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 $C_{\text{key}}$ 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 $H$ by the users, and the MI between the maximum likelihood estimates (MLE) of $H$ 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 $C_{\text{key}}$.

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 $W_{BA} \in \mathbb{C}^{N_B \times N_A}$, and that from Bob to Alice by $W_{AB} = W_{BA}^T$ where the reciprocal property is applied. Let $P_A \in \mathbb{C}^{N_A \times T}$ be the pilot matrix transmitted by Alice over $T$ symbol intervals, and similarly

![Fig. 1. System model for secret key generation between Alice and Bob](image-url)
\( \mathbf{P}_B \in \mathbb{C}^{N_B \times T} \) 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

\[
\mathbf{Y}_A = \mathbf{W}_{AB} \mathbf{P}_B + \mathbf{N}_A \\
\mathbf{Y}_B = \mathbf{W}_{AB}^{T} \mathbf{P}_A + \mathbf{N}_B
\]

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. \( \mathcal{C}\mathcal{N}(0,1) \) entries. Furthermore, we assume that \( \mathbf{W}_{AB} = \mathbf{R}_B^{1/2} \mathbf{H}^{T} \mathbf{R}_B^{1/2} \) with \( \mathbf{R}_A = \mathbf{R}_B \mathbf{R}_A^{T} \) and \( \mathbf{R}_B = \mathbf{R}_B \mathbf{R}_B^{T} \) as known channel correlation matrices but with \( \mathbf{H} \in \mathbb{C}^{N_A \times N_B} \) consisting of i.i.d \( \mathcal{C}\mathcal{N}(0, \sigma^2) \) 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 \( 2T \leq T_c \).

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 

\[
C_{\text{key}} = I(\mathbf{Y}_A;\mathbf{Y}_B)
\]

In this paper, we will address how to choose \( \mathbf{P}_A \) and \( \mathbf{P}_B \) to maximize \( C_{\text{key}} \). Specifically, we consider the following

\[
\begin{align*}
\max_{\mathbf{P}_A, \mathbf{P}_B} & \quad C_{\text{key}} \\
\text{s.t.} & \quad \text{Tr}(\mathbf{P}_A \mathbf{P}_A^{H}) \leq T \mathbf{P}_A, \quad \text{Tr}(\mathbf{P}_B \mathbf{P}_B^{H}) \leq T \mathbf{P}_B
\end{align*}
\]

where \( \mathbf{P}_A \) and \( \mathbf{P}_B \) 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 \( T \geq \max \{N_A, N_B\} \) 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 \( C_{\text{key}} \)

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

\[
\begin{align*}
\mathbf{y}_A &= \mathbf{G}_A \mathbf{h} + \mathbf{n}_A \\
\mathbf{y}_B &= \mathbf{G}_A \mathbf{h} + \mathbf{n}_B
\end{align*}
\]

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

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

\[
\begin{align*}
\mathbf{h}_A &= \mathbf{K}_{y,y}^{-1} \mathbf{y}_A = \sigma^2 \mathbf{G}_B^{H} (\sigma^2 \mathbf{G}_B \mathbf{G}_B^{H} + I)^{-1} (\mathbf{G}_B \mathbf{h} + \mathbf{n}_A) \\
\mathbf{h}_B &= \mathbf{K}_{y,y}^{-1} \mathbf{y}_B = \sigma^2 \mathbf{G}_A^{H} (\sigma^2 \mathbf{G}_A \mathbf{G}_A^{H} + I)^{-1} (\mathbf{G}_A \mathbf{h} + \mathbf{n}_B)
\end{align*}
\]

with

\[
\mathbf{K}_{(\mathbf{h}_A, \mathbf{h}_B)} = \begin{bmatrix} \mathbf{K}_{\mathbf{h}_A} & \mathbf{K}_{\mathbf{h}_A, \mathbf{h}_B} \\ \mathbf{K}_{\mathbf{h}_A, \mathbf{h}_B}^{T} & \mathbf{K}_{\mathbf{h}_B} \end{bmatrix}
\]

and \( \mathbf{K}_{\mathbf{h}_A} = \sigma^2 \mathbf{G}_B^{H}(\sigma^2 \mathbf{G}_B \mathbf{G}_B^{H} + I)^{-1} \mathbf{G}_B \), \( \mathbf{K}_{\mathbf{h}_B} = \sigma^2 \mathbf{G}_A^{H}(\sigma^2 \mathbf{G}_A \mathbf{G}_A^{H} + I)^{-1} \mathbf{G}_A \), \( \mathbf{K}_{\mathbf{h}_A, \mathbf{h}_B} = \sigma^2 \mathbf{G}_B^{H}(\sigma^2 \mathbf{G}_B \mathbf{G}_B^{H} + I)^{-1} \mathbf{G}_B \mathbf{G}_A^{H}(\sigma^2 \mathbf{G}_A \mathbf{G}_A^{H} + I)^{-1} \mathbf{G}_A \). Note that we have applied that \( \mathbf{K}_{\mathbf{h}_A} \) and \( \mathbf{K}_{\mathbf{h}_B} \) are both invertible.\(^{[6a]}\)\(I(\mathbf{h}_A;\mathbf{h}_B) = h(\mathbf{h}_A) + h(\mathbf{h}_B) - h(\mathbf{h}_A, \mathbf{h}_B)
\]

\[
\begin{align*}
&= \log_2 |\mathbf{K}_{\mathbf{h}_A}| + \log_2 |\mathbf{K}_{\mathbf{h}_B}| - \log_2 |\mathbf{K}_{(\mathbf{h}_A, \mathbf{h}_B)}| \\
&= -\log_2 (1 - \mathbf{K}_{\mathbf{h}_B}^{-1} \mathbf{K}_{\mathbf{h}_B}^{-1} \mathbf{K}_{\mathbf{h}_A, \mathbf{h}_B}^{-1}) \\
&= -\log_2 (1 - \sigma^{-4} \mathbf{K}_{\mathbf{h}_A}^{-1} \mathbf{h}_A)
\end{align*}
\]
descending order respectively. From (10), we have $R_d^T = \bar{U}_A \bar{A}_d^T$ and $R_R^T = \bar{U}_B \bar{A}_R^T$. Based on (9) and (10), we have

$$P_A = (R_A^T U_A A_V A_H^T)^*$$
$$P_B = (R_B^T U_B A_B V_H^T)^*$$

(11)

In the following, we will recall $G_B = (P_B^T R_B \otimes R_A^T)$ and $G_A = (R_A^T \otimes P_A^T R_A^T)$, apply the property $(A \otimes B) (C \otimes D) = (AC \otimes BD)$, and use the definitions of $A_d^T = A_d^T A_A$ and $A_B^T = A_B A_A$. It follows from (9) and (10) that

$$K_{R,B} = \sigma^d G_H^T (\sigma^2 G_A G_H^T + I)^{-1} G_A$$
$$= \sigma^d (\bar{U}_B \bar{A}_B \otimes U_A A_B A_V^H (\sigma^2 \bar{U}_B \bar{A}_B \bar{U}_B^H$$
$$\otimes V_A A_B^T A_A V_H^T + I)^{-1} (\bar{U}_B \bar{A}_B \otimes V_A A_B^T A_A V_H^T$$
$$= \sigma^d (\sigma^2 I + (\bar{A}_B \otimes U_A A_B^T A_A V_H^T)^{-1} - 1$$

(12)

$$K_{R,A} = \sigma^d G_H^T (\sigma^2 G_B G_H^T + I)^{-1} G_B$$
$$= \sigma^d (U_B A_B V_H^T \otimes \bar{A}_B \bar{U}_B^H$$
$$\otimes \bar{U}_A \bar{A}_A \bar{U}_A^H + I)^{-1} (V_B A_B^T V_H^T \otimes \bar{U}_A \bar{A}_A^T V_H^T$$
$$= \sigma^d (\sigma^2 I + (U_B A_B^T V_H^T \otimes \bar{A}_A \bar{A}_A^T)^{-1} - 1$$

(13)

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

$$I(\bar{h}_A; \bar{h}_B) = -\log_2 \frac{\sigma^d (\sigma^2 I + (\bar{A}_B \otimes U_A A_B^2 \bar{A}_B A_H^T)^{-1} - 1$$

(14)

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

$$I(\bar{h}_A; \bar{h}_B) = \log_2 [\sigma^d (\sigma^2 I + (\bar{A}_B \otimes U_A A_B \bar{A}_B A_H^T)^{-1} - 1$$

+ $\log_2 [\sigma^d I + (U_B A_B V_H^T \otimes \bar{A}_A \bar{A}_A^T)^{-1} - 1$$

- $\log_2 [(\sigma^2 I + (U_B A_B V_H^T \otimes \bar{A}_B \bar{A}_B^T A_A V_H^T)^{-1} - 1$$

- $\log_2 [\sigma^d (\sigma^2 I + (U_B A_B V_H^T \otimes \bar{A}_B \bar{A}_B^T A_A V_H^T)^{-1} - 1$$

- $\log_2 [\sigma^d (\sigma^2 I + (U_B A_B V_H^T \otimes \bar{A}_A \bar{A}_A^T)^{-1} - 1$$

- $\log_2 [\sigma^d (\sigma^2 I + (U_B A_B V_H^T \otimes \bar{A}_A \bar{A}_A^T)^{-1} - 1$$

(15a)

$$= \log_2 [\sigma^d (\sigma^2 I + (\bar{A}_B \otimes U_A A_B \bar{A}_B A_H^T)^{-1} - 1$$

(15b)

where $ \in \in U_B \otimes U_A^H$. Here, (15a) is due to $\log_2 [\sigma^d \sigma^d I + (\bar{A}_B \otimes U_A A_B \bar{A}_B A_H^T)^{-1} - 1$$

(15b) is due to $\log_2 [\sigma^d I + (\bar{A}_B \otimes U_A A_B \bar{A}_B A_H^T)^{-1} - 1$$

Further, the optimization of $U_A$ and $U_B$ can be formulated as

$$\{U_{A,opt}, U_{B,opt}\} = \arg \min_{U_A, U_B} \log_2 [\sigma^d (\sigma^2 I + (\bar{A}_B \otimes U_A A_B \bar{A}_B A_H^T)^{-1} - 1$$

(16)

According to [12], we have:

**Lemma 1:** Given Hermitian matrices $A, C \in \mathbb{C}^{n \times n}$ and $B, D \in \mathbb{C}^{m \times m}$ with the corresponding diagonal eigenvalue matrices $A_{\alpha}, A_{\gamma}, A_{\beta}, A_{\lambda}$, and the diagonal elements in each diagonal matrix are in descending order. Then

$$|A \otimes B + C \otimes D| \geq \min_{P_1, P_2} |A_{\alpha} \otimes A_{\beta} + A_{c,P_1} \otimes A_{d,P_2}|$$

(17)

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

From Lemma 1, we have:

**Corollary 1:** If $A, B, C, D$ are positive semi-definite Hermitian matrices, then

$$|A \otimes B + C \otimes D| \leq \max_{P_1, P_2} |A_{\alpha} \otimes A_{\beta} + A_{c,P_1} \otimes A_{d,P_2}|$$

(18)

where elements in $\tilde{A}_c$ and $\tilde{A}_d$ are in ascending order.

**Proof:** Denote $\lambda_{c,s}$ and $\lambda_{c,t}$ as two elements in $A_{c,p_1}$ where $s < t$. Define two permutations $\lambda_{c,p_1} \geq \lambda_{c,p_1}$ and $\lambda_{c,p_1} \leq \lambda_{c,p_1}$. From (17) we have

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

(20)

For any given $j$ we have

$$\lambda_{a,b,i} + \lambda_{c,p_1,j} + \lambda_{c,p_1,j} + \lambda_{c,p_1,j} + \lambda_{c,p_1,j} \leq 0$$

(21)

Therefore, as elements in $\tilde{A}_c$ are in descending order, a descending $A_{c,p_1}$ will minimize the right hand side of (20). Similarly, as elements in $\tilde{A}_c$ are in descending order, a descending $A_{d,p_2}$ 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 $U_{A,opt} = I$ and $U_{B,opt} = I$, which are optimal for the objective function in (2).

Let $A_B^2 = \text{diag} \{A_{b,1}, \ldots, A_{b,N_B}\}$, $A_A^2 = \text{diag} \{A_{a,1}, \ldots, A_{a,N_A}\}$, and $A_B = \text{diag} \{A_{b,1}, \ldots, A_{b,N_B}\}$. Also let $C_A = A_A^{-1} A_B^2$ and $C_B = A_B^{-1} A_B^2$ with their elements denoted by $c_{a,i} = \lambda_{a,i} / \lambda_{a,i}$ and $c_{b,j} = \lambda_{b,j} / \lambda_{b,j}$. Therefore, we can optimize. With $U_A = I$ and $U_B = I$, (15b) becomes

$$I(\bar{h}_A; \bar{h}_B) = \log_2 [\sigma^d I + (\bar{A}_B \otimes U_A A_B \bar{A}_B A_H^T)^{-1} - 1$$

(18)

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

$$TR(P_A H) = TR(\bar{A}_A^{-1} U_A A_B^2 U_B^H)$$

$$\geq TR(\bar{A}_A^{-1} A_B^2) = \sum_{i=1}^{N_A} \sum_{i=1}^{N_A} c_{a,i}$$

(23)
Algorithm 1 Bisection section search

**Input:**
\( \lambda_0, \hat{\lambda}, P_A, P_B, T, \eta \)

Accuracy threshold \( \varepsilon_1, \varepsilon_2 \)

**Initialization** \( k = 0, c_b^{(k)} = \frac{T P_B}{N_B} 1_{N_B}, c_b^{(k)} = \frac{T P_B}{N_B} 1_{N_B} \).

1: repeat
2: Given \( c_b^{(k)} \), do bisection search of \( \mu \) and obtain solution \( c_a^{(k+1)} \) to meet the power constraint \( \sum_{i=1}^{N_A} c_{a,i} - T P_A \leq \varepsilon_1 \);
3: Given \( c_b^{(k+1)} \), do bisection search of \( \nu \) and obtain solution \( c_b^{(k+1)} \) to meet the power constraint \( \sum_{j=1}^{N_B} c_{b,j} - T P_B \leq \varepsilon_1 \);
4: until \( ||[c_b^{(k)}] - [c_b^{(k-1)}]|| \leq \varepsilon_2 \)
5: return \( \{c_a^{(k)}, c_b^{(k)}\} \)

\[ T r(P_B P_B^H) \geq T r(A_B^{-1} A_B^2) = \sum_{j=1}^{N_B} \lambda_{b,j} = \sum_{j=1}^{N_B} c_{b,j} \]  \hspace{1cm} (24)

where the equalities in the inequalities hold when \( U_B = I_{N_B} \) and \( U_A = I_{N_A} \) [13]. Namely, \( U_B = I_{N_B} \) and \( U_A = I_{N_A} \) are also optimal for the constraint in (22).

It is obvious that the unitary matrices \( V_A \) and \( V_B \) do not affect neither the objective function nor the constraint in (22). Without lose of generality, we can set them to be the identity matrices. Also note that by choosing \( V_A = V_B = I_T \), we have ensured that there is a common row subspace between \( P_A \) and \( P_B \) of the dimension \( \min \{N_A, N_B\} \) if \( A_A^2 \) and \( A_B^2 \) have the full ranks \( N_A \) and \( N_B \) respectively. It will also be shown that each of the optimal \( A_A^2 \) and \( A_B^2 \) is of full rank and always contains descending entries.

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

\[ \begin{align*}
\max_{c_a > 0, c_b > 0} & \sum_{j=1}^{N_B} \sum_{i=1}^{N_A} f_{i,j}(c_{a,i}, c_{b,j}) \\
\text{s.t.} & \sum_{i=1}^{N_A} c_{a,i} \leq T P_A, \sum_{j=1}^{N_B} c_{b,j} \leq T P_B 
\end{align*} \]  \hspace{1cm} (25a)

(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 \( \sum_{i=1}^{N_A} c_{a,i} = T P_A \) and \( \sum_{j=1}^{N_B} c_{b,j} = T P_B \).

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

\[ \begin{bmatrix}
\sigma^2 \lambda_{a,i} \lambda_{a,j} (\phi_{i,j} - \phi_{a,i,j}) & -\sigma^2 \lambda_{a,i} \lambda_{a,j} (\phi_{i,j} - \phi_{a,i,j}) \\
-\sigma^2 \lambda_{a,i} \lambda_{a,j} (\phi_{i,j} - \phi_{a,i,j}) & \sigma^2 \lambda_{a,i} \lambda_{a,j} (\phi_{i,j} - \phi_{a,i,j})
\end{bmatrix} \]  \hspace{1cm} (26)

where \( \phi_{a,i,j} = (1 + \sigma^2 \lambda_{a,i} \lambda_{a,j} c_{a,i} c_{a,j})^2, \phi_{b,i,j} = (1 + \sigma^2 \lambda_{a,i} \lambda_{a,j} c_{b,i} c_{b,j})^2, \phi_{a,i,j} = (1 + \sigma^2 \lambda_{a,i} \lambda_{a,j} c_{a,i} c_{a,j})^2 \).

This matrix is positive definite if and only if \( c_{a,i} c_{b,j} \geq \frac{1}{2 \sigma^2 \lambda_{a,i} \lambda_{a,j}} \). This means that when \( T P_A \) and \( T P_B \) 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 \( c_a \) and \( c_b \) 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

\[ L = \sum_{j=1}^{N_B} \sum_{i=1}^{N_A} f_{i,j}(c_{a,i}, c_{b,j}) - \mu \left( \sum_{i=1}^{N_A} c_{a,i} - T P_A \right) + \alpha^T c_a \]  \hspace{1cm} (27)

And the corresponding KKT conditions are

\[ \begin{align*}
\frac{\partial L}{\partial c_{a,i}} &= \frac{1}{\ln 2} \sum_{j=1}^{N_B} g_{i,j}(c_{a,i}, c_{b,j}) - \mu = 0 \\
\sum_{i=1}^{N_A} c_{a,i} &\leq T P_A, \mu \sum_{i=1}^{N_A} c_{a,i} - T P_A = 0, \mu \geq 0 \\
c_{a,i} &> 0, \alpha^T c_a = 0, \alpha \geq 0
\end{align*} \]  \hspace{1cm} (28)

where

\[ g_{i,j}(x,y) = \frac{\sigma^4 \lambda_{a,i} \lambda_{b,j} y}{(1 + \sigma^2 \lambda_{a,i} \lambda_{b,j} x)(1 + \sigma^2 \lambda_{a,i} \lambda_{b,j} x + \sigma^2 \lambda_{a,i} \lambda_{b,j} y)} \]  \hspace{1cm} (29)

In phase two, similar KKT conditions can be found. From (28) we can see that \( \mu \) 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 \( \lambda_{a,i} \) and a decreasing function of \( \lambda_{b,j} \). Given any \( c_{a,i} \), the solution from (28) is \( c_{b,j} \), 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 \( c_{b,j} \geq c_{b,j+1} \). If \( c_{b,j} < c_{b,j+1} \) then \( \mu \ln 2 = \sum_{j=1}^{N_B} g_{i,j}(c_{a,i}, c_{b,j}) > \sum_{j=1}^{N_B} g_{i,j}(c_{a,i+1}, c_{b,j}) \geq \sum_{j=1}^{N_B} g_{i,j}(c_{a,i+1}, c_{b,j}) \mu \ln 2 \), which is not possible.)

Similarly, \( c_{b,j} \geq c_{b,j+1} \). Therefore, the diagonal elements of the optimal solutions of \( A_A^2 \) and \( A_B^2 \) are also in descending order respectively.

**IV. ASYMPTOTIC ANALYSIS**

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

\[ I(\hat{h}_A; \hat{h}_B) = \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \log_2 \left( \frac{1 + T P \sigma^2 \lambda_{a,i} \lambda_{b,j} \hat{c}_{a,i} \hat{c}_{b,j}}{1 + T P \sigma^2 \lambda_{a,i} \lambda_{b,j} \hat{c}_{a,i} \hat{c}_{b,j} + T P \sigma^2 \lambda_{a,i} \lambda_{b,j} \hat{c}_{a,i} \hat{c}_{b,j}} \right) \]  \hspace{1cm} (30)

**A. High Power Case**

For large \( P \), (30) can be approximated as

\[ I(\hat{h}_A; \hat{h}_B) \approx \sum_{j=1}^{N_B} \sum_{i=1}^{N_A} \log_2 (\hat{c}_{a,i} \hat{c}_{b,j}) + \sum_{j=1}^{N_B} \sum_{i=1}^{N_A} \log_2 (T P \sigma^2 \lambda_{a,i} \lambda_{b,j}) \]  \hspace{1cm} (31)

From (31), we know that the degree of freedom per channel realization is \( \lim_{P \to \infty} \frac{\phi(\hat{c}_{a,i} \hat{c}_{b,j} \hat{a} \hat{b})}{\log_2 P} = N_A N_B \).
Also, \(-\frac{\partial^2 \phi_1}{\partial \alpha_i^2} = -\sum_j (c_{a,i} - c_{b,j})^2 \geq 0\), which means that \(-\phi_1\) is a convex function of \(c_a\). Meanwhile, \(-\phi_1\) is a symmetric function of \(c_a\). Therefore, \(\phi_1\) is a Schur-concave function [13] of \(c_a\), and then we have \(\phi_1(1_{N_A}, c_b, \lambda_A, \lambda_b) \geq \phi_1(c_a, c_b, \lambda_A, \lambda_b)\) with any descending \(c_a\). Similar idea can be applied to show that (31) is also a Schur-concave function of \(c_b\). Therefore, the optimal power allocation in the high power case is such that \(c_a = \frac{1}{n_A} 1_{N_A}\) and \(c_b = \frac{1}{n_B} 1_{N_B}\).

Also, by applying the same argument, one can easily prove that (31) is also a Schur-concave function of \(\lambda_a\) and \(\lambda_b\) respectively. Therefore, when \(\lambda_a = 1_{N_A}\) and \(\lambda_b = 1_{N_B}\), (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 \(P = 0\):

\[
I(\mathbf{h}_A; \mathbf{h}_B) = I(\hat{\mathbf{h}}_A; \hat{\mathbf{h}}_B)|_{P=0} + \frac{1}{2} I(\hat{\mathbf{h}}_A; \hat{\mathbf{h}}_B)|_{P=0} P^2 + o(P^2)
\]

(32)

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

\[
\hat{I}(\hat{\mathbf{h}}_A; \hat{\mathbf{h}}_B)|_{P=0} = 2 \ln 2 \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} \sigma^4 \lambda_{a,i}^2 \lambda_{b,j}^2 T^2 \tilde{c}_{a,i} \tilde{c}_{b,j} \Delta \phi_2(c_a, c_b, \lambda_a, \lambda_b)
\]

(33)

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

Also, \(\phi_2(c_a, c_b, \lambda_a, \lambda_b)\) is a Schur-concave function of \(\lambda_a\) and \(\lambda_b\) is a Schur-convex function of \(\lambda_a\). 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) \(I_{\text{MSKR}}\) - maximum secret key rate (MSKR) from (25); (2) \(I_{\text{MCEE}}\) - secret key rate based on minimum channel estimation error (MCEE); and (3) \(I_{\text{U}}\) - secret key rate based on uniform power allocation \(C_i = \frac{P_T}{N_i} I\), \(i \in \{A, B\}\). Define the channel correlation matrix as \(\rho[i,j] = r|i-j|\) 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 \(P_A T = P_B T = P_T\) and \(T \geq \max\{N_A, N_B\} = 8\). In Fig. 2, we show the secret key rate (SKR) of \(I_{\text{MSKR}}\) 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 \(I_{\text{MCEE}}\) and \(I_{\text{U}}\) 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.

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

\[ J = Tr(\mathcal{E}( (h - \hat{h}_A)(h - \hat{h}_A)^H) ) \]

\[ = Tr( K_h - K_{h,y} K_y^H K_{y,h} ) \]

\[ = Tr( A^2 I - \sigma^4 G_B (\sigma^2 G_B G_B^H + I)^{-1} G_B^H ) \]

\[ = Tr( (I + \sigma^2 G_B G_B^H)^{-1} (I + \sigma^2 G_B G_B^H)^{-1} ) \] (34a)

\[ = Tr( (I + \sigma^2 U_B A_B^2 U_B^H \otimes \hat{A})^{-1} (I + \sigma^2 U_B A_B^2 U_B^H \otimes \hat{A})^{-1} ) \] (34b)

\[ = \sum_{j=1}^{NB} \sum_{i=1}^{NA} \sigma^2 \left( \frac{1}{1 + \sigma^2 c_{b,j} T P_B \lambda_{b,j} \lambda_{a,i}} \right), \] s.t. \( \sum_{j=1}^{NB} c_{b,j} \leq 1 \) (34c)

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 \( c_b = c_b / (T P_B) = \lambda_b \lambda_b^{-1} / (T P_B) \). Since (34) is invariant to \( U_B \) and \( V_B \), then (24) implies that the optimal \( U_B \) and \( V_B \) are the identity matrices. Then the optimization problem with respect to \( c_b \) becomes

\[ \min_{c_b} \sum_{j=1}^{NB} \sum_{i=1}^{NA} \frac{\sigma^2}{1 + \sigma^2 c_{b,j} T P_B \lambda_{b,j} \lambda_{a,i}}, \] s.t. \( \sum_{j=1}^{NB} c_{b,j} \leq 1 \) (35)

The corresponding Lagrangian function is

\[ \mathcal{L} = \sum_{j=1}^{NB} \sum_{i=1}^{NA} \frac{\sigma^2}{1 + \sigma^2 c_{b,j} T P_B \lambda_{b,j} \lambda_{a,i}} + \mu \left( \sum_{j=1}^{NB} c_{b,j} - 1 \right) \] (36)

The KKT conditions are

\[ \left\{ \begin{array}{l}
\frac{\partial \mathcal{L}}{\partial c_{b,j}} = \sum_{i=1}^{NA} \frac{\sigma^2 T P_B \lambda_{b,j} \lambda_{a,i}}{1 + \sigma^2 c_{b,j} T P_B \lambda_{b,j} \lambda_{a,i}} - \mu = 0 \\
\sum_{j=1}^{NB} c_{b,j} \leq 1, \mu \left( \sum_{j=1}^{NB} c_{b,j} - 1 \right) = 0, \mu \geq 0
\end{array} \right. \] (37)

which can be solved by using bisection in terms of \( \mu \), where \( \sum_{j=1}^{NB} c_{b,j} = 1 \) must be satisfied. When the power is high, \( P_B \to \infty \), the first equation in (37) can be approximated as

\[ \frac{\partial \mathcal{L}}{\partial c_{b,j}} = \sum_{i=1}^{NA} \frac{\sigma^2}{1 + \sigma^2 c_{b,j} T P_B \lambda_{b,j} \lambda_{a,i}} - \mu = 0, \] and one can see that the optimal pilot here in the high power case is not necessarily uniform in power distribution since each \( c_{b,j} \) is depending on \( \lambda_{b,j} \). On the other hand, when \( P_B \to 0 \), (34c) can be approximated by its first order Taylor series expansion at point \( P_B = 0 \):

\[ J(P_B) \approx J(0) + \frac{\partial J}{\partial P_B} \bigg|_{P_B=0} P_B = N_A N_B \sigma^2 \left( \sum_{j=1}^{NB} \lambda_{b,j} c_{b,j} P_B \right) \]

where, unlike (32), the first order term is not zero. To minimize (38), the optimal solution is such that all power is allocated to \( \tilde{c}_{b,j} \). 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 \( P_A \) and \( P_B \) designed here also have a common row subspace of the dimension \( \min\{N_A, N_B\} \), which satisfies the requirement of ANECE.

REFERENCES


