Optimal Pilots for Maximal Capacity of Secret Key Generation

Qiping Zhu, Yingbo Hua
University of California, Riverside, CA 92521, USA
Email: qzhu005@ucr.edu, yhua@ece.ucr.edu

***Abstract*—Through a wireless channel, two users can transmit pilot signals to each other, and then utilize the received signals to agree upon a secret key through communications in a public channel without leaking any secret about this key to anyone else. This paper addresses the optimization of the two pilots to maximize the capacity of secret key generation for a MIMO channel with any given receive and transmit correlation matrices. This study shows how the globally optimal pilots can be obtained if the transmission power is either low or high. For an arbitrary transmission power, the algorithm developed is at least locally optimal. Comparison to pilots based on minimum channel estimation errors and uniform power distribution is also presented. It is also shown that the designed pilots meet the requirement for anti-eavesdropping channel estimation.**

***Index Terms*—Secret key generation, pilot design, channel state information, MIMO channel**

I. Introduction

A secret key shared by any two legitimate users is essential for wireless security purposes such as authentication of each other and protection of information via encryption for each other. But such a secret must be periodically updated between the users or otherwise it could be cracked by attackers over time. One of the ways to update the secret shared by two users is to exploit the correlation between two signals received by the users via secret key agreement protocol without leaking any information to eavesdropper (Eve) [1]–[9].

In a recent work [10], a method called anti-eavesdropping channel estimation (ANECE) is proposed. For two users equipped with full-duplex MIMO transceivers, ANECE lets both users transmit their packets concurrently where the pilots in both packets share a common subspace so that Eve is unable to obtain its CSI, which then degrades substantially Eve's capacity to detect the secret information in the packets exchanged between the users. Corresponding to the pilots in the packets, the two users receive correlated signals due to the reciprocal nature of their common channel. These two signals can be utilized by the users via secret key agreement protocol to generate a shared secret information in addition to the secret in the packets.

In this paper, we study how to find the optimal pilots to maximize the capacity C_{key} of secret key generation for a MIMO channel with any given transmit and receive correlation matrices. For this preliminary work, we consider the case of two users subject to a power constraint, and the resulting optimal pilots always share a common row subspace as required by ANECE. Furthermore, this work has the following novelties against the known results in the literature:

- In [5], [8], a similar problem was addressed. But [5] assumes the transmit and receive correlation matrices of the users to be the identity matrices and [8] assumes them to be diagonal matrices.
- In [9], non-diagonal transmit and receive correlation matrices were considered. But they did not establish the optimality of a pair of diagonalization matrices, which is however established in this paper.
- This paper provides a novel insight into the equivalence among the mutual information (MI) between observed signals by the users, the MI between the minimum mean squared error (MMSE) estimates of a common channel matrix $\mathbf{H}$ by the users, and the MI between the maximum likelihood estimates (MLE) of $\mathbf{H}$ by the users.
- Furthermore, this paper establishes the optimal pilots under either high or low transmission power P, an algorithm to compute locally optimal pilots for any given P, and a detailed analysis of the effect of channel correlations on C_{key}.

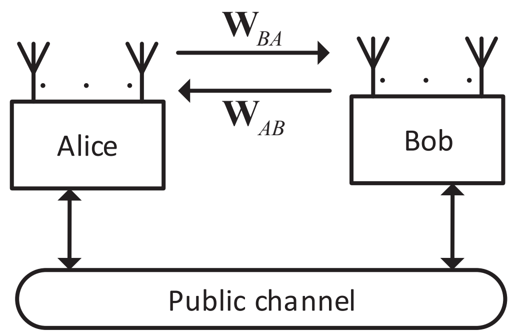

Fig. 1. System model for secret key generation between Alice and Bob

II. System Model

The system under consideration is illustrated in Fig. 1 where the two users are referred to as Alice and Bob equipped with N_A and N_B antennas respectively. The channel matrix from Alice to Bob is denoted by $\mathbf{W}_{BA} \in \mathbb{C}^{N_B \times N_A}$, and that from Bob to Alice is by $\mathbf{W}_{BA} = \mathbf{W}_{AB}^T$ where the reciprocal property is applied. Let $\mathbf{P}_A \in \mathbb{C}^{N_A \times T}$ be the pilot matrix transmitted by Alice over T symbol intervals, and similarly

The authors are with This work was supported in part by the Army Research Office under Grant Number W911NF-17-1-0581. The views and conclusions contained in this document are those of the authors.

$\mathbf{P}_B \in \mathbb{C}^{N_B \times T}$ is the pilot matrix from Bob. Regardless of whether both users are operating in full-duplex or half-duplex mode, we assume that the signals received by Alice and Bob can be represented by $\mathbf{Y}_A \in \mathbb{C}^{N_A \times T}$ and $\mathbf{Y}_B \in \mathbb{C}^{N_B \times T}$ respectively, and

$$\begin{aligned} \mathbf{Y}_A &= \mathbf{W}_{AB}\mathbf{P}_B + \mathbf{N}_A \\ \mathbf{Y}_B &= \mathbf{W}_{AB}^T\mathbf{P}_A + \mathbf{N}_B \end{aligned} \tag{1}$$

where $\mathbf{N}_A \in \mathbb{C}^{N_A \times T}$ and $\mathbf{N}_B \in \mathbb{C}^{N_B \times T}$ are noise matrices with i.i.d. $\mathcal{CN}(0,1)$ entries. Furthermore, we assume that $\mathbf{W}_{AB} = \mathbf{R}_A^{\frac{1}{2}}\mathbf{H}\mathbf{R}_B^{\frac{T}{2}}$ with $\mathbf{R}_A = \mathbf{R}_A^{\frac{1}{2}}\mathbf{R}_A^{\frac{H}{2}}$ and $\mathbf{R}_B = \mathbf{R}_B^{\frac{1}{2}}\mathbf{R}_B^{\frac{H}{2}}$ as known channel correlation matrices but with $\mathbf{H} \in \mathbb{C}^{N_A \times N_B}$ consisting of i.i.d $\mathcal{CN}(0,\sigma^2)$ entries. If full-duplex is applied, we need T to be no larger than the channel coherent time T_c (measured in number of sampling intervals). If half-duplex is applied, we need $2T \le T_c$.

Provided that Eve is more than half-wavelength away from both users, we can assume that the receive CSI at Eve is independent of the CSI matrix $\mathbf{H}$ between the users. Then, it is known [11, Th. 4.1] that the secret key capacity in bits per realization of $\mathbf{H}$ achievable by secret key agreement protocol over many channel coherent periods is given by $C_{key} = I(\mathbf{Y}_A;\mathbf{Y}_B)$. In this paper, we will address how to choose $\mathbf{P}_A$ and $\mathbf{P}_B$ to maximize C_{key}. Specifically, we consider the following

$$\begin{aligned} &\max_{\mathbf{P}_A,\mathbf{P}_B} C_{key} \\ &s.t.\ Tr(\mathbf{P}_A\mathbf{P}_A^H) \le TP_A,\ Tr(\mathbf{P}_B\mathbf{P}_B^H) \le TP_B \\ &\quad \mathbf{P}_A\mathbf{P}_A^H \succ \mathbf{0},\ \mathbf{P}_B\mathbf{P}_B^H \succ \mathbf{0} \end{aligned} \tag{2}$$

where P_A and P_B are the averaged powers used by Alice and Bob respectively,and the strict positive-definite conditions are used here due to Proposition 1 and its application later. We will see that the resulting optimal $\mathbf{P}_A$ and $\mathbf{P}_B$ with $T \ge \max\{N_A, N_B\}$ are such that they share a common row subspace of the dimension $\min\{N_A, N_B\}$ which is ideal for ANECE [10].

III. Analysis and Maximization of C_{key}

We can replace $\mathbf{Y}_A$ and $\mathbf{Y}_B$ by $\mathbf{y}_A = vec(\mathbf{Y}_A)$ and $\mathbf{y}_B = vec(\mathbf{Y}_B^T)$. Then, using $vec(\mathbf{XYZ}) = (\mathbf{Z}^T \otimes \mathbf{X})vec(\mathbf{Y})$, it follows that

$$\begin{aligned} \mathbf{y}_A &= \mathbf{G}_B\mathbf{h} + \mathbf{n}_A \\ \mathbf{y}_B &= \mathbf{G}_A\mathbf{h} + \mathbf{n}_B \end{aligned} \tag{3}$$

where $\mathbf{h} = vec(\mathbf{H})$, $\mathbf{n}_A = vec(\mathbf{N}_A)$, $\mathbf{n}_B = vec(\mathbf{N}_B)$, $\mathbf{G}_B = (\mathbf{P}_B^T\mathbf{R}_B^{\frac{1}{2}} \otimes \mathbf{R}_A^{\frac{1}{2}})$, and $\mathbf{G}_A = (\mathbf{R}_B^{\frac{1}{2}} \otimes \mathbf{P}_A^T\mathbf{R}_A^{\frac{1}{2}})$. It is obvious that $C_{key} = I(\mathbf{y}_A;\mathbf{y}_B)$. Furthermore, we will show that $C_{key} = I(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B)$ where $\hat{\mathbf{h}}_A$ and $\hat{\mathbf{h}}_B$ are the MMSE estimates of $\mathbf{h}$ by Alice and Bob respectively.

Let $\mathbf{K}_{\mathbf{x},\mathbf{y}} = \mathcal{E}\{\mathbf{x}\mathbf{y}^H\}$ for any two random vectors $\mathbf{x}$ and $\mathbf{y}$, and $\mathbf{K}_{\mathbf{x}} = \mathbf{K}_{\mathbf{x},\mathbf{x}}$. It follows that

$$\begin{aligned} \hat{\mathbf{h}}_A &= \mathbf{K}_{\mathbf{h},\mathbf{y}_A}\mathbf{K}_{\mathbf{y}_A}^{-1}\mathbf{y}_A \\ &= \sigma^2\mathbf{G}_B^H(\sigma^2\mathbf{G}_B\mathbf{G}_B^H + \mathbf{I})^{-1}(\mathbf{G}_B\mathbf{h} + \mathbf{n}_A) \end{aligned} \tag{4}$$

$$\begin{aligned} \hat{\mathbf{h}}_B &= \mathbf{K}_{\mathbf{h},\mathbf{y}_B}\mathbf{K}_{\mathbf{y}_B}^{-1}\mathbf{y}_B \\ &= \sigma^2\mathbf{G}_A^H(\sigma^2\mathbf{G}_A\mathbf{G}_A^H + \mathbf{I})^{-1}(\mathbf{G}_A\mathbf{h} + \mathbf{n}_B) \end{aligned} \tag{5}$$

$$\begin{aligned} I(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B) &= h(\hat{\mathbf{h}}_A) + h(\hat{\mathbf{h}}_B) - h(\hat{\mathbf{h}}_A,\hat{\mathbf{h}}_B) \\ &= \log_2|\mathbf{K}_{\hat{\mathbf{h}}_A}| + \log_2|\mathbf{K}_{\hat{\mathbf{h}}_B}| - \log_2|\mathbf{K}_{\{\hat{\mathbf{h}}_A,\hat{\mathbf{h}}_B\}}| \\ &= -\log_2|\mathbf{I} - \mathbf{K}_{\hat{\mathbf{h}}_B}^{-1}\mathbf{K}_{\hat{\mathbf{h}}_B,\hat{\mathbf{h}}_A}\mathbf{K}_{\hat{\mathbf{h}}_A}^{-1}\mathbf{K}_{\hat{\mathbf{h}}_A,\hat{\mathbf{h}}_B}| \qquad (6a) \\ &= -\log_2|\mathbf{I} - \sigma^{-4}\mathbf{K}_{\hat{\mathbf{h}}_A}\mathbf{K}_{\hat{\mathbf{h}}_B}| \qquad (6b) \end{aligned}$$

with

$$\mathbf{K}_{\{\hat{\mathbf{h}}_A,\hat{\mathbf{h}}_B\}} = \begin{bmatrix} \mathbf{K}_{\hat{\mathbf{h}}_A} & \mathbf{K}_{\hat{\mathbf{h}}_A,\hat{\mathbf{h}}_B} \\ \mathbf{K}_{\hat{\mathbf{h}}_B,\hat{\mathbf{h}}_A} & \mathbf{K}_{\hat{\mathbf{h}}_B} \end{bmatrix} \tag{7}$$

and $\mathbf{K}_{\hat{\mathbf{h}}_A} = \sigma^4\mathbf{G}_B^H(\sigma^2\mathbf{G}_B\mathbf{G}_B^H + \mathbf{I})^{-1}\mathbf{G}_B$, $\mathbf{K}_{\hat{\mathbf{h}}_B} = \sigma^4\mathbf{G}_A^H(\sigma^2\mathbf{G}_A\mathbf{G}_A^H + \mathbf{I})^{-1}\mathbf{G}_A$, $\mathbf{K}_{\hat{\mathbf{h}}_A,\hat{\mathbf{h}}_B} = \sigma^6\mathbf{G}_B^H(\sigma^2\mathbf{G}_B\mathbf{G}_B^H + \mathbf{I})^{-1}\mathbf{G}_B\mathbf{G}_A^H(\sigma^2\mathbf{G}_A\mathbf{G}_A^H + \mathbf{I})^{-1}\mathbf{G}_A$. Note that we have applied that $\mathbf{K}_{\hat{\mathbf{h}}_A}$ and $\mathbf{K}_{\hat{\mathbf{h}}_B}$ are both invertible. And (6a) is based on the fact that $\left|\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^H & \mathbf{C} \end{bmatrix}\right| = |\mathbf{A}||\mathbf{C} - \mathbf{B}^H\mathbf{A}^{-1}\mathbf{B}| = |\mathbf{C}||\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^H|$ where $\mathbf{A}$ and $\mathbf{C}$ are invertible, and (6b) is based on $\mathbf{K}_{\hat{\mathbf{h}}_A,\hat{\mathbf{h}}_B} = \sigma^{-2}\mathbf{K}_{\hat{\mathbf{h}}_A}\mathbf{K}_{\hat{\mathbf{h}}_B}$. The following is a generalization of a result in the SISO case shown in [3] and complements a fact that C_{key} also equals to the mutual information between the MLEs of $\mathbf{h}$ by Alice and Bob [4].

***Proposition** 1:* If $\mathbf{R}_A$, $\mathbf{R}_B$, $\mathbf{P}_A^T$ and $\mathbf{P}_B^T$ all have full collum ranks (which requires $T \ge \max\{N_A, N_B\}$), then we have $I(\mathbf{y}_A;\mathbf{y}_B) = I(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B)$.

Proof: With the stated conditions, both $\mathbf{K}_{\hat{\mathbf{h}}_A}$ and $\mathbf{K}_{\hat{\mathbf{h}}_B}$ are invertible. Similar to (6a), we have

$$\begin{aligned} &I(\mathbf{y}_A;\mathbf{y}_B) \\ &= -\log_2|\mathbf{I} - \mathbf{K}_{\mathbf{y}_B}^{-1}\mathbf{K}_{\mathbf{y}_B,\mathbf{y}_A}\mathbf{K}_{\mathbf{y}_A}^{-1}\mathbf{K}_{\mathbf{y}_A,\mathbf{y}_B}| \\ &= -\log_2|\mathbf{I} - \sigma^4(\sigma^2\mathbf{G}_A\mathbf{G}_A^H + \mathbf{I})^{-1} \\ &\qquad \times \mathbf{G}_A\mathbf{G}_B^H(\sigma^2\mathbf{G}_B\mathbf{G}_B^H + \mathbf{I})^{-1}\mathbf{G}_B\mathbf{G}_A^H| \end{aligned} \tag{8}$$

which can be further shown to be $I(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B)$ based on (6b). Here the second equality is based on $\mathbf{K}_{\mathbf{y}_A} = \sigma^2\mathbf{G}_B\mathbf{G}_B^H + \mathbf{I}$, $\mathbf{K}_{\mathbf{y}_B} = \sigma^2\mathbf{G}_A\mathbf{G}_A^H + \mathbf{I}$ and $\mathbf{K}_{\mathbf{y}_A,\mathbf{y}_B} = \sigma^2\mathbf{G}_B\mathbf{G}_A^H$. ■

In order to further simplify (6b), we introduce the following singular value decompositions (SVD):

$$\begin{aligned} \mathbf{R}_A^{\frac{H}{2}}\mathbf{P}_A^* &= \mathbf{U}_A\mathbf{\Lambda}_A\mathbf{V}_A^H \\ \mathbf{R}_B^{\frac{H}{2}}\mathbf{P}_B^* &= \mathbf{U}_B\mathbf{\Lambda}_B\mathbf{V}_B^H \end{aligned} \tag{9}$$

$$\begin{aligned} \mathbf{R}_A &= \tilde{\mathbf{U}}_A\tilde{\mathbf{\Lambda}}_A\tilde{\mathbf{U}}_A^H \\ \mathbf{R}_B &= \tilde{\mathbf{U}}_B\tilde{\mathbf{\Lambda}}_B\tilde{\mathbf{U}}_B^H \end{aligned} \tag{10}$$

where $\mathbf{U}_A \in \mathbb{C}^{N_A \times N_A}$, $\mathbf{\Lambda}_A \in \mathbb{R}^{N_A \times T}$, $\mathbf{V}_A \in \mathbb{C}^{T \times T}$, $\mathbf{U}_B \in \mathbb{C}^{N_B \times N_B}$, $\mathbf{\Lambda}_B \in \mathbb{R}^{N_B \times T}$ and $\mathbf{V}_B \in \mathbb{C}^{T \times T}$ are the matrices to be optimized as they are functions of the pilots. With $T \ge \max\{N_A, N_B\}$, we denote the singular values in (9) as $\mathbf{\Lambda}_A = [diag\{\sqrt{\lambda_{a,1}}, \dots, \sqrt{\lambda_{a,N_A}}\}, \mathbf{0}_{N_A \times (T-N_A)}]$ and $\mathbf{\Lambda}_B = [diag\{\sqrt{\lambda_{b,1}}, \dots, \sqrt{\lambda_{b,N_B}}\}, \mathbf{0}_{N_B \times (T-N_B)}]$ while the non-zero singular values $\{\lambda_{a,i}\}$ and $\{\lambda_{b,j}\}$ are in descending order respectively. Also note that $\tilde{\mathbf{U}}_A \in \mathbb{C}^{N_A \times N_A}$, $\tilde{\mathbf{\Lambda}}_A \in \mathbb{R}^{N_A \times N_A}$, $\tilde{\mathbf{U}}_B^H \in \mathbb{C}^{N_B \times N_B}$, $\tilde{\mathbf{\Lambda}}_B \in \mathbb{R}^{N_B \times N_B}$ are known matrices where $\tilde{\mathbf{\Lambda}}_A = diag\{\tilde{\lambda}_{a,1}, \dots, \tilde{\lambda}_{a,N_A}\}$ and $\tilde{\mathbf{\Lambda}}_B = diag\{\tilde{\lambda}_{b,1}, \dots, \tilde{\lambda}_{b,N_B}\}$ with $\sum_i \tilde{\lambda}_{a,i} = N_A$ and $\sum_j \tilde{\lambda}_{b,j} = N_B$. The elements $\{\tilde{\lambda}_{a,i}\}$ and $\{\tilde{\lambda}_{b,i}\}$ are in descending order respectively. From (10), we have $\mathbf{R}_A^{\frac{1}{2}} = \tilde{\mathbf{U}}_A\tilde{\mathbf{\Lambda}}_A^{\frac{1}{2}}$ and $\mathbf{R}_B^{\frac{1}{2}} = \tilde{\mathbf{U}}_B\tilde{\mathbf{\Lambda}}_B^{\frac{1}{2}}$, and based on (9) and (10), we have

$$\begin{aligned}\mathbf{P}_A &= (\mathbf{R}_A^{-\frac{H}{2}}\mathbf{U}_A\mathbf{\Lambda}_A\mathbf{V}_A^H)^* \\ \mathbf{P}_B &= (\mathbf{R}_B^{-\frac{H}{2}}\mathbf{U}_B\mathbf{\Lambda}_B\mathbf{V}_B^H)^*\end{aligned} \tag{11}$$

In the following, we will recall $\mathbf{G}_B = (\mathbf{P}_B^T\mathbf{R}_B^{\frac{1}{2}} \otimes \mathbf{R}_A^{\frac{1}{2}})$ and $\mathbf{G}_A = (\mathbf{R}_B^{\frac{1}{2}} \otimes \mathbf{P}_A^T\mathbf{R}_A^{\frac{1}{2}})$, apply the property $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC}\otimes\mathbf{BD})$, and use the definitions of $\mathbf{\Lambda}_A^2 = \mathbf{\Lambda}_A\mathbf{\Lambda}_A^T$ and $\mathbf{\Lambda}_B^2 = \mathbf{\Lambda}_B\mathbf{\Lambda}_B^T$. It follows from (9) and (10) that

$$\begin{aligned}\mathbf{K}_{\hat{\mathbf{h}}_B} &= \sigma^4\mathbf{G}_A^H(\sigma^2\mathbf{G}_A\mathbf{G}_A^H+\mathbf{I})^{-1}\mathbf{G}_A \\ &= \sigma^4(\tilde{\mathbf{\Lambda}}_B^{\frac{1}{2}}\tilde{\mathbf{U}}_B^H \otimes \mathbf{U}_A\mathbf{\Lambda}_A\mathbf{V}_A^H)(\sigma^2\tilde{\mathbf{U}}_B\tilde{\mathbf{\Lambda}}_B\tilde{\mathbf{U}}_B^H \\ &\quad \otimes \mathbf{V}_A\mathbf{\Lambda}_A^T\mathbf{\Lambda}_A\mathbf{V}_A^H+\mathbf{I})^{-1}(\tilde{\mathbf{U}}_B\tilde{\mathbf{\Lambda}}_B^{\frac{1}{2}} \otimes \mathbf{V}_A\mathbf{\Lambda}_A^T\mathbf{U}_A^H) \\ &= \sigma^4(\sigma^2\mathbf{I}+(\tilde{\mathbf{\Lambda}}_B\otimes\mathbf{U}_A\mathbf{\Lambda}_A^2\mathbf{U}_A^H)^{-1})^{-1}\end{aligned} \tag{12}$$

$$\begin{aligned}\mathbf{K}_{\hat{\mathbf{h}}_A} &= \sigma^4\mathbf{G}_B^H(\sigma^2\mathbf{G}_B\mathbf{G}_B^H+\mathbf{I})^{-1}\mathbf{G}_B \\ &= \sigma^4(\mathbf{U}_B\mathbf{\Lambda}_B\mathbf{V}_B^H \otimes \tilde{\mathbf{\Lambda}}_A^{\frac{1}{2}}\tilde{\mathbf{U}}_A^H)(\sigma^2\mathbf{V}_B\mathbf{\Lambda}_B^T\mathbf{\Lambda}_B\mathbf{V}_B^H \\ &\quad \otimes \tilde{\mathbf{U}}_A\tilde{\mathbf{\Lambda}}_A\tilde{\mathbf{U}}_A^H+\mathbf{I})^{-1}(\mathbf{V}_B\mathbf{\Lambda}_B^T\mathbf{U}_B^H \otimes \tilde{\mathbf{U}}_A\tilde{\mathbf{\Lambda}}_A^{\frac{1}{2}}) \\ &= \sigma^4(\sigma^2\mathbf{I}+(\mathbf{U}_B\mathbf{\Lambda}_B^2\mathbf{U}_B^H\otimes\tilde{\mathbf{\Lambda}}_A)^{-1})^{-1}\end{aligned} \tag{13}$$

Plugging (12) and (13) into (6b), we have

$$\begin{aligned}&I(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B) \\ &= -\log_2\left|\mathbf{I}-\sigma^4(\sigma^2\mathbf{I}+(\tilde{\mathbf{\Lambda}}_B\otimes\mathbf{U}_A\mathbf{\Lambda}_A^2\mathbf{U}_A^H)^{-1})^{-1}\right. \\ &\qquad \left.\times(\sigma^2\mathbf{I}+(\mathbf{U}_B\mathbf{\Lambda}_B^2\mathbf{U}_B^H\otimes\tilde{\mathbf{\Lambda}}_A)^{-1})^{-1}\right|\end{aligned} \tag{14}$$

Furthermore, (14) can be reorganized into the following form:

$$\begin{aligned}&I(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B) \\ &= \log_2\left|\sigma^2\mathbf{I}+(\tilde{\mathbf{\Lambda}}_B\otimes\mathbf{U}_A\mathbf{\Lambda}_A^2\mathbf{U}_A^H)^{-1}\right| \\ &\quad + \log_2\left|\sigma^2\mathbf{I}+(\mathbf{U}_B\mathbf{\Lambda}_B^2\mathbf{U}_B^H\otimes\tilde{\mathbf{\Lambda}}_A)^{-1}\right| \\ &\quad - \log_2\left|(\sigma^2\mathbf{I}+(\tilde{\mathbf{\Lambda}}_B\otimes\mathbf{U}_A\mathbf{\Lambda}_A^2\mathbf{U}_A^H)^{-1})\right. \\ &\qquad \left.\times(\sigma^2\mathbf{I}+(\mathbf{U}_B\mathbf{\Lambda}_B^2\mathbf{U}_B^H\otimes\tilde{\mathbf{\Lambda}}_A)^{-1})-\sigma^4\mathbf{I}\right| \qquad (15a) \\ &= \log_2|\mathbf{I}+\sigma^2\tilde{\mathbf{\Lambda}}_B\otimes\mathbf{\Lambda}_A^2| + \log_2|\mathbf{I}+\sigma^2\mathbf{\Lambda}_B^2\otimes\tilde{\mathbf{\Lambda}}_A| \\ &\quad - \log_2|\mathbf{I}+\sigma^2\tilde{\mathbf{\Lambda}}_B\otimes\mathbf{\Lambda}_A^2+\sigma^2\mathbf{U}(\mathbf{\Lambda}_B^2\otimes\tilde{\mathbf{\Lambda}}_A)\mathbf{U}^H| \qquad (15b)\end{aligned}$$

where $\mathbf{U} \triangleq \mathbf{U}_B\otimes\mathbf{U}_A^H$. Here, (15a) is due to $-\log_2|\mathbf{I}-\mathbf{A}^{-1}\mathbf{B}^{-1}| = \log_2|\mathbf{A}|+\log_2|\mathbf{B}|-\log_2|\mathbf{AB}-\mathbf{I}|$, and (15b) is due to $\log_2|\mathbf{I}+\mathbf{A}^{-1}| = \log_2|\mathbf{I}+\mathbf{A}|-\log_2|\mathbf{A}|$ when $\mathbf{A}$ and $\mathbf{B}$ are invertible. From (15), the optimization of $\mathbf{U}_A$ and $\mathbf{U}_B$ can be formulated as

$$\begin{aligned}&\{\mathbf{U}_{A,opt},\mathbf{U}_{B,opt}\} \\ &= arg\min_{\mathbf{U}_A,\mathbf{U}_B}\log_2|\mathbf{I}+\sigma^2\tilde{\mathbf{\Lambda}}_B\otimes\mathbf{\Lambda}_A^2+\sigma^2\mathbf{U}(\mathbf{\Lambda}_B^2\otimes\tilde{\mathbf{\Lambda}}_A)\mathbf{U}^H|\end{aligned} \tag{16}$$

According to [12], we have:

***Lemma** 1:* Given Hermitian matrices $\mathbf{A},\mathbf{C}\in\mathbb{C}^{n\times n}$ and $\mathbf{B},\mathbf{D}\in\mathbb{C}^{m\times m}$ with the corresponding diagonal eigenvalue matrices $\mathbf{\Lambda}_a$, $\mathbf{\Lambda}_c$, $\mathbf{\Lambda}_b$, $\mathbf{\Lambda}_d$. And the diagonal elements in each diagonal matrix are in descending order. Then

$$|\mathbf{A}\otimes\mathbf{B}+\mathbf{C}\otimes\mathbf{D}| \geq \min_{P_1,P_2}|\mathbf{\Lambda}_a\otimes\mathbf{\Lambda}_b+\mathbf{\Lambda}_{c,P_1}\otimes\mathbf{\Lambda}_{d,P_2}| \tag{17}$$

$$|\mathbf{A}\otimes\mathbf{B}+\mathbf{C}\otimes\mathbf{D}| \leq \max_{P_1,P_2}|\mathbf{\Lambda}_a\otimes\mathbf{\Lambda}_b+\mathbf{\Lambda}_{c,P_1}\otimes\mathbf{\Lambda}_{d,P_2}| \tag{18}$$

where the minimum or maximum are taken over all the permutations $\{P_1,P_2\}$.

From Lemma 1, we have:

Corollary 1: If $\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{D}$ are positive semi-definite Hermitian matrices, then

$$|\mathbf{A}\otimes\mathbf{B}+\mathbf{C}\otimes\mathbf{D}| \geq |\mathbf{\Lambda}_a\otimes\mathbf{\Lambda}_b+\mathbf{\Lambda}_c\otimes\mathbf{\Lambda}_d| \tag{19a}$$

$$|\mathbf{A}\otimes\mathbf{B}+\mathbf{C}\otimes\mathbf{D}| \leq |\mathbf{\Lambda}_a\otimes\mathbf{\Lambda}_b+\bar{\mathbf{\Lambda}}_c\otimes\bar{\mathbf{\Lambda}}_d| \tag{19b}$$

where elements in $\bar{\mathbf{\Lambda}}_c$ and $\bar{\mathbf{\Lambda}}_d$ are in ascending order.

Proof: Denote $\lambda_{c,s}$ and $\lambda_{c,l}$ as two elements in $\mathbf{\Lambda}_{c,P_1}$ where $s<l$. Define two permutations $\lambda_{c,p_s}\geq\lambda_{c,p_l}$ and $\lambda_{c,p_s'}\leq\lambda_{c,p_l'}$ where $\lambda_{c,p_s}=\lambda_{c,p_l'}$, $\lambda_{c,p_l}=\lambda_{c,p_s'}$. From (17) we have

$$|\mathbf{A}\otimes\mathbf{B}+\mathbf{C}\otimes\mathbf{D}| \geq \min_{P_1,P_2}\prod_{j=1}^{m}\prod_{i=1}^{n}(\lambda_{a,i}\lambda_{b,j}+\lambda_{c,P_{1,i}}\lambda_{d,P_{2,j}}) \tag{20}$$

For any given j we have

$$\begin{aligned}&(\lambda_{a,s}\lambda_{b,j}+\lambda_{c,p_s}\lambda_{d,P_{2,j}})(\lambda_{a,l}\lambda_{b,j}+\lambda_{c,p_l}\lambda_{d,P_{2,j}}) \\ &\quad -(\lambda_{a,s}\lambda_{b,j}+\lambda_{c,p_s'}\lambda_{d,P_{2,j}})(\lambda_{a,l}\lambda_{b,j}+\lambda_{c,p_l'}\lambda_{d,P_{2,j}}) \\ &= \lambda_{a,s}\lambda_{b,j}\lambda_{c,p_l}\lambda_{d,P_{2,j}}+\lambda_{c,p_s}\lambda_{d,P_{2,j}}\lambda_{a,l}\lambda_{b,j} \\ &\quad -\lambda_{a,s}\lambda_{b,j}\lambda_{c,p_l'}\lambda_{d,P_{2,j}}-\lambda_{c,p_s'}\lambda_{d,P_{2,j}}\lambda_{a,l}\lambda_{b,j} \\ &= \lambda_{d,P_{2,j}}\lambda_{b,j}(\lambda_{a,s}-\lambda_{a,l})(\lambda_{c,p_l}-\lambda_{c,p_s})\leq 0\end{aligned} \tag{21}$$

Therefore, as elements in $\mathbf{\Lambda}_a$ are in descending order, a descending $\mathbf{\Lambda}_{c,P_1}$ will minimize the right hand side of (20). Similarly, as elements in $\mathbf{\Lambda}_b$ are in descending order, a descending $\mathbf{\Lambda}_{d,P_2}$ will also minimize the right hand side of (20). The proof of (19a) is done, and (19b) can be proved in a similar manner. ■

Applying Corollary 1 to (16) yields $\mathbf{U}_{A,opt}=\mathbf{I}$ and $\mathbf{U}_{B,opt}=\mathbf{I}$, which are optimal for the objective function in (2).

Let $\mathbf{\Lambda}_A^2 = diag\{\lambda_{a,1},\ldots,\lambda_{a,N_A}\}$, $\mathbf{\Lambda}_B^2 = diag\{\lambda_{b,1},\ldots,\lambda_{b,N_B}\}$, $\tilde{\mathbf{\Lambda}}_A = diag\{\tilde{\lambda}_{a,1},\ldots,\tilde{\lambda}_{a,N_A}\}$ and $\tilde{\mathbf{\Lambda}}_B = diag\{\tilde{\lambda}_{b,1},\ldots,\tilde{\lambda}_{b,N_B}\}$. Also let $\mathbf{C}_A = \tilde{\mathbf{\Lambda}}_A^{-1}\mathbf{\Lambda}_A^2$ and $\mathbf{C}_B = \tilde{\mathbf{\Lambda}}_B^{-1}\mathbf{\Lambda}_B^2$ with their elements denoted by $c_{a,i}=\lambda_{a,i}/\tilde{\lambda}_{a,i}$ and $c_{b,j}=\lambda_{b,j}/\tilde{\lambda}_{b,j}$, which remain to be optimized. With $\mathbf{U}_A=\mathbf{I}$ and $\mathbf{U}_B=\mathbf{I}$, (15b) becomes

$$\begin{aligned}&I(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B) \\ &= \log_2|\mathbf{I}+\sigma^2\tilde{\mathbf{\Lambda}}_B\otimes\mathbf{C}_A\tilde{\mathbf{\Lambda}}_A| + \log_2|\mathbf{I}+\sigma^2\mathbf{C}_B\tilde{\mathbf{\Lambda}}_B\otimes\tilde{\mathbf{\Lambda}}_A| \\ &\quad - \log_2|\mathbf{I}+\sigma^2\tilde{\mathbf{\Lambda}}_B\otimes\mathbf{C}_A\tilde{\mathbf{\Lambda}}_A+\sigma^2\mathbf{C}_B\tilde{\mathbf{\Lambda}}_B\otimes\tilde{\mathbf{\Lambda}}_A| \qquad (22a) \\ &= \sum_{j=1}^{N_B}\sum_{i=1}^{N_A}\log_2\left(\frac{(1+\sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}c_{a,i})(1+\sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}c_{b,j})}{1+\sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}c_{a,i}+\sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}c_{b,j}}\right) \\ &\triangleq \sum_{j=1}^{N_B}\sum_{i=1}^{N_A}f_{i,j}(c_{a,i},c_{b,j}) \qquad (22b)\end{aligned}$$

For the constraint in (2), we now use (10) and (11) to yield

$$\begin{aligned}Tr(\mathbf{P}_A\mathbf{P}_A^H) &= Tr(\tilde{\mathbf{\Lambda}}_A^{-1}\mathbf{U}_A\mathbf{\Lambda}_A^2\mathbf{U}_A^H) \\ &\geq Tr(\tilde{\mathbf{\Lambda}}_A^{-1}\mathbf{\Lambda}_A^2) = \sum_{i=1}^{N_A}\frac{\lambda_{a,i}}{\tilde{\lambda}_{a,i}} = \sum_{i=1}^{N_A}c_{a,i}\end{aligned} \tag{23}$$

Algorithm 1 Bisection section search

Input:

$\tilde{\boldsymbol{\lambda}}_a, \tilde{\boldsymbol{\lambda}}_b, P_A, P_B$,T;

Accuracy threshold ϵ_1, ϵ_2.

Initialization $k=0$, $\mathbf{c}_a^{(k)} = \frac{TP_A}{N_A}\mathbf{1}_{N_A}, \mathbf{c}_b^{(k)} = \frac{TP_B}{N_B}\mathbf{1}_{N_B}$.

repeat

Given $\mathbf{c}_b^{(k)}$, do bisection search of μ and obtain solution $\mathbf{c}_a^{(k+1)}$ to meet the power constraint $|\sum_{i=1}^{N_A} c_{a,i} - TP_A| \le \epsilon_1$; Given $\mathbf{c}_a^{(k+1)}$, do bisection search of ν and obtain solution $\mathbf{c}_b^{(k+1)}$ to meet the power constraint $|\sum_{j=1}^{N_B} c_{b,j} - TP_B| \le \epsilon_1$.

$k = k+1$.

until $\|[\mathbf{c}_a^{(k)}, \mathbf{c}_b^{(k)}] - [\mathbf{c}_a^{(k-1)}, \mathbf{c}_b^{(k-1)}]\| \le \epsilon_2$

return $\{\mathbf{c}_a^{(k)}, \mathbf{c}_b^{(k)}\}$

$$Tr(\mathbf{P}_B\mathbf{P}_B^H) \ge Tr(\tilde{\mathbf{\Lambda}}_B^{-1}\mathbf{\Lambda}_B^2) = \sum_{j=1}^{N_B} \frac{\lambda_{b,j}}{\tilde{\lambda}_{b,j}} = \sum_{j=1}^{N_B} c_{b,j} \tag{24}$$

where the equalities in the inequities hold when $\mathbf{U}_B = \mathbf{I}_{N_B}$ and $\mathbf{U}_A = \mathbf{I}_{N_A}$ [13]. Namely, $\mathbf{U}_B = \mathbf{I}_{N_B}$ and $\mathbf{U}_A = \mathbf{I}_{N_A}$ are also optimal for the constraint in (2).

It is obvious that the unitary matrices $\mathbf{V}_A$ and $\mathbf{V}_B$ do not affect neither the objective function nor the constraint in (2). Without lose of generality, we can set them to be the identity matrices. Also note that by choosing $\mathbf{V}_A = \mathbf{V}_B = \mathbf{I}_T$, we have ensured that there is a common row subspace between $\mathbf{P}_A$ and $\mathbf{P}_B$ of the dimension $\min\{N_A, N_B\}$ if $\mathbf{\Lambda}_A^2$ and $\mathbf{\Lambda}_B^2$ have the full ranks N_A and N_B respectively. It will also be shown that each of the optimal $\mathbf{\Lambda}_A^2$ and $\mathbf{\Lambda}_B^2$ is of full rank and always contains descending entries.

With the above results, we have now transformed (2) into

$$\max_{\mathbf{c}_a > \mathbf{0}, \mathbf{c}_b > \mathbf{0}} \quad \sum_{j=1}^{N_B}\sum_{i=1}^{N_A} f_{i,j}(c_{a,i}, c_{b,j}) \tag{25a}$$

$$s.t. \quad \sum_{i=1}^{N_A} c_{a,i} \le TP_A, \ \sum_{j=1}^{N_B} c_{b,j} \le TP_B \tag{25b}$$

It is easy to verify that $f(c_{a,i}, c_{b,j})$ is a monotonically increasing function of $c_{a,i}$ and $c_{b,j}$ respectively. So, the optimal solutions must satisfy $\sum_{i=1}^{N_A} c_{a,i} = TP_A$ and $\sum_{j=1}^{N_B} c_{b,j} = TP_B$.

However, $-f_{i,j}(c_{a,i}, c_{b,j})$ is not always convex of $c_{a,i}$ and $c_{b,j}$. The Hessian matrix of $-f_{i,j}(c_{a,i}, c_{b,j})$ is

$$\begin{bmatrix} \frac{\sigma^4\tilde{\lambda}_{a,i}^2\tilde{\lambda}_{b,j}^2(\varphi_{i,j} - \phi_{a,i,j})}{\phi_{a,i,j}\varphi_{i,j}} & -\frac{\sigma^4\tilde{\lambda}_{a,i}^2\tilde{\lambda}_{b,j}^2}{\varphi_{i,j}} \\ -\frac{\sigma^4\tilde{\lambda}_{a,i}^2\tilde{\lambda}_{b,j}^2}{\varphi_{i,j}} & \frac{\sigma^4\tilde{\lambda}_{a,i}^2\tilde{\lambda}_{b,j}^2(\varphi_{i,j} - \phi_{b,i,j})}{\phi_{b,i,j}\varphi_{i,j}} \end{bmatrix} \tag{26}$$

where $\phi_{a,i,j} = (1 + \sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}c_{a,i})^2$, $\phi_{b,i,j} = (1 + \sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}c_{b,j})^2$ and $\varphi_{i,j} = (1 + \sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}c_{a,i} + \sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}c_{b,j})^2$. This matrix is positive definite if and only if $c_{a,i}c_{b,j} \ge \frac{1}{2\sigma^4\tilde{\lambda}_{a,i}^2\tilde{\lambda}_{b,j}^2}$. This means that when TP_A and TP_B are large, the Hessian matrix of $-f_{i,j}(c_{a,i}, c_{b,j})$ is typically positive definite and hence $-f_{i,j}(c_{a,i}, c_{b,j})$ is typically convex. In this high power case, the problem (25) is convex and the globally optimal solution is available. In general, $f_{i,j}(c_{a,i}, c_{b,j})$ is a concave function with respect to $c_{a,i}$ and $c_{b,j}$ individually. To obtain locally optimal solution to (25), we can apply a two-phase iteration method, i.e., optimizing $\mathbf{c}_a$ and $\mathbf{c}_b$ alternately until convergence. The discussion of the following two-phase algorithm is similar to that in [9].

In phase one, the Lagrangian function with respect to $c_{a,i}$ is

$$\mathcal{L} = \sum_{j=1}^{N_B}\sum_{i=1}^{N_A} f_{i,j}(c_{a,i}, c_{b,j}) - \mu\left(\sum_{i=1}^{N_A} c_{a,i} - TP_A\right) + \boldsymbol{\alpha}^T\mathbf{c}_a \tag{27}$$

And the corresponding **KKT** conditions are

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial c_{a,i}} = \dfrac{1}{\ln 2}\displaystyle\sum_{j=1}^{N_B} g_{i,j}(c_{a,i}, c_{b,j}) - \mu = 0 \\ \displaystyle\sum_{i=1}^{N_A} c_{a,i} \le TP_A, \ \mu(\sum_{i=1}^{N_A} c_{a,i} - TP_A) = 0, \ \mu \ge 0 \\ \mathbf{c}_a > \mathbf{0}, \ \boldsymbol{\alpha}^T\mathbf{c}_a = 0, \ \boldsymbol{\alpha} \ge \mathbf{0} \end{cases} \tag{28}$$

where

$$g_{i,j}(x,y) = \frac{\sigma^4\tilde{\lambda}_{a,i}^2\tilde{\lambda}_{b,j}^2 y}{(1 + \sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}x)(1 + \sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}x + \sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}y)} \tag{29}$$

In phase two, similar **KKT** conditions can be found. From (28) we can see that μ is a monotonically decreasing function of $c_{a,i}$. Therefore, we can use a bisection search to solve (28). An efficient algorithm to solve (25) is shown in Algorithm 1.

From (29), we know that $g_{i,j}(c_{a,i}, c_{b,j})$ is an increasing function of $\tilde{\lambda}_{a,i}$ and a decreasing function of $c_{a,i}$. Given any $\mathbf{c}_b$, the solution from (28) is $\mathbf{c}_a^*$, which must satisfy $\sum_{j=1}^{N_B} g_{i,j}(c_{a,i}^*, c_{b,j}) = \mu \ln 2$. Hence, one can verify that $c_{a,i}^* \ge c_{a,i+1}^*$. (If $c_{a,i}^* < c_{a,i+1}^*$ then $\mu \ln 2 = \sum_{j=1}^{N_B} g_{i,j}(c_{a,i}^*, c_{b,j}) > \sum_{j=1}^{N_B} g_{i,j}(c_{a,i+1}^*, c_{b,j}) \ge \sum_{j=1}^{N_B} g_{i+1,j}(c_{a,i+1}^*, c_{b,j}) = \mu \ln 2$, which is not possible.) Similarly, $c_{b,i}^* \ge c_{b,i+1}^*$. Therefore, the diagonal elements of the optimal solutions of $\mathbf{\Lambda}_A^2$ and $\mathbf{\Lambda}_B^2$ are also in descending order respectively.

IV. Asymptotic Analysis

Assume $P_A = P_B = P$. Define $\check{c}_{a,i} = \frac{c_{a,i}}{TP}$ and $\check{c}_{b,i} = \frac{c_{b,i}}{TP}$. Then, the power constraints become $\sum_{i=1}^{N_A} \check{c}_{a,i} = 1$ and $\sum_{j=1}^{N_B} \check{c}_{b,j} = 1$. And (22) now becomes

$$I(\hat{\mathbf{h}}_A; \hat{\mathbf{h}}_B) = \sum_{j=1}^{N_B}\sum_{i=1}^{N_A} \log_2\left(\frac{(1 + TP\sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}\check{c}_{a,i})(1 + TP\sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}\check{c}_{b,j})}{1 + TP\sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}\check{c}_{a,i} + TP\sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}\check{c}_{b,j}}\right) \tag{30}$$

A. High Power Case

For large P, (30) can be approximated as

$$\begin{aligned} &I(\hat{\mathbf{h}}_A; \hat{\mathbf{h}}_B) \\ &\approx \sum_{j=1}^{N_B}\sum_{i=1}^{N_A} \log_2\left(\frac{\check{c}_{a,i}\check{c}_{b,j}}{\check{c}_{a,i} + \check{c}_{b,j}}\right) + \sum_{j=1}^{N_B}\sum_{i=1}^{N_A} \log_2(TP\sigma^2\tilde{\lambda}_{a,i}\tilde{\lambda}_{b,j}) \\ &\triangleq \phi_1(\check{\mathbf{c}}_a, \check{\mathbf{c}}_b, \tilde{\boldsymbol{\lambda}}_a, \tilde{\boldsymbol{\lambda}}_b) \end{aligned} \tag{31}$$

From (31), we know that the degree of freedom per channel realization is $\lim_{P\to\infty} \frac{\phi_1(\check{\mathbf{c}}_a, \check{\mathbf{c}}_b, \tilde{\boldsymbol{\lambda}}_a, \tilde{\boldsymbol{\lambda}}_b)}{\log_2 P} = N_A N_B$.

Also, $-\frac{\partial^2\phi_1}{\partial\check{c}_{a,i}^2} = -\sum_j\left(\frac{1}{(\check{c}_{a,i}+\check{c}_{b,j})^2}-\frac{1}{\check{c}_{a,i}^2}\right)\geq 0$, which means that $-\phi_1$ is a convex function of $\check{\mathbf{c}}_a$. Meanwhile, $-\phi_1$ is a symmetric function of $\check{\mathbf{c}}_a$. Therefore, ϕ_1 is a Schur-concave function [13] of $\check{\mathbf{c}}_a$, and then we have $\phi_1(\mathbf{1}_{N_A},\check{\mathbf{c}}_b,\tilde{\boldsymbol{\lambda}}_a,\tilde{\boldsymbol{\lambda}}_b)\geq\phi_1(\check{\mathbf{c}}_a,\check{\mathbf{c}}_b,\tilde{\boldsymbol{\lambda}}_a,\tilde{\boldsymbol{\lambda}}_b)$ with any descending $\check{\mathbf{c}}_a$. Similar idea can be applied to show that (31) is also a Schur-concave function of $\check{\mathbf{c}}_b$. Therefore, the optimal power allocation in the high power case is such that $\check{\mathbf{c}}_a=\frac{1}{N_A}\mathbf{1}_{N_A}$ and $\check{\mathbf{c}}_b=\frac{1}{N_B}\mathbf{1}_{N_B}$.

Also, by applying the same argument, one can easily prove that (31) is also a Schur-concave function of $\tilde{\boldsymbol{\lambda}}_a$ and $\tilde{\boldsymbol{\lambda}}_b$ respectively. Therefore, when $\tilde{\boldsymbol{\lambda}}_a=\mathbf{1}_{N_A}$ and $\tilde{\boldsymbol{\lambda}}_b=\mathbf{1}_{N_B}$, (31) is maximized. In other words, in the high power case, less correlated channel yields a higher secret key rate.

B. Low Power Case

For small P, we can approximate (30) by its second-order Taylor series expansion at point $P=0$:

$$I(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B)=I(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B)|_{P=0}+\dot{I}(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B)|_{P=0}P + \frac{1}{2}\ddot{I}(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B)|_{P=0}P^2+o(P^2) \quad (32)$$

where $\dot{I}(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B)$ and $\ddot{I}(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B)$ are the first and second order derivatives of (30) regarding to P. It can be easily proved that $\dot{I}(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B)|_{P=0}=0$ while $\ddot{I}(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B)|_{P=0}$ can be expressed as

$$\begin{aligned}&\ddot{I}(\hat{\mathbf{h}}_A;\hat{\mathbf{h}}_B)|_{P=0}\\&=\frac{2}{\ln 2}\sum_i^{N_A}\sum_j^{N_B}\sigma^4\tilde{\lambda}_{a,i}^2\tilde{\lambda}_{b,j}^2T^2\check{c}_{a,i}\check{c}_{b,j}\triangleq\phi_2(\check{\mathbf{c}}_a,\check{\mathbf{c}}_b,\tilde{\boldsymbol{\lambda}}_a,\tilde{\boldsymbol{\lambda}}_b)\end{aligned} \quad (33)$$

To maximize (32), we just need to maximize the term (33). Based on (33) we have $\frac{\partial\phi_2}{\partial\check{c}_{a,i}}=\sigma^4T^2\tilde{\lambda}_{a,i}^2\sum_{j=1}^{N_B}\tilde{\lambda}_{b,j}^2\check{c}_{b,j}$. Since $\{\tilde{\lambda}_{a,i}\}$ is in descending order, we know that $\phi_2(\check{\mathbf{c}}_a,\check{\mathbf{c}}_b,\tilde{\boldsymbol{\lambda}}_a,\tilde{\boldsymbol{\lambda}}_b)$ is a Schur-convex function of $\check{\mathbf{c}}_a$ with descending entries, which means it is maximized by putting almost all of the power to $\check{c}_{a,1}$. The reason that "almost all" instead of "all" is used here is to ensure the positive condition on $\mathbf{c}_a$. Same conclusion can be drawn to $\check{c}_{b,1}$ for maximizing $\phi_2(\check{\mathbf{c}}_a,\check{\mathbf{c}}_b,\tilde{\boldsymbol{\lambda}}_a,\tilde{\boldsymbol{\lambda}}_b)$. That is, in the low power case, almost all of the power should be allocated to the strongest stream.

Also, $\phi_2(\check{\mathbf{c}}_a,\check{\mathbf{c}}_b,\tilde{\boldsymbol{\lambda}}_a,\tilde{\boldsymbol{\lambda}}_b)$ is a Schur-convex function of $\tilde{\boldsymbol{\lambda}}_a$ and $\tilde{\boldsymbol{\lambda}}_b$ is a Schur-convex function of $\tilde{\boldsymbol{\lambda}}_b$. Therefore, in low power region, a higher channel correlation leads to a higher secret key rate.

V. Simulation Results

In this section, we provide a numerical comparison of the secret key rates based on three choices of the pilots: (1) I_{MSKR} - maximum secret key rate (MSKR) from (25); (2) I_{MCEE} - secret key rate based on minimum channel estimation error (MCEE); and (3) I_U - secret key rate based on uniform power allocation $\mathbf{C}_i=\frac{P_iT}{N_i}\mathbf{I}$, $i=\{A,B\}$. Define the channel correlation matrix as $[\mathbf{R}]_{i,j}=r^{|i-j|}$ where $r\in[0,1]$ is the correlation coefficient. We assume that Alice and Bob have the same channel correlation $r_A=r_B=r$, the same antenna numbers $N_A=N_B=8$, and the channel variance $\sigma^2=1$. We also let $P_AT=P_BT=P_T$ and $T\geq\max\{N_A,N_B\}=8$. In Fig. 2, we show the secret key rate (SKR) of I_{MSKR} (in bits per realization of $\mathbf{H}$) with two different correlations $r=0.4$ and $r=0.8$. As expected from the previous analysis, in the low power region, a higher correlation yields a higher secret key rate, but in the high power region, the opposite is true. In Fig. 3, we show the SKR ratios of $\frac{I_{MCEE}}{I_{MSKR}}$ and $\frac{I_U}{I_{MSKR}}$ with $r=0.8$. As the power increases, the uniform power pilots become closer to the optimal. We also see that the pilots based on MCEE are nearly optimal in the low power case. This is because the pilot design based on MCEE with channel correlation also allocates all the power to the strongest stream in the low power case. But the pilot design based on MCEE does not lead to uniform power allocation in high power case [14]. A brief discussion of MCEE is shown in appendix A.

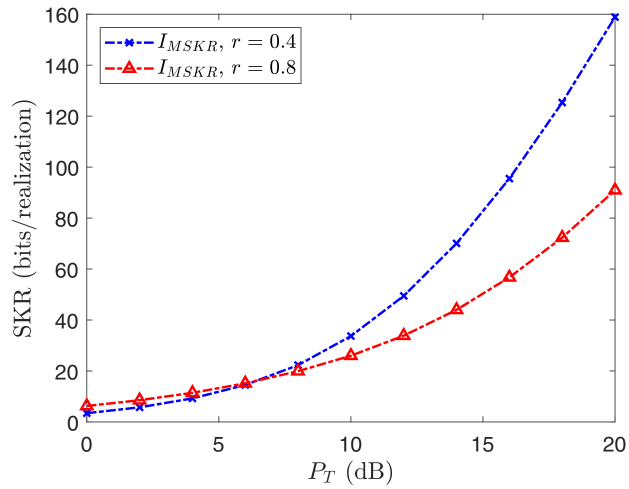

Fig. 2. Secret key rate with two different correlations $r=0.4$ and $r=0.8$

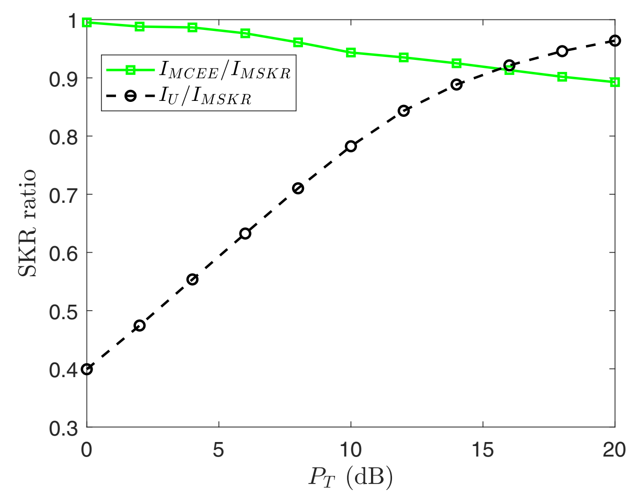

Fig. 3. Secret key rate ratio with correlation $r=0.8$

VI. Conclusion

We have developed an algorithm to compute the optimal pilots that maximize the capacity of secret key generation, shown that our algorithm yields the globally optimal solution for high and low power cases, compared the capacity performance of the optimal pilots with that of pilots based on minimum channel estimation errors or uniform power distribution, and shown that the optimal pilots designed here also meet the requirement for anti-eavesdropping channel estimation (ANECE) [10].

APPENDIX

A. Optimal Pilot for Minimum Channel Estimation Error

Regarding to the MMSE channel estimation by Alice (4), the power of the estimation error is

$$\begin{aligned}
J &= Tr\big(\mathcal{E}\{(\mathbf{h}-\hat{\mathbf{h}}_A)(\mathbf{h}-\hat{\mathbf{h}}_A)^H\}\big) \\
&= Tr\big(\mathbf{K}_\mathbf{h}-\mathbf{K}_{\mathbf{h},\mathbf{y}_A}\mathbf{K}_{\mathbf{y}_A}^{-1}\mathbf{K}_{\mathbf{y}_A,\mathbf{h}}\big) \\
&= Tr\big(\sigma^2\mathbf{I}-\sigma^4\mathbf{G}_B(\sigma^2\mathbf{G}_B\mathbf{G}_B^H+\mathbf{I})^{-1}\mathbf{G}_B^H\big) \\
&= Tr\big(\sigma^2(\mathbf{I}+\sigma^2\mathbf{G}_B^H\mathbf{G}_B)^{-1}\big) &\text{(34a)} \\
&= Tr\big(\sigma^2(\mathbf{I}+\sigma^2\mathbf{U}_B\mathbf{\Lambda}_B^2\mathbf{U}_B^H\otimes\tilde{\mathbf{\Lambda}}_A)^{-1}\big) &\text{(34b)} \\
&= \sum_{j=1}^{N_B}\sum_{i=1}^{N_A}\frac{\sigma^2}{1+\sigma^2\check{c}_{b,j}TP_B\tilde{\lambda}_{b,j}\tilde{\lambda}_{a,i}} &\text{(34c)}
\end{aligned}$$

where (34a) is based on matrix inverse lemma, (34b) is from using the SVD in (9) and (10), and (34c) is from the previous definition $\check{\mathbf{c}}_b=\mathbf{c}_b/(TP_B)=\boldsymbol{\lambda}_b\tilde{\boldsymbol{\lambda}}_b^{-1}/(TP_B)$. Since (34) is invariant to $\mathbf{U}_B$ and $\mathbf{V}_B$, then (24) implies that the optimal $\mathbf{U}_B$ and $\mathbf{V}_B$ are the identity matrices. Then the optimization problem with respect to $\check{\mathbf{c}}_b$ becomes

$$\min_{\check{\mathbf{c}}_b}\ \sum_{j=1}^{N_B}\sum_{i=1}^{N_A}\frac{\sigma^2}{1+\sigma^2\check{c}_{b,j}TP_B\tilde{\lambda}_{b,j}\tilde{\lambda}_{a,i}},\quad s.t.\sum_{j=1}^{N_B}\check{c}_{b,j}\le 1 \tag{35}$$

The corresponding Lagrangian function is

$$\mathcal{L}=\sum_{j=1}^{N_B}\sum_{i=1}^{N_A}\frac{\sigma^2}{1+\sigma^2\check{c}_{b,j}TP_B\tilde{\lambda}_{b,j}\tilde{\lambda}_{a,i}}+\mu\Big(\sum_{j=1}^{N_B}\check{c}_{b,j}-1\Big) \tag{36}$$

The **KKT** conditions are

$$\begin{cases}
\dfrac{\partial\mathcal{L}}{\partial\check{c}_{b,j}}=\displaystyle\sum_{i=1}^{N_A}\frac{\sigma^4TP_B\tilde{\lambda}_{b,j}\tilde{\lambda}_{a,i}}{(1+\sigma^2\check{c}_{b,j}TP_B\tilde{\lambda}_{b,j}\tilde{\lambda}_{a,i})^2}-\mu=0 \\
\displaystyle\sum_{j=1}^{N_B}\check{c}_{b,j}\le 1,\ \mu\Big(\sum_{j=1}^{N_B}\check{c}_{b,j}-1\Big)=0,\ \mu\ge 0
\end{cases} \tag{37}$$

which can be solved by using bisection in terms of μ, where $\sum_{j=1}^{N_B}\check{c}_{b,j}=1$ must be satisfied. When the power is high, $P_B\to\infty$, the first equation in (37) can be approximated as $\frac{\partial\mathcal{L}}{\partial\check{c}_{b,j}}=\sum_{i=1}^{N_A}\frac{1}{\check{c}_{b,j}^2TP_B\tilde{\lambda}_{b,j}\tilde{\lambda}_{a,i}}-\mu=0$, and one can see that the optimal pilot here in the high power case is not necessarily uniform in power distribution since each $\check{c}_{b,j}^*$ is depending on $\tilde{\lambda}_{b,j}$. On the other hand, when $P_B\to 0$, (34c) can be approximated by its first order Taylor series expansion at point $P_B=0$:

$$J(P_B)\approxeq J(0)+\dot{J}(0)P_B=N_AN_B\sigma^2-\sigma^4N_AT\sum_{j=1}^{N_B}\tilde{\lambda}_{b,j}\check{c}_{b,j}P_B \tag{38}$$

where, unlike (32), the first order term is not zero. To minimize (38), the optimal solution is such that all power is allocated to $\check{c}_{b,1}$. The above discussion is about the pilot to be used by Bob. The pilot to be used by Alice can be determined similarly.

Furthermore, it is easy to verify that the optimal pilots $\mathbf{P}_A$ and $\mathbf{P}_B$ designed here also have a common row subspace of the dimension $\min\{N_A,N_B\}$, which satisfies the requirement of ANECE.

REFERENCES

[1] R. Ahlswede and I. Csiszar, “Common randomness in information theory and cryptography. I. Secret sharing,” *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, Jul 1993.

[2] R. Wilson, D. Tse, and R. A. Scholtz, “Channel identification: Secret sharing using reciprocity in ultrawideband channels,” *IEEE Trans. Inf. FORENSICS Secur.*, vol. 2, no. 3, pp. 270–275, 2007.

[3] T.-H. Chou, S. C. Draper, and A. M. Sayeed, “Key Generation Using External Source Excitation: Capacity, Reliability, and Secrecy Exponent,” *IEEE Trans. Inf. Theory*, vol. 58, no. 4, pp. 2455–2474, Apr 2012.

[4] L. Lai, Y. Liang, and H. V. Poor, “A Unified Framework for Key Agreement Over Wireless Fading Channels,” *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 2, pp. 480–490, Apr 2012.

[5] E. A. Jorswieck, A. Wolf, and S. Engelmann, “Secret key generation from reciprocal spatially correlated MIMO channels,” *2013 IEEE Globecom Workshops (GC Wkshps)*, pp. 1245–1250, 2013.

[6] A. Khisti, “Secret-Key Agreement Over Non-Coherent Block-Fading Channels With Public Discussion,” *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7164–7178, Dec 2016.

[7] J. W. Wallace and R. K. Sharma, “Automatic secret keys from reciprocal MIMO wireless channels: Measurement and analysis,” *IEEE Trans. Inf. Forensics Secur.*, vol. 5, no. 3, pp. 381–392, 2010.

[8] K. Chen, B. B. Natarajan, and S. Shattil, “Secret Key Generation Rate With Power Allocation in Relay-Based LTE-A Networks,” *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 11, pp. 2424–2434, Nov 2015.

[9] B. T. Quist and M. A. Jensen, “Maximization of the Channel-Based Key Establishment Rate in MIMO Systems,” *IEEE Trans. Wirel. Commun.*, vol. 14, no. 10, pp. 5565–5573, 2015.

[10] Y. Hua, “Advanced Properties of Full-Duplex Radio for Securing Wireless Network,” *IEEE Trans. Signal Process.*, vol. 67, no. 1, pp. 120–135, Jan 2019.

[11] M. Bloch and J. Barros, *Physical-layer security: from information theory to security engineering*. Cambridge University Press, 2011.

[12] M. Fiedler, “Bounds for the Determinant of the Sum of Hermitian Matrices,” *Proc. Am. Math. Soc.*, vol. 30, no. 1, p. 27, Sep 1971.

[13] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, ser. Springer Series in Statistics. New York, NY: Springer New York, 2011.

[14] E. Björnson and B. Ottersten, “A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance,” *IEEE Trans. Signal Process.*, vol. 58, no. 3 PART 2, pp. 1807–1820, 2010.