Artificial-Noise-Aided Nonlinear Secure Transmission for Multiuser Multi-Antenna Systems with Finite-Rate Feedback

Liang Sun, Member, IEEE, Rui Wang, Member, IEEE, Zhiyao Tang, and Victor C. M. Leung, Fellow, IEEE

Abstract—We consider low-complexity transceiver design for secure communications of multiuser multi-antenna system with an external multi-antenna eavesdropper. We propose to employ Tomlinson Harashima precoder (THP) to simultaneously transmit message-bearing signals and artificial noise (AN) according to the quantized channel state information (CSI). Based on random vector quantization of the channel vectors, we obtain analytical approximations of the ergodic secrecy rate (ESR) of each legitimate receiver (LR) and the ergodic secrecy sum rate of the system with arbitrary system parameters. Based on the obtained analytical results, the near-optimal power allocation to the information signals and AN can be obtained using numerical method. We show using information-theoretical method that, besides the advantage in the ESR over the linear precoding scheme, THP can reduce the supported rate of the eavesdropper’s channel by preventing the eavesdropper from obtaining the legitimate channels’ (quantized) CSI. The ESR loss of each LR compared to the perfect CSI case will increase without bound for a fixed number of feedback bits. We also derive a feedback bit scaling law to solve this problem. Finally, numerical results are provided to verify our analytical results and also the advantage of the proposed secure nonlinear transceiver over the corresponding linear scheme.

Index Terms—Physical layer security, artificial noise, nonlinear precoding, limited feedback, secrecy (sum) rate, performance analysis

I. INTRODUCTION

With the rapid growth of wireless communication systems in recent years, there is ever-increasing amount of privacy-sensitive data transmitted over wireless networks. However, signals can be easily intercepted by unauthorized users due to the broadcast nature of wireless media. The traditional cryptography-based secrecy methods are becoming even more challenging, due to the fast development of computing capabilities of new devices. Since the pioneering works in [1] and [2], where the wiretap channels and the so-called secrecy capacity were characterized for degraded channels and non-degraded broadcast channels respectively, achieving security in the physical layer has been receiving more and more attentions rather than from a purely information-theoretic perspective.

A. Previous Works

The research of the wiretap channel has also been extended to multi-antenna systems [3–5]. Regarding the employment of linear beamforming or precoding in secure communications, the work in [3] obtained the secrecy capacity of a single-user (SU) multiple-input-single-output (MISO) wiretap channel with multiple eavesdroppers in colluding fashion and showed that transmit beamforming was the optimal solution. The results were generalized to multiple-input-multiple-output (MIMO) wiretap channel in [4], where a precoding technique was proposed as a generalization of beamforming to achieve secrecy capacity. [5] obtained the ergodic secrecy sum rate (ESSR) of the simple regularized linear zero-forcing (ZF) precoding scheme for a multiuser MISO (MU-MISO) system under the assumption of perfect channel state information of the legitimate channels at the transmitter (denoted as CSIT).

Many of the research works in the context of secrecy of multi-antenna systems have assumed that the transmitter has perfect CSI of the legitimate channels, and in some cases also perfect knowledge of the eavesdroppers’ channels. In practice, however, the eavesdroppers’ channels are generally unknown to the transmitter, especially when the eavesdroppers are “passive” and keep silent. In addition, there is the practical constraint of limited feedback channels from the legitimate receivers (LRs) to the transmitter (which may be also insecure), particularly in widely-used frequency division duplex (FDD) systems, for which the CSI is typically obtained based on pilot training at the receiver, and conveyed to the transmitter through digital feedback. Hence, only partial or quantized version of the legitimate channels’ CSI is available at the transmitter. Thus, it is important to study secure transmission design under such conditions with limited feedback constraints.

Besides multi-antenna technology, another very effective approach to enhance security is to transmit artificial noise (AN) or cooperative jamming noise together with the confidential-message-bearing signals, especially for the scenarios that the CSI of the eavesdropper’s channels is unknown to the transmitter. There has been a considerable
body of literatures that investigated the design of AN-aided linear schemes in various scenarios with different qualities of CSIT [6–10]. With AN-aided and cooperative jamming schemes, information signals are beamformed/precoded in the direction of legitimate channels and AN is beamformed/precoded in the null space of the legitimate channels, or directed to the signal space that causes as little interference to LRs as possible. To be specific, [6] proposed AN generation methods for two communication scenarios with perfect CSIT of legitimate channel: SU MIMO system and the system with multiple amplify-forward relays acting as multiple antennas. Assuming full CSIT, only statistical CSI of the eavesdropper’s channel and a large number of transmit antennas, the authors in [7] derived approximated bounds of the ergodic secrecy rate (ESR) for AN-aided secure linear ZF precoding in MIMO systems with a multi-antenna eavesdropper. Assuming quantized legitimate channel’s CSI at the transmitter and considering beamforming with artificial noise in fast fading channels, [8] obtained optimal power allocation to maximize the ESR in two asymptotic regions, while [9] proposed to maximize the throughput under a connection outage constraint on the legitimate channel and a secrecy outage constraint against the eavesdropper. Cooperative jamming as the generalized idea of the AN was also studied in [10] with quantized CSIT.

All above works employed linear processing at the multi-antenna transmitter. It has been shown in the previous literatures that nonlinear Tomlinson Harashima precoding (THP) is in essence superior to the linear precoding [11, 12]. This is mainly due to the fact that the linear precoding can cause more serious power increase at the transmitter than that of THP, and thus incurs an inevitable capacity loss. Thus, THP can achieve a good tradeoff between complexity and rate performance over linear precoding and dirty paper coding (DPC) scheme [13]. As far as we know, there have been very limited works previously exploiting practical nonlinear precoding (e.g. THP) for secure transceiver design in multi-antenna systems with the exceptions in [14–18]. To be specific, [14] designed a nonlinear THP in a SU MIMO wiretap channel with a multi-antenna eavesdropper to guarantee certain mean-squared-error (MSE) for the intended LR. [15, 16] designed nonlinear successive optimization THP based on the perfect CSIT of the main channels of multiple LRs for MU-MIMO systems with multiple eavesdroppers. Only 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. Our previous work in [17] proposed to employ THP with a coexisting cooperative jamming helper. 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 contrast to the CSI models in [14–16], we consider FDD systems in which there exists limited-rate feedback channel from each LR to the transmitter and assume each LR can perfectly estimate the main channel. Our scheme employs a direct method which only requires the quantized channel direction information\(^1\) (CDI) of legitimate channels at the transmitter and is robust to distribution of channel fading and quantization method. Moreover, we would like to clarify that the performance of the proposed designs in [14–16] were only examined through simulations, which led to the fact that the power allocation could only be determined based on these simulations. These simulations significantly increase the implementation cost of the proposed scheme. In our previous paper [18], ESR performance was derived based on quantization cell approximation method which led to the obtained results much over the exact values. Moreover, the power allocation optimization was not considered therein. The above problems with the previous work will be studied and solved in this paper and new analytical results will be obtained.

\^1The channel direction information of a \( M \)-dimensional channel vector \( \mathbf{h} \) is defined as \( \bar{\mathbf{h}} = \frac{\mathbf{h}}{\| \mathbf{h} \|} \).
loss. Then, to maintain a constant ESR loss compared to the perfect CSI case, we obtain the number of feedback bits of each LR needs to scale logarithmically with the transmit SNR and linearly with the number of antennas of the legitimate transmitter.

C. Notations

\( C \) denotes the set of complex numbers. \( Z \) denotes the set of integers. \( \mathbb{E}[\cdot] \) represents expectation with respect to random variable \( X \). \( I_N \) denotes \( N \times N \) identity matrix. \( \text{diag} \{ \mathbf{m} \} \) denotes the diagonal matrix consisting of the elements of vector \( \mathbf{m} \) as the diagonal elements. \( \text{diag} \{ \mathbf{M} \} \) denotes the vector consisting of the diagonal elements of matrix \( \mathbf{M} \). For \( \mathbf{M} \in \mathbb{C}^{n \times n} \), we write \( \mathbf{M} \geq 0 \), if \( \mathbf{M} \) is Hermitian positive-definite. \( |x| \) denotes expectation of random variable \( x. \) \( \mu_x \) and \( \sigma^2_x \) denote mean and variance of \( x. \) \( C_{x,y} \) denote the covariance of random variables \( x \) and \( y. \) \( \Pr(E) \) denotes the probability of event \( E. \) \( \mathbf{Y} \) denotes the approximated value of \( \mathbf{Y}. \) \( \hat{\mathbf{Y}} \) denotes the set composed of the diagonal elements of matrix \( \mathbf{Y}. \) \( \hat{\mathbf{H}}_k \) represents the CDI of estimation \( \hat{\mathbf{H}}_k \).

Then, LR \( k \) quantizes the CDI of estimation \( \hat{\mathbf{H}}_k = \frac{\mathbf{h}_k}{||\mathbf{h}_k||} \) from a pre-determined quantization codebook \( \mathcal{W}_k \) with \( 2^{B_k} \) codewords according to \( \hat{\mathbf{H}}_k = \arg \max_{\mathbf{W}_k} \{ ||\hat{\mathbf{H}}_k\mathbf{a}^H||^2 \} \) and feeds back to Alice the index of \( \hat{\mathbf{H}}_k \) with \( B_k \) bits via an error- and delay-free feedback channel.

However, in general, it is very difficult if not intractable to obtain any analytical results when considering both channel estimation at the receiver side and quantized CSI feedback at the transmitter. Then, it is impossible to further analytically study the impact of limited CSI feedback on the system performance and to optimize. Moreover, by increasing the power of training signals, the channel estimation error can be controlled within a certain range and its impact on the system performance can be much smaller than that caused by limited CSI feedback. Thus, most previous related works ignored channel estimation error [8–10]. For the tractability of analysis and without causing confusion of the description, we will follow most of the previous related works by assuming each LR can perfectly estimate the channel and LR \( k \) quantizes the CDI \( \hat{\mathbf{H}}_k = \frac{\mathbf{h}_k}{||\mathbf{h}_k||} \) as \( \mathbf{h}_k = \arg \max_{\mathbf{W}_k} \{ ||\hat{\mathbf{H}}_k\mathbf{a}^H||^2 \} \). The impact of imperfect channel estimation on the ESR performance of our proposed scheme will be studied by numerical results in Section VI.

The codebook for each LR’s channel is designed offline. Since the optimal quantization codebook is generally unknown, we use random vector quantization (RVQ) codebook, in which the codewords are independently and identically distributed on the complex unit sphere [20, 21]. Its performance is closed to the optimal quantization codebook [20, 22]. Then, we have \( \hat{\mathbf{H}}_k = \mathbf{h}_k \cos \theta_k + \mathbf{h}_k \sin \theta_k, \) where \( \cos^2 \theta_k = ||\mathbf{h}_k\mathbf{a}^H||^2 \). \( \mathbf{H} \) can be decomposed as [12]

\[
\mathbf{H} = \mathbf{G} \left( \mathbf{\Phi} \mathbf{H} + \mathbf{\Omega} \right),
\]

where \( \mathbf{\Gamma} = \text{diag}(\rho_1, \cdots, \rho_K) \) with \( \rho_k = ||\mathbf{h}_k||, \) \( \mathbf{\Phi} = \text{diag}(\cos \theta_1, \cdots, \cos \theta_K), \) \( \mathbf{\Omega} = \text{diag}(\sin \theta_1, \cdots, \sin \theta_K), \) \( \mathbf{H} = [\mathbf{h}_1^T, \cdots, \mathbf{h}_K^T]^T \) and \( \mathbf{\hat{H}} = [\mathbf{h}_1^T, \cdots, \mathbf{h}_K^T]^T \) are \( K \times N_A \) matrices. Moreover, following the very similar model that has been used in many previous works (e.g. [10]), we consider a system with delay-tolerant traffic, where the coding block of each LR’s messages is long enough, such that we can use ergodic secrecy (sum) rate as the performance metric.

II. SYSTEM AND SIGNAL MODEL

We consider a downlink MU-MISO system with flat fading channels operating in FDE mode, where the transmitter (Alice) with \( N_A \) antennas simultaneously sends independent confidential messages to the \( K \) single-antenna LRs. There is an external passive \( N_E \)-antenna eavesdropper (Eve) in the vicinity of the MU-MISO system attempting to eavesdrop on the confidential messages of all LRs. Let \( \mathbf{H} = [\mathbf{h}_1^T, \cdots, \mathbf{h}_K^T]^T \in \mathbb{C}^{K \times N_A} \) denote the fading channel matrix consisting of all channel vectors from Alice to multiple LRs. The received signals of all LRs can be written as

\[
y_b = \mathbf{H} \mathbf{t} + \mathbf{n}_b, \tag{1}
\]

where \( y_b = [y_{b,1}, y_{b,2}, \cdots, y_{b,K}]^T \in \mathbb{C}^{K \times 1} \) consists of the received signals at all LRs with \( y_{b,k} \) denoting the received signal of LR \( k, \mathbf{t} \in \mathbb{C}^{N_A \times 1} \) denotes the transmit signal vector at Alice which is a mixture of information-bearing signals for all LRs and AN signal, and \( \mathbf{n}_b \) is the corresponding additive noise vector. The received signal at Eve is

\[
y_e = \mathbf{G} \mathbf{t} + \mathbf{n}_e, \tag{2}
\]

where \( \mathbf{G} \in \mathbb{C}^{N_E \times N_A} \) is the channel matrix from Alice to Eve.

In practise, Alice learns about the partial knowledge of each LR’s main channel through channel estimation and the CSI feedback. During the channel estimation, each antenna of Alice transmits training symbol vector with the length of \( T_r \) which are orthogonal to each other. \( T_r \geq N_A \) needs to be satisfied to obtain a reliable channel estimation [19]. Given the received vector corresponding to the \( T_r \) training symbols, LR \( k (k = 1, \cdots, K) \) can estimate the legitimate channel using linear minimum mean square error estimation. Then, the main channel of LR \( k \) can be written in terms of the estimate \( \hat{\mathbf{H}}_k \) and estimation noise \( \mathbf{m}_k \) as [19] \( \hat{\mathbf{H}}_k = \mathbf{h}_k + \mathbf{m}_k, \) where \( \mathbf{m}_k \sim \mathcal{CN}(0_{N_A \times 1}, \sigma^2_{m,k} \mathbf{I}_{N_A}) \) is independent of the estimation and is Gaussian with covariance \( \sigma^2_{m,k} = \frac{1}{1 + \frac{T_r}{\tau} ||\mathbf{h}_k||^2} \). Then, LR \( k \) quantizes the CDI of estimation \( \hat{\mathbf{H}}_k = \hat{\mathbf{h}}_k \)}
the largest integer not larger than $z$ [11]. With the effect of the modulo operation, the channel input signal vector can be equivalently written as

$$x_1 = s_1, \quad x_k = s_k + d_k - \sum_{i=1}^{k-1} |B|_{k,i,x_i}, \quad k = 2, \ldots, K,$$

(4)

where $x_k$, $k = 1, 2, \ldots, K$, are outputs of THP and $d_k \in \{2\tau_k(p_1 + \sqrt{-1} p_Q) | p_1, p_Q \in \mathbb{Z}\}$ is selected properly to restrict the real and imaginary parts of $x_k$ into the square region $R_k = \{x + \sqrt{-1} y | x, y \in (-\tau_k, \tau_k)\}$ [11], where $\tau_k$ corresponds to the modulation format of $s_k$. The effective signal vector $v = [v_1, v_2, \ldots, v^K]^T = s + d$ can be represented as

$$v = D x,$$

(5)

where $D = B + I$. It is shown in [11], without signal shaping for channel signal $x$ the elements of $x$ asymptotically follow uncorrelated and uniform distribution in $M_k$ as the constellation size increases. For $M_k$-order square constellation and $E\{s s^H\} = I$, the variance of $x_k$ for $k \geq 2$ can be accurately approximated as $E\{|x_k|^2\} = \frac{M_k}{M_k - 1}$ and $E\{|x_1|^2\} = E\{|s_1|^2\} = 1$ [11]. The power increase can be neglected for moderate to large constellation sizes. Without loss of generality, in the following we will ignore this effect.

Besides THP, there is a semi-unitary feedforward precoding matrix $F \in C^{N \times K}$ performed prior to transmission, i.e., $t_k = F x$. With the quantized CDI of all LRs in the compact channel matrix $H$, Alice obtains the LQ decomposition of $H$ as $H = R Q$, where $R = [R_{i,j}] \in C^{K \times K}$ is a lower triangular matrix and $Q = [q_{K,1}^T, q_{K,2}^T, \ldots, q_{K,K}^T]^T \in C^{K \times N}$ is a semi-unitary matrix with orthonormal rows. Then, using ZF criterion, we have $F = Q^H$ and $B = (\text{diag}\{\hat{R}\})^{-1} \hat{R} - I$ [11, 12]. When $N_A > K$, Alice can send the information signals in conjunction with an AN signal to disturb the signal reception at Eve. The total transmit power at Alice is $P$. A fraction $\phi$ of $P$ is allocated to the information-bearing signal and the left is allocated to AN. Assuming equal power allocation to the $K$ LRs, the transmitted signal vector can be re-written as

$$t = \sqrt{\frac{\phi P}{K}} F x + \sqrt{\frac{(1 - \phi) P}{N_A - K}} N u,$$

(6)

where the semi-unitary matrix $N \in C^{N \times (N_A - K)}$ is the precoding matrix for AN vector $u$, and $u$ is distributed as $u \sim \mathcal{CN}(0_{(N_A-K)\times 1}, \mathbb{I}_{(N_A-K)})$. In addition, a null constraint is imposed on the AN signal such that the columns of $N$ constitute an orthonormal basis for the null space of $H$, i.e. $HN = 0$, or equivalently $Q N = 0$, and $U = Q^H N$ is an orthonormal basis. Here, due to the absence of Eve’s CSI, all of the AN power is uniformly allocated to the null space of $H$.

Prior to the modulo operation, each LR equalizes the channel gain by dividing the received signal a factor $e_k$. Then, the signals of all users after equalization can be written in vector form as $r_b = E y_b$, where $E = \text{diag}(e_1, \ldots, e_K)$. The diagonal matrix consisting of all scaling factors at the LRs is given by

$$E = \sqrt{\frac{K}{\phi P}} \left( F \Phi \text{diag}\{\hat{R}\} \right)^{-1}.$$

(7)

With the processing of $E$, the equalized signal vector at all LRs, $r_b$, can be further written as

$$r_b = E y_b = \sqrt{\frac{K}{\phi P}} \left( F \Phi \text{diag}\{\hat{R}\} \right)^{-1} \left[ H \left( \sqrt{\frac{\phi P}{K}} F x + \sqrt{\frac{(1 - \phi) P}{N_A - K}} N u \right) + n_b \right]$$

$$+ \sqrt{\frac{1}{\phi P}} \left( F \Phi \text{diag}\{\hat{R}\} \right)^{-1} \Omega H N u$$

$$+ \sqrt{\frac{K}{\phi P}} \left( F \Phi \text{diag}\{\hat{R}\} \right)^{-1} n_b,$$

(8)

where we have used the relationship $HN = 0$ and $v = (\text{diag}\{\hat{R}\})^{-1} \hat{R} x$. Then, LR $k$ can estimate the symbol $s_k$ as

$$\hat{s}_k = \text{MOD}_r_k [v_k + i_k + \hat{n}_k] = \text{MOD}_r_k [s_k + i_k + \hat{n}_k]$$

(9)

where $i_k = \frac{\sin \theta_k}{\rho_{F_k} \cos \theta_k} h_k Q^H x + \sqrt{\frac{(1 - \phi) K}{\phi (N_A - K)}} \frac{\sin \theta_k}{\rho_{F_k} \cos \theta_k} \hat{n}_k N u$ and $\hat{n}_k = \frac{1}{K} E \Phi^H \rho_{F_k} \cos \theta_k \hat{n}_k$.

We note that, without signal shaping, the uncorrelated uniform distribution of $x$’s elements leads to the result that the rate can be up to 1.53 dB from the channel capacity (i.e., shaping loss) [23, Section 3.2.9]. Additionally, the interference term $i_k$ is non-Gaussian distributed given the channel matrices, which makes it intractable to obtain explicit analytical ergodic rates. By combining signal shaping (e.g. trellis shaping) with THP into an entity, the signals $x_k (k = 1, 2, \ldots, K)$ which are distributed close to uncorrelated Gaussian (restricted to the region of $R_k$) can be obtained to reduce the shaping loss. The detailed algorithm is more involved which can be found in [23, Section 5.4.1], but it is not the focus of this work. It is self-evident that the signals with high-energy occur very rarely for exact Gaussian-distributed signals. Additionally, it has been noted in [23, Page 129] that $x$ can be accurately approximated as Gaussian-distributed for moderate and high constellation sizes. Thus, for the tractability of performance analysis, we approximate $x$ as being with Gaussian distribution in this work. Moreover, it was also noted in [23, Page 129] that the additive noise (plus interference in our case) is signal-dependent and no longer exactly Gaussian due to the modulo operation at the receiver. This also makes the exact performance analysis intractable. However, for moderate to high SNRs this effect can be neglected and the end-to-end behavior can be well approximated by the additive Gaussian interference and noise model. This approximation had previously been adopted for the capacity analysis of THP in [11] and [12] for the systems with perfect CSI and quantized CSI, respectively. We will follow the above approximations with secrecy rate performance analysis in
\[
\hat{\gamma}_k = \frac{1}{\sin^2 \theta_k \|\hat{h}_k Q H\|^2 + \frac{(1-\phi)K}{\phi(N_A-K)}\sin^2 \theta_k \|\hat{h}_k N\|^2 + \frac{K}{\phi P} \frac{\sigma^2_{b,k}}{\|\hat{r}_{k,k}\|^2 \cos^2 \theta_k} + \frac{\sigma^2_{b,k}}{\|\hat{r}_{k,k}\|^2 \cos^2 \theta_k}}
\]

Section IV. Then, the output signal-to-interference-plus-noise ratio (SINR) for LR \(k\), \(\gamma_k\), can be approximately written as (10) at the top of this page, where \(\sigma^2_{b,k}\) is the variance of additive noise at LR \(k\).

We now compare the computational complexity of the proposed THP scheme and the corresponding ZF precoding. It is known that the complexities of both THP and ZF are polynomials of the number of transmit antennas \(N_A\) and the number of users \(K\) [24]. Particularly, without considering the complexity to implement artificial noise, the complexity of the proposed THP is \(10K^3 + 8K^2N_A - 2K^2 + 4K + 2\tau(2K+2K^2+N_A(4K-1)-4)\), where \(\tau\) is the length of the channel coherence interval during which the channel does not change. Whereas, the complexity of the ZF precoding is \(4K^3 + 2KN_A(4K-1) + K(4N_A-1)(K+1) + 8K^2 + 6K + 2\tau N_A(4K-1)\) [24]. Moreover, for both THP and ZF precoding the complexities to implement artificial noise in the null space of all LRs’ quantized channels are the same. Here, the computational complexity is expressed in terms of the required number of FLOPs corresponding to the number of complex-valued multiplications and additions. It is assumed one complex-valued multiplication and one complex-valued addition require 6 and 2 FLOPs respectively [25]. Thus, we can see that, the proposed THP and ZF precoding has the same level of computational complexity.

### B. Eavesdropping

Before proceeding, we first note that according to (4), \(x_k\) (\(k \geq 2\)) is a randomized version of \(s_k\) by channel (B) with THP. In the scenario where Eve doesn’t know the quantized CSI of all LRs (i.e., \(H\)) or he doesn’t know the processing method at the transmitter (Then, he does not know the real composite effective channel matrix. We call this general scenario for security.), it is impossible for Eve to recover \(s_k\) (except that \(x_1 = s_1\)) with any possible method. Most previous linear transceiver scheme had considered this general scenario for security, where Eve only knows the composite effect of physical channel with linear precoding of himself. In this work, for the very rigorous scenario for security, we assume the feedback channel is not secure such that Eve can obtain both \(H\) and \(G\), and Eve also knows the processing method at Alice. In this scenario, employing proper wiretap coding at Alice, Eve cannot recover the original message but can only guess one. Without the correct \(s_k\)’s, THP can introduce additional MU interference at Eve to mask confidential messages of LR \(k\) for \(k \geq 2\). Thus, the original confidential message of LR \(k\) (\(k \geq 2\)) can be secure even without wiretap coding as long as \(s_1\) is secure using proper wiretap coding. This can be identified as an unique advantage of THP over the linear precoding schemes for secure communications.

Instead of considering any concrete decoder used by Eve, we will study in the next section an upper bound of the ergodic rate of \(s_k\) at Eve by any possible decoder with the knowledge of \(H\), \(G\), and the processing method of Alice. Although we don’t consider any concrete eavesdropping method here, we would like to mention some possible methods in the previous literatures. The received signal at Eve can be written as

\[
y_e = \sqrt{\frac{\phi P}{K}} Q H^T D^{-1} v + \sqrt{\frac{(1-\phi)P}{N_A-K}} G N u + n_e, \quad (11)
\]

which can be identified as a MIMO channel with co-channel interference and noise. When Eve can obtain the CSI of his own channel (G) and the quantized CSI of legitimate channels (H), different decoders with different complexities can be used to detect \(v\) (s), such as the linear decoders based on ZF or MMSE criterion and the optimal maximum-likelihood (ML) decoder that minimizes the average error probability whose complexity is exponentially increasing with the dimension of \(v\) (or \(s\)). Sphere decoders is a class of algorithms that can approximate the performance of ML decoder with a polynomial complexity within a wide range of system parameters [26]. However, we notice that the original messages are implicitly embedded in \(v\) (recall that \(v_k = s_k + d_k\) is periodically shift of \(s_k\)). In addition, according to (4), \(d_k\) is determined depending on the first \(k - 1\) elements of \(x\) and \(H\) (B). Thus, compared with those used in the original MIMO channel without THP, all previous decoder algorithms designed for MIMO channel need to be adjusted to this change, which makes the optimal and the approximated ML decoders even more complicated.

**Remark 1:** The above assumptions of the knowledge of the quantized CSI of the legitimate channels and the processing method used by Alice will result in an upper bound on the achievable ergodic rate of Eve, and thus, the worst-case secrecy rate of each LR. Since we only consider low-complexity design which is only based on the CDI of instantaneous channels realization at the transmit side, Alice chooses a fixed value for \(\phi\).

Next, we will study the secrecy performance of the proposed artificial-noise-aided secure nonlinear transceiver.

### IV. Ergodic Secrecy Rate Analysis and Power Allocation

In this section, we analytically study the ESR of each LR for the typical fading scenario, where all channels are mutually independent and spatially uncorrelated.
Rayleigh fading, i.e., $\mathbf{h}_k \sim \mathcal{CN}(0_{1 \times N_A}, \mathbf{I}_{N_A})$, $\mathbf{G} \sim \mathcal{CN}(0_{N_B \times N_A}, \mathbf{I}_{N_B} \otimes \mathbf{I}_{N_A})$. In addition, as noted in [6, 9], the receiver noise power at Eve may not be known by Alice and the LRs. Thus, to guarantee secure communication, we consider the worst-case scenario where the noise variance at Eve is arbitrarily small, i.e., $\sigma_e^2 \to 0$. This assumption in practise also results in an upper bound on the ergodic rate for the message of LR $k$ at Eve (denoted as $R_{e,k}^{LFB}$).

The ESR of LR $k$ can be written as $[6]$ $R_{e,k}^{LFB} = \left[R_{e,k}^{FB} - R_{e,k}^{LFB}\right]^+$, where $R_{e,k}^{FB} = \mathbb{E}\left[\log_2(1 + \rho_k)\right]$ is the ergodic rate of LR $k$’s messages at Alice and $R_{e,k}^{LFB}$ is the maximum ergodic rate of LR $k$’s messages achieved by any possible method at Eve for any condition. Then, the ESSR is given by $R_{sec,k} = \sum_{k=1}^{K} R_{e,k}^{LFB}$. Before proceeding, we first present some distribution results related to the signals, interferers and AN leakage. These results are very useful for the secrecy (sum) rate analysis.

\textbf{Lemma 1:} For $1 < K < N_A$, the random variables $\varepsilon_k = \|\mathbf{h}_k \mathbf{Q}^H\|^2$, for $k = 1, 2, \cdots, K$, follow the same beta distribution with shape $(K - 1)$ and $(N_A - K)$, which is denoted as $\varepsilon_k \sim \text{Beta}(K - 1, N_A - K)$, and the probability distribution function (PDF) is given by

$$f_{\varepsilon_k}(x) = \frac{1}{\beta(K - 1, N_A - K)} x^{K - 2}(1 - x)^{N_A - K - 1}, \tag{12}$$

where $\beta(a, b) = \int_0^1 x^{a-1}(1 - x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ is beta function [27]. Specially, when $K = 1$ there is no interference term. The random variables $\varepsilon_k = \|\mathbf{h}_k \mathbf{N}\|^2$, for $k = 1, 2, \cdots, K$, follow beta distribution with shape $(N_A - K)$ and $(K - 1)$. For $K = 1$, $\varepsilon_1 = \|\mathbf{h}_1 \mathbf{N}\|^2 = 1$.

The random variable $\|\mathbf{h}_k \mathbf{N}\|^2$ follows beta distribution with shape $(N_A - k + 1)$ and $(k - 1)$, whose PDF is

$$f_{\|\mathbf{h}_k \mathbf{N}\|^2}(x) = \frac{1}{\beta(N_A - k + 1, k - 1)} x^{N_A - k - 1}(1 - x)^{k - 2}. \tag{13}$$

For $k = 1$, $\|\mathbf{h}_1 \mathbf{N}\|^2 = 1$.

\textbf{Proof:} The proof of (12) is provided in [12, Lemma 1]. The distribution of $\varepsilon_k$ follows by the following important relation: $\|\mathbf{h}_k \mathbf{N}\|^2 + \|\mathbf{h}_k \mathbf{Q}^H\|^2 = \|\mathbf{h}_k \mathbf{U}\|^2 = 1$. The distribution of $\|\mathbf{h}_k \mathbf{N}\|^2$ in (13) can be proved using the facts that $\|\mathbf{h}_k \mathbf{N}\|^2$ and $\|\mathbf{h}_k \mathbf{Q}^H\|^2$ have the same distribution, which was proved in Appendix C of [12], and $\rho_k^2$ follows $\chi^2_{2N_A}$ distribution.

In the following, we will only focus on the general situation that $1 < K < N_A$. However, it is much easier to obtain the results for the special case of $K = 1$.

We can rewrite $R_{e,k}^{LFB}$ as

$$R_{e,k}^{LFB} = \mathbb{E}\left[\log_2\left(\frac{\phi P}{K\sigma_k^2} \rho_k^2 \|\mathbf{h}_k \mathbf{Q}^H\|^2 \cos^2 \theta_k\right)\right] + \frac{\phi P}{K\sigma_k^2} \rho_k^2 \|\mathbf{h}_k \mathbf{Q}^H\|^2 \sin^2 \theta_k \left\{\begin{array}{l}
(1 - \rho_k^2) \left\|\mathbf{h}_k \mathbf{N}\|^2 + 1\right\} \right.
- \mathbb{E}\left[\log_2\left(\frac{\phi P}{K\sigma_k^2} \rho_k^2 \|\mathbf{h}_k \mathbf{Q}^H\|^2 \sin^2 \theta_k \left\|\mathbf{h}_k \mathbf{N}\|^2 + 1\right\} \right]
+ \frac{\phi P}{K\sigma_k^2} \rho_k^2 \|\mathbf{h}_k \mathbf{Q}^H\|^2 , \tag{14}
\end{array}\right.$$

Now, with Lemma 1 and some distribution results in [21], the exact distribution of each random term in each $\log_2$ of (14) is known. However, all random terms are correlated to each other. Thus, it is intractable to obtain the exact closed-form results of two expectations in (14). Instead, we would like to obtain some tight approximation of (14) which is given in the following theorem.

\textbf{Theorem 1:} $R_{e,k}^{LFB} \geq R_{b,k}^{approx}$, where, for $k = 1$,

$$R_{b,k}^{approx} = \frac{K - 1}{N_A - 1} 2^{B_k} \sum_{m=1}^{N_A-1} \beta\left(\frac{m}{N_A - 1} + 1, 2B_k - 1\right)
- \log_2\left(\frac{e}{N_A - 1}\right) \sum_{i=1}^{N_A-1} \beta\left(2B_k, \frac{i}{N_A - 1}\right)
+ \log_2\left(\frac{1}{\left(1 - \phi \frac{\sigma_{b,k}}{\mu(K - 1)}\right) N_A-K} \right), \tag{15}\right.$$

for $k \geq 2$,

$$R_{b,k}^{approx} = p_k 2^{B_k} \sum_{m=1}^{N_A-1} \beta\left(\frac{m}{N_A - 1} + 1, 2B_k - 1\right)
- \log_2\left(\frac{e}{N_A - 1}\right) \sum_{i=1}^{N_A-1} \beta\left(2B_k, \frac{i}{N_A - 1}\right)
+ \log_2\left(p_k \frac{1}{(1 - \phi \frac{\sigma_{b,k}}{\mu(K - 1)} \frac{1}{N_A - K})} \right), \tag{16}\right.$$

where $\sigma_k = \frac{1}{\sigma_{b,k}}$, $\mu_k = 1 - d_k 2^{B_k} \beta\left(2B_k, \frac{N_A}{N_A - 1}\right)$ with $d_1 = \frac{N_A - K}{N_A - 1}$ and $d_k = 1 - c_k$ for $k \geq 2$,

$$p_k = \sum_{j=K-1}^{N_A-2} \left(\frac{N_A - 2}{j}\right) \frac{\beta(N_A - k + j, N_A + k - j - 3)}{\beta(N_A - k + 1, k - 1)}, \tag{17}\right.$$

$$c_k = \frac{K - 1}{N_A - 1} \sum_{j=K}^{N_A-1} \left(\frac{N_A - 1}{j}\right) \frac{\beta(N_A - k + j, N_A + k - j - 2)}{\beta(N_A - k + 1, k - 1)}. \tag{18}\right.$$
\[ (u_1, u_2, v_1, v_2, w), \] and \[ z_1 = (\mu_{x_1}, \mu_{y_1}, \sigma_{x_1}^2, \sigma_{y_1}^2, C_{x_1, y_1}) \] 
\[ z_k = (\mu_{x_k}, \mu_{y_k}, \sigma_{x_k}^2, \sigma_{y_k}^2, C_{x_k, y_k} | \epsilon_k), \] \[ z_{0,k} = (\mu_{x_0,k}, \mu_{y_0,k}, \sigma_{x_0,k}^2, \sigma_{y_0,k}^2, C_{x_0,k, y_0}) \] for \( k = 2, \ldots, K \), 
\[ \nu_{0,k} = (\mu_{x_0,k}, \mu_{y_0,k}, \sigma_{x_0,k}^2, \sigma_{y_0,k}^2, C_{x_0,k, y_0}) \] for \( k = 1, \ldots, K \) 

with
\[ \mu_{x_k} = \sigma_{x_k}^2 = \frac{N_A - k + 1}{N_A} \sigma_k, \]
\[ \mu_{x_0,k} = \frac{N_A - 1}{N_A} \sigma_k, \]
\[ \sigma_{y_k}^2 = \frac{(N_A + 1)(N_A - K + 1)(N_A - K)}{N_A - 1} \times \frac{N_A^{2}(N_A - K)^2}{(N_A - 1)^2} \sigma_k^2, \]
\[ \sigma_{x_0,k}^2 = \frac{(N_A + 1)K(K - 1)}{N_A - 1} \times \frac{N_A^{2}(K - 1)(N_A - K)}{N_A - 1} \times \frac{N_A^{2}(K - 1)(N_A - K)}{N_A - 1} \sigma_k^2, \]
\[ C_{x_0,k, y_0} = \frac{(N_A + 1)(K - 1)(N_A - K)}{N_A - 1} \times \frac{N_A^{2}(K - 1)(N_A - K)}{N_A - 1} \sigma_k^2, \]
\[ C_{x_1, y_1} = \frac{(N_A - K)N_A}{N_A - 1} \sigma_k, \] and for \( k \geq 2 \)
\[ C_{x_k, y_k} | \epsilon_k = \left[ \left( \frac{N_A + 1}{N_A} \right) N_A \bar{k}_k \right] \sigma_k, \] 
\[ C_{x_{0,k}, y_{0,k}} | \epsilon_k = \left[ \left( \frac{N_A + 1}{N_A} \right) N_A \bar{k}_k \right] \sigma_k, \] 
\[ \bar{k}_k = \frac{N_A - k + 1}{N_A} \sum_{j=0}^{N_A - k + 2} \left( \frac{N_A}{j} \right) \times \frac{\beta(2N_A - K - j + 1, K + j - 1)}{\beta(N_A - K, K - 1)}, \] and
\[ \bar{k}_0 = \frac{(N_A - k + 1)(N_A - K)}{(N_A - 1)^2} - \bar{k}_k. \]

Proof: See Appendix A.

Remark 2: The result in Theorem 1 is obtained mainly using Lemma 5, which is based Taylor expansion of function \( g(x, y) \). For our case, the function \( g(x, y) = \ln(1 + Ax + By) \) is with the reminder of \( R(x, y) = 3(1 + A[\mu_x + \sigma_x^2, y - \mu_y)]^3 \). The expectation of \( R(x, y) \) is the gap from the real values induced by the approximation. This term can be positive or negative depending on the concrete system settings, but very hard to analytically quantify. As we can see from the simulation results that this approximation can be very accurate especially for high-resolution quantization scenario, which illustrates our analytical approximation is reasonable.

Recall that we consider the worst-case scenario where \( \sigma_e^2 \to 0 \). In this case, in order to achieve positive ESR Alice has to transmit AN with non-zero power and cannot allocate full power to AN, i.e., \( 0 < \phi < 1 \). Then, we rewrite the received signal at Eve as
\[ y_e = \sqrt{\frac{\phi P}{K}G_Q x} + \sqrt{\frac{(1 - \phi)P}{N_A - K}G_N u}, \] where \( G_Q = GQ^H \) and \( G_N = GN \). In the following, we implicitly assume Eve can acquire information of \( G_Q \) and \( G_N \). 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. Then, we have the following results.

Lemma 2: Given \( \hat{H} \), we have
\[ I(s_k; y_e | \hat{H}) = \frac{1}{K} I(s; y_e | \hat{H}) = \frac{1}{K} I(x; y_e | \hat{H}), \] for all \( s_k \), and \( I(x; y_e | \hat{H}) \) can be upper bounded as
\[ I(x; y_e | \hat{H}) \leq R_{e, \text{sum}} \]
\[ \neq \log_2 \left[ \det \left( \frac{\phi(N_A - K)}{(1 - \phi)K}G_Q G_Q^H + G_N G_N^H \right) \right] - \log_2 \left[ \det (G_N G_N^H) \right], \]
where the upper bound in (22) is achieved when \( x \) is Gaussian distributed which can be approximately achieved by combining THP and signal shaping. Moreover, we have
\[ I(s; y_e | \hat{H}) > I(s; y_e). \]

Proof: See Appendix B.

Remark 3: \( I(s; y_e | \hat{H}) > I(s; y_e) \) illustrates, when \( \hat{H} \) is unknown to Eve, the supported rate for the message-bearing signals over the eavesdropping channel is less than that when Eve does know the quantized CSI \( \hat{H} \), and thus the secrecy rate performance of the system can be enhanced if Eve can be prevented from obtaining the legitimate channels’ (quantized) CSI. Or equivalently, we conclude that the randomness of the legitimate channels’ CSI can help to enhance the secrecy of the system by employing THP. Whereas for linear precoding schemes (e.g., linear ZF precoding), since Eve can directly obtain through channel estimation the effective channel that combines physical channel \( G \) with the precoding matrix, the further knowledge of the legitimate channels’ CSI is no help for eavesdropping. Thus, \( I(s; y_e | \hat{H}) = I(s; y_e) \) holds for linear precoding.

Remark 4: It is noted that the quantized CSI does not necessarily need to be kept confidential from Eve over the feedback channels. On the one hand, in fact, the analytical secrecy performance obtained (for the downlink channels) is for the scenario where the quantized CSI is not kept confidential from Eve. It is obvious the system can work properly even though the system does not take efforts to protect the quantized CSI from Eve. On the other hand, the system can also employ certain proper physical layer
security method\textsuperscript{a} at the feedback channels to keep quantized CSI confidential from Eve. We have theoretically proved how the secrecy rate performance of the system is increased in the later case.

Next, we will derive $R_{c,k}^{LF}$. According to Lemma 2, we have

$$R_{c,k}^{LF} = E \left[ \max_{p \in \mathcal{P}(s,x)} I(s_k; y_e | \mathbf{H}) \right] = \frac{1}{N} E[\mathbb{E}_{\text{Rave,sum}}],$$

where the maximization is taken over all possible input distributions of $p \in \mathcal{P}(s,x | \mathbf{H})$. Thus, in the following we first derive the expression of $\mathbb{E}_{\text{Rave,sum}}$.

Remark 5: From (22), it is required that the matrix $\mathbf{G}_N \mathbf{G}_N^H$ is invertible. Otherwise, the supported rate of Eve’s channel can be infinite and it is possible for Eve to obtain confidential messages. In order to guarantee this, it is required that $N_A \geq N_E + K$.

It is easy to see that $\mathbb{E}_{\text{Rave,sum}}$ can be written as

$$\mathbb{E}_{\text{Rave,sum}} = \mathbb{E} \left[ \log_2 \det (\mathbf{W}_N) - \log_2 \det (\mathbf{W}_N) \right] .$$

where $\mathbf{W}_N = \mathbf{G}_N \mathbf{G}_N^H$ and $\mathbf{W}_{QN} = \mathbf{G}_N \mathbf{Q}_N \mathbf{G}_N^H$ with $\mathbf{G}_N = [\mathbf{Q}_G \mathbf{G}_N^H]$ and

$$\Sigma_{QN} = \begin{bmatrix} \phi(N_E-K) & I_K & 0 \\ 0 & I_{N_A-K} \end{bmatrix} .$$

Then, $\mathbf{W}_N$ and $\mathbf{W}_{QN}$ follow complex central Wishart distribution, i.e., $\mathbf{W}_N \sim W_{NE}(N_A - K, I_{N_A - K})$ and $\mathbf{W}_{QN} \sim W_{NE} (N_A, \Sigma_{QN})$. Denote the joint PDFs of the non-zero ordered eigenvalues of complex matrices $\mathbf{W}_N$ and $\mathbf{W}_{QN}$ as $f_1(\lambda)$ and $f_2(\lambda)$ respectively, where the $N_E$ non-zero eigenvalues are denoted as $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{N_E})$.

Using the result in [29, 30], the joint PDFs of $\mathbf{W}_N$ and $\mathbf{W}_{QN}$ can be obtained and given in the following lemma.

Lemma 3: $f_1(\lambda)$ is given as

$$f_1(\lambda) = K_N \det [\mathbf{V}(\lambda)] \prod_{i=1}^{N_E} e^{-\lambda_i N_A - N_E - K} ,$$

where $\mathbf{V}(\lambda)$ is a $N_E \times N_E$ Vandermonde matrix with elements $\{ \mathbf{V}(\lambda) \}_{i,j} = \lambda_i^{j-1}$ and $K_N = \frac{\Gamma(N_E+1)}{\Gamma(N_A-K+1)}$.

The $m(a) = \prod_{i=1}^{a-1}(a-i)!$ is the normalized complex multivariate gamma function. $(a,k) = (a(a-1) \ldots (a-k+1), a)_0 = 1.$

(a) For the scenario that $\phi \neq \frac{\mathcal{N}_A}{\mathcal{N}_E}$, $f_2(\lambda)$ is given as

$$f_2(\lambda) = K_{QN1} \det [\mathbf{V}(\lambda)] \prod_{i=1}^{N_E} e^{-\lambda_i N_A - N_E - K} ,$$

where the $(i,j)$-th element of $N_A \times N_A$ matrix $\mathbf{V}(\lambda, z)$ is given by

$$\{ \mathbf{V}(\lambda, z) \}_{i,j} = \chi_{m(i-1)} e^{-\omega(1) \lambda_j} ,$$

$$\lambda_i = 1, \ldots, m_1; j = 1, \ldots, N_E$$

$$\lambda_{N_A-j} e^{-\omega(2)} ,$$

$i = 1, \ldots, m_1; j = 1, \ldots, N_E$

$$N_A - j \chi_{m(i-1)} e^{-\omega(1)} ,$$

$i = 1, \ldots, m_1; j = N_E + 1, \ldots, N_A$

$$N_A - j \chi_{m(i-1)} e^{-\omega(2)} ,$$

$i = 1, \ldots, m_1; j = N_E + 1, \ldots, N_A$.

(b) For the scenario that $\phi = \frac{K_N}{A}$, $\mathbb{E}_{\text{Rave,sum}}$ is given by

$$\mathbb{E}_{\text{Rave,sum}} = K_{QN2} \sum_{p=1}^{N_E} \det (\mathbf{V}(\lambda)) \prod_{i=1}^{N_E} e^{-\lambda_i N_A - N_E} ,$$

where $K_{QN2} = \frac{\Gamma(N_E+1)}{\Gamma(N_A-K+1) \Gamma(N_A-K)}$.

With the results of joint PDFs of the ordered eigenvalues in Lemma 3, $\mathbb{E}_{\text{Rave,sum}}$ can be obtained in closed-form and given in the following theorem.

Theorem 2: (a) For the scenario that $\phi \neq \frac{K_N}{A}$, $\mathbb{E}_{\text{Rave,sum}}$ is given by

$$\mathbb{E}_{\text{Rave,sum}} = K_{QN1} \sum_{p=1}^{N_E} \det (\mathbf{V}(\lambda)) - K_N \sum_{p=1}^{N_E} \det (\mathbf{V}(\lambda)) ,$$

(b) For the scenario that $\phi = \frac{K_N}{A}$, $\mathbb{E}_{\text{Rave,sum}}$ is given by

$$\mathbb{E}_{\text{Rave,sum}} = K_{QN2} \sum_{p=1}^{N_E} \det (\mathbf{V}(\lambda)) - K_N \sum_{p=1}^{N_E} \det (\mathbf{V}(\lambda)) ,$$

where $K_{QN2} = \frac{\Gamma(N_E+1)}{\Gamma(N_A-K+1) \Gamma(N_A-K)}$.

Proof: With joint PDFs obtained in Lemma 3, $\mathbb{E}_{\text{Rave,sum}}$ can be obtained as

$$\mathbb{E}_{\text{Rave,sum}} = \int \cdots \int \left[ \prod_{i=1}^{N_E} \log_2 (\lambda_i) \right] [f_2(\lambda) - f_1(\lambda)] d\lambda ,$$

\textsuperscript{a}Considering that the feedback channels are with limited capacity, it is not necessarily the one in this paper.
\[ R_{b,k}^{\infty} \triangleq \lim_{P/\sigma_{b,k}^2 \to \infty} R_{b,k}^{\text{LFB}} = \mathbb{E} \left\{ \log_2 \left( 1 + \frac{\phi}{K} \frac{|r_{k,k}|^2}{\|h_k Q^H\|^2 + \frac{(1-\phi)}{\lambda_k} \|h_k N\|^2 \sin^2 \theta_k} \right) \right\}, \]  

(31)

where the integral region is \( \mathcal{D} = \{ \lambda : \lambda_1 > \lambda_2 > \cdots > \lambda_{N_k} \geq 0 \} \). The final results can be obtained by following the similar steps in [31] with the joint PDFs in (26), (27) or (28) and [27, 3.351.3; 4.352.1].

We notice that \( \mathbb{E} [R_{\text{eve,sum}}] (R_{b,k}^{\text{LFB}}) \) given by (29) and (30) do not depend on the number of feedback bits \( B_k \). In fact, similar to the cooperative jamming scheme in [10], it is easy to show the distribution of AN signals at Eve does not change with the quality of CSI of all legitimate channels (determined by \( B_k \)).

**Remark 6:** Since the distributions of \( G_Q \) and \( G_N \) are independent with that of \( H \), with (22) it is not difficult to see \( R_{b,k}^{\text{LFB}} \) in (24) does not depend on the distribution of \( H \).

With the analytical results of \( R_{b,k}^{\text{LFB}} \) in Theorem 1 and \( \mathbb{E} [R_{\text{eve,sum}}] \) in Theorem 2, using Lemma 2, \( R_{\text{sec,sum}} \) can be approximated as

\[ R_{\text{sec,sum}} \approx R_{\text{sec,sum}}^{\text{approx}} \triangleq \sum_{k=1}^{K} R_{\text{sec,k}}^{\text{approx}}, \]

(32)

where \( R_{\text{sec,k}}^{\text{approx}} = \left[ R_{b,k}^{\text{approx}} - \frac{1}{K} \mathbb{E} [R_{\text{eve,sum}}] \right] \) is the ESR approximation of LR \( k \). Moreover, we would like to note that the corresponding results for the scheme using THP without AN (with perfect or quantized CSIT) can be easily obtained by using the existing results of THP in [12] (without considering security).

**Remark 7:** According to (14) and the derivation of \( R_{b,k}^{\text{approx}} \) in Theorem 1, it is easy to see the corresponding result of \( R_{b,k}^{\text{approx}} \) for the general case with arbitrary channel variances can be obtained by substituting \( P/\sigma_{b,k}^2 \) instead of \( P \) into Theorem 1, where \( \sigma_{b,k}^2 \) is the variance of each element of \( h_k \).

In addition, from (22) it is not difficult to see that the variance of \( G \) does not affect the result of \( \mathbb{E} [R_{\text{eve,sum}}] (R_{b,k}^{\text{LFB}}) \).

In general, ESRs are usually complex functions of \( \phi \) in practical systems. The heuristic methods in the previous works usually turned to some special cases (e.g., large-antenna systems, high-resolution quantization, high-SNR region or even system with two antennas [5, 7, 8]) to obtain closed-form approximations of the optimal \( \phi \). However, the obtained result cannot be applied to general system parameters. Instead, in the following we still consider the general system settings. We can obtain the following analytical results which are useful to obtain near-optimal \( \phi \).

**Lemma 4:** Given the number of feedback bits of each LR \( B_k \), \( R_{b,k}^{\text{LFB}} \) as functions of \( \phi \) is concave for \( 0 < \phi < 1 \), and \( \mathbb{E} [R_{\text{eve,sum}}] (R_{b,k}^{\text{LFB}}) \) is convex in the region \((\epsilon, 1)\) for some \( 0 \leq \epsilon < 0.5 \) and is concave in the region \((0, \epsilon)\), where \( \epsilon \) depends on system parameters.

**Proof:** See Appendix C. \( \square \)

We will further study the results in this lemma by numerical results in Section VI.

**Remark 8:** Although we can obtain a closed-from expression of \( \frac{\partial^2 \mathbb{E} [R_{\text{eve,sum}}]}{\partial \phi^2} \) using (29), (30) in the very direct way, it is expected that the result is intractable to observe the convexity or concavity property of \( \mathbb{E} [R_{\text{eve,sum}}(\phi)] \) for the systems with arbitrary system parameters due to the complicated expression.

According to Lemma 4, we conclude that \( R_{\text{sec,k}}^{\text{approx}} = \left[ R_{b,k}^{\text{approx}} - \frac{1}{K} \mathbb{E} [R_{\text{eve,sum}}] \right] \) is concave function for \( \epsilon < \phi \leq 1 \) with general system settings, where \( \epsilon \) is as defined in Lemma 4. However, it is intractable to strictly prove the concavity or convexity of \( R_{\text{sec,k}}^{\text{approx}} \) for \( 0 < \phi \leq \epsilon \) with general system settings. Moreover, as illustrated in Remark 2, \( R_{\text{approx}} \) as the approximation of \( R_{\text{LFB}}^{\text{approx}} \) is obtained using affine functions. Thus, it is intractable to strictly prove the concavity or convexity of \( R_{\text{sec,k}}^{\text{approx}} \) also. As we will see from the numerical results in Section VI, \( R_{\text{sec,k}}^{\text{approx}} \) is indeed a concave function in \((0, 1)\) for the systems considered, and \( R_{\text{sec,k}}^{\text{approx}} \) can approximate and track the real values well, especially for the systems with large number of feedback bits. These results motivate us to propose obtaining the sub-optimal solution of \( \phi \) by solving \( \max_{\phi \in (0,1)} \sum_{k=1}^{K} \left[ R_{b,k}^{\text{approx}} - \frac{1}{K} \mathbb{E} [R_{\text{eve,sum}}(\phi)] \right] \) using some numerical method, such as golden section search and parabolic interpolation.

V. SCALING OF THE NUMBER OF FEEDBACK BITS

In the above section, we have derived an approximation of \( R_{b,k}^{\text{LFB}} \) and the exact close-form expression of \( R_{b,k}^{\text{LFB}} \). According to LR \( k \)'s SINR given by (10) and the expression of \( R_{\text{sec,k}}^{\text{approx}} \) given in Theorem 2, it is easy to observe that, in the high SNR regime, \( R_{b,k}^{\text{LFB}} \) converges to (31) shown at the top of this page, which is bounded for any fixed values of \( \phi \) and \( B_k \), and is independent of \( P/\sigma_{b,k}^2 \). The same result also holds for \( R_{\text{sec,k}}^{\text{LFB}} \). Moreover, there is no AN leakage at each LR with perfect CDI. It has been shown in [12] that the ESR of LR \( k \) with perfect CDI is \( R_{b,k}^{\text{perfCDI}} \triangleq \mathbb{E} \left\{ \log_2 \left( 1 + \frac{\phi P}{K \sigma_{b,k}^2 r_k^2} \|r_k\|^2 \right) \right\} \), which increases without bound as SNR increases. Certainly, we can obtain an approximation of each LR's ESR (and also the ESSR) in this high SNR regime scenario by the similarly method as that used in Theorem 1, or obtain an upper bound using the similar method in [12]. However, we focus on the problem that the ESR loss of each LR compared to the perfect CDI case will become arbitrarily large for a fixed \( B_k \). To resolve this problem, we derive in this section the value of \( B_k \) that is needed for maintaining a constant ESR loss.

First, we define the ESR loss of LR \( k \) \((k = 1, \cdots, K)\) as

\[ \Delta R_{\text{sec,k}} = \max_{\phi \in \mathcal{R}_{\text{sec,k}}^{\text{perfCDI}}} \left\{ R_{\text{sec,k}}^{\text{perfCDI}} - R_{\text{sec,k}}^{\text{LFB}}(\phi) \right\} \]

(33)

where \( \phi^* = \arg \max_{\phi} R_{\text{sec,k}}^{\text{LFB}}(\phi) \). Note that according to the above analysis, \( R_{\text{sec,k}}^{\text{LFB}}(\phi) \) becomes independent of \( P/\sigma_{b,k}^2 \) as \( P/\sigma_{b,k}^2 \to \infty \), when the other system parameters fixed.
follows that $\phi^*$ approaches to a value which is independent with the SNR $P/\sigma^2_{b,k}$. Moreover, it is easy to see

$$R_{sec,k}^{lfb}(\phi^*) - R_{b,k}^{lfb}(\phi^*) \leq R_{sec,k}^{per CDI}(\phi^*) - R_{b,k}(\phi^*).$$

(34)

By combing (33) and (34), we show in Appendix D the following theorem:

**Theorem 3**: The ESR loss between perfect and quantized CDI is upper-bounded as $\Delta R_{sec,k} \leq \Delta R_{sec,k}^{u}$ where $\Delta R_{sec,k}^{u}$ is given by

$$\Delta R_{sec,k}^{u} = \log_2 \left( \frac{1 + \phi^*(N_A - k + 1) \frac{P}{\sigma_{b,k}^2}}{K} \right)$$

$$- \log_2 \left( 1 + \frac{\phi^*(N_A - k + 1) \frac{P}{\sigma_{b,k}^2}}{(K - \phi^*) \frac{P}{\sigma_{b,k}^2} - \frac{1}{\sigma_{b,k}^2} + K} \right) \right).$$

(35)

To characterize a sufficient condition of the scaling of feedback rate of LR $k$, we set the right hand side (RHS) of (35) to be the maximum allowable gap of $\Delta R_{th,k}$. After some simple manipulations, we can see the implication of the above theorem: For $\Delta R_{sec,k} \leq \Delta R_{th,k}$, $B_k$ is chosen to scale with the increase of $\frac{P}{\sigma_{b,k}^2}$ as (36) at the top of this page.

Note that the RHS of (36) approaches

$$B_k \triangleq (N_A - 1) \left[ \log_2 \left( \frac{P}{\sigma^2_{b,k}} \right) - \log_2 \left( \frac{2}{K} \right) \right] \right).$$

(37)

as $\frac{P}{\sigma^2_{b,k}} \rightarrow \infty$ for the fixed $N_A$, $\Delta R_{th,k}$ and $K$. Therefore, to maintain a constant ESR loss, $B_k$ needs scale logarithmically with $P/\sigma^2_{b,k}$ and linearly with $N_A - 1$. Moreover, we can observe from the second term in the RHS of (37) that, the setting of $B_k$ can be approximated decreased linearly with the stem $(N_A - 1)\Delta R_{th,k}$.

**VI. NUMERICAL STUDIES**

For the numerical studies, we consider the wiretap MISO MU-MISO systems with different system settings. We denote $PK/(\sum_{k=1}^{K} \sigma^2_{b,k})$ as the SNR of the systems. The number of antennas of Alice can be $N_A = 6$ or $N_A = 4$ for different systems. In the following we first study the performance of $R_{b,k}^{lfb}$, $R_{c,k}^{lfb}$ and ESSR as functions of $\phi$.

We first show some numerical and analytical results of $R_{b,k}^{lfb}$, $R_{c,k}^{lfb}$ and ESSR as functions of $\phi$ for the systems with different values of $N_A$, $N_E$, $K$ and $B_k$, based on which we will study the property of these functions and verify our theoretical analysis in Section IV. Particularly, Fig. 1(a) and 1(b) show $R_{b,k}^{lfb}$, $R_{c,k}^{lfb}$ as functions of $\phi$ for systems with $N_A = 4$ and $N_A = 6$ respectively. As we can see, each $R_{b,k}^{lfb}$ and $R_{c,k}^{approx}$ is concave for $0 < \phi \leq 1$. In addition, $R_{c,k}^{lfb}$ is convex in $\phi \in [x, 1]$ for certain $0 < x < 0.5$ and is concave for $0 < \phi \leq x$. These results agree with our theoretical analysis in Section IV. In addition, we can see the analytical results of $R_{b,k}^{approx}$ can approximate the exact values well.

Fig. 2 and Fig. 3 show the ESSR performance of the systems with different number of antennas ($N_A = 4, 6$), different number of feedback bits of each LR and different SNRs ($P = 15, 25$ dB). We can see each ESSR $R_{sec,sum}$ and each $R_{approx}$ is concave function of $\phi$. In addition, we can see although the obtained analytical approximations can be a little over or under the exact values for some system settings, they are very close to the exact values, especially for large number of $B_k$. Thus, we can employ some numerical method to obtain near optimal solution to maximize each ESSR. Moreover, by comparing these figures, we can see...
the optimal $\phi$ increases as SNR increases. But, with fixed $N_A$, it changes very little as the number of feedback bits of each LR increases. At last, we see increasing the transmit antennas at Alice with fixed feedback bits of each LR does not necessarily increases the ESSR performance. At SNR of 15 dB, increasing $N_A$ can increase ESSR. In the contrast, increasing $N_A$ decreases ESSR at SNR of 25 dB.

Fig. 4(a) - Fig. 4(c) compare the ESSR performance of the THP scheme and linear ZF precoding combined with AN (this was widely studied in the existing literatures) as functions of $\phi$ for the systems with different system settings. We can observe that, for a fixed system setting, the THP outperforms linear ZF precoding scheme for any given power allocation. In addition, very interestingly, the maximum performances of the two schemes reach the top with almost the same power allocation $\phi$, and the maximum performance gap between the two schemes reaches approximately at this top. By comparing results in Fig. 4(b) and Fig. 4(c), we find the increase in the number of Eve’s antennas has great reduction on the ESSR performance.

In Fig. 5, we show the simulated performance results and the obtained analytical approximation of ESSR in (32) of our proposed scheme for a system with $N_A = 6$, $K = 4$ and $N_E = 2$. As the baseline, we also show the performance of linear ZF precoding. For comparisons, we show the simulated performance of the THP with perfect CSIT at Alice and also the corresponding results of THP with quantized CSIT. First, we need to note that, for the MU-MISO wiretap channel with the limited CSI feedback model considered here, to the best of our knowledge, there is no analytical ESR result for the combination of linear ZF precoding with AN allowing for arbitrary number of $N_A$, $N_E$ and $K$ (satisfying the constraint caused by the assumption of zero-power noise at Eve), arbitrary system SNR, arbitrary power allocation coefficient $\phi$, and arbitrary number of feedback bits $B_k$. In this case, there is no previous result that we can turn to to obtain the optimal power allocation coefficients for the corresponding linear ZF precoding scheme with AN. However, as we have observed from the simulation results in Fig. 4 that the optimal power allocation coefficients for THP and ZF are almost the same. Thus, in the simulations we set the optimal $\phi$ of linear ZF precoding scheme the same as that of THP scheme. Particularly, for the systems with perfect CSI at the transmitter, there is no AN received at each LR and it has been shown in the main content of this work that the quality of the CSI has
no effect on the distribution of received AN at Eve. Thus, the analytical ergodic rates of the legitimate channels (exact values or approximations) are available for both THP (e.g. [12]) and ZF precoding (e.g. [32]) and the upper bound of the ergodic rate of Eve’s channel is the same as that of the system with quantized CSI at the transmitter for both THP and ZF precoding. The optimal $\phi$ for the system with perfect CSI at the transmitter is obtained using numerical method based on these analytical results. Not surprisingly, there is a significant gap between the performance with perfect CSIT and quantized CSIT. Moreover, our proposed secure nonlinear precoding algorithm outperforms linear ZF precoding algorithm for all settings. Still, the obtained analytical results of ESSR can approximate the exact values well.

In the following, we show the effect of increasing the number of Eve’s transmit antennas on the ESSR performance through numerical results. In Fig. 6, we show the numerical results for the system where $N_A = 12$ and $K = 4$ are fixed and $N_E$ increases (denoted as Experiment 1) and also for the system where the number of LR $K = 4$ is fixed and $N_A$ needs to be increased as $N_E$ increases (denoted as Experiment 2). For Experiment 1, we set $B_k = 20$ and let $N_E$ increase from 2 to 8 (in order to satisfy the constraint $N_A \geq N_E + K$). For Experiment 2, we set $N_A = N_E + K$ and let the number of feedback bits of each LR increase with $N_A$ as $B_k = 2(N_A - 1)$ in order to keep the accuracy of CSI quantization consistent [20, 22].

For both experiments the SNR of the systems is set to be 25 dB and the power allocation $\phi$ is optimized based the obtained analytical ESSR results. We can see the natural result from the curves that the number of Eve’s antennas $N_E$ has severely detrimental effect on the ESSR performance for both THP and linear ZF precoding. Particularly, given $N_A$ and the fixed accuracy of CSI quantization, the ESSRs of both precoding schemes decrease almost linearly as $N_E$ varies from 2 to 6 and quickly approaches zero as $N_E$ approaches the maximum allowable number. When $N_A$ increases as $N_E$ whilst satisfying the relationship $N_A = N_E + K$, the ESSRs of both precoding schemes approaches zeros even faster than the former case. Thus, for the system to operate properly, $N_A$ should increase much faster than $N_E$ (faster than in the linear mode) whilst keeping the accuracy of CSI quantization.

We show in Fig. 7 the ESR of each LR and the ESSR of the system versus system SNR for the system with $N_A = 4$ and $K = 2$ and $N_E = 2$ and $\phi = 0.6$. The number of feedback bits of LR $k$ for $k = 1, 2$ is set to scale according to the lower bound in (36) in order to maintain a constant ESR loss compared to the perfect CDI case. Here we examine the case of $\Delta R_{th,k} = 2$ bits for both LRs. Note that as suggested by Fig. 3(a), the optimum $\phi^*$ in (36) is very close to 0.6 as the number of feedback bits $B_k$ increases. We can see the ESR losses of both LRs are well controlled as expected by the obtained bit scaling law. The ESSR of the system are also maintained within the pre-defined ESSR loss of 4 bits as the SNR in increases.

At last, we further take into consideration the impact of imperfect channel estimation at each LR on the system ESR performance. We consider in Fig. 8 the ESSR performance of the systems with $N_A = 6$, $K = 4$, $N_E = 2$, and different combinations of the number of feedback bits and the training symbol length $T_r$. For comparisons, we also plot the results of the systems with perfect channel estimation at each LR. We see the expected results that the ESSR increases as the training symbol lengths $T_r$ increases. Moreover, the ESSR improvement by increasing $T_r$ becomes marginal as the system SNR increases. The performance of the systems with channel estimation error in the high SNR region is very close to that of the system with perfect channel estimation. Since increasing the training symbol lengths $T_r$ means reducing the transmission time of the information symbols, further reducing the channel estimation error by increasing $T_r$ is not necessary in high SNR region.
Moreover, we derived the scaling law of the number of each optimal power allocation to the message signals and AN. To use numerical optimization methods to obtain a sub-optimal power allocation scheme can be improved by keeping the legitimate channels’ allowable gap of ESR loss of that LR as the SNR increases. Numerical results have been shown to verify the advantage of the proposed secure nonlinear precoding over the previous linear ZF precoding scheme, the tightness of the derived analytical ESSR approximation and the effectiveness of the feedback bit scaling law.

**VII. CONCLUSIONS**

In this paper, we have enriched the research of the AN-aided secure communications in multiuser multi-antenna systems. Employing THP based on quantized CSIT through limited feedback, we have proposed an AN-aided nonlinear transceiver for secure communications in MU-MISO systems. For the scenario where the quantized CSI can be obtained by the passive eavesdropper, we also derived the analytical approximations for the ergodic rate of each LR’s messages over legitimate channel and the ESSR with general settings. Using information-theoretical method, we have proved the ESSR performance of the proposed THP scheme can be improved by keeping the legitimate channels’ (quantized) CSI confidential from eavesdropper. Through theoretic analysis and numerical simulations, we have also investigated the convexity and concavity properties for the ergodic rate of each LR’s messages, the worst-case ergodic rate of each LR’s messages at Eve and the ESSR. Based on the obtained convexity/concavity properties and the analytical approximation of the ESSR, we have proposed to use numerical optimization methods to obtain a sub-optimal power allocation to the message signals and AN. Moreover, we derived the scaling law of the number of each LR’s feedback bits to maintain a pre-determined maximum

**APPENDIX**

**A. Proof of Theorem 1**

$R_{LFB}^{b,k}$ of (14) can be re-written as follows:

$$R_{LFB}^{b,k} = \mathbb{E} \left\{ \log_2 \left[ 1 + \frac{\phi P}{K \sigma_{b,k}^2} \rho^2 r_{b,k} \right] \right\}$$

$$\times \left( \cos^2 \theta_k + \frac{\phi P}{K \sigma_{b,k}^2} \rho^2 \sin^2 \theta_k \| \hat{h}_{k,Q} H \| ^2 \right)$$

$$\times \left( \left( 1 - \frac{\phi P}{K \sigma_{b,k}^2} \rho^2 r_{b,k} \right) \left( 1 + \frac{\phi P}{K \sigma_{b,k}^2} \rho^2 r_{b,k} \right) \right)$$

$$+ \left( \left( 1 - \frac{\phi P}{K \sigma_{b,k}^2} \rho^2 r_{b,k} \right) \left( 1 + \frac{\phi P}{K \sigma_{b,k}^2} \rho^2 r_{b,k} \right) \right)$$

$$\mathbb{E} \left\{ \log_2 \left[ 1 + \frac{\phi P}{K \sigma_{b,k}^2} r_{b,k} \| \hat{h}_{k,Q} N \| ^2 \right] \right\} + \sin^2 \theta_k \right\}$$

- $\mathbb{E} \left\{ \log_2 \left[ 1 + \frac{\phi P}{K \sigma_{b,k}^2} \rho^2 \sin^2 \theta_k \| \hat{h}_{k,Q} H \| ^2 \right] \right\}$
\[
\begin{align*}
E \left[ \log_2 \left( \cos^2 \theta_1 + \frac{1 + \phi_P K \rho_{b,k}^2}{(N_A - K) \sigma_{b,k}^2} H  \frac{\rho_{b,k}^2}{(N_A - K) \sigma_{b,k}^2} \| \hat{h}_k \hat{Q}^H \|^2 \right) \sin^2 \theta_1 \right] \\
\geq E \left[ \log_2 \left( \cos^2 \theta_1 + \| \hat{h}_k \hat{Q}^H \|^2 \sin^2 \theta_1 + \frac{1 + \phi_P (N_A - K) \sigma_{b,k}^2 \rho_{b,k}^2}{(N_A - K) \sigma_{b,k}^2} \| \hat{h}_k \hat{Q}^H \|^2 \sin^2 \theta_1 \right) \right] \\
\geq E \left[ \log_2(\cos^2 \theta_1) + \log_2 \left( 1 + \| \hat{h}_k \hat{Q}^H \|^2 \sin^2 \theta_1 \right) \right] + \log_2 \left( 1 + \frac{1 + \phi_P (N_A - K) \sigma_{b,k}^2 \rho_{b,k}^2}{(N_A - K) \sigma_{b,k}^2} \| \hat{h}_k \hat{Q}^H \|^2 \sin^2 \theta_1 \right) \right] \\
\geq E \left[ \log_2(\cos^2 \theta_1) + \| \hat{h}_k \hat{Q}^H \|^2 \sin^2 \theta_1 + \log_2 \left( 1 + \frac{1 + \phi_P (N_A - K) \sigma_{b,k}^2 \rho_{b,k}^2}{(N_A - K) \sigma_{b,k}^2} \| \hat{h}_k \hat{Q}^H \|^2 \sin^2 \theta_1 \right) \right] \\
= E \left[ \log_2(\cos^2 \theta_1) + \| \hat{h}_k \hat{Q}^H \|^2 \sin^2 \theta_1 + \log_2 \left( 1 + \frac{1 + \phi_P (N_A - K) \sigma_{b,k}^2 \rho_{b,k}^2}{(N_A - K) \sigma_{b,k}^2} \| \hat{h}_k \hat{Q}^H \|^2 \sin^2 \theta_1 \right) \right] \\
- \log_2 \left( 1 + \phi_P \rho_{b,k}^2 \right) \\
\end{align*}
\]

Moreover, it was shown in [21] that the PDF of \( \cos^2 \theta_k \) is given by \( f_{\cos^2 \theta_k}(x) = 2B_k(N_A - 1) [1 - (1 - x)^{N_A - 1}] - 1 \), \( 0 \leq x \leq 1 \). Then, using the above PDF and change of variables, after some manipulations we can obtain

\[
E \left[ \sin^2 \theta_{1,k} \right] = \int_0^1 \frac{1 - x}{x} f_{\cos^2 \theta_{1,k}}(x) dx \\
= \int_0^1 \frac{1 - x}{x} 2B_k(N_A - 1) [1 - (1 - x)^{N_A - 1}]^{2b_k - 1} \times (1 - x)^{N_A - 2} dx \\
= 2B_k(N_A - 1) \int_0^1 \sum_{m=0}^1 t^{m+N_A-1} [1 - t^{N_A - 1}]^{2b_k - 2} dt \\
= 2B_k N_A - 2 \sum_{m=0}^1 \beta \left( \frac{m + N_A}{N_A - 1}, \frac{2B_k - 1}{2B_k - 1} \right). 
\]

Then, \( \mu_k \) follows from the above results.

For \( k \geq 2 \), we have (44) and (45) at the top of the next page, where, again, (44) is obtained by using Jensen’s inequality and (45) is obtained by using the inequality \( \log_2(1 + x) \geq x \) for \( 0 \leq x \leq 1 \). \( \mu_k = E \left[ \cos^2 \theta_k + \| \hat{h}_k \hat{Q}^H \|^2 \sin^2 \theta_k \right] \). Using the properties of the independent \( \varepsilon_k = \| \hat{h}_k \hat{Q}^H \|^2 \) and \( \hat{r}_{k,k}^2 \) and the regularized incomplete beta function \( I_x(a, b) \)

\[
I_y(a, b - a + 1) = \sum_{j=a}^{b} \binom{b}{j} y^j (1 - y)^{b-j}, 
\]

we can obtain

\[
p_k = \Pr(\varepsilon_k) = \Pr \left\{ \| \hat{h}_k \hat{Q}^H \|^2 \leq \hat{r}_{k,k}^2 \right\} 
\]
\[
\mathbb{E} \left[ \log_2 \left( \cos^2 \theta_k + \frac{1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{Q}^H \|}{1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2} \sin^2 \theta_k \right) \right]
\]
\[
\geq p_k \mathbb{E} \left[ \log_2 \left( \cos^2 \theta_k + \frac{\| \hat{h}_k \hat{Q}^H \|^2}{1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2} \sin^2 \theta_k \right) \right] + \mathbb{E} \left[ \log_2 \left( 1 + \frac{\| \hat{h}_k \hat{Q}^H \|^2}{1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2} \sin^2 \theta_k \right) \right] E_k \]
\[
+ (1 - p_k) \mathbb{E} \left[ \log_2 \left( 1 + \frac{\| \hat{h}_k \hat{Q}^H \|^2}{1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2} \sin^2 \theta_k \right) \right] E_k \]
\[
\geq p_k \mathbb{E} \left[ \log_2 \left( \cos^2 \theta_k + \frac{\| \hat{h}_k \hat{Q}^H \|^2}{1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2} \sin^2 \theta_k \right) \right] + \mathbb{E} \left[ \log_2 \left( 1 + \frac{\| \hat{h}_k \hat{Q}^H \|^2}{1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2} \sin^2 \theta_k \right) \right] E_k \]
\[
+ (1 - p_k) \mathbb{E} \left[ \log_2 \left( 1 + \frac{\| \hat{h}_k \hat{Q}^H \|^2}{1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2} \sin^2 \theta_k \right) \right] E_k \]  
(44)
\[
\geq p_k \mathbb{E} \left[ \log_2 \left( \cos^2 \theta_k + \frac{\| \hat{h}_k \hat{Q}^H \|^2}{1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2} \sin^2 \theta_k \right) \right] \]
\[
+ \mathbb{E} \left[ \log_2 \left( 1 + \frac{\| \hat{h}_k \hat{Q}^H \|^2}{1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2} \sin^2 \theta_k \right) \right] E_k \]
\[
+ (1 - p_k) \mathbb{E} \left[ \log_2 \left( 1 + \frac{\| \hat{h}_k \hat{Q}^H \|^2}{1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2} \sin^2 \theta_k \right) \right] E_k \]
\[
+ p_k \mathbb{E} \left[ \log_2 \left( \cos^2 \theta_k + \frac{\| \hat{h}_k \hat{Q}^H \|^2}{1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2} \sin^2 \theta_k \right) \right] - \mathbb{E} \left[ \log_2 \left( 1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2 \sin^2 \theta_k \right) \right] E_k \]  
(45)

which is given by (17). \( \Pr(\tilde{E}_k) = 1 - p_k \). Similarly, for \( k \geq 2 \), we can obtain

\[
\mathbb{E} \left[ \frac{||\hat{h}_k \hat{Q}^H||^2}{|\tilde{r}_{k,k}|^2} \big| E_k \right] = \int_0^1 \int_0^1 \int_0^y x f_{\tilde{r}_{k,k}}(x) f_{\tilde{r}_{k,k}}(y) dy dx dy
\]

\[
= \int_0^1 \int_0^y \frac{x(y-N_{\hat{h},k} (1-\gamma)^{K-1})}{\beta(N_{\hat{h},k} + 1, K-1)} dy dx
\]

\[
= \frac{K-1}{N_{\hat{h},k} - K} \sum_{j=K}^{N_{\hat{h},k}-1} \frac{\left( N_{\hat{h},k} - j \right)}{\beta(N_{\hat{h},k} + 1, K-1)} dy\]

\[
\int_0^1 y^j |1 - y|^{N_{\hat{h},k} - j - 1} dy
\]

\[
\beta(N_{\hat{h},k} - j + 1, K-1) dy
\]

which is given by (18), where we have used again the property in (46). Then, \( \mu_k \) and \( \mathbb{E} \left[ \log_2 \left( \cos^2 \theta_k + \frac{\| \hat{h}_k \hat{Q}^H \|^2}{1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2} \sin^2 \theta_k \right) \right] - \mathbb{E} \left[ \log_2 \left( 1 + \frac{\partial P}{K\sigma_{\hat{h},k}^2} \rho_k^2 \| \hat{h}_k \hat{N} \|^2 \sin^2 \theta_k \right) \right] E_k \] in (45) follows from the obtained results above.

Notice that, since the random terms inside each \( \log_2 \) are correlated with each other, the expectations of the \( \log_2 \) terms in (41) (44) and (45) are very difficult to obtain. Thus, we will approximate each expectation using the following result from [28, pp. 215].

**Lemma 5**: Let \( x, y \) be two random variables with means \( \mu_x, \mu_y \), variances \( \sigma_x^2, \sigma_y^2 \) and covariance \( C_{x,y} \). If the function \( g(x, y) \) is sufficiently smooth near the point \( (\mu_x, \mu_y) \), then the mean of \( g(x, y) \) can be approximated as

\[
\mathbb{E}[g(x, y)] \approx g(\mu_x, \mu_y) + \frac{1}{2} \left( \frac{\partial^2 g(\mu_x, \mu_y)}{\partial x^2} \sigma_x^2 + \frac{\partial^2 g(\mu_x, \mu_y)}{\partial y^2} \sigma_y^2 + 2 \frac{\partial^2 g(\mu_x, \mu_y)}{\partial x \partial y} C_{x,y} \right).
\]

(47)

In this work, \( g(x, y) \) is defined as \( g(x, y) = \ln(1 + Ax + By) \), where \( A \geq 0 \) and \( B \geq 0 \), but \( AB > 0 \). It is easy to obtain

\[
\frac{\partial^2 g(\mu_x, \mu_y)}{\partial x^2} = - \frac{A^2}{(1 + A\mu_x + B\mu_y)^2}
\]

(48)

\[
\frac{\partial^2 g(\mu_x, \mu_y)}{\partial y^2} = - \frac{B^2}{(1 + A\mu_x + B\mu_y)^2}
\]

(49)

\[
\frac{\partial^2 g(\mu_x, \mu_y)}{\partial x \partial y} = - \frac{AB}{(1 + A\mu_x + B\mu_y)^2}
\]

(50)

Define \( x_k = \rho_k^2 |\tilde{r}_{k,k}|^2 \) with \( |\tilde{r}_{1,1}|^2 = 1 \), \( y_k = \rho_k^2 \| \hat{h}_k \hat{N} \|^2 \sin^2 \theta_k \). It has been shown in [12]
that $x_k$ is distributed as $\text{Gamma}(N_A - k + 1, 1)$. The moments of $\rho_k^2$ are $E[\rho_k^{2m}] = \Gamma(m + N_A)/(\Gamma(N_A))$ [28]. Thus, $E[\rho_k^4] = (N_A + 1)N_A$. According to distribution of $|\hat{r}_{k,k}|^2$ given in Lemma 1, we have $E[|\hat{r}_{k,k}|^2] = N_A - k + 1$.

Then, we have $\mu_{x_k} = E[x_k] = N_A - k + 1, \sigma_{x_k}^2 = \text{Var}(x_k) = N_A - k + 1, \mu_{y_k} = E[y_k] = \frac{N_A}{N_A - 1}$. $E[y_k^2] = E[\rho_k^2] = E[|\hat{r}_{k,k}|^4] E[\sin^2 \theta_k] = \frac{(N_A + 1)(N_A - K + 1)(N_A - K)K\beta^2}{(N_A - 1)(N_A - K)^2}$, respectively.

Thus, $\sigma_{x_k}^2 = \text{Var}(x_k) = E[y_k^2] - (E[y_k])^2$ and $E[x_k y_k | \mathcal{E}_k] = E[\rho_k^2 | \mathcal{E}_k] |\hat{r}_{k,k}|^2 E[N_k^2] E[\sin^2 \theta_k] E[\mathcal{E}_k] = E[\rho_k^2] |\hat{r}_{k,k}|^2 E[N_k^2] E[\sin^2 \theta_k] E[\mathcal{E}_k]$, where

$E[|\hat{r}_{k,k}|^4] E[\sin^2 \theta_k] E[\mathcal{E}_k] = \int_0^1 f_{\rho_k}(x) \int_0^1 y \rho_k^2 |\hat{r}_{k,k}|^2 \sin^2 \theta_k |\hat{E}_k|^2 dxdy = \int_0^1 \frac{x^{N_A - K - 1}}{\beta(N_A - K, K - 1)} \beta(N_A - k + 1, K - 1) \times (N_A - k + 1) dx = \frac{N_A - k + 1}{N_A} \sum_{j=N_A-k+2}^{N_A} \binom{N_A}{j} \frac{N_A}{N_A - 1} \times \frac{N_A}{N_A - 1} \frac{\beta(N_A - k + 1, K - 1)}{\beta(N_A - K, K - 1)},$ (54)

which is given in (19), where again we used (46). Substituting into (54) the results obtained above and with Lemma 1 in [20], we have $E[x_k y_k | \mathcal{E}_k] = (N_A + 1)N_A K\beta^2/N_A - 1$. With (54), we have

$E[|\hat{r}_{k,k}|^4] E[\sin^2 \theta_k] E[\mathcal{E}_k] = \frac{(N_A - k + 1)(N_A - K)}{N_A} - \kappa_k \triangleq \hat{\kappa}_k.$ (55)

Similarly, we have $E[x_k y_k | \mathcal{E}_k] = (N_A + 1)N_A K\beta^2/N_A - 1$. Then, the conditional covariance of $x_k$ and $y_k$ can be obtained as

$C_{x_k, y_k | \mathcal{E}_k} = E[x_k y_k | \mathcal{E}_k] - \mu_{x_k} \mu_{y_k} = \frac{(N_A - k + 1)(N_A - K)}{N_A} - \kappa_k \triangleq \hat{\kappa}_k.$

B. The Proof of Lemma 2

Firstly, we have $I(x; y_e) = H(y_e) - H(y_e | x)$, and

$H(y_e) \leq \log_2 \left( \frac{1}{\pi e} N_e \det \left( \frac{\phi P}{K} G_Q G_Q^H + \frac{1 - \phi}{N_A - K} G_N G_N^H \right) \right).$ (56)

where the inequality is due to the fact that, with the average power constraint of $x$, the maximum differential entropy of $y_e$ is achieved if and only if $x$ is a circularly symmetric complex Gaussian vector. In addition, we have

$H(y_e | x) = \log_2 \left( \frac{1}{\pi e} N_e \det \left( \frac{1 - \phi}{N_A - K} G_N G_N^H \right) \right).$ (57)
(22) follows by combining (56) and (57).

According to (4), given quantized CSI $\mathbf{H}$, $x$ can be uniquely determined from $s$ by THP and vice versa. We denote the THP operation as a one-to-one mapping $T$ from $(s, \mathbf{H})$ to $x$, which is given as $x = T(s, \mathbf{H})$. In the following, we prove that, when Eve knows $\mathbf{H}$ (given $\mathbf{H}$) and also knows the processing method of Alice, $s, y, x_e$ form a Markov chain and vice versa, i.e., $s \leftrightarrow x \leftrightarrow y_e$ and $x \leftrightarrow s \rightarrow y_e$. Thus, according to data-processing inequality [34], we obtain $I \left(x; y_e | \mathbf{H} \right) = I \left(s; y_e | \mathbf{H} \right)$. First, since $p(y_e | s, \mathbf{H}, x) = p(y_e | s, \mathbf{H})$, we have $p(x, y_e | s, \mathbf{H}) = p(x | s, \mathbf{H})p(x | s, \mathbf{H}) = p(y_e | s, \mathbf{H})p(x | s, \mathbf{H})$, which means $x$ and $y_e$ are conditionally independent given $s$ and $\mathbf{H}$. In fact, given $s, \mathbf{H}$ the conditional probability mass function of $x$ is $p(x | s, \mathbf{H}) = \left\{ \begin{array}{ll} 1 & x = T(s, \mathbf{H}) \\ 0 & x \neq T(s, \mathbf{H}) \end{array} \right.$.

Thus, $x \leftrightarrow s \leftrightarrow y_e$ holds. For the same reason, it is easy to see $p(s, y_e | x, \mathbf{H}) = p(s | x, \mathbf{H})p(y_e | x, \mathbf{H}) = p(y_e | x, \mathbf{H})p(s | x, \mathbf{H})$, which means $s$ and $y_e$ are conditionally independent given $x$ and $\mathbf{H}$. Thus, $s \leftrightarrow x \leftrightarrow y_e$ holds.

Using the properties of mutual information and differential entropy, it is easy to prove [34]

$$I \left(s; y_e | \mathbf{H} \right) = \sum_{i=1}^{K} I \left(s_i; y_e | \mathbf{H}, s_1, s_2, \cdots, s_{i-1} \right)$$

$$= \sum_{i=1}^{K} \left[H \left(s_i | \mathbf{H}, s_1, s_2, \cdots, s_{i-1} \right) - H \left(s_i | \mathbf{H}, y_e, s_1, s_2, \cdots, s_{i-1} \right) \right].$$

Since $s_k$ for $k = 1, 2, \cdots, K$ and $\mathbf{H}$ are independent with each other, we conclude that $H(s_k | \mathbf{H}, s_1, s_2, \cdots, s_{i-1} ) = H(s_k)$ and $H(s_1 | \mathbf{H}, y_e, s_1, s_2, \cdots, s_{i-1} ) = H(s_1 | \mathbf{H}, y_e)$. Thus, we have $I \left(s; y_e | \mathbf{H} \right) = \sum_{k=1}^{K} I \left(s_k; y_e | \mathbf{H} \right)$. In addition, it is obvious that the order of $s_k$ ($k = 1, 2, \cdots, K$) in $s$ does not change the distribution of $x$, thus does not change $I( x; y_e)$ and $I \left(s; y_e | \mathbf{H} \right)$.

Thus, we conclude that $I \left(s_k; y_e | \mathbf{H} \right)$ for all $k$ are equal.

Moreover, $y_e$ depends on $x$ and also $\mathbf{H}$ through $\bar{Q}$ from LQ decomposition of $\mathbf{H}$, and $x$ is determined from $(\mathbf{H}, s)$ through THP, i.e., $\mathbf{H}$ and $s$ are implicitly related through $y_e$. Thus, given $y_e$, $s$ implicitly becomes conditionally independent with $\mathbf{H}$. Then, we have $H(s | y_e) > H(s | \mathbf{H}, y_e)$, since condition reduces entropy [34]. Then, we can prove

$$I \left(s; y_e | \mathbf{H} \right) > I \left(s; y_e, \mathbf{H} \right)$$

as follows.

$$I \left(s; y_e | \mathbf{H} \right) = H \left(s | \mathbf{H} \right) - H \left(s | y_e, \mathbf{H} \right)$$

$$> H \left(s \right) - H \left(s | y_e \right) = I \left(s; y_e \right).$$

C. Proof of Lemma 4

First, using (10) we can rewrite $R_{b,k}^{LFB}$ as

$$R_{b,k}^{LFB} \left(\phi \right) = E \left[ \log_2 \left( 1 + \frac{D_1 \phi}{D_2 \phi + D_3} \right) \right],$$

where $D_1 = \frac{p_{\mathbf{y}^Q \mathbf{x}^Q}^2 |E|_{k,k}^2 \cos^2 \theta_k}{K \sigma_{s,k}^2}$, $D_2 = p_{\mathbf{y}^Q \mathbf{x}^Q}^2 \sin^2 \theta_k \hat{b}_k \hat{Q}^Q \|2 \|^2 \frac{p_{\mathbf{y}^Q \mathbf{x}^Q}^2 |E|_{k,k}^2 \sin^2 \theta_k |\mathbf{h}_k \mathbf{N}^2}{(N-K)\sigma_{s,k}^2}$, and $D_3 = \frac{p_{\mathbf{y}^Q \mathbf{x}^Q}^2 |E|_{k,k}^2 \sin^2 \theta_k |\mathbf{h}_k \mathbf{N}^2|^2 + 1}{1}$, all of which are not related to $\phi$. Define

$$g(\phi, x, y, z) \triangleq \log_2 \left( 1 + \frac{\phi}{D_2 \phi + D_3} \right) f_{D_1, D_2, D_3} (x, y, z),$$

where $D_1, D_2, D_3 (x, y, z)$ is the joint PDF of $D_1, D_2, D_3$, which is obviously a continuous function of $x, y, z$. Thus, $R_{b,k}^{LFB} (\phi)$ can be written as $R_{b,k}^{LFB} (\phi) = \int \int g(\phi, x, y, z) dxdydz$, where the integration region is according to the joint distribution of $D_1, D_2, D_3$.

It is obvious that $g(\phi, x, y, z)$, $\phi(x, y, z)$, and $\phi(x, y, z)$ are continuous functions of variables $\phi$, $x$, $y$, and $z$ for the defined regions. It can be found in most mathematical analysis textbooks (e.g. [35]) that the order of integration and differentiation of $g(\phi, x, y, z)$ is interchangeable. After some algebra manipulations, we can obtain (61) at the top of the next page. Thus, $R_{b,k}^{LFB}$ is concave for $0 < \phi \leq 1$.

For the same reason, we can obtain

$$\frac{d^2 E[R_{\text{eve,sum}}(\phi)]}{d\phi^2}$$

as follows. First, from (22) we rewrite $E[R_{\text{eve,sum}}(\phi)]$ as

$$E \left[ R_{\text{eve,sum}}(\phi) \right] = E \left[ \log_2 \left( 1 + \frac{\phi(N-A-K)}{(1-\phi)A} (G_Q G_Q^H) \right) \right] \times (G_N G_N^H)^{-1} \left[ \sum \lambda_i \right].$$

where $L$ is the number of non-zero eigenvalues of matrix $(G_Q G_Q^H)$, $(G_N G_N^H)^{-1} \left[ \sum \lambda_i \right]$ and (62) is due to the fact that $(G_N G_N^H)^{-1} \left[ \sum \lambda_i \right]$ with probability of 1. After some manipulations, we can obtain

$$\frac{d^2 E[R_{\text{eve,sum}}(\phi)]}{d\phi^2} = \left[ \sum \phi_i \right],$$

where $\phi_i = \frac{N-A-K}{K} \lambda_i$.

The lemma follows by observing the following facts. First, it is easy to see when $0 \leq \phi < 1$, $\frac{d^2 E[R_{\text{eve,sum}}(\phi)]}{d\phi^2} > 0$. Thus, $E[R_{\text{eve,sum}}(\phi)]$ is a convex function. Moreover, for $0 < \phi < 0.5$, the numerator in (63) $(2\phi - 1) t_i + 2(1 - \phi)$ $\geq 0$ if $\lambda_i < \frac{K}{N-A-K} \frac{2(1-\phi)}{1-2\phi}$. In addition, $\frac{K}{N-A-K} \frac{2(1-\phi)}{1-2\phi}$ is an increasing function of $\phi$, and $\frac{K}{N-A-K} \frac{2(1-\phi)}{1-2\phi} \to +\infty$ as $\phi \to 0^-$, and $\frac{K}{N-A-K} \frac{2(1-\phi)}{1-2\phi} \to 2\frac{K}{N-A-K}$ as $\phi \to 0^+$. It is obvious that $\frac{d^2 E[R_{\text{eve,sum}}(\phi)]}{d\phi^2}$ is a continuous function of $\phi$. Thus, there must be a certain number $\varepsilon < 0.5$ that, for $\varepsilon < \phi < 1$, $\frac{d^2 E[R_{\text{eve,sum}}(\phi)]}{d\phi^2} > 0$. However, it is quite intractable to obtain the closed-form expression of $\frac{d^2 E[R_{\text{eve,sum}}(\phi)]}{d\phi^2}$ based on (63). Thus, we conclude that for certain $0 \leq \varepsilon < 0.5$, $E[R_{\text{eve,sum}}(\phi)]$ is a convex function in $[\varepsilon, 1)$, and is a concave function in $(0, \varepsilon]$, where $\varepsilon$ depends on system parameters. We will further study this by numerical results in Section VI.
\[
\frac{d^2 R^LFB_{b,k}(\phi)}{d\phi^2} = \int \int \frac{d^2 g(\phi, x, y, z)}{d\phi^2} \, dx \, dy \, dz
\]
\[
= \mathbb{E}_{\{D_1, D_2, D_3\}} \left[ -\frac{D_1 D_3 [D_2 ((D_1 + D_2)\phi + D_3) + (D_1 + D_2)(D_2\phi + D_3)]}{(D_2\phi + D_3)^2 [(D_1 + D_2)\phi + D_3]^2} \right] \leq 0. \quad (61)
\]

\[
R^LFB_{b,k}(\phi^*) \geq \mathbb{E} \left[ \log_2 \left( 1 + \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right) \right] - \mathbb{E} \left[ \log_2 \left( 1 + \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right) \right]
\]
\[
= \mathbb{E} \left[ \log_2 \left( 1 + \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right) \right] - \mathbb{E} \left[ \log_2 \left( 1 + \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right) \right] \quad \text{(a)}
\]
\[
\leq \mathbb{E} \left[ \log_2 \left( 1 + \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right) \right] - \mathbb{E} \left[ \log_2 \left( 1 + \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right) \right]
\]
\[
= 2 B_b \sum_{m=0}^{N_A-2} \frac{\Gamma \left( \frac{m+N_A}{N_A-1} \right)}{\Gamma \left( \frac{m+1}{N_A-1} \right)} \frac{\Gamma \left( \frac{2B_b}{N_A-1} \right)}{\Gamma \left( \frac{2B_b}{N_A-1} \right)} = 2 B_b \sum_{m=0}^{N_A-2} \frac{\Gamma \left( \frac{m+N_A}{N_A-1} \right)}{\Gamma \left( \frac{m+1}{N_A-1} \right) \Gamma \left( \frac{2B_b}{N_A-1} \right)} \leq 1 \quad \text{(64)}
\]

\section*{D. An Upper Bound of Each LR’s ESR Loss}

According to (10), \( R^LFB_{b,k}(\phi^*) \) can be lower bounded by Jensen’s inequality as (64) at the top of this page. Thus, according to (34) we have
\[
\Delta R_{sec,k} \leq \mathbb{E} \left[ \log_2 \left( 1 + \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right) \right] - \mathbb{E} \left[ \log_2 \left( 1 + \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right) \right]
\]
\[
= \log_2 \left( 1 + \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right) \left( 1 + \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right)
\]
\[
= \log_2 \left( 1 + \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right) \left( 1 - \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right)
\]
\[
= \log_2 \left( 1 + \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right) \left( 1 - \frac{\phi^* P}{K \sigma^2_{b,k}} \rho_k^2 |f_{b,k}|^2 \cos^2 \theta_k \right)
\]
\[
\approx \mathbb{E} \left[ \sin^2 \theta_k \right] - \frac{\Gamma \left( \frac{m+N_A}{N_A-1} \right)}{\Gamma \left( \frac{m+1}{N_A-1} \right) \Gamma \left( \frac{2B_b}{N_A-1} \right)} = \frac{\Gamma \left( \frac{m+N_A}{N_A-1} \right)}{\Gamma \left( \frac{m+1}{N_A-1} \right) \Gamma \left( \frac{2B_b}{N_A-1} \right)}
\]
\[
\leq 1 \quad \text{(65)}
\]
\[
\] where (a) follows from Jensen’s inequality and \( \mathbb{E} \left[ \log_2 \left( 1 + X \right) \right] \geq \log_2 \left( 1 + \frac{1}{e} \right) \) and (65) follows from \( \mathbb{E} \left[ \frac{1}{\rho_k^2} \right] = \frac{1}{N_A-1} \). \( \mathbb{E} \left[ \sin^2 \theta_k \right] \) is obtained in Appendix A and given by (43). Then, using \( \beta(x, y) = \frac{\Gamma(\phi)(y)}{(x+y)} \) and \( \Gamma(x+1) = x \Gamma(x) \), we have
\[
\mathbb{E} \left[ \frac{\sin^2 \theta_k}{\cos^2 \theta_k} \right] = \frac{1}{N_A-1}
\]

\begin{thebibliography}{9}
\end{thebibliography}
A. Papoulis and S. U. Pillai, Probability, Random Variables and

M. Chiani, M. Z. Win, and H. Shin, “MIMO Networks: The effects

I. S. Gradshteyn and I. M. Ryzhik, Table of Integrals, Series, and

M. Arakawa, “Computational Workloads for Commonly Used Signal

L. Sun, R. Wang, H. Wang, and V. C. Leung, “A novel nonlinear

G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, “Multiuser

S. Zarei, W. Gerstacker, and R. Schober, “Low-complexity hybrid

C. K. Au-Yeung and D. J. Love, “On the performance of random vec-

N. Jindal, “MIMO broadcast channels with finite-rate feedback,” IEEE

X. Lu, K. Zu, and R. Lamare, “Lattice-reduction aided successive

S. Zarei, W. Gerstacker, and R. Schober, “Low-complexity hybrid

C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber,

M. Pei, A. L. Swindlehurst, D. Ma, and J. Wei, “Adaptive limited

C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber,

M. Sun and M. Lei, “Quantized CSI-based Tomlinson-Harashima pre-

S. Goel and R. Negi, “Guaranteeing secrecy using artificial noise,” IEEE


