q)^2({r_{\rm eff}}^*)^2}\right)^{d/2}(1+o_d(1)) \tag{87}$$

Rearranging,

$${r_{\rm eff}}^* = (d\pi)^{\frac{1}{2d}}\frac{\sqrt{d}q^{\beta_0}}{\sqrt{2\pi e}}(1+o_d(1))^{1/d} \tag{88}$$

Let the packing radius of $\Lambda_0^{(d)*}$ be $r_{\rm pack}(\Lambda_0^{(d)*}) = \gamma(d){r_{\rm eff}}^*$. From the definition of the packing radius, $\gamma(d)<1$ for all d. Again, since the dual lattice is good for packing, $\lim_{d\to\infty}\gamma(d)\geq 1/2$. Also, since $o_d(1)\to 0$ as $d\to\infty$, we have $(1+o_d(1))^{1/d} = (1+o_d(1))$. Therefore, we have,

$$r_{\rm eff}(\Lambda_0^{(d)})r_{\rm pack}(\Lambda_0^{(d)*}) = \gamma(d)r_{\rm eff}(\Lambda_0^{(d)})(d\pi)^{(1/2d)}\frac{\sqrt{d}q^{\beta_0}}{\sqrt{2\pi e}}(1+o_d(1))$$

Substituting for q^{β_0} from (86) in the above equation, we get

$$\frac{r_{\text{eff}}(\Lambda_0^{(d)})r_{\text{pack}}(\Lambda_0^{(d)*})}{d} = \gamma(d)(d\pi)^{(1/d)}\frac{1}{2\pi e}(1+o_d(1)) \tag{89}$$

Therefore, as $d \to \infty$, the above expression converges to a value greater than or equal to $1/4\pi e$. Using $r_{\text{pack}}(\hat{\Lambda}_0^{(d)}) = 2\pi r_{\text{pack}}(\Lambda_0^{(d)*})$, we get Lemma 16. □

Appendix F: Proof of Theorem 19

The following lemma from [20] will be used in the proof.

Lemma 29 ([20], Lemma 4)**.** *Let Λ be a lattice in $\mathbb{R}^d$. Then, for all $\mathbf{z} \in \mathbb{R}^d$, and $\kappa > 0$,*

$$\frac{1-\epsilon_\Lambda(\kappa)}{1+\epsilon_\Lambda(\kappa)} \le \frac{g_{\kappa,\mathbf{z}}(\Lambda)}{g_\kappa(\Lambda)} \le 1.$$

For ease of notation, we will suppress the index d in $\epsilon^{(d)}$, $\Lambda_0^{(d)}$ and $\Lambda^{(d)}$. We will find upper and lower bounds for $p_{U+V}(\mathbf{u})$ and $p_{U+V|\mathbf{x}}(\mathbf{u})$, and then use these to get an upper bound on the absolute value of the difference between the two.

For a message X chosen at node $\mathtt{A}$, let $\mathbf{x}$ be the coset representative of X from $\Lambda \cap \mathcal{V}(\Lambda_0)$. For any subset $S \subseteq \mathbb{R}^d$, let $\mathbf{1}_S(\cdot)$ denote the indicator function of S, i.e., $\mathbf{1}_S(\mathbf{u})$ is 1 if $\mathbf{u} \in S$, and 0 otherwise. From (40), with $\kappa = \sqrt{\mathcal{P}}$, we have

$$p_{U|\mathbf{x}}(\mathbf{u}) = \frac{g_{\sqrt{\mathcal{P}}}(\mathbf{u})}{g_{\sqrt{\mathcal{P}},-\mathbf{x}}(\Lambda_0)}\mathbf{1}_{\Lambda_0+\mathbf{x}}(\mathbf{u}). \tag{90}$$

Let $\mathbb{G}_X := \Lambda \cap \mathcal{V}(\Lambda_0)$, and $M := |\mathbb{G}^{(d)}| = |\mathbb{G}_X|$. Since the messages are uniformly distributed,

$$p_U(\mathbf{u}) = \sum_{\mathbf{x}\in\mathbb{G}_X} \frac{g_{\sqrt{\mathcal{P}}}(\mathbf{u})}{g_{\sqrt{\mathcal{P}},-\mathbf{x}}(\Lambda_0)}\frac{\mathbf{1}_{(\Lambda_0+\mathbf{x})}(\mathbf{u})}{M}. \tag{91}$$

The flatness factor, $\epsilon_{\Lambda_0}(\theta)$ is a decreasing function of θ [20, Remark 3]. Therefore, $\epsilon_{\Lambda_0}(\sqrt{\mathcal{P}}) < \epsilon_{\Lambda_0}(\sqrt{\mathcal{P}/2}) = \epsilon$, and using Lemma 29,

$$\frac{g_{\sqrt{\mathcal{P}}}(\mathbf{u})}{g_{\sqrt{\mathcal{P}}}(\Lambda_0)} \le \frac{g_{\sqrt{\mathcal{P}}}(\mathbf{u})}{g_{\sqrt{\mathcal{P}},-\mathbf{x}}(\Lambda_0)} \le \frac{g_{\sqrt{\mathcal{P}}}(\mathbf{u})}{g_{\sqrt{\mathcal{P}}}(\Lambda_0)}\frac{1+\epsilon}{1-\epsilon}.$$

Using this in (91), we get for $\mathbf{u} \in \Lambda$,

$$\frac{g_{\sqrt{\mathcal{P}}}(\mathbf{u})}{Mg_{\sqrt{\mathcal{P}}}(\Lambda_0)} \le p_U(\mathbf{u}) \le \frac{g_{\sqrt{\mathcal{P}}}(\mathbf{u})}{Mg_{\sqrt{\mathcal{P}}}(\Lambda_0)}\frac{1+\epsilon}{1-\epsilon}. \tag{92}$$

We will require bounds on $g_{\sqrt{\mathcal{P}}}(\Lambda)$ in the proof. Rearranging the terms above,

$$\left(\frac{1-\epsilon}{1+\epsilon}\right) p_U(\mathbf{u}) M g_{\sqrt{\mathcal{P}}}(\Lambda_0) \le g_{\sqrt{\mathcal{P}}}(\mathbf{u}) \le p_U(\mathbf{u}) M g_{\sqrt{\mathcal{P}}}(\Lambda_0).$$

Since p_U is a pmf supported over Λ, and $\sum_{\mathbf{u}\in\Lambda} p_U(\mathbf{u}) = 1$, we can get

$$\left(\frac{1-\epsilon}{1+\epsilon}\right) M g_{\sqrt{\mathcal{P}}}(\Lambda_0) \le g_{\sqrt{\mathcal{P}}}(\Lambda) \le M g_{\sqrt{\mathcal{P}}}(\Lambda_0). \tag{93}$$

It can be similarly verified that for any $\mathbf{a} \in \mathbb{R}^n$,

$$\left(\frac{1-\epsilon}{1+\epsilon}\right) M g_{\sqrt{\frac{\mathcal{P}}{2}},\mathbf{a}}(\Lambda_0) \leq g_{\sqrt{\frac{\mathcal{P}}{2}},\mathbf{a}}(\Lambda) \leq M g_{\sqrt{\frac{\mathcal{P}}{2}},\mathbf{a}}(\Lambda_0). \tag{94}$$

We establish some more notation for convenience. Let

$$\alpha(\mathbf{w}) := \frac{g_{\sqrt{2\mathcal{P}}}(\mathbf{w})}{M g_{\sqrt{\mathcal{P}}}(\Lambda_0)} \frac{g_{\sqrt{\frac{\mathcal{P}}{2}}}(\Lambda_0)}{g_{\sqrt{\mathcal{P}}}(\Lambda_0)}, \tag{95}$$

$$\beta(\mathbf{x},\mathbf{w}) := \left(\frac{g_{\sqrt{\frac{\mathcal{P}}{2}},\frac{\mathbf{w}}{2}-\mathbf{x}}(\Lambda_0)}{g_{\sqrt{\frac{\mathcal{P}}{2}}}(\Lambda_0)}\right)\left(\frac{g_{\sqrt{\mathcal{P}},-\mathbf{x}}(\Lambda_0)}{g_{\sqrt{\mathcal{P}}}(\Lambda_0)}\right)^{-1}. \tag{96}$$

We can bound $p_{U+V|\mathbf{x}}$ and p_{U+V} as follows.

Lemma 30. *For any lattice point $\mathbf{w} \in \Lambda$, and any $\mathbf{x} \in \mathbb{G}_X$, we have*

$$\left(\frac{1-\epsilon}{1+\epsilon}\right)\alpha(\mathbf{w}) \leq p_{U+V}(\mathbf{w}) \leq \left(\frac{1+\epsilon}{1-\epsilon}\right)^2 \alpha(\mathbf{w}) \tag{97}$$

$$\beta(\mathbf{x},\mathbf{w})\alpha(\mathbf{w}) \leq p_{U+V|\mathbf{x}}(\mathbf{w}) \leq \left(\frac{1+\epsilon}{1-\epsilon}\right)\beta(\mathbf{x},\mathbf{w})\alpha(\mathbf{w}). \tag{98}$$

Proof: Let $\mathbf{x}$ be any fine lattice point from $\mathbb{G}_X$. Then,

$$p_{U+V|\mathbf{x}}(\mathbf{w}) = \sum_{\mathbf{t}\in\Lambda_0+\mathbf{x}} p_{U|\mathbf{x}}(\mathbf{t}) p_V(\mathbf{w}-\mathbf{t})$$

Using (90) and (92) in the above equation, we obtain

$$\sum_{\mathbf{t}\in\Lambda_0+\mathbf{x}} \frac{g_{\sqrt{\mathcal{P}}}(\mathbf{t})}{g_{\sqrt{\mathcal{P}},-\mathbf{x}}(\Lambda_0)} \frac{g_{\sqrt{\mathcal{P}}}(\mathbf{w}-\mathbf{t})}{M g_{\sqrt{\mathcal{P}}}(\Lambda_0)} \leq p_{U+V|\mathbf{x}}(\mathbf{w}) \leq \sum_{\mathbf{t}\in\Lambda_0+\mathbf{x}} \frac{g_{\sqrt{\mathcal{P}}}(\mathbf{t})}{g_{\sqrt{\mathcal{P}},-\mathbf{x}}(\Lambda_0)} \frac{g_{\sqrt{\mathcal{P}}}(\mathbf{w}-\mathbf{t})}{M g_{\sqrt{\mathcal{P}}}(\Lambda_0)} \left(\frac{1+\epsilon}{1-\epsilon}\right). \tag{99}$$

Consider the term

$$\begin{aligned}
\sum_{\mathbf{t}\in\Lambda_0+\mathbf{x}} \frac{g_{\sqrt{\mathcal{P}}}(\mathbf{t})}{g_{\sqrt{\mathcal{P}},-\mathbf{x}}(\Lambda_0)} \frac{g_{\sqrt{\mathcal{P}}}(\mathbf{w}-\mathbf{t})}{M g_{\sqrt{\mathcal{P}}}(\Lambda_0)} &= \frac{1}{g_{\sqrt{\mathcal{P}},-\mathbf{x}}(\Lambda_0)} \frac{1}{M g_{\sqrt{\mathcal{P}}}(\Lambda_0)} \sum_{\mathbf{t}\in\Lambda_0+\mathbf{x}} \frac{e^{\left(-\frac{\|\mathbf{t}\|^2}{2\mathcal{P}} - \frac{\|\mathbf{t}-\mathbf{w}\|^2}{2\mathcal{P}}\right)}}{(2\pi\sqrt{\mathcal{P}})^d} \\
&= \frac{1}{M g_{\sqrt{\mathcal{P}}}(\Lambda_0) g_{\sqrt{\mathcal{P}},-\mathbf{x}}(\Lambda_0)} \sum_{\mathbf{t}\in\Lambda_0+\mathbf{x}} \frac{e^{\left(-\frac{\|\mathbf{w}\|^2}{4\mathcal{P}} - \frac{\|\mathbf{t}-\frac{\mathbf{w}}{2}\|^2}{\mathcal{P}}\right)}}{(2\pi\sqrt{\mathcal{P}})^d} \\
&= \frac{g_{\sqrt{2\mathcal{P}}}(\mathbf{w})}{M g_{\sqrt{\mathcal{P}}}(\Lambda_0) g_{\sqrt{\mathcal{P}},-\mathbf{x}}(\Lambda_0)} \sum_{\mathbf{t}\in\Lambda_0+\mathbf{x}} g_{\sqrt{\frac{\mathcal{P}}{2}},\frac{\mathbf{w}}{2}}(\mathbf{t}) \\
&= \frac{g_{\sqrt{2\mathcal{P}}}(\mathbf{w})}{M g_{\sqrt{\mathcal{P}}}(\Lambda_0)} \frac{g_{\sqrt{\frac{\mathcal{P}}{2}},\frac{\mathbf{w}}{2}-\mathbf{x}}(\Lambda_0)}{g_{\sqrt{\mathcal{P}},-\mathbf{x}}(\Lambda_0)}
\end{aligned} \tag{100}$$

Substituting this in (99), and writing this in terms of α and β, we obtain (98). Similarly, bounding both p_U and p_V from above and below using (92), proceeding as above, and finally using (94) to bound $g_{\sqrt{\frac{\mathcal{P}}{2}},\frac{\mathbf{w}}{2}}(\Lambda)$, we get (97). ∎

Observe that $\beta(\mathbf{x},\mathbf{w})$ in (96) is a ratio of two terms, both of which can be bounded using Lemma 29 to get

$$\left(\frac{1-\epsilon}{1+\epsilon}\right) \leq \beta(\mathbf{x},\mathbf{w}) \leq \left(\frac{1+\epsilon}{1-\epsilon}\right). \tag{101}$$

Let $\overline{p}_{U+V}$ and $\underline{p}_{U+V}$ respectively denote the upper and lower bounds for p_{U+V} in (97), and let $\overline{p}_{U+V|\mathbf{x}}$ and $\underline{p}_{U+V|\mathbf{x}}$ respectively denote the upper and lower bounds for $p_{U+V|\mathbf{x}}$ in (98). Then, we can say that $|p_{U+V|\mathbf{x}}(\mathbf{w}) - p_{U+V}(\mathbf{w})|$ is less than or equal to the maximum of $|\overline{p}_{U+V|\mathbf{x}}(\mathbf{w}) - \underline{p}_{U+V}(\mathbf{w})|$ and $|\underline{p}_{U+V|\mathbf{x}}(\mathbf{w}) - \overline{p}_{U+V}(\mathbf{w})|$.

Substituting for $|\overline{p}_{U+V|\mathbf{x}}(\mathbf{w}) - \underline{p}_{U+V}(\mathbf{w})|$, we get

$$|\overline{p}_{U+V|\mathbf{x}}(\mathbf{w}) - \underline{p}_{U+V}(\mathbf{w})| = \alpha(\mathbf{w}) \left(\frac{1-\epsilon}{1+\epsilon}\right) \left| \left(\frac{1+\epsilon}{1-\epsilon}\right)^2 \beta(\mathbf{x},\mathbf{w}) - 1 \right|. \tag{102}$$

However, from (101), we see that

$$1 < \left(\frac{1+\epsilon}{1-\epsilon}\right) \leq \left(\frac{1+\epsilon}{1-\epsilon}\right)^2 \beta(\mathbf{x},\mathbf{w}) \leq \left(\frac{1+\epsilon}{1-\epsilon}\right)^3,$$

and for $\epsilon \leq 1/2$, we have $\left(\frac{1+\epsilon}{1-\epsilon}\right)^3 \leq 1 + 64\epsilon$. Therefore,

$$|\overline{p}_{U+V|\mathbf{x}}(\mathbf{w}) - \underline{p}_{U+V}(\mathbf{w})| \leq \alpha(\mathbf{w}) \left(\frac{1-\epsilon}{1+\epsilon}\right) 64\epsilon. \tag{103}$$

Similarly, expressing $|\underline{p}_{U+V|\mathbf{x}}(\mathbf{w}) - \overline{p}_{U+V}(\mathbf{w})|$ in terms of α and β, and using the fact that $((1-\epsilon)/(1+\epsilon))^3 \geq 1 - 8\epsilon$ for $\epsilon < 1/2$, we get

$$|\underline{p}_{U+V|\mathbf{x}}(\mathbf{w}) - \overline{p}_{U+V}(\mathbf{w})| \leq \alpha(\mathbf{w}) \left(\frac{1+\epsilon}{1-\epsilon}\right)^2 8\epsilon. \tag{104}$$

Rearranging (97), and observing that $\sum_{\mathbf{w}\in\Lambda} p_{U+V}(\mathbf{w}) = 1$, we have

$$\left(\frac{1-\epsilon}{1+\epsilon}\right)^2 \leq \sum_{\mathbf{w}\in\Lambda} \alpha(\mathbf{w}) \leq \left(\frac{1+\epsilon}{1-\epsilon}\right). \tag{105}$$

Combining (103) and (104), and summing over $\mathbf{w}$,

$$\mathbb{V}(p_{U+V}, p_{U+V|\mathbf{x}}) \leq \sum_{\mathbf{w}\in\Lambda} \alpha(\mathbf{w}) \max\left\{ \left(\frac{1-\epsilon}{1+\epsilon}\right) 64\epsilon, \left(\frac{1+\epsilon}{1-\epsilon}\right)^2 8\epsilon \right\},$$

and using (105) to bound $\sum_{\mathbf{w}\in\Lambda} \alpha(\mathbf{w})$ from above, we get

$$\mathbb{V}(p_{U+V}, p_{U+V|\mathbf{x}}) \leq \max\left\{ 64\epsilon, \left(\frac{1+\epsilon}{1-\epsilon}\right)^3 8\epsilon \right\} \leq \max\left\{64\epsilon, 27 \times 8\epsilon\right\},$$

since $\epsilon \leq 1/2$. Therefore,

$$\mathbb{V}(p_{U+V}, p_{U+V|\mathbf{x}}) \leq 216\epsilon,$$

thereby completing the proof. □

Appendix G: Proof of Lemma 20

The average transmit power per dimension is given by

$$\frac{1}{d}\mathbb{E}\|\mathbf{U}\|^2 = \frac{1}{d}\mathbb{E}\|\mathbf{V}\|^2 = \frac{1}{d}\sum_{\lambda\in\Lambda^{(d)}} \|\lambda\|^2 p_U(\lambda)$$

Using inequalities (92) and (93) from Appendix F, we can bound the average transmit power in terms of $\epsilon^{(d)} := \epsilon_{\Lambda_0^{(d)}}(\sqrt{\mathcal{P}/2})$ as follows:

$$\left(\frac{1-\epsilon^{(d)}}{1+\epsilon^{(d)}}\right)\frac{1}{d}\sum_{\lambda\in\Lambda^{(d)}} \|\lambda\|^2 \frac{g_{\sqrt{\mathcal{P}}}(\lambda)}{g_{\sqrt{\mathcal{P}}}(\Lambda^{(d)})} \le \frac{1}{d}\mathbb{E}\|\mathbf{U}\|^2 \le \left(\frac{1+\epsilon^{(d)}}{1-\epsilon^{(d)}}\right)\frac{1}{d}\sum_{\lambda\in\Lambda^{(d)}} \|\lambda\|^2 \frac{g_{\sqrt{\mathcal{P}}}(\lambda)}{g_{\sqrt{\mathcal{P}}}(\Lambda^{(d)})}. \tag{106}$$

Now, Lemma 6 from [20] shows that as $d\to\infty$, the term $\frac{1}{d}\sum_{\lambda\in\Lambda^{(d)}} \|\lambda\|^2 \frac{g_{\sqrt{\mathcal{P}}}(\lambda)}{g_{\sqrt{\mathcal{P}}}(\Lambda^{(d)})}$ converges to $\mathcal{P}$, provided $\epsilon_{\Lambda^{(d)}}(\sqrt{\mathcal{P}}/2)\to 0$. Therefore, if both $\epsilon^{(d)}$ and $\epsilon_{\Lambda^{(d)}}(\sqrt{\mathcal{P}}/2)$ tend to 0 as $d\to\infty$, then the average transmit power per dimension converges to $\mathcal{P}$. □

Appendix H: Proof of Lemma 23

Recall that for $t\in\{1,2,\ldots,N\}$, $\Theta_{k,t}$ was defined to be the set $\{J_k, J_{k-1}, W_k[2m+k-1] : 1\le m\le t\}$. Making repeated use of the chain rule of mutual information (see e.g., [7]), we see that

$$\begin{aligned}
\mathcal{I}(X_1,\ldots,X_N;\Theta_{k,N}) &= \sum_{t=1}^{N}\mathcal{I}(X_t;\Theta_{k,N}\big|X_1,\ldots,X_{t-1}) \\
&= \sum_{t=1}^{N}\left[\mathcal{I}(X_t;J_k,J_{k-1}\big|X_1,\ldots,X_{t-1}) + \sum_{n=1}^{N}\mathcal{I}(X_t;W_k[2n+k-1]\big|X_1,\ldots,X_{t-1},\Theta_{k,n-1})\right] \\
&= \sum_{t=1}^{N}\sum_{n=1}^{N}\mathcal{I}(X_t;W_k[2n+k-1]\big|X_1,\ldots,X_{t-1},\Theta_{k,n-1}),
\end{aligned} \tag{107}$$

where the last step follows from the fact that $\mathcal{I}(X_t;J_k,J_{k-1}\big|X_1,\ldots,X_{t-1}) = 0$ for $1\le t\le N$, since the messages and the jamming signals are independent. We first show that conditioned on $X_1,\ldots,X_{t-1}$, the message X_t depends only on the vector received in the $(2t+k-1)$st phase, $\mathbf{W}_k[2t+k-1]$, and is independent of the rest of the terms in $\Theta_{k,N}$. We do this in two steps. First, we make the observation that when conditioned on the first $t-1$ messages, X_t does not depend on the variables obtained by the kth relay in the first $2t+k-2$ phases. We then prove that when conditioned on $X_1, X_2\ldots,X_{t-1}$, the message X_t is independent of everything received after the $(2t+k-1)$st phase. In proving the following propositions, we make use of the fact that if X, Y, and Z are random variables distributed over a finite group $\mathbb{G}$, with X being uniformly distributed over $\mathbb{G}$ and independent of (Y,Z), then $X\oplus Y$ is uniformly distributed over $\mathbb{G}$ and independent of Z.

Proposition 31. *Let $1\le t\le N$, and $n,l\in\{1,2,\ldots,t\}$. Then, the message X_t is conditionally independent of $\Theta_{k,n-1}$ given $X_1, X_2,\ldots,X_{l-1}$.*

Proof: From (58), for any $1\le m\le N$, we have $\mathbf{W}_k[2m+k-1] = \mathbf{V}_{k-1}[2m+k-1] + \mathbf{V}_{k+1}[2m+k-1]$. From (57), we know that $\mathbf{V}_{k-1}[2m+k-1] = \mathcal{E}\big((\oplus_{p=1}^{m}X_p)\oplus J_{k+m-1}\big)$, so that $\mathbf{V}_{k-1}[2m+k-1]$ is a

function of $X_1, \ldots, X_m$ and J_{k+m-1}. Similarly, $\mathbf{V}_{k+1}[2m+k-1]$, is a function of $X_1, \ldots, X_{m-1}$ and J_{k+m}. Therefore, for $n \in \{1, 2, \ldots, N\}$, $\Theta_{k,n-1}$ consists of random variables which are all functions of $X_1, \ldots, X_{n-1}$ and $J_{k-1}, \ldots, J_{k+n-1}$, which are all independent of X_t for $n \leq t$ (even when conditioned on the first $l-1 < t$ messages). Therefore, for all $n, l \leq t$, $\Theta_{k,n-1}$ and X_t are conditionally independent given $X_1, \ldots, X_{l-1}$. ■

The following proposition says that for $n > t$, $\mathbf{W}_k[2n+k-1]$ gives no information about the message X_t even when conditioned on the information obtained in the past.

Proposition 32. *Let $1 \leq t < n \leq N$. The vector $\mathbf{W}_k[2n+k-1]$ received by the kth relay in the $(2n+k-1)$st phase is independent of $X_1, \ldots, X_t$ and $\Theta_{k,n-1}$.*

Proof: From Proposition 31 and the fact that the messages are independent, we have for all $t < n$, that X_n is independent of $\Theta_{k,n-1}$ and $X_1, \ldots, X_t$. Furthermore, since X_n is uniformly distributed over $\mathbb{G}^{(d)}$, the term $(\oplus_{p=1}^{n} X_p \oplus J_{k+n-1})$, and hence, $\mathbf{V}_{k-1}[2n+k-1]$ is independent of $\Theta_{k,n-1}$ and $X_1, \ldots, X_t$.

A similar observation that J_{k+n} is independent of $\Theta_{k,n-1}$ and $X_1, \ldots, X_t$, tells us that $\oplus_{p=1}^{n-1} X_p \oplus J_{k+n}$, and hence, $\mathbf{V}_{k+1}[2n+k-1]$ is also independent of $\Theta_{k,n-1}$ and $X_1, \ldots, X_t$. Therefore, $\mathbf{W}_k[2n+k-1]$ is independent of $\Theta_{k,n-1}$ and $X_1, \ldots, X_t$. ■

We now evaluate the terms in (107). We can write

$$\mathcal{I}(X_t; W_k[2n+k-1] \big| X_1, \ldots, X_{t-1}, \Theta_{k,n-1}) = \mathcal{H}(X_t \big| X_1, \ldots, X_{t-1}, \Theta_{k,n-1}) - \mathcal{H}(X_t \big| X_1, \ldots, X_{t-1}, \Theta_{k,n}).$$

From Proposition 31, we have for all $t > n$, $\mathcal{H}(X_t \big| X_1, \ldots, X_{t-1}, \Theta_{k,n}) = \mathcal{H}(X_t \big| X_1, \ldots, X_{t-1}, \Theta_{k,n-1}) = \mathcal{H}(X_t)$. Therefore,

$$\mathcal{I}(X_t; W_k[2n+k-1] \big| X_1, \ldots, X_{t-1}, \Theta_{k,n-1}) = 0 \tag{108}$$

for all $1 \leq n < t \leq N$. Using Proposition 32, for all $n > t$, $\mathcal{H}(W_k[2n+k-1] \big| X_1, \ldots, X_{t-1}, \Theta_{k,n-1}) = \mathcal{H}(W_k[2n+k-1] \big| X_1, \ldots, X_t, \Theta_{k,n-1}) = \mathcal{H}(W_k[2n+k-1])$, and hence

$$\mathcal{I}(X_t; W_k[2n+k-1] \big| X_1, \ldots, X_{t-1}, \Theta_{k,n-1}) = 0 \tag{109}$$

for all $1 \leq t < n \leq N$. Therefore, (107) reduces to

$$\mathcal{I}(X_1, \ldots, X_N; \Theta_{k,N}) = \sum_{t=1}^{N} \mathcal{I}(X_t; W_k[2t+k-1] \big| X_1, \ldots, X_{t-1}, \Theta_{k,t-1}). \tag{110}$$

The mutual information $\mathcal{I}(X_t; W_k[2t+k-1] \big| X_1, \ldots, X_{t-1}, \Theta_{k,t-1})$ can be written in terms of conditional entropies as

$$\begin{aligned} \mathcal{I}(X_t; W_k[2t+k-1] \big| X_1, \ldots, X_{t-1}, \Theta_{k,t-1}) &= \mathcal{H}(\mathbf{W}_k[2t+k+1] \big| X_1, \ldots, X_{t-1}, \Theta_{k,t-1}) \\ &\quad - \mathcal{H}(\mathbf{W}_k[2t+k+1] \big| X_1, \ldots, X_t, \Theta_{k,t-1}). \end{aligned} \tag{111}$$

Let us evaluate each of the terms on the right hand side. Consider the second term,

$$\begin{aligned} \mathcal{H}(W_k[2t+k+1] \big| X_1, \ldots, X_t, \Theta_{k,t-1}) &\geq \mathcal{H}(\mathbf{W}_k[2t+k+1] \big| X_1, \ldots, X_t, \oplus_{p=1}^{t} X_p \oplus J_{k+t-1}, \Theta_{k,t-1}) \\ &= \mathcal{H}(\mathbf{W}_k[2t+k+1] \big| \oplus_{p=1}^{t} X_p \oplus J_{k+t-1}) \end{aligned} \tag{112}$$

The first step is true because conditioning reduces entropy. The second step requires more justification. We have $\mathbf{W}_k[2t+k-1] = \mathbf{V}_{k-1}[2t+k-1] + \mathbf{V}_{k+1}[2t+k-1]$, where $\mathbf{V}_{k-1}[2t+k-1] = \mathcal{E}(\oplus_{p=1}^{t} X_p \oplus J_{k+t-1})$, and $\mathbf{V}_{k+1}[2t+k-1] = \mathcal{E}(\oplus_{p=1}^{t-1} X_p \oplus J_{k+t})$. Given $\oplus_{p=1}^{t} X_p \oplus J_{k+t-1}$, the term $\mathbf{V}_{k-1}[2t+k-1]$ is independent of $X_1, \ldots, X_t, \Theta_{k,t-1}$. The jamming signal, J_{k+t} is independent of $\Theta_{k,t-1}$ (since all the terms in $\Theta_{k,t-1}$ depend only on the first $k+t-1$ jamming signals), all the first t messages, and $\oplus_{p=1}^{t} X_p \oplus J_{k+t-1}$. Hence, $\mathbf{V}_{k+1}[2t+k-1]$ is also independent of $\Theta_{k,t-1}$, the first t messages and $\oplus_{p=1}^{t} X_p \oplus J_{k+t-1}$, thus justifying (112).

Define $X := \oplus_{p=1}^{t} X_p \oplus J_{k+t-1}$, and $Y := \oplus_{p=1}^{t-1} X_p \oplus J_{k+t}$. Recall that $\mathcal{E}(X)$ denotes the encoded form of X, i.e., $\mathcal{E}(X)$ is a random variable distributed over $\Lambda^{(d)}$ according to (40). Then, we have,

$$\mathcal{H}(W_k[2t+k+1] \big| X_1, \ldots, X_t, \Theta_{k,t-1}) \geq \mathcal{H}(\mathcal{E}(X) + \mathcal{E}(Y) \big| X),$$

From Proposition 32, the first term of (111), $\mathcal{H}(W_k[2t+k+1] \big| X_1, \ldots, X_{t-1}, \Theta_{k,t-1}) = \mathcal{H}(W_k[2t+k+1])$. This, in turn, is equal to $\mathcal{H}\left(\mathcal{E}(X) + \mathcal{E}(Y)\right)$. Therefore, $\mathcal{I}(X_t; W_k[2t+k-1] \big| X_1, \ldots, X_{t-1}, \Theta_{k,t-1})$ is bounded above by $\mathcal{I}(X; \mathcal{E}(X) + \mathcal{E}(Y))$, and the random variables X and Y are independent and uniformly distributed over $\mathbb{G}^{(d)}$. From Theorem 19 and Lemma 18, $\mathcal{I}(X; \mathcal{E}(X) + \mathcal{E}(Y))$ is upper bounded by $\epsilon^{(d)} \left(\log_2 |\mathbb{G}^{(d)}| - \log_2 \epsilon^{(d)}\right)$. Substituting this in (110), the lemma follows. □

References

[1] S. Agrawal and S. Vishwanath, "On the secrecy rate of interference networks using structured codes," *Proc. 2009 IEEE Int. Symp. Information Theory*, Seoul, Korea, pp. 2091–2095.

[2] I.-J. Baik and S.-Y. Chung, "Network coding for two-way relay channels using lattices," in *Proc. IEEE Int. Conf. Communications*, Beijing, China, 2008, pp. 3898–3902.

[3] A. Barvinok, *Math 669: Combinatorics, Geometry and Complexity of Integer Points*. [Online]. Available: http://www.math.lsa. umich.edu/ barvinok/latticenotes669.pdf .

[4] J.-C. Belfiore and F. Oggier, "Secrecy gain: A wiretap lattice code design," in *Proc. 2010 Int. Symp. Information Theory and Its Applications*, Taichung, Taiwan, pp. 174–178.

[5] J.-C. Belfiore, "Lattice codes for the compute-and-forward protocol: The flatness factor," in *Proc. 2011 Information Theory Workshop*, Paraty, Brazil, pp. 1–4.

[6] J.H. Conway and N.J. Sloane, *Sphere Packings, Lattices and Groups*, New York: Springer-Verlag, 1988.

[7] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley-Interscience, 1996.

[8] R. Cramer, Y. Dodis, S. Fehr, C. Padro, and D. Wichs, "Detection of algebraic manipulation with applications to robust secret sharing and fuzzy extractors," *Adv. Cryptology*, vol. 4965, pp. 471–488, 2008.

[9] I. Csiszár and P. Narayan, "Secrecy capacities for multiple terminals," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3047–3061, Dec. 2004.

[10] W. Ehm, T. Gneiting, and D. Richards, "Convolution roots of radial positive definite functions with compact support," *Trans. AMS*, vol. 356, no. 11, pp. 4655–4685, May 2004.

[11] U. Erez and R. Zamir, "Error exponents of modulo additive noise channels with side information at the transmitter," *IEEE Trans. Inf. Theory*, vol. 47, pp. 210–218, Jan. 2001.

[12] U. Erez and R. Zamir, "Achieving 1/2log(1+SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[13] U. Erez, S. Litsyn, and R. Zamir, "Lattices which are good for (almost) everything," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3401–3416, Oct. 2005.

[14] A. Elbert and A. Laforgia, "An asymptotic relation for the zeros of Bessel functions," *J. Math. Analysis and Applications*, vol. 98, no. 2, pp. 502–510, 1984.

[15] W. Feller, *An Introduction to Probability Theory and Its Applications, Vol. 2*, 2nd ed. New York: Wiley, 1971.

[16] X. He and A. Yener, "Providing secrecy with lattice codes," *Proc. 46th Annual Allerton Conf. on Communication, Control, and Computing*, Monticello, IL, 2008, pp. 1199–1206.

[17] X. He and A. Yener, "Strong secrecy and reliable Byzantine detection in the presence of an untrusted relay," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, Jul. 2013.

[18] I.N. Herstein, *Topics in Algebra*, 2nd ed. New York: Wiley, 1975.

[19] N. Kashyap, V. Shashank, and A. Thangaraj, "Secure computation in a bidirectional relay," in *Proc. 2012 IEEE Int. Symp. Information Theory*, Cambridge, MA, pp. 1162–1166.

[20] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehlé, "Semantically secure lattice codes for the Gaussian wiretap channel," submitted for publication. [Online]. Available: http://arxiv.org/abs/1210.6673.

[21] E. Lukacs, *Characteristic Functions*, 2nd ed. London, U.K.: Griffin, 1970.

[22] U. Maurer and S. Wolf. "Information-theoretic key agreement: From weak to strong secrecy for free," *Proc. EUROCRYPT–2000 on Advances in Cryptology*, vol. 1807, pp. 351–368, Springer, 2000.

[23] 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.

[24] B. Nazer and M. Gastpar, "Reliable physical layer network coding," *Proc. of the IEEE*, vol. 99, no. 3, pp. 438–460, Mar. 2011.

[25] S. Nitinawarat and P. Narayan, "Secret key generation for correlated Gaussian sources," *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3373–3391, Jun. 2012.

[26] F. Oggier, P. Solé, and J.-C. Belfiore, "Lattice codes for the wiretap Gaussian channel: construction and analysis," submitted for publication. [Online]. Available: http://arxiv.org/abs/1103.4086.

[27] G. Poltyrev, "On coding without restrictions for the AWGN channel," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 409–417, Mar. 1994.

[28] P. Popovski and H. Yomo, "Physical network coding in two-way wireless relay channels," in *Proc. IEEE Int. Conf. Communications*, Glasgow, Scotland, 2007, pp. 707–712.

[29] S. Ramanujan, "A proof of Bertrand's postulate," *J. Indian Math. Soc.*, vol. 11, pp. 181–182, 1919.

[30] R.M. Roth, *Introduction to Coding Theory*, Cambridge, U.K.: Cambridge University Press, 2006.

[31] H. Rubin and T.M. Sellke, "Zeroes of infinitely differentiable characteristic functions," in *A Festschrift for Herman Rubin*, Anirban DasGupta, ed., Institute of Mathematical Statistics Lecture Notes – Monograph Series, vol. 45, pp. 164–170, 2004.

[32] E.M. Stein and G.L. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton, NJ: Princeton Univ. Press, 1971.

[33] F.G. Tricomi, "Sulle funzioni di Bessel di ordine e argomento pressoché uguali," *Atti Accad. Sci. Torino Cl. Sci. Fis. Mat. Natur.*, vol. 83, pp. 3–20, 1949.

[34] M. Wilson, K. Narayanan, H. Pfister, and A. Sprintson, "Joint physical layer coding and network coding for bidirectional relaying," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5641–5654, Nov. 2010.

[35] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975

[36] S. Zhang and S.-C. Liew, "Channel coding and decoding in a relay system operated with physical-layer network coding," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 5, pp. 788–796, Jun. 2009.

[37] R. Zhang, L. Song, Z. Han, B. Jiao, and M. Debbah, "Physical layer security for two way relay communications with friendly jammers," *Proc. IEEE Global Communications Conf.*, Miami, FL, 2010, pp. 1–6.