Nonlinear-Precoded Multiuser Secure Transmission with Cooperative Jamming and Adaptive Limited Feedback

Liang Sun, *Member*, *IEEE*, Rui Wang, *Member*, *IEEE*,
Wei Wang, *Member*, *IEEE*, and Victor C. M. Leung, *Fellow*, *IEEE*

***Abstract*— This paper studies secure transceiver design for multiuser multi-antenna systems with an external passive eavesdropper and a cooperative jamming helper. Due to the finite-rate constraint of feedback channels, only quantized channel state information (CSI) of the legitimate users is available at the transmitter and the helper. A nonlinear-precoded secure transmission strategy is proposed using Tomlinson-Harashima precoding at the transmitter and null-space beamforming at the helper based on the quantized CSI. The accurate closed-form expression of an approximation for the ergodic rate of each legitimate user is obtained using quantization cell approximation for random vector quantization of the channels. Assuming the quantized CSI of the legitimate channels and the helper's channel at the eavesdropper, closed-form expression of an upper bound of the ergodic rate of each user's message at the eavesdropper is also derived. Then, a closed-form expression of an approximation for the worst-case ergodic secrecy sum rate follows. We also theoretically show that, besides the advantage in ergodic rate over the linear precoding scheme, when the quantized CSI is *not* available at the eavesdropper, Tomlinson-Harashima precoding is also more effective in degrading the received signal quality at the eavesdropper, and thus is more capable to enhance the secrecy of the systems compared with the linear precoding scheme. Given the total bandwidth constraint of CSI-feedback channels, an adaptive feedback bit allocation algorithm is proposed for each legitimate user to optimize the ergodic secrecy rate performance. Numerical results illustrate that our proposed nonlinear-precoded secure transmission strategy outperforms the corresponding linear precoding scheme, and significant advantage can be achieved by adaptively allocating the total available feedback bits.**

***Index Terms*— Physical layer security, secrecy (sum) rate, jamming, nonlinear precoding, limited feedback.**

Manuscript received Jan 21, 2018; revised May 14, 2018; accepted July 21, 2018. Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org. The work of L. Sun was supported in part by National Natural Science Foundation of China under grant 61671003 and Fundamental Research Funds for the Central Universities under grant YWF-18-BJ-Y-176. The work of R. Wang was supported by National Natural Science Foundation of China under grant 61771232. The work of W. Wang was supported in part by the National Science Foundation of China under Grant 91738202. The work of Victor C. M. Leung was supported by the Canadian Natural Sciences and Engineering Research Council through grants RGPIN-2014-06119 and RGPAS-462031-2014. The associate editor coordinating the review of this paper and approving it for publication was Dr. Walaa Hamouda. (Corresponding author: W. Wang)

L. Sun is with the Beihang University, Beijing, China (email: {eelsun}@buaa.edu.cn), and previously was with WiNMoS Lab at the Department of Electrical and Computer Engineering, the University of British Columbia, Vancouver, BC, Canada.

R. Wang is with the South University of Science and Technology of China, Shenzhen, China (email: {wang.r}@sustc.edu.cn).

W. Wang is with the Huazhong University of Science and Technology, Wuhan, China (email:{weiwangw}@hust.edu.cn).

V. C. M. Leung is with WiNMoS Lab at the Department of Electrical and Computer Engineering, the University of British Columbia, Vancouver, BC, Canada (email: {vleung}@ece.ubc.ca).

I. Introduction

As an alternative to the traditional cryptography-based technologies, physical layer security (PLS) is an emerging technology that enables the exchange of confidential messages over wireless channel by exploiting the physical characteristics of channel fading and source. It has attracted significant attentions from the research community recently [1–7]. The pioneering works by Wyner [8] and I. Csisźar and J. Körner in [9] introduced the wiretap channel model and showed that, when the legitimate user's channel is "more capable" compared to the eavesdropping channel, a positive perfect information rate (secrecy rate, SR) can be achieved for degraded channels and the non-degraded broadcast channels. In the context of PLS, the signal processing methods to achieve security for various contemporary architectures, such as multiple-input-multiple-output (MIMO) systems and relay systems, have been widely studied in [1–7].

In multi-antenna systems, combining controlled artificial noise (AN) or cooperative jamming noise (CJN) together with information signals, the spatial-multiplexing capability is explored to simultaneously enhance legitimate channels' strength and reduce the eavesdropping channel capacity, thus improve the SR significantly [1, 4–7]. Along this line, for single-user (SU) multiple-input-single-output (MISO) systems, the algorithms that combine linear precoding/beamforming with AN and the corresponding ergodic secrecy rate (ESR) were studied for the systems with perfect channel state information at the transmitter (CSIT) in [1], and for systems with limited CSI feedback in [6, 7]. For multiuser (MU) MIMO systems, the ESR performance of the linear secure MU precoding schemes based on zero-forcing (ZF) criterion was studied in [2] for large-scale MIMO systems. In spite of low complexity, the above linear precoding/beamforming schemes inevitably incur capacity loss. Nonlinear precoding provides an alternative approach that offers the potential for rate improvements over linear schemes [10–12]. Particularly, nonlinear Tomlinson-Harashima precoding (THP) can achieve a very good trade-off between performance improvement and complexity increase [10, 11].

It is well known that CSIT is important in wireless communications. However, due to the implementation issues in practise, only partial CSI will be available at the transmit side. For example, limited CSI feedback is widely used in frequency division duplex systems, where the CSI is estimated at the receiver using pilot training, and conveyed to the transmitter through the limited feedback channels. The effects of quantized CSI on the transceiver design and the performance have been extensively studied for both SU and MU multi-antenna systems without considering security

[13–15]. Moreover, it is almost impossible in general for the transmitter to know instantaneous CSI of eavesdropper's channel in practice, especially if eavesdropper is "passive" and keeps silent. Particularly, the schemes proposed in [1–3] all assumed perfect CSI of the legitimate channels at the transmitter for precoding design. With secrecy constraint, some recent works have attempted to relax the perfect CSI assumption. The work in [4, 16] focused on the robust signal processing method design with Gaussian distributed channel estimation errors. [5, 6] studied the impact of quantized channel direction information (CDI) on the ESR. The work in [7] maximized the ergodic secrecy sum rate (ESSR) under a connection outage constraint on the legitimate channel and a secrecy outage constraint against eavesdropping.

As far as we know, compared with linear processing methods there have been very few PLS works employing practical nonlinear precoding (e.g. THP) in multiple-antenna systems, except for [17–21]. Particularly, [17] considered SU MIMO systems with a multiple-antenna eavesdropper and designed a nonlinear THP to guarantee a certain quality-of-service level for the intended user in terms of mean-squared-error (MSE). Partial transmit power was allocated to message-bearing signals in order to achieve the target MSE and the remaining power was allocated to AN to degrade the eavesdropper's channel. Thus, the secrecy is not primary consideration of [17]. [18] and [19] designed nonlinear successive optimization THP based on the perfect CSI of the main channels of multiple legitimate users (LUs) for MU-MIMO systems with multiple eavesdroppers. For the scenario with Gaussian error imperfect CSIT assumption, AN is added to enhance the secrecy performance of the system. Through computer simulations, the ESR as well as bit error rate performance of the proposed method is evaluated and compared with other traditional linear precoders for system with both perfect and imperfect CSI of the main channels. Based on the quantized CSIT, [20] proposed to transmit LU's signals pre-processed by THP together with AN in a MU-MISO system. The advantage of THP over the corresponding linear scheme was illustrated through both the theoretical analysis and some simulation results. Our previous work [21] proposed a similar scheme as this work that employed THP with a coexisting cooperative jamming helper. However, the eavesdropper in the system was assumed to employ a genie-aided eavesdropping method. Thus, the performance obtained is underestimated for any possible practical scenario.

In this work, we aim to add some new results to nonlinear secure transceiver for multi-antenna broadcast systems. We consider a scenario where a multiple-antenna transmitter simultaneously communicates with multiple single-antenna LUs, while there is a passive external multiple-antenna eavesdropper attempting to eavesdrop on the confidential messages of all LUs. The transmitter employs nonlinear THP for secure transmission and to reduce the interference between the MU signals. Moreover, there is a cooperative helper with multiple antennas to enhance the secrecy of the legitimate channels via null-space beamforming approach intended to produce jamming noise signals that are ideally interference-free to all LUs. In contrast to the assumption of CSI in the related works of [17–19], we consider there exist limited-rate feedback channels from each LU to both the transmitter and the cooperative helper and assume each LU can perfectly estimate the instantaneous main channel and the helper's channel. Then, the transmitter and the helper both design the precoding/beaforming based on the quantized CSI conveyed from multiple LUs. In contrast to considering channel-distribution and quantization-method dependent designs (such as the scheme in [22]), we focus on designing low-complexity method in this work, which is fit for easy implementation in practical systems and robust to distributions of channel fading [12]. Thus, our scheme will employ a more direct method which is only based on quantized CDI at the transmitter.

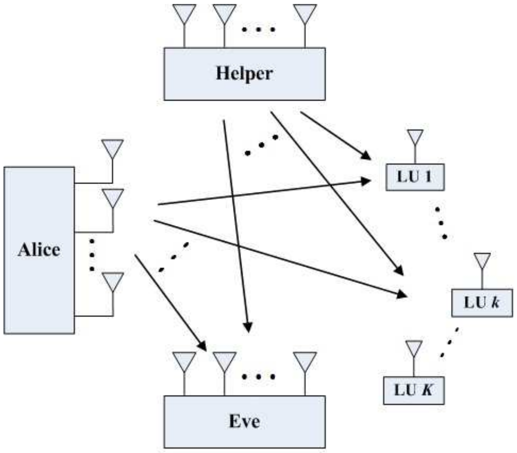


Fig. 1: System model.

For the study of the ESR and the ESSR, we consider the worst case that the eavesdropper can acquire not only the perfect CSI of the channels from the transmitter and the helper, but also the quantized CSI of the main channels of all LUs. The obtained secrecy performance can be viewed as a lower bound of the ESR/ESSR in practise. Using random vector quantization (RVQ) codebooks [13] and quantization cell approximation for RVQ [15], we first derive some closed-form expressions of approximations for both the ESR of each LU and the ESSR. We also derive a closed-form upper bound on the mean loss of the ESR for each LU due to limited CSI feedback. Based on the results, we obtain a closed-form solution of the optimized bit allocation to the main channel of each LU and the helper's channel. Our numerical results illustrate that the proposed nonlinear-precoded secure transmission strategy based on ZF criterion outperforms the linear ZF precoding scheme for all system settings. The numerical results also illustrate the optimized feedback bit allocation can further improve the ESSR performance of the system.

Notation: $\mathbb{C}$ denotes the set of complex numbers. $\mathbb{E}_X\{\cdot\}$ represents expectation with respect to random variable X. $x \overset{d.}{=} y$ denote random variables x and y have the same distribution. $\jmath = \sqrt{-1}$. $\mathbf{I}_N$ denotes $N \times N$ identity matrix. diag$\{\mathbf{M}\}$ denotes the vector consisting of the diagonal elements of matrix $\mathbf{M}$. $[\mathbf{M}]_{i,j}$ denotes the (i, j)-th element of matrix $\mathbf{M}$. $[\mathbf{M}]_{i:j,k:l}$ denote the submatrix of $\mathbf{M}$ obtained by extracting rows i through row j and column k through column l.

II. System and Channel Models

We consider a downlink MU-MISO wiretap broadcast system as illustrated in Fig. 1, where a multiple-antenna

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2018.2864178, IEEE Transactions on Vehicular Technology

transmitter (Alice) simultaneously sends individual confidential messages to K single-antenna LUs, while there is an external passive multiple-antenna eavesdropper (Eve) attempting to eavesdrop on the confidential information all LUs. The message for each LU does not need to be kept confidential among the other LUs, but must be protected from the external eavesdropper. There is a friendly multiple-antenna jammer (Helper) that has no knowledge about the confidential messages and aids the secure communication by sending jamming Gaussian noise to degrade the received signal quality at Eve. We assume there are N_a antennas with Alice, N_h antennas with Helper and N_e antennas with Eve.

Let $\mathbf{t} \in \mathcal{C}^{N_a \times 1}$ denote the information-bearing confidential signal vector transmitted by Alice and $\mathbf{z} \in \mathcal{C}^{N_h}$ denote the Gaussian jamming signal generated by Helper. The signal vectors received at LU k and Eve are given by

$$y_{b,k} = \mathbf{h}_{b,k}\mathbf{t} + \mathbf{g}_{b,k}\mathbf{z} + n_{b,k}, \tag{1}$$

and

$$\mathbf{y}_e \;\; = \;\; \mathbf{H}_e\mathbf{t} + \mathbf{G}_e\mathbf{z} + \mathbf{n}_e, \tag{2}$$

respectively, where $\mathbf{h}_{b,k} \in \mathcal{C}^{1\times N_a}$ and $\mathbf{g}_{b,k} \in \mathcal{C}^{1\times N_h}$ ($k = 1, 2, \cdots, K$) are the channel vectors from Alice and Helper to LU k respectively, $\mathbf{H}_e \in \mathcal{C}^{N_e \times N_a}$ and $\mathbf{G}_e \in \mathcal{C}^{N_e \times N_h}$ are channel matrices from Alice and Helper to Eve respectively, $n_{b,k} \sim \mathcal{CN}(0, \sigma_{b,k}^2)$ and $\mathbf{n}_e \sim \mathcal{CN}\left(\mathbf{0}_{N_e \times 1}, \sigma_e^2 \mathbf{I}_{N_e}\right)$ are additive noises at LU k and Eve. All channels are assumed to be fast block-faded and remain constant during a time block of channel estimation, CSI feedback and data transmission. Using similar model as in many previous works, we consider a scenario with delay-tolerant traffics, where the coding block of each LU's messages is long enough, such that we can use ESSR/ESR as our performance metric [6].

For the limited CSI feedback model, we assume LU k can perfectly estimate each realization of $\mathbf{h}_{b,k}$ and $\mathbf{g}_{b,k}$, and Eve is capable to perfectly estimate $\mathbf{H}_e$ and $\mathbf{G}_e$, but Alice and Helper can only obtain quantized CSI via distinct finite-rate feedback channels from each LU. Since Eve works passively, the instantaneous CSI of the Eve's channel is unknown at Alice, whereas the distribution of the CSI of the Eve's channel is assumed to be available at Alice[1]. Given the distinct quantization codebooks $\mathcal{W}_i$ with 2^{B_i} codewords and $\mathcal{V}_i$ with 2^{D_i} codewords ($i = 1, 2, \cdots, K$), of which $\mathcal{W}_i$ (the codebook for the channel from Alice to LU i) is known to Alice and the respective LU, and $\mathcal{V}_i$ (the codebook for the channel from Helper to LU i) is known to Helper and the respective LU. LU k ($k = 1, 2, \cdots, K$) quantizes the channel direction vectors $\bar{\mathbf{h}}_{b,k} = \frac{\mathbf{h}_{b,k}}{\|\mathbf{h}_{b,k}\|}$ and $\bar{\mathbf{g}}_{b,k} = \frac{\mathbf{g}_{b,k}}{\|\mathbf{g}_{b,k}\|}$ as $\hat{\mathbf{h}}_{b,k} = \arg\max_{\mathbf{a}\in\mathcal{W}_k}\{|\bar{\mathbf{h}}_{b,k}\mathbf{a}^H|^2\}$ and $\hat{\mathbf{g}}_{b,k} = \arg\max_{\mathbf{b}\in\mathcal{V}_k}\{|\bar{\mathbf{g}}_{b,k}\mathbf{b}^H|^2\}$ respectively. We assume the corresponding indices of the selected codewords can be fed back ideally to Alice and Helper [13]. Note that only the CDI is quantized and fed back, since the proposed transmission scheme does not require information of $\| \mathbf{h}_{b,k} \|$ and $\| \mathbf{g}_{b,k} \|$.

III. Nonlinear-Precoded Secure Transceiver Design and Eavesdropping

A. Nonlinear Precoding

In this work, we propose to employ nonlinear THP at Alice for interference pre-subtraction between multiple LUs. Let $\mathbf{s} = [s_1, s_2, \cdots, s_K]^T$ represent the modulated symbol vector for all LUs, where[2] s_k is the modulated symbol for LU k with $\mathbb{E}\{|s_k|^2\} = 1$. At Alice's side, $\mathbf{s}$ is fed into a structure with a backward square matrix $\mathbf{B} \in \mathcal{C}^{K\times K}$, which is strictly lower triangular to allow data precoding in a recursive fashion [10]. The design of precoding matrix $\mathbf{B}$ is based on the CSI of the LUs available at Alice's side. Since $\mathbf{B}$ is a function of the channels, the instantaneous power of the output signal vector can be greatly increased. Thus, a modulo operation $\mathrm{MOD}_{\tau_k}(z) = z - \tau_k \left\lfloor \frac{z+\tau_k}{2\tau_k} \right\rfloor$ is introduced here to ensure that the transmit symbol x_k is mapped into the square region $\mathcal{R}_k = \{x + \sqrt{-1}\ y | x, y \in (-\tau_k, \tau_k)\}$, where τ_k is a modulation specific parameter[3] and $\lfloor x \rfloor$ is the largest integer not exceeding x [10]. Considering the effect of the modulo operation, the precoding procedure can be equivalently given as follows:

$$x_1 = s_1, \quad x_k = s_k + d_k - \sum_{l=1}^{k-1} [\mathbf{B}]_{k,l} x_l, \;\; k = 2, \ldots, K, \tag{3}$$

where $x_k, k = 1, 2, \cdots, K$, are the outputs of THP and $d_k \in \left\{2\tau_k(p_I + \sqrt{-1}\ p_Q) |\ p_I, p_Q \in \mathbb{Z}\right\}$ is properly selected to ensure the real and imaginary parts of x_k to fall into $\mathcal{R}$ [10]. Let $\mathbf{x} = [x_1, x_2, \cdots, x_K]^T$ and effective signal vector be $\mathbf{v} = [v_1, v_2, \cdots, v_K]^T$ with $v_k = s_k + d_k$. Equivalently, we have $\mathbf{v} = \mathbf{s} + \mathbf{d} = \mathbf{C}\mathbf{x}$, where $\mathbf{C} \triangleq \mathbf{B} + \mathbf{I}$. With THP, the elements of $\mathbf{x}$ can be accurately approximated to be independent and uniformly distributed over the Voronoi region of $\mathcal{R}$ [10]. The power of $\mathbf{x}$ is somewhat increased compared to the original symbols from the constellation which is quantified by the *precoding loss*. The power increase decreases as alphabet size increases and can be neglected for moderate to large values of M_k. Thus, we assume $\mathbb{E}\left\{\mathbf{x}\mathbf{x}^H\right\} = \mathbf{I}$ as in [10].

B. Secure Transceiver Design with Quantized CSI

In addition to THP, a spatial channel pre-equalization is performed prior to transmission using a feedforward precoding matrix $\mathbf{F} \in \mathcal{C}^{N_a \times K}$, i.e., $\mathbf{t} = \mathbf{F}\mathbf{x}$. Throughout this work, we assume equal power allocation to the K LUs. The received signals of all LUs can be written in vector form as

$$\mathbf{y}_b = \sqrt{\frac{P_a}{K}} \mathbf{H}_b \mathbf{F} \mathbf{C}^{-1} \mathbf{v} + \mathbf{G}_b \mathbf{z} + \mathbf{n}_b,$$

where $\mathbf{y}_b = [y_{b,1}, y_{b,2}, \cdots, y_{b,K}]^T$ and $\mathbf{n}_b = [n_{b,1}, n_{b,2}, \cdots, n_{b,K}]^T$ are received signal vector at

[1] This assumption has been extensively used in the previous literatures on PLS, e.g. [1, 2, 5–7].

[2] As a default premise in many previous works about PLS [1, 3, 6, 16], s_k carries both bits representing confidential message and random bits to confuse Eve [8, 9], which is obtained by wiretap coding matching to LU k's and Eve's channels.

[3] The modulo device of the THP precoder reduces the transmit signals to a well prescribed range. The operation is tightly related to the used signal constellation. For illustration purpose, we only present the operation with square constellations in the content. τ_k is determined according to the used modulation order and format [10]. We refer to [11] for the operation of THP with more modulation formats.

all LUs and the corresponding additive Gaussian noise vector. $\mathbf{H}_b = \left[\mathbf{h}_{b,1}^T, \cdots, \mathbf{h}_{b,K}^T\right]^T \in \mathcal{C}^{K\times N_a}$ is the fading channel matrix consisting of all channel vectors from Alice to multiple LUs and $\mathbf{G}_b = \left[\mathbf{g}_{b,1}^T, \cdots, \mathbf{g}_{b,K}^T\right]^T \in \mathcal{C}^{K\times N_h}$ is the fading channel matrix consisting of all channel vectors from Helper to multiple LUs. Each LU compensates for the channel gain by dividing by a factor e_k prior to the modulo operation.

As presented in Section II, Alice obtains the quantized downlink CSI through the limited feedback by each LU. Using the result in [13], for LU k we have $\bar{\mathbf{h}}_{b,k} = \hat{\mathbf{h}}_{b,k}\cos\theta_{1,k} + \tilde{\mathbf{h}}_{b,k}\sin\theta_{1,k}$ and $\hat{\mathbf{g}}_{b,k}\cos\theta_{2,k} + \tilde{\mathbf{g}}_{b,k}\sin\theta_{2,k}$, where $\cos^2\theta_{1,k} = |\bar{\mathbf{h}}_{b,k}\hat{\mathbf{h}}_{b,k}^H|^2$, $\tilde{\mathbf{h}}_{b,k} \in \mathbb{C}^{1\times N_a}$ is a unit norm vector isotropically distributed in the orthogonal complement subspace of $\hat{\mathbf{h}}_{b,k}$ and independent of $\sin\theta_{1,k}$. $\tilde{\mathbf{g}}_{b,k} \in \mathbb{C}^{1\times N_h}$ and $\cos^2\theta_{2,k}$ are defined similarly. Then, $\mathbf{H}_b$ and $\mathbf{G}_b$ can be decomposed as $\mathbf{H}_b = \mathbf{\Gamma}\left(\mathbf{\Phi}\hat{\mathbf{H}}_b + \mathbf{\Omega}\tilde{\mathbf{H}}_b\right)$ and $\mathbf{G}_b = \mathbf{\Sigma}\left(\mathbf{\Upsilon}\hat{\mathbf{G}}_b + \mathbf{\Psi}\tilde{\mathbf{G}}_b\right)$, where $\mathbf{\Gamma} = \mathrm{diag}\big(\rho_1, \cdots, \rho_K\big)$ with $\rho_k = \|\mathbf{h}_{b,k}\|$ and $\mathbf{\Sigma} = \mathrm{diag}\big(\xi_1, \cdots, \xi_K\big)$ with $\xi_k = \|\mathbf{g}_{b,k}\|$. $\mathbf{\Phi} = \mathrm{diag}\left(\cos\theta_{1,1}, \cdots, \cos\theta_{1,K}\right)$, $\mathbf{\Omega} = \mathrm{diag}\big(\sin\theta_{1,1}, \cdots, \sin\theta_{1,K}\big)$, $\mathbf{\Upsilon} = \mathrm{diag}\left(\cos\theta_{2,1}, \cdots, \cos\theta_{2,K}\right)$ and $\mathbf{\Psi} = \mathrm{diag}\big(\sin\theta_{2,1}, \cdots, \sin\theta_{2,K}\big)$. $\hat{\mathbf{H}}_b = \left[\hat{\mathbf{h}}_{b,1}^T, \cdots, \hat{\mathbf{h}}_{b,K}^T\right]^T$, $\tilde{\mathbf{H}}_b = \left[\tilde{\mathbf{h}}_{k,1}^T, \cdots, \tilde{\mathbf{h}}_{b,K}^T\right]^T$, $\hat{\mathbf{G}}_b = \left[\hat{\mathbf{g}}_{b,1}^T, \cdots, \hat{\mathbf{g}}_{b,K}^T\right]^T$, $\tilde{\mathbf{G}}_b = \left[\tilde{\mathbf{g}}_{k,1}^T, \cdots, \tilde{\mathbf{g}}_{b,K}^T\right]^T$.

Define the LQ decomposition of compact channel matrix $\hat{\mathbf{H}}_b$ as $\hat{\mathbf{H}}_b = \hat{\mathbf{R}}\hat{\mathbf{Q}}$, where the matrices $\hat{\mathbf{R}} = [\hat{r}_{i,j}] \in \mathcal{C}^{K\times K}$ is a lower left triangular matrix and $\hat{\mathbf{Q}} = [\hat{\mathbf{q}}_1^T, \hat{\mathbf{q}}_2^T, \cdots, \hat{\mathbf{q}}_K^T]^T \in \mathcal{C}^{K\times N_a}$ is a semi-unitary matrix with orthonormal rows that satisfies $\hat{\mathbf{Q}}\hat{\mathbf{Q}}^H = \mathbf{I}_K$. Then, using ZF criterion, the feedforward precoding matrices is obtained as $\mathbf{F} = \hat{\mathbf{Q}}^H$ and the THP is given as $\mathbf{B} = \left(\mathrm{diag}\left\{\hat{\mathbf{R}}\right\}\right)^{-1}\hat{\mathbf{R}} - \mathbf{I}$ [10, 12]. The diagonal matrix consisting of the scaling factors at the multiple LUs is $\mathbf{E} = \sqrt{\frac{K}{P}}\left(\mathbf{\Gamma\Phi}\ \mathrm{diag}\left\{\hat{\mathbf{R}}\right\}\right)^{-1}$. We note that the transmitter only needs to inform LU k the corresponding $\hat{r}_{k,k}$ to obtain the scaling factor, since γ_k and $\cos\theta_{1,k}$ is known by LU k. In addition, the jamming noise signal $\mathbf{z}$ can be written as $\mathbf{z} = \sqrt{\frac{P_h}{N_h - K}}\hat{\mathbf{\Gamma}}_h\mathbf{u}$, where the null constraint is imposed on $\mathbf{z}$ such that $\hat{\mathbf{\Gamma}}_h \in \mathcal{C}^{N_h\times(N_h-K)}$ is an orthonormal basis for the null space of $\hat{\mathbf{G}}_b$, i.e., $\hat{\mathbf{G}}_b\hat{\mathbf{\Gamma}}_h = \mathbf{0}$, and $\mathbf{u} \sim \mathcal{CN}\big(\mathbf{0}, \mathbf{I}_{(N_h-K)}\big)$ such that the transmit power of Helper is $\mathbb{E}\left[\mathbf{z}^H\mathbf{z}\right] = \frac{P_h}{N_h-K}\mathbb{E}\left[\mathbf{u}^H\hat{\mathbf{\Gamma}}_h^H\hat{\mathbf{\Gamma}}_h\mathbf{u}\right] = \frac{P_h}{N_h-K}\mathbb{E}\left[\mathbf{u}^H\mathbf{u}\right] = P_h$. Following many previous related works employing AN (e.g. [6, 7]), here, the transmit power of Helper is uniformly assigned to the null space of $\hat{\mathbf{G}}_b$ due to the absence of Eve's CSI. Then, the signals of all LUs after compensation (before modulo operation) can be written in vector form as

$$
\begin{aligned}
\hat{\mathbf{v}} &= \mathbf{E}\sqrt{\frac{P_a}{K}}\mathbf{\Gamma}\left(\mathbf{\Phi}\hat{\mathbf{H}}_b + \mathbf{\Omega}\tilde{\mathbf{H}}_b\right)\mathbf{F}\mathbf{x} \\
&+ \sqrt{\frac{P_h}{N_h - K}}\mathbf{E}\mathbf{\Sigma}\left(\mathbf{\Upsilon}\hat{\mathbf{G}}_b + \mathbf{\Psi}\tilde{\mathbf{G}}_b\right)\hat{\mathbf{\Gamma}}_h\mathbf{u} + \mathbf{E}\mathbf{n}_b \\
&= \mathbf{v} + \left(\mathbf{\Phi}\ \mathrm{diag}\left\{\hat{\mathbf{R}}\right\}\right)^{-1}\mathbf{\Omega}\tilde{\mathbf{H}}_b\hat{\mathbf{Q}}^H\mathbf{x} + \sqrt{\frac{P_hK}{(N_h - K)P_a}} \\
&\times \left(\mathbf{\Gamma\Phi}\ \mathrm{diag}\left\{\hat{\mathbf{R}}\right\}\right)^{-1}\left(\mathbf{\Sigma\Psi}\tilde{\mathbf{G}}_b\hat{\mathbf{\Gamma}}_h\mathbf{u} + \mathbf{n}_b\right), \qquad (4)
\end{aligned}
$$

where we have used the relationship $\hat{\mathbf{G}}_b\hat{\mathbf{\Gamma}}_h = \mathbf{0}$ and $\mathbf{v} = \left(\mathrm{diag}\left\{\hat{\mathbf{R}}\right\}\right)^{-1}\hat{\mathbf{R}}\mathbf{x}$, and $\hat{\mathbf{v}} = [\hat{v}_1, \hat{v}_2, \cdots, \hat{v}_K]^T$ with $\hat{v}_k$ denoting the estimate of the modified data v_k. In (4), the first term is the useful signal vector for all LUs and the second term is the interference signal caused by the quantized CSI and the effective additive noise. An estimate of LU k's data symbol s_k can be obtained as

$$
\hat{s}_k = \mathrm{MOD}_{\tau_k}\left[v_k + y_k + \hat{n}_k\right] = \mathrm{MOD}_{\tau_k}\left[s_k + y_k + \hat{n}_k\right],
$$

where $y_k \triangleq \frac{\sin\theta_{1,k}}{\hat{r}_{k,k}\cos\theta_{1,k}}\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\mathbf{x} + \sqrt{\frac{P_hK}{(N_h-K)P_a}}\frac{\xi_k\sin\theta_{2,k}}{\rho_k\cos\theta_{1,k}\hat{r}_{k,k}}\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\mathbf{u}$ and $\hat{n}_k \triangleq \sqrt{\frac{K}{P_a}}\frac{1}{\rho_k\hat{r}_{k,k}\cos\theta_{1,k}}n_k$. It is easy to see that, conditioned on the CSI, $\hat{n}_k \sim \mathcal{CN}\big(0, \frac{K}{P_a}\frac{\sigma_{b,k}^2}{\rho_k^2|\hat{r}_{k,k}|^2\cos^2\theta_{1,k}}\big)$.

We note that, *without signal shaping* for channel signal $\mathbf{x}$, the elements of $\mathbf{x}$ can be well approximated as to be independent and uniformly distributed over $\mathcal{R}$ [10], which leads to the result that the achievable rate can be up to 1.53 dB from the channel capacity. To reduce this shaping loss, $\mathbf{x}$ with close to uncorrelated Gaussian distribution restricted to the region of $\mathcal{R}$ can be obtained by combining signal shaping (e.g. trellis shaping) with THP into an entity in a proper way [11]. But the detailed algorithm is more involved and is not the focus of this work. Thus, for the tractability of performance analysis, we approximate $\mathbf{x}$ as being with Gaussian distribution in this work. It had also been noted in [11] that, the effect of the modulo operation at receiver can be neglected for moderate to high signal-to-noise ratios (SNRs) and the end-to-end behavior can be well approximated by the additive Gaussian (interference and) noise model. This approximation had previously been adopted in [10] for the capacity analysis of THP. Thus, we will follow the above approximations with secrecy rate analysis in Section IV.

Following the above assumptions of Gaussian distributed $\mathbf{x}$ as in [10], the output signal-to-interference-plus-noise ratio (SINR) $\hat{\gamma}_k$ for LU k can be written as (5) at the top of the next page.

C. Eavesdropping

Before proceeding, we first note that, since $\mathbf{x}$ is obtained from symbol vector using THP based on the random quantized CSI, if Eve does not know the quantized CSI of all LUs (i.e., $\hat{\mathbf{H}}_b$), or he does not know the processing method of Alice, he cannot the detect s_k with any possible method with the exception of $x_1 = s_1$. And the LU 1's message $x_1 = s_1$ can be protected by employing proper wiretap coding. This is an unique advantage of THP over the linear precoding schemes to enhance secure communications.

The received signals at Eve in (2) can be rewritten as

$$
\mathbf{y}_e^{lfb} = \sqrt{\frac{P_a}{K}}\mathbf{H}_e\hat{\mathbf{Q}}^H\mathbf{x} + \sqrt{\frac{P_h}{N_h - K}}\mathbf{G}_e\hat{\mathbf{\Gamma}}_h\mathbf{u} + \mathbf{n}_e, \quad (6)
$$

which can be identified as a MIMO channel with co-channel interference and noise. In real world, different

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2018.2864178, IEEE Transactions on Vehicular Technology

$$\hat{\gamma}_k = \frac{1}{\frac{\sin^2\theta_{1,k}}{|\hat{r}_{k,k}|^2\cos^2\theta_{1,k}}\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2 + \frac{P_h K}{(N_h-K)P_a}\frac{\xi_k^2\sin^2\theta_{2,k}}{\rho_k^2\cos^2\theta_{1,k}|\hat{r}_{k,k}|^2}\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2 + \frac{K}{P_a}\frac{\sigma_{b,k}^2}{\rho_k^2|\hat{r}_{k,k}|^2\cos^2\theta_{1,k}}}$$

$$= \frac{\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2\cos^2\theta_{1,k}}{\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2\sin^2\theta_{1,k}\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2 + \frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\sin^2\theta_{2,k}\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2 + 1}. \quad (5)$$

decoders with different complexities can be used, such as the low-complexity linear decoders and the optimal maximum-likelihood (ML) decoder whose complexity is exponentially increasing with the dimension of $\mathbf{x}$ (or $\mathbf{s}$) [23]. In this work, instead of studying any concrete decoder, we will focus on studying the maximum achievable ergodic rate of s_k at Eve for the very rigorous scenario of security, where Eve is assumed to have the knowledge of the processing method of Alice and the CSI of $\mathbf{H}_e$, $\mathbf{G}_e$, $\hat{\mathbf{H}}_b$ and $\hat{\mathbf{G}}_b$. Notice that these assumptions result in the worst-case ESR performance of each LU (and also the worst-case ESSR of the system).

IV. Ergodic Secrecy (Sum) Rate Analysis

For the secrecy performance analysis of the proposed scheme described above, all channels are assumed to be mutually independent and spatially uncorrelated Rayleigh fading[4], i.e., $\mathbf{h}_{b,k} \sim \mathcal{CN}\left(\mathbf{0}_{1\times N_a}, \mathbf{I}_{N_a}\right)$, $\mathbf{g}_{b,k} \sim \mathcal{CN}\left(\mathbf{0}_{1\times N_h}, \mathbf{I}_{N_h}\right)$, $\mathbf{H}_e \sim \mathcal{CN}\big(\mathbf{0}_{N_e\times N_a}, \mathbf{I}_{N_e}\otimes\mathbf{I}_{N_a}\big)$, $\mathbf{G}_e \sim \mathcal{CN}\big(\mathbf{0}_{N_e\times N_h}, \mathbf{I}_{N_e}\otimes\mathbf{I}_{N_h}\big)$. For tractability, we further consider the quantization cell approximation for RVQ employed in [13, 15, 24]. Any reasonably well-designed codebook should perform at least as well as RVQ, which gives a performance lower bound of average rate. This quantization cell approximation also provides an accurate performance indication for any well-designed quantization codebook [15]. But the codebook design is not the focus of this work. Before proceeding, we first derive some distribution results related to the target signals, the interfers and the CJN leakage. These results are very useful for the ESSR analysis and the optimization of feedback bit allocation.

A. Some Preliminary Results

Some distribution results related to the target signals, the interfers and the CJN leakage are given in the following lemma.

Lemma 1: For $1 < K < N_a$ and $K < N_h$, the random variables $\eta_k = \|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2$ for $k = 1, \cdots, K$ follow the same beta distribution with shape $(N_h - K)$ and $(K-1)$ which is denoted as $\eta_k \sim \mathrm{Beta}(N_h - K, K - 1)$. In addition, the probability density function (PDF) of η_k is given as

$$f_{\eta_k}(x) = \frac{1}{\beta(N_h-K, K-1)}x^{N_h-K-1}(1-x)^{K-2}, \quad (7)$$

where $\beta(a,b) = \int_0^1 t^{a-1}t^{b-1}\mathrm{d}t = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is beta function [25].

The random variables $\varepsilon_k = \|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2$ for $k = 2, \cdots, K$ follow the same beta distribution with shape $(K-1)$ and $(N_a - K)$ which is denoted as $\varepsilon_k \sim \mathrm{Beta}(K-1, N_a-K)$ with PDF $f_{\varepsilon_k}(x) = \frac{1}{\beta(K-1,N_a-K)}x^{K-2}(1-x)^{N_a-K-1}$. Particularly, when $K = 1$ there is no MU interference term. When $K = N_a$, $\varepsilon_k = \|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2$ is equal to 1 which is a constant. The random variable $|\hat{r}_{k,k}|^2$ $(k = 2, \cdots, K)$ follows beta distribution with shape $(N_a - k + 1)$ and $(k-1)$ with PDF $f_{|\hat{r}_{k,k}|^2}(x) = \frac{1}{\beta(N_a-k+1,k-1)}x^{N_a-k}(1-x)^{k-2}$. For $k = 1$, $|\hat{r}_{k,k}|^2 = 1$.

Proof: See Appendix A. □

B. Ergodic Secrecy (Sum) Rate Approximation

In this subsection, we study the ESSR focusing on the general scenario that $1 < K < N_a$. The results for the special cases of $K = 1$ and $K = N_a$ also can be directly obtained from the derived results. In fact, the work in [6] can be viewed as a special case of our scheme with $K = 1$.

The ESR of LU k with quantized CSI at Alice is given by [1, 9] $R_{sec,k}^{lfb} = \left[R_{b,k}^{lfb} - R_{e,k}^{lfb}\right]^+$, where $R_{b,k}^{lfb} = \mathbb{E}_{\mathbf{H}_b,\mathbf{G}_b,\mathbf{\Phi},\mathbf{\Upsilon}}\left[1 + \hat{\gamma}_k\right]$ is the ergodic rate of LU k's messages over Alice-LU k channel and $R_{e,k}^{lfb}$ is the maximum ergodic rate of LU k's messages over Alice-Eve channel achieved by any possible method at Eve. Although the exact distribution of each random term in the denominator and numerator of (5) is known, since the terms $\rho_k^2|\hat{r}_{k,k}|$ and $\cos^2\theta_{k,1}$ in the numerator are correlated with the terms ρ_k^2 and $\sin^2\theta_{k,1}$ in denominator respectively, it is rather difficult if not impossible to obtain the exact distribution of $\hat{\gamma}_k$ in the very direct way. Alternatively, we will derive an approximation of $R_{b,k}^{lfb}$ in closed-form expression. Using (5), $R_{b,k}^{lfb}$ can be re-expressed as in (8) shown at the top of the next page.

Now, as stated above, the first two terms in the first $\log_2$ of (8) are correlated to each other. Thus, the very direct method to obtain the exact closed-form result of the first expectation in (8) by first obtaining the distribution of the *sum* inside the $\log_2$ is very difficult. Instead, we would like to obtain some tight approximation of (8), which are given the following theorem.

Theorem 1: $R_{b,k}^{lfb} \gtrsim R_{\mathrm{low},k}^{lfb}$, where for $k = 1$, $R_{\mathrm{low},k}^{lfb}$ is given by

$$\begin{aligned} R_{\mathrm{low},k}^{lfb} &= H\left(\mathbf{m}_1, \boldsymbol{\zeta}_2\right) - H\left(\mathbf{m}_2, \boldsymbol{\zeta}_3\right) + \frac{K-1}{N_a-1}2^{B_k} \\ &\quad \times \sum_{m=1}^{N_a-1}\beta\left(\frac{m}{N_a-1} + 1, 2^{B_k} - 1\right) \\ &\quad - \frac{\log_2(e)}{N_a-1}\sum_{i=1}^{N_a-1}\beta\left(2^{B_k}, \frac{i}{N_a-1}\right). \end{aligned} \quad (9)$$

For $k \geq 2$, $R_{\mathrm{low},k}^{lfb}$ is given by

$$R_{\mathrm{low},k}^{lfb} = (1 - A_k)H\left(\mathbf{m}_1, \boldsymbol{\zeta}_1\right) + A_k\ H\left(\mathbf{m}_1, \boldsymbol{\zeta}_2\right) -$$

[4] The variances of all channels are determined by large-scale fading and antenna gains. We first present the analytical results assuming all channel elements' variances are normalized to 1. We will illustrate later how to obtain the results for the general case with arbitrary channel variances based on the derived results.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2018.2864178, IEEE Transactions on Vehicular Technology

$$\begin{aligned} R_{b,k}^{lfb} &= \mathbb{E}_{\mathbf{H}_b,\mathbf{G}_b,\mathbf{H}_e,\mathbf{G}_e,\mathbf{\Phi},\mathbf{\Upsilon}}\Bigg[\log_2\left(1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2\cos^2\theta_{1,k}+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2\sin^2\theta_{1,k}\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2\right. \\ &\quad \left.+\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\sin^2\theta_{2,k}\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2\right)-\log_2\left(1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2\sin^2\theta_{1,k}\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2\right. \\ &\quad \left.+\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\sin^2\theta_{2,k}\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2\right)\Bigg]. \end{aligned} \tag{8}$$

$$H(\mathbf{m},\boldsymbol{\zeta}) = \begin{cases} \log_2(e)\prod_{i=1}^2\frac{1}{\zeta_i^{m_i}}\sum_{j=1}^2\sum_{l=1}^{m_j}\Xi_{j,l}(\mathbf{m},\boldsymbol{\zeta})\,\zeta_j^{m_j-l+1}\exp\left(\frac{1}{\zeta_j}\right)\sum_{q=1}^{m_j-l+1}E_q\left(\frac{1}{\zeta_j}\right) & \zeta_1\neq\zeta_2 \\ \log_2(e)\exp\left(\frac{1}{\zeta_1}\right)\sum_{q=1}^{m_1+m_2}E_q\left(\frac{1}{\zeta_1}\right) & \zeta_1=\zeta_2 \end{cases} \tag{11}$$

$$C_k = \frac{\beta(K,N_a-K)\left[\sum_{j=K}^{N_a-1}\binom{N_a-1}{j}\beta(N_a-k+j,N_a+k-j-2)\right]}{\beta(K-1,N_a-K)\beta(N_a-k+1,k-1)}. \tag{13}$$

$$\begin{aligned} &H(\mathbf{m}_2,\boldsymbol{\zeta}_3)+A_k\left[C_k2^{B_k}\sum_{m=1}^{N_a-1}\beta\left(\frac{m}{N_a-1}+1,2^{B_k}-1\right)\right. \\ &\qquad \left.-\frac{\log_2(e)}{N_a-1}\sum_{i=1}^{N_a-1}\beta\left(2^{B_k},\frac{i}{N_a-1}\right)\right], \end{aligned} \tag{10}$$

where $H(\mathbf{m},\boldsymbol{\zeta})$ is defined as (11) at the top of this page with $\Xi_{j,l}(\mathbf{m},\boldsymbol{\zeta}) = (-1)^{l+1}\binom{m_{3-j}+l-2}{l-1}\left(\frac{1}{\zeta_{3-j}}-\frac{1}{\zeta_j}\right)^{-(m_{3-j}+l-1)}$, and $\mathbf{m}=(m_1,m_2)$ and $\boldsymbol{\zeta}=(\zeta_1,\zeta_2)$. $\mathbf{m}_1=(N_a-k+1,N_h-K)$ and $\mathbf{m}_2=(K-1,N_h-K)$. $\boldsymbol{\zeta}_1=(\frac{1}{\vartheta_{a,k}},\frac{\delta_{2,k}}{\vartheta_{h,k}})$, $\boldsymbol{\zeta}_2=\left(\frac{1}{\vartheta_{a,k}},\frac{\delta_{2,k}}{\vartheta_{h,k}\nu_{1,k}}\right)$ and $\boldsymbol{\zeta}_3=\left(\frac{\delta_{1,k}}{\vartheta_{a,k}},\frac{\delta_{2,k}}{\vartheta_{h,k}}\right)$. $\vartheta_a=\frac{K\sigma_{b,k}^2}{P_a}$, $\vartheta_{h,k}=\frac{(N_h-K)\sigma_{b,k}^2}{P_h}$. $\delta_{1,k}=2^{-\frac{B_k}{N_a-1}}$ and $\delta_{2,k}=2^{-\frac{D_k}{N_h-1}}$. $\nu_{1,1}=\mathbb{E}\big[\cos^2\theta_{1,1}+\|\tilde{\mathbf{h}}_{b,1}\hat{\mathbf{Q}}^H\|^2\sin^2\theta_{1,1}\big]=1-\frac{N_a-K}{N_a-1}2^{B_1}\beta\left(2^{B_1},\frac{N_a}{N_a-1}\right)$. For $k\geq 2$, $\nu_{1,k}=\mathbb{E}\left[\cos^2\theta_{1,k}+\frac{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{|\hat{r}_{k,k}|^2}\sin^2\theta_{1,k}\middle|\frac{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{|\hat{r}_{k,k}|^2}\leq 1\right]=1-(1-C_k)2^{B_k}\beta\left(2^{B_k},\frac{N_a}{N_a-1}\right)$. The coefficients A_k is

$$\begin{aligned} A_k &= \sum_{j=K-1}^{N_a-2}\binom{N_a-2}{j} \\ &\quad\times\frac{\beta(N_a-k+j+1,N_a+k-j-3)}{\beta(N_a-k+1,k-1)} \end{aligned} \tag{12}$$

and C_k is given by (13) at the top of this page. $E_l(x)=\int_1^\infty\frac{e^{-xt}}{t^l}\mathrm{d}t$ is generalized exponential integral.

Proof: See Appendix B. □

In the following, we will derive a closed-form expression of $R_{e,k}^{lfb}$. To satisfy the strict secrecy, *we implicitly assume Eve can acquire the information of* $\hat{\mathbf{H}}_e\triangleq\mathbf{H}_e\hat{\mathbf{Q}}^H$ *and* $\hat{\mathbf{G}}_e\triangleq\mathbf{G}_e\hat{\mathbf{\Gamma}}_h$. Without causing ambiguity, given the above channel matrices the following conditional mutual information and the differential entropies will be implicitly written without these random variables. First, we will show the following lemma.

Lemma 2: Given the CSI of $\hat{\mathbf{H}}_b$ and $\hat{\mathbf{G}}_b$ at Eve, we have

$$I\left(s_k;\mathbf{y}_e^{lfb}|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right) = \frac{1}{K}I\left(\mathbf{s};\mathbf{y}_e^{lfb}|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right) \tag{14}$$

$$= \frac{1}{K}I\left(\mathbf{x};\mathbf{y}_e^{lfb}|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right), \tag{15}$$

for all k, and $I\left(\mathbf{x};\mathbf{y}_e^{lfb}|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)$ can be upper bounded by the instantaneous rate $R_{e,\text{sum}}^{\text{ins}}$ as

$$\begin{aligned} I\left(\mathbf{x};\mathbf{y}_e^{lfb}|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)\leq R_{e,\text{sum}}^{\text{ins}} &\triangleq \log_2\det\left(\frac{P_a}{K}\hat{\mathbf{H}}_e\hat{\mathbf{H}}_e^H\right. \\ &\quad\left.+\frac{P_h}{N_h-K}\hat{\mathbf{G}}_e\hat{\mathbf{G}}_e^H+\sigma_e^2\mathbf{I}\right) \\ &\quad-\log_2\det\left(\frac{P_h}{N_h-K}\hat{\mathbf{G}}_e\hat{\mathbf{G}}_e^H+\sigma_e^2\mathbf{I}\right), \end{aligned} \tag{16}$$

where the upper bound is achieved when $\mathbf{x}$ is Gaussian distributed which can be approximately achieved combing THP and signal shaping. Moreover, we have

$$I\left(\mathbf{s};\mathbf{y}_e^{lfb}|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right) > I\left(\mathbf{s};\mathbf{y}_e^{lfb}\right). \tag{17}$$

Proof: See Appendix C. □

Remark 1: $I\left(\mathbf{s};\mathbf{y}_e^{lfb}|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right) > I\left(\mathbf{s};\mathbf{y}_e^{lfb}\right)$ illustrates the supported rate of Eve's channel can be degraded by preventing Eve obtaining the quantized CSI of $\hat{\mathbf{H}}_b$ and $\hat{\mathbf{G}}_b$ and thus improve the secrecy performance of the system. Whereas for linear precoding schemes (e.g. linear ZF precoding), since Eve can directly obtain the effective channels combing physical channels $\mathbf{H}_e$ and $\mathbf{G}_e$ with the precoding matrices through channel estimation, the quantized CSI $\hat{\mathbf{H}}_b$ and $\hat{\mathbf{G}}_b$ at Eve does not change the secrecy performance of the system, i.e., $I\left(\mathbf{s};\mathbf{y}_e^{lfb}|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)=I\left(\mathbf{s};\mathbf{y}_e^{lfb}\right)$.

In the following, we will derive $R_{e,k}^{lfb}$. According to *Lemma* 2, we have

$$R_{e,k}^{lfb}=\mathbb{E}\left[\max_{p(\mathbf{s},\mathbf{x})}I\left(s_k;\mathbf{y}_e^{lfb}|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)\right]=\frac{1}{K}\mathbb{E}[R_{e,\text{sum}}^{\text{ins}}], \tag{18}$$

where the maximization is taken over all possible input distributions $p(\mathbf{s},\mathbf{x})$. Thus, we first derive a closed-form expression of $\mathbb{E}[R_{e,\text{sum}}^{\text{ins}}]$. It is easy to see that $\mathbb{E}[R_{e,\text{sum}}^{\text{ins}}]$ can be written as

$$\mathbb{E}\left[R_{e,\text{sum}}^{\text{ins}}\right] = \mathbb{E}\Big[\log_2\left[\det\left(\mathbf{W}_{SI}+\sigma_e^2\mathbf{I}\right)\right]$$

$$-\log_2\left[\det\left(\frac{P_h}{N_h-K}\mathbf{W}_I+\sigma_e^2\mathbf{I}\right)\right]\Bigg], \quad (19)$$

where $\mathbf{W}_{SI} = \mathbf{H}_{SI}\mathbf{\Sigma}_{SI}\mathbf{H}_{SI}^H$ and $\mathbf{W}_I = \hat{\mathbf{G}}_e\hat{\mathbf{G}}_e^H$ with $\mathbf{H}_{SI} = \left[\hat{\mathbf{H}}_e\ \hat{\mathbf{G}}_e\right] \in \mathcal{C}^{N_e\times N_h}$ and $\mathbf{\Sigma}_{SI} = \text{blockdiag}\left\{\frac{P_a}{K}\mathbf{I}_K, \frac{P_h}{N_h-K}\mathbf{I}_{N_h-K}\right\}$. It is required that the covariance matrix of the interference plus noise $\frac{P_h}{N_h-K}\mathbf{W}_I+\sigma_e^2\mathbf{I}$ is invertible for Eve. Otherwise, Eve will be able to eliminate the CJN, resulting $R_{e,k}^{lfb}=\infty$. In order to guarantee the covariance matrix of the interference plus noise is invertible even for high signal-to-noise ratio (SNR) regime $(P_a/\sigma_e^2, P_h/\sigma_e^2\to\infty)$, it is required that $N_h \geq N_e+K$.

With uncorrelated Rayleigh fading channels, we have $\hat{\mathbf{H}}_e \sim \mathcal{CN}\left(\mathbf{0}_{N_e\times K}, \mathbf{I}_{N_e}\otimes\mathbf{I}_K\right)$, $\hat{\mathbf{G}}_e \sim \mathcal{CN}\left(\mathbf{0}_{N_e\times(N_h-K)}, \mathbf{I}_{N_e}\otimes\mathbf{I}_{N_h-K}\right)$. Thus, $\mathbf{W}_{SI}$ and $\mathbf{W}_I$ have complex central Wishart distributions, i.e., $\mathbf{W}_{SI}\sim\mathcal{W}_{N_e}\left(N_h,\mathbf{\Sigma}_{SI}\right)$ and $\mathbf{W}_I\sim\mathcal{W}_{N_e}\left(N_h-K,\mathbf{I}_{N_h-K}\right)$. Denote the joint PDFs of the non-zero ordered eigenvalues of complex matrices $\mathbf{W}_{SI}$ and $\mathbf{W}_I$ as $f_1(\boldsymbol{\lambda})$ and $f_2(\boldsymbol{\lambda})$ respectively, where the N_e non-zero eigenvalues are denoted as $\boldsymbol{\lambda}=(\lambda_1,\lambda_2,\cdots,\lambda_{N_e})$. Then, $\mathbb{E}\left[R_{e,\text{sum}}^{\text{ins}}\right]$ in (19) can be rewritten as

$$\mathbb{E}\left[R_{e,\text{sum}}^{\text{ins}}\right] = \int\cdots\int_{\mathcal{D}}\left[\sum_{k=1}^{N_e}\log_2(\sigma_e^2+\lambda_k)\right]f_1(\boldsymbol{\lambda})\mathrm{d}\boldsymbol{\lambda}$$

$$-\int\cdots\int_{\mathcal{D}}\left[\sum_{k=1}^{N_e}\log_2\left(\frac{P_h}{N_h-K}\lambda_k+\sigma_e^2\right)\right]f_2(\boldsymbol{\lambda})\mathrm{d}\boldsymbol{\lambda}, \quad (20)$$

where the integral region is $\mathcal{D}=\{\boldsymbol{\lambda}\,|\lambda_1>\lambda_2>\cdots>\lambda_{N_e}\geq 0\}$. Thus, we first need the joint PDFs, which can be obtained using results in [26, 27] and given in the following lemma.

Lemma 3: $(1a)$ For the scenario that $\frac{P_a}{K}\neq\frac{P_h}{N_h-K}$, $f_1(\boldsymbol{\lambda})$ is given as

$$f_1(\boldsymbol{\lambda}) = K_1\det\left[\mathbf{V}(\boldsymbol{\lambda})\right]\det\left[\mathbf{\Upsilon}(\boldsymbol{\lambda},\boldsymbol{\mu})\right], \quad (21)$$

where $K_1=\frac{(-1)^{(N_h-N_e)N_e}}{\Gamma_{(N_e)}(N_e)\Gamma_{(N_h-K)}(N_h-K)\Gamma_{(K)}(K)}\times\frac{\left(\frac{K}{P_a}\right)^{KN_e}\left(\frac{N_h-K}{P_h}\right)^{(N_h-K)N_e}}{\left|\frac{K}{P_a}-\frac{N_h-K}{P_h}\right|^{K(N_h-K)}}$, and $\boldsymbol{\mu}=[\mu_{(1)},\mu_{(2)}]$ with $\mu_{(1)}=\max\{\frac{K}{P_a},\frac{N_h-K}{P_h}\}$ and $\mu_{(2)}=\min\{\frac{K}{P_a},\frac{N_h-K}{P_h}\}$. The multiplicity of $\mu_{(i)}$, $m_i\in\{K,N_h-K\}$, is determined according to which eigenvalue $(\frac{P_a}{K}$ or $\frac{P_h}{N_h-K})$ $\frac{1}{\mu_{(i)}}$ is equal to and the multiplicity of this eigenvalue. $\mathbf{V}(\boldsymbol{\lambda})$ is a $N_e\times N_e$ Vandermonde matrix with elements $\left[\mathbf{V}(\boldsymbol{\lambda})\right]_{i,j}=\lambda_j^{i-1}$. The $N_h\times N_h$ matrix $\mathbf{\Upsilon}(\boldsymbol{\lambda},\boldsymbol{\mu})$ is given by

$$\{\mathbf{\Upsilon}(\boldsymbol{\lambda},\boldsymbol{\mu})\}_{i,j}$$

$$=\begin{cases}(-\lambda_j)^{m_1-i}e^{-\mu_{(1)}\lambda_j}, \\ i=1,\cdots,m_1; j=1,\cdots,N_e \\ (-\lambda_j)^{N_h-i}e^{-\mu_{(2)}\lambda_j}, \\ i=m_1+1,\cdots,N_h; j=1,\cdots,N_e \\ \left[N_h-j\right]_{(m_1-i)}\mu_{(1)}^{m_2+i-j}, \\ i=1,\cdots,m_1; j=N_e+1,\cdots,N_h \\ \left[N_h-j\right]_{(N_h-i)}\mu_{(2)}^{i-j}, \\ i=m_1+1,\cdots,N_h; j=N_e+1,\cdots,N_h.\end{cases} \quad (22)$$

In addition, $\Gamma_m(a)=\prod_{i=1}^m(a-i)!$ is the normalized complex multivariate gamma function. $[a]_k=a(a-1)\cdots(a-k+1)$, $[a]_0=1$.

$(1b)$ For the scenario that $\frac{P_a}{K}=\frac{P_h}{N_h-K}$, $f_1(\boldsymbol{\lambda})$ is given as

$$f_1(\boldsymbol{\lambda}) = K_2\left[\det\left(\mathbf{V}(\mu\boldsymbol{\lambda})\right)\right]^2\prod_{i=1}^{N_e}e^{-\mu\lambda_i}(\mu\lambda_i)^{N_h-N_e}, \quad (23)$$

where $\mu\triangleq\frac{K}{P_a}=\frac{N_h-K}{P_h}$ and $K_2=\frac{\mu^{N_e}}{\Gamma_{N_e}(N_e)\Gamma_{N_e}(N_h)}$.
(2) $f_2(\boldsymbol{\lambda})$ is given as

$$f_2(\boldsymbol{\lambda}) = L\det\left[\mathbf{V}(\boldsymbol{\lambda})\right]^2\prod_{i=1}^{N_e}e^{-\lambda_i}\lambda_i^{N_h-N_e-K}, \quad (24)$$

where $L=\frac{1}{\Gamma_{N_e}(N_e)\Gamma_{N_e}(N_h-K)}$.

Following the similar steps in [26] to evaluate the integrals in (20) with *Lemma* 3, we can obtain the closed-from expressions of $R_{e,k}^{lfb}$ given in the following theorem.

Theorem 2: $R_{e,k}^{lfb}=\frac{1}{K}\mathbb{E}[R_{e,\text{sum}}^{\text{ins}}]$, where for the scenario that $(1a)$ $\frac{P_a}{K}\neq\frac{P_h}{N_h-K}$,

$$\mathbb{E}[R_{e,\text{sum}}^{\text{ins}}] = K_1\sum_{p=1}^{N_e}\det\left(\mathbf{\Theta}_{1,(p)}\right)-L\sum_{p=1}^{N_e}\det\left(\mathbf{\Xi}_{(p)}\right), \quad (25)$$

and for the scenario that $(1b)$ $\frac{P_a}{K}=\frac{P_h}{N_h-K}$,

$$\mathbb{E}[R_{e,\text{sum}}^{\text{ins}}] = K_2\sum_{p=1}^{N_e}\det\left(\mathbf{\Theta}_{2,(p)}\right)-L\sum_{p=1}^{N_e}\det\left(\mathbf{\Xi}_{(p)}\right) \quad (26)$$

Here, $\mathbf{\Theta}_{1,(p)}\in\mathcal{R}^{N_h\times N_h}$ is given by

$$\mathbf{\Theta}_{1,(p)}=\begin{bmatrix}\mathbf{A}_{(p)} & \mathbf{C}\\ \mathbf{B}_{(p)} & \mathbf{D}\end{bmatrix}, \quad (27)$$

with the (i,j)-th elements of matrices $\mathbf{A}_{(p)}\in\mathcal{R}^{m_1\times N_e}$, $\mathbf{B}_{(p)}\in\mathcal{R}^{m_2\times N_e}$, $\mathbf{C}\in\mathcal{R}^{m_1\times(N_h-N_e)}$ and $\mathbf{D}\in\mathcal{R}^{m_2\times(N_h-N_e)}$ are given respectively as

$$\left[\mathbf{A}_{(p)}\right]_{i,j}=(-1)^{(m_1-i)}\Gamma(m_1-i+j)\ \mu_{(1)}^{-(m_1-i+j)}$$

$$\times T_{j,p}\Big(\log_2(e)\exp\left(\sigma_e^2\mu_{(1)}\right)\sum_{q=1}^{m_1-i+j}E_q\left(\sigma_e^2\mu_{(1)}\right)\Big), \quad (28)$$

$$\left[\mathbf{B}_{(p)}\right]_{i,j}=(-1)^{m_2-i}\Gamma(m_2-i+j)\mu_{(2)}^{-(m_2-i+j)}$$

$$\times T_{j,p}\Big(\log_2(e)\exp\left(\sigma_e^2\mu_{(2)}\right)\sum_{q=1}^{m_2-i+j}E_q\left(\sigma_e^2\mu_{(2)}\right)\Big), \quad (29)$$

$$\left[\mathbf{C}\right]_{i,j}=\left[N_h-N_e-j\right]_{(m_1-i)}\mu_{(1)}^{m_2-N_e+i-j} \quad (30)$$

$$\left[\mathbf{D}\right]_{i,j}=\left[N_h-N_e-j\right]_{(m_2-i)}\mu_{(2)}^{m_1-N_e+i-j}, \quad (31)$$

where the function $T_{j,p}(x)$ is defined as $T_{j,p}(x)=\begin{cases}x & j=p\\ 1 & j\neq p\end{cases}$. The (i,j)-th element of $\mathbf{\Theta}_{2,(p)}\in\mathcal{R}^{N_e\times N_e}$ is given by

$$\left[\mathbf{\Theta}_{2,(p)}\right]_{i,j}=\frac{\Gamma(i+j+N_h-N_e-1)}{\mu}T_{j,p}\Big(\log_2(e)$$

$$\times\exp(\sigma_e^2\mu)\sum_{q=1}^{i+j+N_h-N_e-1}E_q(\sigma_e^2\mu)\Big). \quad (32)$$

The (i,j)-th element of $\mathbf{\Xi}_{(p)}\in\mathcal{R}^{N_e\times N_e}$ is given by

$$\left[\mathbf{\Xi}_{(p)}\right]_{i,j}=\Gamma(i+j+N_h-N_e-K-1)$$

$$\times T_{j,p}\left(\log_2(e)\exp\left(\frac{\sigma_e^2(N_h-K)}{P_h}\right)\right.$$

$$\times \sum_{q=1}^{i+j+N_h-N_e-K-1} E_q\left(\frac{\sigma_e^2(N_h-K)}{P_h}\right)\Bigg). \quad (33)$$

The other symbols are the same as defined in *Lemma* 3. We notice that the ergodic rate of each LU's messages at Eve does not dependent on B_k and D_k.

Remark 2: Since the distributions of $\hat{\mathbf{H}}_e$ and $\hat{\mathbf{G}}_e$ are independent with those of $\hat{\mathbf{H}}_b$ and $\hat{\mathbf{G}}_b$, with (16) it is not difficult to see $R_{e,k}^{lfb}$ in (18) does not dependent on the distributions of $\hat{\mathbf{H}}_b$ and $\hat{\mathbf{G}}_b$.

Remark 3: According to (8) and the derivation of $R_{b,k}^{\text{approx}}$ in *Theorem* 1, it is easy to see the corresponding results for the general case with *arbitrary* channel variances can be obtained by substituting $P_a\sigma_{h,k}^2$ and $P_h\sigma_{g,k}^2$ instead of P_a and P_h respectively into *Theorem* 1, where $\sigma_{h,k}^2$ and $\sigma_{g,k}^2$ are the variances of $\mathbf{h}_k$'s and $\mathbf{g}_k$'s elements respectively. In addition, it is not difficult to see, still, $\mathbf{W}_{SI}$ and $\mathbf{W}_I$ have the Wishart distribution as described above with different correlation matrices $\mathbf{\Sigma}_{SI} = \text{blockdiag}\left\{\frac{P_a\sigma_{he}^2}{K}\mathbf{I}_K, \frac{P_h\sigma_{ge}^2}{N_h-K}\mathbf{I}_{N_h-K}\right\}$ and $\mathbf{\Sigma}_I = \sigma_{ge}^2\mathbf{I}_{N_h-K}$, where σ_{he}^2 and σ_{ge}^2 are the variances of $\mathbf{H}_e$'s and $\mathbf{G}_e$'s elements respectively. Thus, $R_{e,k}^{lfb}$ for the general case with arbitrary channel variances can be obtained by substituting $P_a\sigma_{he,k}^2$ and $P_h\sigma_{ge,k}^2$ instead of P_a and P_h respectively into *Theorem* 2.

With the analytical results of *Theorem* 1 and *Theorem* 2, the ESSR with the limited CSI feedback, $R_{\text{sec,sum}}^{lfb} = \sum_{k=1}^{K} R_{sec,k}^{lfb}$, can be approximated as

$$R_{\text{sec,sum}}^{lfb} \gtrsim \sum_{k=1}^{K}[R_{b,k}^{\text{approx}} - R_{e,k}^{lfb}]^+. \quad (34)$$

Remark 4: We note that the cell approximation method, which has been used for performance analysis of the systems with limited feedback in many previous works and also in this work [13, 15], can lead to a little higher rate performance than the real values for RVQ. This fact was also shown by the numerical results in the very related PLS work of [6] (see Fig. 3 - Fig. 5 in [6]) which employed RVQ. Due to the above reasons, our derived analytical approximation of $R_{b,k}^{lfb}$ (and also ESRs) can be a little larger than the real values for certain system settings. However, as indicated in [15], the analytical performance obtained using the cell approximation method can be accurate for any well-designed codebook other than RVQ. In addition, the terms which we have discarded in deriving the approximation of $R_{b,k}^{lfb}$ in the proof of *Theorem* 1 are very small. Thus, our analytical results can still approximate the exact values well and is useful for the optimization of bit allocation in the next section.

V. Ergodic Secrecy Rate loss and Adaptive Limited Feedback

Another important performance metric is ESR loss due to the quantized CSI at the transmitter. Here, we derive an upper bound of the ESR loss of each LU, based on which we optimize the number of feedback bits for the legitimate channel and Helper's channel to each LU with a constraint of the feedback channel capacity of each LU.

A. Ergodic Secrecy Rate Loss

Using the similar notations to the previous ones, with perfect CSI at Alice, the precoder given by $\mathbf{F} = \mathbf{Q}^H$ is obtained from the LQ decomposition $\mathbf{H}_b = \mathbf{RQ}$. The diagonal matrix $\mathbf{E}$ consisting of the scaling factors of all LUs is given as $\mathbf{E} = \sqrt{\frac{K}{P}}\mathbf{\Delta}$ with $\mathbf{\Delta} = [\text{diag}(\mathbf{R})]^{-1} = \text{diag}\left(r_{1,1}^{-1}, \cdots, r_{K,K}^{-1}\right)$. The feedback matrix reads $\mathbf{B} = \mathbf{\Delta}\mathbf{H}_b\mathbf{F} - \mathbf{I} = \mathbf{\Delta}\mathbf{R} - \mathbf{I}$. In addition, the jamming noise $\mathbf{z}$ now becomes $\mathbf{z} = \sqrt{\frac{P_h}{N_h-K}}\mathbf{\Gamma}_h\mathbf{u}$, where $\mathbf{\Gamma}_h$ is an orthonormal basis for the null space of $\mathbf{G}_b$. The ESR loss of LU k is given by $\Delta R_{sec,k} = R_{sec,k}^{per} - R_{sec,k}^{lfb}$, where $R_{sec,k}^{per}$ denotes the ESR of LU k's messages with perfect CSI at Alice, i.e., $R_{sec,k}^{per} = \left[R_{b,k}^{per} - R_{e,k}^{per}\right]^+$ with $R_{b,k}^{per} = \mathbb{E}[\log_2(1+\gamma_k)]$, $\gamma_k = \frac{P_a}{K}|r_{k,k}|^2$. $R_{e,k}^{per}$ is defined for the case with perfect CSI in the same way as $R_{e,k}^{lfb}$.

Previously, we have seen that $R_{e,k}^{lfb}$ does not depend on the feedback bits B_k and D_k. In fact, since $\hat{\mathbf{Q}} \overset{d.}{=} \mathbf{Q}$, $\hat{\mathbf{\Gamma}}_h \overset{d.}{=} \mathbf{\Gamma}_h$ and they are independent with $\mathbf{H}_e$ and $\mathbf{G}_e$, it follows that $\acute{\mathbf{H}}_e \triangleq \mathbf{H}_e\mathbf{Q}^H \overset{d.}{=} \hat{\mathbf{H}}_e$ and $\acute{\mathbf{G}}_e \triangleq \mathbf{G}_e\mathbf{\Gamma}_h \overset{d.}{=} \hat{\mathbf{G}}_e$. Thus, $R_{e,k}^{per} = R_{e,k}^{lfb}$. Since it is clear that $R_{sec,k}^{per} \geq R_{sec,k}^{lfb}$, the ESR loss $\Delta R_{sec,k}$ can be upper bounded as $\Delta R_{sec,k} \leq R_{b,k}^{per} - R_{b,k}^{lfb}$ [6]. Using the existing results with *Theorem* 1, we can obtain an upper bound of the ESR loss in the following theorem.

Theorem 3: The ESR loss of LU k between the perfect CSI and the quantized CSI at Alice is upper-bounded as $\Delta R_{sec,k} \leq \Delta R_{sec,k}^{up}$, where $\Delta R_{sec,k}^{up}$ is given by

$$\begin{aligned}\Delta R_{sec,k}^{up} \;=\; & \log_2\Bigg(1+\frac{(K-1)N_a}{(N_a-1)\vartheta_{a,k}}2^{B_k}\beta(2^{B_k},\frac{N_a}{N_a-1}) \\ & +\frac{(N_h-K)N_h}{(N_h-1)\vartheta_{h,k}}2^{D_k}\beta(2^{D_k},\frac{N_h}{N_h-1})\Bigg) \\ & +\varpi_k 2^{B_k}\beta(2^{B_k},\frac{N_a}{N_a-1}) \qquad (35)\end{aligned}$$

with $\varpi_1 = \frac{N_a-K}{N_a-1}$ and $\varpi_k = A_k(1-C_k)$ for $k \geq 2$, where the other symbols are as those defined in *Theorem* 1.

Proof: See Appendix D. □

It is easy to check that, given the transmit power, $\Delta R_{sec,k}^{up} \rightarrow 0$ as B_k goes to ∞. Thus, the rate approximation $R_{b,k}^{lfb}$ in *Theorem* 1 and ESR lower bound in (34) is asymptotically tight as B_k goes to 0.

B. Adaptive Feedback Bit Allocation Algorithm

In this subsection, we minimize the upper bound on the ESR loss in (35), with respect to the constraint on the total feedback bits assigned to the legitimate channel and Helper's channel at each LU, i.e., $B_t = B_k + D_k$. For analytical tractability, we will simplify (35) in high-resolution regime (B_k, D_k are large), and then derive the bit allocation.

First, by applying Gautschi's inequalities for the gamma function [28] $\left(\frac{1}{x+1}\right)^{1-s} < \frac{\Gamma(x+s)}{\Gamma(x+1)} < \left(\frac{1}{x}\right)^{1-s}, 0 \leq s < 1, x > 0$ with $x = 2^{D_k} - 1 + \frac{1}{N_h-1}$ and $s = 1 - \frac{1}{N_h-1}$, we have $\left(\frac{1}{2^{D_k}+\frac{1}{N_h-1}}\right)^{\frac{1}{N_h-1}} < \frac{\Gamma\left(2^{D_k}\right)}{\Gamma\left(2^{D_k}+\frac{1}{N_h-1}\right)} < \left(\frac{1}{2^{D_k}-1+\frac{1}{N_h-1}}\right)^{\frac{1}{N_h-1}}$. Using Gautschi's inequalities with $x\Gamma(x) = \Gamma(x+1)$, we can approximate as $2^{D_k}\beta\left(2^{D_k}, \frac{N_h}{N_h-1}\right) \approx \Gamma\left(1+\frac{1}{N_h-1}\right)2^{-\frac{D_k}{N_h-1}}$. Similarly, we have $2^{B_k}\beta\left(2^{B_k}, \frac{N_a}{N_a-1}\right)$

$\approx \Gamma\left(1+\frac{1}{N_a-1}\right)2^{-\frac{B_k}{N_a-1}}$, $\beta\left(2^{B_k},\frac{i}{N_a-1}\right) \approx \Gamma\left(\frac{i}{N_a-1}\right)2^{-\frac{B_k}{N_a-1}}$ and $2^{B_k}\sum_{i=1}^{N_a-1}\frac{\Gamma\left(2^{B_k}-1\right)\Gamma\left(\frac{i}{N_a-1}+1\right)}{\Gamma\left(2^{B_k}+\frac{i}{N_a-1}\right)} \approx \Gamma\left(\frac{1}{N_a-1}+1\right)2^{-\frac{B_k}{N_a-1}}$. Thus, we can approximate the upper bound of the ESR loss of LU k as $\Delta R_{sec,k}^{up} \approx \log_2\left(1+\alpha_k 2^{-\frac{B_k}{N_a-1}}+\beta_k 2^{-\frac{D_k}{N_h-1}}\right) + \varsigma_k 2^{-\frac{B_k}{N_a-1}} = \tilde{\Delta}R_{sec,k}$, where $\alpha_k = \frac{(K-1)N_a\Gamma(1+\frac{1}{N_a-1})}{(N_a-1)\vartheta_{a,k}} > 0$ and $\beta_k = \frac{(N_h-K)N_h\Gamma(1+\frac{1}{N_h-1})}{(N_h-1)\vartheta_{h,k}} > 0$. $\varsigma_k = (\log_2(e)A_k - \varpi_k)\,\Gamma(1+\frac{1}{N_a-1})$. It is easy to check that $\varsigma_k > 0$ for all k. Then, the feedback bit allocation problem can be written as

$$\underset{B_k,D_k}{\text{minimize}} \quad \tilde{\Delta}R_{sec,k} \tag{36}$$

$$\text{subject to} \quad B_k + D_k = B_t. \tag{37}$$

The optimum bit allocation method according to the problem in (36) - (37) is given in the following algorithm, which are proved in Appendix E.

Algorithm 1: (1) For the case that $N_h = N_a$, consider the cubic equation

$$E_{0,k}x^3 + E_{1,k}x^2 + E_{2,k}x + E_{3,k} = 0, \tag{38}$$

where $E_{0,k} = \frac{\beta_k}{N_h-1}2^{-\frac{B_t}{N_h-1}}$, $E_{1,k} = -\frac{\ln(2)\beta_k\varsigma_k}{N_a-1}2^{-\frac{B_t}{N_h-1}}$ and $E_{2,k} = -\frac{\ln(2)\varsigma_k+\alpha_k}{N_a-1}$, $E_{3,k} = -\frac{\ln(2)\alpha_k\varsigma_k}{N_a-1}$. The equation has a unique *positive real* root $x^\star$, which can be obtained in closed form using the well known root formula for cubic equation. If $0 < x^\star \leq 1$, then the optimal bit allocation is $B_k^o = 0$ and $D_k^o = B_t$. If $x^\star > 1$, the minimum value of the upper bound $\tilde{\Delta}R_{sec,k}$ is achieved at the unique stationary point $B_k^\star = (N_a-1)\log_2(x^\star)$. Then, the practical optimal bit allocation can be found by checking the one or two integer values in $[0, B_t]$ nearest to $B_k^\star$.
(2) For the general case that $N_h \neq N_a$, consider the equation

$$E_{0,k}y^{\frac{2}{N_a-1}+\frac{1}{N_h-1}} + E_{1,k}y^{\frac{1}{N_a-1}+\frac{1}{N_h-1}} + E_{2,k}y^{\frac{1}{N_a-1}} + E_{3,k} = 0. \tag{39}$$

Generally, there is no simple closed-form solution to (39). One method to obtain the practical optimal bit allocation is to first obtain the unique *positive real* root $x^\star$ of equation (39) using numerical method and then follow the same method used in the case with $N_a = N_h$. Here, the unique stationary point $B_k^\star = \log_2(x^\star)$.
Note that the parameters α_k, ς_k and β_k are the functions of the transmit power P_a and P_h, the number of transmit antennas N_a and N_h, and the channel and noise statistics reflected in $\sigma_{b,k}^2$ (See footnote 3.). Thus, the optimal feedback bit allocation remains fixed as long as the channel and noise statistics are constant and the transmit power allocation is given.

VI. Numerical Results

For the simulations, we consider a MU MISO system with $N_a = N_h = 5$, $N_e = 2$, $K = 2$ and $\sigma_e^2 = 1$. In Fig. 2 - Fig. 4, we will consider the systems with equal noise power, where $\sigma_{b,k}^2 = 1$ for all k. In Fig. 5, we will consider the systems with both equal and unequal noise power. All the channel coefficients are distributed as assumed in Section IV with variances equal to 1. Since our focus is the rigorous scenario of security, we assume Eve can obtain $\hat{\mathbf{H}}_b$ for all numerical results. The other assumptions are the same as described in the previous sections for system using nonlinear THP. For the clearness of the figures, in the following we will only show the results for the rigorous secrecy scenario but not for any other possibly practical scenario where the Eve cannot obtain $\hat{\mathbf{H}}_b$ and $\hat{\mathbf{G}}_b$. However, as proved in *Lemma* 2, better performance can be expected for nonlinear THP (and larger gap between the performance curves of THP and ZF precoding) in any practical scenario. For the performance of the system using linear ZF precoding, we assume Eve can perfectly obtain the effective channel matrices $\mathbf{H}_b\mathbf{W}_{zf}$ and $\mathbf{G}_b\mathbf{\Gamma}_{zf}$, where $\mathbf{W}_{zf}$ is the precoding matrix obtained at Alice based on $\hat{\mathbf{H}}_b$ and ZF criterion, and $\mathbf{\Gamma}_{zf}$ is the precoding matrix for AN obtained at Helper satisfying the null constraint $\hat{\mathbf{G}}_b\mathbf{\Gamma}_{zf} = 0$. The Eve employs the optimal maximum ratio combining receiver that achieve the maximum rate over the Eve's channel for each LU's messages[5].

Fig. 2 shows the simulation results of ESSRs achieved by our proposed nonlinear precoding scheme and the linear ZF precoding scheme versus P_a with $P_h = \alpha P_a$ for $\alpha = 0.5, 5$ and $B_t = 20$ with *equal* bit allocation method. The corresponding analytical results of (34) based on *Theorem* 1 and *Theorem* 2 are also plotted. We can observe that, for different transmit power of Helper, the proposed nonlinear precoding scheme outperforms the linear precoding scheme for all system settings even in the rigorous secrecy scenario.

Fig. 3 plots the simulation results of the ESSRs and the corresponding analytical results given by (34) versus P_h for different values of B_t (B_k) with *equal* bit allocation method. We can see the natural result that the ESSRs increase as the quantization bits of CDI increases. We also see that, given B_t, the ESSRs do not always increase as P_h increases. This can be explained as, when the CSI at Helper is not perfect, the Helper's CJN degrades the received signal quality at both Eve's and each LU's sides. When P_h is too large, the degradation at each LU's side can be even greater than that at Eve's side. Moreover, we also observe that the optimal transmit power of Helper needs to be decreased as B_t increases. This can be explained as when more feedback bits are available, the desired signal space and the null space of Helper's channels can be more accurately characterized. Thus, Helper's CJN becomes more effective. Moreover, it has been shown in our previous work [12] that nonlinear precoding suffers from quantized CSI more than linear precoding does. However, when SNR is not large or the feedback quantization resolution is high enough, THP can still achieve better ergodic rate than that of the linear precoding. The similar result can be observed by comparing the ergodic secrecy rates of the systems employing THP and those of the systems employing linear precoding in Fig. 3.

In Fig. 4, we plot the simulation results of the ESSRs and the corresponding closed-form analytical results in (34) versus the number of feedback bits allocated to Alice for two different relative Alice-Helper transmit power allocations, where the total number of feedback bits available to each LU is constrained as $B_t = 24$ bits. We can see, for a fixed transmit power of Alice, more bits need to be allocated to Helper to achieve better ESSR performance as P_h increases.

[5] Notice that, when Eve can not recover the original confidential messages, this is indeed the optimal receiver for eavesdropping.

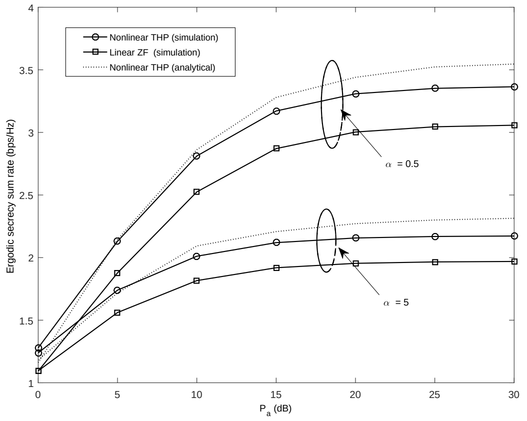


Fig. 2: ESSR versus P_a for $P_h = \alpha P_a$ and $B_t = 20$ with equal bit allocation method.

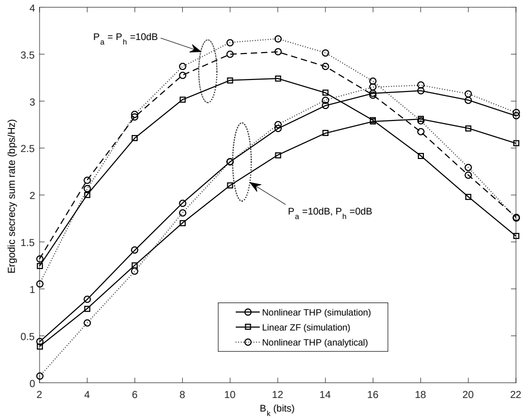


Fig. 4: ESSR versus B_k with $B_t = 24$.

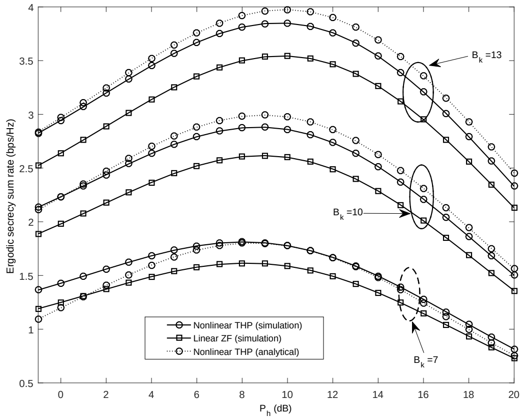


Fig. 3: ESSR versus P_h with different values of B_t and $P_a = 10$dB.

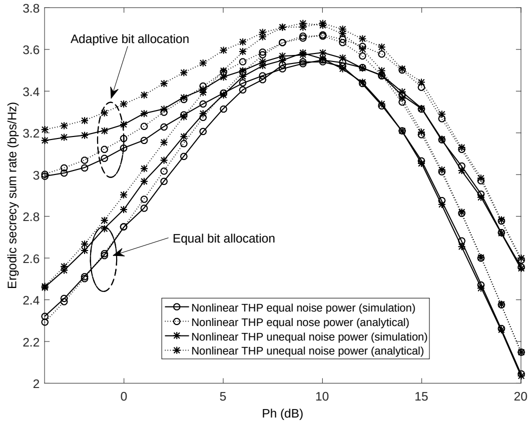


Fig. 5: ESSR achieved by adaptive and equal bit allocation methods versus P_h with $B_t = 24$ and $P_a = 10$dB.

This result agrees with those shown in Fig. 3. In addition, we can observe, with total feedback bits for each LU, better secrecy rate performance can be expected for system with more P_h by adaptively allocating feedback bits for legitimate channel and Helper's channel to each LU. From Fig. 3 and 4, we can see, although there is some gap between the curves of the simulation results and the corresponding analytical results, the analytical results can track the real values well.

At last, we study the performance of the proposed bit allocation algorithm for the systems with both equal and unbalanced noise power. For the system with unbalanced noise power, we assume $\sigma_{b,1}^2 = 1.5$, $\sigma_{b,2}^2 = 0.5$, such that the total noise power is the same as that of the system with equal noise power. Fig. 5 compares the ESSRs versus P_h achieved by the proposed nonlinear-precoded secure transmission scheme with equal feedback bit allocation and the proposed adaptive strategy given by *Algorithm* 1, where $B_t = 24$ bits and $P_a = 10$ dB are fixed. We can see that the proposed adaptive strategy performs better than the equal bit allocation strategy for the systems with both equal noise power and unbalanced noise power. Also, for some special cases, the proposed adaptive algorithm happens to align with equal bit allocation scheme.

VII. Conclusions

In this paper, we have investigated secure communications in multiuser multi-antenna systems with a passive multiple-antenna eavesdropper and a cooperative helper. Using THP at the transmitter and null-space beamforming at the helper, a nonlinear-precoded secure transmission strategy has been proposed based on quantized CSI of the downlink channels from the transmitter and the helper to each LU. Based on RVQ, we have derived closed-form expressions of the approximations for the ergodic rate of each LU and the ESSR. We have also derived an upper bound of the ESR loss of each LU due to quantized CSI. Then, considering a constraint on the total feedback bits of each LU for the legitimate channel and the helper's channel, we have obtained an adaptive bit allocation strategy to minimize the obtained upper bound on the ESR loss. We have showed in theory that, besides the advantage in the ergodic rate over the linear precoding scheme, the nonlinear precoder is also more effective to degrade the received signal quality at any possible eavesdropper. Numerical results have been shown to illustrate our analysis. The numerical results have also demonstrated that the obtained feedback bit allocation algorithm can lead to further improvement in the ESR.

Appendix

A. Proof of Lemma 1

Let LQ decomposition of $\hat{\mathbf{G}}_b$ be $\hat{\mathbf{G}}_b = \mathbf{LU}$, where $\mathbf{L}$ is a $K \times K$ unit lower triangular matrix with the (i,j)-th element $l_{i,j}$, and $\mathbf{U} = \left[\mathbf{u}_1^T, \cdots, \mathbf{u}_K^T\right]^T$ is a $K \times N_h$ semi-unitary matrix whose rows are the orthonormal basis of the subspace spanned by quantized channel vectors $\hat{\mathbf{g}}_{b,k}$ $(k = 1, \cdots, K)$. With RVQ, $\hat{\mathbf{g}}_{b,k}$ $(k = 1, \cdots, K)$ are independently and isotropically distributed on the N_h–dimensional complex unit sphere due to the i.i.d. Rayleigh fading. Thus, the orthonormal basis $\mathbf{u}_1, \cdots, \mathbf{u}_K$ have no preference of direction, i.e., $\mathbf{U}$ is isotropically distributed in the $K \times N_h$ semi-unitary space. Thus, for the derivation of the distribution of $\eta_k = \|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2$, we can assume $\mathbf{u}_i = \mathbf{e}_i$ for $i = 1, \cdots, K$ w.l.g., where $\mathbf{e}_i$ is the i-th row of the identity matrix $\mathbf{I}_{N_h}$. Since $\hat{\mathbf{G}}_b\hat{\mathbf{\Gamma}}_h = \mathbf{0}$, we have $\mathbf{U}\hat{\mathbf{\Gamma}}_h = \mathbf{0}$. Thus, we can let $\hat{\mathbf{\Gamma}}_h = \left[\mathbf{e}_{K+1}^H, \cdots, \mathbf{e}_{N_h}^H\right]$. In addition, we know $\tilde{\mathbf{g}}_{b,k}$ is uniformly distributed in the null space of $\hat{\mathbf{g}}_{b,k}$, i.e., $\tilde{\mathbf{g}}_{b,k} \perp \hat{\mathbf{g}}_{b,k}$, and $\hat{\mathbf{g}}_{b,1} = l_{1,1}\mathbf{e}_1$. Thus, we can write $\tilde{\mathbf{g}}_{b,1} = [0, \boldsymbol{\alpha}_1]$ with $\boldsymbol{\alpha}_1 = [\tilde{g}_{b,1}(2), \cdots, \tilde{g}_{b,1}(N_h)]$ being unit vector isotropically distributed in $\mathcal{C}^{N_h-1}$. Then, we have $\eta_1 = \sum_{i=K+1}^{N_h} |\tilde{g}_{b,1}(i)|^2$. Using the similar method of [12, Apendix A], it can be proved that $\eta_k \overset{d.}{=} \eta_1$ for $k \neq 1$. In the following, we will focus on the distribution of η_1.

From the result in [29] we know the joint PDF of *arbitrary* m elements $(m = 1, 2, \cdots, M-1)$ of a isotropically distributed unit vector $\boldsymbol{\alpha}$ in $\mathcal{C}^M$ is

$$p(\boldsymbol{\alpha}^{(m)}) = \frac{\Gamma(M)}{\pi^m \Gamma(M-m)}(1 - \boldsymbol{\alpha}^{(m)^H}\boldsymbol{\alpha}^{(m)})^{M-m-1}, \quad (40)$$

where $\boldsymbol{\alpha}^{(m)} = [\alpha_1^{(m)}, \alpha_2^{(m)}, \cdots, \alpha_m^{(m)}]^T$ is $m \times 1$ vector composed of arbitrary m elements of $\boldsymbol{\alpha}$. Let $r = \|\boldsymbol{\alpha}^{(m)}\|_2$, we will use the methodology given in [30] for the calculation of distribution of r^2. For each j, let $\alpha_j^{(m)} = x_j + \jmath\, y_j$. Now, we use the following transformation of variables: $x_1 = r\sin\varphi_1\sin\varphi_2\cdots\sin\varphi_{2m-3}\sin\varphi_{2m-2}\sin\varphi_{2m-1}$, $y_1 = r\sin\varphi_1\sin\varphi_2\cdots\sin\varphi_{2m-3}\sin\varphi_{2m-2}\cos\varphi_{2m-1}$, $x_2 = r\sin\varphi_1\sin\varphi_2\cdots\sin\varphi_{2m-3}\cos\varphi_{2m-2}$, $y_2 = r\sin\varphi_1\sin\varphi_2\cdots\cos\varphi_{2m-3}$, $\cdots$, $x_m = r\sin\varphi_1\cos\varphi_2, y_m = r\cos\varphi_1$, where $r > 0$, $0 \leq \varphi_i \leq \pi$ for $i = 1, 2, \cdots, 2m-2$ and $0 \leq \varphi_i \leq 2\pi$ for $i = 2m-1$. The Jacobian of this transformation can be easily obtained as $J = r^{2m-1}\sin^{2m-2}\varphi_1\sin^{2m-3}\varphi_2\cdots\sin\varphi_{2m-2}$. In addition, we have $x_j = r_j\cos\psi_j, y_j = r_j\sin\psi_j$ and $r_1 = r\cos\varphi_1\cos\varphi_2\cdots\cos\varphi_{m-3}\cos\varphi_{m-2}\cos\varphi_{m-1}$, $r_2 = r\cos\varphi_1\cos\varphi_2\cdots\cos\varphi_{m-3}\cos\varphi_{m-2}\sin\varphi_{m-1}$, $r_3 = r\cos\varphi_1\cos\varphi_2\cdots\cos\varphi_{m-3}\sin\varphi_{m-2}, \cdots, r_{m-1} = r\cos\varphi_1\sin\varphi_2, r_m = r\sin\varphi_1$, where, since $r_i > 0$, it follows that $0 \leq \varphi_i \leq \frac{\pi}{2}$ for $j = 1, \cdots, m$. Thus, $\alpha_j^{(m)} = r_j e^{\psi_j}$ for $j = 1, \cdots, m$. Applying these results in (40), we have $p(\boldsymbol{\alpha}^{(m)})d\boldsymbol{\alpha}^{(m)} = p(r, \varphi_1, \varphi_2, \cdots, \varphi_{2m-1})drd\varphi_1$ $d\varphi_2\cdots, d\varphi_{2m-1}$. Thus, the PDF of r^2 is $p_{r^2}(z) = \frac{p_r(\sqrt{z})}{2\sqrt{z}} = \frac{\Gamma(M)}{\pi^m\Gamma(M-m)}\frac{(1-z)^{M-m-1}}{2\sqrt{z}}\sqrt{z}^{2m-1}\int_0^{\pi}\sin^{2m-2}\varphi_1 d\varphi_1$ $\int_0^{\pi}\sin^{2m-3}\varphi_2 d\varphi_2\cdots\int_0^{\pi}\sin\varphi_{2m-2}d\varphi_{2m-2}\int_0^{2\pi}d\varphi_{2m-1}$. Simplifying $p_{r^2}(z)$ using the following results $\int_0^{\frac{\pi}{2}}\sin^{2m}xdx = \frac{\pi}{2}\frac{(2m-1)!!}{(2m)!!}, \int_0^{\frac{\pi}{2}}\sin^{2m-1}xdx = \frac{(2m-2)!!}{(2m-1)!!}, \int_0^{\pi}\sin^k xdx = 2\int_0^{\frac{\pi}{2}}\sin^k xdx$, we have

$$p_{r^2}(z) = \frac{\Gamma(M)}{\Gamma(M-m)\Gamma(m)}(1-z)^{M-m-1}z^{m-1},$$

which is beta distribution with shape m and $(M-m)$. (7) follows from $p_{r^2}(z)$.

We obtain LQ decomposition of $\hat{\mathbf{H}}_b$ applying Gram-Schmidt orthogonalization to the row vectors of $\hat{\mathbf{H}}_b$. Then, we have the following important relations: $\hat{\mathbf{h}}_{b,k} = \sum_{i=1}^k \hat{r}_{k,i}\hat{\mathbf{q}}_i$, where $\hat{r}_{k,i} = \hat{\mathbf{h}}_{b,i}\hat{\mathbf{q}}_i^H$ for $k = 2, \cdots, k$ and $|r_{k,k}|^2 = 1 - \sum_{i=1}^{k-1}|r_{k,i}|^2$. According to the above analysis, for LU k, we can assume $\hat{\mathbf{q}}_i = \mathbf{i}_i$ for $i = 1, \cdots, k-1$ w.l.g., where $\mathbf{i}_i$ is the i-th row of the identity matrix $\mathbf{I}_{N_a}$. With this assumption we have $|\hat{r}_{k,k}|^2 = 1 - \sum_{i=1}^{k-1}|\hat{\mathbf{h}}_{b,k}(i)|^2$, where $\hat{\mathbf{h}}_{b,k}(i)$ is the i-th elements of $\hat{\mathbf{h}}_{b,k}$. Thus, using the result of $p_{r^2}(z)$, it is easy to obtain the distribution of $|\hat{r}_{k,k}|^2$ follows Beta distribution with shape $(N_a - k + 1)$ and $(k-1)$. At last, the distribution of ε_k was obtained and given in [12].

B. Proof of Theorem 1

We can rewrite $R_{b,k}^{lfb}$ in (8) as

$$\begin{aligned} R_{b,k}^{lfb} = \mathbb{E}\Bigg[& \log_2\left(1 + \frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2\right) \\ & + \log_2\Bigg(\cos^2\theta_{1,k} + \frac{1 + \frac{P_a}{K\sigma_{b,k}^2}\rho_k^2\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{1 + \frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2}\sin^2\theta_{1,k} \\ & + \frac{\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2}{1 + \frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2}\sin^2\theta_{2,k}\Bigg) \\ & - \log_2\Bigg(1 + \frac{P_a}{K\sigma_{b,k}^2}\rho_k^2\sin^2\theta_{1,k}\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2 \\ & + \frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\sin^2\theta_{2,k}\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2\Bigg)\Bigg] \end{aligned} \quad (41)$$

In the following, we look at each term individually.

First, we show that $Z_k = \rho_k^2|\hat{r}_{k,k}|^2$ is distributed as Gamma$(N_a - k + 1, 1)$. According to the result in [31], if W_1 and W_2 are independent gamma random variables with $W_1 \sim$ Gamma(n_1, λ) and $W_2 \sim$ Gamma(n_2, λ), then $\omega = \frac{W_1}{W_1+W_2} \sim$ Beta(n_1, n_2). In addition, we know $(W_1 + W_2) \sim$ Gamma$(n_1 + n_2, \lambda)$. Thus, if W is a gamma random variable distributed as $W \sim$ Gamma(n, λ), where $n = n_1 + n_2$, then $\omega W \sim$ Gamma(n_1, λ). It is well known that $\rho_k^2 \sim$ Gamma$(N_a, 1)$. $Z_k \sim$ Gamma$(N_a - k + 1, 1)$ follows from the above results and the distribution of $|\hat{r}_{k,k}|^2$ given in *Lemma* 1. Moreover, according to the results in [15], $\rho_k^2\sin^2\theta_{1,k} \sim \delta_{1,k}$Gamma$(N_a - 1, 1)$ and $\xi_k^2\sin^2\theta_{2,k} \sim \delta_{2,k}$Gamma$(N_h - 1, 1)$. With the above discussions, we know $\rho_k^2\sin^2\theta_{1,k}\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2 \sim \delta_{1,k}Y_1$ and $\xi_k^2\sin^2\theta_{2,k}\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2 \sim \delta_{2,k}Y_2$, where Y_1 and Y_2 are two independent gamma random variables, i.e., $Y_1 \sim$ Gamma$(K - 1, 1)$ and $Y_2 \sim$ Gamma$(N_h - K, 1)$. As we will see later, we don't need to obtain $\mathbb{E}\left[\log_2\left(1 + \frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2\right)\right]$ to reach the final result.

Now, we look at the second $\log_2$ term. Let $\mathcal{E}_k$ and $\bar{\mathcal{E}}_k$ denote the events that $\mathcal{E}_k = \left\{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2 \geq |\hat{r}_{k,k}|^2\right\}$ and

$$\begin{aligned}
&\mathbb{E}\Bigg[\log_2\Bigg(\cos^2\theta_{1,1}+\frac{1+\frac{P_a}{K\sigma_{b,1}^2}\rho_1^2\|\tilde{\mathbf{h}}_{b,1}\hat{\mathbf{Q}}^H\|^2}{1+\frac{P_a}{K\sigma_{b,1}^2}\rho_1^2|\hat{r}_{1,1}|^2}\sin^2\theta_{1,1}+\frac{\frac{P_h}{(N_h-K)\sigma_{b,1}^2}\xi_1^2\|\tilde{\mathbf{g}}_{b,1}\hat{\mathbf{\Gamma}}_h\|^2}{1+\frac{P_a}{K\sigma_{b,1}^2}\rho_1^2|\hat{r}_{1,1}|^2}\sin^2\theta_{2,1}\Bigg)\Bigg]\\
\geq\;&\mathbb{E}\Bigg[\log_2\Bigg(\cos^2\theta_{1,1}+\|\tilde{\mathbf{h}}_{b,1}\hat{\mathbf{Q}}^H\|^2\sin^2\theta_{1,1}+\frac{\frac{P_h}{(N_h-K)\sigma_{b,1}^2}\xi_1^2\|\tilde{\mathbf{g}}_{b,1}\hat{\mathbf{\Gamma}}_h\|^2}{1+\frac{P_a}{K\sigma_{b,1}^2}\rho_1^2}\sin^2\theta_{2,1}\Bigg)\Bigg]\\
\geq\;&\mathbb{E}\Bigg[\log_2\left(\cos^2\theta_{1,1}\right)+\log_2\left(1+\|\tilde{\mathbf{h}}_{b,1}\hat{\mathbf{Q}}^H\|^2\frac{\sin^2\theta_{1,1}}{\cos^2\theta_{1,1}}\right)\Bigg]\\
&+\mathbb{E}\Bigg[\log_2\Bigg(1+\frac{\frac{P_h}{(N_h-K)\sigma_{b,1}^2}\xi_1^2\|\tilde{\mathbf{g}}_{b,1}\hat{\mathbf{\Gamma}}_h\|^2}{\left(1+\frac{P_a}{K\sigma_{b,1}^2}\rho_1^2\right)\mathbb{E}\left[\cos^2\theta_{1,1}+\|\tilde{\mathbf{h}}_{b,1}\hat{\mathbf{Q}}^H\|^2\sin^2\theta_{1,1}\right]}\sin^2\theta_{2,1}\Bigg)\Bigg] \qquad (42)\\
\geq\;&\mathbb{E}\Bigg[\log_2\left(\cos^2\theta_{1,1}\right)+\|\tilde{\mathbf{h}}_{b,1}\hat{\mathbf{Q}}^H\|^2\frac{\sin^2\theta_{1,1}}{\cos^2\theta_{1,1}}\Bigg]+\mathbb{E}\Bigg[\log_2\left(1+\frac{Z_1}{\vartheta_{a,1}}+\frac{\delta_{2,1}Y_2}{\vartheta_{h,1}\nu_{1,1}}\right)-\log_2\left(1+\frac{Z_1}{\vartheta_{a,1}}\right)\Bigg] \qquad (43)
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}\Bigg[\log_2\Bigg(\cos^2\theta_{1,k}+\frac{1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2}\sin^2\theta_{1,k}+\frac{\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2}{1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2}\sin^2\theta_{2,k}\Bigg)\Bigg]\\
&\geq \Pr\left\{\mathcal{E}_k\right\}\;\mathbb{E}\Bigg[\log_2\Bigg(1+\frac{\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2}{1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2}\sin^2\theta_{2,k}\Bigg)\Bigg|\mathcal{E}_k\Bigg]\\
&\quad+\Pr\left\{\bar{\mathcal{E}}_k\right\}\mathbb{E}\Bigg[\log_2\Bigg(\cos^2\theta_{1,k}+\frac{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{|\hat{r}_{k,k}|^2}\sin^2\theta_{1,k}+\frac{\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2}{1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2}\sin^2\theta_{2,k}\Bigg)\Bigg|\bar{\mathcal{E}}_k\Bigg]\\
&\geq \Pr\left\{\mathcal{E}_k\right\}\;\mathbb{E}\Bigg[\log_2\Bigg(1+\frac{\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2}{1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2}\sin^2\theta_{2,k}\Bigg)\Bigg|\mathcal{E}_k\Bigg]+\Pr\left\{\bar{\mathcal{E}}_k\right\}\mathbb{E}\Bigg[\log_2\Bigg(\cos^2\theta_{1,k}+\frac{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{|\hat{r}_{k,k}|^2}\sin^2\theta_{1,k}\Bigg)\\
&\quad+\log_2\left(1+\frac{\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2\sin^2\theta_{2,k}}{\left(1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2\right)\mathbb{E}\left[\left(\cos^2\theta_{1,k}+\frac{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{|\hat{r}_{k,k}|^2}\sin^2\theta_{1,k}\right)\big|\bar{\mathcal{E}}_k\right]}\right)\Bigg|\bar{\mathcal{E}}_k\Bigg]. \qquad (44)
\end{aligned}$$

$\bar{\mathcal{E}}_k=\left\{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2<|\hat{r}_{k,k}|^2\right\}$. It is easy to prove that, for $\mathcal{E}_k$, $1\leq\frac{1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2}\leq\frac{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{|\hat{r}_{k,k}|^2}$ and for $\bar{\mathcal{E}}_k$, $\frac{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{|\hat{r}_{k,k}|^2}\leq\frac{1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2}<1$. Thus, for $k=1$, since $|\hat{r}_{1,1}|^2=1$, we have (42) and (43) at the top of this page, where (42) follows from Jensen's inequality and (43) follows from the fact that $\log_2(1+x)\geq x$ for $0\leq x\leq 1$. It was shown in [14] that the PDF of $\cos^2\theta_{1,k}$ is given by $f_{\cos^2\theta_{1,k}}(x)=N_k(N_a-1)\left[1-(1-x)^{N_a-1}\right]^{N_k-1}(1-x)^{N_a-2}$, $0\leq x\leq 1$, where $N_k=2^{B_k}$. Using this PDF and change of variables, after some manipulations, we can obtain $\mathbb{E}\left[\frac{\sin^2\theta_{1,k}}{\cos^2\theta_{1,k}}\right]=2^{B_k}\sum_{m=0}^{N_a-2}\beta\left(\frac{m+N_a}{N_a-1},2^{B_k}-1\right)$. The mean of $\log_2\left(\cos^2\theta_{1,k}\right)$ has been obtained in [32] as $\mathbb{E}\left\{\log_2\left(\cos^2\theta_{1,k}\right)\right\}=-\frac{\log_2(e)}{N_a-1}\sum_{i=1}^{N_a-1}\beta\left(2^{B_k},\frac{i}{N_a-1}\right)$. Using the distribution result in *Lemma* 1, we have $\mathbb{E}_{\mathbf{H}_b,\{\mathcal{W}_k\}}\left\{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2\right\}=\frac{K-1}{N_a-1}$. $\mathbb{E}\Big[\log_2\left(\cos^2\theta_{1,1}\right)+\|\tilde{\mathbf{h}}_{b,1}\hat{\mathbf{Q}}^H\|^2\frac{\sin^2\theta_{1,1}}{\cos^2\theta_{1,1}}\Big]$ can be obtained by combing the above results. Similarly, for $2\leq k\leq K$, we have (44) at the top of this page.

Since $\varepsilon_k=\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2$ is independent with $|\hat{r}_{k,k}|^2$, using the property that $I_y\left(a,b-a+1\right)=\sum_{j=a}^{b}\binom{b}{j}y^j(1-y)^{b-j}$, $\Pr\left\{\bar{\mathcal{E}}_k\right\}$ can be obtained as

$$\begin{aligned}
&\Pr\left\{\bar{\mathcal{E}}_k\right\}=\Pr\{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2\leq|\hat{r}_{k,k}|^2\}\\
=\;&\int_0^1\Pr\{\varepsilon_k\leq y\}f_{|\hat{r}_{k,k}|^2}(y)\mathrm{d}y\\
=\;&\int_0^1 I_y\left(K-1,N_a-K\right)\frac{y^{N_a-k}(1-y)^{k-2}}{\beta(N_a-k+1,k-1)}\mathrm{d}y\\
=\;&\int_0^1\sum_{j=K-1}^{N_a-2}\binom{N_a-2}{j}y^j(1-y)^{N_a-2-j}\\
&\times\frac{y^{N_a-k}(1-y)^{k-2}}{\beta(N_a-k+1,k-1)}\mathrm{d}y=A_k, \qquad (45)
\end{aligned}$$

which is given by (12). Then, $\Pr\left\{\mathcal{E}_k\right\}=1-\Pr\left\{\bar{\mathcal{E}}_k\right\}$. Similarly, for $k\geq 2$, we can obtain

$$\begin{aligned}
\mathbb{E}\left[\frac{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{|\hat{r}_{k,k}|^2}\big|\bar{\mathcal{E}}_k\right]&=\int_0^1\int_0^y\frac{x}{y}f_{\varepsilon_k}(x)f_{|\hat{r}_{k,k}|^2}(y)\mathrm{d}x\mathrm{d}y\\
=\;&\int_0^1\left(\int_0^y\frac{x}{y}\frac{x^{K-2}(1-x)^{N_a-K-1}}{\beta(K-1,N_a-K)}\mathrm{d}x\right)
\end{aligned}$$

$$
\begin{aligned}
&\mathbb{E}\left[\log_2\left(1+\frac{\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2\sin^2\theta_{2,k}}{\left(1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2\right)\left(\cos^2\theta_{1,k}+\frac{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{|\hat{r}_{k,k}|^2}\sin^2\theta_{1,k}\right)}\right)\Bigg|\bar{\mathcal{E}}_k\right] \\
\geq\;& \mathbb{E}\left[\log_2\left(1+\frac{\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2\sin^2\theta_{2,k}}{\left(1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2\right)\mathbb{E}\Big[\left(\cos^2\theta_{1,k}+\frac{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{|\hat{r}_{k,k}|^2}\sin^2\theta_{1,k}\right)\big|\bar{\mathcal{E}}_k\Big]}\right)\right] \qquad (48)\\
=\;& \mathbb{E}\Bigg[\log_2\left(1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2+\frac{P_h}{\nu_{1,k}(N_h-K)\sigma_{b,k}^2}\xi_k^2\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2\sin^2\theta_{2,k}\right)-\log_2\left(1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2\right)\Bigg] \\
=\;& \mathbb{E}\Bigg[\log_2\left(1+\frac{1}{\vartheta_{a,k}}Z_k+\frac{\delta_{2,k}}{\vartheta_{h,k}\nu_{1,k}}Y_2\right)\Bigg]-\mathbb{E}\Bigg[\log_2\left(1+\frac{1}{\vartheta_{a,k}}Z_k\right)\Bigg], \qquad (49)
\end{aligned}
$$

$$
\begin{aligned}
&\times\frac{y^{N_a-k}(1-y)^{k-2}}{\beta(N_a-k+1,k-1)}\mathrm{d}y \\
=\;& \frac{\beta(K,N_a-K)}{\beta(K-1,N_a-K)}\int_0^1 I_y\left(K,N_a-K\right) \\
&\times\frac{y^{N_a-k-1}(1-y)^{k-2}}{\beta(N_a-k+1,k-1)}\mathrm{d}y \\
=\;& \frac{\beta(K,N_a-K)}{\beta(K-1,N_a-K)}\sum_{j=K}^{N_a-1}\binom{N_a-1}{j}\int_0^1 y^j \\
&\times(1-y)^{N_a-1-j}\frac{y^{N_a-k-1}(1-y)^{k-2}}{\beta(N_a-k+1,k-1)}\mathrm{d}y = C_k
\end{aligned}
$$

which is given by (13). In addition, according to [14], we have $\mathbb{E}_{\mathbf{h}_{b,k},\mathcal{W}_k}\left\{\cos^2\theta_{1,k}\right\} = 1-2^{B_k}\beta\left(2^{B_k},\frac{N_a}{N_a-1}\right)$. Combining the above results, it is easy to obtain $\nu_{1,k}$ given in the main context. Moreover, still using the property that $\log_2(1+x)\geq x$ for $0\leq x\leq 1$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\log_2\left(\cos^2\theta_{1,k}+\frac{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{|\hat{r}_{k,k}|^2}\sin^2\theta_{1,k}\right)\Bigg|\bar{\mathcal{E}}_k\right] \\
&\geq \mathbb{E}\left[\log_2\left(\cos^2\theta_{1,k}\right)\right]+\mathbb{E}\left[\frac{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{|\hat{r}_{k,k}|^2}\Bigg|\bar{\mathcal{E}}_k\right] \\
&\times\mathbb{E}\left[\frac{\sin^2\theta_{1,k}}{\cos^2\theta_{1,k}}\right], \qquad (46)
\end{aligned}
$$

which can be expressed using the obtained results above. In addition, we have

$$
\begin{aligned}
&\mathbb{E}\Bigg[\log_2\left(1+\frac{\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2}{1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2}\sin^2\theta_{2,k}\right)\Bigg|\mathcal{E}_k\Bigg] \\
=\;& \mathbb{E}\Bigg[\log_2\left(1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2+\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\right. \\
&\left.\times\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2\sin^2\theta_{2,k}\right)\Bigg]-\mathbb{E}\Bigg[\log_2\left(1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2\right)\Bigg] \\
=\;& \mathbb{E}\Bigg[\log_2\left(1+\frac{1}{\vartheta_{a,k}}Z_k+\frac{\delta_{2,k}}{\vartheta_{h,k}}Y_2\right)\Bigg] \\
&-\mathbb{E}\Bigg[\log_2\left(1+\frac{1}{\vartheta_{a,k}}Z_k\right)\Bigg], \qquad (47)
\end{aligned}
$$

and (48) (49) shown at the top of this page, where (48) follows from Jensen's inequality.

For the general case that $\frac{\delta_{1,k}}{\vartheta_{a,k}}\neq\frac{\delta_{2,k}}{\vartheta_{h,k}}$, using the result in [33], the PDF of $J_k = \frac{P_a}{K\sigma_{b,k}^2}\rho_k^2\sin^2\theta_{1,k}\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2 + \frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\sin^2\theta_{2,k}\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2 \overset{d.}{=} \frac{\delta_{1,k}}{\vartheta_{a,k}}Y_1 + \frac{\delta_{2,k}}{\vartheta_{h,k}}Y_2$ is given by $f_{J_k}(y) = \prod_{i=1}^2\frac{1}{\zeta_i^{m_i}}\sum_{j=1}^2\sum_{l=1}^{m_j}\frac{\Xi_{j,l}(\mathbf{m},\boldsymbol{\zeta})}{(m_j-l)!}y^{m_j-l}e^{-\frac{y}{\zeta_j}}U(y)$ with $\mathbf{m} = (m_1,m_2) = \mathbf{m}_2 = (K-1,N_h-K)$ and $\boldsymbol{\zeta} = (\zeta_1,\zeta_2) = \boldsymbol{\zeta}_3 = \left(\frac{\delta_{1,k}}{\vartheta_{a,k}},\frac{\delta_{2,k}}{\vartheta_{h,k}}\right)$, where $U(y)$ is Heaviside step function. Thus,

$$
\begin{aligned}
&\mathbb{E}\Bigg[\log_2\left(1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2\sin^2\theta_{1,k}\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2\right. \\
&\left.+\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\sin^2\theta_{2,k}\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2\right)\Bigg] \qquad (50)\\
=\;& \int_0^\infty\log_2\left(1+y\right)f_{J_k}(y)\mathrm{d}y \\
=\;& \prod_{i=1}^2\frac{1}{\zeta_i^{m_i}}\sum_{j=1}^2\sum_{l=1}^{m_j}\frac{\Xi_{j,l}\left(\mathbf{m}_2,\boldsymbol{\zeta}_3\right)}{(m_j-l)!} \\
&\times\int_0^\infty\log_2\left(1+y\right)y^{m_j-l}e^{-\frac{y}{\zeta_j}}\mathrm{d}y \\
=\;& \log_2(e)\prod_{i=1}^2\frac{1}{\zeta_i^{m_i}}\sum_{j=1}^2\sum_{l=1}^{m_j}\Xi_{j,l}\left(\mathbf{m}_2,\boldsymbol{\zeta}_3\right)\zeta_j^{m_j-l+1} \\
&\times\exp\left(\frac{1}{\zeta_j}\right)\sum_{q=1}^{m_j-l+1}E_q\left(\frac{1}{\zeta_j}\right) = H\left(\mathbf{m}_2,\boldsymbol{\zeta}_3\right). \qquad (51)
\end{aligned}
$$

For the special case that $\zeta_1 = \zeta_2 = \frac{\delta_{1,k}}{\vartheta_{a,k}} = \frac{\delta_{2,k}}{\vartheta_{h,k}}$, $J_k \overset{d.}{=} \frac{\delta_{1,k}}{\vartheta_{a,k}}(Y_1+Y_2)$. Since $Y_1+Y_2\sim$ Gamma$(m_1+m_2,1)$, (50) is given by

$$
\begin{aligned}
&\int_0^\infty\log_2\left(1+\frac{\delta_{1,k}}{\vartheta_{a,k}}y\right)\frac{y^{m_1+m_2-2}e^{-y}}{\Gamma(m_1+m_2-1)}\mathrm{d}y \\
=\;& \log_2(e)\exp\left(\frac{\vartheta_{a,k}}{\delta_{1,k}}\right)\sum_{i=1}^{m_1+m_2-1}E_i\left(\frac{\vartheta_{a,k}}{\delta_{1,k}}\right) \qquad (52)\\
=\;& H\left(\mathbf{m}_2,\boldsymbol{\zeta}_3\right)
\end{aligned}
$$

where (51) and (52) follow from the result in [34] that

$$
\begin{aligned}
&\int_0^\infty\ln\left(1+ax\right)\exp\left(-\mu x\right)x^{q-1}\mathrm{d}x \\
=\;& \frac{(q-1)!\exp\left(\frac{\mu}{a}\right)}{\mu^q}\sum_{i=1}^q E_i\left(\frac{\mu}{a}\right). \qquad (53)
\end{aligned}
$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2018.2864178, IEEE Transactions on Vehicular Technology

Similarly, the first expectations in (49) and (47) can be respectively obtained as $\mathbb{E}\left[\log_2\left(1+\frac{1}{\vartheta_{a,k}}Z_k+\frac{\delta_{2,k}}{\vartheta_{h,k}}Y_2\right)\right]=H\left(\mathbf{m}_1,\boldsymbol{\zeta}_1\right)$ and $\mathbb{E}\left[\log_2\left(1+\frac{1}{\vartheta_{a,k}}Z_k+\frac{\delta_{2,k}}{\vartheta_{h,k}\nu_{1,k}}Y_2\right)\right]=H\left(\mathbf{m}_1,\boldsymbol{\zeta}_2\right)$. (10) follows by substituting all above results in (41).

C. Proof of Lemma 2

Firstly, we have $I\left(\mathbf{x};\mathbf{y}_e^{lfb}\right)=H(\mathbf{y}_e^{lfb})-H(\mathbf{y}_e^{lfb}|\mathbf{x})$, and

$$H(\mathbf{y}_e^{lfb})\leq \log_2\det\left(\frac{P_a}{K}\hat{\mathbf{H}}_e\hat{\mathbf{H}}_e^H+\frac{P_h}{N_h-K}\hat{\mathbf{G}}_e\hat{\mathbf{G}}_e^H+\sigma_e^2\mathbf{I}\right),\tag{54}$$

where the inequality is due to the fact that, with the average power constraint of $\mathbf{x}$, the maximum differential entropy of $\mathbf{y}_e^{lfb}$ is achieved if and only if $\mathbf{x}$ is a circularly symmetric complex Gaussian vector ($\mathbf{u}$ is already circularly symmetric complex Gaussian distributed). For the same reason, we have

$$H(\mathbf{y}_e^{lfb}|\mathbf{x})=\log_2\det\left(\frac{P_h}{N_h-K}\hat{\mathbf{G}}_e\hat{\mathbf{G}}_e^H+\sigma_e^2\mathbf{I}\right).\tag{55}$$

(16) follows by combing (54) and (55).

According to (3), given the quantized CSI $\hat{\mathbf{H}}_b$, $\mathbf{x}$ can be uniquely determined from $\mathbf{s}$ by THP and *vice versa*. Thus, we can denote the THP operation as a *one-to-one* mapping T from $(\mathbf{s},\hat{\mathbf{H}}_b)$ to $\mathbf{x}$, which is given as $\mathbf{x}=T(\mathbf{s},\hat{\mathbf{H}}_b)$ and $\mathbf{s}=T^{-1}(\mathbf{s},\hat{\mathbf{H}}_b)$. According to data-processing inequality [35], $I\left(\mathbf{x};\mathbf{y}_e^{lfb}\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)=I\left(\mathbf{s};\mathbf{y}_e^{lfb}\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)$. In fact, we can prove when Eve knows $\hat{\mathbf{H}}_b$ and also knows the processing method of Alice, $\mathbf{s},\mathbf{x},\mathbf{y}_e^{lfb}$ form a Markov chain and *vice versa*, i.e., $\mathbf{s}\leftrightarrow\mathbf{x}\leftrightarrow\mathbf{y}_e^{lfb}$ and $\mathbf{x}\leftrightarrow\mathbf{s}\leftrightarrow\mathbf{y}_e^{lfb}$ as follows. First, since $p(\mathbf{y}_e^{lfb}|\mathbf{s},\hat{\mathbf{H}}_b,\mathbf{x})=p(\mathbf{y}_e^{lfb}|\mathbf{s},\hat{\mathbf{H}}_b)$, we have $p(\mathbf{x},\mathbf{y}_e^{lfb}|\mathbf{s},\hat{\mathbf{H}}_b)=p(\mathbf{y}_e^{lfb}|\mathbf{x},\mathbf{s},\hat{\mathbf{H}}_b)p(\mathbf{x}|\mathbf{s},\hat{\mathbf{H}}_b)=p(\mathbf{y}_e^{lfb}|\mathbf{s},\hat{\mathbf{H}}_b)p(\mathbf{x}|\mathbf{s},\hat{\mathbf{H}}_b)$, which means $\mathbf{x}$ and $\mathbf{y}_e^{lfb}$ are conditionally independent given $\mathbf{s}$ and $\hat{\mathbf{H}}_b$. In fact, given $\mathbf{s}$ and $\hat{\mathbf{H}}_b$ the conditional probability mass function of $\mathbf{x}$ is $p(\mathbf{x}|\mathbf{s},\hat{\mathbf{H}}_b)=\begin{cases}1 & \mathbf{x}=T(\mathbf{s},\hat{\mathbf{H}}_b)\\ 0 & \mathbf{x}\neq T(\mathbf{s},\hat{\mathbf{H}}_b)\end{cases}$. Thus, $\mathbf{x}\leftrightarrow\mathbf{s}\leftrightarrow\mathbf{y}_e^{lfb}$ holds. Similarly, it is easy to see $p(\mathbf{s},\mathbf{y}_e^{lfb}|\mathbf{x},\hat{\mathbf{H}}_b)=p(\mathbf{y}_e^{lfb}|\mathbf{s},\hat{\mathbf{H}}_b,\mathbf{x})p(\mathbf{s}|\mathbf{x},\hat{\mathbf{H}}_b)=p(\mathbf{y}_e^{lfb}|\hat{\mathbf{H}}_b,\mathbf{x})p(\mathbf{s}|\mathbf{x},\hat{\mathbf{H}}_b)$, which means $\mathbf{s}$ and $\mathbf{y}_e^{lfb}$ are conditionally independent given $\mathbf{x}$ and $\hat{\mathbf{H}}_b$. Thus, $\mathbf{s}\leftrightarrow\mathbf{x}\leftrightarrow\mathbf{y}_e^{lfb}$ holds.

Using the properties of mutual information and differential entropy [35], we have

$$\begin{aligned}&I\left(\mathbf{s};\mathbf{y}_e^{lfb}\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)\\&=\sum_{i=1}^{K}I\left(s_i;\mathbf{y}_e^{lfb}\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b,s_1,s_2,\cdots,s_{i-1}\right)\\&=\sum_{i=1}^{K}\Big[H\left(s_i\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b,s_1,s_2,\cdots,s_{i-1}\right)\\&-H\left(s_i\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b,\mathbf{y}_e^{lfb},s_1,\cdots,s_{i-1}\right)\Big].\end{aligned}\tag{56}$$

Since s_k for $k=1,2,\cdots,K$, $\hat{\mathbf{H}}_b$ and $\hat{\mathbf{G}}_b$ are independent with each other, we conclude that $H(s_i|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b,s_1,s_2,\cdots,s_{i-1})=H\left(s_i\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)=H\left(s_i\right)$ and $H\left(s_i\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b,\mathbf{y}_e^{lfb},s_1,s_2,\cdots,s_{i-1}\right)=H\left(s_i\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b,\mathbf{y}_e^{lfb}\right)$. Combing these results with (56), we have $I\left(\mathbf{s};\mathbf{y}_e^{lfb}\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)=\sum_{k=1}^{K}I\left(s_k;\mathbf{y}_e^{lfb}\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)$. In addition, it is obvious that the order of s_k $(k=1,2,\cdots,K)$ in $\mathbf{s}$ does not change the distribution of $\mathbf{x}$, thus does not change $I\left(\mathbf{x};\mathbf{y}_e^{lfb}\right)$ and $I\left(\mathbf{s};\mathbf{y}_e^{lfb}\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)$. Thus, we conclude that $I\left(s_k;\mathbf{y}_e\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)$ for all k are equal.

Moreover, $\mathbf{y}_e^{lfb}$ depends on $\hat{\mathbf{H}}_b$ through $\hat{\mathbf{Q}}$ from LQ decomposition of $\hat{\mathbf{H}}_b$ and also $\mathbf{x}$, and $\mathbf{x}$ is determined from $(\hat{\mathbf{H}}_b,\mathbf{s})$. Given $\mathbf{y}_e^{lfb}$, $\hat{\mathbf{H}}_b$ and $\mathbf{s}$ are related through $\mathbf{y}_e^{lfb}$ given by (6). Thus, $\mathbf{s}$ implicitly becomes conditionally dependent with $\hat{\mathbf{H}}_b$. Then, we have $H\left(\mathbf{s}|\mathbf{y}_e^{lfb}\right)>H\left(\mathbf{s}|\mathbf{y}_e^{lfb},\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)$, since condition reduces entropy [35]. Thus, we can prove $I\left(\mathbf{s};\mathbf{y}_e^{lfb}\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)>I\left(\mathbf{s};\mathbf{y}_e^{lfb}\right)$ as follows.

$$\begin{aligned}&I\left(\mathbf{s};\mathbf{y}_e^{lfb}\middle|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)\\=\;&H\left(\mathbf{s}|\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)-H\left(\mathbf{s}|\mathbf{y}_e^{lfb},\hat{\mathbf{H}}_b,\hat{\mathbf{G}}_b\right)\\>\;&H\left(\mathbf{s}\right)-H\left(\mathbf{s}|\mathbf{y}_e^{lfb}\right)=I\left(\mathbf{s};\mathbf{y}_e^{lfb}\right).\end{aligned}\tag{57}$$

D. Proof of Theorem *3*

It is known that $\gamma_k\sim\chi^2_{N_a-k+1}$, or equivalently $\sim$ Gamma(N_a-k+1) [36]. According to the proof of *Theorem* 1, $\gamma_k\overset{d.}{=}\rho_k^2|\hat{r}_{k,k}|^2$. Thus, it follows that $R_{b,k}^{per}=\mathbb{E}\Big[\log_2\big(1+\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2|\hat{r}_{k,k}|^2\big)\Big]$. Combining this fact with (41), (42) and (43), and using Jensen's inequality, for $k=1$, we have

$$\begin{aligned}\Delta R_{sec,1}\leq&\log_2\Bigg(1+\mathbb{E}\bigg[\frac{P_a}{K\sigma_{b,1}^2}\rho_1^2\sin^2\theta_{1,1}\|\tilde{\mathbf{h}}_{b,1}\hat{\mathbf{Q}}^H\|^2\\+&\frac{P_h}{(N_h-K)\sigma_{b,1}^2}\xi_1^2\sin^2\theta_{2,1}\|\tilde{\mathbf{g}}_{b,1}\hat{\mathbf{\Gamma}}_h\|^2\bigg]\Bigg)\\-&\mathbb{E}\left[\log_2\left(\cos^2\theta_{1,1}\right)+\|\tilde{\mathbf{h}}_{b,1}\hat{\mathbf{Q}}^H\|^2\frac{\sin^2\theta_{1,1}}{\cos^2\theta_{1,1}}\right],\end{aligned}\tag{58}$$

where we have omitted the term related to $\sin^2\theta_{2,1}$ in (43). Similarly, for $k\geq 2$, combing (41), (44) and (46), and using Jensen's inequality, we have

$$\begin{aligned}\Delta R_{sec,k}\leq&\log_2\Bigg(1+\mathbb{E}\bigg[\frac{P_a}{K\sigma_{b,k}^2}\rho_k^2\sin^2\theta_{1,k}\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2\\&+\frac{P_h}{(N_h-K)\sigma_{b,k}^2}\xi_k^2\sin^2\theta_{2,k}\|\tilde{\mathbf{g}}_{b,k}\hat{\mathbf{\Gamma}}_h\|^2\bigg]\Bigg)-\Pr\left\{\bar{\mathcal{E}}_k\right\}\\&\Bigg\{\mathbb{E}\left[\log_2\left(\cos^2\theta_{1,k}\right)\right]+\mathbb{E}\left[\frac{\|\tilde{\mathbf{h}}_{b,k}\hat{\mathbf{Q}}^H\|^2}{|\hat{r}_{k,k}|^2}\middle|\bar{\mathcal{E}}_k\right]\\&\times\mathbb{E}\left[\frac{\sin^2\theta_{1,k}}{\cos^2\theta_{1,k}}\right]\Bigg\},\end{aligned}\tag{59}$$

The final result follows from substituting the existing results in Appendix B into (58) and (59).

E. Proof of Algorithm *1*

First, we can rewrite $\tilde{\Delta}R_{sec,k}$ as $\tilde{\Delta}R_{sec,k}=\log_2\left(\beta_k2^{-\frac{B_t}{N_h-1}}2^{\left(\frac{1}{N_a-1}+\frac{1}{N_h-1}\right)B_k}+2^{\frac{B_k}{N_a-1}}+\alpha_k\right)-\frac{B_k}{N_a-1}+\varsigma_k2^{-\frac{B_k}{N_a-1}}$. Notice that, treating B_k,D_k as real numbers, $\tilde{\Delta}R_{sec,k}$ is continuous and differentiable

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2018.2864178, IEEE Transactions on Vehicular Technology

$$\frac{d\tilde{\Delta}R_{sec,k}}{dB_k} = \frac{\frac{\beta_k}{N_h-1}2^{-\frac{B_t}{N_h-1}}2^{\left(\frac{2}{N_a-1}+\frac{1}{N_h-1}\right)B_k} - \frac{\ln(2)\beta_k\varsigma_k}{N_a-1}2^{-\frac{B_t}{N_h-1}}2^{\left(\frac{1}{N_a-1}+\frac{1}{N_h-1}\right)B_k} - \frac{\ln(2)\varsigma_k+\alpha_k}{N_a-1}2^{\frac{B_k}{N_a-1}} - \frac{\ln(2)\alpha_k\varsigma_k}{N_a-1}}{\left[\beta_k 2^{-\frac{B_t}{N_h-1}}2^{\left(\frac{1}{N_a-1}+\frac{1}{N_h-1}\right)B_k} + 2^{\frac{B_k}{N_a-1}} + \alpha_k\right]2^{\frac{B_k}{N_a-1}}}. \tag{60}$$

in B_k. It obvious that $-\frac{B_k}{N_a-1} + \varsigma_k 2^{-\frac{B_k}{N_a-1}}$ is convex function of B_k. Let $L_k(B_k) = \log_2\left(\beta_k 2^{-\frac{B_t}{N_h-1}}2^{\left(\frac{1}{N_a-1}+\frac{1}{N_h-1}\right)B_k} + 2^{\frac{B_k}{N_a-1}} + \alpha_k\right)$. In the following, we prove the term $L_k(B_k)$ is convex in B_k. First, after some manipulations we can obtain

$$\frac{1}{L_k'(B_k)} = \frac{(N_a-1)(N_h-1)}{N_a+N_h-2} + \frac{\frac{N_a-1}{N_a+N_h-2} + \alpha_k 2^{-\frac{B_k}{N_a-1}}}{\frac{N_a+N_h-2}{(N_a-1)(N_h-1)}\beta_k 2^{-\frac{B_t}{N_h-1}}2^{\frac{B_k}{N_h-1}} + \frac{1}{N_a-1}}.$$

It is easy to see that $L_k'(B_k) > 0$ and is a monotonic increasing function of B_k, or equivalently $L_k''(B_k) > 0$. Thus, $L_k(B_k)$ is a convex function of B_k. It follows that $\tilde{\Delta}R_{sec,k}$ is a convex function of B_k. Thus, $\frac{d\tilde{\Delta}R_{sec,k}}{dB_k}$ is a monotonic increasing function of B_k, which is obtained as (60) shown at the top of this page. The value of B_k that minimizes $\tilde{\Delta}R_{sec,k}$ will be a single global optimal value and is obtained by equating the numerator of (60) to zero (Since the denominator is positive.). Observing the numerator of (60), we find, for the case that $N_a = N_h$, the unique global optimal value B_k can be obtained by solving the cubic equation (38) with $x = 2^{\frac{B_k}{N_a-1}} = 2^{\frac{B_k}{N_h-1}} > 0$ for real number B_k. $\frac{d\tilde{\Delta}R_{sec,k}}{dB_k} = 0$ has a unique real solution. The practical optimal bit allocation can be obtained as described in Algorithm 1. Similarly, for the general case that $N_a \neq N_h$, the unique global optimal value B_k can be obtained by first solving (39) with $y = 2^{B_k} > 0$ and then following the same method for the case $N_a = N_h$.

References

[1] S. Goel and R. Negi, "Guaranteeing secrecy using artificial noise," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2180–2189, Jun. 2008.

[2] G. Geraci, M. Egan, J. Yuan, A. Razi, and I. B. Collings, "Secrecy sum-rates for multi-user MIMO regularized channel inversion precoding," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3472–3482, Nov. 2012.

[3] J. Huang, A. Mukherjee, and A. L. Swindlehurst, "Secure communication via an untrusted non-regenerative relay in fading channels," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2536–2550, Jan. 2013.

[4] M. Pei, J. Wei, K. K. Wong, and X. Wang, "Masked beamforming for multiuser MIMO wiretap channels with imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 544–549, Feb. 2012.

[5] S. C. Lin, T. H. Chang, Y. L. Liang, Y. W. Hong, and C. Y. Chi, "On the impact of quantized channel feedback in guaranteeing secrecy with artificial noise: The noise leakage problem," *IEEE Trans. Wireless Commun.*, vol. 10, no. 3, pp. 901–915, Mar. 2011.

[6] M. Pei, A. L. Swindlehurst, D. Ma, and J. Wei, "Adaptive limited feedback for MISO wiretap channels with cooperative jamming," *IEEE Trans. Signal Process.*, vol. 62, no. 4, pp. 993–1004, Feb. 2014.

[7] X. Zhang, M. R. McKay, X. Zhou, and R. W. Heath Jr., "Artificial-noise-aided secure multi-antenna transmission with limited feedback," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2742–2754, May 2015.

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

[9] I. Csisźar and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 24, no. 3, pp. 339–348, May. 1978.

[10] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305–1316, Jul. 2004.

[11] R. F. H. Fischer, *Precoding and Signal Shaping for Digital Transmission*, 1st ed. USA: New York: Wiley, 2002.

[12] L. Sun and M. Lei, "Quantized CSI-based Tomlinson-Harashima precoding in multiuser MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1118–1126, March 2013.

[13] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inform. Theory*, vol. 52, pp. 5045–5060, Nov. 2006.

[14] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. Wireless Commun.*, vol. 6, pp. 458–462, Feb. 2007.

[15] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1478–1491, Sep. 2007.

[16] A. Mukherjee and A. L. Swindlehurst, "Robust beamforming for security in MIMO wiretap channels with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 351–361, Jan. 2011.

[17] L. Zhang, Y. Cai, B. Champagne, and M. Zhao, "Tomlinson-Harashima precoding design in MIMO wiretap channels based on the MMSE criterion," in *Proc. of 2015 IEEE International Conference on Communication Workshop (ICCW)*, 2015, pp. 470–474.

[18] X. Lu, R. C. de Lamare, and K. Zu, "Successive Optimization Tomlinson-Harashima Precoding Strategies for Physical-Layer Security in Wireless Networks," *Eurasip Journal on Wireless Communications and Networking*, Oct. 2016.

[19] X. Lu, K. Zu, and R. Lamare, "Lattice-reduction aided successive optimization Tomlinson-Harashima precoding strategies for physical-layer security in wireless networks," in *Proc. of Sensor Signal Processing for Defence (SSPD)*, 2014, pp. 1–5.

[20] L. Sun, R. W. Wang, and V. C. Leung, "Artificial-noise-aided nonlinear secure transmission for MU-MISO wiretap channel with quantized CSIT," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, May 2017.

[21] L. Sun, R. Wang, H. Wang, and V. C. Leung, "A novel nonlinear secure transmission design for MU-MISO systems with limited feedback," in *Proc. of IEEE International Conference on Communications (ICC)*, May 2016.

[22] I. Slim, A. Mezghani, and J. A. Nossek, "Quantized CDI based Tomlinson Harashima precoding for broadcast channels," in *Proc. IEEE Int. Conf. on Commun. (ICC)*, Jun. 2011, pp. 1–5.

[23] M. O. Damen, H. E. Gamal, and G. Caire, "On maximum likelihood detection and the search for the closest lattice point," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2389–2402, Oct. 2003.

[24] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple antenna systems," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2562–2579, Oct. 2003.

[25] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th ed. New York: Academic, 2000.

[26] M. Chiani, M. Z. Win, and H. Shin, "MIMO Networks: The effects of interference," *IEEE Trans. Inform. Theory*, vol. 56, no. 1, pp. 993–1004, Feb. 2010.

[27] M. A. Girshick, "On the sampling theory of roots of determinantal equations," *Annals of Math. Statistics*, vol. 10, pp. 203–204, 1939.

[28] W. Gautschi, "Some elementary inequalities relating to the gamma and incomplete gamma function," *J. Math. Phys.*, vol. 38, pp. 77–81, 1959.

[29] T. L. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[30] M. G. Kendall, *A Course in the Geometry of N-dimensions*, 1st ed. London, U.K.: Charles Griffin Co., Ltd., 1961.

[31] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous univariate distributions, vol 1 and vol. 2*, 2nd ed. Wiley series in probability and mathematical statistics, New York: Wiley and Sons, 1994.

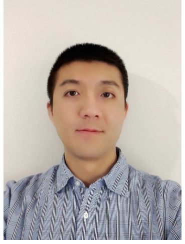
[32] R. Bhagavatula and R. W. Heath, “Adaptive bit partitioning for multicell intercell interference nulling with delayed limited feedback,” *IEEE Trans. Signal Process.*, vol. 59, no. 8, pp. 3824–3836, Aug. 2011.

[33] S. Amari and R. Misra, “Closed-from expressions for distribution of sum of exponential random variables,” *IEEE Trans. on Reliability*, vol. 46, pp. 519–522, 1997.

[34] M. S. Alouini and A. Goldsmith, “Capacity of Rayleigh-fading channels under different adaptive transmission and diversity techniques,” *IEEE Trans. Veh. Technol.*, vol. 48, pp. 1165–1181, Jul. 1999.

[35] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 3rd ed. New York: Wiley, 2006.

[36] A. K. Gupta and D. K. Nagar, *Matrix variate distributions.* in Monographs and Surveys in pure and applied mathematics, Boca Raton, FL: Chapman and Hall/CRC, 2000.

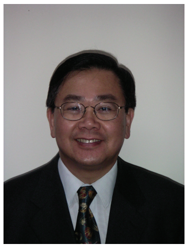


Liang Sun (S’08, M’12) received Ph.D. degree in Electrical Engineering from the Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2010. From 2010 to 2014, he was a researcher with the Alcatel-Lucent Shanghai Bell and the NEC Corporation. From September 2014 to December 2015, he had been a postdoctoral research fellow with the Department of Electrical and Computer Engineering, the University of British Columbia, Vancouver, BC, Canada. Since 2016, he has joined the Beihang University as associate professor. He also held visiting positions at the Waseda University, Japan, and the South University of Science and Technology of China, Shenzhen, China. His research interests include information theory and signal processing; in particular design and analysis of communication systems, MIMO, multiuser and heterogenous wireless networks, cooperative communications, limited feedback techniques, and cognitive systems.

Dr. Sun received a 2010 Young Author Best Paper Award by the IEEE Signal Processing Society and Hong Kong Telecom Institute of Information Technology Post-Graduate Excellence Scholarships.

Rui Wang received his Bachelor’s Degree at the University of Science and Technology of China (USTC) in 2004. Then he got Ph.D. degree in wireless communications at the Hong Kong University of Science and Technology (HKUST) in 2008. From 2009 to 2012, he was a senior research engineer in Huawei Technologies, Co., Ltd. Since 2012, he has joined the Southern University of Science and Technology (SUSTech) as an associate professor. He is also serving as the Vice-Chair of IEEE Shenzhen Joint SPS-Comsoc Chapter. Dr. Wang has research experience in both academia and industry. He has published over 20 papers on the top level IEEE journals in the areas of wireless radio resource optimization, cellular interference management, stochastic optimization of information and communication systems and etc. Moreover, he also participated in the development of 5G systems, and has contributed more than 20 US patent application and 50 Chinese patent application (30 of them have been granted).

Wei Wang (S’10-M’16) is currently a professor in School of Electronic Information and Communications, Huazhong University of Science and Technology. He received his Ph.D. degree in Department of Computer Science and Engineering from Hong Kong University of Science and Technology. He served as editors of KSII Transactions on Internet and Information Systems, International Journal of Communication Systems China Communications, guest editors of Wireless Communications and Mobile Computing, IEEE COMSOC MMTC Communications, and TPC of INFOCOM, GBLOBECOM, etc. His research interests include PHY/MAC designs and mobile computing in wireless systems.



Victor C. M. Leung (S75, M89, SM97, F03) received the B.A.Sc. (Hons.) degree in electrical engineering from the University of British Columbia (UBC) in 1977, when he was awarded the APEBC Gold Medal as the head of the graduating class in the Faculty of Applied Science. He attended graduate school at UBC on a Canadian Natural Sciences and Engineering Research Council Postgraduate Scholarship and received the Ph.D. degree in electrical engineering in 1982.

From 1981 to 1987, Dr. Leung was a Senior Member of Technical Staff and satellite system specialist at MPR Teltech Ltd., Canada. In 1988, he was a Lecturer in the Department of Electronics at the Chinese University of Hong Kong. He returned to UBC as a faculty member in 1989, and currently holds the positions of Professor and TELUS Mobility Research Chair in Advanced Telecommunications Engineering in the Department of Electrical and Computer Engineering. Dr. Leung has co-authored more than 1100 journal/conference papers, 40 book chapters, and co-edited 14 book titles. Several of his papers had been selected for best paper awards. His research interests are in the broad areas of wireless networks and mobile systems.

Dr. Leung is a registered Professional Engineer in the Province of British Columbia, Canada. He is a Fellow of IEEE, the Royal Society of Canada, the Engineering Institute of Canada, and the Canadian Academy of Engineering. He was a Distinguished Lecturer of the IEEE Communications Society. He is serving on the editorial boards of the IEEE Transactions on Green Communications and Networking, IEEE Transactions on Cloud Computing, IEEE Network, IEEE Access, Computer Communications, and several other journals, and has previously served on the editorial boards of the IEEE Journal on Selected Areas in Communications Wireless Communications Series and Series on Green Communications and Networking, IEEE Transactions on Wireless Communications, IEEE Transactions on Vehicular Technology, IEEE Transactions on Computers, IEEE Wireless Communications Letters, and Journal of Communications and Networks. He has guest-edited many journal special issues, and provided leadership to the organizing committees and technical program committees of numerous conferences and workshops. He received the IEEE Vancouver Section Centennial Award, the 2011 UBC Killam Research Prize, the 2017 Canadian Award for Telecommunications Research, and the 2018 IEEE TGCC Distinguished Technical Achievement Recognition Award. He co-authored papers that won the 2017 IEEE ComSoc Fred W. Ellersick Prize, the 2017 IEEE Systems Journal Best Paper Award, and the 2018 IEEE CSIM Best Journal Paper Award.