

Semantically Secure Lattice Codes for Compound MIMO Channels

Antonio Campello, Cong Ling and Jean-Claude Belfiore

Abstract—We consider compound multi-input multi-output (MIMO) wiretap channels where minimal channel state information at the transmitter (CSIT) is assumed. Code construction is given for the special case of isotropic mutual information, which serves as a conservative strategy for general cases. Using the flatness factor for MIMO channels, we propose lattice codes universally achieving the secrecy capacity of compound MIMO wiretap channels up to a constant gap (measured in nats) that is equal to the number of transmit antennas. The proposed approach improves upon existing works on secrecy coding for MIMO wiretap channels from an error probability perspective, and establishes information theoretic security (in fact semantic security). We also give an algebraic construction to reduce the code design complexity, as well as the decoding complexity of the legitimate receiver. Thanks to the algebraic structures of number fields and division algebras, our code construction for compound MIMO wiretap channels can be reduced to that for Gaussian wiretap channels, up to some additional gap to secrecy capacity.

I. INTRODUCTION

Due to the open nature of the wireless medium, wireless communications are inherently vulnerable to eavesdropping attacks. Information theoretic security offers additional protection for wireless data, since it only relies on the physical properties of wireless channels, thus representing a competitive/complementary approach to security compared to traditional cryptography.

The fundamental wiretap channel model was first introduced by Wyner [1]. In this seminal paper, Wyner defined the secrecy capacity and presented the idea of coset coding to encode both data and random bits to mitigate eavesdropping. In recent years, the quest for the secrecy capacity of many classes of channels has been one of the central topics in wireless communications [2], [3], [4], [5], [6], [7], [8].

In the information theory community, a commonly used secrecy notion is *strong secrecy*: the mutual information $\mathbb{I}(M; Z^T)$ between the confidential message M and the channel output Z^T should vanish when the code length $T \rightarrow \infty$.

This work was presented in part at the International Zurich Seminar on Communications (IZS) 2018 and in part at the International Symposium on Turbo Codes and Iterative Information Processing (ISTC) 2016.

A. Campello is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: a.campello@imperial.ac.uk).

C. Ling is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: cling@iee.org).

J.-C. Belfiore is with the Mathematical and Algorithmic Sciences Lab, France Research Center, Huawei Technologies (e-mail: belfiore@telecom-paristech.fr).

This common assumption of uniformly distributed messages was relaxed in [9], which considered the concept of *semantic security*: for *any* message distribution, the advantage obtained by an eavesdropper from its received signal vanishes for large block lengths. This notion is motivated by the fact that the plaintext can be fixed and arbitrary.

For the Gaussian wiretap channel, [10] introduced the *secrecy gain* of lattice codes while [11] proposed semantically secure lattice codes based on the lattice Gaussian distribution. To obtain semantic security, the *flatness factor* of a lattice was introduced in [11] as a fundamental criterion which implies that conditional outputs are indistinguishable for different input messages. Using a random coding argument, it was shown that there exist families of lattice codes which are *good for secrecy*, meaning that their flatness factor vanishes. Such families achieve semantic security for rates up to $1/2$ nat from the secrecy capacity.

Compared to the Gaussian wiretap channel, the cases of fading and multi-input multi-output (MIMO) wiretap channels are more technically challenging. The fundamental limits of fading wireless channels with secrecy constraints have been investigated in [12], [13], [2], where the achievable rates and the secrecy outage probability were given. The secrecy capacity of the MIMO wiretap channel was derived in [14], [15], [16], [17], assuming full channel state information at the transmitter (CSIT). A code design in this setting was given in [18] by reducing to scalar Gaussian codes. Although CSIT is sometimes available for the legitimate channel, it is hardly possible that it would be available for the eavesdropping channel. An achievability result was given in [19] for varying MIMO wiretap channels with no CSI about the wiretapper, under the condition that the wiretapper has less antennas than the legitimate receiver. Schaefer and Loyka [20] studied the secrecy capacity of the *compound* MIMO wiretap channel, where a transmitter has no knowledge of the realization of the eavesdropping channel, except that it remains fixed during the transmission block and belongs to a given set (the *compound set*). The compound model represents a well-accepted reasonable approach to information theoretic security, which assumes minimal CSIT of the eavesdropping channel [21], [22], [23]. It can also model a multicast channel with several eavesdroppers, where the transmitter sends information to all legitimate receivers while keeping it secret from all eavesdroppers [21].

When it comes to code design for fading and MIMO wiretap channels, an error probability criterion was used in several prior works [24], [25], [26], while information theoretic security was only addressed recently with the help of flatness

factors [27], [28]. In particular, [28] established strong secrecy over MIMO wiretap channels for secrecy rates that are within a constant gap from the secrecy capacity.

A. Main Contributions

In this paper, we propose universal codes for compound Gaussian MIMO wiretap channels that complement the recent work reported in [28]. The key method is discrete Gaussian shaping and a “direct” proof of the universal flatness of the eavesdropper’s lattice. This method is similar to that used in [29] to approach the capacity of compound MIMO channels so that the present paper can be considered a companion paper of [29] for wiretap channels. Note that [28] used an “indirect” proof, which was based on an upper bound on the smoothing parameter in terms of the minimum distance of the dual lattice. Besides considering different channel models ([28] is focused on ergodic stationary channels although it also briefly addresses compound channels), the code constructions of this paper and [28] are also different: the construction of [28] is based on a particular sequence of algebraic number fields with increasing degrees, while the algebraic construction of this work combines algebraic number fields of fixed degree and random error correcting codes of increasing lengths. The proposed construction enjoys a significantly smaller gap to secrecy capacity, as well as lower decoding complexity, than [28], over compound MIMO wiretap channels.

We focus on a compound channel formed by the set of all matrices with the same white-input capacity (see (3) for the precise model). Our lattice coding scheme universally achieves rates (in nats) up to $(C_b - C_e - n_a)^+$, where C_b is the capacity of the legitimate channel, C_e is the capacity of the eavesdropper channel, n_a is the number of transmit antennas and $(x)^+ = \max\{x, 0\}$. We believe the n_a -nat gap is an artifact of our proof technique based on the flatness factor, which may be removed by improving the flatness-factor method. This is left as an open problem for future research.

For this special compound model, we also show how to extend the analysis in order to accommodate number-of-antenna mismatch, *i.e.*, security is valid *regardless* of the number of antennas at the eavesdropper¹. This is a very appealing property, since the number of receive antennas of an eavesdropper may be unknown to the transmitter.

We present two techniques to prove universality of the proposed lattice codes. The first technique is based on Construction A (see Sect. V-A for the definition) and the usual argument for compound channels [30], [31], which combines fine quantization of the channel space with mismatch encoding for quantized states. This method is a generic proof of the existence of good codes which potentially incurs large blocklengths and performance loss. The second technique is based on algebraic lattices and assumes that the codes admit an “algebraic reduction” and can absorb the channel state. In fact, any code which is good for the *Gaussian* wiretap channel can be coupled with this second technique, as long as it also possesses an additional algebraic structure (for precise terms

see Definition 6). It is inspired by previous works on algebraic reduction for fading and MIMO channels [32], [33], which are revisited here in terms of secrecy.

B. Relation to Previous Works

An idea of approaching the secrecy capacity of fading wiretap channels using nested lattice codes was outlined in [34]. Code construction for compound wiretap channels has been further developed in [35], which leads to the current work where proof details are given.

The technique for establishing universality of the codes in [20] over the compound MIMO channel with (uncountably) infinite uncertainty sets consists of quantizing the channel space and designing a (random Gaussian) codebook for the quantized channels. This method is similar to the proof of Theorem 1 in the present paper.

Compound MIMO channels *without* secrecy constraints have been considered earlier in [30], [31], [36] for random codebooks. Lattice codes are shown to achieve the optimal diversity-multiplexing tradeoff for MIMO channels in [37]. More recently it was proven that precoded integer forcing [38] achieves the compound capacity up to a gap, while algebraic lattice codes [29] achieve the compound capacity with ML decoding and a gap to the compound capacity of MIMO channels with reduced decoding complexity. As mentioned above, some techniques (generalized Construction A and channel quantization) of this paper are similar to those used in [29].

C. Organization

The technical content of this paper is organized as follows. In Section II we discuss the main problem and notions of security. In Section III, we introduce the main notation on lattices and discrete Gaussians, stating generalized versions of known results for correlated Gaussian distributions. In Section IV we give an overview of the main coding scheme and analyze the information leakage and reliability. The proof of universality, however, is postponed until Section V, where we show that lattice codes can achieve vanishing information leakage under semantic security through the two aforementioned techniques. Section VI concludes the paper with a discussion of other compound models and future work.

D. Notation

Matrices and column vectors are denoted by upper and lowercase boldface letters, respectively. For a matrix \mathbf{A} , its Hermitian transpose, inverse, determinant and trace are denoted by \mathbf{A}^\dagger , \mathbf{A}^{-1} , $|\mathbf{A}|$ and $\text{tr}(\mathbf{A})$, respectively. We denote the Frobenius norm of a matrix by $\|\mathbf{A}\|_F \triangleq \sqrt{\text{tr}(\mathbf{A}^\dagger \mathbf{A})}$ and the spectral norm (*i.e.*, 2-norm) by $\|\mathbf{A}\| \triangleq \sqrt{\lambda_1}$, where λ_1 is the largest eigenvalue of $\mathbf{A}^\dagger \mathbf{A}$. \mathbf{I} denotes the identity matrix. We write $\mathbf{A} \succeq \mathbf{0}$ for a symmetric matrix \mathbf{A} if it is positive semi-definite. Similarly, we write $\mathbf{A} \succeq \mathbf{B}$ if $(\mathbf{A} - \mathbf{B}) \succeq \mathbf{0}$. We use the standard asymptotic notation $f(x) = O(g(x))$ when $\limsup_{x \rightarrow \infty} |f(x)|/g(x) < \infty$, $f(x) = o(g(x))$ when $\lim_{x \rightarrow \infty} f(x)/g(x) = 0$, $f(x) = \Omega(g(x))$ when

¹Previous works [24], [28] required that the number of the eavesdropper’s antennas be greater than or equal to n_a .

$\liminf_{x \rightarrow \infty} f(x)/g(x) > 0$, and $f(x) = \omega(g(x))$ when $\lim_{x \rightarrow \infty} f(x)/g(x) = \infty$. Finally, in this paper, the logarithm is taken with respect to base e (where e is the Neper number) and information is measured in nats.

II. PROBLEM STATEMENT

Consider the following wiretap model. A transmitter (Alice) sends information through a MIMO channel to a legitimate receiver (Bob) and is eavesdropped by an illegitimate user (Eve). The channel equations for Bob and Eve read:

$$\begin{aligned} \underbrace{\mathbf{Y}_b}_{n_b \times T} &= \underbrace{\mathbf{H}_b}_{n_b \times n_a} \underbrace{\mathbf{X}}_{n_a \times T} + \underbrace{\mathbf{W}_b}_{n_b \times T} \\ \underbrace{\mathbf{Y}_e}_{n_e \times T} &= \underbrace{\mathbf{H}_e}_{n_e \times n_a} \underbrace{\mathbf{X}}_{n_a \times T} + \underbrace{\mathbf{W}_e}_{n_e \times T}, \end{aligned} \quad (1)$$

where n_a is the number of transmit antennas, n_b (n_e , resp.) is the number of receive antennas for Bob (Eve, resp.), T is the coherence time, and \mathbf{W}_b (\mathbf{W}_e , resp.) has circularly symmetric complex Gaussian i.i.d. entries with variance σ_b^2 (σ_e^2 , resp.) per complex dimension. We can vectorize (1) in a natural way:

$$\begin{aligned} \underbrace{\mathbf{y}_b}_{n_b T \times 1} &= \underbrace{\mathcal{H}_b}_{n_b T \times n_a T} \underbrace{\mathbf{x}}_{n_a T \times 1} + \underbrace{\mathbf{w}_b}_{n_b T \times 1} \\ \underbrace{\mathbf{y}_e}_{n_e T \times 1} &= \underbrace{\mathcal{H}_e}_{n_e T \times n_a T} \underbrace{\mathbf{x}}_{n_a T \times 1} + \underbrace{\mathbf{w}_e}_{n_e T \times 1}, \end{aligned} \quad (2)$$

where \mathcal{H}_b and \mathcal{H}_e are the block diagonal matrices

$$\begin{aligned} \mathcal{H}_b &= \mathbf{I}_T \otimes \mathbf{H}_b = \begin{pmatrix} \mathbf{H}_b & & & \\ & \mathbf{H}_b & & \\ & & \ddots & \\ & & & \mathbf{H}_b \end{pmatrix}, \\ \mathcal{H}_e &= \mathbf{I}_T \otimes \mathbf{H}_e = \begin{pmatrix} \mathbf{H}_e & & & \\ & \mathbf{H}_e & & \\ & & \ddots & \\ & & & \mathbf{H}_e \end{pmatrix}. \end{aligned}$$

For convenience, we denote the transmit signal-to-noise ratio (SNR) in Bob and Eve's channels by

$$\rho_b \triangleq \frac{P}{\sigma_b^2} \text{ and } \rho_e \triangleq \frac{P}{\sigma_e^2},$$

respectively, where P is the power constraint, i.e., the transmitted signal satisfies $\mathbb{E}[\mathbf{x}^\dagger \mathbf{x}] \leq n_a T P$.

We assume that the channel realizations ($\mathbf{H}_b, \mathbf{H}_e$) are unknown to Alice but belong to a compound set $\mathcal{S} = \mathcal{S}_b \times \mathcal{S}_e \in \mathbb{C}^{n_b \times n_a} \times \mathbb{C}^{n_e \times n_a}$. From the security perspective, we further make the conservative assumption that Eve knows both \mathbf{H}_b and \mathbf{H}_e . Under this general scenario the (strong) secrecy capacity is bounded by [20]:

$$C_s \geq \max_{\mathbf{R}} \min_{\mathbf{H}_b, \mathbf{H}_e} \left(\log |\mathbf{I} + \sigma_b^{-1} \mathbf{H}_b^\dagger \mathbf{H}_b \mathbf{R}| - \log |\mathbf{I} + \sigma_e^{-1} \mathbf{H}_e^\dagger \mathbf{H}_e \mathbf{R}| \right) \quad (\text{iv}^*)$$

where the minimum is over all realizations in \mathcal{S} and the maximum over the matrices $\mathbf{R} \succeq 0$ such that $\text{tr}(\mathbf{R}) \leq n_a P$.

Suppose that \mathcal{S}_b and \mathcal{S}_e are the set of channels with the same isotropic mutual information, i.e.,

$$\begin{aligned} \mathcal{S}_b &= \left\{ \mathbf{H}_b \in \mathbb{C}^{n_b \times n_a} : |\mathbf{I} + \rho_b \mathbf{H}_b^\dagger \mathbf{H}_b| = e^{C_b} \right\}, \\ \mathcal{S}_e &= \left\{ \mathbf{H}_e \in \mathbb{C}^{n_e \times n_a} : |\mathbf{I} + \rho_e \mathbf{H}_e^\dagger \mathbf{H}_e| = e^{C_e} \right\}, \end{aligned} \quad (3)$$

for fixed $C_b, C_e \geq 0$. In this case, the bound gives $C_s \geq (C_b - C_e)^+$. The worst case is achieved by taking a specific "isotropic" realization $\mathbf{H}_b^\dagger \mathbf{H}_b = \alpha_b \mathbf{I}$, $\mathbf{H}_e^\dagger \mathbf{H}_e = \alpha_e \mathbf{I}$, where α_b and α_e are such that \mathbf{H}_b and \mathbf{H}_e belong to \mathcal{S}_b and \mathcal{S}_e , respectively. From this we conclude that $C_s = C_b - C_e$. The goal of this paper is to construct universal lattice codes that approach the secrecy capacity C_s with semantic security. As a corollary, the semantic security capacity and the strong secrecy capacity of the compound set $\mathcal{S}_b \times \mathcal{S}_e$ coincide.

A practical motivation to consider the compound model (3) is the following. Firstly, notice that the secrecy capacity is the same if we replace the equality in the definition of \mathcal{S}_b and \mathcal{S}_e with upper/lower bounds; more precisely the secrecy capacity of the channel with compound set $\bar{\mathcal{S}}_e \times \bar{\mathcal{S}}_b$, where

$$\begin{aligned} \bar{\mathcal{S}}_b &= \left\{ \mathbf{H}_b \in \mathbb{C}^{n_b \times n_a} : |\mathbf{I} + \rho_b \mathbf{H}_b^\dagger \mathbf{H}_b| \geq e^{C_b} \right\}, \\ \bar{\mathcal{S}}_e &= \left\{ \mathbf{H}_e \in \mathbb{C}^{n_e \times n_a} : |\mathbf{I} + \rho_e \mathbf{H}_e^\dagger \mathbf{H}_e| \leq e^{C_e} \right\}, \end{aligned} \quad (4)$$

is the same as for $\mathcal{S}_e \times \mathcal{S}_b$. Note that the sets $\mathcal{S}_b, \mathcal{S}_e$ and $\bar{\mathcal{S}}_e$ are compact whereas $\bar{\mathcal{S}}_b$ is not. In other words, universal codes are robust, in the sense that only a lower bound on the legitimate channel capacity and an upper bound on the eavesdropper channel are needed. From the security perspective, this is a safe strategy in the scenario where the capacities are not known precisely. Even if Bob and Eve's channels are random, an acceptable secrecy-outage probability can be guaranteed by setting C_b and C_e properly. Then, the problem still boils down to the design of universal codes for the compound model (3).

A. Notions of Security

A secrecy code for the compound MIMO channel can be formally defined as follows.

Definition 1. An (R, R', T) -secrecy code for a compound MIMO channel with set $\mathcal{S} = \mathcal{S}_b \times \mathcal{S}_e$ consists of

- (i) A set of messages $\mathcal{M}_T = \{1, \dots, e^{TR}\}$ (the secret message rate R is measured in nats and e^{TR} is assumed to be an integer for convenience).
- (ii) An auxiliary (not necessarily uniform) source U taking values in \mathcal{U}_T with entropy $R' = H(U)$.
- (iii) A stochastic encoding function $f_T : \mathcal{M}_T \times \mathcal{U}_T \rightarrow \mathbb{C}^{n_a \times T}$ satisfying the power constraint

$$\frac{1}{T} \text{tr} \left(\mathbb{E} [f_T(m, U)^\dagger f_T(m, U)] \right) \leq n_a P, \quad (5)$$

for any $m \in \mathcal{M}_T$.

- (iv) A decoding function $g_T : \mathcal{S}_b \times \mathbb{R}^{n_b \times T} \rightarrow \mathcal{M}_T$ with output $\hat{m} = g_T(s_b, \mathbf{Y}_b)$.

A pair $(s_b, s_e) \in \mathcal{S}_b \times \mathcal{S}_e$ is referred to as a channel state (or channel realization). To ensure reliability for all channel

states we require a sequence of codes whose error probability for message M vanishes uniformly:

$$\mathbb{P}_{\text{err}|M} \triangleq \mathbb{P}(\hat{M} \neq M) \rightarrow 0, \forall s_b \in \mathcal{S}_b, \text{ as } T \rightarrow \infty. \quad (6)$$

Let p_M be a message distribution over \mathcal{M}_T . For strong secrecy, p_M is usually assumed to be uniform; however, this assumption is not sufficient from the viewpoint of semantic security, which is the standard notion of security in modern cryptography. Let \mathbf{Y}_e be the output of the channel to the eavesdropper, who is omniscient. The following security notions are adapted from [9], [11] and should hold in the limit $T \rightarrow \infty$:

- **Mutual Information Security (MIS)**: Unnormalized mutual information

$$\mathbb{I}(M; \mathbf{Y}_e) \rightarrow 0 \quad (7)$$

for any message distribution p_M and for all $s_e \in \mathcal{S}_e$.

- **Semantic Security (SEMANTICS)**: Adversary's advantage

$$\sup_{f, p_M} \left\{ \max_{m'} \mathbb{P}(f(M) = f(m') | \mathbf{Y}_e) - \max_{m''} \mathbb{P}(f(M) = f(m'')) \right\} \xrightarrow{T \rightarrow \infty} 0$$

for any function f from M to finite sequences of bits in $\{0, 1\}^*$, and all $s_e \in \mathcal{S}_e$.

- **Distinguishing Security (DISTS)**: The maximum variational distance

$$\max_{m', m'' \in \mathcal{M}_T} \mathbb{V}(p_{\mathbf{Y}_e|m'}, p_{\mathbf{Y}_e|m''}) \rightarrow 0 \text{ for all } s_e \in \mathcal{S}_e.$$

We stress that all three notions require a sequence of codes to be *universally* secure for all channel states. Treating these notions as classes, we have the inclusions $\text{MIS} \subseteq \text{SEMANTICS} = \text{DISTS}$, *i.e.*, the sequences of codes satisfying DISTS are the same as the ones satisfying SEMANTICS and also include those satisfying MIS [11, Prop. 1]. Moreover, if in the above notions we require that the convergence rate is $o(1/T)$, the three sets coincide. We thus define universally secure codes as follows.

Definition 2. *A sequence of codes of rate R is universally secure for the MIMO wiretap channel if for all $(s_b, s_e) \in \mathcal{S}$, it satisfies the reliability condition (6) and mutual information security (7) uniformly.*

Then, semantic security follows as a corollary, which is a direct consequence of established relations between MIS and SEMANTICS [9]:

Corollary 1. *The sequence of codes given in Definition 2 is semantically secure for the compound MIMO wiretap channel.*

In what follows we proceed to construct universally secure codes for the MIMO wiretap channel using lattice coset codes.

III. CORRELATED DISCRETE GAUSSIAN DISTRIBUTIONS

In this subsection, we exhibit essential results and concepts for the definition and analysis of our lattice coding scheme.

A. Preliminary Lattice Definitions

A (complex) lattice Λ with generator matrix $\mathbf{B}_c \in \mathbb{C}^{n \times 2n}$ is a discrete additive subgroup of \mathbb{C}^n given by

$$\Lambda = \mathcal{L}(\mathbf{B}_c) = \{ \mathbf{B}_c \mathbf{x} : \mathbf{x} \in \mathbb{Z}^{2n} \}. \quad (8)$$

A complex lattice has an equivalent real lattice generated by the matrix obtained by stacking real and imaginary parts of matrix \mathbf{B}_c :

$$\mathbf{B}_r = \begin{pmatrix} \Re(\mathbf{B}_c) \\ \Im(\mathbf{B}_c) \end{pmatrix} \in \mathbb{R}^{2n \times 2n}.$$

A *fundamental region* $\mathcal{R}(\Lambda)$ for Λ is any interior-disjoint region that tiles \mathbb{C}^n through translates by vectors of Λ . For any $\mathbf{y}, \mathbf{x} \in \mathbb{C}^n$ we say that $\mathbf{y} = \mathbf{x} \pmod{\Lambda}$ iff $\mathbf{y} - \mathbf{x} \in \Lambda$. By convention, we fix a fundamental region and denote by $\mathbf{y} \pmod{\Lambda}$ the unique representative $\mathbf{x} \in \mathcal{R}(\Lambda)$ such that $\mathbf{y} = \mathbf{x} \pmod{\Lambda}$. The volume of Λ is defined as the volume of a fundamental region for the equivalent real lattice, given by $V(\Lambda) = |\mathbf{B}_r|$.

Throughout this text, for convenience, we also use the matrix-notation of lattice points. If $\Lambda \subset \mathbb{C}^{nT}$ is a full-rank lattice, the matrix form representation of $\mathbf{x} = (x_1, \dots, x_{nT}) \in \Lambda$ is

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \cdots & x_T \\ x_{T+1} & x_{T+2} & \cdots & x_{2T} \\ x_{2T+1} & x_{2T+2} & \cdots & x_{3T} \\ \vdots & \vdots & \ddots & \vdots \\ x_{(n-1)T+1} & x_{(n-1)T+2} & \cdots & x_{nT} \end{pmatrix}.$$

The *dual* Λ^* of a complex lattice is defined as

$$\Lambda^* = \{ \mathbf{x} \in \mathbb{C}^n : \Re \langle \mathbf{x}, \mathbf{y} \rangle \in \mathbb{Z} \text{ for all } \mathbf{y} \in \Lambda \}.$$

B. The Flatness Factor

The flatness factor has been introduced in [11], and will be used here to bound the information leakage of information transmission of our coding scheme.

The p.d.f. of the complex Gaussian centered at $\mathbf{c} \in \mathbb{C}^n$ is defined as

$$f_{\sigma, \mathbf{c}}(\mathbf{x}) = \frac{1}{(\pi\sigma^2)^n} e^{-(\mathbf{x}-\mathbf{c})^\dagger (\mathbf{x}-\mathbf{c})/\sigma^2}.$$

We write $f_{\sigma, \Lambda}(\mathbf{x})$ for the sum of $f_{\sigma, \mathbf{c}}(\mathbf{x})$ over $\mathbf{c} \in \Lambda$. The *flatness factor* of a lattice quantifies the distance between $f_{\sigma, \Lambda}(\mathbf{x})$ and the uniform distribution over $\mathcal{R}(\Lambda)$ and, as we will see, bounds the amount of leaked information in a lattice coding scheme.

Definition 3 (Flatness factor for spherical Gaussian distributions). *For a lattice Λ and a parameter σ , the flatness factor is defined by:*

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

where $\mathcal{R}(\Lambda)$ is a fundamental region of Λ .

For a complex lattice $\Lambda \subset \mathbb{C}^n$, let $\gamma_\Lambda(\sigma) = \frac{V(\Lambda)^{\frac{1}{n}}}{\sigma^2}$ be the volume-to-noise ratio (VNR). We recall the formulas of the

flatness factor and smoothing parameter, adapted to complex lattices. The flatness factor can be written as [11, Prop. 2]:

$$\epsilon_{\Lambda}(\sigma) = \left(\frac{\gamma_{\Lambda}(\sigma)}{\pi} \right)^n \Theta_{\Lambda} \left(\frac{1}{\pi\sigma^2} \right) - 1 = \Theta_{\Lambda^*}(\pi\sigma^2) - 1, \quad (9)$$

where Θ_{Λ} is the *theta series* of the lattice Λ .

Definition 4 (Smoothing parameter [39]). *For a lattice Λ and $\varepsilon > 0$, the smoothing parameter is defined by the function $\eta_{\varepsilon}(\Lambda) = \sqrt{2\pi}\sigma$, for the smallest $\sigma > 0$ such that $\sum_{\lambda^* \in \Lambda^* \setminus \{0\}} e^{-\pi^2\sigma^2\|\lambda^*\|^2} \leq \varepsilon$.*

When we have a correlated Gaussian distribution with covariance matrix Σ

$$f_{\sqrt{\Sigma}, \mathbf{c}}(\mathbf{x}) = \frac{1}{\pi^n |\Sigma|} \exp \left\{ -(\mathbf{x} - \mathbf{c})^T \Sigma^{-1} (\mathbf{x} - \mathbf{c}) \right\}, \quad (10)$$

the flatness factor is similarly defined.

Definition 5 (Flatness factor for correlated Gaussian distributions).

$$\epsilon_{\Lambda}(\sqrt{\Sigma}) \triangleq \max_{\mathbf{x} \in \mathcal{R}(\Lambda)} |V(\Lambda) f_{\sqrt{\Sigma}, \Lambda}(\mathbf{x}) - 1|$$

where $\mathcal{R}(\Lambda)$ is a fundamental region of Λ .

The usual smoothing parameter in Definition 4 is a scalar. To extend its definition to matrices, we say $\sqrt{2\pi}\Sigma \succeq \eta_{\varepsilon}(\Lambda)$ if $\epsilon_{\Lambda}(\sqrt{\Sigma}) \leq \varepsilon$. This induces a partial order because $\epsilon_{\Lambda}(\sqrt{\Sigma_1}) \leq \epsilon_{\Lambda}(\sqrt{\Sigma_2})$ if $\Sigma_1 \succeq \Sigma_2$.

When $\mathbf{c} = 0$ we ignore the index and write $f_{\sqrt{\Sigma}, 0}(\mathbf{x}) = f_{\sqrt{\Sigma}}(\mathbf{x})$. For a covariance matrix Σ we define the generalized-volume-to-noise ratio as

$$\gamma_{\Lambda}(\sqrt{\Sigma}) = \frac{V(\Lambda)^{1/n}}{|\Sigma|^{1/n}}.$$

Clearly, the effect of correlation on the flatness factor may be absorbed if we use a new lattice $\frac{\sqrt{\Sigma}}{\sigma} \cdot \Lambda$, i.e., $\epsilon_{\Lambda}(\sigma) = \epsilon_{\frac{\sqrt{\Sigma}}{\sigma} \cdot \Lambda}(\sqrt{\Sigma})$. From this, and from the expression of the flatness factor, we have

$$\begin{aligned} \epsilon_{\Lambda}(\sqrt{\Sigma}) &= \frac{V(\Lambda)}{\pi^n |\Sigma|} \sum_{\lambda \in \Lambda} e^{-\lambda^T \Sigma^{-1} \lambda} - 1 \\ &= \left(\frac{\gamma_{\sqrt{\Sigma^{-1}\Lambda}(\sigma^2)}}{\pi} \right)^n \Theta_{\sqrt{\Sigma^{-1}\Lambda}} \left(\frac{1}{\pi\sigma^2} \right) - 1. \end{aligned}$$

In our applications, the matrix Σ will be determined by the channel realization (1). Figure 1 shows the effect of fading on the lattice Gaussian function. A function (10) which is flat over the Gaussian channel (corresponding to $\Sigma = \mathbf{I}$) need not be flat for a channel in deep fading (corresponding to an ill-conditioned Σ), in which case an eavesdropper could clearly distinguish one dimension of the signal.

C. The Discrete Gaussian Distribution

In order to define our coding scheme, we need a last element, which is the distribution of the sent signals. To this end, we define the *discrete Gaussian distribution* $\mathcal{D}_{\Lambda+\mathbf{c}, \sqrt{\Sigma}}$

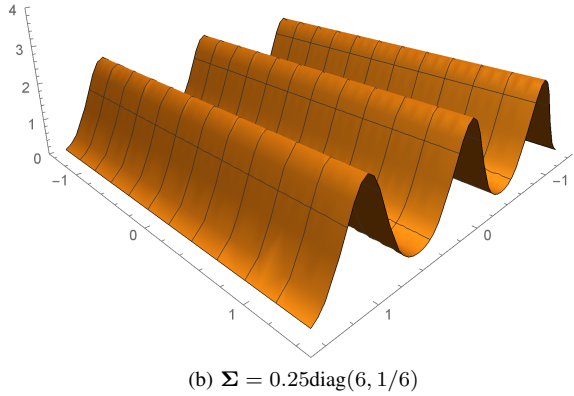
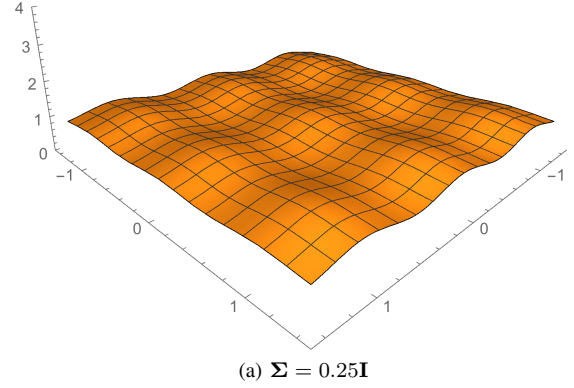


Fig. 1: Illustration of the periodic Gaussian function for the lattice \mathbb{Z}^2 and different covariance matrices with same determinant.

as the distribution assuming values on $\Lambda + \mathbf{c}$, such that the probability of each point $\lambda + \mathbf{c}$ is given by

$$\mathcal{D}_{\Lambda+\mathbf{c}, \sqrt{\Sigma}}(\lambda + \mathbf{c}) = \frac{f_{\sqrt{\Sigma}}(\lambda + \mathbf{c})}{f_{\sqrt{\Sigma}, \Lambda}(\mathbf{c})}.$$

Its relation to the continuous Gaussian distribution can be shown via the smoothing parameter or the flatness factor. For instance, a vanishing flatness factor guarantees that the power per dimension of $\mathcal{D}_{\Lambda+\mathbf{c}, \sigma\mathbf{I}}$ is approximately σ^2 [11, Lemma 6].

The next proposition says that the sum of a continuous Gaussian and a discrete Gaussian is approximately a continuous Gaussian, provided that the flatness factor is small. The proof can be found in [28, Appendix I-A]:

Lemma 1. *Given \mathbf{x}_1 sampled from the discrete Gaussian distribution $\mathcal{D}_{\Lambda+\mathbf{c}, \sqrt{\Sigma_1}}$ and \mathbf{x}_2 sampled from the continuous Gaussian distribution $f_{\sqrt{\Sigma_2}}$. Let $\Sigma_0 = \Sigma_1 + \Sigma_2$ and let $\Sigma_3^{-1} = \Sigma_1^{-1} + \Sigma_2^{-1}$. If $\sqrt{\Sigma_3} \succeq \eta_{\varepsilon}(\Lambda)$ for $\varepsilon \leq \frac{1}{2}$, then the distribution g of $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ is close to $f_{\sqrt{\Sigma_0}}$:*

$$g(\mathbf{x}) \in f_{\sqrt{\Sigma_0}}(\mathbf{x}) [1 - 4\varepsilon, 1 + 4\varepsilon].$$

IV. CODING SCHEME AND ANALYSIS

A. Overview

Given a pair of nested lattices $\Lambda_e^T \subset \Lambda_b^T \subset \mathbb{C}^{n_a T}$ such that

$$\frac{1}{T} \log |\Lambda_b^T / \Lambda_e^T| = R,$$

the transmitter maps a message m to a coset of Λ_e^T in quotient $\Lambda_b^T / \Lambda_e^T$, then samples a point from that coset. Concretely, one can use a one-to-one map ϕ such that $\phi(m) = \lambda_m$, where λ_m is a representative of the coset and then samples the signal $\mathbf{x} \sim \mathcal{D}_{\Lambda_e^T + \lambda_m, \sigma_s}$, broadcasting it to the channels. A block diagram for the transmission until the front-end receivers Bob and Eve is depicted in Figure 2a.

In order to find pairs of sequences of nested lattices Λ_b^T and Λ_e^T we employ constructions of lattices from error-correcting codes. The analysis and full construction are explained in Section V. Essentially, the lattice Λ_b^T controls reliability and has to be chosen in such a way that it is *universally good* for the legitimate compound channel. The lattice Λ_e^T controls the information leakage to the eavesdropper, and has to be chosen in such a way that the flatness factor vanishes universally for any eavesdropper realization (universally good for secrecy). The main result of this section is the following theorem, stating the existence of schemes with vanishing probability of error and vanishing information leakage for all pairs of realizations in the compound set $\mathcal{S}_b \times \mathcal{S}_e$.

Theorem 1. *There exists a sequence of pairs of nested lattices $(\Lambda_b^T, \Lambda_e^T)_{T=1}^\infty$, $\Lambda_b^T \subset \Lambda_e^T \subset \mathbb{C}^{n_a T}$ such that as $T \rightarrow \infty$, the lattice coding scheme universally achieves any secrecy rate*

$$R < (C_b - C_e - n_a)^+.$$

Moreover, we show that both the probability of error and information leakage in Theorem 1 vanishes uniformly for all realizations.

B. The Eavesdropper Channel: Security

For a *fixed* realization \mathbf{H}_e , the key element for bounding the information leakage is the following lemma [11, Lem 2]:

Lemma 2. *Suppose that there exists a probability density function q taking values in $\mathbb{C}^{n_e \times T}$ such that $\mathbb{V}(p_{\mathbf{Y}_e|m}, q_{\mathbf{Y}_e}) \leq \varepsilon_T$ for all $m \in \mathcal{M}_T$. Then, for all message distributions, the information leakage is bounded as:*

$$\mathbb{I}(M; \mathbf{Y}_e) \leq 2n_e T \varepsilon_T R - 2\varepsilon_T \log 2\varepsilon_T. \quad (11)$$

We will show that if the distribution is sufficiently flat, then $\mathbf{Y}_e|m$ is statistically close to a multivariate Gaussian for any $m \in \mathcal{M}_T$. Let us assume for now that \mathbf{H}_e is an invertible square matrix (we next show how to reduce the other cases to this one). In this case, given a message m , we have

$$\mathcal{H}_e \mathbf{x} \sim \mathcal{D}_{\mathcal{H}_e(\Lambda_e^T + \lambda_m), \sqrt{(\mathcal{H}_e \mathcal{H}_e^\dagger) \sigma_s^2}}.$$

According to Lemma 1, the distribution of $\mathcal{H}_e \mathbf{x} + \mathbf{w}_e$ is within variational distance $4\varepsilon_T$ from the normal distribution

$\mathcal{N}(0, \sqrt{\Sigma_0})$, where $\varepsilon_T = \varepsilon_{\mathcal{H}_e \Lambda_e^T}(\sqrt{\Sigma_3})$ and

$$\Sigma_0 = (\mathcal{H}_e \mathcal{H}_e^\dagger) \sigma_s^2 + \sigma_e^2 \mathbf{I}, \quad \Sigma_3^{-1} = (\mathcal{H}_e \mathcal{H}_e^\dagger)^{-1} \sigma_s^{-2} + \sigma_e^{-2} \mathbf{I}. \quad (12)$$

We thus have the following bound for the information leakage ((11) with ε_T replaced by $4\varepsilon_T$):

$$\mathbb{I}(M; \mathbf{Y}_e) \leq 8n_e T \varepsilon_T R - 8\varepsilon_T \log 8\varepsilon_T. \quad (13)$$

Therefore, if $\varepsilon_T = \varepsilon_{\mathcal{H}_e \Lambda_e^T}(\sqrt{\Sigma_3}) = o(1/T)$, the leakage vanishes as T increases *for the specific realization* \mathcal{H}_e . To achieve strong secrecy universally, we must, however, ensure the existence of a lattice with vanishing flatness factor for *all* possible Σ_3 . We postpone the universality discussion to Section V where it is proven that a vanishing flatness factor is possible simultaneously for all $\mathbf{H}_e \in \mathcal{S}_e$ and $\gamma_{\mathcal{H}_e \Lambda_e^T}(\sqrt{\Sigma_3}) < \pi$. This condition implies that semantic security is possible for any VNR,

$$\gamma_{\mathcal{H}_e \Lambda_e^T}(\sqrt{\Sigma_3}) = \frac{|\mathcal{H}_e^\dagger \mathcal{H}_e|^{1/n_a T} V(\Lambda_e^T)^{1/n_a T}}{|\Sigma_3|^{1/n_a T}} < \pi, \quad (14)$$

$$V(\Lambda_e^T)^{1/n_a T} < |\mathbf{I} + \rho_e \mathbf{H}_e^\dagger \mathbf{H}_e|^{-1/n_a} \pi \sigma_s^2 = (\pi \sigma_s^2) e^{-C_e/n_a}. \quad (15)$$

Number-of-Antenna Mismatch. The above analysis assumed that $n_e = n_a$, *i.e.*, the number of eavesdropper receive antennas is *equal* to the the number of transmit antennas. Although analytically simpler, this assumption is not reasonable in practice, since we expect a compound scheme to perform well for any number of eavesdropper antennas. We show next how to reduce the other cases to the square case.

(i) $n_e < n_a$: Recall that the signal received by the eavesdropper is given in matrix form by

$$\mathbf{Y}_e = \mathbf{H}_e \mathbf{X} + \mathbf{W}_e.$$

Let $\tilde{\mathbf{H}}_e \in \mathbb{C}^{(n_a - n_e) \times n_a}$ be a completion of \mathbf{H}_e such that

$$\bar{\mathbf{H}}_e = \begin{pmatrix} \mathbf{H}_e \\ \beta \tilde{\mathbf{H}}_e \end{pmatrix},$$

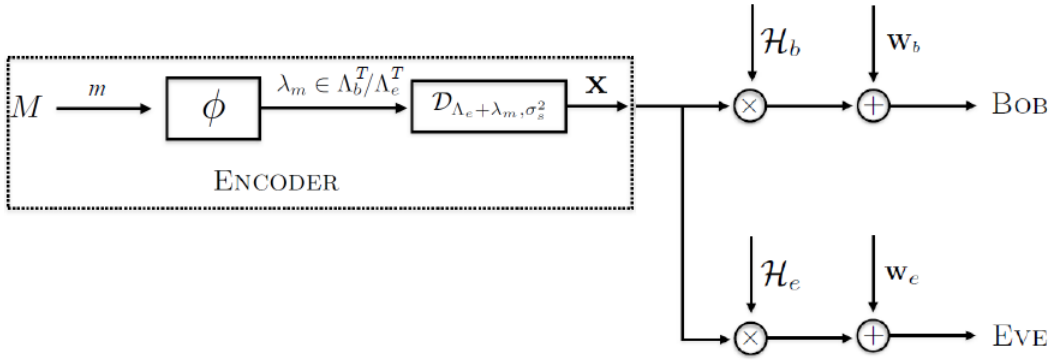
is a full-rank square matrix and $\beta > 0$ is some small number. Let $\tilde{\mathbf{W}}_e \in \mathbb{C}^{(n_a - n_e) \times T}$ be a matrix corresponding to circularly symmetric Gaussian noise. Consider the following surrogate MIMO channel:

$$\begin{pmatrix} \mathbf{Y}_e \\ \tilde{\mathbf{Y}}_e \end{pmatrix} = \begin{pmatrix} \mathbf{H}_e \\ \beta \tilde{\mathbf{H}}_e \end{pmatrix} \mathbf{X} + \begin{pmatrix} \mathbf{W}_e \\ \tilde{\mathbf{W}}_e \end{pmatrix},$$

where $\tilde{\mathbf{H}}_e$ is scaled so that the capacity of the new channel is arbitrarily close to the original one. Indeed for any full rank completion $\tilde{\mathbf{H}}_e$, from the matrix determinant lemma, we have

$$|\mathbf{I} + \rho_e \bar{\mathbf{H}}_e^\dagger \bar{\mathbf{H}}_e| = |\mathbf{I} + \rho_e \mathbf{H}_e^\dagger \mathbf{H}_e| \times |\mathbf{I} + \beta^2 \tilde{\mathbf{H}}_e (|\mathbf{I} + \rho_e \mathbf{H}_e^\dagger \mathbf{H}_e|)^{-1} \tilde{\mathbf{H}}_e^\dagger| \geq e^{C_e}. \quad (16)$$

Therefore, by letting $\beta \rightarrow 0$, the left-hand side tends to e^{C_e} . For any signal \mathbf{X} , the information leakage of the surrogate channel is strictly greater than the original one. Indeed, the the eavesdropper's original channel is stochastically degraded with respect to the augmented one, thus $\mathbb{I}(M; (\mathbf{Y}_e, \tilde{\mathbf{Y}}_e)) \geq \mathbb{I}(M; \mathbf{Y}_e)$. A universally secure code for the $n_a \times n_a$ MIMO compound channel will have vanishing information leakage



(a) Block diagram of the wiretap coding scheme.

(b) Block diagram of Bob's receiver, where \mathbf{F}_b is the MMSE-GDFE matrix and \mathbf{R}_b^{-1} is the inverse linear operator that maps cosets of $\mathbf{R}_b \Lambda_b^T / \mathbf{R}_b \Lambda$ into cosets of $\Lambda_b^T / \Lambda_e^T$.

Fig. 2: Encoding and decoding over the compound wiretap channel.

for the surrogate $n_a \times n_a$ channel (for *any* completion) and therefore will also be secure for the original $n_e \times n_a$ channel.

(ii) $n_e > n_a$: Performing a rectangular QR factorization of \mathbf{H}_e we have:

$$\mathbf{H}_e = \mathbf{Q} \begin{pmatrix} \hat{\mathbf{R}} \\ \mathbf{0} \end{pmatrix},$$

where $\mathbf{Q} \in \mathbb{C}^{n_e \times n_e}$ and $\hat{\mathbf{R}} \in \mathbb{C}^{n_a \times n_a}$ are square matrices. Therefore the eavesdropper's received signal is equivalent to

$$\begin{aligned} \mathbf{Y}_e &= \mathbf{Q} \begin{pmatrix} \hat{\mathbf{R}} \\ \mathbf{0} \end{pmatrix} \mathbf{X} + \begin{pmatrix} \mathbf{W}_e^{(1)} \\ \mathbf{W}_e^{(2)} \end{pmatrix} \\ \Leftrightarrow \mathbf{Q}^\dagger \mathbf{Y}_e &= \begin{pmatrix} \hat{\mathbf{R}} \\ \mathbf{0} \end{pmatrix} \mathbf{X} + \begin{pmatrix} \tilde{\mathbf{W}}_e^{(1)} \\ \tilde{\mathbf{W}}_e^{(2)} \end{pmatrix}, \end{aligned}$$

where the components of the noise matrices $\tilde{\mathbf{W}}_e^{(1)}$, $\tilde{\mathbf{W}}_e^{(2)}$ are i.i.d. Gaussian. The leakage is therefore the same as for the square channel $\hat{\mathbf{R}}$ and a universal code will also achieve vanishing leakage for the non-square channel.

C. The Legitimate Channel: Reliability

It was shown in [29] that if $\mathbf{X} \sim \mathcal{D}_{\Lambda_b^T, \sigma_s}$, then the maximum-a-posteriori (MAP) decoder for the signal \mathbf{Y}_b is equivalent to lattice decoding of $\mathbf{F}_b \mathbf{Y}_b$, where \mathbf{F}_b is the MMSE-GDFE matrix to be defined in the sequel. We cannot claim directly that $\mathbf{X} \sim \mathcal{D}_{\Lambda_b^T, \sigma_s}$, since the message distribution in \mathcal{M}_T need not be uniform. Nonetheless, we show that reliability is still possible for all individual messages.

The full decoding process is depicted in Figure 2b. Bob first applies a filtering matrix \mathbf{F}_b so that

$$\tilde{\mathbf{Y}}_b = \mathbf{F}_b \mathbf{Y}_b = \mathbf{R}_b \mathbf{X} + \mathbf{W}_{b,\text{eff}},$$

where $\mathbf{R}_b^\dagger \mathbf{R}_b = \mathbf{H}_b^\dagger \mathbf{H}_b + \rho_b^{-1} \mathbf{I}$ and $\mathbf{F}_b^\dagger \mathbf{R}_b = \rho_b^{-1} \mathbf{H}_b$, and the effective noise is

$$\mathbf{W}_{b,\text{eff}} = (\mathbf{F}_b \mathbf{H}_b - \mathbf{R}_b) \mathbf{X} + \mathbf{F}_b \mathbf{W}_b.$$

The next step is to decode $\tilde{\mathbf{Y}}_b$ in $\mathbf{R}_b \Lambda_b^T$, in order to obtain $Q_{\mathbf{R}_b \Lambda_b^T}(\tilde{\mathbf{Y}}_b)$, which is then remapped into the element of the coset $\mathbf{R}_b \Lambda_b^T / \mathbf{R}_b \Lambda_e^T$ through the operation $\text{mod } \mathbf{R}_b \Lambda_e^T$. We can then invert the linear transformation associated to \mathbf{R}_b (notice that \mathbf{R}_b has full rank) in order to obtain the coset in $\Lambda_b^T / \Lambda_e^T$ and re-map it to the message space \mathcal{M}_T through ϕ^{-1} .

In the first step, from Lemma 1, the effective noise $\mathbf{W}_{b,\text{eff}}$ is statistically close to a Gaussian noise with covariance:

$$\begin{aligned} \Sigma_{b,\text{eff}} &= \sigma_s^2 (\mathbf{F}_b \mathbf{H}_b - \mathbf{R}_b) (\mathbf{F}_b \mathbf{H}_b - \mathbf{R}_b)^\dagger + \sigma_b^2 \mathbf{F}_b \mathbf{F}_b^\dagger \\ &= \rho_b^{-2} \sigma_s^2 \mathbf{R}_b^{-\dagger} \mathbf{R}_b^{-1} + \sigma_b^2 \rho_b^{-2} \mathbf{R}_b^{-\dagger} \mathbf{H}_b^\dagger \mathbf{H}_b \mathbf{R}_b^{-1} \\ &= \sigma_b^2 \mathbf{R}_b^{-\dagger} (\rho_b^{-1} \mathbf{I} + \mathbf{H}_b^\dagger \mathbf{H}_b) \mathbf{R}_b^{-1} = \sigma_b^2 \mathbf{I}. \end{aligned} \quad (17)$$

provided that $\varepsilon_{(\mathbf{F}_b \mathbf{H}_b - \mathbf{R}_b) \Lambda_e^T}(\Sigma_{b,\text{inv}})$ is small, where

$$\Sigma_{b,\text{inv}}^{-1} = (\sigma_s^2 (\mathbf{F}_b \mathbf{H}_b - \mathbf{R}_b) (\mathbf{F}_b \mathbf{H}_b - \mathbf{R}_b)^\dagger)^{-1} + (\sigma_b^2 \mathbf{F}_b \mathbf{F}_b^\dagger)^{-1}. \quad (18)$$

The probability of error given *any* message m is thus bounded by

$$\mathbb{P}_{\text{err}|m} \leq (1 + 4\varepsilon_{(\mathbf{F}_b \mathbf{H}_b - \mathbf{R}_b) \Lambda_e^T}(\Sigma_{b,\text{inv}})) P(\tilde{\mathbf{W}}_{b,\text{eff}} \notin \mathcal{V}(\mathbf{R}_b \Lambda_b^T)), \quad (19)$$

where each entry of $\tilde{\mathbf{W}}_{b,\text{eff}}$ is i.i.d. normal with variance σ_b^2 . Therefore, if we guarantee that $\varepsilon_{(\mathbf{F}_b \mathbf{H}_b - \mathbf{R}_b) \Lambda_e^T}(\Sigma_{b,\text{inv}})$ is bounded and if we choose a universally good lattice, the probability vanishes for all possible \mathbf{R}_b . This is possible [29] provided that

$$\gamma_{\mathbf{R}_b \Lambda_b^T}(\sigma_b) > \pi e, \quad (20)$$

namely,

$$V(\Lambda_b^T)^{1/n_a T} > \left| \mathbf{I} + \rho_b \mathbf{H}_b^\dagger \mathbf{H}_b \right|^{-1/n_a} \pi e \sigma_s^2 = (\pi e \sigma_s^2) e^{-C_b/n_a}. \quad (21)$$

However, the evaluation of $\Sigma_{b,\text{inv}}$ is cumbersome and implies an extra condition for the flatness of Λ_e^T . Next we show, instead, how to circumvent this problem by using the fact that the effective noise is “asymptotically” sub-Gaussian with covariance matrix $\sigma_b^2 \mathbf{I}$. We say that a centred random vector $\mathbf{w} \in \mathbb{R}^n$ is sub-Gaussian with (proxy) parameter σ if

$$\log E[e^{t\langle \mathbf{w}, \mathbf{u} \rangle}] \leq \frac{t^2 \sigma^2}{2}$$

for all $t \in \mathbb{R}$ and all unit norm vectors $\mathbf{u} \in \mathbb{R}^n$.

Lemma 3 ([28]). *Let \mathbf{x} be a random vector with distribution $\mathcal{D}_{\Lambda_e^T + \lambda_m, \sigma_s}$, and let $\varepsilon' = \varepsilon_{\Lambda_e^T}(\sigma_s)$. For any matrix \mathbf{A} and any vector $\mathbf{u} \in \mathbb{C}^{n_b T}$, we have:*

$$E[e^{\Re\{\mathbf{x}^\dagger \mathbf{A} \mathbf{u}\}}] \leq \left(\frac{1 + \varepsilon'}{1 - \varepsilon'} \right) e^{\frac{\sigma_s^2}{4} \|\mathbf{A} \mathbf{u}\|^2}.$$

Notice that the average power per dimension of a sub-Gaussian random variable is always less than or equal to its parameter σ_s^2 . Moreover, the sum of two sub-Gaussians is also a sub-Gaussian (for more properties, the reader is referred to [28]). The above lemma, along with (18), allows us to establish that $\mathbf{W}_{b,\text{eff}}$ is almost sub-Gaussian with parameter σ_b^2 . Therefore, as long as $\varepsilon' \approx 0$ the probability of error tends to zero if we choose Λ_b^T to be universally AWGN-good.

D. Proof of Theorem 1: Achievable Secrecy Rates

From the previous subsections, semantic security is achievable if Λ_b^T and Λ_e^T satisfy:

- 1) Reliability (20): $\gamma_{\mathbf{R}_b \Lambda_b^T}(\sigma_b) > \pi e$
- 2) Secrecy (14): $\gamma_{\mathcal{H}_e \Lambda_e^T}(\sqrt{\Sigma_3}) < \pi$
- 3) Sub-Gaussianity of equivalent noise and power constraint: $\varepsilon_{\Lambda_e^T}(\sigma_s) \rightarrow 0$.

From and (21), the first two conditions can be satisfied for rates up to

$$\log |\mathbf{I} + \rho_b \mathbf{H}_b^\dagger \mathbf{H}_b| - \log |\mathbf{I} + \rho_e \mathbf{H}_e^\dagger \mathbf{H}_e| - n_a$$

nats per channel use, but the last conditions may, *a priori*, limit these rates to certain SNR regimes. Fortunately, if condition 2) is satisfied, we automatically satisfy the condition for $\varepsilon_{\Lambda_e^T}(\sigma_s) \rightarrow 0$, since

$$\frac{V(\Lambda_e^T)^{1/n_a T}}{\sigma_s^2} \leq \frac{V(\Lambda_e^T)^{1/n_a T}}{e^{-C_e/n_a} \sigma_s^2} < \pi.$$

Therefore, if $(\Lambda_b^T, \Lambda_e^T)$ is a sequence of nested lattices, where

- 1) Λ_b^T is universally good for the compound channel with set \mathcal{S}_b ,
- 2) Λ_e^T is universally secure for the compound channel with set \mathcal{S}_e ,

then nested lattice Gaussian coding achieves any secrecy rate up to

$$R \leq (C_b - C_e - n_a)^+.$$

The *existence* of such nested pairs is proved subsequently in Section V and Appendix B, which concludes the proof of Theorem 1.

In fact using a method in [40] we can further reduce the gap to approximately $n_a \log(e/2)$. We conjecture that this gap can be completely removed with tighter bounds for the variational distance between the discrete and continuous Gaussians. This is left as an open question.

Remark 1. *Theorem 1 is also a slight improvement on the main result of [11, Theorem 5] in the sense that one of the conditions on the SNR of Bob ($\rho_b > e$) is not needed any longer. Indeed, for the Gaussian channel, $n_a = 1$ and the SNR condition for non-zero secrecy rates is $C_b > C_e + 1$, which is equivalent to*

$$\frac{1 + \rho_b}{1 + \rho_e} > e.$$

V. UNIVERSALLY FLAT GAUSSIANS

The results in the previous section require the existence of sequences of lattices which are universally good for the wiretap channel. More specifically, we need a sequence Λ_b^T which is universally AWGN-good and a sequence Λ_e^T whose leakage vanishes for *all* channel realizations of the eavesdropper. The first condition was studied in [29], where it was shown, through a compactness argument, that random lattices are universal. In this section we deal with the second condition and prove the existence of lattices Λ_e^T which are universally good for secrecy of the MIMO channel.

Two methods are provided to establish the main result. The first method relies solely on random lattice coding arguments and achieves secrecy capacity up to a gap of n_a nats per channel use. The second method is based on algebraic reductions and exhibits a larger gap (by a factor of $\omega(n_a \log n_a)$) to capacity, but has the appealing property of reducing the problem to the one of constructing secrecy-good lattices for the *AWGN channel*, making it potentially more useful in practice.

A. Construction A

Construction A (or “mod- p ”) lattices are certainly the simplest choice for constructing pairs of nested lattices, however generalizations based on algebraic lattices may offer greater flexibility in the code design, which could be leveraged to obtain better decoding complexity, diversity, or other parameters. Moreover, the coding scheme in Section V-C entails an extra condition on the ensemble, which can be satisfied by assuming an algebraic structure. A general “flexible” construction can be defined via “generalized reductions”. Let $\psi : \Lambda_{\text{base}} \rightarrow \mathbb{F}_p^T$ be a surjective homomorphism from a base lattice Λ_{base} of complex dimension $N \geq T$ to the vector space \mathbb{F}_p^T (also referred to as a *reduction*). Define the lattice $\Lambda(\mathcal{C})$ as the pre-image of a linear code \mathcal{C} ,

$$\Lambda(\mathcal{C}) = \psi^{-1}(\mathcal{C}).$$

If \mathcal{C} has length T and dimension k , the volume of $\Lambda(\mathcal{C})$ equals to $p^{T-k} V(\Lambda_{\text{base}})$. For instance if $N = T$, $\Lambda_{\text{base}} = \mathbb{Z}[i]^T$ the

mapping ψ is the reduction modulo p :

$$\begin{aligned} \psi(a_1 + b_1 i, a_2 + b_2 i, \dots, a_T + b_T i) = \\ (a_1 \pmod{p}, b_1 \pmod{p}, a_2 \pmod{p}, b_2 \pmod{p}, \dots, \\ a_T \pmod{p}, b_T \pmod{p}), \end{aligned} \quad (22)$$

we recover an analogue of Loeliger's (mod- p) Construction A [41]. In this case we obtain a nested lattice between $\mathbb{Z}[i]^T$ and $p\mathbb{Z}[i]^T$. More refined "direct" constructions can be obtained by using number theory and prime ideals of $\mathbb{Z}[i]$. For instance, if Λ_{base} is the embedding of the ring of integers of a number field and ψ is the reduction modulo a prime ideal we can recover the constructions in [29]. Notice that, for this construction, if $\mathcal{C}_1 \subset \mathcal{C}_2$, we obtain two nested lattices $\Lambda(\mathcal{C}_1) \subset \Lambda(\mathcal{C}_2)$.

It was shown in [42] that if $\{\psi\}$ is an infinite sequence of mappings, under mild conditions² the ensemble of lattices averaged over all linear codes \mathcal{C} of same dimension k satisfies the *Minkowski-Hlawka theorem*, namely:

$$\lim_{p \rightarrow \infty} \mathbb{E}_{\mathcal{C}} \left[\sum_{\mathbf{x} \in \beta \Lambda(\mathcal{C}) \setminus \{\mathbf{0}\}} f(\mathbf{x}) \right] = V^{-1} \int_{\mathbb{C}^N} f(\mathbf{x}),$$

where $\beta = V^{1/2N} (p^{T-k} V(\Lambda))^{1/2N}$ is a constant so that all lattices have volume V . The result holds for any integrable function f which decays sufficiently fast (in particular any function upper bounded by a constant times $1/(\|\mathbf{x}\| + 1)^{2N+\delta}$ for some $\delta > 0$). Clearly the Gaussian probability density function satisfies this restriction.

B. Lattices Which Are Good for Secrecy

In what follows we will apply the generalized version of Construction A to construct a sequence of lattices Λ_e^T which is good for secrecy, *i.e.*, which has vanishing flatness factor for all eavesdropper channel realizations. As usual, T will denote the blocklength (cf. Equation (1)), N will be set to $n_a T$ (the complex dimension of the coding lattice) and $k < T$ is any positive integer.

Using the above Minkowski-Hlawka theorem, there exists an ensemble of lattices \mathbb{L} of volume V such that

$$\mathbb{E}_{\mathbb{L}} \left[\sum_{\mathbf{x} \in \Lambda \setminus \{\mathbf{0}\}} f(\mathbf{x}) \right] \leq V^{-1} \int_{\mathbb{C}^{n_a T}} f(\mathbf{x}) d\mathbf{x} + \varepsilon, \quad (23)$$

for any $\varepsilon > 0$. Equation (23) implies that

$$\mathbb{E}_{\mathbb{L}} \left[\sum_{\mathbf{x} \in \Lambda \setminus \{\mathbf{0}\}} e^{-\mathbf{x}^\dagger \Sigma^{-1} \mathbf{x}} \right] \leq V^{-1} \pi^{n_a T} |\Sigma| + \varepsilon,$$

therefore

$$\mathbb{E}_{\mathbb{L}} \left[\varepsilon_{\Lambda}(\sqrt{\Sigma}) \right] \leq \frac{V(1 + \varepsilon)}{\pi^{n_a T} |\Sigma|}.$$

Hence as long as ε is bounded and $V^{1/T}/\pi|\Sigma|^{1/T}$ is bounded by a constant less than 1, the flatness factor tends to zero exponentially in the proposed lattice coding scheme. The

²More specifically, it is required that the sequence of lattices corresponding to the kernels of ψ has a non-vanishing Hermite parameter.

condition for ε can be achieved, for instance, by choosing p sufficiently large in Construction A.

Lemma 4 (Universally Flat Lattice Gaussians). *Let $\mathcal{H}_e = \mathbf{H}_e \otimes \mathbf{I}$ and $\Sigma_3^{-1} = (\mathcal{H}_e \mathcal{H}_e^\dagger)^{-1} \sigma_s^{-2} + \sigma_e^{-2} \mathbf{I}$, as in Equation (12). For any $\gamma < \pi$, there exists a sequence of lattices $\Lambda_e^T \subset \mathbb{C}^{n_a T}$ with $\gamma_{\mathcal{H}_e \Lambda_e^T}(\sqrt{\Sigma_3}) \leq \gamma$ and universally vanishing flatness factor, *i.e.*,*

$$\lim_{T \rightarrow \infty} \varepsilon_{\mathcal{H}_e \Lambda_e^T}(\sqrt{\Sigma_3}) = 0 \text{ for all } \mathbf{H}_e \in \mathcal{S}_e.$$

Moreover, the convergence rate is exponential, *i.e.*, for all $\mathbf{H}_e \in \mathcal{S}_e$, $\varepsilon_{\mathcal{H}_e \Lambda_e^T}(\sqrt{\Sigma_3}) = e^{-\Omega(T)}$.

Proof: The proof is analogous to the quantization argument for the probability of error in [29], which, in turn follows [36].

(i) *Fixed \mathbf{H}_e .* If \mathbb{L} is a Minkowski-Hlawka ensemble with volume V , then

$$\mathbb{E}_{\mathbb{L}} [\varepsilon_{\mathcal{H}_e \Lambda_e^T}(\Sigma_3)] \leq (1 + \varepsilon) \left(\frac{\gamma_{\mathcal{H}_e \Lambda_e^T}(\sqrt{\Sigma_3})}{\pi} \right)^{n_a T}$$

which guarantees a sequence Λ_e^T (at this point, possibly depending on \mathbf{H}_e) with vanishing flatness factor as long as $\gamma < \pi$.

(ii) *Finite set.* Let $\mathcal{S}_Q \subset \mathcal{S}_e$ be a finite subset of \mathcal{S}_e with cardinality Q . We have

$$\begin{aligned} \mathbb{E}_{\mathbb{L}} \left[\max_{\mathbf{H}_e \in \mathcal{S}_Q} \varepsilon_{\mathcal{H}_e \Lambda_e^T}(\Sigma_3) \right] &\leq \mathbb{E}_{\mathbb{L}} \left[\sum_{\mathbf{H}_e \in \mathcal{S}_Q} \varepsilon_{\mathcal{H}_e \Lambda_e^T}(\Sigma_3) \right] \\ &= Q(1 + \varepsilon) \left(\frac{\gamma_{\mathcal{H}_e \Lambda_e^T}(\sqrt{\Sigma_3})}{\pi} \right)^{n_a T} \rightarrow 0 \end{aligned}$$

which guarantees a sequence Λ_e^T with exponentially vanishing flatness factor for any $\mathbf{H}_e \in \mathcal{S}_e$.

(iii) *Quantization step.* By quantizing the channel space, we can extend step (ii) into a universal code for any channel in \mathcal{S}_e . This analysis is described in Appendix A. Here we provide a sketch of the argument. Suppose \mathcal{S}_Q is a δ -covering for \mathcal{S}_e , *i.e.*, for all $\mathbf{H}_e \in \mathcal{S}_e$, there exists $\mathbf{H}_q \in \mathcal{S}_Q$ such that $\|\mathbf{H}_e - \mathbf{H}_q\| \leq \delta$. From the compactness of \mathcal{S}_e , such a covering exists for any arbitrarily small $\delta > 0$, and the size of the covering depends only on n_a , which is fixed for the whole transmission. Furthermore, the theta series is a continuous function of \mathbf{H}_e , which implies that the flatness factor in different channel realizations are also close. From this, we can choose δ independently of T that guarantees that the total exponent is negative. Therefore, the flatness factor tends to zero uniformly as $T \rightarrow \infty$. ■

The above proof does not rely on a specific realization but rather on the knowledge of the compact *compound set* \mathcal{S}_e . It is reminiscent of a widely used technique in coding for compound channels (*e.g.*, [36]). Essentially, an encoder develops a code for Q_δ channels, where Q_δ is the cardinality of a good quantizer of the channel space. However the quantization Q_δ may increase the effective blocklength for a target information leakage. Moreover, the proof does not give

us insights on how to effectively quantize \mathcal{S}_e , making algebraic approaches appealing in practice.

Lemma 4 shows the existence of universally flat Gaussians or, in other words, the existence of a sequence of lattices Λ_e^T which are good for secrecy. Recall that in our construction IV-A, we required Λ_e^T to be nested with $\Lambda_b^T \supset \Lambda_e^T$, where Λ_b^T is a sequence of lattices which are good for the legitimate compound channel. The existence of Λ_b^T was proven in [29]. In Appendix B we argue that both conditions can be achieved by a nested pair $(\Lambda_b^T, \Lambda_e^T)$ which is the last missing part of the proof of Theorem 1.

C. Algebraic Approach

Following [33], we now define a lattice admitting algebraic reduction.

Definition 6 (EU Decomposition). *We say that Λ admits algebraic reduction if for any unit determinant matrix $\mathbf{A} \in \mathbb{C}^{n_a \times n_a}$ there exists a matrix decomposition of the form $\mathbf{A} = \mathbf{E}\mathbf{U}$, where \mathbf{E} and \mathbf{U} are also unit-determinant satisfying the following properties:*

- 1) $\mathbf{U}\Lambda = \Lambda$,
- 2) $\|\mathbf{E}^{-1}\|_F \leq \alpha$ for some absolute constant α that does not depend on \mathbf{A} .

The Golden Code is one example of a lattice that admits algebraic reduction [33]. Lattices built from generalized versions of Construction A based on number fields and division algebras also admit a similar reduction (if necessary we may relax requirement 1) to include equivalence instead of equality). This property was used in [29] to achieve capacity of the infinite compound MIMO channel. Note that α grows with n_a . See [29, Theorem 3] for an upper bound on α in the case of number fields, and [33] in the case of division algebras. Next, we show that an ensemble of lattices satisfying Definition 6 achieves the secrecy capacity of the compound MIMO channel up to a constant gap.

Recall the following relation between the spectral norm and the Frobenius norm:

$$\|\mathbf{I} \otimes \mathbf{A}\| = \|\mathbf{A}\| \leq \|\mathbf{A}\|_F,$$

for the identity matrix \mathbf{I} of any dimension.

Lemma 5. *Suppose that $\Lambda \subset \mathbb{C}^{n_a T}$ is such that its dual lattice Λ^* admits algebraic reduction. Then for $\mathbf{A} \in \mathbb{C}^{n_a \times n_a}$,*

$$\varepsilon_\Lambda(\sqrt{\mathbf{I}_T \otimes \mathbf{A}}) \leq \varepsilon_\Lambda(|\mathbf{A}|^{1/2n_a}/\alpha).$$

Proof: From the Poisson summation formula and the expression for the flatness factor (9):

$$\begin{aligned} \varepsilon_\Lambda(\sqrt{\mathbf{I}_T \otimes \mathbf{A}}) &= \sum_{\lambda \in \Lambda^* \setminus \{0\}} e^{-\pi^2 \lambda^\dagger (\mathbf{I}_T \otimes \mathbf{A}) \lambda} \\ &= \sum_{\lambda \in \Lambda^* \setminus \{0\}} e^{-\pi^2 \|\sqrt{\mathbf{I}_T \otimes \mathbf{A}} \lambda\|^2} \\ &= \sum_{\lambda \in \Lambda^* \setminus \{0\}} e^{-\pi^2 |\mathbf{A}|^{1/n_a} \left\| \frac{\sqrt{\mathbf{I}_T \otimes \mathbf{A}}}{(\sqrt{|\mathbf{A}|})^{1/n_a}} \lambda \right\|^2}. \end{aligned} \quad (24)$$

Upon decomposing $\frac{\sqrt{\mathbf{A}}}{(\sqrt{|\mathbf{A}|})^{1/n_a}} = \mathbf{E}\mathbf{U}$ as in Definition 6, the last equation becomes

$$\begin{aligned} &\sum_{\lambda \in \Lambda^* \setminus \{0\}} e^{-\pi^2 |\mathbf{A}|^{1/n_a} \|(\mathbf{I} \otimes \mathbf{E})\lambda\|^2} \stackrel{(a)}{\leq} \sum_{\lambda \in \Lambda^* \setminus \{0\}} e^{-\pi^2 |\mathbf{A}|^{1/n_a} \frac{\|\lambda\|^2}{\|\mathbf{E}^{-1}\|^2}} \\ &\stackrel{(b)}{\leq} \sum_{\lambda \in \Lambda^* \setminus \{0\}} e^{-\pi^2 |\mathbf{A}|^{1/n_a} \frac{\|\lambda\|^2}{\|\mathbf{E}^{-1}\|_F^2}} \stackrel{(c)}{\leq} \sum_{\lambda \in \Lambda^* \setminus \{0\}} e^{-\pi^2 |\mathbf{A}|^{1/n_a} \frac{\|\lambda\|^2}{\alpha^2}} \\ &= \varepsilon_\Lambda(|\mathbf{A}|^{1/2n_a}/\alpha), \end{aligned} \quad (25)$$

where (a) is due to the bound $\|\lambda\| \leq \|(\mathbf{I} \otimes \mathbf{E})^{-1}\| \|(\mathbf{I} \otimes \mathbf{E})\lambda\|$ and the fact that $\|(\mathbf{I} \otimes \mathbf{E})^{-1}\| = \|(\mathbf{I} \otimes \mathbf{E}^{-1})\| = \|\mathbf{E}^{-1}\|$, (b) is due to the inequality between the 2-norm and the Frobenius norm and (c) follows from Definition 6. ■

Since

$$\varepsilon_{\mathcal{H}_e \Lambda_e^T}(\sqrt{\Sigma_3}) = \varepsilon_{\Lambda_e^T}(\sqrt{\Sigma}),$$

where $\Sigma^{-1} = \mathcal{H}_e^\dagger \Sigma_3^{-1} \mathcal{H}_e$ and $\Sigma_3^{-1} = (\mathcal{H}_e \mathcal{H}_e^\dagger)^{-1} \sigma_s^{-2} + \sigma_e^{-2} \mathbf{I}$ is block-diagonal, we can apply the above lemma. Therefore, if we construct an ensemble of lattices such that their duals admit algebraic reduction for some constant $\alpha > 0$, then there exist lattices with vanishing flatness factor $\varepsilon_{\mathcal{H}_e \Lambda_e^T}(\sqrt{\Sigma_3})$ provided that

$$\varepsilon_{\Lambda_e^T} \left(\sqrt{\sigma_s^2 \alpha^{-2} e^{-C_e/n_a}} \right) \rightarrow 0. \quad (26)$$

This can be achieved if:

$$V(\Lambda_e^T)^{1/n_a T} < \pi e^{-C_e/n_a} \alpha^{-2} \sigma_s^2. \quad (27)$$

Notice that the right-hand side of (27) depends only on the determinant of Σ_3 or on the capacity of the eavesdropper channel, not on any individual realization. For this condition to hold, we only need a sequence of secrecy-good lattices for a surrogate eavesdropper channel with smaller noise variance (by a factor α^{-2}). Therefore, by combining (21) and (27), we arrive at the the following result:

Theorem 2. *Let $(\Lambda_b^T, \Lambda_e^T)$ be a sequence of nested lattices where: (i) Λ_b^T is universally good for the compound MIMO channel and (ii) Λ_e^T satisfies Definition 6 and is secrecy good for the AWGN channel (Condition (26)). Then nested lattice Gaussian coding achieves any secrecy rate up to*

$$R \leq (C_b - C_e - n_a - 2n_a \log(\alpha))^+.$$

Notice that the gap has a different nature than the one in the previous subsection. It consists of two parts: n_a due to the same restriction on the flatness factor in Theorem 1, and $\log \alpha$ due to algebraic reduction. Although we have conjectured that the gap in Theorem 1 can be essentially removed, this is not the case for $\log \alpha$ in Theorem 2. Indeed, since α cannot be smaller than $\sqrt{n_a}$ [29, Theorem 3], this gap is always larger than $n_a \log n_a$. However, the code construction can be reduced to the problem of finding good lattices for the Gaussian wiretap channel (with some additional algebraic structure), making the design potentially more practical.

Notice also that this strategy is closely related to the ‘‘decoupled design’’ for compound MIMO channels [29, Sect. VI]. Both strategies can indeed be combined, *i.e.*, Bob’s code can also benefit from algebraic reduction. In this case both the original channel decoder and the code design can be greatly simplified, at the cost of an extra gap (*i.e.*, an extra factor $2n_a \log(\alpha)$) to the compound capacity.

VI. DISCUSSION

In this paper, we have presented a construction of nested lattice codes to achieve the secrecy capacity of compound MIMO wiretap channels, up to a gap equal to the number of transmit antennas. Compared to [28], the construction in this work is not only more practical, but also enjoys a smaller gap. With algebraic reduction, further simplification has been made, at the cost of an extra gap to the secrecy capacity. Interestingly, the algebraic approach reaffirms the important role of the dual lattice of Λ_e^T in wiretap channels, firstly discovered in [28].

a) Encoding and decoding: Encoding and decoding are not much different from those of lattice codes for compound MIMO channels in [29]. The generalized Construction A employed in this paper may be viewed as a concatenated code, where the inner code is a lattice with some desired properties, while the outer code is an error correction code. Therefore, decoding can be run successively, which greatly reduces the decoding complexity. As for encoding, the discrete Gaussian shaping can be facilitated by choosing a nice base lattice, *e.g.*, a rotated \mathbb{Z}^{n_a} lattice whose Gaussian shaping is easy. There are highly efficient algorithms for Gaussian shaping over specific lattices [43], but more research is needed for Gaussian shaping over generalized Construction A. Practical implementation of the proposed codes is left as future work.

b) Comparison to other compound models: When the channel \mathbf{H}_b is known and the eavesdropper channel has bounded norm, [20] has shown that the eavesdropper’s worst channel is also isotropic. In this case the capacity can be achieved by decomposing the channels into different independent substreams with appropriate power, and applying independent coding for the Gaussian channel. This is also the case when \mathbf{H}_b has a linear uncertainty. In these cases, a combination of correct power allocation and a similar argument to Lemma 4 shows that semantic secrecy is also achievable by random lattice codes. On the other hand, the algebraic approach (Theorem 2) heavily relies on the fact that the channels in \mathcal{S}_e have the same white-input mutual information.

c) Finite-length performance: The results of this work are based on asymptotic analysis as $T \rightarrow \infty$. The practical performance of the proposed universal codes at finite block lengths warrants an investigation. In particular, how large T is required to approach the promised gap in practice? For given T , how far do practical codes perform from secrecy capacity? It may be a challenging problem to design good, practical universal codes.

As a further perspective, one may consider an ‘‘outage’’ analysis of the MIMO wiretap channel in a finite blocklength regime, where the channel matrices \mathbf{H}_b and \mathbf{H}_e may be

random. In other words, one may analyze the probability that the code rate R exceeds the secrecy capacity. In such scenarios, we believe that lattices with the non-vanishing determinant property will be able to provide universal bounds for the outage probability. We leave it as an open problem.

APPENDIX A

QUANTIZATION OF CHANNEL SPACE

In this appendix we show bounds on the flatness factor in the quantized channel space, formalizing part (iii) in the proof of Lemma 4. Instead of performing the quantization directly in the eavesdropper space \mathcal{S}_b , we will consider the corresponding covariance matrices. Following the notation of Lemma 4, we have:

$$\varepsilon_{\mathcal{H}_e \Lambda_e^T}(\sqrt{\mathbf{\Sigma}_3}) = \varepsilon_{\Lambda_e^T}(\sqrt{\mathbf{\Sigma}}),$$

where $\mathbf{\Sigma}^{-1} = \mathcal{H}_e^\dagger \mathbf{\Sigma}_3^{-1} \mathcal{H}_e$ and $\mathbf{\Sigma}_3^{-1} = (\mathcal{H}_e \mathcal{H}_e^\dagger)^{-1} \sigma_s^{-2} + \sigma_e^{-2} \mathbf{I}$. Let Ω_e be the space of co-variance matrices of the form $\mathbf{\Sigma}$, where \mathbf{H}_e can be any matrix in the space of eavesdropper matrices \mathcal{S}_b :

$$\Omega_e = \{ \mathbf{\Sigma} = (\mathcal{H}_e^\dagger \mathbf{\Sigma}_3^{-1} \mathcal{H}_e)^{-1} : \mathcal{H}_e = \mathbf{I}_T \otimes \mathbf{H}_e, \mathbf{H}_e \in \mathcal{S}_e \}.$$

By using the definition of the flatness factor, we can show the following:

Lemma 6. *Let $\mathbf{\Sigma}, \bar{\mathbf{\Sigma}} \in \Omega_e$ be two matrices satisfying $\|\mathbf{\Sigma} - \bar{\mathbf{\Sigma}}\| \leq \delta$. If δ is sufficiently small, then $\bar{\mathbf{\Sigma}} - \delta \mathbf{I}$ is positive-definite and*

$$\varepsilon_{\Lambda_e^T}(\sqrt{\mathbf{\Sigma}}) \leq \varepsilon_{\Lambda_e^T}(\sqrt{\bar{\mathbf{\Sigma}} - \delta \mathbf{I}}).$$

Proof: For any $\boldsymbol{\lambda} \in \mathbb{C}^{n_a T}$ we have $|\boldsymbol{\lambda}^\dagger (\mathbf{\Sigma} - \bar{\mathbf{\Sigma}}) \boldsymbol{\lambda}| \leq \|\boldsymbol{\lambda}\|^2 \delta$. Therefore

$$\begin{aligned} \varepsilon_{\Lambda_e^T}(\sqrt{\mathbf{\Sigma}}) &= \sum_{\boldsymbol{\lambda} \in \Lambda^* \setminus \{0\}} e^{-\pi^2 \boldsymbol{\lambda}^\dagger \mathbf{\Sigma} \boldsymbol{\lambda}} \\ &= \sum_{\boldsymbol{\lambda} \in \Lambda^* \setminus \{0\}} e^{-\pi^2 \boldsymbol{\lambda}^\dagger (\mathbf{\Sigma} - \bar{\mathbf{\Sigma}}) \boldsymbol{\lambda}} e^{-\pi^2 \boldsymbol{\lambda}^\dagger \bar{\mathbf{\Sigma}} \boldsymbol{\lambda}} \\ &\leq \sum_{\boldsymbol{\lambda} \in \Lambda^* \setminus \{0\}} e^{\pi^2 \delta \boldsymbol{\lambda}^\dagger \boldsymbol{\lambda}} e^{-\pi^2 \boldsymbol{\lambda}^\dagger \bar{\mathbf{\Sigma}} \boldsymbol{\lambda}} = \varepsilon_{\Lambda_e^T}(\sqrt{\bar{\mathbf{\Sigma}} - \delta \mathbf{I}}). \end{aligned}$$

Suppose now that \mathcal{S}_δ is a δ -quantizer for Ω_e with cardinality Q_δ , *i.e.*, for all $\mathbf{\Sigma}$ there exists $\bar{\mathbf{\Sigma}} \in \mathcal{S}_\delta$ such that $\|\mathbf{\Sigma} - \bar{\mathbf{\Sigma}}\| \leq \delta$. For any $\mathbf{\Sigma}$ we have:

$$\begin{aligned} \mathbb{E}[\varepsilon_{\Lambda_e^T}(\sqrt{\mathbf{\Sigma}})] &\leq \mathbb{E}[\varepsilon_{\Lambda_e^T}(\sqrt{\bar{\mathbf{\Sigma}} - \delta \mathbf{I}})] \leq \mathbb{E}[\max_{\bar{\mathbf{\Sigma}} \in \mathcal{S}_\delta} \varepsilon_{\Lambda_e^T}(\sqrt{\bar{\mathbf{\Sigma}} - \delta \mathbf{I}})] \\ &\leq \mathbb{E}[\sum_{\bar{\mathbf{\Sigma}} \in \mathcal{S}_\delta} \varepsilon_{\Lambda_e^T}(\sqrt{\bar{\mathbf{\Sigma}} - \delta \mathbf{I}})] \\ &\leq Q_\delta (1 + \varepsilon_T) \left(\frac{\gamma_{\Lambda_e^T}(\sqrt{\bar{\mathbf{\Sigma}} - \delta \mathbf{I}})}{\pi} \right)^{n_a T} \\ &= Q_\delta (1 + \varepsilon_T) f(\delta) \left(\frac{\gamma_{\Lambda_e^T}(\sqrt{\mathbf{\Sigma}})}{\pi} \right)^{n_a T}, \end{aligned}$$

where

$$f(\delta) = \frac{|\mathcal{S}_\delta|}{|\mathbf{\Sigma} - \delta \mathbf{I}|} = \frac{1}{|\mathbf{I} - \delta \mathbf{\Sigma}^{-1}|}.$$

The last upper bound is universal, in the sense that it does not depend on the specific realization \mathbf{H}_e . Note that if the VNR condition is satisfied, namely $\gamma_{\Lambda_e^T}(\sqrt{\Sigma}) < \pi$, then the term $(\gamma_{\Lambda_e^T}(\sqrt{\Sigma})/\pi)^{n_a T}$ decays exponentially in T with exponent given by

$$c_1 = -n_a \log(\gamma_{\Lambda_e^T}(\sqrt{\Sigma})/\pi).$$

From this, we obtain the bound

$$\begin{aligned} \mathbb{E}[\varepsilon_{\Lambda_e^T}(\sqrt{\Sigma})] &\leq (1 + \varepsilon_T) Q_\delta e^{-T(c_1 - \frac{\log f(\delta)}{T})} \\ &= (1 + \varepsilon_T) Q_\delta e^{-T(c_1 - \log |\mathbf{I} - \delta(\sigma_e^{-2} \mathbf{H}_e^\dagger \mathbf{H}_e + \sigma_s^{-2} \mathbf{I})|)}, \end{aligned} \quad (28)$$

which holds for any $\Sigma \in \Omega_e$. We can therefore choose a small δ (independently of T) such that the total exponent is negative. Since Q_δ does not depend on T , and ε_T can be made arbitrarily small, we obtain an exponential decay of the flatness factor.

APPENDIX B SIMULTANEOUS GOODNESS

From Section V, the construction of universally secure codes boils down to finding a sequence of pairs of nested lattices $\Lambda_b^T \subset \Lambda_e^T$ such that

- Λ_b^T has vanishing probability of error: $\mathbb{P}_{\Lambda_b}(\mathbf{R}_b) \triangleq \mathbb{P}(\tilde{\mathbf{W}}_{b,\text{eff}} \notin \mathcal{V}(\mathbf{R}_b \Lambda_b^T)) \rightarrow 0$ as $T \rightarrow \infty$;
- Λ_e^T has vanishing flatness factor: $\varepsilon_{\Lambda_e^T}(\sqrt{\Sigma}) \rightarrow 0$ as $T \rightarrow \infty$,

where we recall that $\tilde{\mathbf{W}}_{b,\text{eff}}$ is the effective noise, sub-Gaussian with co-variance matrix $\sigma_b^2 \mathbf{I}$, $\mathbf{R}_b^\dagger \mathbf{R}_b = \mathbf{H}_b^\dagger \mathbf{H}_b + \rho_b^{-1} \mathbf{I}$, and

$$\Sigma^{-1} = \mathcal{H}_e^\dagger \Sigma_3^{-1} \mathcal{H}_e, \text{ with } \Sigma_3^{-1} = (\mathcal{H}_e \mathcal{H}_e^\dagger)^{-1} \sigma_s^{-2} + \sigma_e^{-2} \mathbf{I}.$$

First suppose that \mathbf{R}_b and Σ are fixed. Let $\Lambda_b^T = \Lambda(\mathcal{C}_b)$ be obtained by choosing \mathcal{C}_b uniformly in the set of all codes with parameters (T, k_b, p) . Let $\Lambda_e^T = \Lambda(\mathcal{C}_e)$ be obtained by expurgating $k_b - k_e$ columns from \mathcal{C}_b . With this process \mathcal{C}_e will be also chosen uniformly from all (T, k_e, p) codes. We have:

$$\begin{aligned} \mathbb{E}_{\mathcal{C}_b}[\max\{\mathbb{P}_{\Lambda_b^T}(\mathbf{R}_b), \varepsilon_{\Lambda_e^T}(\sqrt{\Sigma})\}] \\ \leq \mathbb{E}_{\mathcal{C}_b}[\mathbb{P}_{\Lambda_b^T}(\mathbf{R}_b) + \varepsilon_{\Lambda_e^T}(\sqrt{\Sigma})] \\ = \mathbb{E}_{\mathcal{C}_b}[\mathbb{P}_{\Lambda_b^T}(\mathbf{R}_b)] + \mathbb{E}_{\mathcal{C}_e}[\varepsilon_{\Lambda_e^T}(\sqrt{\Sigma})] \rightarrow 0. \end{aligned} \quad (29)$$

Convergence of both terms in the last equation is guaranteed to be exponentially fast. Indeed:

- The term $\mathbb{E}_{\mathcal{C}_b}[\mathbb{P}_{\Lambda_b^T}(\mathbf{R}_b)]$ tends to zero exponentially provided that $\gamma_{\mathbf{R}_b}(\Lambda_b) > \pi e$, due to AWGN-goodness of Λ_b^T .
- The term $\mathbb{E}_{\mathcal{C}_e}[\varepsilon_{\Lambda_e^T}(\sqrt{\Sigma})]$ tends to zero exponentially provided that $\gamma_{\Lambda_e^T}(\sqrt{\Sigma}) \rightarrow 0$, due to Appendix A, Equation (28).

Furthermore, by considering the quantized channel spaces, similarly to Appendix A, we conclude that the convergence is universal. Therefore, there exists a pair of lattices Λ_b^T, Λ_e^T where Λ_b^T is universally AWGN-good and Λ_e^T is universally secrecy-good, and Theorem 1 follows.

Remark 2. *Although the above argument only demonstrates the existence of a pair of good lattices, it is possible to show a concentration result on the performance of the ensemble of nested lattices. Suppose some exponential bound e^{-cT} on (29) for some $c > 0$. Then, using Markov's inequality, we have that for the ensemble of nested lattices considered,*

$$\mathbb{P}(\mathbb{P}_{\Lambda_b^T}(\mathbf{R}_b) + \varepsilon_{\Lambda_e^T}(\sqrt{\Sigma}) > e^{-(c-c')T}) < e^{-c'T}, \quad \forall 0 < c' < c. \quad (30)$$

That is, with probability higher than $1 - e^{-c'T}$ over the choice of \mathcal{C}_b , (29) stays below $e^{-(c-c')T}$. In other words, most of these nested lattices have a performance concentrating around e^{-cT} .

ACKNOWLEDGMENT

The authors would like to thank Laura Luzzi and Roope Vehkalahti for helpful discussions.

REFERENCES

- [1] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, pp. 1355–1387, Oct. 1975.
- [2] M. Bloch, J. Barros, M. R. D. Rodrigues, and S. W. McLaughlin, "Wireless information-theoretic security," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2515–2534, June 2008.
- [3] M. Bloch and J. Barros, *Physical Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.
- [4] Y. Liang, H. Poor, and S. Shamai, *Information Theoretic Security*. Foundations and Trends in Communications and Information Theory, Now Publishers, 2009.
- [5] H. Mahdaviifar and A. Vardy, "Achieving the secrecy capacity of wiretap channels using polar codes," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6428–6443, Oct. 2011.
- [6] T. C. Gulcu and A. Barg, "Achieving secrecy capacity of the wiretap channel and broadcast channel with a confidential component," *IEEE Trans. Inform. Theory*, vol. 63, no. 2, pp. 1311–1324, Feb. 2017.
- [7] Y.-P. Wei and S. Ulukus, "Polar coding for the general wiretap channel," in *Proc. 2015 IEEE Inform. Theory Workshop*, Jerusalem, Israel, April 2015, pp. 1–5.
- [8] H. Tyagi and A. Vardy, "Universal hashing for information-theoretic security," *Proc. IEEE*, vol. 103, no. 10, pp. 1781–1795, Oct. 2015.
- [9] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Proc. CRYPTO 2012*, ser. Lecture Notes in Computer Science, vol. 7417. Springer-Verlag, pp. 294–311.
- [10] F. Oggier, P. Solé, and J.-C. Belfiore, "Lattice codes for the wiretap Gaussian channel: Construction and analysis," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5690–5708, Oct. 2016.
- [11] C. Ling, L. Luzzi, J.-C. Belfiore, and D. Stehlé, "Semantically secure lattice codes for the Gaussian wiretap channel," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6399–6416, Oct. 2014.
- [12] J. Barros and M. R. D. Rodrigues, "Secrecy capacity of wireless channels," in *2006 IEEE International Symposium on Information Theory*, July 2006, pp. 356–360.
- [13] Y. Liang, H. V. Poor, and S. Shamai, "Secrecy capacity region of fading broadcast channels," in *2007 IEEE International Symposium on Information Theory*, June 2007, pp. 1291–1295.
- [14] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4961–4972, Aug. 2011.
- [15] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas—Part II: The MIMOME wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5515–5532, Nov. 2010.
- [16] T. Liu and S. Shamai, "A note on the secrecy capacity of the multiple-antenna wiretap channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2547–2553, June 2009.
- [17] S. Loyka and C. D. Charalambous, "Optimal signaling for secure communications over Gaussian MIMO wiretap channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7207–7215, Dec. 2016.
- [18] A. Khina, Y. Kochman, and A. Khisti, "The MIMO wiretap channel decomposed," *IEEE Trans. Inf. Theory*, vol. 64, no. 2, pp. 1046–1063, Feb. 2018.

- [19] X. He and A. Yener, "MIMO wiretap channels with unknown and varying eavesdropper channel states," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6844–6869, Nov. 2014.
- [20] R. F. Schaefer and S. Loyka, "The secrecy capacity of compound Gaussian MIMO wiretap channels," *IEEE Tran. Inf. Theory*, vol. 61, no. 10, pp. 5535–5552, 2015.
- [21] Y. Liang, G. Kramer, H. Poor, and S. Shamai, "Compound wiretap channels," *Eurasip Journal on Wireless Communications and Networking*, vol. 2009, 2009.
- [22] I. Bjelaković, H. Boche, and J. Sommerfeld, "Secrecy results for compound wiretap channels," *Problems of Information Transmission*, vol. 49, no. 1, pp. 73–98, 2013.
- [23] A. Khisti, "Interference alignment for the multiantenna compound wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 2976–2993, May 2011.
- [24] J.-C. Belfiore and F. Oggier, "An error probability approach to MIMO wiretap channels," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3396–3403, August 2013.
- [25] D. Karpuk, A.-M. Ernvall-Hytönen, C. Hollanti, and E. Viterbo, "Probability estimates for fading and wiretap channels from ideal class Zeta functions," *Advances in Mathematics of Communication*, vol. 9, no. 4, pp. 391–413, 2015.
- [26] W. Kositwattanakarn, S. S. Ong, and F. Oggier, "Construction of a lattices over number fields and block fading (wiretap) coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2273–2282, May 2015.
- [27] H. Mirghasemi and J.-C. Belfiore, "Lattice code design criterion for MIMO wiretap channels," in *IEEE Information Theory Workshop (ITW)*, 2015.
- [28] L. Luzzi, R. Vehkalahti, and C. Ling, "Almost universal codes for MIMO wiretap channels," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7218–7241, Nov 2018.
- [29] A. Campello, C. Ling, and J.-C. Belfiore, "Universal lattice codes for MIMO channels," *IEEE Trans. Inf. Theory*, vol. 64, no. 12, pp. 7847–7865, Dec 2018.
- [30] W. L. Root and P. P. Varaiya, "Capacity of classes of Gaussian channels," *SIAM Journal on Applied Mathematics*, vol. 16, no. 6, pp. 1350–1393, 1968.
- [31] S. Loyka and C. D. Charalambous, "A general formula for compound channel capacity," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3971–3991, July 2016.
- [32] G. Rekaya, J.-C. Belfiore, and E. Viterbo, "A very efficient lattice reduction tool on fast fading channels," in *Proceedings of the International Symposium on Information Theory and its Applications (ISITA), Parma, Italy*, 2004.
- [33] L. Luzzi, G. R.-B. Othman, and J.-C. Belfiore, "Algebraic reduction for the Golden code," *Advances in Mathematics of Communications*, vol. 6, no. 1, pp. 1–26, 2012.
- [34] C. Ling, "Achieving capacity and security in wireless communications with lattice codes," in *2016 9th International Symposium on Turbo Codes and Iterative Information Processing (ISTC)*, Sept. 2016, pp. 171–175.
- [35] A. Campello, C. Ling, and J.-C. Belfiore, "Semantically secure lattice codes for compound MIMO channels," in *2018 International Zurich Seminar on Information and Communication (IZS)*, Feb. 2018, pp. 171–175.
- [36] J. Shi and R. Wesel, "A study on universal codes with finite block lengths," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3066–3074, Sept. 2007.
- [37] H. El Gamal, G. Caire, and M. Damen, "Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 968–985, June 2004.
- [38] O. Ordentlich and U. Erez, "Precoded integer-forcing universally achieves the MIMO capacity to within a constant gap," *IEEE Trans. Inf. Theory*, vol. 61, no. 1, pp. 323–340, Jan. 2015.
- [39] D. Micciancio and O. Regev, "Worst-case to average-case reductions based on Gaussian measures," *SIAM J. Comput.*, vol. 37, no. 1, pp. 267–302, Apr. 2007.
- [40] H. Mirghasemi and J.-C. Belfiore, "The semantic secrecy rate of the lattice Gaussian coding for the Gaussian wiretap channel," in *2014 IEEE Information Theory Workshop (ITW 2014)*, Nov 2014, pp. 112–116.
- [41] H.-A. Loeliger, "Averaging bounds for lattices and linear codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 6, pp. 1767–1773, Nov. 1997.
- [42] A. Campello, "Random ensembles of lattices from generalized reductions," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, pp. 5231–5239, July 2018.
- [43] A. Campello and J.-C. Belfiore, "Sampling algorithms for lattice Gaussian codes," in *24th International Zurich Seminar on Communications (IZS)*, Zurich, Switzerland, March 2016, pp. 165–169.