## Title

Physical Layer Security with Full-Duplex Radio in Wireless Networks

## Author

Zhu, Qiping

## Publication Date

2019

UNIVERSITY OF CALIFORNIA
RIVERSIDE

# Physical Layer Security with Full-Duplex Radio in Wireless Networks

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Qiping Zhu

September 2019

Dissertation Committee:

Dr. Yingbo Hua, Chairperson
Dr. Salman Asif
Dr. Amit K. Roy-Chowdhury

The Dissertation of Qiping Zhu is approved:

Committee Chairperson

University of California, Riverside

## Acknowledgments

I am grateful to my advisor Dr. Yingbo Hua, who is always being patient to me, teaches me with his erudite knowledge and affects me with his rigorous personality. Without his help, I would not have been here.

I would like to thank Dr. Amit K. Roy-Chowdhury, Dr. Salman Asif, Dr. Shaolei Ren for their generous help as being committee members for my dissertation evaluation and defense.

I would like to thank my labmates Lei Chen, Peixi Liu, Qingpeng Liang, Shuo Wu, Reza Sohrabi for their kindly help.

To my parents for all the support.

# ABSTRACT OF THE DISSERTATION

Physical Layer Security with Full-Duplex Radio in Wireless Networks

by

Qiping Zhu

Doctor of Philosophy, Graduate Program in Electrical Engineering
University of California, Riverside, September 2019
Dr. Yingbo Hua, Chairperson

Physical layer security (PLS) is an approach that provides secrecy based on information-theoretic model which does not account for any computation capability assumption or pre-installed standardized secret key generation algorithm and it is a good additional protection on the top of the existing security scheme. This work includes four different topics for improving PLS with full-duplex radio. In the first topic, we develop a fast algorithms for computing power allocations in subcarriers, subject to power and rate constraints, to maximize the secrecy capacity of a three-node network consisting of a source, a full-duplex capable destination and an eavesdropper. The optimal power allocation at the destination is found to be significant especially when its power budget is high. The second topic is about the analysis of a two-phase scheme for secret information transmission with the technique of anti-eavesdropping channel estimation (ANECE) which can degenerate Eve's channel estimation with any number of antennas at Eve. The analysis is based on the assumption that everyone has prior statistical knowledge of its channel state information (CSI) and it yields lower and upper bounds on secure degrees of freedom as functions of the number of antennas on Eve and the size of information packet. For the third topic, we present optimal designs of the pilots for ANECE based on two criteria. The first is to optimize the minimum mean square error (MMSE) channel estimation for the users, and the second is to maximize the mutual information between the pilot-driven signals observed by the users. Closed-form optimal pilots are shown under both criteria but subject to a symmetric and isotropic condition. Algorithms for computing the optimal pilots are shown for general cases. In the fourth topic, we analyze the secure degree of freedom and the asymptotic expression of the achievable secret key rate from a two-phase key generation scheme which consists of channel training phase and secure information transmission phase. Based on the asymptotic secret key rate, we develop an efficient algorithm for coherence time allocation between the two phases to maximize the achievable secret key rate.

# Contents

# List of Figures

# Chapter 1

# Introduction

## 1.1 Motivations

Security of wireless networks is of paramount importance in today's world as billions of people around the globe are dependent upon these networks for a myriad of activities for their businesses and lives. A secure wireless network should satisfy the requirements of authenticity, confidentiality, integrity, and availability [1]. Among these issues, confidentiality is of particular interest to many researchers in recent years which also is the main focus in this work. For convenience, we will refer to confidentiality as security and vice versa.

The traditional way to keep information confidential from unauthorized persons and/or devices is via cryptography at upper layers of the network, which include the asymmetric-key method (involving a pair of public key and private key) and the symmetric-key method (involving a secret key shared between two legitimate users). As the computing capabilities of modern computers (including quantum computers) rapidly improve, the asymmetric-key method is increasingly vulnerable as this method relies upon computational complexity for security. Moreover, for the emerging decentralized networks, i.e. 5G network and IoT network, devices with different computation capability may randomly connect in or leave the network at any instance. Therefore, establishing standardized cryptography key management and distribution become very challenging [2].

Different to the traditional cryptography methods, physical layer security (PLS) is a technique that does not account for any assumption of computation capability from the adversary and it is under information-theoretic security model that the information leakage to eavesdropper can be precisely expressed which is a function of the channel quality. PLS can provide either secret key establishment or direct secret information transmission between the users, which make it a good additional protection for the existing security scheme. There are two complementary approaches in PLS [3]: wiretap channel model and secret key agreement. The former requires one user to transmit secret information directly to the other and the performance measurement is called secrecy capacity. The latter requires users to use their (correlated) observations and public discussion agreement to establish a secret key and the performance measurement is called secret key capacity.

## 1.2 Wiretap Channel Model

Wiretap channel model is first introduced by Wyner [4] and it requires the eavesdropping channel is noisier or less capable than the channel between the users. Later, Csiszar and Korner extend the theory to broadcast channel with confidential and common messages [5]. Particularly, the quantity to measure the secret message is secrecy capacity and it is defined as

$$\mathcal{C}_{WT} = \max_{p_{VX}}(I(V;Y) - I(V;Z)) \tag{1.1}$$

where $X$ is the source codewords, $Y$ is the received codewords at legitimate node, $Z$ is the received codewords at eavesdropping node and $V$ is the random variable accounts for randomization in the encoder. Equation (1.1) is based on the Markov chain $V \rightarrow X \rightarrow YZ$ and $\mathcal{C}_{WT}$ is maximization regarding to the joint distribution $p_{VX}$. If eavesdropping channel is less capable than the channel between the legitimate nodes, then (1.1) is maximized by $V = X$ and it can be simplified as

$$\mathcal{C}_{WT} = \max_{p_X}(I(X;Y) - I(X;Z)). \tag{1.2}$$

## 1.3 Secret Key Agreement

The theory of physical layer secret key agreement is first established by Maurer [6], Ahlswede and Csiszar [7, 8]. There two different models for key generation: source model and channel model. In source model, all the parties observe their individual realizations of a random source which is assumed to be outside the control of all parties. Denote $X$, $Y$, $Z$ are the observed signals at the two legitimate nodes and eavesdropping node respectively, then through public discussion the secret key capacity satisfies

$$I(X;Y) - \min\{I(X;Z), I(Y;Z)\} \leq \mathcal{C}^s_{key} \leq \max\{I(X;Y), I(X;Y|Z)\} \tag{1.3}$$

which can be recognized as $\mathcal{C}^s_{key} = I(X;Y)$ when $Z$ is independent of $(X,Y)$.

For the channel model, it can be viewed as wiretap channel model enhanced with public discussion. Denote $X$ as the transmitted secure codeword, $Y$, $Z$ as the received codeword at legitimated node and eavesdropping node respectively, then the secret key capacity for channel model satisfies

$$\max\left\{\max_{p_X}(I(X;Y)-I(X;Z)),\max_{p_X}(I(X;Y)-I(Y;Z))\right\}\leq \mathcal{C}_{key}^{c}\leq \max_{p_X}\min\{I(X;Y),I(X;Y|Z)\} \tag{1.4}$$

## 1.4 Contributions

In this work, we study four different topics for improving wireless physical layer security with full-duplex radio. Part of this work has been included in [9, 10, 11, 12, 13]

In chapter 2, we will consider a three-node multi-subcarrier network consisting of a source (Alice), a destination (Bob) with full-duplex and an eavesdropper (Eve). Bob is able to receive the signal from Alice and at the same time to transmit a jamming noise against Eve. Under normal circumstances where for example some public information is shared and all nodes can communicate friendly with each other and their channel state information be made available to all, we assume that Alice and Bob know their channel amplitudes with respect to Eve during secure data transmissions. We will utilize the knowledge of channel amplitudes in computing power allocations for maximum secrecy capacity and develop fast algorithms for this purpose. Unlike [14, 15, 16], we take into account the residual self-interference at Bob which is a more realistic model [17, 18, 19, 20, 21]. Another unique feature of this part is that we consider both power and rate constraints in maximizing the secrecy capacity while most of the prior works on physical layer security such as [15, 16, 22, 23, 24, 25, 26, 27, 28, 29, 30] only considered power constraints. In order to transmit a packet from Alice to Bob, a preselected data rate for the packet should be guaranteed.

In chapter 3, we analyze a two-phase scheme for secret information transmission proposed in [31]. In the first phase, an anti-eavesdropping channel estimation (ANECE) method is applied which allows users to find their channel state information (CSI) but suppresses Eve's ability to obtain its CSI. In the second phase, secret information is transmitted between Alice and Bob while Eve has little or no knowledge of its CSI. We will assume that Eve has a prior statistical knowledge of its CSI. With every node knowing its statistical model of CSI , we use mutual information to analyze the secret rate of the network, from which lower and upper bounds on the secure degrees of freedom (SDoF) are derived. These bounds are simple functions of the number of antennas on Eve. In literature, [32] studies the SDoF with perfect CSI at both receiver and eavesdropper but no CSI in transmitter, [33] derives the SDoF analysis based on two-phases scheme but no full duplex is involved. Our result has not been discovered in the literature and it is significant for understanding the property of ANECE.

In chapter 4, we present optimal designs of the pilot signals subject to ANECE requirement, which is to suppress Eve's channel estimation. We will consider two criteria for optimality: 1) minimizing the sum of mean squared errors (MSE) of the minimum-mean-squared-error (MMSE) channel estimation at each and every user, and 2) maximizing the sum of the pair-wise mutual information (MI) between the signals excited by the pilots and observed by all users. The first criterion is useful since the best channel estimation at each user allows the best detection of the information symbols transmitted subsequently following the pilots. The second criterion is also useful since the MI between two signals observed by two users is the capacity of secret key generation based on the two signals assuming that Eve's CSI is independent of the (reciprocal) CSI between the two users [34, 35, 36]. The novelty of our works includes: 1) Closed-form optimal pilots are presented under a symmetric and isotropic condition where each user has the same number of antennas, the same noise variance, the same transmit power and the independent and identically distributed (i.i.d.) channel coefficients; and 2) Algorithms for computing the optimal pilots for any other choices of the above parameters. The closed-form optimal pilots and the computed optimal pilots are compared with the previous choice shown in [31]. The algorithm for optimal MMSE channel estimation is an extension of [37] from two users to more than two users. The algorithm for maximum MI extends [38] from two users to more than two users. These extensions are significant contributions while they are subject to the ANECE requirement.

In chapter 5, we analyze the achievable secret key rate for a two-phase key generation scheme. In phase one, users will transmit pilot signals and all the nodes can successfully estimate their CSI. In phase two, users will transmit secret information. With the help of public discussion, the signal received in users from both phases will be utilized for key generation. We consider the users are equipped with full-duplex radio and multiple antennas. Such system model is an extension to the SISO system in [35, 36]. We show that full-duplex system can achieve higher SDoF compared to half-duplex when Eve's antenna number is no larger than the total transmitting antenna number from the users. The asymptotic secret key rate is derived and an effective algorithm for coherence time allocation between the two phases that maximizes the secret key rate is given. The simulation results show the relationship of Eve's antenna number and the coherence time allocation.

## 1.5 Notations

Vectors and matrices are represented by bold lower case and bold upper case respectively. The $n \times n$ identity matrix is $\mathbf{I}_n$ or simply $\mathbf{I}$ when its dimension is obvious. The trace, expectation, differential, natural logarithm, base-2 logarithm, determinant, transpose, conjugate, conjugated transpose and Kronecker product are respectively $Tr$, $\mathcal{E}$, $\partial$, ln, $\log_2$, $|\cdot|$, $^T$, $^*$, $^H$ and $\otimes$. The $n \times m$ real field and $n \times m$ complex field are $\mathbb{R}^{n \times m}$ and $\mathbb{C}^{n \times m}$. All other notations are defined in the context.

# Chapter 2

# Fast Power Allocation for Secure Communication with Full-Duplex Radio

## 2.1 System Model

A three-node wireless network is shown as Fig. 2.1. This is an ad hoc network where every node uses the same frequency band to communicate with other nodes. In this network (or a snapshot of this network), the source (Alice) plans to transmit some sensitive information to its legitimate destination (Bob) while a potential eavesdropper (Eve) is to be prevented from “wiretapping” the transmission. We assume that the channel on each link consists of $N$ orthogonal subcarriers and the fading on each subcarrier is flat. To actively deteriorate the SINR at Eve, Bob will use its full-duplex capacity to transmit

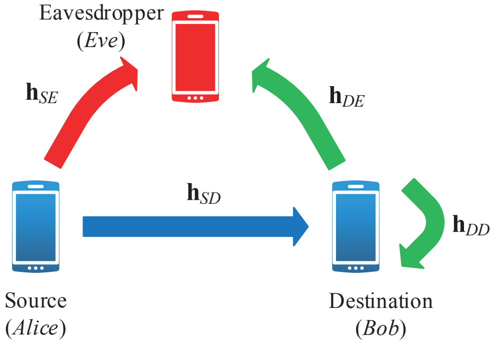

Figure 2.1: A three-node wireless network with a full-duplex destination.

interference noise in the same channel where at the same time it receives the signal from Alice. Potentially, all nodes could work in full duplex. But this would make the network much more complicated. If all nodes only work in half duplex, then this is a conventional network for which the conventional methods can be applied. The setting of our problems is somewhere in between the two extremes.

Let $\mathbf{x}_S(t) \in \mathbb{C}^{N\times 1}$ be the signal vector of approximately i.i.d. symbols of zero mean and unit variance) to be transmitted by Alice and $\mathbf{x}_D(t) \in \mathbb{C}^{N\times 1}$ be the jamming noise vector of approximately i.i.d. symbols of zero mean and unit variance) to be transmitted by Bob. Then the signal vectors to be received by Bob and Eve can be respectively expressed as:

$$\mathbf{y}_D(t) = \mathbf{h}_{SD} \circ \sqrt{\mathbf{p}_S} \circ \mathbf{x}_S(t) + \sqrt{\rho}\mathbf{h}_{DD} \circ \sqrt{\mathbf{p}_D} \circ \mathbf{x}_D(t) + \mathbf{n}_D(t),$$

$$\mathbf{y}_E(t) = \mathbf{h}_{SE} \circ \sqrt{\mathbf{p}_S} \circ \mathbf{x}_S(t) + \mathbf{h}_{DE} \circ \sqrt{\mathbf{p}_D} \circ \mathbf{x}_D(t) + \mathbf{n}_E(t),$$

where $\mathbf{h}_{SD} \in \mathbb{C}^{N\times 1}$ is the channel response vector from Alice to Bob, $\mathbf{h}_{SE} \in \mathbb{C}^{N\times 1}$ is that from Alice to Eve, $\mathbf{h}_{DE} \in \mathbb{C}^{N\times 1}$ is that from Bob to Eve, and $\mathbf{h}_{DD} \in \mathbb{C}^{N\times 1}$ is the self-interference channel response vector of Bob. $\mathbf{p}_S \in \mathbb{R}^{N\times 1}$ and $\mathbf{p}_D \in \mathbb{R}^{N\times 1}$ are the transmitting power vectors of Alice and Bob respectively; $\sqrt{\mathbf{p}_S}$ and $\sqrt{\mathbf{p}_D}$ denote the element-wise square roots of $\mathbf{p}_S$ and $\mathbf{p}_D$, respectively. Both $\mathbf{n}_D(t) \in \mathbb{C}^{N\times 1}$ and $\mathbf{n}_E(t) \in \mathbb{C}^{N\times 1}$ are independent white Gaussian noise of zero mean and unit variance. The symbol '$\circ$' denotes the *Hadamard product* (i.e., element-wise product). And $\rho$ is the self-interference attenuation factor.

Let $p_S^{(n)}$ denote the $n$th element of $\mathbf{p}_S$, and other similar notations are defined accordingly. The SINRs of the $n$th subcarrier at Bob and Eve are respectively:

$$\gamma_D^{(n)} = \frac{A_n x_n}{1 + B_n y_n} \quad \text{and} \quad \gamma_E^{(n)} = \frac{C_n x_n}{1 + D_n y_n}, \tag{2.1}$$

where $A_n = |h_{SD}^{(n)}|^2$, $B_n = \rho|h_{DD}^{(n)}|^2$, $C_n = |h_{SE}^{(n)}|^2$, $D_n = |h_{DE}^{(n)}|^2$, $x_n = p_S^{(n)}$ and $y_n = p_D^{(n)}$.

Note that we will assume that the channel amplitudes $A_n$, $B_n$, $C_n$ and $D_n$, $\forall n$ are available for computing power allocations. None of the channel phases is required. In practice, the amplitudes are much slower in changing and much easier to estimate than the phases are. Since the channel amplitudes have a large coherence time, any data transmission from Eve to Alice and bob could allow Alice and Bob to know the required channel amplitude responses from Alice and Bob to Eve via the reciprocal propositionerty. We also assume that Alice and Bob are fully cooperative.

Now secrecy capacity of the system in bits per channel use is known as [39]:

$$\mathcal{C}(\mathbf{x}, \mathbf{y}) = \frac{1}{N} \sum_{n=1}^{N} \max\{0, \Delta\mathcal{R}_n(x_n, y_n)\}, \tag{2.2}$$

where $\Delta\mathcal{R}_n(x_n, y_n) = \log(1+\gamma_D^{(n)}) - \log(1+\gamma_E^{(n)})$. The pre-multiplier $1/N$ in (2.2) should be removed if the $N$ subcarriers are spatial subcarriers (due to use of multiple antennas) instead of temporal subcarriers (due to time and/or frequency divisions). This paper is concerned about maximizing the secrecy capacity $\mathcal{C}(\mathbf{x}, \mathbf{y})$ through power allocations at both Alice and Bob. And most of the technical details are aimed to reduce the computational complexity.

In relation to $\mathcal{C}(\mathbf{x},\mathbf{y})$, we define $\tilde{\mathcal{R}}_s(\mathbf{x},\mathbf{y})$ as:

$$\tilde{\mathcal{R}}_s(\mathbf{x},\mathbf{y}) = \max\{0, \Delta\mathcal{R}(\mathbf{x},\mathbf{y})\}, \tag{2.3}$$

where $\Delta\mathcal{R}(\mathbf{x},\mathbf{y}) = \frac{1}{N}\sum_{n=1}^{N}\Delta\mathcal{R}_n(x_n, y_n)$.

Shown below are three important proposition. proposition 1 will be used to simplify the secrecy capacity as an objective function from a form of "summation of maximums" to a form of "maximum of sums". proposition 2 is a precursor of proposition 3, the latter of which provides a necessary condition to determine whether a subcarrier at Bob needs to be allocated with nonzero power.

**Proposition 1** $\mathcal{C}(\mathbf{x},\mathbf{y})$ *is no less than* $\tilde{\mathcal{R}}_s(\mathbf{x},\mathbf{y})$*, and* $\max(\mathcal{C}(\mathbf{x},\mathbf{y})) = \max(\tilde{\mathcal{R}}_s(\mathbf{x},\mathbf{y}))$ *subject to* $\sum_{n=1}^{N} x_n \leq P_S$ *and* $\sum_{n=1}^{N} y_n \leq P_D$*.*

**Proposition 2** *For any given* $x_n \in (0, +\infty)$*, there is at most one stationary point for* $\Delta\mathcal{R}_n(x_n, y_n)$ *with regard to* $y_n \in (0, +\infty)$*.*

**Proposition 3** *For any given* $x_n$*,* $\forall n$*, a necessary condition that the optimal value of* $y_n$ *is nonzero is that* $\frac{B_n}{D_n} < 1$ *and* $\frac{A_n}{C_n} > \frac{B_n}{D_n}$*.*

See the proof in [10] for proposition 1 - 3.

## 2.2 Power allocation under power constraints

In this section, we consider the problem of power allocation for maximization of secrecy capacity subject to power-only constraints. Specifically, we consider the following problem:

$$\max_{\mathbf{x},\mathbf{y}} \quad \mathcal{C}(\mathbf{x},\mathbf{y}) \tag{2.4a}$$

$$\begin{aligned} s.t. \quad & \sum_{n=1}^{N} x_n \leq P_S, \sum_{n=1}^{N} y_n \leq P_D, \\ & x_n \geq 0, y_n \geq 0, \forall n \in \mathsf{N}. \end{aligned} \tag{2.4b}$$

where we assume the power budget $P_S$ at source and the power budget $P_D$ at destination. Note that $\mathsf{N} \doteq \{1, ..., N\}$.

With proposition 1, the power allocation problem (2.4a) can be transformed equivalently to:

$$\begin{aligned} \max_{\mathbf{x},\mathbf{y}} \quad & \Delta\mathcal{R}(\mathbf{x},\mathbf{y}) \\ s.t. \quad & \textit{Power constraint } (2.4b). \end{aligned} \tag{2.5}$$

Solving this non-convex optimization problem (2.5) directly is still difficult. We will treat this problem in two phases: in phase one, we optimally allocate the source power for a given destination power vector; and in phase two, we optimally allocate the destination power for a given source power vector. The two phases will be iterated until convergence. Note that since the two-phase iteration algorithm increases the same (upper bounded) objective function at each iteration and each phase, this algorithm is guaranteed to be locally convergent. Such a propositionerty is a special case of one that is discussed in [40].

In the following two subsections, the two phases of the two-phase algorithm are discussed separately in detail.

### 2.2.1 Source power allocation

With a fixed destination power allocation, the source power allocation problem from (2.5) is:

$$\begin{aligned} \max_{\mathbf{x}} \quad & \frac{1}{N}\sum_{n=1}^{N}\log(1+\alpha_n x_n) - \frac{1}{N}\sum_{n=1}^{N}\log(1+\beta_n x_n) \\ s.t. \quad & \sum_{n=1}^{N} x_n \le P_S, x_n \ge 0, \forall n \in \mathsf{N}. \end{aligned} \tag{2.6}$$

where

$$\alpha_n = \frac{A_n}{1+B_n y_n} \quad \text{and} \quad \beta_n = \frac{C_n}{1+D_n y_n}. \tag{2.7}$$

The above problem is still non-convex due to the non-convex cost function. But we will be able to find the solution to this problem by finding the solution to its KKT conditions as follows.[1] The Lagrangian function of the problem can be written as:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \upsilon) = -\frac{1}{N}\sum_{n=1}^{N}\log\left(\frac{1+\alpha_n x_n}{1+\beta_n x_n}\right) - \boldsymbol{\lambda}^T\mathbf{x} + \upsilon(\sum_{n=1}^{N} x_n - P_S). \tag{2.8}$$

The solution to the problem (2.6) must satisfy the following KKT conditions [41]:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x_n} = -\varphi_n(x_n) - \lambda_n + \upsilon = 0, \\ \displaystyle\sum_{n=1}^{N} x_n \le P_S, \upsilon \ge 0, \upsilon(\sum_{n=1}^{N} x_n - P_S) = 0, \\ x_n \ge 0, \lambda_n \ge 0, \lambda_n x_n = 0, \forall n \in \mathsf{N}, \end{cases} \tag{2.9}$$

where

$$\varphi_n(x_n) = \frac{1}{N}\frac{\alpha_n}{1+\alpha_n x_n} - \frac{1}{N}\frac{\beta_n}{1+\beta_n x_n}. \tag{2.10}$$

[1] In general, the KKT conditions are necessary conditions for the optimal solution. But for all convex problems and some non-convex problems, the KKT conditions are both necessary and sufficient conditions for the optimal solution. When the solution to the KKT conditions is unique, it must be the optimal solution to the original problem. When KKT conditions (of a non-convex problem) have more than one solutions, one has to be innovative to exploit other propositionerties associated with the optimal solution to rule out the non-optimal solutions if possible.

Before solving these KKT conditions, we introduce the following proposition:

**Proposition 4** *Let $\mathbf{x}^\dagger$ be the solution of the source power allocation phase. Then, for any $n$, if $\alpha_n \leq \beta_n$, then $x_n^\dagger = 0$. Furthermore, we have either $\sum_{n=1}^{N} x_n^\dagger = 0$ or $\sum_{n=1}^{N} x_n^\dagger = P_S$.*

See proof in [10] for proposition 4 and follow that we have $x_n^\dagger = 0$ for $n \in \{n|\alpha_n \leq \beta_n, n \in \mathsf{N}\}$, and for the remaining subcarriers, the power allocation results can be obtained by solving the following simplified KKT conditions:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial x_n} = -\varphi_n(x_n) - \lambda_n + \upsilon = 0, \\ x_n \geq 0, \lambda_n \geq 0, \lambda_n x_n = 0, \forall n \in \Theta_y, \\ \displaystyle\sum_{n \in \Theta_y} x_n = P_S, \Theta_y \doteq \{n|\alpha_n > \beta_n, n \in \mathsf{N}\}. \end{cases} \tag{2.11}$$

It can be verified that $\frac{\partial \varphi_n(x_n)}{\partial x_n} < 0, \forall n \in \Theta_y$. From the first equation in (2.11), we know that $\upsilon$ is a decreasing function of $x_n$,$\forall n \in \Theta_y$. Thus, these simplified KKT conditions can be solved by a bisection search algorithm as shown in the table of Algorithm 1[2]. This algorithm is similar to a solution in [39].

[2] For KKT conditions, all the Lagrange multipliers (such as $\upsilon$ and $\lambda_n$) associated with the inequalities must be non-negative. For a given $\upsilon$ and $\lambda_n = 0$, the solution to $\varphi_n(x_n^\dagger) = \upsilon$ may or may not be positive. If there is a positive solution of $x_n$, the corresponding $\lambda_n$ is zero as assumed in the first place. If there is no positive solution of $x_n$, the corresponding optimal solution of $x_n$ is zero and the corresponding $\lambda_n$ should be positive (although its actual value is now useless).Also, $\varphi_n(x_n^\dagger) = \upsilon$ is equivalent to an quadratic equation which has two roots, only one of the two roots can be greater than or equal to 0, which is the valid solution.

**Algorithm 1** Source power allocation algorithm - solution to (2.11):

```
Input:
    $A_n, B_n, C_n, D_n, y_n, \forall n \in \mathsf{N}$; Source power constraint $P_S$; Accuracy threshold $\varepsilon$.
Output:
    $\upsilon^+ = \max_{n \in \Theta_y} \{\varphi_n(0)\}$; $\upsilon^- = \max_{n \in \Theta_y} \{\varphi_n(P_S)\}$;
Temporary variable $\mu = 0$; $x_1^\dagger = x_2^\dagger =, ..., = x_N^\dagger = 0$.
while ($|P_S - \mu| > \varepsilon$) do
    $\upsilon = \frac{\upsilon^- + \upsilon^+}{2}$;
    for $n \in \Theta_y$ do
        if $\upsilon \geq \varphi_n(0)$ then
            $x_n^\dagger = 0$;
        else
            Solve $\varphi_n(x_n^\dagger) = \upsilon$ (By solving an equivalent quadratic equation. There is only one
            positive solution to this equation due to the nature of the function $\varphi_n(x_n)$.) and set
            $x_n^\dagger = x_n$;
        end if
    end for
    $\mu = \sum_{n \in \Theta_y} x_n^\dagger$;
    if $\mu > P_S$ then
        $\upsilon^- = \upsilon$;
    else
        $\upsilon^+ = \upsilon$;
    end if
end while
return $x_1^\dagger, x_2^\dagger, ..., x_N^\dagger$.
```

### 2.2.2 Destination power allocation

With a given source power allocation, the destination power allocation problem from (2.5) is as follows:

$$\begin{aligned} \max_{\mathbf{y}} \quad & \frac{1}{N}\sum_{n=1}^{N}\left(\log(1+\frac{A_n x_n}{1+B_n y_n}) - \log(1+\frac{C_n x_n}{1+D_n y_n})\right) \\ s.t. \quad & \sum_{n=1}^{N} y_n \le P_D, y_n \ge 0, \forall n \in \mathsf{N}. \end{aligned} \tag{2.12}$$

By proposition 3, the above problem is equivalent to:

$$\begin{aligned} \max_{\mathbf{y}} \quad & \frac{1}{N}\sum_{n\in\Phi}(\log(1+\frac{A_n x_n}{1+B_n y_n}) - \log(1+\frac{C_n x_n}{1+D_n y_n})) \\ s.t. \quad & \sum_{n\in\Phi} y_n \le P_D, y_n \ge 0, \forall n \in \Phi, \end{aligned} \tag{2.13}$$

where

$$\Phi \doteq \{n | n \in \mathsf{N}, \frac{B_n}{D_n} < 1, \frac{A_n}{C_n} > \frac{B_n}{D_n}\}, \tag{2.14}$$

and $y_n = 0, \forall n \notin \Phi$.

The above problem is once again non-convex. To find its solution, we will consider its KKT conditions. The Lagrangian function of this problem is:

$$\begin{aligned} & \mathcal{L}(\mathbf{y}, \boldsymbol{\lambda}, \upsilon) = \\ & -\frac{1}{N}\sum_{n\in\Phi}(\log(1+\frac{A_n x_n}{1+B_n y_n}) - \log(1+\frac{C_n x_n}{1+D_n y_n})) \\ & -\boldsymbol{\lambda}^T\mathbf{y} + \upsilon(\sum_{n\in\Phi} y_n - P_D). \end{aligned} \tag{2.15}$$

The KKT conditions of (2.13) are:

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial y_n} = -\psi_n(y_n) - \lambda_n + \upsilon = 0, \\ \displaystyle\sum_{n \in \Phi} y_n \leq P_D, \upsilon \geq 0, \upsilon(\sum_{n \in \Phi} y_n - P_D) = 0, \\ y_n \geq 0, \lambda_n \geq 0, \lambda_n y_n = 0, \quad \forall n \in \Phi, \end{cases} \tag{2.16}$$

where

$$\begin{aligned} \psi_n(y_n) = \frac{1}{N} \frac{\partial \Delta \mathcal{R}_n}{\partial y_n} = \frac{1}{N} \Bigg( & \frac{B_n}{1 + B_n y_n + A_n x_n} \\ & - \frac{B_n}{1 + B_n y_n} - \frac{D_n}{1 + D_n y_n + C_n x_n} + \frac{D_n}{1 + D_n y_n} \Bigg). \end{aligned} \tag{2.17}$$

From (2.17), we know that the region of interest for $y_n$ is where $\psi_n(y_n) > 0$. In this region, $\psi_n(y_n)$ is decreasing with increasing $y_n$:

**Proposition 5** *$\psi_n(y_n)$ is decreasing with increasing $y_n$ as long as $\psi_n(y_n) > 0$ if $\frac{A_n}{B_n} > \frac{B_n}{D_n}$ and $\frac{B_n}{D_n} < 1$.*

See [10] for the proof of proposition 5 and follows that the KKT conditions in (2.16) can be solved with a bisection search of $\upsilon$ as shown in Algorithm 2.

**Algorithm 2** Destination power allocation algorithm - solution to (2.16)

```
Input:
    $A_n, B_n, C_n, D_n, x_n, \forall n \in \Phi$; Destination power constraint $P_D$; Accuracy threshold $\varepsilon$.
Output:
    $\upsilon^+ = \max_{n \in \Phi}\{\psi_n(0)\}$; $\upsilon^- = \max\{0, \max_{n \in \Phi}\{\psi_n(P_D)\}\}$;
for $n \in \Phi$ do
    if $\psi_n(0) \leq 0$ then
        $y_n^\dagger = 0$;
    else
        Solve $\psi_n(y_n^\dagger) = 0$ by solving an equivalent 4th-order polynomial which has only one positive root. The roots of 4th-order polynomial have closed-form expressions.;
    end if
end for
if $(\sum y_n^\dagger > P_D$ or $\upsilon^- > 0)$ then
    Temporary variable $\mu = 0$; $y_n^\dagger = 0, \forall n \in \Phi$.
    Do bisection search of $\upsilon$ and obtain solution $y_n^\dagger, \forall n$ to meet the power constraint $|\sum_{n \in \Phi} y_n^\dagger - P_D| \leq \epsilon$. The algorithm is similar to algorithm 1.
end if
return $y_n^\dagger, \forall n \in \Phi$.
```

## 2.3 Power allocation under power and rate constraints

In this section, we consider power allocation for maximizing the secrecy capacity of the three-node network subject to power constraints as well as a source-to-destination data rate constraint. Namely, we consider the following non-convex problem:

$$
\begin{aligned}
\max_{\mathbf{x},\mathbf{y}} \quad & \frac{1}{N}\sum_{n\in\Theta_y}\Delta\mathcal{R}_n(x_n,y_n) \\
s.t. \quad & \frac{1}{N}\sum_{n=1}^{N}\log(1+\frac{A_n x_n}{1+y_n B_n}) \geq \mathcal{C}_{SD}, \\
& \textit{Power constraint } (2.4b).
\end{aligned}
\tag{2.18}
$$

where $\mathcal{C}_{SD}$ is the required source-to-destination rate. (In the scenario of the key transmission, this rate should be the rate of the data packet containing the key.) The set $\Theta_y$ is the same set defined in (2.11), and $\Delta\mathcal{R}_n(x_n, y_n) < 0, \forall n \notin \Theta_y$. This is why the sum in the objective function is over $n \in \Theta_y$. However, due to the rate constraint, the optimal $x_n$ may be positive for some $n \notin \Theta_y$. So, the sum in the rate constraint must still be done over all $n \in \mathsf{N}$. The larger is the secrecy capacity (the first line in (2.18)), the more secure is the data rate from the source to the destination (the second line in (2.18)). The data packet transmitted from source to destination should be encoded at the source (and decoded at the destination) *jointly across all subcarriers* (not separately on each subcarrier).

Although the rate constraint introduces a complex situation where $x_n, \forall n$ and $y_n, \forall n$ now have a shared constraint, the two-phase iteration method is still applicable. Each of the two phases is discussed next.

### 2.3.1 Source power allocation

In this phase, $\mathbf{y}$ is fixed and the optimization problem (2.18) reduces to the following convex problem:

$$\begin{aligned} \max_{\mathbf{x}} \quad & \frac{1}{N}\sum_{n\in\Theta_y}[\log(1+\alpha_n x_n)-\log(1+\beta_n x_n)] \\ s.t. \quad & \frac{1}{N}\sum_{n=1}^{N}\log(1+\alpha_n x_n)\geq \mathcal{C}_{SD}, \\ & \sum_{n=1}^{N} x_n \leq P_S, x_n\geq 0, \forall n\in \mathsf{N}. \end{aligned} \tag{2.19}$$

where $\alpha_n$ and $\beta_n$ are defined in (2.7). The Lagrangian function of this problem is:

$$\begin{aligned} \mathcal{L}(\mathbf{x},\lambda,\boldsymbol{\mu},\upsilon) = & -\frac{1}{N}\sum_{n\in\Theta_y}(\log(1+\alpha_n x_n)-\log(1+\beta_n x_n)) \\ & +\lambda(\mathcal{C}_{SD}-\frac{1}{N}\sum_{n=1}^{N}\log(1+\alpha_n x_n))-\boldsymbol{\mu}^T\mathbf{x}+\upsilon(\sum_{n=1}^{N}x_n-P_S). \end{aligned} \tag{2.20}$$

The KKT conditions of (2.19) are

$$\begin{cases} \frac{\partial\mathcal{L}}{\partial x_n} = -\bar{\varphi}_n(x_n)-\frac{\lambda}{N}\frac{\alpha_n}{1+\alpha_n x_n}-\mu_n+\upsilon=0, \\ \lambda\geq 0, \frac{1}{N}\sum_{n=1}^{N}\log(1+\alpha_n x_n)\geq\mathcal{C}_{SD}, \\ \lambda(\frac{1}{N}\sum_{n=1}^{N}\log(1+\alpha_n x_n)-\mathcal{C}_{SD})=0, \\ x_n\geq 0, \mu_n\geq 0, \mu_n x_n = 0, \forall n\in\mathsf{N}, \\ \upsilon\geq 0, \sum_{n=1}^{N}x_n\leq P_S, \upsilon(\sum_{n=1}^{N}x_n-P_S)=0, \end{cases} \tag{2.21}$$

where $\bar{\varphi}_n(x_n)=\varphi_n(x_n)$ as defined by (2.10) for $n\in\Theta_y$, and $\bar{\varphi}_n(x_n)=0$ for $n\notin\Theta_y$. From the first equation in (2.21), we see that if $\lambda$ is fixed, $\upsilon$ is a decreasing function of $x_n$, and if $\upsilon$ is fixed, $\lambda$ is an increasing function of $x_n$. Hence, the conditions of (2.21) can be solved by a two-dimensional bisection search as summarized in the table of Algorithm 3. The bisection search of $\upsilon$ is to meet the power constraint, and the bisection search of $\lambda$ is to meet the rate constraint. For each given pair of $\upsilon$ and $\lambda$, the first equation in (2.21) is equivalent to a quadratic equation of $x_n$ and hence has a closed-form solution for $x_n$.

**Algorithm 3** Algorithm to solve the problem (2.19) by solving the KKT conditions (2.21), which uses 2-D bisection search for $\upsilon$ and $\lambda$

```
Input:
    $A_n, B_n, C_n, D_n, y_n, \forall n \in \mathsf{N}$; Source power constraint $P_S$; SD capacity constraint $\mathcal{C}_{SD}$; Accuracy threshold $\varepsilon$, $\zeta$.
Output:
Set $\lambda = 0$ (i.e., removing the rate constraint), do the search for $\upsilon$ and $\mathbf{x}$ (similar to Algorithm 1);
Calculate SD capacity $C(\mathbf{x})$;
if $C(\mathbf{x}) > \mathcal{C}_{SD}$ then
    return $\mathbf{x}$ (This means that the rate constraint is satisfied by the solution without the rate constraint even imposed.);
else
    Two-Dimensional bisection search: Do bisection search for $\upsilon > 0$ to meet the power constraint up to the precision $\varepsilon$. For each given $\upsilon$, do bisection search for $\lambda > 0$ to meet the rate constraint up to the precision $\zeta$. For each given pair of $\upsilon$ and $\lambda$, find $x_n \geq 0$ as the solution to the first equation in (2.21) for each $n \in \mathsf{N}$.
    return $\mathbf{x}$.
end if
```

### 2.3.2 Destination power allocation

In this phase, $\mathbf{x}$ is fixed and the problem (2.18) reduces to the following (still non-convex) problem:

$$\begin{aligned} \max_{\mathbf{y}} \quad & \frac{1}{N}\sum_{n\in\Theta_y}\left(\log(1+\frac{A_n x_n}{1+B_n y_n}) - \log(1+\frac{C_n x_n}{1+D_n y_n})\right) \\ s.t. \quad & \frac{1}{N}\sum_{n=1}^{N}\log(1+\frac{A_n x_n}{1+B_n y_n}) \geq \mathcal{C}_{SD}, \\ & \sum_{n=1}^{N} y_n \leq P_D, y_n \geq 0, \forall n \in \mathsf{N}. \end{aligned} \tag{2.22}$$

By proposition 3, the problem (2.22) can be rewritten as

$$\begin{aligned} \max_{\mathbf{y}} \quad & \frac{1}{N}\sum_{n\in\Psi_y}\left(\log(1+\frac{A_n x_n}{1+B_n y_n}) - \log(1+\frac{C_n x_n}{1+D_n y_n})\right) \\ s.t. \quad & \frac{1}{N}\sum_{n\in\Psi_y}\log(1+\frac{A_n x_n}{1+B_n y_n}) \geq \tilde{\mathcal{C}}_{SD}, \\ & \sum_{n\in\Psi_y} y_n \leq P_D, y_n \geq 0, \forall n \in \Psi_y, \end{aligned} \tag{2.23}$$

where

$$\Psi_y = \Theta_y \cap \Phi, \tag{2.24}$$

$$\begin{aligned} \tilde{\mathcal{C}}_{SD} &= \mathcal{C}_{SD} - \frac{1}{N}\sum_{n\in\Psi_y^{\perp}}\log(1+\frac{A_n x_n}{1+B_n y_n}), \\ \Psi_y^{\perp} &= \{n | n \in \mathsf{N}, n \notin \Psi_y\}, \end{aligned} \tag{2.25}$$

and $y_n = 0, \forall n \in \Psi_y^{\perp}$.

Because the set $\Psi_y$ is a function of $y_n$, $\forall n$, we will use the following approach to determine $\Psi_y$:

We start with the largest possible set of $\Psi_y$ which is $\Psi_y^{(0)} = \Phi$. Then, for any given $\Psi_y = \Psi_y^{(k)}$, solve the problem (2.23), substitute the solution $\mathbf{y}^{(k)}$ into the equation (2.24) to obtain a new $\Psi_y^{(k+1)}$. If $\Psi_y^{(k)} = \Psi_y^{(k+1)}$, stop, and $\mathbf{y}^{(k)}$ is the solution; otherwise, let $\Psi_y = \Psi_y^{(k+1)}$, and continue the iteration.

Now the main challenge is how to solve the problem (2.23) with a given $\Psi_y$. Since solving the exact KKT conditions of (2.23) is very tedious even if feasible, we will now use a sequential convex programming (SCP) method [42] to relax the nonconvex rate constraint into a convex one by sequential linearization. Let

$$F(\mathbf{y}) = \frac{1}{N} \sum_{n \in \Psi_y} \log(1 + \frac{A_n x_n}{1 + B_n y_n}). \tag{2.26}$$

By the first order Taylor's series expansion around $\mathbf{y} = \mathbf{y}^{(k)}$, $F(\mathbf{y})$ can be approximated as:

$$\begin{aligned} F_T(\mathbf{y}, \mathbf{y}^{(k)}) &= F(\mathbf{y}^{(k)}) + (\nabla F(\mathbf{y}^{(k)}))^T (\mathbf{y} - \mathbf{y}^{(k)}) \\ &= F(\mathbf{y}^{(k)}) + \frac{1}{N} \sum_{n \in \Psi_y} \phi_n \cdot (y_n - y_n^{(k)}), \end{aligned} \tag{2.27}$$

where $\phi_n = -\frac{B_n}{1+B_n y_n^{(k)}} + \frac{B_n}{1+B_n y_n^{(k)} + A_n x_n}$.

We compute the updated estimate $\mathbf{y}^{(k+1)}$ by the following:

$$\begin{aligned} \mathbf{y}^{(k+1)} = \arg\max_{\mathbf{y}} & \left\{ \frac{1}{N} \sum_{n \in \Psi_y} \left( \log(1 + \frac{A_n x_n}{1 + B_n y_n}) - \log(1 + \frac{C_n x_n}{1 + D_n y_n}) \right) \right\} \\ s.t. \qquad & F_T(\mathbf{y}, \mathbf{y}^{(k)}) \geq \tilde{\mathcal{C}}_{SD}, \\ & \sum_{n \in \Psi_y} y_n \leq P_D, y_n \geq 0, \forall n \in \Psi_y. \end{aligned} \tag{2.28}$$

The Lagrangian function of this problem is:

$$\begin{aligned} \mathcal{L}(\mathbf{y}, \lambda, \boldsymbol{\mu}, \upsilon) = &- \frac{1}{N} \sum_{n \in \Psi_y} (\log(1 + \frac{A_n x_n}{1 + B_n y_n}) - \log(1 + \frac{C_n x_n}{1 + D_n y_n})) \\ &- \boldsymbol{\mu}^T \mathbf{y} + \upsilon(\sum_{n \in \Psi_y} y_n - P_D) + \lambda \left( \tilde{\mathcal{C}}_{SD} - F_T(\mathbf{y}, \mathbf{y}^{(k)}) \right). \end{aligned} \tag{2.29}$$

The KKT conditions of (2.28) are:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial y_n} = -\psi_n(y_n) - \frac{\lambda}{N}\phi_n - \mu_n + \upsilon = 0, \\ y_n \geq 0, \mu_n \geq 0, \mu_n y_n = 0, \quad \forall n \in \Psi_y, \\ \upsilon \geq 0, \sum_{n\in\Psi_y} y_n \leq P_D, \upsilon(\sum_{n\in\Psi_y} y_n - P_D) = 0, \\ \lambda \geq 0, F_T(\mathbf{y}, \mathbf{y}^{(k)}) - \tilde{\mathcal{C}}_{SD} \geq 0 \\ \lambda\left(\tilde{\mathcal{C}}_{SD} - F_T(\mathbf{y}, \mathbf{y}^{(k)})\right) = 0, \end{cases} \tag{2.30}$$

where $\psi_n(y_n)$ is defined in Eq. (2.17). From the first condition of (2.30), one can verify by using propositionosition 5 that $\lambda$ and $\upsilon$ are each monotonic functions of $y_n$ as long as $\psi_n(y_n) > 0$. So, the KKT conditions in (2.30) can be solved by a 2-D bisection algorithm which is similar to algorithm 3 but omitted here. Every new solution of $y_n, \forall n$ needs to be used to update the problem (2.28) until convergence.

## 2.4 Numerical Results

In this section, we present the simulation results based on our proposed algorithms. In the simulation, all channel magnitudes are Rayleigh distributed with unit mean square, and the self-interference attenuation factor $\rho$ is set to be 0.5 unless stated otherwise.

### 2.4.1 With power - only constraints

With $N = 8$ and $P_S = P_D = P$, shown in Fig. 2.2 are four curves of averaged secrecy capacity versus the power $P$. The "UB" means "asymptotical limit at high power", "JA" means "joint optimal power allocation at both source and destination", "DA" means "optimal destination power allocation while uniform source power allocation", "SA" means “optimal source power allocation while uniform destination power allocation”, and “UA” means “uniform power allocation at both source and destination”. We see that in the very low power region, “optimal source power allocation” has an advantage over “optimal destination power allocation”. This is because at low power, the SINR on each subcarrier (see (2.1)) is dominated by the source power and the destination power has little effect.

While in the high power region, “optimal destination power allocation” is much more effective than “optimal source power allocation”. This is because at higher power, the uniform source power allocation approaches its optimal allocation, and hence optimal destination power allocation subject to uniform source power allocation approaches the joint optimality at both source and destination. However, the uniform destination power allocation is generally not optimal at high power. We see indeed that the results for “optimal destination power allocation” and “joint optimal power allocation” achieve the same upper bound at high power. The effect of the optimal destination power allocation at high power is very significant.

Shown in Fig. 2.3 are results for a varying level of self-interference channel magnitude. Clearly, the less the self-interference, the higher secrecy capacity achievable.

The two-phase iterations typically take less than 5 iterations to converge. The bisection search within each of the two phases converges rapidly (exponentially fast) as expected.

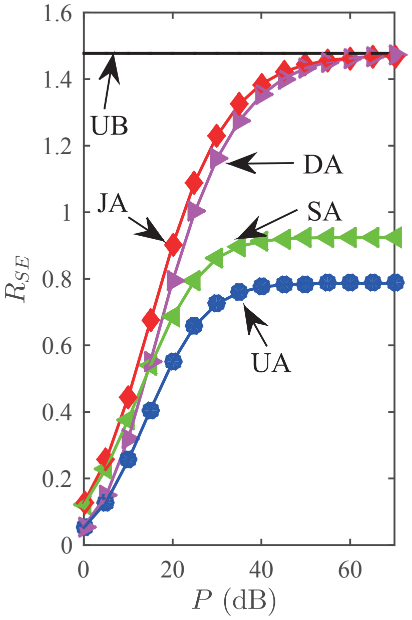

Figure 2.2: Secrecy capacity vs. power budget $P = P_S = P_D$ ($\rho = 0.5$)

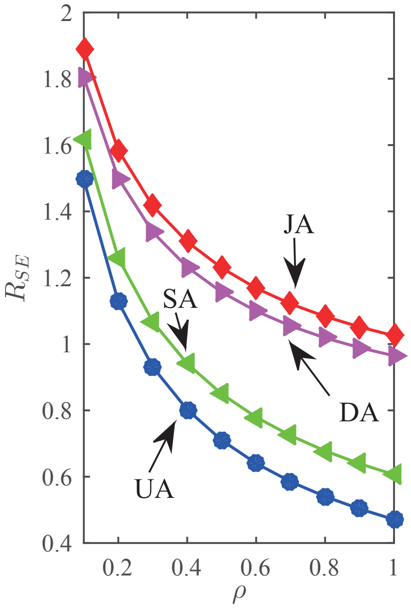

Figure 2.3: The secrecy capacity vs. self-interference attenuation factor $\rho$ ($P_S = P_D =$ 30dB)

### 2.4.2 With power and rate constraints

Shown in Figs. 2.4 and 2.5 is a comparison of three different cases in terms of the secrecy capacity against Eve (Figs. 2.4) and the Alice-to-Bob data rate (Fig. 2.5) for a specific realization of all channels where $A_n < C_n, \forall n \in \mathsf{N}$ (i.e., Eve has a stronger channel from Alice than Bob has from Alice for all subcarriers).

In case I, the data rate is maximized subject to power constraints at Alice and Bob but there is no secrecy capacity constraint. The resulting data rate is denoted by $\mathcal{C}_{SD,I}$ (which is obtained by the standard waterfilling algorithm). And the resulting secrecy capacity $\mathcal{R}_{SE,I}$ is zero for this channel realization as expected.

In case II, the secrecy capacity is maximized subject to power constraints at Alice and Bob and also a Alice-to-Bob rate constraint. The constrained rate (i.e., the lower bound on the rate) is set at $\mathcal{C}_{SD}^{\dagger} = 0.9\mathcal{C}_{SD,I}$. The corresponding achieved rate is denoted by $\mathcal{C}_{SD,II}$, the curve of which is, as expected, indistinguishable from that of $\mathcal{C}_{SD}^{\dagger}$. The resulting secrecy capacity is denoted by $\mathcal{R}_{SE,II}$, which is large and not far from that of case III.

In case III, the secrecy capacity is maximized with power-only constraints at Alice and Bob but no rate constraint. The resulting secrecy capacity is denoted by $\mathcal{R}_{SE,III}$ and the resulting data rate is $\mathcal{C}_{SD,III}$.

We see that because of the rate constraint, case II results in a much better tradeoff between the source-to-destination data rate and the network's secrecy capacity than the other two cases.

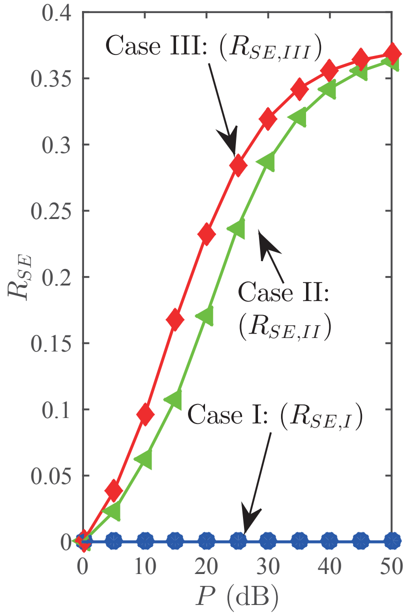

Figure 2.4: The achieved secrecy capacity under power and rate constraints ($\mathcal{C}^{\dagger}_{SD} = 0.9\mathcal{C}_{SD,I}$

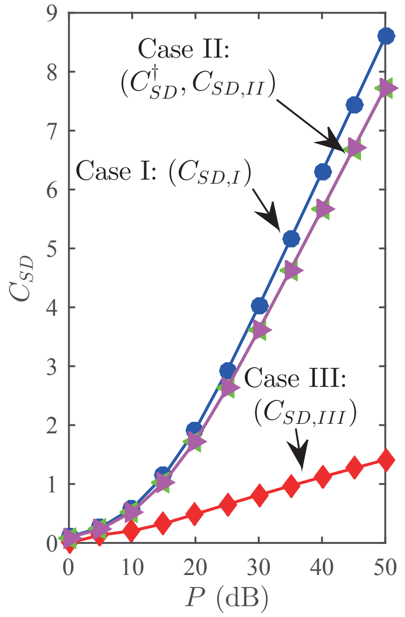

Figure 2.5: The achieved transmission rate between Alice and Bob when the secrecy rate is maximized.

## 2.5 Conclusion

In this chapter, we have studied fast power allocation algorithms for maximizing secrecy capacity of a three-node network subject to both power and rate constraints. The rate constraint along with self-interference of the full-duplex destination makes this study unique from many previous works.

# Chapter 3

# Secure Degree of Freedom Analysis for MIMOME Network with Anti-Eavesdropper Channel Estimation

## 3.1 Introduction

First we introduce the anti-eavesdropping channel estimation (ANECE) scheme that proposed by [31]. Consider a wireless network with all the legitimate users equipped with full duplex radio. In the channel estimation period, all the legitimate users will exchange pilot signals (which is known to everyone) concurrently. The pilots are such that they excite all dimensions of the receive CSI for each user but leave a subspace of Eve's

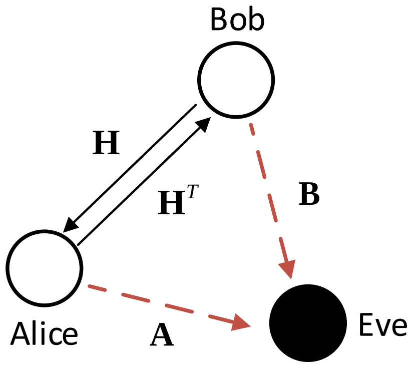

Figure 3.1: A three-node system with two legitimate full-duplex transceivers and one passive eavesdropper.

receive CSI unexcited. In other words, the composite pilot matrix seen by any user in such that allows consistent estimation of the receive CSI at this user, but the composite pilot matrix seen by Eve has a rank deficiency that makes a subspace of Eve's receive CSI unobservable by Eve. With such pilots, Eve will be unable to correctly estimate its own channels and therefore can not successfully decode the subsequent information packet.

In this chapter, we will show the analysis of secure degree of freedom of one-way secure transmission and two-way secure transmission subjected to using the ANECE style pilots.

## 3.2 Secure Degree of Freedom Analyses

Consider a block Rayleigh fading channel for which Alice and Bob first conduct ANECE by transmitting their pilot signals $\mathbf{p}_A(k)$ and $\mathbf{p}_B(k)$ concurrently (in full-duplex mode) where $k = 1, \cdots, K_1$ ($K_1$ is the length of the pilot), and then transmit information to each other (over $K_2$ samples). The system is shown in Fig. 3.1. For information transmission, we will consider a one-way transmission and a two-way transmission separately.

### 3.2.1 Channel estimation

Define $\mathbf{P}_i = [\mathbf{p}_i(1), \cdots, \mathbf{p}_i(K_1)]$ where $i = A, B$. then the corresponding signals received by Alice, Bob and Eve can be expressed as

$$\mathbf{Y}_A = \mathbf{H}^T \mathbf{P}_B + \mathbf{N}_A \tag{3.1a}$$

$$\mathbf{Y}_B = \mathbf{H}\mathbf{P}_A + \mathbf{N}_B \tag{3.1b}$$

$$\mathbf{Y}_E = \mathbf{A}\mathbf{P}_A + \mathbf{B}\mathbf{P}_B + \mathbf{N}_E \tag{3.1c}$$

where $\mathbf{H}$ is the reciprocal channel matrix between Alice and Bob, and all the noise matrices consist of i.i.d. $\mathcal{CN}(0,1)$. Here, the self-interferences at Alice and Bob are assumed to be negligible.

It is known and easy to show that for the best performance of the maximum likelihood (ML) estimation (or the MMSE estimation as shown later) of $\mathbf{H}$ by Bob, $\mathbf{P}_A$ should be such that $\mathbf{P}_A\mathbf{P}_A^H = \frac{K_1 P_A}{N_A}\mathbf{I}_{N_A}$. Similarly, $\mathbf{P}_B$ should be such that $\mathbf{P}_B\mathbf{P}_B^H = \frac{K_1 P_B}{N_B}\mathbf{I}_{N_B}$.

In the following analyses, we assume that $\mathbf{H}$, $\mathbf{A}$ and $\mathbf{B}$ all consist of i.i.d. zero mean complex Gaussian elements with variance 1, $a$ and $b$ respectively (from one coherence block to another). $a$ and $b$ is assumed to be known to everyone.

Without loss of generality, let $N_A \geq N_B$. Without affecting the channel estimation performance at Alice and Bob, but maximizing the difficulty of channel estimation for Eve, we let the row span of $\mathbf{P}_B$ be part of the row span of $\mathbf{P}_A$. More specifically, we can write $\mathbf{P}_A = \sqrt{\frac{K_1 P_A}{N_A}}[\mathbf{I}_{N_A}, \mathbf{0}_{N_A \times (K_1 - N_A)}]\mathbf{\Gamma}$ and $\mathbf{P}_B = \sqrt{\frac{K_1 P_B}{N_B}}[\mathbf{I}_{N_B}, \mathbf{0}_{N_B \times (K_1 - N_B)}]\mathbf{\Gamma}$ where $\mathbf{\Gamma}$ can be any $K_1 \times K_1$ unitary matrix. In this way, any estimates of $\mathbf{A}$ and $\mathbf{B}$ by Eve, denoted by $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$, are ambiguous in that $[\sqrt{a}\hat{\mathbf{A}}, \sqrt{b}\hat{\mathbf{B}}]$ can be added to $\mathbf{\Theta}[\mathbf{C}_A, \mathbf{C}_B]$ without affecting Eve's observation $\mathbf{Y}_E$ where $\mathbf{\Theta} \in \mathbb{C}^{N_E \times N_B}$ is arbitrary and $[\mathbf{C}_A, \mathbf{C}_B][\mathbf{P}_A^T, \mathbf{P}_B^T]^T = 0$.

Let $\mathbf{h} = vec(\mathbf{H})$, $\mathbf{a} = vec(\mathbf{A})$, $\mathbf{b} = vec(\mathbf{B})$, $\mathbf{y}_A = vec(\mathbf{Y}_A^T)$, $\mathbf{y}_B = vec(\mathbf{Y}_B)$, $\mathbf{n}_A = vec(\mathbf{N}_A^T)$ and $\mathbf{n}_B = vec(\mathbf{N}_B)$. Note $vec(\mathbf{XYZ}) = (\mathbf{Z}^T \otimes \mathbf{X})vec(\mathbf{Y})$. Then (3.1) becomes

$$\mathbf{y}_A = (\mathbf{I}_{N_A} \otimes \mathbf{P}_B^T)\mathbf{h} + \mathbf{n}_A \tag{3.2a}$$

$$\mathbf{y}_B = (\mathbf{P}_A^T \otimes \mathbf{I}_{N_B})\mathbf{h} + \mathbf{n}_B \tag{3.2b}$$

$$\mathbf{y}_E = (\mathbf{P}_A^T \otimes \mathbf{I}_{N_E})\mathbf{a} + (\mathbf{P}_B^T \otimes \mathbf{I}_{N_E})\mathbf{b} + \mathbf{n}_E. \tag{3.2c}$$

It is known that the minimum-mean-squared-error (MMSE) estimate of a vector $\mathbf{x}$ from another vector $\mathbf{y}$ is $\hat{\mathbf{x}} = \mathbf{K}_{\mathbf{x},\mathbf{y}}\mathbf{K}_{\mathbf{y}}^{-1}\mathbf{y}$ with $\mathbf{K}_{\mathbf{x},\mathbf{y}} = \mathcal{E}\{\mathbf{x}\mathbf{y}^H\}$ and $\mathbf{K}_{\mathbf{y}} = \mathcal{E}\{\mathbf{y}\mathbf{y}^H\}$. And the error $\Delta\mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}$ has the covariance matrix $\mathbf{K}_{\Delta\mathbf{x}} = \mathbf{K}_{\mathbf{x}} - \mathbf{K}_{\mathbf{x},\mathbf{y}}\mathbf{K}_{\mathbf{y}}^{-1}\mathbf{K}_{\mathbf{x},\mathbf{y}}^H$.

Let $\hat{\mathbf{h}}_A$ be the MMSE estimate of $\mathbf{h}$ by Alice, and $\Delta\mathbf{h}_A = \mathbf{h} - \hat{\mathbf{h}}_A$ be its error. Similar notations are defined for Bob and Eve. It is easy to show that the covariance matrices of the errors of these estimates are, respectively, $\mathbf{K}_{\Delta\mathbf{h}_A} = \sigma_A^2\mathbf{I}_{N_A N_B}$, $\mathbf{K}_{\Delta\mathbf{h}_B} = \sigma_B^2\mathbf{I}_{N_A N_B}$, $\mathbf{K}_{\Delta\mathbf{a}} = \sigma_{EA}^2\mathbf{I}_{N_A N_E}$ and $\mathbf{K}_{\Delta\mathbf{b}} = \sigma_{EB}^2\mathbf{I}_{N_B N_E}$ where $\sigma_A^2 = \frac{1}{1+K_1 P_B/N_B}$, $\sigma_B^2 = \frac{1}{1+K_1 P_A/N_A}$, $\sigma_{EA}^2 = \frac{bK_1 P_B/N_B + 1}{(aK_1 P_A/N_A + bK_1 P_B/N_B)+1}$ and $\sigma_{EB}^2 = \frac{aK_1 P_A/N_A + 1}{(aK_1 P_A/N_A + bK_1 P_B/N_B)+1}$.

### 3.2.2 One-way information transmission

Now assume that following the pilots (over $K_1$ samples) transmitted by Alice and Bob in full-duplex mode, Alice transmits information (over $K_2$ samples) to Bob in half-duplex mode. Namely, while the first phase is in full-duplex, the second phase is in half-duplex. In the second phase, Bob and Eve receive

$$\begin{aligned} \mathbf{Y}_B &= \mathbf{H}\mathbf{S}_A + \mathbf{N}_B \\ \mathbf{Y}_E &= \mathbf{A}\mathbf{S}_A + \mathbf{N}_E \end{aligned} \tag{3.3}$$

where $\mathbf{S}_A = [\mathbf{s}_A(1), \ldots, \mathbf{s}_A(K_2)]$. The corresponding vector forms of the above are

$$\mathbf{y}_B = (\mathbf{I}_{K_2} \otimes \mathbf{H})\bar{\mathbf{s}}_A + \mathbf{n}_B \tag{3.4a}$$

$$\mathbf{y}_E = (\mathbf{I}_{K_2} \otimes \mathbf{A})\bar{\mathbf{s}}_A + \mathbf{n}_E \tag{3.4b}$$

where $\bar{\mathbf{s}}_A = vec(\mathbf{S}_A)$ (which is assumed to be independent of all channel parameters). Then an achievable secrecy rate in bits/s/Hz in phase 2 from Alice to Bob (conditional on the MMSE channel estimation in phase 1) is

$$\mathcal{R}_{one} = \frac{1}{K_2}\big(I(\bar{\mathbf{s}}_A; \mathbf{y}_B|\hat{\mathbf{h}}_B) - I(\bar{\mathbf{s}}_A; \mathbf{y}_E|\hat{\mathbf{a}})\big)^+ \tag{3.5}$$

To analyze $\mathcal{R}_{one}$, we now assume $P_A = P_B = P$ (which holds for both phases 1 and 2) and that $\mathbf{s}_A(k)$ are i.i.d. with $\mathcal{CN}(0, \frac{P_A}{N_A}\mathbf{I}_{N_A})$. We also use $\hat{\mathbf{H}}_B = ivec(\hat{\mathbf{h}}_B) \in \mathbb{C}^{N_B \times N_A}$ (i.e., $\hat{\mathbf{h}}_B = vec(\hat{\mathbf{H}}_B)$).

We will next derive lower and upper bounds on $\mathcal{R}_{one}$. To do that, we need to obtain lower and upper bounds on $I(\bar{\mathbf{s}}_A; \mathbf{y}_B|\hat{\mathbf{h}}_B)$ and those on $I(\bar{\mathbf{s}}_A; \mathbf{y}_E|\hat{\mathbf{a}})$.

First, we have

$$\begin{aligned} I(\bar{\mathbf{s}}_A;\mathbf{y}_B|\hat{\mathbf{h}}_B) &= h(\bar{\mathbf{s}}_A|\hat{\mathbf{h}}_B) - h(\bar{\mathbf{s}}_A|\mathbf{y}_B,\hat{\mathbf{h}}_B) \\ &= h(\bar{\mathbf{s}}_A) - h(\bar{\mathbf{s}}_A|\mathbf{y}_B,\hat{\mathbf{h}}_B). \end{aligned} \tag{3.6}$$

It is known that $h(\bar{\mathbf{s}}_A) = \log\left[(\pi e)^{N_A K_2}\left|\frac{P_A}{N_A}\mathbf{I}_{N_A K_2}\right|\right]$. It is also known [43] that for a random vector $\mathbf{s}\in\mathbb{C}^{n\times 1}$ and another random vector $\mathbf{w}$, $h(\mathbf{s}|\mathbf{w}) \le \log\left[(\pi e)^n|\mathbf{K}_{\mathbf{s}|\mathbf{w}}|\right]$ where $\mathbf{K}_{\mathbf{s}|\mathbf{w}} = \mathbf{K}_{\mathbf{s}} - \mathbf{K}_{\mathbf{s},\mathbf{w}}(\mathbf{K}_{\mathbf{w}})^{-1}\mathbf{K}_{\mathbf{s},\mathbf{w}}$ which is the covariance matrix of the MMSE estimation of $\mathbf{s}$ from $\mathbf{w}$. Note that $\mathbf{y}_B = (\mathbf{I}_{K_2}\otimes\hat{\mathbf{H}}_B)\bar{\mathbf{s}}_A + (\mathbf{I}_{K_2}\otimes\Delta\mathbf{H}_B)\bar{\mathbf{s}}_A + \mathbf{n}_B$. Then conditional on $\hat{\mathbf{H}}_B$ (which is independent of $\bar{\mathbf{s}}_A$), the covariance matrix of the MMSE estimate of $\bar{\mathbf{s}}_A$ from $\mathbf{y}_B$ is $\mathbf{K}_{\bar{\mathbf{s}}_A|\mathbf{y}_B,\hat{\mathbf{h}}_B} = \frac{P_A}{N_A}\mathbf{I}_{N_A K_2} - \frac{P_A^2}{N_A^2}(\mathbf{I}_{K_2}\otimes\hat{\mathbf{H}}_B^H)(\frac{P_A}{N_A}(\mathbf{I}_{K_2}\otimes\hat{\mathbf{H}}_B\hat{\mathbf{H}}_B^H)+\mathbf{K}_B+\mathbf{I}_{N_B K_2})^{-1}(\mathbf{I}_{K_2}\otimes\hat{\mathbf{H}}_B)$ where $\mathbf{K}_B = \mathcal{E}\{(\mathbf{I}_{K_2}\otimes\Delta\mathbf{H}_B)\bar{\mathbf{s}}_A\bar{\mathbf{s}}_A^H(\mathbf{I}\otimes\Delta\mathbf{H}_B^H)\} = \frac{P_A}{1+K_1 P_A/N_A}\mathbf{I}_{N_B K_2}$. Using $|\mathbf{I}_{r_A}+\mathbf{AB}| = |\mathbf{I}_{r_B}+\mathbf{BA}|$ where $r_A$ and $r_B$ are the numbers of rows of $\mathbf{A}$ and $\mathbf{B}$ respectively, one can verify that $\log|\mathbf{K}_{\bar{\mathbf{s}}_A|\mathbf{y}_B,\hat{\mathbf{h}}_B}| = N_A K_2\log\frac{P_A}{N_A} + \log|\mathbf{K}_B+\mathbf{I}_{N_B K_2}| - \log|\frac{P_A}{N_A}(\mathbf{I}_{K_2}\otimes\hat{\mathbf{H}}_B\hat{\mathbf{H}}_B^H)+\mathbf{K}_B+\mathbf{I}_{N_B K_2}| = N_A K_2\log\frac{P_A}{N_A} - K_2\log|\mathbf{I}_{N_B} + \frac{P_A/N_A}{1+\frac{P_A}{1+K_1 P_A/N_A}}\hat{\mathbf{H}}_B\hat{\mathbf{H}}_B^H|$. Applying the above results to (3.6) yields

$$\begin{aligned} & I(\bar{\mathbf{s}}_A;\mathbf{y}_B|\hat{\mathbf{h}}_B) \\ & \ge \log|\frac{P_A}{N_A}\mathbf{I}_{N_A K_2}| - \mathcal{E}\{\log|\mathbf{K}_{\bar{\mathbf{s}}_A|\mathbf{y}_B,\hat{\mathbf{h}}_B}|\} \\ & = K_2\mathcal{E}\{\log|\mathbf{I}_{N_B} + \frac{P_A/N_A}{1+\frac{P_A}{1+K_1 P_A/N_A}}\hat{\mathbf{H}}_B\hat{\mathbf{H}}_B^H|\} \\ & \triangleq \mathcal{R}_B^-. \end{aligned} \tag{3.7}$$

To derive an upper bound on $I(\bar{\mathbf{s}}_A;\mathbf{y}_B|\hat{\mathbf{h}}_B)$, we now write

$$I(\bar{\mathbf{s}}_A;\mathbf{y}_B|\hat{\mathbf{h}}_B) = h(\mathbf{y}_B|\hat{\mathbf{h}}_B) - h(\mathbf{y}_B|\hat{\mathbf{h}}_B,\bar{\mathbf{s}}_A). \tag{3.8}$$

Here, we have

$$h(\mathbf{y}_B|\hat{\mathbf{h}}_B) \leq \mathcal{E}\{\log[(\pi e)^{N_B K_2}|\frac{P_A}{N_A}(\mathbf{I}_{K_2} \otimes \hat{\mathbf{H}}_B \hat{\mathbf{H}}_B^H) + \mathbf{K}_B + \mathbf{I}_{N_B K_2}|]\}$$

$$= K_2 \mathcal{E}\{\log[(\pi e)^{N_B}|\frac{P_A}{N_A}(\hat{\mathbf{H}}_B \hat{\mathbf{H}}_B^H) + (1 + \frac{P_A}{1 + K_1 P_A/N_A})\mathbf{I}_{N_B}|]\} \tag{3.9}$$

and

$$h(\mathbf{y}_B|\hat{\mathbf{h}}_B, \bar{\mathbf{s}}_A) = \mathcal{E}\{\log[(\pi e)^{N_B K_2}|\frac{1}{1 + K_1 P_A/N_A}(\mathbf{S}_A^T \mathbf{S}_A^* \otimes \mathbf{I}_{N_B}) + \mathbf{I}_{N_B K_2}|]\}$$

$$= N_B \mathcal{E}\{\log[(\pi e)^{K_2}|\frac{1}{1 + K_1 P_A/N_A}(\mathbf{S}_A^T \mathbf{S}_A^*) + \mathbf{I}_{K_2}|]\}. \tag{3.10}$$

Note that conditional on $\hat{\mathbf{h}}_B$ and $\bar{\mathbf{s}}_A$ the covariance matrix of $\mathbf{y}_B$ is invariant to $\hat{\mathbf{h}}_B$. Now define

$$\mathbf{M}_A = \begin{cases} \frac{N_A}{P_A}\mathbf{S}_A^T \mathbf{S}_A^*, & K_2 < N_A \\ \frac{N_A}{P_A}\mathbf{S}_A^* \mathbf{S}_A^T, & K_2 \geq N_A \end{cases} \tag{3.11}$$

which is a full rank matrix for any $N_A$ and $K_2$ and a self-product of $\sqrt{\frac{N_A}{P_A}}\mathbf{S}_A$ with i.i.d. $\mathcal{CN}(0,1)$ entries. Also define $t_A = \min\{N_A, K_2\}$ and $r_A = \max\{N_A, K_2\}$. It follows that (as part of $h(\mathbf{y}_B|\hat{\mathbf{h}}_B, \bar{\mathbf{s}}_A)$)

$$\mathcal{E}\{\log|\frac{1}{1 + K_1 P_A/N_A}(\mathbf{S}_A^T \mathbf{S}_A^*) + \mathbf{I}_{K_2}|\}$$

$$= \mathcal{E}\{\log|\frac{P_A/N_A}{1 + K_1 P_A/N_A}\mathbf{M}_A + \mathbf{I}_{t_A}|\}$$

$$\geq t_A \mathcal{E}\{\log(1 + |\frac{P_A/N_A}{1 + K_1 P_A/N_A}\mathbf{M}_A|^{\frac{1}{t_A}})\} \tag{3.12a}$$

$$= t_A \mathcal{E}\big\{\log(1 + \frac{P_A/N_A}{1 + K_1 P_A/N_A}\exp(\frac{1}{t_A}\ln|\mathbf{M}_A|))\big\}$$

$$\geq t_A \log\big(1 + \frac{P_A/N_A}{1 + K_1 P_A/N_A}\exp(\frac{1}{t_A}\mathcal{E}\{\ln|\mathbf{M}_A|\})\big) \tag{3.12b}$$

$$= t_A \log\big(1 + \frac{P_A/N_A}{1 + K_1 P_A/N_A}\exp(\frac{1}{t_A}\sum_{j=1}^{t_A}\sum_{k=1}^{r_A - j}\frac{1}{k} - \gamma)\big) \tag{3.12c}$$

where (3.12a) is due to the matrix Minkowski's inequality $|\mathbf{X}+\mathbf{Y}|^{1/n} \geq |\mathbf{X}|^{1/n}+|\mathbf{Y}|^{1/n}$ where $\mathbf{X}$ and $\mathbf{Y}$ are $n \times n$ positive definite matrices [44], (3.12b) is due to the Jensen's inequality and that $\log(1+ae^x)$ is a convex function of $x$ when $a > 0$, and (3.12c) is based on [45, Th.1] where $\gamma \approxeq 0.57721566$ is Euler's constant. Defining $e_A = \exp(\frac{1}{t_A}\sum_{j=1}^{t_A}\sum_{k=1}^{r_A-j}\frac{1}{k} - \gamma)$ and applying the above results since (3.8), we have from (3.8) that

$$\begin{aligned} &I(\bar{\mathbf{s}}_A;\mathbf{y}_B|\hat{\mathbf{h}}_B) \\ &\leq K_2\mathcal{E}\{\log|\mathbf{I}_{N_B} + \frac{P_A/N_A\hat{\mathbf{H}}_B\hat{\mathbf{H}}_B^H}{1+\frac{P_A}{1+K_1P_A/N_A}}|\} + N_B\log\left(\frac{(1+\frac{P_A}{1+K_1P_A/N_A})^{K_2}}{\left(1+\frac{P_A/N_A}{1+K_1P_A/N_A}e_A\right)^{t_A}}\right) \\ &\triangleq \mathcal{R}_B^+ \end{aligned} \tag{3.13}$$

From (3.7) and (3.13) we see that the difference between the upper and lower bounds on $I(\bar{\mathbf{s}}_A;\mathbf{y}_B|\hat{\mathbf{h}}_B)$ is the second term in (3.13).

To consider $I(\bar{\mathbf{s}}_A;\mathbf{y}_E|\hat{\mathbf{a}})$ in (3.5), we let $\hat{\mathbf{A}} = ivec(\hat{\mathbf{a}})$. Similar to the discussions leading to (3.7) and (3.13), one can verify that

$$I(\bar{\mathbf{s}}_A;\mathbf{y}_E|\hat{\mathbf{a}}) \geq K_2\mathcal{E}\{\log|\mathbf{I}_{N_E} + \frac{P_A/N_A\hat{\mathbf{A}}\hat{\mathbf{A}}^H}{1+P_A\sigma_{EA}^2}|\} \triangleq \mathcal{R}_E^- \tag{3.14}$$

and

$$\begin{aligned} &I(\bar{\mathbf{s}}_A;\mathbf{y}_E|\hat{\mathbf{a}}) \\ &\leq \mathcal{R}_E^- + N_E\log\left(\frac{(1+P_A\sigma_{EA}^2)^{K_2}}{\left(1+(P_A\sigma_{EA}^2/N_A)e_A\right)^{t_A}}\right) \\ &\triangleq \mathcal{R}_E^+ \end{aligned} \tag{3.15}$$

When $P_A = P_B = P \to \infty$, we have $\sigma_{EA}^2 \to \frac{bN_A}{aN_B+bN_A}$, $\sigma_B^2 \to 0$, $\mathcal{E}\{\hat{a}_i\hat{a}_i^*\} \to \frac{aN_B}{aN_B+bN_A}$ and $\mathcal{E}\{\hat{h}_{B,i}\hat{h}_{B,i}^*\} \to 1$. From [46, Th.2], we know that $\mathcal{E}\{\log|\mathbf{I}_r + \frac{P}{t}\mathbf{X}\mathbf{X}^H|\} \to \min(r,t)\log P + o(\log P)$ as $P \to \infty$ where the entries of $\mathbf{X} \in \mathbb{C}^{r\times t}$ are i.i.d. $\mathcal{CN}(0,1)$.

Therefore, from (3.7) and (3.13),

$$\lim_{P\to\infty}\frac{\mathcal{R}_B^-}{\log P}=\lim_{P\to\infty}\frac{\mathcal{R}_B^+}{\log P}=K_2\min\{N_A,N_B\} \tag{3.16}$$

And from (3.14) and (3.15), we have

$$\lim_{P\to\infty}\frac{\mathcal{R}_E^-}{\log P}=0 \tag{3.17}$$

and

$$\lim_{P\to\infty}\frac{\mathcal{R}_E^+}{\log P}=\begin{cases}0, & K_2\leq N_A\\ \\ N_E(K_2-N_A), & K_2>N_A\end{cases} \tag{3.18}$$

Combining (3.16), (3.17) and (3.18) and using $\mathcal{R}_{one}^+\triangleq\frac{1}{K_2}[\mathcal{R}_B^+-\mathcal{R}_E^-]^+$ and $\mathcal{R}_{one}^-\triangleq\frac{1}{K_2}[\mathcal{R}_B^- -\mathcal{R}_E^+]^+$ (i.e., $\mathcal{R}_{one}^-\leq\mathcal{R}_{one}\leq\mathcal{R}_{one}^+$), we have

$$\begin{aligned}&\lim_{P\to\infty}\frac{\mathcal{R}_{one}^-}{\log P}\\ &=\begin{cases}\min\{N_A,N_B\}, & K_2\leq N_A\\ \left(\min\{N_A,N_B\}-\dfrac{N_E}{K_2}(K_2-N_A)\right)^+, & K_2>N_A\end{cases}\end{aligned} \tag{3.19}$$

and

$$\lim_{P\to\infty}\frac{\mathcal{R}_{one}^+}{\log P}=\min\{N_A,N_B\}. \tag{3.20}$$

Note that $\lim_{P\to\infty}\frac{\mathcal{R}_{one}}{\log P}$ is called the secure degrees of freedom of the one-way information transmission. From (3.19) and (3.20), we see that when $K_2\leq N_A$, we have $\lim_{P\to\infty}\frac{\mathcal{R}_{one}}{\log P}=\min\{N_A,N_B\}$ which equals the degrees of freedom of the main channel capacity from Alice to Bob. This supports and complements a conclusion from [31] where the analyses did not use the complete statistical model of $\mathbf{H}$, $\mathbf{A}$ and $\mathbf{B}$. We also see from (3.19) that if $K_2>N_A$, the above lower bound on secure degrees of freedom decreases linearly as $N_E$ increases.

### 3.2.3 Two-way information transmission

Now we consider a two-way (full-duplex) communication in the second phase where the signals received by Alice, Bob and Eve in a coherence period are

$$\begin{aligned} \mathbf{Y}_A &= \mathbf{H}^T\mathbf{S}_B + \mathbf{N}_A \\ \mathbf{Y}_B &= \mathbf{H}\mathbf{S}_A + \mathbf{N}_B \\ \mathbf{Y}_E &= \mathbf{A}\mathbf{S}_A + \mathbf{B}\mathbf{S}_B + \mathbf{N}_E \end{aligned} \tag{3.21}$$

where $\mathbf{S}_A = [\mathbf{s}_A(1), \ldots, \mathbf{s}_A(K_2)]$ and $\mathbf{s}_A(t) \sim \mathcal{CN}(0, \frac{P_A}{N_A}\mathbf{I})$. Similarly $\mathbf{S}_B = [\mathbf{s}_B(1), \ldots, \mathbf{s}_B(K_2)]$ and $\mathbf{s}_B(t) \sim \mathcal{CN}(0, \frac{P_B}{N_B}\mathbf{I})$. Note that all information symbols from Alice and Bob are i.i.d.. The vectorized forms of (3.21) are

$$\begin{aligned} \mathbf{y}_A &= (\mathbf{I}_{K_2} \otimes \mathbf{H}^T)\bar{\mathbf{s}}_B + \mathbf{n}_A \\ \mathbf{y}_B &= (\mathbf{I}_{K_2} \otimes \mathbf{H})\bar{\mathbf{s}}_A + \mathbf{n}_B \\ \mathbf{y}_E &= (\mathbf{I}_{K_2} \otimes \mathbf{A})\bar{\mathbf{s}}_A + (\mathbf{I}_{K_2} \otimes \mathbf{B})\bar{\mathbf{s}}_B + \mathbf{n}_E \end{aligned} \tag{3.22}$$

where both $\bar{\mathbf{s}}_A$ and $\bar{\mathbf{s}}_B$ are assumed to be independent of all channel parameters. Conditional on the MMSE channel estimation in phase 1, an achievable secrecy rate in phase 2 by the two-way wiretap channel is (e.g., see [47]):

$$\begin{aligned} \mathcal{R}_{two} =& \frac{1}{K_2}\big(I(\bar{\mathbf{s}}_B; \mathbf{y}_A|\hat{\mathbf{h}}_A) + I(\bar{\mathbf{s}}_A; \mathbf{y}_B|\hat{\mathbf{h}}_B) \\ & - I(\bar{\mathbf{s}}_A, \bar{\mathbf{s}}_B; \mathbf{y}_E|\hat{\mathbf{a}}, \hat{\mathbf{b}})\big)^+ \end{aligned} \tag{3.23}$$

The following analyses is similar to the previous section, for which we will only provide the key steps and results.

From (3.7) and (3.13), we already know a pair of lower and upper bounds on $I(\bar{\mathbf{s}}_A; \mathbf{y}_B|\hat{\mathbf{h}}_B)$. To show a similar pair of lower and upper bounds on $I(\bar{\mathbf{s}}_B; \mathbf{y}_A|\hat{\mathbf{h}}_A)$, we let $\hat{\mathbf{H}}_A = ivec(\hat{\mathbf{h}}_A)$. One can verify that

$$I(\bar{\mathbf{s}}_B; \mathbf{y}_A|\hat{\mathbf{h}}_A) \geq K_2 \mathcal{E}\{\log |\mathbf{I}_{N_A} + \frac{P_B/N_B}{1 + \frac{\sigma^2 P_B}{1+\sigma^2 T_1 P_B/N_B}} \hat{\mathbf{H}}_A^T \hat{\mathbf{H}}_A^*|\} \triangleq \mathcal{R}_A^- \tag{3.24}$$

and

$$\begin{aligned} I(\bar{\mathbf{s}}_B; \mathbf{y}_A|\hat{\mathbf{h}}_A) &\leq \mathcal{R}_A^- + N_A \log \left( \frac{(1 + \frac{P_B}{1+K_1 P_B/N_B})^{K_2}}{(1 + \frac{P_B/N_B}{1+K_1 P_B/N_B} e_B)^{t_B}} \right) \\ &\triangleq \mathcal{R}_A^+ \end{aligned} \tag{3.25}$$

where $e_B = \exp(\frac{1}{t_B} \sum_{j=1}^{t_B} \sum_{k=1}^{r_B - j} \frac{1}{k} - \gamma)$, $t_B = \min\{N_B, K_2\}$ and $r_B = \max\{N_B, K_2\}$.

For $I(\bar{\mathbf{s}}_A, \bar{\mathbf{s}}_B; \mathbf{y}_E|\hat{\mathbf{a}}, \hat{\mathbf{b}})$, we use $\hat{\mathbf{B}} = ivec(\hat{\mathbf{b}})$ (similar to $\hat{\mathbf{A}}$). One can verify that $\mathbf{K}_{\mathbf{y}_E|\hat{\mathbf{a}},\hat{\mathbf{b}}} = \frac{P_A}{N_A}(\mathbf{I}_{K_2} \otimes \hat{\mathbf{A}}\hat{\mathbf{A}}^H) + \frac{P_B}{N_B}(\mathbf{I}_{K_2} \otimes \hat{\mathbf{B}}\hat{\mathbf{B}}^H) + \mathbf{K}_{EA} + \mathbf{K}_{EB} + \mathbf{I}_{N_E K_2}$ where $\mathbf{K}_{EA} = \mathcal{E}\{(\mathbf{I}_{K_2} \otimes \Delta\mathbf{A})\bar{\mathbf{s}}_A \bar{\mathbf{s}}_A^H (\mathbf{I}_{K_2} \otimes \Delta\mathbf{A})^H\} = \sigma_{EA}^2 P_A \mathbf{I}_{N_E K_2}$ and $\mathbf{K}_{EB} = \mathcal{E}\{(\mathbf{I}_{K_2} \otimes \Delta\mathbf{B})\bar{\mathbf{s}}_B \bar{\mathbf{s}}_B^H (\mathbf{I}_{K_2} \otimes \Delta\mathbf{B})^H\} = \sigma_{EB}^2 P_B \mathbf{I}_{N_E K_2}$. Also note that $\mathbf{y}_E = (\mathbf{S}_A^T \otimes \mathbf{I}_{N_E})\mathbf{h}_{EA} + (\mathbf{S}_B^T \otimes \mathbf{I}_{N_E})\mathbf{h}_{EB} + \mathbf{n}_E$. Then,

$$\begin{aligned} &I(\bar{\mathbf{s}}_A, \bar{\mathbf{s}}_B; \mathbf{y}_E|\hat{\mathbf{a}}, \hat{\mathbf{b}}) \\ &= h(\mathbf{y}_E|\hat{\mathbf{a}}, \hat{\mathbf{b}}) - h(\mathbf{y}_E|\hat{\mathbf{a}}, \hat{\mathbf{b}}, \bar{\mathbf{s}}_A, \bar{\mathbf{s}}_B) \\ &\leq \mathcal{E}\{\log[(\pi e)^{K_2 N_E} |\mathbf{K}_{\mathbf{y}_E|\hat{\mathbf{a}},\hat{\mathbf{b}}}|]\} - h(\mathbf{y}_E|\hat{\mathbf{a}}, \hat{\mathbf{b}}, \bar{\mathbf{s}}_A, \bar{\mathbf{s}}_B) \\ &= \mathcal{E}\{\log |\mathbf{K}_{\mathbf{y}_E|\hat{\mathbf{a}},\hat{\mathbf{b}}}|\} - \mathcal{E}\{\log |\sigma_{EA}^2 (\mathbf{S}_A^T \mathbf{S}_A^* \otimes \mathbf{I}_{N_E}) \\ &\quad + \sigma_{EB}^2 (\mathbf{S}_B^T \mathbf{S}_B^* \otimes \mathbf{I}_{N_E}) + \mathbf{I}_{N_E K_2}|\} \\ &= K_2 \mathcal{E}\{\log |\frac{P_A}{N_A} \hat{\mathbf{A}}\hat{\mathbf{A}}^H + \frac{P_B}{N_B} \hat{\mathbf{B}}\hat{\mathbf{B}}^H + (1 + P_A \sigma_{EA}^2 + P_B \sigma_{EB}^2)\mathbf{I}_{N_E}|\} \\ &\quad - N_E \mathcal{E}\{\log |\sigma_{EA}^2 \mathbf{S}_A^T \mathbf{S}_A^* + \sigma_{EB}^2 \mathbf{S}_B^T \mathbf{S}_B^* + \mathbf{I}_{K_2}|\} \end{aligned} \tag{3.26}$$

Define $\mathbf{S}_{AB} = [\check{\mathbf{S}}_A^T, \check{\mathbf{S}}_B^T] \in \mathbb{C}^{K_2 \times (N_A+N_B)}$ where $\mathbf{S}_A = \frac{P_A}{N_A}\check{\mathbf{S}}_A$ and $\mathbf{S}_B = \frac{P_B}{N_B}\check{\mathbf{S}}_B$. Define $\mathbf{T} = diag\{\sigma_{EA}^2 \frac{P_A}{N_A}\mathbf{I}_{N_A}, \sigma_{EB}^2 \frac{P_B}{N_B}\mathbf{I}_{N_B}\}$. Then we can rewrite the last term from (3.26) as $\mathcal{E}\{\log|\sigma_{EA}^2 \mathbf{S}_A^T\mathbf{S}_A^* + \sigma_{EB}^2\mathbf{S}_B^T\mathbf{S}_B^* + \mathbf{I}_{K_2}|\} = \mathcal{E}\{\log|\mathbf{I}_{K_2} + \mathbf{S}_{AB}\mathbf{T}\mathbf{S}_{AB}^H|\}$.

For $K_2 < N_A + N_B$, we have

$$\begin{aligned}
&\mathcal{E}\{\log|\mathbf{I}_{K_2} + \mathbf{S}_{AB}\mathbf{T}\mathbf{S}_{AB}^H|\} \\
&\geq K_2\mathcal{E}\{\log(1 + |\mathbf{S}_{AB}\mathbf{T}\mathbf{S}_{AB}^H|^{\frac{1}{K_2}})\} \\
&= K_2\mathcal{E}\big\{\log\big(1 + \exp\big(\frac{1}{K_2}\ln|\mathbf{S}_{AB}\mathbf{T}\mathbf{S}_{AB}^H|\big)\big)\big\} \\
&\geq K_2\mathcal{E}\big\{\log\big(1 + \exp\big(\frac{1}{K_2}\ln\sigma_{min}^{2K_2}|\mathbf{S}_{AB}\mathbf{S}_{AB}^H|\big)\big)\big\} \\
&\geq K_2\log\big(1 + \sigma_{min}^2 e_{E1}\big)
\end{aligned} \tag{3.27}$$

where $e_{E1} = \exp(\frac{1}{K_2}\sum_{j=1}^{K_2}\sum_{k=1}^{N_A+N_B-j}\frac{1}{k} - \gamma)$. The second inequality in (3.27) is from the fact (see [48, Th. 3]) that $|\mathbf{S}_{AB}\mathbf{T}\mathbf{S}_{AB}^H| \geq \sigma_{min}^{2K_2}|\mathbf{S}_{AB}\mathbf{S}_{AB}^H|$ where $\sigma_{min}^2 = \min\{\sigma_{EA}^2\frac{P_A}{N_A}, \sigma_{EB}^2\frac{P_B}{N_B}\}$. Similarly, for $K_2 \geq N_A + N_B$, we have

$$\begin{aligned}
&\mathcal{E}\{\log|\mathbf{I} + \mathbf{S}_{AB}\mathbf{T}\mathbf{S}_{AB}^H|\} \\
&= \mathcal{E}\{\log|\mathbf{I} + \mathbf{T}\mathbf{S}_{AB}^H\mathbf{S}_{AB}|\} \\
&\geq (N_A+N_B)\mathcal{E}\big\{\log\big(1 + |\mathbf{T}|^{\frac{1}{N_A+N_B}}exp\big(\frac{1}{N_A+N_B}\ln|\mathbf{S}_{AB}^H\mathbf{S}_{AB}|\big)\big)\big\} \\
&\geq (N_A+N_B)\log\big(1 + |\mathbf{T}|^{\frac{1}{N_A+N_B}}e_{E2}\big)
\end{aligned} \tag{3.28}$$

where $e_{E2} = \exp(\frac{1}{N_A+N_B}\sum_{j=1}^{N_A+N_B}\sum_{k=1}^{K_2-j}\frac{1}{k} - \gamma)$ Therefore, using (3.27) and (3.28), we have from (3.26) that

$$\begin{aligned} & I(\bar{\mathbf{s}}_A, \bar{\mathbf{s}}_B; \mathbf{y}_E | \hat{\mathbf{a}}, \hat{\mathbf{b}}) \\ & \leq K_2 \mathcal{E}\{\log | \frac{\frac{P_A}{N_A}\hat{\mathbf{A}}\hat{\mathbf{A}}^H + \frac{P_B}{N_B}\hat{\mathbf{B}}\hat{\mathbf{B}}^H}{1 + P_A\sigma_{EA}^2 + P_B\sigma_{EB}^2} + \mathbf{I}|\} \\ & + \begin{cases} K_2 N_E \log\left(\frac{1 + P_A\sigma_{EA}^2 + P_B\sigma_{EB}^2}{1 + \sigma_{min}^2 e_{E1}}\right), \ K_2 \leq N_A + N_B \\ N_E \log\left(\frac{(1 + P_A\sigma_{EA}^2 + P_B\sigma_{EB}^2)^{K_2}}{\left(1 + |\mathbf{T}|^{\frac{1}{N_A+N_B}} e_{E2}\right)^{N_A+N_B}}\right), \ K_2 > N_A + N_B \end{cases} \\ & \triangleq \mathcal{R}_{E,t}^+ \end{aligned} \tag{3.29}$$

One can also verify $I(\bar{\mathbf{s}}_A, \bar{\mathbf{s}}_B; \mathbf{y}_E | \hat{\mathbf{a}}, \hat{\mathbf{b}}) \geq K_2 \mathcal{E}\{\log | \frac{\frac{P_A}{N_A}\hat{\mathbf{A}}\hat{\mathbf{A}}^H + \frac{P_B}{N_B}\hat{\mathbf{B}}\hat{\mathbf{B}}^H}{1 + P_A\sigma_{EA}^2 + P_B\sigma_{EB}^2} + \mathbf{I}|\} \triangleq \mathcal{R}_{E,t}^-$ which is the first term in (3.29).

When $P_A = P_B = P \to \infty$, we have $\sigma_{EA}^2 \to \frac{bN_A}{aN_B + bN_A}$, $\sigma_{EB}^2 \to \frac{aN_B}{aN_B + bN_A}$, $\sigma_A^2 \to 0$, $\sigma_B^2 \to 0$, $\mathcal{E}\{\hat{a}_i\hat{a}_i^*\} \to \frac{aN_B}{aN_B + bN_A}$, $\mathcal{E}\{\hat{b}_i\hat{b}_i^*\} \to \frac{bN_A}{aN_B + bN_A}$, $\mathcal{E}\{\hat{h}_{A,i}\hat{h}_{A,i}^*\} \to 1$, $\mathcal{E}\{\hat{h}_{B,i}\hat{h}_{B,i}^*\} \to 1$, $\sigma_{min}^2 = P\min\{\frac{\sigma_{EA}^2}{N_A}, \frac{\sigma_{EB}^2}{N_B}\}$ and $|\mathbf{T}|^{\frac{1}{N_A+N_B}} = P((\frac{\sigma_{EA}^2}{N_A})^{N_A}(\frac{\sigma_{EB}^2}{N_B})^{N_B})^{1/(N_A+N_B)}$.

Then, similar to (3.16), we have

$$\lim_{P\to\infty} \frac{\mathcal{R}_A^-}{\log P} = \lim_{P\to\infty} \frac{\mathcal{R}_A^+}{\log P} = K_2 \min\{N_A, N_B\} \tag{3.30}$$

One can also verify that

$$\lim_{P\to\infty} \frac{\mathcal{R}_{E,t}^-}{\log P} = 0 \tag{3.31}$$

and

$$\lim_{P\to\infty} \frac{\mathcal{R}_{E,t}^+}{\log P} = \begin{cases} 0, & K_2 \leq N_A + N_B \\ N_E(K_2 - N_A - N_B), & K_2 > N_A + N_B \end{cases} \tag{3.32}$$

Now applying (3.16), (3.30), (3.31) and (3.32), and using $\mathcal{R}_{two}^+ \triangleq \frac{1}{K_2}[\mathcal{R}_A^+ + \mathcal{R}_B^+ - \mathcal{R}_{E,t}^-]^+$

and $\mathcal{R}_{two}^{-} \triangleq \frac{1}{K_2}[\mathcal{R}_A^{-} + \mathcal{R}_B^{-} - \mathcal{R}_{E,t}^{+}]^{+}$ as upper and lower bounds on $\mathcal{R}_{two}$, we have

$$\begin{aligned}&\lim_{P\to\infty} \frac{\mathcal{R}_{two}^{-}}{\log P}\\ &= \begin{cases} 2\min\{N_A, N_B\}, & K_2 \leq N_A + N_B \\ \left(2\min\{N_A, N_B\} - \dfrac{N_E}{K_2}(K_2 - N_A - N_B)\right)^{+}, & K_2 > N_A + N_B \end{cases}\end{aligned} \tag{3.33}$$

and

$$\lim_{P\to\infty} \frac{\mathcal{R}_{two}^{+}}{\log P} = 2\min\{N_A, N_B\} \tag{3.34}$$

We see that if $K_2 \leq N_A + N_B$, $\lim_{P\to\infty} \frac{\mathcal{R}_{two}}{\log P} = 2\min\{N_A, N_B\}$ which equals the degrees of freedom of the full-duplex channel between Alice and Bob. And if $K_2 > N_A + N_B$, the above lower bound on $\lim_{P\to\infty} \frac{\mathcal{R}_{two}}{\log P}$ decreases linearly as $N_E$ increases. We see an advantage of two-way information transmission over one-way information transmission.

## 3.3 Conclusion

In this chapter we analyzed the full-duplex MIMOME network subject to the application of anti-eavesdropping channel estimation (ANECE) in a two-phase scheme. Assuming that a statistical model of CSI anywhere is known everywhere, we derived lower and upper bounds on the secure degrees of freedom of the network, which reveal clearly how the number of antennas on Eve affect these bounds. In particular, for $1 \leq K_2 \leq N_A$ in one-way information transmission or $1 \leq K_2 \leq N_A + N_B$ in two-way information transmission, the lower and upper bounds coincide and equal to those of the channel capacity between Alice and Bob.

# Chapter 4

# Optimal Pilot Design for Anti-Eavesdropper Channel Estimation

## 4.1 Introduction

From chapter 3 we see the advantage of using ANECE. In this chapter, we will try to investigate the pilot signals design subject to ANECE since only heuristic pilot design has been proposed from the original work [31]. We will consider two criteria for optimality: 1) minimizing the sum of mean squared errors (MSE) of the minimum-mean-squared-error (MMSE) channel estimation at each and every user, and 2) maximizing the sum of the pair-wise mutual information (MI) between the signals excited by the pilots and observed by all users.

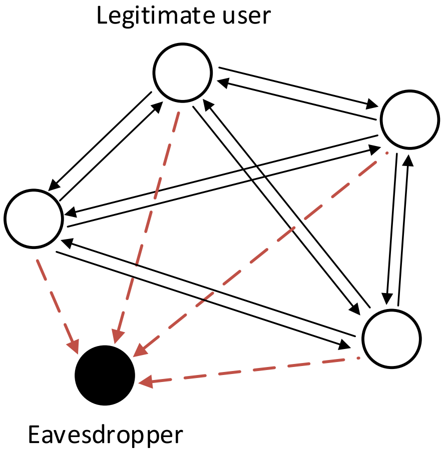

Figure 4.1: Multiple full-duplex multi-antenna users perform ANECE.

## 4.2 System Model

As illustrated in Fig 4.1, we consider a wireless network of $M$ legitimate full-duplex multi-antenna users and a passive multi-antenna eavesdropper (Eve). Let $N_i$ be the number of antennas on user $i$, and $N_E$ be the number of antennas on Eve. According to ANECE [31], all users concurrently transmit their pilots $\mathbf{p}_i(k)$ over a time window $k = 1, \cdots, K$ with $i$ corresponding to user $i$. These pilots are designed in such a way (see below) that all users can reliably estimate their own channel matrices but Eve cannot.

Specifically, let the signal received by user $i$ over the time window be $\mathbf{Y}_i \in \mathbb{C}^{N_i \times K}$, and the signal received by Eve be $\mathbf{Y}_E \in \mathbb{C}^{N_E \times K}$. It follows that

$$\mathbf{Y}_i = \sum_{j \neq i}^{M} \mathbf{R}_i^{\frac{1}{2}} \mathbf{H}_{i,j} \mathbf{R}_j^{\frac{T}{2}} \mathbf{P}_j + \mathbf{N}_i \tag{4.1a}$$

$$\mathbf{Y}_E = \sum_{i=1}^{M} \mathbf{H}_{E,i} \mathbf{P}_i + \mathbf{N}_E \tag{4.1b}$$

where $\mathbf{P}_i = [\mathbf{p}_i(1), \cdots, \mathbf{p}_i(K)]$ is the pilot matrix sent by user $i$, $\mathbf{R}_i^{\frac{1}{2}} \mathbf{H}_{i,j} \mathbf{R}_j^{\frac{T}{2}}$ is the overall channel matrix from user $j$ to user $i$, and $\mathbf{H}_{E,j}$ is the overall channel matrix from user $j$ to Eve. And $\|\mathbf{H}_{E,i}\mathbf{P}_i\|$ for any $i$ is assumed to be not negligible compared to $\|\sum_{j \neq i} \mathbf{H}_{E,j}\mathbf{P}_j\|$. Here, $\mathbf{R}_i = \mathbf{R}_i^{\frac{1}{2}} \mathbf{R}_i^{\frac{H}{2}}$ is the receive/transmit channel correlation matrix of user $i$ (of full rank and known to all users and Eve). We assume that $\mathbf{H}_{E,j}$ for any $j$ is independent of $\mathbf{H}_{i,m}$ for any $i$ and $m$. Furthermore, $\mathbf{H}_{i,j}$ consists of independent and identically distributed (i.i.d.) zero-mean unit-variance complex Gaussian ($\mathcal{CN}(0,1)$) elements. Finally, $\mathbf{N}_i$ includes all residual self-interference at user $i$ and consists of i.i.d. $\mathcal{CN}(0, \sigma_i^2)$ entries.

Now define $N_T = \sum_{i=1}^{M} N_i$, $\bar{\mathbf{P}} = [\mathbf{P}_1^T, \cdots, \mathbf{P}_M^T]^T$, $\bar{\mathbf{P}}_{(i)}$ as $\bar{\mathbf{P}}$ without $\mathbf{P}_i$, $\bar{\mathbf{R}} = diag[\mathbf{R}_1, \cdots, \mathbf{R}_M]$, $\bar{\mathbf{R}}_{(i)}$ as $\bar{\mathbf{R}}$ without $\mathbf{R}_i$, $\bar{\mathbf{H}}_{(i)}$ as the horizontal stack of $\mathbf{H}_{i,j}$ for all $j \neq i$, and $\bar{\mathbf{H}}_E = [\mathbf{H}_{E,1}, \cdots, \mathbf{H}_{E,M}]$. Then (4.1) can be rewritten as

$$\mathbf{Y}_i = \mathbf{R}_i^{\frac{1}{2}} \bar{\mathbf{H}}_{(i)} \bar{\mathbf{R}}_{(i)}^{\frac{T}{2}} \bar{\mathbf{P}}_{(i)} + \mathbf{N}_i \tag{4.2a}$$

$$\mathbf{Y}_E = \bar{\mathbf{H}}_E \bar{\mathbf{P}} + \mathbf{N}_E. \tag{4.2b}$$

For ANECE, we need to choose the (publicly known) pilots such that $rank(\bar{\mathbf{P}}_{(i)}) = N_T - N_i$ (i.e., the rows of $\bar{\mathbf{P}}_{(i)}$ are independent for every $i$) and $rank(\bar{\mathbf{P}}) = r \leq N_T - 1$ (i.e., the rows of $\bar{\mathbf{P}}$ are not independent). It is clear from (4.2) that the first rank constraint allows each user to obtain a consistent estimate of its channel matrix while the second rank constraint creates a subspace of Eve's channel matrix for which there is no consistent estimation. To explain the latter, we know that there is a $(N_T - r) \times N_T$ matrix $\mathbf{C}$ such that $\mathbf{C}\bar{\mathbf{P}} = 0$ and hence $\bar{\mathbf{H}}_E\bar{\mathbf{P}} = (\bar{\mathbf{H}}_E + \mathbf{\Theta}_0\mathbf{C})\bar{\mathbf{P}}$ for any $\mathbf{\Theta}_0 \in \mathbb{C}^{N_E \times (N_T - r)}$. Because of this, the subsequent exchange of information between users is better protected than otherwise [31].

This paper discusses the optimal designs of the pilots subject to the above rank constraints. We will consider two design criteria: one is based on MMSE channel estimation, and the other is based on maximal MI between observations. A discussion of maximum likelihood (ML) channel estimation is included in the end of the next section.

## 4.3 For Optimal MMSE Channel Estimation

Define $\mathbf{S}_i$ as the $N_i \times N_T$ selection matrix such that $\mathbf{S}_i\bar{\mathbf{P}} = \mathbf{P}_i$, and $\bar{\mathbf{S}}_{(i)}$ as the vertical stack of $\mathbf{S}_j$ for all $j \neq i$. Note that $\bar{\mathbf{R}}_{(i)}^{\frac{T}{2}}\bar{\mathbf{P}}_{(i)} = \bar{\mathbf{S}}_{(i)}\bar{\mathbf{R}}^{\frac{T}{2}}\bar{\mathbf{P}}$. Also using $vec(\mathbf{XYZ}) = (\mathbf{Z}^T \otimes \mathbf{X})vec(\mathbf{Y})$, (4.2a) becomes

$$\mathbf{y}_i = \bar{\mathbf{G}}_i^H\bar{\mathbf{h}}_i + \mathbf{n}_i \tag{4.3}$$

where $\mathbf{y}_i = vec(\mathbf{Y}_i)$, $\bar{\mathbf{h}}_i = vec(\bar{\mathbf{H}}_{(i)})$, $\mathbf{n}_i = vec(\mathbf{N}_i)$ and $\bar{\mathbf{G}}_i = (\bar{\mathbf{S}}_{(i)}\bar{\mathbf{R}}^{\frac{H}{2}}\bar{\mathbf{P}}^* \otimes \mathbf{R}_i^{\frac{H}{2}})$.

Let $\mathbf{K}_{\mathbf{x},\mathbf{y}} = \mathcal{E}\{\mathbf{x}\mathbf{y}^H\}$ be the correlation matrix between two random vectors $\mathbf{x}$ and $\mathbf{y}$, and $\mathbf{K}_{\mathbf{x}} = \mathbf{K}_{\mathbf{x},\mathbf{x}}$. The MMSE estimate of $\bar{\mathbf{h}}_i$ by user $i$ is

$$\hat{\bar{\mathbf{h}}}_i = \mathbf{K}_{\bar{\mathbf{h}}_i,\mathbf{y}_i}\mathbf{K}_{\mathbf{y}_i}^{-1}\mathbf{y}_i = \bar{\mathbf{G}}_i(\bar{\mathbf{G}}_i^H\bar{\mathbf{G}}_i + \sigma_i^2\mathbf{I})^{-1}\mathbf{y}_i. \tag{4.4}$$

Define $\Delta\bar{\mathbf{h}}_i = \bar{\mathbf{h}}_i - \hat{\bar{\mathbf{h}}}_i$. Then the MSE of $\hat{\bar{\mathbf{h}}}_i$ is

$$\begin{aligned}
\mathtt{MSE}_i &= Tr(\mathcal{E}\{\Delta\bar{\mathbf{h}}_i\Delta\bar{\mathbf{h}}_i^H\}) = Tr(\mathbf{K}_{\bar{\mathbf{h}}_i} - \mathbf{K}_{\bar{\mathbf{h}}_i,\mathbf{y}_i}\mathbf{K}_{\mathbf{y}_i}^{-1}\mathbf{K}_{\mathbf{y}_i,\bar{\mathbf{h}}_i}) \\
&= Tr\left(\mathbf{I} - \bar{\mathbf{G}}_i(\bar{\mathbf{G}}_i^H\bar{\mathbf{G}}_i + \sigma_i^2\mathbf{I})^{-1}\bar{\mathbf{G}}_i^H\right) \\
&= Tr\left(\left(\mathbf{I} + \frac{1}{\sigma_i^2}\bar{\mathbf{G}}_i\bar{\mathbf{G}}_i^H\right)^{-1}\right) \qquad (4.5)
\end{aligned}$$

where the last equality is based on the well known matrix inverse lemma.

Now we have the following criterion for pilot design:

$$\min_{\bar{\mathbf{P}}} \quad J_M = \sum_{i=1}^{M} \mathtt{MSE}_i \qquad (4.6)$$

$$s.t.\ Tr(\mathbf{P}_i\mathbf{P}_i^H) \le KP_i,\ i = 1, \ldots, M,$$

$$rank(\bar{\mathbf{P}}) \le r$$

where $N_T - N_{min} \le r \le N_T - 1$ with $N_{min} = \min_i N_i$.

Since $\bar{\mathbf{R}}$ is known and nonsingular, we can choose

$$\bar{\mathbf{R}}^{\frac{H}{2}}\bar{\mathbf{P}}^* = \bar{\mathbf{F}}\bar{\mathbf{V}} \qquad (4.7)$$

where $\bar{\mathbf{V}} \in \mathbb{C}^{r\times K}$ is any semi-unitary matrix satisfying $\bar{\mathbf{V}}\bar{\mathbf{V}}^H = \mathbf{I}_r$, and $\bar{\mathbf{F}} \in \mathbb{C}^{N_T\times r}$ is now what we need to design. Namely,

$$\bar{\mathbf{P}} = \bar{\mathbf{R}}^{-\frac{T}{2}}\bar{\mathbf{F}}^*\bar{\mathbf{V}}^* \qquad (4.8)$$

which meets the rank constraint. To further simplify (4.6), we use the eigenvalue decomposition (EVD):

$$\mathbf{R}_i = \tilde{\mathbf{U}}_i\tilde{\mathbf{\Lambda}}_i\tilde{\mathbf{U}}_i^H \qquad (4.9)$$

where $\tilde{\mathbf{\Lambda}}_i = diag\{\tilde{\lambda}_{i,1}, \ldots, \tilde{\lambda}_{i,N_i}\}$ with $\sum_l \tilde{\lambda}_{i,l} = N_i$. The diagonal elements in $\tilde{\mathbf{\Lambda}}_i$ are in descending order. From (4.9), we have $\mathbf{R}_i^{\frac{1}{2}} = \tilde{\mathbf{U}}_i\tilde{\mathbf{\Lambda}}_i^{\frac{1}{2}}$.

With (4.7) and (4.9), the cost function in (4.6) becomes

$$J_M = \sum_{i=1}^{M} Tr\left(\left[\mathbf{I} + \frac{1}{\sigma_i^2}(\tilde{\mathbf{\Lambda}}_i \otimes \bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T)\right]^{-1}\right) \tag{4.10}$$

where we have used $Tr([\mathbf{I}+\mathbf{X}\otimes\mathbf{Y}]^{-1}) = Tr([\mathbf{I}+\mathbf{Y}\otimes\mathbf{X}]^{-1})$, and hence (4.6) becomes

$$\min_{\bar{\mathbf{F}}} \quad J_M \tag{4.11}$$

$$s.t.\ Tr(\mathbf{S}_i\bar{\mathbf{R}}^{-\frac{H}{2}}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{R}}^{-\frac{1}{2}}\mathbf{S}_i^T) \leq KP_i,\ i = 1, \ldots, M.$$

This problem is non-convex in general. We will next treat it in three separate situations. We will first present a general algorithm for $M \geq 2$, then a specialized (efficient) algorithm for $M = 2$, and finally closed-form solutions of the optimal pilots under the case of $M \geq 2$, $N_i = N$, $P_i = P$, $\sigma_i^2 = \sigma^2$ and $\mathbf{R}_i = \mathbf{I}_N$. The invariance of the above parameters to $i$ is called a symmetric condition, and $\mathbf{R}_i = \mathbf{I}_N$ is an isotropic condition.

### 4.3.1 General algorithm for $M \geq 2$

To solve the problem (4.11) with $M \geq 2$, we can apply the logarithmic barrier method [41]. With the barrier coefficient $t$, we define

$$g_1(\bar{\mathbf{F}}) = tJ_M + \sum_{i=1}^{M} \mathcal{B}_i(\bar{\mathbf{F}}) \tag{4.12}$$

where

$$\mathcal{B}_i(\bar{\mathbf{F}}) = -\ln(\psi_i(\bar{\mathbf{F}})) \tag{4.13}$$

and $\psi_i(\bar{\mathbf{F}}) = KP_i - Tr(\mathbf{S}_i\bar{\mathbf{R}}^{-\frac{H}{2}}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{R}}^{-\frac{1}{2}}\mathbf{S}_i^T)$. Then, (4.11) is approximated by

$$\min_{\bar{\mathbf{F}}} \quad g_1(\bar{\mathbf{F}}) \tag{4.14}$$

The gradient of a real-valued function $f(\mathbf{X})$ with respect to a complex matrix $\mathbf{X}$ is denoted and defined as $\nabla f(\mathbf{X}) = \frac{\partial f(\mathbf{X})}{\partial \mathbf{X}} = \frac{\partial f(\mathbf{X})}{\partial \Re(\mathbf{X})} + j\frac{\partial f(\mathbf{X})}{\partial \Im(\mathbf{X})}$. One can verify that $\nabla g_1(\bar{\mathbf{F}}) = t\nabla J_M(\bar{\mathbf{F}}) + \sum_{i=1}^{M} \nabla \mathcal{B}_i(\bar{\mathbf{F}})$ where

$$\nabla J_M(\bar{\mathbf{F}}) = -2\sum_{i=1}^{M}\sum_{l=1}^{N_i} \frac{\tilde{\lambda}_{i,l}}{\sigma_i^2}\bar{\mathbf{S}}_{(i)}^T(\mathbf{I} + \frac{\tilde{\lambda}_{i,l}}{\sigma_i^2}\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T)^{-2}\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}, \tag{4.15}$$

$$\nabla \mathcal{B}_i(\bar{\mathbf{F}}) = 2\left(\frac{\bar{\mathbf{R}}^{-\frac{1}{2}}\mathbf{S}_i^T\mathbf{S}_i\bar{\mathbf{R}}^{-\frac{H}{2}}\bar{\mathbf{F}}}{\psi_i(\bar{\mathbf{F}})}\right). \tag{4.16}$$

Algorithm 4 shown in the table solves (4.14) using gradient descent where $\bar{\mathbf{F}}$ is initially set to be $\sqrt{\mathbf{D}}\mathbf{Q}_t \in \mathbb{C}^{N_T \times r}$, $\mathbf{Q}_t$ is the $N_T \times N_T$ discrete Fourier transform (DFT) matrix without the last $(N_T - r)$ columns and $\mathbf{D} = diag\{d_1\mathbf{1}_{N_1}^T, \ldots, d_M\mathbf{1}_{N_M}^T\} \in \mathbb{R}^{N_T \times N_T}$ is a positive definite matrix for power controlling. This initialization is based on the pilots proposed in [31].

### 4.3.2 Special algorithm for $M = 2$

When $M = 2$, we can develop an efficient algorithm with guaranteed global optimality. This algorithm has a simple connection with that in [37] as shown next.

Denote the two users by the indices $i = 1$ and $i = 2$. Now the cost function is $J_2$ given by (4.10) with $M = 2$. Notice that $\bar{\mathbf{S}}_{(1)}\bar{\mathbf{F}} = \mathbf{S}_2\bar{\mathbf{F}} \in \mathbb{C}^{N_2 \times r}$ and $\bar{\mathbf{S}}_{(2)}\bar{\mathbf{F}} = \mathbf{S}_1\bar{\mathbf{F}} \in \mathbb{C}^{N_1 \times r}$, which do not have any shared entry. Let us now use the following singular value decompositions (SVDs) to reparameterize $\bar{\mathbf{F}}$:

$$\begin{cases} \bar{\mathbf{S}}_{(2)}\bar{\mathbf{F}} = \mathbf{U}_1\mathbf{\Lambda}_1\mathbf{V}_1^H, \\ \bar{\mathbf{S}}_{(1)}\bar{\mathbf{F}} = \mathbf{U}_2\mathbf{\Lambda}_2\mathbf{V}_2^H \end{cases} \tag{4.17}$$

**Algorithm 4** Solving (4.14) with increasing $t$.

```
Input:
    $r, \bar{\mathbf{R}}, N_i, \sigma_i, P_i$, $T$, for $i = 1, \ldots, M$;
    Accuracy thresholds: $\epsilon_1$, $\epsilon_2$, $N_p$.
    Initialization: $t > 0$, $\mu > 1$, and $\bar{\mathbf{F}}^{(0)} = \sqrt{\mathbf{D}}\mathbf{Q}_t$.
repeat
    p=0;
    repeat
        Compute the derivatives $\frac{\partial g_1(\bar{\mathbf{F}}^{(p)})}{\partial \bar{\mathbf{F}}^{(p)}}$.
        Choose step size $\gamma^{(p)}$ via backtracking line search [41].
        Update $\bar{\mathbf{F}}^{(p+1)} = \bar{\mathbf{F}}^{(p)} - \gamma^{(p)} \nabla g_1(\bar{\mathbf{F}}^{(p)})$.
        p = p+1.
    until $\|\nabla g_1(\bar{\mathbf{F}}^{(p)}) - \nabla g_1(\bar{\mathbf{F}}^{(p-1)})\| \leq \epsilon_2$ or $p \geq N_p$
    $\bar{\mathbf{F}}^{(0)} = \bar{\mathbf{F}}^{(p)}$, $t = \mu t$.
until $\frac{M}{t} < \epsilon_1$
return $\bar{\mathbf{F}}^{(p)}$
```

where $\mathbf{U}_1 \in \mathbb{C}^{N_1 \times N_1}$, $\mathbf{\Lambda}_1 \in \mathbb{R}^{N_1 \times r}$, $\mathbf{V}_1 \in \mathbb{C}^{r \times r}$, $\mathbf{U}_2 \in \mathbb{C}^{N_2 \times N_2}$, $\mathbf{\Lambda}_2 \in \mathbb{R}^{N_2 \times r}$ and $\mathbf{V}_2 \in \mathbb{C}^{r \times r}$. All of these matrices need to be optimized as they all affect the pilots. With $r \geq \max\{N_1, N_2\}$, we denote the singular value matrices in (4.17) as

$\mathbf{\Lambda}_1 = [diag\{\lambda_{1,1}, \ldots, \lambda_{1,N_1}\}, \mathbf{0}_{N_1 \times (r-N_1)}]$ and $\mathbf{\Lambda}_2 = [diag\{\lambda_{2,1}, \ldots, \lambda_{2,N_2}\}, \mathbf{0}_{N_2 \times (r-N_2)}]$ where the diagonal elements in each matrix are in descending order. Using (4.8) and (4.17), we have

$$\bar{\mathbf{P}} = \bar{\mathbf{R}}^{-\frac{T}{2}}[(\mathbf{U}_1\mathbf{\Lambda}_1\mathbf{V}_1^H)^T, (\mathbf{U}_2\mathbf{\Lambda}_2\mathbf{V}_2^H)^T]^H \bar{\mathbf{V}}^*. \tag{4.18}$$

Let $\mathbf{\Lambda}_1^2 = diag\{\lambda_{1,1}^2, \ldots, \lambda_{1,N_1}^2\}$ and $\mathbf{\Lambda}_2^2 = diag\{\lambda_{2,1}^2, \ldots, \lambda_{2,N_2}^2\}$. Also let $\mathbf{C}_1 = \tilde{\mathbf{\Lambda}}_1^{-1}\mathbf{\Lambda}_1^2$ and

$\mathbf{C}_2 = \tilde{\mathbf{\Lambda}}_2^{-1}\mathbf{\Lambda}_2^2$. Then one can verify that $J_2$ becomes

$$J_2 = Tr((\mathbf{I} + \frac{1}{\sigma_1^2}(\tilde{\mathbf{\Lambda}}_1 \otimes \mathbf{C}_2\tilde{\mathbf{\Lambda}}_2)^{-1})$$

$$+ Tr((\mathbf{I} + \frac{1}{\sigma_2^2}(\tilde{\mathbf{\Lambda}}_2 \otimes \mathbf{C}_1\tilde{\mathbf{\Lambda}}_1))^{-1}) \tag{4.19}$$

which is invariant to $\mathbf{U}_1$, $\mathbf{V}_1$, $\mathbf{U}_2$ and $\mathbf{V}_2$. Only $\mathbf{C}_1$ and $\mathbf{C}_2$ remain to be optimized as far as the cost function is concerned.

For the power constraints, we see that for $i = 1, 2$,

$$Tr(\mathbf{P}_i\mathbf{P}_i^H) = Tr(\tilde{\mathbf{\Lambda}}_i^{-1}\mathbf{U}_i\mathbf{\Lambda}_i^2\mathbf{U}_i^H) \geq Tr(\tilde{\mathbf{\Lambda}}_i^{-1}\mathbf{\Lambda}_i^2) = Tr(\mathbf{C}_i) \tag{4.20}$$

where the equality in $\geq$ holds when $\mathbf{U}_i = \mathbf{I}_{N_i}$ [49, H.1.h].

Therefore, both the cost and the power constraints are optimized by choosing $\mathbf{U}_i$ and $\mathbf{V}_i$ with $i = 1, 2$ to be the identity matrices. So, (4.11) becomes

$$\min_{\mathbf{C}_1,\mathbf{C}_2} \; J_2 \tag{4.21}$$

$$s.t. \; Tr(\mathbf{C}_1) \leq KP_1, \; Tr(\mathbf{C}_2) \leq KP_2.$$

where $J_2$ is shown in (4.19). Here $\mathbf{C}_1$ and $\mathbf{C}_2$ are completely decoupled from each other. Each of the two decoupled problems can be solved by following [37, 50]. It is obvious that if $\tilde{\mathbf{\Lambda}}_i$ is proportional to the identity matrix, so is the optimal $\mathbf{C}_i$.

### 4.3.3 Closed-form solution

For $M \geq 2$, we now consider the (previously mentioned) symmetric and isotropic case, i.e., $N_i = N$, $P_i = P$, $\sigma_i^2 = \sigma^2$ and $\mathbf{R}_i = \mathbf{I}_N$. Furthermore, we consider $r = (M-1)N$ which yields the maximal dimensional of the subspace of Eve's CSI that is not identifiable by Eve. Then from (4.10), $J_M = N\sum_{i=1}^{M} Tr\big((\mathbf{I} + \frac{1}{\sigma^2}\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T)^{-1}\big)$. Also the power constraints become $Tr(\mathbf{S}_i\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_i^T) \leq KP,\ i = 1, \ldots, M$. The corresponding Lagrangian function is

$$\mathcal{L} = J_M + \sum_{i=1}^{M} \mu_i(Tr(\mathbf{S}_i\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_i^T) - KP) \tag{4.22}$$

and the KKT conditions [41] are

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial \bar{\mathbf{F}}} = \dfrac{\partial J_M}{\partial \bar{\mathbf{F}}} + 2\displaystyle\sum_{i=1}^{M} \mu_i \mathbf{S}_i^T\mathbf{S}_i\bar{\mathbf{F}} = 0 \\ Tr(\mathbf{S}_i\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_i^T) \leq KP,\ i = 1, \ldots, M \\ \mu_i(Tr(\mathbf{S}_i\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_i^T) - KP) = 0,\ \mu_i \geq 0,\ i = 1, \ldots, M \end{cases} \tag{4.23}$$

It is shown below that a set of (equally optimal) solutions to (4.23) are given by the $NM \times NM$ discrete Fourier transform (DFT) matrix $\mathbf{Q}$ with any $N$ equally spaced columns removed.

**Theorem 6** *Let* $\mathbf{Q}$ *be such that its* $(l+1, k+1)$*th element is* $(\mathbf{Q})_{l+1,k+1} = w_{NM}^{lk}$ *with* $w_{NM} = e^{-j2\pi\frac{1}{MN}}$, $0 \leq l \leq NM-1$ *and* $0 \leq k \leq NM-1$*. Let* $\mathbf{Q}_m$ *consist of* $N$ *equally spaced columns of* $\mathbf{Q}$ *as follows:*

$$\mathbf{Q}_m = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ w_{MN}^{m} & w_{MN}^{m+M} & \cdots & w_{MN}^{m+(N-1)M} \\ \vdots & & & \vdots \\ w_{MN}^{m(NM-1)} & w_{MN}^{(m+M)(NM-1)} & \cdots & w_{MN}^{(m+(N-1)M)(NM-1)} \end{bmatrix}. \tag{4.24}$$

*Also let* $\bar{\mathbf{Q}}_m$ *be* $\mathbf{Q}$ *without the columns in* $\mathbf{Q}_m$*. Then, a solution to* (4.23) *is* $\bar{\mathbf{F}} = \sqrt{\frac{KP}{N^2(M-1)}}\bar{\mathbf{Q}}_m$ *where* $m$ *can be any integer in* $[0, M-1]$*.*

**Proof.** See Appendix 4.7.1. ■

For $M = 2$, the theorem yields $\mathbf{P}_i = \mathbf{S}_i\bar{\mathbf{F}}^*\bar{\mathbf{V}}^*$ that satisfies $\mathbf{P}_i\mathbf{P}_i^H = \frac{KP}{N}\mathbf{I}_N$ where $i = 1, 2$ (easy to verify). These pilots are known to be globally optimal. For $M \geq 3$, our numerical simulations using the previously developed algorithm did not yield any result better than that from Theorem 6 subject to the conditions in the theorem.

**For optimal ML channel estimation**

The ML estimate of $\bar{\mathbf{h}}_i$ is $\hat{\bar{\mathbf{h}}}_{i,ML} = (\bar{\mathbf{G}}_i\bar{\mathbf{G}}_i^H)^{-1}\bar{\mathbf{G}}_i\mathbf{y}_i$ and its covariance matrix is $\mathbf{C}_{i,ML} = \sigma_i^2(\bar{\mathbf{G}}_i\bar{\mathbf{G}}_i^H)^{-1} = \sigma_i^2(\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T \otimes \mathbf{R}_i^{\frac{H}{2}}\mathbf{R}_i^{\frac{1}{2}})^{-1}$. We can design the optimal pilots by minimizing $J_{M,ML} = \sum_{i=1}^M Tr(\mathbf{C}_{i,ML})$ subject to the same power constraints as before.

If $N_i = N$, $P_i = P$, $\sigma_i^2 = \sigma^2$, $\mathbf{R}_i = \mathbf{I}_N$ and $r = (M-1)N$, one can verify that $J_{M,ML}$ equals $J_M$ as $\sigma^2$ becomes small or equivalently $KP$ becomes large. Hence, the optimal pilots from Theorem 6 also apply here (which can also be proved directly by following a similar procedure used for Theorem 6).

## 4.4 For Maximum Mutual Information

Given $\mathbf{Y}_i$ at user $i$ for all $i$ as shown in (4.2a), every pair of users can follow a secret key generation protocol [35, 36] to produce a (shared) secret key. This secret key can be a useful by-product of ANECE which was originally designed to protect the information directly transmitted between users [31]. If $\mathbf{Y}_E$ received by Eve as shown in (4.2b) or equivalently the Eve's channel matrix $\bar{\mathbf{H}}_E$ is independent of all channel matrices between users, the capacity of the secret key (in bits per channel coherence period) achievable between user $i$ and user $j$ is known [34, Th. 4.1] to be $I(\mathbf{Y}_i;\mathbf{Y}_j)$ which is the mutual information between $\mathbf{Y}_i$ and $\mathbf{Y}_j$. So, it is also meaningful to design the optimal pilots as follows:

$$\begin{aligned} \max_{\bar{\mathbf{P}}} \quad & I_M = \sum_{i=1}^{M-1}\sum_{j=i+1}^{M} I(\mathbf{Y}_i;\mathbf{Y}_j) \\ s.t.\ & Tr(\mathbf{P}_i\mathbf{P}_i^H) \leq KP_i,\ i = 1,\ldots,M \\ & rank(\bar{\mathbf{P}}) \leq r, \end{aligned} \tag{4.25}$$

which is in contrast to (4.6). The above problem is also non-convex. We will treat it next in three separate situations as before.

### 4.4.1 General algorithm for $M \geq 2$

From (4.1a), we can write

$$\begin{cases} \mathbf{y}_i = \sum_{j\neq i}^{M}(\bar{\mathbf{P}}^T\bar{\mathbf{R}}^{\frac{1}{2}}\mathbf{S}_j^T \otimes \mathbf{R}_i^{\frac{1}{2}})\mathbf{h}_{i,j} + \mathbf{n}_i \\ \mathbf{y}_{T,j} = \sum_{i\neq j}^{M}(\mathbf{R}_j^{\frac{1}{2}} \otimes \bar{\mathbf{P}}^T\bar{\mathbf{R}}^{\frac{1}{2}}\mathbf{S}_i^T)\mathbf{h}_{i,j} + \mathbf{n}_{T,j} \end{cases} \tag{4.26}$$

where $\mathbf{y}_i = vec(\mathbf{Y}_i)$, $\mathbf{y}_{T,j} = vec(\mathbf{Y}_j^T)$, $\mathbf{H}_{i,j} = \mathbf{H}_{j,i}^T$, $\mathbf{h}_{i,j} = vec(\mathbf{H}_{i,j})$, $\mathbf{n}_i = vec(\mathbf{N}_i)$ and $\mathbf{n}_{T,j} = vec(\mathbf{N}_j^T)$. Clearly we have $I(\mathbf{Y}_i;\mathbf{Y}_j) = I(\mathbf{y}_i;\mathbf{y}_{T,j})$.

Recall $\bar{\mathbf{G}}_i = (\bar{\mathbf{S}}_{(i)}\bar{\mathbf{R}}^{\frac{H}{2}}\bar{\mathbf{P}}^* \otimes \mathbf{R}_i^{\frac{H}{2}})$. Also define $\bar{\mathbf{G}}_{T,j} = (\mathbf{R}_j^{\frac{H}{2}} \otimes \bar{\mathbf{S}}_{(j)}\bar{\mathbf{R}}^{\frac{H}{2}}\bar{\mathbf{P}}^*)$, $\mathbf{G}_{i,j} = (\mathbf{S}_j\bar{\mathbf{R}}^{\frac{H}{2}}\bar{\mathbf{P}}^* \otimes \mathbf{R}_i^{\frac{H}{2}})$ and $\mathbf{G}_{T,j,i} = (\mathbf{R}_j^{\frac{H}{2}} \otimes \mathbf{S}_i\bar{\mathbf{R}}^{\frac{H}{2}}\bar{\mathbf{P}}^*)$. From (4.26) , one can verify that

$$\mathbf{K}_{\mathbf{y}_i} = \sigma_i^2\mathbf{I} + \bar{\mathbf{G}}_i^H\bar{\mathbf{G}}_i \tag{4.27}$$

$$\mathbf{K}_{\mathbf{y}_{T,j}} = \sigma_j^2\mathbf{I} + \bar{\mathbf{G}}_{T,j}^H\bar{\mathbf{G}}_{T,j} \tag{4.28}$$

$$\mathbf{K}_{\mathbf{y}_i,\mathbf{y}_{T,j}} = \mathbf{G}_{i,j}^H\mathbf{G}_{T,j,i} \tag{4.29}$$

$$\mathbf{K}_{\mathbf{y}_{T,j},\mathbf{y}_i} = \mathbf{G}_{T,j,i}^H \mathbf{G}_{i,j}. \tag{4.30}$$

Also note

$$\begin{aligned} I(\mathbf{y}_i; \mathbf{y}_{T,j}) &= h(\mathbf{y}_i) + h(\mathbf{y}_{T,j}) - h(\mathbf{y}_i, \mathbf{y}_{T,j}) \\ &= \log_2 |\mathbf{K}_{\mathbf{y}_i}| + \log_2 |\mathbf{K}_{\mathbf{y}_{T,j}}| - \log_2 |\mathbf{K}_{\{\mathbf{y}_i, \mathbf{y}_{T,j}\}}| \\ &= -\log_2 |\mathbf{I} - \mathbf{K}_{\mathbf{y}_{T,j}}^{-1} \mathbf{K}_{\mathbf{y}_{T,j},\mathbf{y}_i} \mathbf{K}_{\mathbf{y}_i}^{-1} \mathbf{K}_{\mathbf{y}_i,\mathbf{y}_{T,j}}| \qquad (4.31\text{a}) \\ &= -\log_2 |\mathbf{I} - (\sigma_j^2 \mathbf{I} + \bar{\mathbf{G}}_{T,j}^H \bar{\mathbf{G}}_{T,j})^{-1} \mathbf{G}_{T,j,i}^H \mathbf{G}_{i,j} (\sigma_i^2 \mathbf{I} + \bar{\mathbf{G}}_i^H \bar{\mathbf{G}}_i)^{-1} \mathbf{G}_{i,j}^H \mathbf{G}_{T,j,i}| \quad (4.31\text{b}) \end{aligned}$$

where

$$\mathbf{K}_{\{\mathbf{y}_i, \mathbf{y}_{T,j}\}} = \begin{bmatrix} \mathbf{K}_{\mathbf{y}_i} & \mathbf{K}_{\mathbf{y}_i,\mathbf{y}_{T,j}} \\ \mathbf{K}_{\mathbf{y}_{T,j},\mathbf{y}_i} & \mathbf{K}_{\mathbf{y}_{T,j}} \end{bmatrix} \tag{4.32}$$

and the last equality in (4.31a) is based on the fact that $\left| \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y}^H & \mathbf{Z} \end{bmatrix} \right| = |\mathbf{X}||\mathbf{Z} - \mathbf{Y}^H \mathbf{X}^{-1} \mathbf{Y}| = |\mathbf{Z}||\mathbf{X} - \mathbf{Y}\mathbf{Z}^{-1}\mathbf{Y}^H|$ with invertible $\mathbf{X}$ and $\mathbf{Z}$.

From (4.26), we can express the MMSE estimates of $\mathbf{h}_{i,j}$ by users $i$ and $j$, respectively, as

$$\begin{cases} \hat{\mathbf{h}}_{ij,i} = \mathbf{K}_{\mathbf{h}_{i,j},\mathbf{y}_i} \mathbf{K}_{\mathbf{y}_i}^{-1} \mathbf{y}_i = \mathbf{G}_{i,j} (\sigma_i^2 \mathbf{I} + \bar{\mathbf{G}}_i^H \bar{\mathbf{G}}_i)^{-1} \mathbf{y}_i \\ \hat{\mathbf{h}}_{ij,j} = \mathbf{K}_{\mathbf{h}_{i,j},\mathbf{y}_{T,j}} \mathbf{K}_{\mathbf{y}_{T,j}}^{-1} \mathbf{y}_{T,j} = \mathbf{G}_{T,j,i} (\sigma_j^2 \mathbf{I} + \bar{\mathbf{G}}_{T,j}^H \bar{\mathbf{G}}_{T,j})^{-1} \mathbf{y}_{T,j}. \end{cases} \tag{4.33}$$

The following lemma is a generalization of a SISO result shown in [51]. It also complements the fact that $I(\mathbf{y}_i; \mathbf{y}_{T,j})$ equals to the mutual information between the ML estimates of $\mathbf{h}_{i,j}$ by users $i$ and $j$ [35].

**Lemma 7** *For each pair of $i$ and $j$, if $\mathbf{S}_j \bar{\mathbf{R}}^{\frac{H}{2}} \bar{\mathbf{P}}^*$, $\mathbf{S}_i \bar{\mathbf{R}}^{\frac{H}{2}} \bar{\mathbf{P}}^*$, $\mathbf{R}_i$, $\mathbf{R}_j$ have all full row ranks (which requires $K \geq \max\{N_i, N_j\}$), then we have $I(\mathbf{y}_i; \mathbf{y}_{T,j}) = I(\hat{\mathbf{h}}_{ij,i}; \hat{\mathbf{h}}_{ij,j})$.*

**Proof.** With the stated conditions, we have $\mathbf{K}_{\hat{\mathbf{h}}_{ij,i}} = \mathbf{G}_{i,j}(\sigma_i^2\mathbf{I} + \bar{\mathbf{G}}_i^H\bar{\mathbf{G}}_i)^{-1}\mathbf{G}_{i,j}^H$, $\mathbf{K}_{\hat{\mathbf{h}}_{ij,j}} = \mathbf{G}_{T,j,i}(\sigma_j^2\mathbf{I} + \bar{\mathbf{G}}_{T,j}^H\bar{\mathbf{G}}_{T,j})^{-1}\mathbf{G}_{T,j,i}^H$, and $\mathbf{K}_{\hat{\mathbf{h}}_{ij,i},\hat{\mathbf{h}}_{ij,j}} = \mathbf{G}_{i,j}(\sigma_i^2\mathbf{I} + \bar{\mathbf{G}}_i^H\bar{\mathbf{G}}_i)^{-1}\mathbf{G}_{i,j}^H\mathbf{G}_{T,j,i}(\sigma_j^2\mathbf{I} + \bar{\mathbf{G}}_{T,j}^H\bar{\mathbf{G}}_{T,j})^{-1}\mathbf{G}_{T,j,i}^H$. Also, $\mathbf{K}_{\hat{\mathbf{h}}_{ij,i},\hat{\mathbf{h}}_{ij,j}} = \mathbf{K}_{\hat{\mathbf{h}}_{ij,i}}\mathbf{K}_{\hat{\mathbf{h}}_{ij,j}}$. Then,

$$\begin{aligned}
&I(\hat{\mathbf{h}}_{ij,i};\hat{\mathbf{h}}_{ij,j})\\
&= -\log_2|\mathbf{I} - \mathbf{K}_{\hat{\mathbf{h}}_{ij,j}}^{-1}\mathbf{K}_{\hat{\mathbf{h}}_{ij,j},\hat{\mathbf{h}}_{ij,i}}\mathbf{K}_{\hat{\mathbf{h}}_{ij,i}}^{-1}\mathbf{K}_{\hat{\mathbf{h}}_{ij,i},\hat{\mathbf{h}}_{ij,j}}|\\
&= -\log_2|\mathbf{I} - \mathbf{K}_{\hat{\mathbf{h}}_{ij,i}}\mathbf{K}_{\hat{\mathbf{h}}_{ij,j}}|\\
&= -\log_2|\mathbf{I} - \mathbf{G}_{i,j}(\sigma_i^2\mathbf{I} + \bar{\mathbf{G}}_i^H\bar{\mathbf{G}}_i)^{-1}\mathbf{G}_{i,j}^H\mathbf{G}_{T,j,i}(\sigma_j^2\mathbf{I} + \bar{\mathbf{G}}_{T,j}^H\bar{\mathbf{G}}_{T,j})^{-1}\mathbf{G}_{T,j,i}^H| = I(\mathbf{y}_i;\mathbf{y}_{T,j})
\end{aligned} \tag{4.34}$$

where the last equation follows from (4.31b) using $\log_2|\mathbf{I} - \mathbf{XY}| = \log_2|\mathbf{I} - \mathbf{YX}|$. ■

Define $\mathbf{\Gamma}_{i,j} = \mathbf{G}_{i,j}(\sigma_i^2\mathbf{I}+\bar{\mathbf{G}}_i^H\bar{\mathbf{G}}_i)^{-1}\mathbf{G}_{i,j}^H$ and $\mathbf{\Gamma}_{T,j,i} = \mathbf{G}_{T,j,i}(\sigma_j^2\mathbf{I}+\bar{\mathbf{G}}_{T,j}^H\bar{\mathbf{G}}_{T,j})^{-1}\mathbf{G}_{T,j,i}^H$. Also using (4.7) and (4.9), one can verify that

$$\mathbf{\Gamma}_{i,j} =(\mathbf{S}_j\bar{\mathbf{F}} \otimes \tilde{\mathbf{\Lambda}}_i^{\frac{1}{2}})(\sigma_i^2\mathbf{I} + \bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}} \otimes \tilde{\mathbf{\Lambda}}_i)^{-1}(\bar{\mathbf{F}}^H\mathbf{S}_j^T \otimes \tilde{\mathbf{\Lambda}}_i^{\frac{1}{2}}) \tag{4.35}$$

$$\mathbf{\Gamma}_{T,j,i} =(\tilde{\mathbf{\Lambda}}_j^{\frac{1}{2}} \otimes \mathbf{S}_i\bar{\mathbf{F}})(\sigma_j^2\mathbf{I} + \tilde{\mathbf{\Lambda}}_j \otimes \bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(j)}^T\bar{\mathbf{S}}_{(j)}\bar{\mathbf{F}})^{-1}(\tilde{\mathbf{\Lambda}}_j^{\frac{1}{2}} \otimes \bar{\mathbf{F}}^H\mathbf{S}_i^T). \tag{4.36}$$

The rank constraint on $\bar{\mathbf{P}}$ is satisfied by using $\bar{\mathbf{F}}$ defined in (4.7). With (4.35) and (4.36), we have

$$I_M = -\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\log_2|\mathbf{I} - \mathbf{\Gamma}_{i,j}\mathbf{\Gamma}_{T,j,i}| \tag{4.37}$$

and (4.25) becomes

$$\begin{aligned}
&\max_{\bar{\mathbf{F}}} \quad I_M\\
&s.t.\ Tr(\mathbf{S}_i\bar{\mathbf{R}}^{-\frac{H}{2}}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{R}}^{-\frac{1}{2}}\mathbf{S}_i^T) \leq KP_i,\ i = 1,\dots,M.
\end{aligned} \tag{4.38}$$

To solve (4.38) by using the logarithmic barrier method, we let

$$g_2(\bar{\mathbf{F}}) = tI_M + \sum_{i=1}^{M} \mathcal{B}_i(\bar{\mathbf{F}}) \tag{4.39}$$

where $t$ is the barrier coefficient and $\mathcal{B}_i(\bar{\mathbf{F}})$ is shown in (4.13). Then we can solve (4.38) by solving the following (with an increasing $t$):

$$\max_{\bar{\mathbf{F}}} \quad g_2(\bar{\mathbf{F}}). \tag{4.40}$$

The algorithm to solve (4.40) is similar to Algorithm 4 and hence omitted here. The way to find the gradient of $g_2(\bar{\mathbf{F}})$ is shown in Appendix 4.7.2.

### 4.4.2 Special algorithm for $M = 2$

For $M = 2$, the problem is similar to one addressed in [38] where an algorithm was developed and its local optimality is stated there. In this following, we effectively readdress the same problem but show some new insights. One of them is the establishment of optimality of two matrices heuristically chosen in [38]. Furthermore, we will present an asymptotical analyses to show the globally optimal solution in high or low power region.

For $M = 2$, we know $\bar{\mathbf{S}}_{(1)} = \mathbf{S}_2$ and $\bar{\mathbf{S}}_{(2)} = \mathbf{S}_1$. Using (4.17), (4.35) and (4.36), we have

$$\begin{aligned}
&\boldsymbol{\Gamma}_{1,2} \\
&= (\mathbf{S}_2\bar{\mathbf{F}} \otimes \tilde{\boldsymbol{\Lambda}}_1^{\frac{1}{2}})(\sigma_1^2\mathbf{I} + \bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(1)}^T\bar{\mathbf{S}}_{(1)}\bar{\mathbf{F}} \otimes \tilde{\boldsymbol{\Lambda}}_1)^{-1}(\bar{\mathbf{F}}^H\mathbf{S}_2^T \otimes \tilde{\boldsymbol{\Lambda}}_1^{\frac{1}{2}}) \\
&= (\mathbf{U}_2 \otimes \mathbf{I})(\boldsymbol{\Lambda}_2 \otimes \tilde{\boldsymbol{\Lambda}}_1^{\frac{1}{2}})(\sigma_1^2\mathbf{I} + \boldsymbol{\Lambda}_2^2 \otimes \tilde{\boldsymbol{\Lambda}}_1)^{-1}(\boldsymbol{\Lambda}_2^T \otimes \tilde{\boldsymbol{\Lambda}}_1^{\frac{1}{2}})(\mathbf{U}_2^H \otimes \mathbf{I}) \qquad (4.41)
\end{aligned}$$

$$\begin{aligned}
&\boldsymbol{\Gamma}_{T,2,1}\\
&= (\tilde{\boldsymbol{\Lambda}}_2^{\frac{1}{2}} \otimes \mathbf{S}_1\bar{\mathbf{F}})(\sigma_2^2\mathbf{I} + \tilde{\boldsymbol{\Lambda}}_2 \otimes \bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(2)}^T\bar{\mathbf{S}}_{(2)}\bar{\mathbf{F}})^{-1}(\tilde{\boldsymbol{\Lambda}}_2^{\frac{1}{2}} \otimes \bar{\mathbf{F}}^H\mathbf{S}_1^T)\\
&= (\mathbf{I} \otimes \mathbf{U}_1)(\tilde{\boldsymbol{\Lambda}}_2^{\frac{1}{2}} \otimes \boldsymbol{\Lambda}_1)(\sigma_2^2\mathbf{I} + \tilde{\boldsymbol{\Lambda}}_2 \otimes \boldsymbol{\Lambda}_1^T\boldsymbol{\Lambda}_1)^{-1}(\tilde{\boldsymbol{\Lambda}}_2^{\frac{1}{2}} \otimes \boldsymbol{\Lambda}_1^T)(\mathbf{I} \otimes \mathbf{U}_1^H). \qquad (4.42)
\end{aligned}$$

It is obvious that both $I_2 = I(\mathbf{y}_1; \mathbf{y}_{T,2}) = -\log_2 |\mathbf{I} - \boldsymbol{\Gamma}_{1,2}\boldsymbol{\Gamma}_{T,2,1}|$ and $Tr(\mathbf{P}_i\mathbf{P}_i^H)$ are invariant to $\mathbf{V}_i$ in (4.17) where $i = 1, 2$. We can set $\mathbf{V}_i = \mathbf{I}_r$. Now (4.38) becomes

$$\begin{aligned}
&\max_{\mathbf{U}_1,\mathbf{U}_2,\boldsymbol{\Lambda}_1,\boldsymbol{\Lambda}_2} \quad I_2 \qquad (4.43)\\
&s.t.\ Tr(\tilde{\boldsymbol{\Lambda}}_1^{-1}\mathbf{U}_1\boldsymbol{\Lambda}_1^2\mathbf{U}_1^H) \leq KP_1,\ Tr(\tilde{\boldsymbol{\Lambda}}_2^{-1}\mathbf{U}_2\boldsymbol{\Lambda}_2^2\mathbf{U}_2^H) \leq KP_2\\
&\boldsymbol{\Lambda}_1 \succ \mathbf{0},\ \boldsymbol{\Lambda}_2 \succ \mathbf{0}.
\end{aligned}$$

Here we have added the positive definite constraints on $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are mild constraints and it leads to the optimal $\mathbf{U}_1$ and $\mathbf{U}_2$ being the identity matrices as shown next.

With $\boldsymbol{\Lambda}_1 \succ \mathbf{0}$ and $\boldsymbol{\Lambda}_2 \succ \mathbf{0}$, (4.42) and (4.41) become $\boldsymbol{\Gamma}_{1,2} = (\mathbf{I} + \sigma_1^2(\mathbf{U}_2\boldsymbol{\Lambda}_2^2\mathbf{U}_2^H \otimes \tilde{\boldsymbol{\Lambda}}_1)^{-1})^{-1}$ and $\boldsymbol{\Gamma}_{T,2,1} = (\mathbf{I} + \sigma_2^2(\tilde{\boldsymbol{\Lambda}}_2 \otimes \mathbf{U}_1\boldsymbol{\Lambda}_1^2\mathbf{U}_1^H)^{-1})^{-1}$, and then the cost function in (4.43) becomes

$$
\begin{aligned}
I_2 &= \log_2 \left|\mathbf{I} + \sigma_2^2(\tilde{\mathbf{\Lambda}}_2 \otimes \tilde{\mathbf{U}}_1\mathbf{\Lambda}_1^2\tilde{\mathbf{U}}_1^H)^{-1}\right| + \log_2 \left|\mathbf{I} + \sigma_1^2(\mathbf{U}_2\mathbf{\Lambda}_2^2\mathbf{U}_2^H \otimes \tilde{\mathbf{\Lambda}}_1)^{-1}\right| \\
&\quad - \log_2 \left|(\mathbf{I} + \sigma_2^2(\tilde{\mathbf{\Lambda}}_2 \otimes \tilde{\mathbf{U}}_1\mathbf{\Lambda}_1^2\tilde{\mathbf{U}}_1^H)^{-1})(\mathbf{I} + \sigma_1^2(\mathbf{U}_2\mathbf{\Lambda}_2^2\mathbf{U}_2^H \otimes \tilde{\mathbf{\Lambda}}_1)^{-1}) - \mathbf{I}\right| && \text{(4.44a)} \\
&= \log_2 |\sigma_2^2\mathbf{I} + \tilde{\mathbf{\Lambda}}_2 \otimes \mathbf{\Lambda}_1^2| + \log_2 |\sigma_1^2\mathbf{I} + \mathbf{\Lambda}_2^2 \otimes \tilde{\mathbf{\Lambda}}_1| \\
&\quad - \log_2 |\sigma_1^2\sigma_2^2\mathbf{I} + \sigma_1^2\tilde{\mathbf{\Lambda}}_2 \otimes \tilde{\mathbf{U}}_1\mathbf{\Lambda}_1^2\tilde{\mathbf{U}}_1^H + \sigma_2^2\mathbf{U}_2\mathbf{\Lambda}_2^2\mathbf{U}_2^H \otimes \tilde{\mathbf{\Lambda}}_1| && \text{(4.44b)} \\
&= \log_2 |\sigma_2^2\mathbf{I} + \tilde{\mathbf{\Lambda}}_2 \otimes \mathbf{\Lambda}_1^2| + \log_2 |\sigma_1^2\mathbf{I} + \mathbf{\Lambda}_2^2 \otimes \tilde{\mathbf{\Lambda}}_1| \\
&\quad - \log_2 |\sigma_1^2\sigma_2^2\mathbf{I} + \sigma_1^2\tilde{\mathbf{\Lambda}}_2 \otimes \mathbf{\Lambda}_1^2 + \sigma_2^2\mathbf{U}(\mathbf{\Lambda}_2^2 \otimes \tilde{\mathbf{\Lambda}}_1)\mathbf{U}^H| && \text{(4.44c)}
\end{aligned}
$$

where $\mathbf{U} \triangleq \mathbf{U}_2 \otimes \tilde{\mathbf{U}}_1^H$. Here, (4.44a) 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 (4.44b) is due to $\log_2 |\mathbf{I} + \mathbf{A}^{-1}| = \log_2 |\mathbf{I} + \mathbf{A}| - \log_2 |\mathbf{A}|$. Then the optimal $\mathbf{U}_1$ and $\mathbf{U}_2$ that maximize (4.44) are given by

$$
\begin{aligned}
&\{\mathbf{U}_{1,opt}, \mathbf{U}_{2,opt}\} \\
&= arg \min_{\mathbf{U}_1,\mathbf{U}_2} \log_2 |\sigma_1^2\sigma_2^2\mathbf{I} + \sigma_1^2\tilde{\mathbf{\Lambda}}_2 \otimes \mathbf{\Lambda}_1^2 + \sigma_2^2\mathbf{U}(\mathbf{\Lambda}_2^2 \otimes \tilde{\mathbf{\Lambda}}_1)\mathbf{U}^H|
\end{aligned} \tag{4.45}
$$

According to [52], we have:

**Lemma 8** *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 $\mathbf{\Lambda}_a$, $\mathbf{\Lambda}_c$, $\mathbf{\Lambda}_b$, $\mathbf{\Lambda}_d$ where the diagonal elements in each diagonal matrix are in descending order. Then*

$$
|\mathbf{A} \otimes \mathbf{B} + \mathbf{C} \otimes \mathbf{D}| \geq \min_{P_1,P_2} |\mathbf{\Lambda}_a \otimes \mathbf{\Lambda}_b + \mathbf{\Lambda}_{c,P_1} \otimes \mathbf{\Lambda}_{d,P_2}| \tag{4.46a}
$$

$$
|\mathbf{A} \otimes \mathbf{B} + \mathbf{C} \otimes \mathbf{D}| \leq \max_{P_1,P_2} |\mathbf{\Lambda}_a \otimes \mathbf{\Lambda}_b + \mathbf{\Lambda}_{c,P_1} \otimes \mathbf{\Lambda}_{d,P_2}| \tag{4.46b}
$$

*where the minimum or maximum are taken over all possible (diagonal-wise) permutations $\{P_1, P_2\}$.*

From Lemma 8, we have:

**Lemma 9** *Let* $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ *be positive semi-definite Hermitian matrices with the corresponding eigenvalue matrices* $\mathbf{\Lambda}_a$*,* $\mathbf{\Lambda}_b$*,* $\mathbf{\Lambda}_c$*,* $\mathbf{\Lambda}_d$ *each of descending diagonal elements. Then*

$$|\mathbf{A} \otimes \mathbf{B} + \mathbf{C} \otimes \mathbf{D}| \geq |\mathbf{\Lambda}_a \otimes \mathbf{\Lambda}_b + \mathbf{\Lambda}_c \otimes \mathbf{\Lambda}_d| \tag{4.47a}$$

$$|\mathbf{A} \otimes \mathbf{B} + \mathbf{C} \otimes \mathbf{D}| \leq |\mathbf{\Lambda}_a \otimes \mathbf{\Lambda}_b + \bar{\mathbf{\Lambda}}_c \otimes \bar{\mathbf{\Lambda}}_d| \tag{4.47b}$$

*where* $\bar{\mathbf{\Lambda}}_c$ *and* $\bar{\mathbf{\Lambda}}_d$ *are respectively* $\mathbf{\Lambda}_c$ *and* $\mathbf{\Lambda}_d$ *but with reversed order of diagonal elements.*

**Proof.** See Appendix 4.7.3 ■

Applying (4.47a) to (4.45) and from (4.20), we have:

**Theorem 10** *For* $M = 2$*,* $\mathbf{U}_{1,opt} = \mathbf{I}$ *and* $\mathbf{U}_{2,opt} = \mathbf{I}$ *are respectively the globally optimal solutions of* $\mathbf{U}_1$ *and* $\mathbf{U}_2$ *(defined in* (4.17)*) to the MI based problem* (4.43)*.*

The above choices of $\mathbf{U}_1$ and $\mathbf{U}_2$ were also used in [38] but they could not establish their optimality. Also note that the optimality of the above choice of $\mathbf{U}_1$ and $\mathbf{U}_2$ was rather obvious (see the discussions of (4.19) and (4.20)) for the MSE based problem (4.6).

Let $\mathbf{C}_1 = \tilde{\mathbf{\Lambda}}_1^{-1}\mathbf{\Lambda}_1^2$ and $\mathbf{C}_2 = \tilde{\mathbf{\Lambda}}_2^{-1}\mathbf{\Lambda}_2^2$ with their diagonal elements denoted by $c_{1,l} = \lambda_{1,l}^2/\tilde{\lambda}_{1,l}$ and $c_{2,k} = \lambda_{2,k}^2/\tilde{\lambda}_{2,k}$. Then (4.44c) becomes

$$\begin{aligned}
&I_2\\
&= \log_2 |\sigma_2^2\mathbf{I} + \tilde{\mathbf{\Lambda}}_2 \otimes \mathbf{C}_1\tilde{\mathbf{\Lambda}}_1| + \log_2 |\sigma_1^2\mathbf{I} + \mathbf{C}_2\tilde{\mathbf{\Lambda}}_2 \otimes \tilde{\mathbf{\Lambda}}_1|\\
&\quad - \log_2 |\sigma_1^2\sigma_2^2\mathbf{I} + \sigma_1^2\tilde{\mathbf{\Lambda}}_2 \otimes \mathbf{C}_1\tilde{\mathbf{\Lambda}}_1 + \sigma_2^2\mathbf{C}_2\tilde{\mathbf{\Lambda}}_2 \otimes \tilde{\mathbf{\Lambda}}_1|\\
&= \sum_{k=1}^{N_2}\sum_{l=1}^{N_1} \log_2 \left( \frac{(\sigma_2^2 + \tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}c_{1,l})(\sigma_1^2 + \tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}c_{2,k})}{\sigma_1^2\sigma_2^2 + \sigma_1^2\tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}c_{1,l} + \sigma_2^2\tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}c_{2,k}} \right)\\
&\triangleq \sum_{k=1}^{N_2}\sum_{l=1}^{N_1} f_{l,k}(c_{1,l}, c_{2,k})
\end{aligned} \tag{4.48}$$

Let $\mathbf{c}_1$ and $\mathbf{c}_2$ be the vectors of the diagonal elements from $\mathbf{C}_1$ and $\mathbf{C}_2$ respectively. Then (4.43) is transformed to

$$\begin{aligned}
\max_{\mathbf{c}_1>\mathbf{0},\mathbf{c}_2>\mathbf{0}} \quad & \sum_{k=1}^{N_2}\sum_{l=1}^{N_1} f_{l,k}(c_{1,l}, c_{2,k})\\
s.t. \quad & \sum_{l=1}^{N_1} c_{1,l} \leq KP_1, \ \sum_{k=1}^{N_2} c_{2,k} \leq KP_2
\end{aligned} \tag{4.49}$$

It is easy to verify that $f(c_{1,l}, c_{2,k})$ is a monotonically increasing function of $c_{1,l}$ and $c_{2,k}$ respectively. So, the optimal solutions must satisfy $\sum_{l=1}^{N_1} c_{1,l} = KP_1$ and $\sum_{k=1}^{N_2} c_{2,k} = KP_2$.

However, $-f_{l,k}(c_{1,l}, c_{2,k})$ is not always convex of $c_{1,l}$ and $c_{2,k}$. The Hessian matrix of $-f_{l,k}(c_{1,l}, c_{2,k})$ is

$$\begin{bmatrix} \frac{\tilde{\lambda}_{1,l}^2\tilde{\lambda}_{2,k}^2(\vartheta_{l,k}-\sigma_1^4\theta_{1,l,k})}{\theta_{1,l,k}\vartheta_{l,k}} & -\frac{\sigma_1^2\sigma_2^2\tilde{\lambda}_{1,l}^2\tilde{\lambda}_{2,k}^2}{\vartheta_{l,k}} \\ -\frac{\sigma_1^2\sigma_2^2\tilde{\lambda}_{1,l}^2\tilde{\lambda}_{2,k}^2}{\vartheta_{l,k}} & \frac{\tilde{\lambda}_{1,l}^2\tilde{\lambda}_{2,k}^2(\vartheta_{l,k}-\sigma_2^4\theta_{2,l,k})}{\theta_{2,l,k}\vartheta_{l,k}} \end{bmatrix} \tag{4.50}$$

where $\theta_{1,l,k} = (\sigma_2^2 + \tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}c_{1,l})^2$, $\theta_{2,l,k} = (\sigma_1^2 + \tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}c_{2,k})^2$ and $\vartheta_{l,k} = (\sigma_1^2\sigma_2^2 + \sigma_1^2\tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}c_{1,l} + \sigma_2^2\tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}c_{2,k})^2$. This matrix is positive semidefinite if and only if $c_{1,l}c_{2,k} \geq \frac{\sigma_1^2\sigma_2^2}{2\tilde{\lambda}_{1,l}^2\tilde{\lambda}_{2,k}^2}$. This means that when $KP_1$ and $KP_2$ are large, the Hessian matrix of $-f_{l,k}(c_{1,l}, c_{2,k})$ is typically positive definite and hence $-f_{l,k}(c_{1,l}, c_{2,k})$ is typically convex. In this high power case, the problem (4.49) is convex and the globally optimal solution is available. In general, $-f_{l,k}(c_{1,l}, c_{2,k})$ is a convex function with respect to $c_{1,l}$ and $c_{2,k}$ individually. To obtain locally optimal solution to (4.49), we can apply a two-phase iteration method, i.e., optimizing $\mathbf{c}_1$ and $\mathbf{c}_2$ alternately until convergence. The discussion of the following two-phase algorithm is similar to that in [38].

In phase one, the Lagrangian function with respect to $c_{1,l}$ is

$$\mathcal{L} = \sum_{k=1}^{N_2}\sum_{l=1}^{N_1} f_{l,k}(c_{1,l}, c_{2,k}) - \mu\Big(\sum_{l=1}^{N_1} c_{1,l} - KP_1\Big) + \boldsymbol{\alpha}^T \mathbf{c}_1 \tag{4.51}$$

And the corresponding KKT conditions are

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial c_{1,l}} = \dfrac{1}{\ln 2}\displaystyle\sum_{k=1}^{N_2} f'_{l,k}(c_{1,l}, c_{2,k}) - \mu = 0 \\ \displaystyle\sum_{l=1}^{N_1} c_{1,l} \leq KP_1,\ \mu(\sum_{l=1}^{N_1} c_{1,l} - KP_1) = 0,\ \mu \geq 0 \\ \mathbf{c}_1 > \mathbf{0},\ \boldsymbol{\alpha}^T\mathbf{c}_1 = 0,\ \boldsymbol{\alpha} \geq \mathbf{0} \end{cases} \tag{4.52}$$

where

$$\begin{aligned} &f'_{l,k}(x,y) \\ &= \frac{\sigma_2^2 \tilde{\lambda}_{1,l}^2 \tilde{\lambda}_{2,k}^2 y}{(\sigma_2^2 + \tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}x)(\sigma_1^2\sigma_2^2 + \sigma_1^2\tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}x + \sigma_2^2\tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}y)} \end{aligned} \tag{4.53}$$

In phase two, similar KKT conditions can be found. From (4.52), we see that $\mu$ is a monotonically decreasing function of $c_{1,l}$. Therefore, we can use a bisection search to solve (4.52). An efficient algorithm to solve (4.49) is shown in Algorithm 5. From (4.53), we know that $f'_{l,k}(c_{1,l}, c_{2,k})$ is an increasing function of $\tilde{\lambda}_{1,l}$ and a decreasing function of $c_{1,l}$. Given any $\mathbf{c}_2$, the solution from (4.52) is $\mathbf{c}_1^*$, which must satisfy $\sum_{k=1}^{N_2} f'_{l,k}(c_{1,l}^*, c_{2,k}) = \mu \ln 2$. Hence, one can verify that $c_{1,l}^* \geq c_{1,l+1}^*$. (If $c_{1,l}^* < c_{1,l+1}^*$ then $\mu \ln 2 = \sum_{k=1}^{N_2} f'_{l,k}(c_{1,l}^*, c_{2,k}) > \sum_{k=1}^{N_2} f'_{l,k}(c_{1,l+1}^*, c_{2,k}) \geq \sum_{k=1}^{N_2} f'_{l+1,k}(c_{1,l+1}^*, c_{2,k}) = \mu \ln 2$, which is not possible.) Similarly,

**Algorithm 5** Bisection section search to solve (4.52)

```
Input:
    $\tilde{\mathbf{\Lambda}}_1, \tilde{\mathbf{\Lambda}}_2, P_1, P_2$, K;
    Accuracy threshold $\epsilon_1$, $\epsilon_2$.
    Initialization $p = 0$, $\mathbf{c}_1^{(p)} = \frac{KP_1}{N_1}\mathbf{1}_{N_1}, \mathbf{c}_2^{(p)} = \frac{KP_2}{N_2}\mathbf{1}_{N_2}$.
repeat
    Given $\mathbf{c}_2^{(p)}$, do bisection search of $\mu$ and obtain solution $\mathbf{c}_1^{(p+1)}$ to meet the power constraint $|\sum_{l=1}^{N_1} c_{1,l} - KP_1| \leq \epsilon_1$; Given $\mathbf{c}_1^{(p+1)}$, do bisection search of $\nu$ and obtain solution $\mathbf{c}_2^{(p+1)}$ to meet the power constraint $|\sum_{k=1}^{N_2} c_{2,k} - KP_2| \leq \epsilon_1$.
    $p = p + 1$.
until $\|[\mathbf{c}_1^{(p)}, \mathbf{c}_2^{(p)}] - [\mathbf{c}_1^{(p-1)}, \mathbf{c}_2^{(p-1)}]\| \leq \epsilon_2$
return $\{\mathbf{c}_1^{(p)}, \mathbf{c}_2^{(p)}\}$
```

$c_{2,k}^* \geq c_{2,k+1}^*$. Therefore, the diagonal elements of the optimal solutions of $\mathbf{\Lambda}_1^2$ and $\mathbf{\Lambda}_2^2$ are also in descending order respectively.

### Asymptotic Analysis

The following theorem shows the globally optimal solution to (4.25) in high or low power region. These solutions are also given by Algorithm 5.

**Theorem 11** *Let $P_1 = P_2 = P$. If $P$ is arbitrarily large, the globally optimal $c_{1,l}$ and $c_{2,k}$ (defined before (4.48)) are invariant to $l$ and $k$ (which will be called "uniform power" allocation), and a less correlated channel yields a higher secret key rate. If $P$ is arbitrarily small, the globally optimal $c_{1,l}$ and $c_{2,k}$ are all arbitrarily small except for $l = k = 1$, and a higher correlated channel yields a higher secret key rate.*

**Proof.** See Appendix 4.7.4. ■

### 4.4.3 Closed-form solution

For $M \geq 2$, we now consider the same symmetric and isotropic case considered before. Without loss of generality, also let $\sigma = 1$. Then applying the matrix inverse lemma to (4.35) and (4.36), we have

$$\mathbf{\Gamma}_{i,j} = (\mathbf{S}_j\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_j^T \otimes \mathbf{I}) - \big((\mathbf{S}_j\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T)(\mathbf{I} + \bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T)^{-1}(\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_j^T)\big) \otimes \mathbf{I} \tag{4.54}$$

$$\mathbf{\Gamma}_{T,j,i} = (\mathbf{I} \otimes \mathbf{S}_i\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_i^T) - \mathbf{I} \otimes \big((\mathbf{S}_i\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(j)}^T)(\mathbf{I} + \bar{\mathbf{S}}_{(j)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(j)}^T)^{-1}(\bar{\mathbf{S}}_{(j)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_i^T)\big) \tag{4.55}$$

Note that $I(\mathbf{y}_i; \mathbf{y}_{T,j}) = -\log_2 |\mathbf{I} - \mathbf{\Gamma}_{i,j}\mathbf{\Gamma}_{T,j,i}|$, $I_M = \sum_{i=1}^{M-1}\sum_{j=i+1}^{M} I(\mathbf{y}_i; \mathbf{y}_{T,j})$ and the power and rank constraints in (4.25) become $Tr(\mathbf{S}_i\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_i^T) \leq KP,\ i = 1, \ldots, M$. Then the Lagrangian function is now

$$\mathcal{L} = I_M - \sum_{i=1}^{M} \mu_i(Tr(\mathbf{S}_i\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_i^T) - KP) \tag{4.56}$$

and the KKT conditions are

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial \bar{\mathbf{F}}} = \dfrac{\partial I_M}{\partial \bar{\mathbf{F}}} - \displaystyle\sum_{i=1}^{M} 2\mu_i \mathbf{S}_i^T\mathbf{S}_i\bar{\mathbf{F}} = 0 \\ Tr(\mathbf{S}_i\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_i^T) \leq KP,\ i = 1, \ldots, M \\ \mu_i(Tr(\mathbf{S}_i\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_i^T) - KP) = 0,\ \mu_i \geq 0,\ i = 1, \ldots, M \end{cases} \tag{4.57}$$

**Theorem 12** *The solutions to* (4.23) *as shown in Theorem 6 are also solutions to* (4.57).

**Proof.** See Appendix 4.7.5. ■

For $M = 2$, the pilots from this theorem satisfy $\mathbf{P}_i\mathbf{P}_i^H = \frac{KP}{N}\mathbf{I}_N$ where $i = 1, 2$, and these pilots are known to be globally optimal for maximal MI [53] under the symmetric and isotropic condition. Also note that for $M \geq 3$, our numerical simulations did not yield any result better than that from Theorem 12 subject to the symmetric and isotropic condition.

## 4.5 Simulation results

To show some simulation results, we let $P_i = P$, $\sigma_i^2 = 1$, $N_i = 4$, $\mathbf{R}_i = \mathbf{R}$, $r = (M-1)N$ and $K \geq r$. We choose the channel correlation matrix to be such that $(\mathbf{R})_{l,k} = R^{|l-k|}$ where $R \in [0,1]$ is the correlation coefficient.

We first use the normalized (i.e., per element of each channel matrix) MSE:

$$\mathcal{J}_M = \frac{J_M}{M(M-1)N^2} \tag{4.58}$$

to compare three different choices of pilots. Since $\mathcal{J}_M$ depends on $R$, we will also write $\mathcal{J}_M = \mathcal{J}_M(R)$. More specifically, we use $\mathcal{J}_{M,opt}(R)$ for the optimal pilot computed from algorithm 4, $\mathcal{J}_{M,c-opt}(R)$ for the conditionally optimal pilot from Theorem 6, and $\mathcal{J}_{M,first}(R)$ for the pilot proposed in [31] (which coincides with that from Theorem 6 if $N_i = N = 1$).

For $M = 3$, Fig. 4.2 shows the normalized MSE vs $0\texttt{dB} \leq KP \leq 70\texttt{dB}$. We see that for high $KP$ all curves of the normalized MSE in log-scale vs $KP$ in dB become parallel straight lines. This is expected since for large enough $KP$ the MSE is proportional to $\frac{1}{KP}$. It is also expected that $\mathcal{J}_{M,opt}(0) = \mathcal{J}_{M,c-opt}(0)$. But we also see that $\mathcal{J}_{M,opt}(R)$ and $\mathcal{J}_{M,c-opt}(R)$ are still rather close to each other even for $R = 0.8$, and they both are substantially better than $\mathcal{J}_{M,first}(R)$ especially at high $KP$.

Using the pilots from Theorem 6, we know that $J_{M,opt}(0) = N\sum_{i=1}^{M} Tr\big((\mathbf{I} + \frac{KP}{N^2(M-1)}\bar{\mathbf{S}}_{(i)}\bar{\mathbf{Q}}_m\bar{\mathbf{Q}}_m^H\bar{\mathbf{S}}_{(i)}^T)^{-1}\big)$, and hence one can verify that

$$\lim_{KP\to\infty} \mathcal{J}_{M,opt}(0) = 2N(1-\frac{1}{M})\frac{1}{KP} \tag{4.59}$$

which is invariant to large $M$. But this limit increases linearly as $N$ increases (because the per-antenna power is $\frac{P}{N}$).

Fig. 4.3 shows $\frac{\mathcal{J}_{M,opt}(0.8)}{\mathcal{J}_{M,opt}(0)}$ vs $M$ and $N$ where $KP = 60$dB. Note that $\frac{\mathcal{J}_{M,opt}(0.8)}{\mathcal{J}_{M,opt}(0)}$ is invariant to large $KP$. From this and other similar plots that we have obtained but not shown here, we see that $\mathcal{J}_{M,opt}(R)$ is also invariant to large $M$ but increases as $N$ increases. Furthermore, $\mathcal{J}_{M,opt}(R)$ increases as $R$ increases within $[0,1)$ in the high power region.

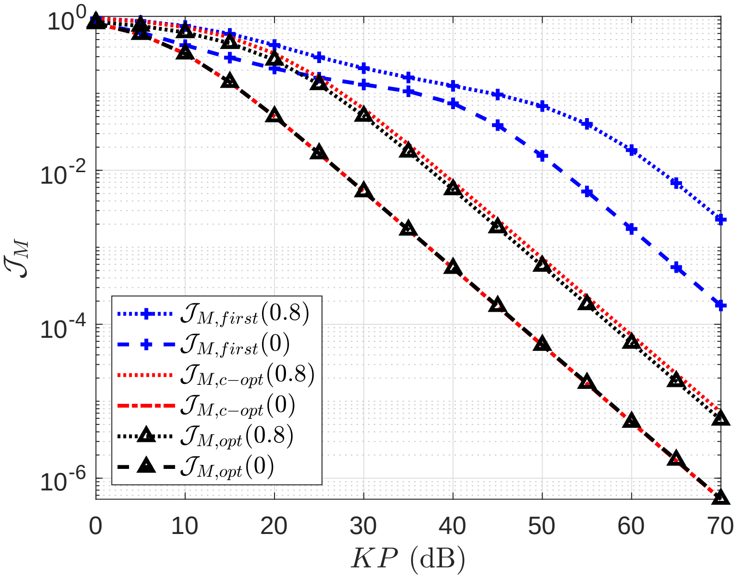

Figure 4.2: Normalized MSE vs $0\texttt{dB} \le KP \le 70\texttt{dB}$ where $M = 3$.

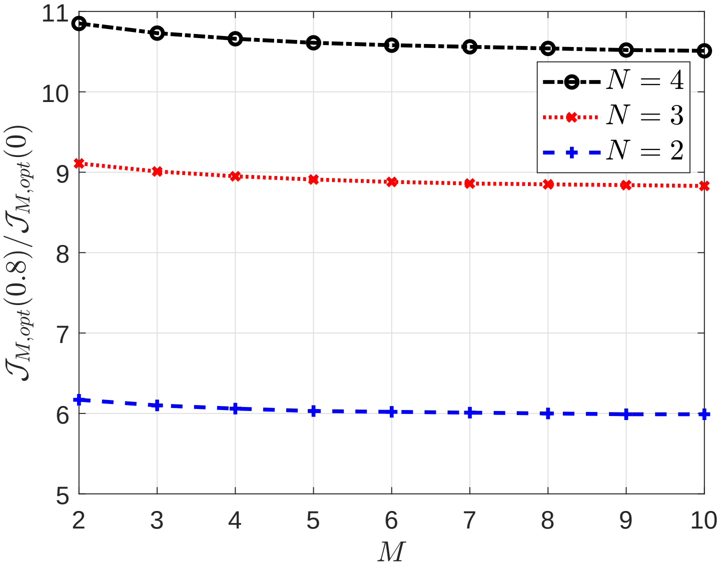

Figure 4.3: $\frac{\mathcal{J}_{M,opt}(0.8)}{\mathcal{J}_{M,opt}(0)}$ vs $M$ and $N$ with $KP = 60\texttt{dB}$.

We now use the normalized (per pair and per degree-of-freedom) MI:

$$\mathcal{I}_M = \frac{I_M}{\frac{M(M-1)N^2}{2}} \tag{4.60}$$

to compare three different choices of pilots. We also write $\mathcal{I}_M = \mathcal{I}_M(R)$. We use $\mathcal{I}_{M,opt}(R)$ for the pilots from (4.25), $\mathcal{I}_{M,c-opt}(R)$ for the pilots from Theorem 12, and $\mathcal{I}_{M,first}(R)$ for the pilots initially suggested in [31].

For $M = 3$, Fig. 4.4 shows $\mathcal{I}_M(R)$ vs $0\texttt{dB} \leq KP \leq 70\texttt{dB}$. Since $\mathcal{I}_M(R)$ is a constant plus $\log_2(KP)$ at high $KP$, we see that all curves here become parallel straight lines when $KP$ is large. Like the MSE case, we also see here that $\mathcal{I}_{M,opt}(R)$ and $\mathcal{I}_{M,c-opt}(R)$ are significantly better than $\mathcal{I}_{M,first}(R)$.

One can verify by using (4.86) and $I_M(0) = -N^2 \log_2(1-\Gamma^2)$ that

$$\lim_{KP\to\infty} \mathcal{I}_{M,opt}(0) = \log_2(\frac{1}{4N}(1+\frac{1}{M-1})) + \log_2(KP) \tag{4.61}$$

which is invariant to large $M$ but decreases as $N$ increases.

Fig. 4.5 shows $\mathcal{I}_{M,opt}(0.8) - \mathcal{I}_{M,opt}(0)$ vs $M$ and $N$ where $KP = 60$dB. Note that $\mathcal{I}_{M,opt}(0.8) - \mathcal{I}_{M,opt}(0)$ is invariant to large $KP$. From this and other similar plots not shown here, we see that $\mathcal{I}_{M,opt}(R)$ is also invariant to large $M$ but decreases as $N$ increases. And $\mathcal{I}_{M,opt}(R)$ decreases as $R$ increases within $[0, 1)$ in the high power region.

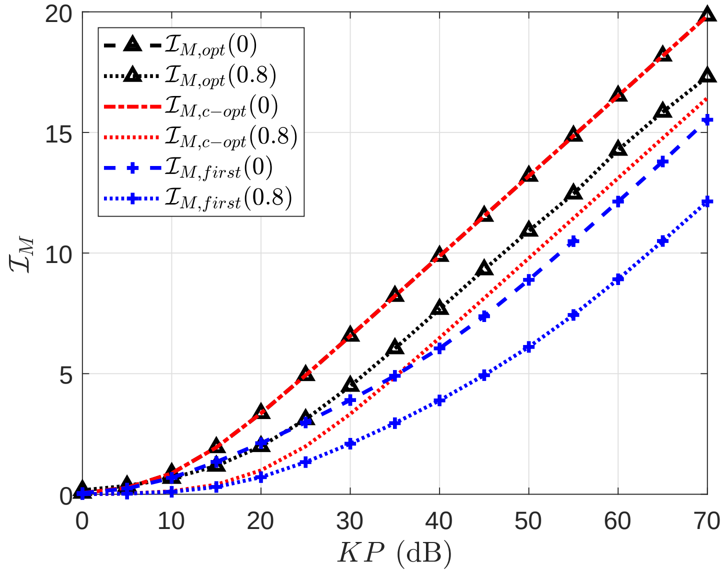

Figure 4.4: Normalized MI $0\texttt{dB} \leq KP \leq 70\texttt{dB}$ with $M = 3$.

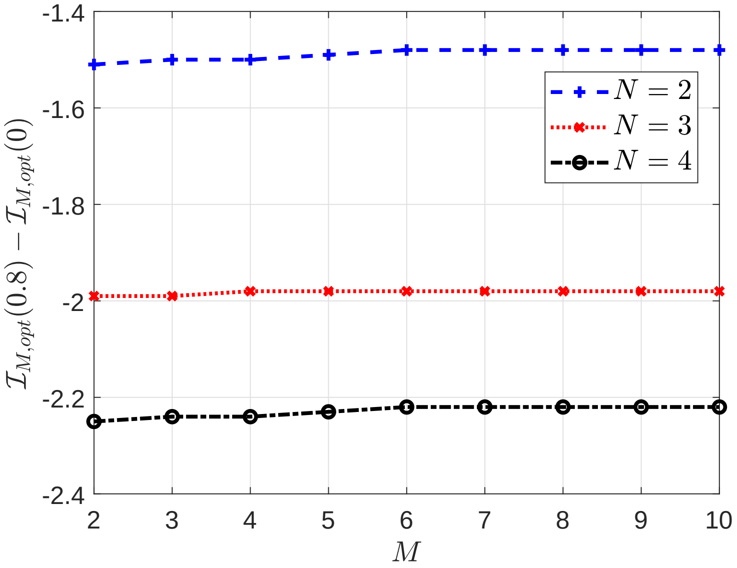

Figure 4.5: $\mathcal{I}_{M,opt}(0.8) - \mathcal{I}_{M,opt}(0)$ vs $M$ and $N$ with $KP = 60\texttt{dB}$.

Finally, let us consider $\mathcal{I}_2 = \frac{I_2}{N^2} = \frac{I(\mathbf{y}_1;\mathbf{y}_2)}{N^2}$ for two users ($M = 2$) based on three choices of pilots, i.e., 1) $\mathcal{I}_{2,opt}(R)$ based on (4.49) which maximizes the mutual information; 2) $\mathcal{I}_{2,MSE}(R)$ based on two-use MMSE channel estimation as in [37]; and 3) $\mathcal{I}_{2,u}(R)$ based on "uniform power" allocation, i.e., $\mathbf{c}_1 = \mathbf{c}_2 = \frac{KP}{N}\mathbf{1}$.

Fig. 4.6 shows $\mathcal{I}_{2,opt}(R)$ (in bits per realization of $\mathbf{H}_{1,2}$) vs $KP$ where $R = 0$ and $R = 0.8$. As expected from the analyses, we see that in the low power region, a higher correlation yields a higher secret key rate, but in the high power region, the opposite is true.

Fig. 4.7 shows $\frac{\mathcal{I}_{2,MSE}(R)}{\mathcal{I}_{2,opt}(R)}$ and $\frac{\mathcal{I}_{2,u}(R)}{\mathcal{I}_{2,opt}(R)}$ vs $KP$ where $R = 0.8$. As expected from theorem 11, we see that as power increases, the uniform power pilots become closer to the optimal, i.e. $\frac{\mathcal{I}_{2,u}(R)}{\mathcal{I}_{2,opt}(R)}$ increases to one. We also see that the pilots based on MMSE channel estimation are nearly optimal in the low power case. This is because when the power becomes low the MMSE based pilot design also allocates all the power to the strongest stream. But the MMSE based pilot design does not lead to uniform power allocation in the high power case [37], which explains the gap at high power. The curve of $\frac{\mathcal{I}_{2,MSE}(R)}{\mathcal{I}_{2,opt}(R)}$ shown here is supported by Theorem 11 but differs from Fig. 7 in [38], the latter of which appears to have an error.

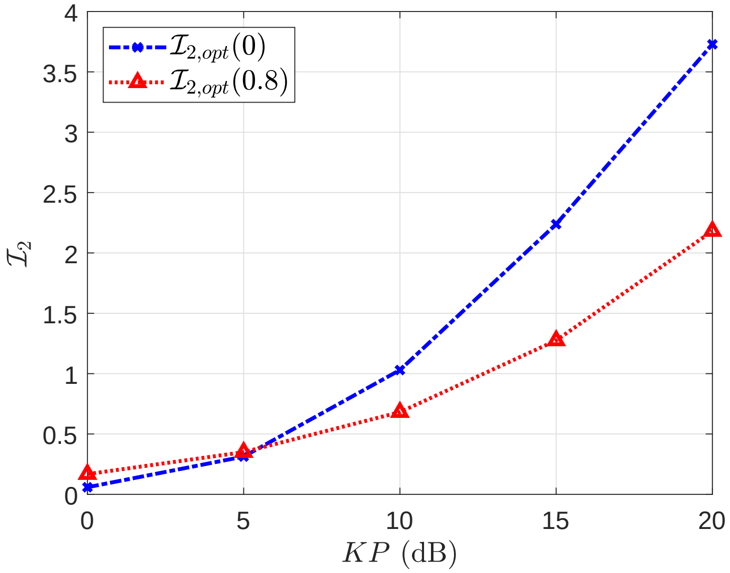

Figure 4.6: Normalized MI with $M = 2$

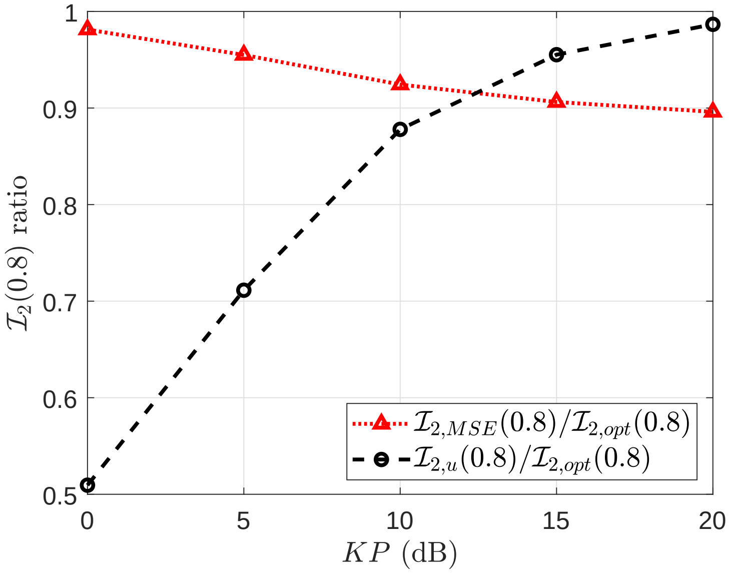

Figure 4.7: Normalized MI ratio with $M = 2$ and correlation $R = 0.8$

## 4.6 Conclusion

In this chapter, we have developed algorithms for computing the optimal pilots used for ANECE under optimal MMSE channel estimation and maximum MI criteria. Each channel matrix is modelled by a known correlation matrix and a matrix of i.i.d. complex Gaussian entries. While the logarithmic barrier gradient method was used to develop algorithms for more than two users, more efficient algorithms were developed for two users. Under the symmetric and isotropic condition, closed-form expression of the optimal pilots was shown (in Theorems 6 and 12) under both optimal MMSE channel estimation and maximum MI criteria. While this closed-form expression coincides with that proposed in [31] for three or more single-antenna users, it is a significant discovery for three or more multi-antenna users. The general algorithms developed for three or more multi-antenna users are also significant contributions beyond the prior works shown in [37] and [38]. We should note however that although the optimal pilots developed in this paper meet the KKT conditions of non-convex problems and there is no other known design that performs better, the global optimality of the optimal pilots from this work is not yet established for most situations such as three or more users. One strategy to prove the global optimality (if true) of the solutions in Theorems 6 and 12 is to find all solutions to the KKT conditions of the non-convex problems and rule out the possibility of better solutions. This is a challenge not yet won.

## 4.7 Proof of Lemma and Theorem

### 4.7.1 Proof of Theorem 6

From (4.24), the $(l+1,k+1)$th element of $\mathbf{Q}_m\mathbf{Q}_m^H$ is

$$\begin{aligned}(\mathbf{Q}_m\mathbf{Q}_m^H)_{l+1,k+1} &= \sum_{n=0}^{N-1} e^{-j2\pi\frac{(l-k)(m+nM)}{NM}} \\ &= e^{-j2\pi\frac{(l-k)m}{NM}} \sum_{n=0}^{N-1} e^{-j2\pi\frac{(l-k)n}{N}} \\ &= \begin{cases} 0, & |l-k| \neq vN \\ Ne^{-j2\pi\frac{(l-k)m}{NM}}, & |l-k| = vN \end{cases} \end{aligned} \tag{4.62}$$

where $v$ is an integer satisfying $0 \leq v \leq M-1$. From (4.62), we know that there are only $M$ non-zero elements on each column or row of $\mathbf{Q}_m\mathbf{Q}_m^H$. More specifically, using $w_M = e^{-j2\pi\frac{1}{M}}$, we have

$$
\begin{aligned}
&\mathbf{Q}_m\mathbf{Q}_m^H \\
&= N\begin{bmatrix} 1 & w_M^{-m} & \cdots & w_M^{-(M-1)m} \\ w_M^{m} & 1 & \cdots & w_M^{-(M-2)m} \\ \vdots & \vdots & \ddots & \vdots \\ w_M^{(M-1)m} & w_M^{(M-2)m} & \cdots & 1 \end{bmatrix} \otimes \mathbf{I}_N && (4.63)\\
&= N\mathbf{q}_m\mathbf{q}_m^H \otimes \mathbf{I}_N && (4.64)
\end{aligned}
$$

where $\mathbf{q}_m = [1, w_M^m, \ldots, w_M^{(M-1)m}]^T$. Since $\mathbf{Q}_m^H\bar{\mathbf{Q}}_m = 0$, we have $(\mathbf{q}_m\mathbf{q}_m^H \otimes \mathbf{I}_N)\bar{\mathbf{Q}}_m = 0$.

For $N_i = N$, we have $\bar{\mathbf{S}}_{(i)} = \mathbf{I}_{M,i} \otimes \mathbf{I}_N$ where $\mathbf{I}_{M,i}$ is $\mathbf{I}_M$ without its $i$th row, and $\mathbf{S}_i = \mathbf{e}_i^T \otimes \mathbf{I}_N, i = 1, \ldots, M$ where $\mathbf{e}_i$ is the $M \times 1$ vector with its $i$th element equal to one. Now assume $\bar{\mathbf{F}} = \sqrt{\alpha_d}\bar{\mathbf{Q}}_m$. Then $\bar{\mathbf{F}}\bar{\mathbf{F}}^H = \alpha_d\bar{\mathbf{Q}}_m\bar{\mathbf{Q}}_m^H = \alpha_d(MN\mathbf{I}_{MN} - \mathbf{Q}_m\mathbf{Q}_m^H) = \alpha_d(MN\mathbf{I}_{MN} - N\mathbf{q}_m\mathbf{q}_m^H \otimes \mathbf{I}_N) = \alpha_d(MN\mathbf{I}_M - N\mathbf{q}_m\mathbf{q}_m^H) \otimes \mathbf{I}_N$, and

$$
\begin{aligned}
&(\mathbf{I}_{(M-1)N} + \bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T)^{-1} \\
&= [\mathbf{I}_{(M-1)N} + \alpha_d(\mathbf{I}_{M,i} \otimes \mathbf{I}_N)(NM\mathbf{I} - N\mathbf{q}_m\mathbf{q}_m^H \otimes \mathbf{I}_N)(\mathbf{I}_{M,i}^T \otimes \mathbf{I}_N)]^{-1} \\
&= ((1 + NM\alpha_d)\mathbf{I}_{(M-1)N} - N\alpha_d\big(\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{I}_{M,i}^T\big) \otimes \mathbf{I}_N)^{-1} && (4.65)\\
&= \frac{\big(\mathbf{I}_{M-1} - \frac{N\alpha_d}{1+NM\alpha_d}\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{I}_{M,i}^T\big)^{-1} \otimes \mathbf{I}_N}{1 + NM\alpha_d} \\
&= \frac{\big(\mathbf{I}_{M-1} + \frac{N\alpha_d}{1+N\alpha_d}\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{I}_{M,i}^T\big) \otimes \mathbf{I}_N}{1 + NM\alpha_d}
\end{aligned}
$$

where the last equality in (4.65) is based on $(\mathbf{I}+\mathbf{x}\mathbf{y}^H)^{-1} = \mathbf{I} - \frac{1}{1+\mathbf{y}^H\mathbf{x}}\mathbf{x}\mathbf{y}$ and $\mathbf{q}_m^H\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\mathbf{q}_m = M - 1$.

Without loss of generality, we now set $\sigma^2 = 1$ since $P$ can be any positive number. Then from (4.15) and the conditions of the theorem, we have

$$\frac{\partial J_M}{\partial \bar{\mathbf{F}}} = -2N \sum_{i=1}^{M} \bar{\mathbf{S}}_{(i)}^T \big(\mathbf{I} + \bar{\mathbf{S}}_{(i)} \bar{\mathbf{F}} \bar{\mathbf{F}}^H \bar{\mathbf{S}}_{(i)}^T\big)^{-2} \bar{\mathbf{S}}_{(i)} \bar{\mathbf{F}} \tag{4.66}$$

where, using (4.65), we have

$$\begin{aligned} & \sum_{i=1}^{M} \bar{\mathbf{S}}_{(i)}^T \big(\mathbf{I}_{(M-1)N} + \bar{\mathbf{S}}_{(i)} \bar{\mathbf{F}} \bar{\mathbf{F}}^H \bar{\mathbf{S}}_{(i)}^T\big)^{-2} \bar{\mathbf{S}}_{(i)} \\ &= \frac{\sum_{i=1}^{M} \bar{\mathbf{S}}_{(i)}^T \big((\mathbf{I}_{M-1} + \frac{N\alpha_d}{1+N\alpha_d} \mathbf{I}_{M,i} \mathbf{q}_m \mathbf{q}_m^H \mathbf{I}_{M,i}^T)^2 \otimes \mathbf{I}\big) \bar{\mathbf{S}}_{(i)}}{(1+NM\alpha_d)^2} \\ &= \frac{\sum_{i=1}^{M} \bar{\mathbf{S}}_{(i)}^T \big((\mathbf{I}_{M-1} + \beta \mathbf{I}_{M,i} \mathbf{q}_m \mathbf{q}_m^H \mathbf{I}_{M,i}^T) \otimes \mathbf{I}_N\big) \bar{\mathbf{S}}_{(i)}}{(1+NM\alpha_d)^2} \\ &= \frac{\sum_{i=1}^{M} \big(\mathbf{I}_{M,i}^T \mathbf{I}_{M,i} + \beta \mathbf{I}_{M,i}^T \mathbf{I}_{M,i} \mathbf{q}_m \mathbf{q}_m^H \mathbf{I}_{M,i}^T \mathbf{I}_{M,i}\big) \otimes \mathbf{I}_N}{(1+NM\alpha_d)^2} \\ &= \frac{\big((M-1+\beta)\mathbf{I}_M + \beta(M-2)\mathbf{q}_m \mathbf{q}_m^H\big) \otimes \mathbf{I}_N}{(1+NM\alpha_d)^2} \end{aligned} \tag{4.67}$$

where $\beta = \frac{2N\alpha_d(1+N\alpha_d)+N^2\alpha_d^2(M-1)}{(1+N\alpha_d)^2} > 0$. The last equality in (4.67) has used $\sum_{i=1}^{M} \mathbf{I}_{M,i}^T \mathbf{I}_{M,i} = (M-1)\mathbf{I}_M$ and

$$\sum_{i=1}^{M} \mathbf{I}_{M,i}^T \mathbf{I}_{M,i} \mathbf{q}_m \mathbf{q}_m^H \mathbf{I}_{M,i}^T \mathbf{I}_{M,i} \ = \mathbf{I}_M + (M-2)\mathbf{q}_m \mathbf{q}_m^H. \tag{4.68}$$

Using $(\mathbf{q}_m \mathbf{q}_m^H \otimes \mathbf{I}_N)\bar{\mathbf{F}} = \mathbf{Q}_m^H \bar{\mathbf{Q}}_m = 0$, (4.66) and (4.67) yield

$$\nabla J_M = -2N \frac{(M-1+\beta)}{(1+NM\alpha_d)^2} \bar{\mathbf{F}} \tag{4.69}$$

Also note that $\sum_{i=1}^{M} \mathbf{S}_i^T \mathbf{S}_i = (\sum_{i=1}^{M} \mathbf{e}_i \mathbf{e}_i^T) \otimes \mathbf{I}_N = \mathbf{I}_M \otimes \mathbf{I}_N = \mathbf{I}_{MN}$. Therefore, the first KKT condition in (4.23) is satisfied by $\mu_i = \frac{N(M-1+\beta)}{(1+NM\alpha_d)^2} > 0$, and all the other KKT conditions are satisfied by $\alpha_d = \frac{KP}{N^2(M-1)}$. Therefore, $\bar{\mathbf{F}} = \sqrt{\frac{KP}{N^2(M-1)}} \bar{\mathbf{Q}}_m$ is a solution to (4.23).

### 4.7.2 The gradient of $g_2(\bar{\mathbf{F}})$ in (4.39)

It follows from (4.39) that $\nabla g_2(\bar{\mathbf{F}}) = -t\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\nabla\log_2|\mathbf{I}-\mathbf{\Gamma}_{i,j}\mathbf{\Gamma}_{T,j,i}| + \sum_{i=1}^{M}\nabla\mathcal{B}_i(\bar{\mathbf{F}})$. Here, $\nabla\mathcal{B}_i(\bar{\mathbf{F}})$ is given by (4.16). To show $\nabla\log_2|\mathbf{I}-\mathbf{\Gamma}_{i,j}\mathbf{\Gamma}_{T,j,i}|$, we first consider

$$\begin{aligned}
&\nabla\log_2|\mathbf{I}-\mathbf{\Gamma}_{i,j}\mathbf{\Gamma}_{T,j,i}|\\
&=\frac{1}{\ln 2\partial\bar{\mathbf{F}}}Tr\left(\mathbf{\Gamma}_{T,j,i}(\mathbf{I}-\mathbf{\Gamma}_{i,j}\mathbf{\Gamma}_{T,j,i})^{-1}\partial\mathbf{\Gamma}_{i,j}\right)\\
&+\frac{1}{\ln 2\partial\bar{\mathbf{F}}}Tr\left((\mathbf{I}-\mathbf{\Gamma}_{i,j}\mathbf{\Gamma}_{T,j,i})^{-1}\mathbf{\Gamma}_{i,j}\partial\mathbf{\Gamma}_{T,j,i}\right)
\end{aligned}\tag{4.70}$$

where we have applied $\partial\ln|\mathbf{X}| = Tr(\mathbf{X}^{-1}\partial\mathbf{X})$, $\partial(\mathbf{XY}) = \partial\mathbf{X}\cdot\mathbf{Y}+\mathbf{X}\cdot\partial\mathbf{Y}$ and $Tr(\mathbf{XY}) = Tr(\mathbf{YX})$.

Using the matrix inverse lemma, (4.35) can be rewritten as

$$\begin{aligned}
\mathbf{\Gamma}_{i,j} &= \frac{1}{\sigma_i^2}(\mathbf{S}_j\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_j^T)\otimes\tilde{\mathbf{\Lambda}}_i - \frac{1}{\sigma_i^4}((\mathbf{S}_j\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T)\otimes\tilde{\mathbf{\Lambda}}_i)\\
&\cdot(\mathbf{I}+\frac{1}{\sigma_i^2}\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T\otimes\tilde{\mathbf{\Lambda}}_i)^{-1}((\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_j^T)\otimes\tilde{\mathbf{\Lambda}}_i)
\end{aligned}\tag{4.71}$$

where each factor or term is a function of $\bar{\mathbf{F}}\bar{\mathbf{F}}^H$, which is useful to simplify the gradient expressions. For example, with respect to the complex matrix $\mathbf{X}$, $\nabla Tr(\mathbf{AXX}^H\mathbf{B}) = 2\mathbf{BAX}$. Let $\mathbf{T}_{i,j}$ be such a permutation matrix that $\mathbf{T}_{i,j}^T[(\mathbf{S}_j\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_j^T)\otimes\tilde{\mathbf{\Lambda}}_i]\mathbf{T}_{i,j} = \tilde{\mathbf{\Lambda}}_i\otimes(\mathbf{S}_j\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_j^T)$. Also define $\tilde{\mathbf{\Gamma}}_{i,j} = \mathbf{T}_{i,j}^T\mathbf{\Gamma}_{T,j,i}(\mathbf{I}-\mathbf{\Gamma}_{i,j}\mathbf{\Gamma}_{T,j,i})^{-1}\mathbf{T}_{i,j}$. Then, one can verify (after a slightly tedious process) that the first term in (4.70) can be written as (without the coefficient $1/\ln 2$):

$$\frac{1}{\partial \bar{\mathbf{F}}} Tr\left(\mathbf{T}_{i,j}\tilde{\boldsymbol{\Gamma}}_{i,j}\mathbf{T}_{i,j}^T\partial\boldsymbol{\Gamma}_{i,j}\right) = 2(\boldsymbol{\Gamma}^{(0)}_{i,j} - \boldsymbol{\Gamma}^{(1)}_{i,j} + \boldsymbol{\Gamma}^{(2)}_{i,j} - \boldsymbol{\Gamma}^{(3)}_{i,j})\bar{\mathbf{F}} \tag{4.72}$$

where

$$\boldsymbol{\Gamma}^{(0)}_{i,j} = \sum_{l=1}^{N_i} \frac{\tilde{\lambda}_{i,l}}{\sigma_i^2}\mathbf{S}_j^T(\tilde{\boldsymbol{\Gamma}}_{i,j})_l\mathbf{S}_j \tag{4.73}$$

$$\boldsymbol{\Gamma}^{(1)}_{i,j} = \sum_{l=1}^{N_i} \frac{\tilde{\lambda}^2_{i,l}}{\sigma_i^4}\bar{\mathbf{S}}^T_{(i)}(\mathbf{I} + \frac{\tilde{\lambda}_{i,l}}{\sigma_i^2}\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}^T_{(i)})^{-1}\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_j^T(\tilde{\boldsymbol{\Gamma}}_{i,j})_l\mathbf{S}_j \tag{4.74}$$

$$\begin{aligned}\boldsymbol{\Gamma}^{(2)}_{i,j} = &\sum_{l=1}^{N_i} \frac{\tilde{\lambda}^3_{i,l}}{\sigma_i^6}\bar{\mathbf{S}}^T_{(i)}(\mathbf{I} + \frac{\tilde{\lambda}_{i,l}}{\sigma_i^2}\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}^T_{(i)})^{-1}\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_j^T \\ &\cdot (\tilde{\boldsymbol{\Gamma}}_{i,j})_l\mathbf{S}_j\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}^T_{(i)}(\mathbf{I} + \frac{\tilde{\lambda}_{i,l}}{\sigma_i^2}\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}^T_{(i)})^{-1}\bar{\mathbf{S}}_{(i)}\end{aligned} \tag{4.75}$$

$$\boldsymbol{\Gamma}^{(3)}_{i,j} = \sum_{l=1}^{N_i} \frac{\tilde{\lambda}^2_{i,l}}{\sigma_i^4}\mathbf{S}_j^T(\tilde{\boldsymbol{\Gamma}}_{i,j})_l\mathbf{S}_j\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}^T_{(i)}(\mathbf{I} + \frac{\tilde{\lambda}_{i,l}}{\sigma_i^2}\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}^T_{(i)})^{-1}\bar{\mathbf{S}}_{(i)} \tag{4.76}$$

and $(\tilde{\boldsymbol{\Gamma}}_{i,j})_l$ is the $l$th $N_j \times N_j$ diagonal block of $\tilde{\boldsymbol{\Gamma}}_{i,j}$.

A similar procedure can be applied to obtain the corresponding (explicit) expression of the second term in (4.70). The details are omitted here.

### 4.7.3 Proof of Lemma 9

To prove (4.47a), we start with (4.46a) which can rewritten as

$$|\mathbf{A} \otimes \mathbf{B} + \mathbf{C} \otimes \mathbf{D}| \geq \min_{P_1,P_2} \prod_{k=1}^{m}\prod_{l=1}^{n}(\lambda_{a,l}\lambda_{b,k} + \lambda_{c,P_1,l}\lambda_{d,P_2,k}) \tag{4.77}$$

where $\lambda_{a,l}$ is the $l$th diagonal element of $\boldsymbol{\Lambda}_a$, and $\lambda_{b,k}$, $\lambda_{c,P_1,l}$ and $\lambda_{d,P_2,k}$ are defined similarly. Every permutation of the diagonal elements of a diagonal matrix can be represented by a sequence of pair-wise permutations (each involving two diagonal elements). To prove

(4.47a), we only need to prove that (1) for every pair of diagonal elements of $\mathbf{\Lambda}_a$ (which are descending) the corresponding pair of diagonal elements of $\mathbf{\Lambda}_{c,P_1}$ must be descending to minimize the right side of (4.77), and (2) for every pair of $\mathbf{\Lambda}_b$ (which are descending) the corresponding pair of diagonal elements of $\mathbf{\Lambda}_{d,P_2}$ must be descending to minimize the right side of (4.77). The proofs of the above two statements are virtually the same. So, we only need to prove the first.

Let $\lambda_{c,P_1,s}$ and $\lambda_{c,P_1,l}$ be two diagonal elements in $\mathbf{\Lambda}_{c,P_1}$ where $s < l$ and $\lambda_{c,P_1,s} \geq \lambda_{c,P_1,l}$ (descending). Let $P_1'$ be another permutation that differs from $P_1$ only for these two elements, i.e., $\lambda_{c,P_1',s} \leq \lambda_{c,P_1',l}$ (ascending), $\lambda_{c,P_1,s} = \lambda_{c,P_1',l}$ and $\lambda_{c,P_1,l} = \lambda_{c,P_1',s}$. To compare the two permutations $P_1$ and $P_1'$, we only need to compare the two factors in (4.77) that are affected from $P_1$ to $P_1'$. The difference between the products of the two factors is

$$\begin{aligned}
&(\lambda_{a,s}\lambda_{b,k} + \lambda_{c,P_1,s}\lambda_{d,P_2,k})(\lambda_{a,l}\lambda_{b,k} + \lambda_{c,P_1,l}\lambda_{d,P_2,k}) \\
&\quad - (\lambda_{a,s}\lambda_{b,k} + \lambda_{c,P_1',s}\lambda_{d,P_2,k})(\lambda_{a,l}\lambda_{b,k} + \lambda_{c,P_1',l}\lambda_{d,P_2,k}) \\
&= \lambda_{a,s}\lambda_{b,k}\lambda_{c,P_1,l}\lambda_{d,P_2,k} + \lambda_{c,P_1,s}\lambda_{d,P_2,k}\lambda_{a,l}\lambda_{b,k} \qquad (4.78)\\
&\quad - \lambda_{a,s}\lambda_{b,k}\lambda_{c,P_1',l}\lambda_{d,P_2,k} - \lambda_{c,P_1',s}\lambda_{d,P_2,k}\lambda_{a,l}\lambda_{b,k} \\
&= \lambda_{d,P_2,k}\lambda_{b,k}(\lambda_{a,s} - \lambda_{a,l})(\lambda_{c,P_1,l} - \lambda_{c,P_1,s}) \leq 0.
\end{aligned}$$

This proves the first statement. The second statement can be proved similarly. Hence (4.47a) is proven. The proof of (4.47b) can be done in a similar manner.

### 4.7.4 Proof of Theorem 11

Define $\check{c}_{1,l} = \frac{c_{1,l}}{KP}$ and $\check{c}_{2,k} = \frac{c_{2,k}}{KP}$. Then, the power constraints become $\sum_{l=1}^{N_1} \check{c}_{1,l} = 1$ and $\sum_{k=1}^{N_2} \check{c}_{2,k} = 1$. And (4.48) now becomes

$$\begin{aligned} & I_2 = \\ & \sum_{k=1}^{N_2}\sum_{l=1}^{N_1} \log_2\left(\frac{(\sigma_2^2 + KP\tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}\check{c}_{1,l})(\sigma_1^2 + KP\tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}\check{c}_{2,k})}{\sigma_1^2\sigma_2^2 + KP\sigma_1^2\tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}\check{c}_{1,l} + KP\sigma_2^2\tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}\check{c}_{2,k}}\right) \end{aligned} \tag{4.79}$$

**High Power Case** For large $P$, (4.79) can be approximated as

$$\begin{aligned} & I_2 \\ & \approx \sum_{k=1}^{N_2}\sum_{l=1}^{N_1} \log_2\left(\frac{KP\tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}\check{c}_{1,l}\check{c}_{2,k}}{\sigma_1^2\check{c}_{1,l} + \sigma_2^2\check{c}_{2,k}}\right) \\ & = \sum_{k=1}^{N_2}\sum_{l=1}^{N_1} \log_2\left(\frac{\check{c}_{1,l}\check{c}_{2,k}}{\sigma_1^2\check{c}_{1,l} + \sigma_2^2\check{c}_{2,k}}\right) + \sum_{k=1}^{N_2}\sum_{l=1}^{N_1} \log_2(KP\tilde{\lambda}_{1,l}\tilde{\lambda}_{2,k}) \\ & \triangleq \phi_1(\check{\mathbf{c}}_1, \check{\mathbf{c}}_2, \tilde{\boldsymbol{\lambda}}_1, \tilde{\boldsymbol{\lambda}}_2) \end{aligned} \tag{4.80}$$

From (4.80), we know that the degrees of freedom per channel realization is $\lim_{P\to\infty} \frac{\phi_1(\check{\mathbf{c}}_1, \check{\mathbf{c}}_2, \tilde{\boldsymbol{\lambda}}_1, \tilde{\boldsymbol{\lambda}}_2)}{\log_2 P} = N_1 N_2$.

Also, $-\frac{\partial^2 \phi_1}{\partial \check{c}_{1,l}^2} = -\sum_j \left(\frac{\sigma_1^4}{(\sigma_1^2\check{c}_{1,l} + \sigma_2^2\check{c}_{2,k})^2} - \frac{1}{\check{c}_{1,l}^2}\right) \geq 0$, which means that $-\phi_1$ is a convex function of $\check{\mathbf{c}}_1$. Meanwhile, $-\phi_1$ is a symmetric function of $\check{\mathbf{c}}_1$. Therefore, $\phi_1$ is a Schur-concave function [49] of $\check{\mathbf{c}}_1$, and then we have $\phi_1(\mathbf{1}_{N_1}, \check{\mathbf{c}}_2, \tilde{\boldsymbol{\lambda}}_1, \tilde{\boldsymbol{\lambda}}_2) \geq \phi_1(\check{\mathbf{c}}_1, \check{\mathbf{c}}_2, \tilde{\boldsymbol{\lambda}}_1, \tilde{\boldsymbol{\lambda}}_2)$ with any $\check{\mathbf{c}}_1$ of descending elements. Similar idea can be applied to show that (4.80) is also a Schur-concave function of $\check{\mathbf{c}}_2$. Therefore, the optimal power allocation in the high power case is such that $\check{\mathbf{c}}_1 = \frac{1}{N_1}\mathbf{1}_{N_1}$ and $\check{\mathbf{c}}_2 = \frac{1}{N_2}\mathbf{1}_{N_2}$.

Also, by applying the same argument, one can easily prove that (4.80) is also a Schur-concave function of $\tilde{\boldsymbol{\lambda}}_1$ and $\tilde{\boldsymbol{\lambda}}_2$ respectively. Therefore, when $\tilde{\boldsymbol{\lambda}}_1 = \mathbf{1}_{N_1}$ and $\tilde{\boldsymbol{\lambda}}_2 = \mathbf{1}_{N_2}$, (4.80) is maximized. In other words, in the high power case, less correlated channel yields a higher secret key rate.

**Low Power Case** For small $P$, we can approximate (4.79) by its second-order Taylor series expansion at point $P = 0$:

$$I_2 = I_2|_{P=0} + \nabla I_2|_{P=0} P + \frac{1}{2}\nabla^2 I_2|_{P=0} P^2 + o(P^2) \tag{4.81}$$

where $\nabla I_2$ and $\nabla^2 I_2$ are the first and second order derivatives of (4.79) with respect to $P$. It can be easily proved that $\nabla I_2|_{P=0} = 0$ and

$$\begin{aligned} &\nabla^2 I_2|_{P=0} \\ &= \frac{2}{\ln 2}\sum_{l=1}^{N_1}\sum_{k=1}^{N_2} \tilde{\lambda}_{1,l}^2 \tilde{\lambda}_{2,k}^2 K^2 \check{c}_{1,l}\check{c}_{2,k} \triangleq \phi_2(\check{\mathbf{c}}_1, \check{\mathbf{c}}_2, \tilde{\boldsymbol{\lambda}}_1, \tilde{\boldsymbol{\lambda}}_2) \end{aligned} \tag{4.82}$$

To maximize (4.81), we just need to maximize the term (4.82). Based on (4.82) we have $\frac{\partial \phi_2}{\partial \check{c}_{1,l}} = K^2 \tilde{\lambda}_{1,l}^2 \sum_{j=1}^{N_2} \tilde{\lambda}_{2,k}^2 \check{c}_{2,k}$. Since $\{\tilde{\lambda}_{1,l}\}$ is in descending order, we know that $\phi_2(\check{\mathbf{c}}_1, \check{\mathbf{c}}_2, \tilde{\boldsymbol{\lambda}}_1, \tilde{\boldsymbol{\lambda}}_2)$ is a Schur-convex function of $\check{\mathbf{c}}_1$ with descending entries, which means it is maximized by putting almost all of the power to $\check{c}_{1,1}$. The reason that “almost all” instead of “all” is used here is to ensure the positive condition on $\mathbf{c}_a$. The same conclusion can be drawn about $\check{c}_{2,1}$ for maximizing $\phi_2(\check{\mathbf{c}}_1, \check{\mathbf{c}}_2, \tilde{\boldsymbol{\lambda}}_1, \tilde{\boldsymbol{\lambda}}_2)$. That is, in the low power case, almost all of the power should be allocated to the strongest stream.

It is also clear that $\phi_2(\check{\mathbf{c}}_1, \check{\mathbf{c}}_2, \tilde{\boldsymbol{\lambda}}_1, \tilde{\boldsymbol{\lambda}}_2)$ is a Schur-convex function of $\tilde{\boldsymbol{\lambda}}_1$ and $\tilde{\boldsymbol{\lambda}}_2$ individually. Therefore, in low power region, a higher channel correlation leads to a higher secret key rate.

### 4.7.5 Proof of Theorem 12

Refer to Appendix 4.7.1. Assume $\bar{\mathbf{F}} = \sqrt{\alpha_d}\bar{\mathbf{Q}}_m$. With (4.64), the first term of $\mathbf{\Gamma}_{i,j}$ in (4.54) can be written as

$$\begin{aligned}
&\mathbf{S}_j\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_j^T \otimes \mathbf{I}_N \\
&= \alpha_d(\mathbf{e}_j^T(MN\mathbf{I}_M - N\mathbf{q}_m\mathbf{q}_m^H)\mathbf{e}_j) \otimes \mathbf{I}_{N^2} \\
&= \alpha_d(M-1)N\mathbf{I}_{N^2}
\end{aligned} \tag{4.83}$$

With (4.65), the second term of $\mathbf{\Gamma}_{i,j}$ in (4.54) becomes

$$\begin{aligned}
&\big((\mathbf{S}_j\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T)(\mathbf{I}_{(M-1)N} + \bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T)^{-1}(\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_j^T)\big) \otimes \mathbf{I}_N \\
&= \alpha_d^2\Bigg(\big((\mathbf{e}_j^T(MN\mathbf{I}_M - N\mathbf{q}_m\mathbf{q}_m^H)\bar{\mathbf{S}}_{(i)}^T) \otimes \mathbf{I}_N\big) \\
&\quad \cdot \Big(\frac{(\mathbf{I}_{M-1} + \frac{N\alpha_d}{1+N\alpha_d}\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{I}_{M,i}^T)}{(1+NM\alpha_d)} \otimes \mathbf{I}_N\Big) \\
&\quad \cdot \big((\bar{\mathbf{S}}_{(i)}(MN\mathbf{I}_M - N\mathbf{q}_m\mathbf{q}_m^H)\mathbf{e}_j) \otimes \mathbf{I}_N\big)\Bigg) \otimes \mathbf{I}_N \\
&= \frac{\alpha_d^2(MN\mathbf{e}_j^T - Nw_M^{(j-1)m}\mathbf{q}_m^H)\mathbf{\Theta}_i(MN\mathbf{e}_j - Nw_M^{-(j-1)m}\mathbf{q}_m)}{1+NM\alpha_d}\mathbf{I}_{N^2}
\end{aligned} \tag{4.84}$$

where $\mathbf{\Theta}_i \triangleq \mathbf{I}_{M,i}^T\mathbf{I}_{M,i} + \frac{N\alpha_d}{1+N\alpha_d}\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}$. Note that $\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}$ is the identity matrix $\mathbf{I}_M$ with its $i$th diagonal element set to zero, and $\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\mathbf{q}_m$ is $\mathbf{q}_m$ with its $i$th element set to zero. Also $\mathbf{e}_j^T\mathbf{\Theta}_i\mathbf{e}_j = 1 + \frac{N\alpha_d}{1+N\alpha_d}$, $\mathbf{e}_j^T\mathbf{\Theta}_i\mathbf{q}_m = w_M^{(j-1)m}(1 + \frac{N\alpha_d}{1+N\alpha_d}(M-1))$, $\mathbf{q}_m^H\mathbf{\Theta}_i\mathbf{e}_j = w_M^{-(j-1)m}(1+\frac{N\alpha_d}{1+N\alpha_d}(M-1))$ and $\mathbf{q}_m^H\mathbf{\Theta}_i\mathbf{q}_m = (M-1)(1+\frac{N\alpha_d}{1+N\alpha_d}(M-1))$. Then, (4.84) becomes

$$\begin{aligned}
&\frac{\alpha_d^2N^2}{1+NM\alpha_d}\big(M^2\mathbf{e}_j^T\mathbf{\Theta}_i\mathbf{e}_j - Mw_M^{-(j-1)m}\mathbf{e}_j^T\mathbf{\Theta}_i\mathbf{q}_m - Mw_M^{(j-1)m}\mathbf{q}_m^H\mathbf{\Theta}_i\mathbf{e}_j + \mathbf{q}_m^H\mathbf{\Theta}_i\mathbf{q}_m\big)\mathbf{I}_{N^2} \\
&= \frac{\alpha_d^2N^2(\frac{N\alpha_d}{1+N\alpha_d} + M^2 - M - 1)}{1+NM\alpha_d}\mathbf{I}_{N^2}
\end{aligned} \tag{4.85}$$

Using (4.83), (4.84) and (4.85), $\mathbf{\Gamma}_{i,j}$ becomes

$$\mathbf{\Gamma}_{i,j} = \frac{\alpha_dMN - N\alpha_d/(1+N\alpha_d)}{1+MN\alpha_d}\mathbf{I}_{N^2} \triangleq \Gamma\mathbf{I}_{N^2} \tag{4.86}$$

where $0 < \Gamma < 1$ which is invariant to $i, j, m$. Similarly, one can verify that $\mathbf{\Gamma}_{T,j,i} = \Gamma \mathbf{I}_{N^2}$. Then we have $(\mathbf{I} - \mathbf{\Gamma}_{i,j}\mathbf{\Gamma}_{T,j,i})^{-1} = (1-\Gamma^2)^{-1}\mathbf{I}_{N^2}$.

Using the above results in (4.70), we have

$$\begin{aligned} &\frac{\partial I(\mathbf{y}_i; \mathbf{y}_{T,j})}{\partial \bar{\mathbf{F}}} \\ &= \frac{1}{\ln 2 \partial \bar{\mathbf{F}}} \left( Tr(\frac{\Gamma}{1-\Gamma^2}\partial \mathbf{\Gamma}_{i,j}) + Tr(\frac{\Gamma}{1-\Gamma^2}\partial \mathbf{\Gamma}_{T,j,i}) \right) \end{aligned} \tag{4.87}$$

Similar to (4.72), the first term in (4.87) (except for a constant factor) can be expressed as

$$\frac{1}{\partial \bar{\mathbf{F}}} Tr\left(\partial \mathbf{\Gamma}_{i,j}\right) = 2\big(\mathbf{\Gamma}_{i,j}^{(0)} - \mathbf{\Gamma}_{i,j}^{(1)} + \mathbf{\Gamma}_{i,j}^{(2)} - \mathbf{\Gamma}_{i,j}^{(3)}\big)\bar{\mathbf{F}} \tag{4.88}$$

where $\mathbf{\Gamma}_{i,j}^{(0)} = N\mathbf{e}_j\mathbf{e}_j^T \otimes \mathbf{I}_N$,

$$\mathbf{\Gamma}_{i,j}^{(1)} = N\bar{\mathbf{S}}_{(i)}^T(\mathbf{I} + \bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T)^{-1}(\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_j^T)\mathbf{S}_j \tag{4.89}$$

$$\begin{aligned} \mathbf{\Gamma}_{i,j}^{(2)} =& N\bar{\mathbf{S}}_{(i)}^T(\mathbf{I} + \bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T)^{-1}(\bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\mathbf{S}_j^T) \\ &\cdot (\mathbf{S}_j\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T)(\mathbf{I} + \bar{\mathbf{S}}_{(i)}\bar{\mathbf{F}}\bar{\mathbf{F}}^H\bar{\mathbf{S}}_{(i)}^T)^{-1}\bar{\mathbf{S}}_{(i)} \end{aligned} \tag{4.90}$$

and $\mathbf{\Gamma}_{i,j}^{(3)} = (\mathbf{\Gamma}_{i,j}^{(1)})^T$. Furthermore, using $\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\mathbf{e}_j\mathbf{e}_j^T = \mathbf{e}_j\mathbf{e}_j^T$ for $i \neq j$ and the previous results under $\bar{\mathbf{F}} = \sqrt{\alpha_d}\bar{\mathbf{Q}}_m$, we have

$$\begin{aligned} &\mathbf{\Gamma}_{i,j}^{(1)} \\ &= \frac{N\alpha_d\big(\mathbf{\Theta}_i(MN\mathbf{I} - N\mathbf{q}_m\mathbf{q}_m^H)\mathbf{e}_j\mathbf{e}_j^T\big) \otimes \mathbf{I}_N}{1 + NM\alpha_d} \\ &= \frac{N\alpha_d\big(MN\mathbf{e}_j\mathbf{e}_j^T - \frac{N}{1+N\alpha_d}\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{e}_j\mathbf{e}_j^T\big) \otimes \mathbf{I}_N}{1 + NM\alpha_d} \end{aligned} \tag{4.91}$$

$$\begin{aligned} &\mathbf{\Gamma}_{i,j}^{(2)} \\ &= \frac{N\alpha_d^2\big(\mathbf{\Theta}_i(MN\mathbf{I} - N\mathbf{q}_m\mathbf{q}_m^H)\mathbf{e}_j\mathbf{e}_j^T(MN\mathbf{I} - N\mathbf{q}_m\mathbf{q}_m^H)\mathbf{\Theta}_i\big) \otimes \mathbf{I}_N}{(1 + NM\alpha_d)^2} \\ &= \frac{\alpha_d^2 N^3}{(1 + NM\alpha_d)^2}\left(M^2\mathbf{e}_j\mathbf{e}_j^T + \frac{1}{(1+N\alpha_d)^2}\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\right. \\ &\left. - \frac{M}{1+N\alpha_d}\mathbf{e}_j\mathbf{e}_j^T\mathbf{q}_m\mathbf{q}_m^H\mathbf{I}_{M,i}^T\mathbf{I}_{M,i} - \frac{M}{1+N\alpha_d}\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{e}_j\mathbf{e}_j^T\right) \otimes \mathbf{I}_N. \end{aligned} \tag{4.92}$$

Some details for the derivation of the second equality in (4.92) are shown later in Appendix 4.7.5.

Similarly, one can verify that $\frac{\partial Tr(\partial \mathbf{\Gamma}_{T,j,i})}{\partial \bar{\mathbf{F}}} = 2(\mathbf{\Gamma}_{j,i}^{(0)} - \mathbf{\Gamma}_{j,i}^{(1)} + \mathbf{\Gamma}_{j,i}^{(2)} - (\mathbf{\Gamma}_{j,i}^{(1)})^T)\bar{\mathbf{F}}$.

Note that

$$\sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \left(\mathbf{e}_j \mathbf{e}_j^T + \mathbf{e}_i \mathbf{e}_i^T\right) \otimes \mathbf{I}_N = (M-1)\mathbf{I}_{MN} \tag{4.93}$$

$$\begin{aligned}&\sum_{i=1}^{M-1} \sum_{j=i+1}^{M} (\mathbf{I}_{M,i}^T \mathbf{I}_{M,i} \mathbf{q}_m \mathbf{q}_m^H \mathbf{e}_j \mathbf{e}_j^T + \mathbf{I}_{M,j}^T \mathbf{I}_{M,j} \mathbf{q}_m \mathbf{q}_m^H \mathbf{e}_i \mathbf{e}_i^T) \\ &= (M-2)\mathbf{q}_m \mathbf{q}_m^H + \mathbf{I}_M\end{aligned} \tag{4.94}$$

$$\sum_{i=1}^{M-1} \sum_{j=i+1}^{M} (\mathbf{I}_{M,i}^T \mathbf{I}_{M,i} \mathbf{q}_m \mathbf{q}_m^H \mathbf{I}_{M,i}^T \mathbf{I}_{M,i} + \mathbf{I}_{M,j}^T \mathbf{I}_{M,j} \mathbf{q}_m \mathbf{q}_m^H \mathbf{I}_{M,j}^T \mathbf{I}_{M,j}) \tag{4.95}$$

$$= (M-1)\mathbf{q}_m \mathbf{q}_m^H + 2\mathbf{I}_M \tag{4.96}$$

Then, with some further manipulations, we obtain

$$\begin{aligned}\frac{\partial I_M}{\partial \bar{\mathbf{F}}} &= \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} \frac{\partial I(\mathbf{y}_i; \mathbf{y}_{T,j})}{\partial \bar{\mathbf{F}}} \\ &= \frac{2N\Gamma}{(1-\Gamma^2)\ln 2}\left(\frac{M-1}{(1+MN\alpha_d)^2} + \frac{2N\alpha_d(1+2N\alpha_d)}{(1+MN\alpha_d)^2(1+N\alpha_d)^2}\right)\bar{\mathbf{F}}.\end{aligned} \tag{4.97}$$

Then one can verify that the first condition in (4.57) is satisfied by (4.97) and $\mu_i = \frac{N\Gamma}{(1-\Gamma^2)\ln 2}(\frac{M-1}{(1+MN\alpha_d)^2} + \frac{2N\alpha_d(1+2N\alpha_d)}{(1+MN\alpha_d)^2(1+N\alpha_d)^2}) > 0$, and all other conditions in (4.57) are satisfied by further choosing $\alpha_d = \frac{KP}{N^2(M-1)}$. Therefore, $\bar{\mathbf{F}} = \sqrt{\frac{KP}{N^2(M-1)}}\bar{\mathbf{Q}}_m$ is a solution to (4.57).

**Derivation of** (4.92)

From the first equality in (4.92), we have

$$\begin{aligned}&\mathbf{\Theta}_i\big(M^2\mathbf{e}_j\mathbf{e}_j^T - M\mathbf{q}_m\mathbf{q}_m^H\mathbf{e}_j\mathbf{e}_j^T - M\mathbf{e}_j\mathbf{e}_j^T\mathbf{q}_m\mathbf{q}_m^H + \mathbf{q}_m\mathbf{q}_m^H\big)\mathbf{\Theta}_i\\ &= M^2\mathbf{\Theta}_i\mathbf{e}_j\mathbf{e}_j^T\mathbf{\Theta}_i - M\mathbf{\Theta}_i\mathbf{e}_j\mathbf{e}_j^T\mathbf{q}_m\mathbf{q}_m^H\mathbf{\Theta}_i\\ &\quad - M\mathbf{\Theta}_i\mathbf{q}_m\mathbf{q}_m^H\mathbf{e}_j\mathbf{e}_j^T\mathbf{\Theta}_i + \mathbf{\Theta}_i\mathbf{q}_m\mathbf{q}_m^H\mathbf{\Theta}_i \qquad (4.98)\end{aligned}$$

Let $\eta = \frac{N\alpha_d}{1+N\alpha_d}$. Each of the four terms in (4.98) can be simplified as follows:

$$\begin{aligned}M^2\mathbf{\Theta}_i\mathbf{e}_j\mathbf{e}_j^T\mathbf{\Theta}_i = M^2\Bigg(&\mathbf{e}_j\mathbf{e}_j^T + \eta\mathbf{e}_j\mathbf{e}_j^T\mathbf{q}_m\mathbf{q}_m^H\mathbf{I}_{M,i}^T\mathbf{I}_{M,i} + \eta\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{e}_j\mathbf{e}_j^T\\ &+ \eta^2\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\Bigg) \qquad (4.99)\end{aligned}$$

$$\begin{aligned}M\mathbf{\Theta}_i\mathbf{e}_j\mathbf{e}_j^T\mathbf{q}_m\mathbf{q}_m^H\mathbf{\Theta}_i = M\Bigg(&(\eta(M-1)+1)\mathbf{e}_j\mathbf{e}_j^T\mathbf{q}_m\mathbf{q}_m^H\mathbf{I}_{M,i}^T\mathbf{I}_{M,i} \qquad (4.100)\\ &+ (\eta^2(M-1)+\eta)\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\Bigg) \qquad (4.101)\end{aligned}$$

$$\begin{aligned}M\mathbf{\Theta}_i\mathbf{q}_m\mathbf{q}_m^H\mathbf{e}_j\mathbf{e}_j^T\mathbf{\Theta}_i = M\Bigg(&(\eta(M-1)+)\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{e}_j\mathbf{e}_j^T\\ &+ (\eta^2(M-1)+\eta)\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\Bigg) \qquad (4.102)\end{aligned}$$

$$\mathbf{\Theta}_i\mathbf{q}_m\mathbf{q}_m^H\mathbf{\Theta}_i = (\eta(M-1)+1)^2\mathbf{I}_{M,i}^T\mathbf{I}_{M,i}\mathbf{q}_m\mathbf{q}_m^H\mathbf{I}_{M,i}^T\mathbf{I}_{M,i} \qquad (4.103)$$

Applying (4.98) - (4.103), the second equality of (4.92) follows.

# Chapter 5

# Two-Phase Secret Key Generation with Full-Duplex Radio

## 5.1 Introduction

Two-phase secret key generation scheme for three-node system has been proposed in [35, 36, 54]. In the first phase, two users will transmit pilot signals and a common secret key can be distilled from those received correlated signals. The corresponding pilot-based secret key rate analysis has been derived in chapter 4. In the second phase, the users will continue to transmit secret message for the rest coherence time and a additional key will be generated. In [35], it applies wiretap channel model for the second phase and [54, 36] use so-called "source emulation" secret key transmission for the second phase, which is essentially the wiretap channel model with public discussions.

In this chapter, we study the achievable secret key rate of two-phase scheme based on three-node MIMO system with full-duplex users. Such model is an extension of the SISO system in [35, 36].

## 5.2 System Model

Now we consider uncorrelated channel model and assume reciprocal channel between Alice and Bob. During the coherence block, we assume Eve's channels are independent to the channel between Alice and Bob. Since the key generation from the first phase has been studied in chapter 4, in the following we will only focus on the key generation in the second phase. In the second phase, the received signals at Alice, Bob and Eve can be expressed as

$$\begin{aligned}\mathbf{y}_A(t) &= \mathbf{H}_{AB}\mathbf{x}_B(t) + \mathbf{n}_A(t)\\ \mathbf{y}_B(t) &= \mathbf{H}_{AB}^T\mathbf{x}_A(t) + \mathbf{n}_B(t)\\ \mathbf{y}_E(t) &= \mathbf{H}_{EA}\mathbf{x}_A(t) + \mathbf{H}_{EB}\mathbf{x}_B(t) + \mathbf{n}_E(t)\end{aligned} \tag{5.1}$$

where $\mathbf{x}_A(t) \in \mathbb{C}^{N_A\times 1}$, $\mathbf{x}_B(t) \in \mathbb{C}^{N_B\times 1}$ are the transmitted signals in the second phases and they are independent Gaussian signal vectors with fixed distribution over the coherence block, i.e. $\mathbf{x}_A(t) \sim \mathcal{CN}(0, \mathbf{Q}_A)$, $\mathbf{x}_B(t) \sim \mathcal{CN}(0, \mathbf{Q}_B)$ and $Tr(\mathbf{Q}_A) \leq P_A$ and $Tr(\mathbf{Q}_B) \leq P_B$. We assume the entries of $\mathbf{H}_{AB}$, $\mathbf{H}_{EA}$ and $\mathbf{H}_{EB}$ are i.i.d zero mean complex Gaussian with variance $\sigma^2$, $\sigma_{AE}^2$ and $\sigma_{BE}^2$ respectively. The entries in $\mathbf{n}_E(t)$, $\mathbf{n}_A(t)$ and $\mathbf{n}_B(t)$ are i.i.d zero mean unit variance complex Gaussian. Here we assume the every node knows the statistic of all the channels.

Define $K_c$ as the coherence period and define $K_1 = \alpha K_c$, $\alpha \in [\frac{\max\{N_A, N_B\}}{K_c}, 1]$ as the channel training period. Then $(1-\alpha)K_c$ will be the period for the second phase key generation. In [36], a achievable secret key rate of SISO system is given and we extend it to MIMO case as following

$$\mathcal{R}_{key} = \frac{1}{K_c} I(\hat{\mathbf{h}}_A; \hat{\mathbf{h}}_B) + (1-\alpha)(\mathcal{R}_{A,s} - \mathcal{R}_{A,p} + \mathcal{R}_{B,s} - \mathcal{R}_{B,p}) \tag{5.2}$$

where $\mathcal{R}_{A,s} = \mathcal{E}\big(\log_2 |\mathbf{I} + \mathbf{H}_{AB}\mathbf{Q}_B\mathbf{H}_{AB}^H - \mathbf{H}_{AB}\mathbf{Q}_B\mathbf{H}_{EB}^H(\mathbf{H}_{EA}\mathbf{Q}_A\mathbf{H}_{EA}^H + \mathbf{H}_{EB}\mathbf{Q}_B\mathbf{H}_{EB}^H + \mathbf{I})^{-1}\mathbf{H}_{EB}\mathbf{Q}_B\mathbf{H}_{AB}^H|\big)$ , $\mathcal{R}_{B,s} = \mathcal{E}\big(\log_2 |\mathbf{I} + \mathbf{H}_{AB}^T\mathbf{Q}_A\mathbf{H}_{AB}^* - \mathbf{H}_{AB}^T\mathbf{Q}_A\mathbf{H}_{EA}^H(\mathbf{H}_{EA}\mathbf{Q}_A\mathbf{H}_{EA}^H + \mathbf{H}_{EB}\mathbf{Q}_B\mathbf{H}_{EB}^H + \mathbf{I})^{-1}\mathbf{H}_{EA}\mathbf{Q}_A\mathbf{H}_{AB}^*|\big)$, $\mathcal{R}_{A,p} = N_A \log_2(1 + \frac{\sigma^2 P_B}{1+\sigma^2\alpha K_c P_B/N_B})$ and $\mathcal{R}_{B,p} = N_B \log_2(1 + \frac{\sigma^2 P_A}{1+\sigma^2\alpha K_c P_A/N_A})$. The term $I(\hat{\mathbf{h}}_A; \hat{\mathbf{h}}_B)$ has been study in chapter 4. The derivation of (5.2) is shown in section 5.7. (5.2) holds when both $\mathbf{Q}_A$ and $\mathbf{Q}_B$ are independent to the channel estimations at Alice and Bob. In other words, $\mathbf{x}_A(t)$ and $\mathbf{x}_B(t)$ only depend on the statistic information of the channel $\mathbf{H}_{AB}$.

## 5.3 Secure Degree of Freedom

From [55, Lemma 2] we know that $\mathcal{R}_{A,s}$ is concave function to $\mathbf{Q}_B$. And because $-\mathbf{X}^{-1}$ is a concave matrix function of $\mathbf{X}$, so $\log_2(\mathbf{Y} - \mathbf{X}^{-1})$ is also a concave function in terms of $\mathbf{X}$ when $\mathbf{Y} - \mathbf{X}^{-1} \succ 0$. Therefore, $\mathcal{R}_{A,s}$ is a convex function of $\mathbf{Q}_A$. Similarly, one can prove that $\mathcal{R}_{B,s}$ is a concave function to $\mathbf{Q}_A$ and $\mathbf{Q}_B$ respectively. Because the unitary matrix will not change the statistic, define $\mathbf{\Sigma}_A$ and $\mathbf{\Sigma}_B$ are the diagonal matrices with eigenvalues of $\mathbf{Q}_A$ and $\mathbf{Q}_B$ respectively. Then $\mathcal{R}_{A,s}$ , $\mathcal{R}_{B,s}$ can be expressed as

$$\begin{aligned}\mathcal{R}_{A,s} = \mathcal{E}\big(&\log_2 |\mathbf{I} + \mathbf{H}_{AB}\mathbf{\Sigma}_B\mathbf{H}_{AB}^H \\ &- \mathbf{H}_{AB}\mathbf{\Sigma}_B\mathbf{H}_{EB}^H(\mathbf{H}_{EA}\mathbf{\Sigma}_A\mathbf{H}_{EA}^H + \mathbf{H}_{EB}\mathbf{\Sigma}_B\mathbf{H}_{EB}^H + \mathbf{I})^{-1}\mathbf{H}_{EB}\mathbf{\Sigma}_B\mathbf{H}_{AB}^H|\big)\end{aligned} \tag{5.3}$$

$$\begin{aligned}\mathcal{R}_{B,s} = \mathcal{E}\big(&\log_2 |\mathbf{I} + \mathbf{H}_{AB}^T\mathbf{\Sigma}_A\mathbf{H}_{AB}^* \\ &- \mathbf{H}_{AB}^T\mathbf{\Sigma}_A\mathbf{H}_{EA}^H(\mathbf{H}_{EA}\mathbf{\Sigma}_A\mathbf{H}_{EA}^H + \mathbf{H}_{EB}\mathbf{\Sigma}_B\mathbf{H}_{EB}^H + \mathbf{I})^{-1}\mathbf{H}_{EA}\mathbf{\Sigma}_A\mathbf{H}_{AB}^*|\big)\end{aligned} \tag{5.4}$$

From (5.3)(5.4) we know $\mathcal{R}_{A,s}$ and $\mathcal{R}_{B,s}$ are symmetric function of both $\mathbf{\Sigma}_A$ and $\mathbf{\Sigma}_B$, therefore $\mathcal{R}_{A,s}$ and $\mathcal{R}_{B,s}$ are Schur-concave function to $\mathbf{\Sigma}_A$ and $\mathbf{\Sigma}_B$. The optimal power distribution to maximize both $\mathcal{R}_{A,s}$ and $\mathcal{R}_{B,s}$ are $\mathbf{\Sigma}_A = \frac{P_A}{N_A}\mathbf{I}_{N_A}$ and $\mathbf{\Sigma}_B = \frac{P_B}{N_B}\mathbf{I}_{N_B}$.

$$\begin{aligned}\mathcal{R}_{A,s} = \mathcal{E}\big\{&\log_2 \big|\mathbf{I} + \frac{P_B}{N_B}\mathbf{H}_{EB}\mathbf{H}_{EB}^H + \frac{P_A}{N_A}\mathbf{H}_{EA}\mathbf{H}_{EA}^H \\ &- \frac{P_B^2}{N_B^2}\mathbf{H}_{EB}\mathbf{H}_{AB}^H(\mathbf{I} + \frac{P_B}{N_B}\mathbf{H}_{AB}\mathbf{H}_{AB}^H)^{-1}\mathbf{H}_{AB}\mathbf{H}_{EB}^H\big|\big\} \\ &+ \mathcal{E}\big\{\log_2 |\mathbf{I} + \frac{P_B}{N_B}\mathbf{H}_{AB}\mathbf{H}_{AB}^H|\big\} - \mathcal{E}\big\{\log_2 |\mathbf{I} + \frac{P_B}{N_B}\mathbf{H}_{EB}\mathbf{H}_{EB}^H + \frac{P_A}{N_A}\mathbf{H}_{EA}\mathbf{H}_{EA}^H|\big\}\end{aligned} \tag{5.5a}$$

$$\begin{aligned}= \mathcal{E}\big\{&\log_2 \big|\mathbf{I} + \frac{P_A}{N_A}\mathbf{H}_{EA}\mathbf{H}_{EA}^H + \frac{P_B}{N_B}\mathbf{H}_{EB}(\mathbf{I} + \frac{P_B}{N_B}\mathbf{H}_{AB}^H\mathbf{H}_{AB})^{-1}\mathbf{H}_{EB}^H\big|\big\} \\ &+ \mathcal{E}\big\{\log_2 |\mathbf{I} + \frac{P_B}{N_B}\mathbf{H}_{AB}\mathbf{H}_{AB}^H|\big\} - \mathcal{E}\big\{\log_2 |\mathbf{I} + \frac{P_B}{N_B}\mathbf{H}_{EB}\mathbf{H}_{EB}^H + \frac{P_A}{N_A}\mathbf{H}_{EA}\mathbf{H}_{EA}^H|\big\}\end{aligned} \tag{5.5b}$$

where (5.5a) are based on matrix determinant of Schur-complement and (5.5b) is based on matrix inverse lemma. $\mathcal{R}_{B,s}$ has similar structure to (5.5) an it is omit here. Based on (5.5) we have the following proposition

**Proposition 13** *Given the secret key rate from the second phase* $\mathcal{R}_{II} = \alpha(\mathcal{R}_{A,s} - \mathcal{R}_{A,p} + \mathcal{R}_{B,s} - \mathcal{R}_{B,p})$ *in* (5.5), *when* $P_A = P_B = P \to \infty$, *we have secrecy degree of freedom (SDoF)* *as*

$$d(\mathcal{R}_{II}) = \begin{cases} 2\min\{N_A, N_B\}, & \max\{N_A, N_B\} \geq N_E \\ 2(N_A + N_B - N_E), & N_A + N_B \geq N_E > \max\{N_A, N_B\} \\ 0, & N_E > N_A + N_B \end{cases} \tag{5.6}$$

**Proof.** By applying SVD we have $\mathbf{H}_{AB}^H \mathbf{H}_{AB} = \mathbf{U}_{AB} diag(\mathbf{\Lambda}_{AB}, \mathbf{0}) \mathbf{U}_{AB}^H$ where $\mathbf{U}_{AB} \in \mathbb{C}^{N_B \times N_B}$ and $\mathbf{\Lambda}_{AB}$ is diagonal matrix with dimension of $\min\{N_A, N_B\}$. Define $\mathbf{U}_{AB} = [\mathbf{U}_1, \mathbf{U}_2]$ where $\mathbf{U}_1 \in \mathbb{C}^{N_B \times \min\{N_B, N_A\}}$ and $\mathbf{U}_2 \in \mathbb{C}^{N_B \times (N_B - N_A)^+}$. Then the SDoF of the first term of $\mathcal{R}_{A,s}$ in (5.5b) is

$$\begin{aligned} &\lim_{P\to\infty} \frac{1}{\log P} \mathcal{E}\{ \log_2 |\mathbf{I} + \frac{P}{N_A} \mathbf{H}_{EA} \mathbf{H}_{EA}^H + \frac{P}{N_B} \mathbf{H}_{EB} (\mathbf{I} + \frac{P}{N_B} \mathbf{H}_{AB}^H \mathbf{H}_{AB})^{-1} \mathbf{H}_{EB}^H |\} \\ &= \lim_{P\to\infty} \frac{1}{\log P} \mathcal{E}\{ \log_2 |\mathbf{I} + \frac{P}{N_A} \mathbf{H}_{EA} \mathbf{H}_{EA}^H + \frac{P}{N_B} \mathbf{H}_{EB} \mathbf{U}_1 (\mathbf{I} + \frac{P}{N_B} \mathbf{\Lambda}_{AB})^{-1} \mathbf{U}_1^H \mathbf{H}_{EB}^H \\ &\quad + \frac{P}{N_B} \mathbf{H}_{EB} \mathbf{U}_2 \mathbf{U}_2^H \mathbf{H}_{EB}^H |\} \\ &= rank(\frac{P}{N_A} \mathbf{H}_{EA} \mathbf{H}_{EA}^H + \frac{P}{N_B} \mathbf{H}_{EB} \mathbf{U}_2 \mathbf{U}_2^H \mathbf{H}_{EB}^H) \\ &= \min\{N_E, N_A + (N_B - N_A)^+\} \end{aligned} \tag{5.7}$$

where the last equation in (5.7) is based on [46, Th.2]. Similarly, the SDoF of the second term and the third term of $\mathcal{R}_{A,s}$ in (5.5b) are $\min\{N_A, N_B\}$ and $\min\{N_E, N_A + N_B\}$ respectively. In terms of $\mathcal{R}_{A,p}$ we have $\lim_{P\to\infty} \frac{\mathcal{R}_{A,p}}{\log_2 P} = \frac{N_A \log_2(1 + \frac{N_B}{\alpha K_c})}{\log_2 P} = 0$. Therefore that SDoF of $\mathcal{R}_{A,s} - \mathcal{R}_{A,p}$ is $d(\mathcal{R}_{A,s}) = \min\{N_E, N_A + (N_B - N_A)^+\} + \min\{N_A, N_B\} - \min\{N_E, N_A + N_B\}$ and similarly we can have $d(\mathcal{R}_{B,s}) = \min\{N_E, N_B + (N_A - N_B)^+\} + \min\{N_A, N_B\} - \min\{N_E, N_A + N_B\}$. Together we have $d(\mathcal{R}_{II}) = \min\{N_E, N_A + (N_B - N_A)^+\} + \min\{N_E, N_B + (N_A - N_B)^+\} + 2\min\{N_A, N_B\} - 2\min\{N_E, N_A + N_B\}$. ■

From [54] we know that under half-duplex, the SDoF of second phase secret key generation will become zero in when $N_E \geq \max\{N_A, N_B\}$.

## 5.4 Asymptotic Forms

In the following we will use random matrix theory to obtain the asymptotic form of $\mathcal{R}_{A,s}$ in (5.5). First we can rewrite it as

$$\mathcal{R}_{A,s} = \mathcal{E}\{\log_2 \left|\mathbf{I} + \mathbf{H}_E\mathbf{T}_1\mathbf{H}_E^H\right|\} - \mathcal{E}\{\log_2 |\mathbf{I} + \mathbf{H}_E\mathbf{T}_0\mathbf{H}_E^H|\} + \mathcal{E}\{\log_2 |\mathbf{I} + \frac{P_B}{N_B}\mathbf{H}_{AB}^H\mathbf{H}_{AB}|\} \tag{5.8}$$

where we define $\mathbf{H}_E = [\frac{1}{\sqrt{N_E\sigma_{AE}^2}}\mathbf{H}_{EA}, \frac{1}{\sqrt{N_E\sigma_{BE}^2}}\mathbf{H}_{EB}]$, $\mathbf{T}_0 = diag(\frac{\sigma_{AE}^2 P_A N_E}{N_A}\mathbf{1}_{N_A}^T, \frac{\sigma_{BE}^2 P_B N_E}{N_B}\mathbf{1}_{N_B}^T)$ and

$$\mathbf{K_1} = \begin{bmatrix} \frac{\sigma_{AE}^2 P_A N_E}{N_A}\mathbf{I}_{N_A} & \\ & \frac{\sigma_{BE}^2 P_B N_E}{N_B}(\mathbf{I} + \frac{P_B}{N_B}\mathbf{H}_{AB}^H\mathbf{H}_{AB})^{-1} \end{bmatrix} \tag{5.9}$$

Similarly, in terms of $\mathcal{R}_{B,s}$ in (5.2),we can reorganize it as

$$\mathcal{R}_{B,s} = \mathcal{E}\{\log_2 \left|\mathbf{I} + \mathbf{H}_E\mathbf{T}_2\mathbf{H}_E^H\right|\} - \mathcal{E}\{\log_2 |\mathbf{I} + \mathbf{H}_E\mathbf{T}_0\mathbf{H}_E^H|\} + \mathcal{E}\{\log_2 |\mathbf{I} + \frac{P_A}{N_A}\mathbf{H}_{AB}^*\mathbf{H}_{AB}^T|\} \tag{5.10}$$

where

$$\mathbf{K_2} = \begin{bmatrix} \frac{\sigma_{AE}^2 P_A N_E}{N_A}(\mathbf{I} + \frac{P_A}{N_A}\mathbf{H}_{AB}^*\mathbf{H}_{AB}^T)^{-1} & \\ & \frac{\sigma_{BE}^2 P_B N_E}{N_B}\mathbf{I}_{N_B} \end{bmatrix} \tag{5.11}$$

Some prerequisite of random matrix theory is shown in section 5.8. We will derive the asymptotic form of $\mathcal{R}_{A,s}$ in the following and $\mathcal{R}_{B,s}$ will have the similar counterpart. Based on lemma 16 in section 5.8, regarding to the first term in (5.8), as $N_E, (N_A+N_B) \to \infty$ with $\frac{N_A+N_B}{N_E} \to \beta_0$, we have the following asymptotic expressions converge almost surely

$$\mathcal{E}\{\frac{1}{N_E}\log_2 \left|\mathbf{I} + \mathbf{H}_E\mathbf{T}_1\mathbf{H}_E^H\right|\} = \beta_0\mathcal{V}_{\mathbf{T}_1}(\eta_1) - \log_2\eta_1 + (\eta_1 - 1)\log_2 e \tag{5.12}$$

and the corresponding $\eta$-transform as

$$1 - \eta_1 = \beta_0(1 - \eta_{\mathbf{T}_1}(\eta_1)) \tag{5.13}$$

In terms of $\mathcal{V}_{\mathbf{T}_1}$, based on the definition of Shannon transform, as $N_A + N_B \to \infty$, we have

$$\begin{aligned}\mathcal{V}_{\mathbf{T}_1}(\eta_1) &= \mathcal{E}_{\mathbf{T}_1}\{\log_2(1+\eta_1\lambda(\mathbf{T}_1))\}\\ &= \frac{1}{N_A+N_B}\Big(N_A\log_2(1+\eta_1\tilde{\sigma}^2_{AE}P_A) + \log_2|\mathbf{I}+\eta_1\tilde{\sigma}^2_{BE}P_B(\mathbf{I}+\frac{P_B}{N_B}\mathbf{H}^H_{AB}\mathbf{H}_{AB})^{-1}|\Big)\\ &= \frac{1}{N_A+N_B}\Big(N_A\log_2(1+\eta_1\tilde{\sigma}^2_{AE}P_A) + \log_2|(\eta_1\tilde{\sigma}^2_{BE}P_B+1)\mathbf{I}+\frac{P_B}{N_B}\mathbf{H}^H_{AB}\mathbf{H}_{AB}|\\ &\quad - \log_2|\mathbf{I}+\frac{P_B}{N_B}\mathbf{H}^H_{AB}\mathbf{H}_{AB}|\Big)\end{aligned}\tag{5.14}$$

where $\tilde{\sigma}^2_{AE} = \frac{\sigma^2_{AE}N_E}{N_A}$ and $\tilde{\sigma}^2_{BE} = \frac{\sigma^2_{BE}N_E}{N_B}$. Particularly, when $N_A, N_B \to \infty$ and $\frac{N_A}{N_B} \to \beta_1$ then we have the second term in (5.14) converges almost surely as

$$\begin{aligned}&\frac{1}{N_B}\log_2|(\eta_1\tilde{\sigma}^2_{BE}P_B+1)\mathbf{I}+\frac{P_B}{N_B}\mathbf{H}^H_{AB}\mathbf{H}_{AB}|\\ &= \log_2(\eta_1\tilde{\sigma}^2_{BE}P_B+1) + \frac{1}{N_B}\log_2|\mathbf{I}+\frac{1}{N_B\sigma^2}\mathbf{H}^H_{AB}\bar{\mathbf{T}}_1\mathbf{H}_{AB}|\\ &= \log_2(\eta_1\tilde{\sigma}^2_{BE}P_B+1) + \beta_1\mathcal{V}_{\bar{\mathbf{T}}_1}(\bar{\eta}_1) - \log_2\bar{\eta}_1 + (\bar{\eta}_1-1)\log_2 e\end{aligned}\tag{5.15}$$

where $\bar{\mathbf{T}}_1 = diag(\frac{\sigma^2P_B}{1+\eta_1\tilde{\sigma}^2_{BE}P_B}\mathbf{1}^T_{N_A})$, $\mathcal{V}_{\bar{\mathbf{T}}_1}(\bar{\eta}_1) = \log_2(1+\bar{\eta}_1\frac{\sigma^2P_B}{1+\eta_1\tilde{\sigma}^2_{BE}P_B})$, $\eta_{\bar{\mathbf{T}}_1} = \frac{1}{1+\bar{\eta}_1\frac{\sigma^2P_B}{1+\eta_1\tilde{\sigma}^2_{BE}P_B}}$ and $1-\bar{\eta}_1 = \beta_1(1-\eta_{\bar{\mathbf{T}}_1})$. For the last term in (5.12), as $N_A, N_B \to \infty$ and $\frac{N_A}{N_B} \to \beta_1$, it will converge to $\mathcal{E}\{\log_2|\mathbf{I}+\frac{P_B}{N_B}\mathbf{H}^H_{AB}\mathbf{H}_{AB}|\}$ which if we plug it into (5.8), it will be canceled with the last term in (5.8). Regarding to $\eta_{\mathbf{T}_1}(\eta_1)$, as $N_A+N_B \to \infty$ we have

$$\begin{aligned}\eta_{\mathbf{T}_1}(\eta_1) &= \mathcal{E}\{\frac{1}{1+\eta_1\lambda(\mathbf{T}_1)}\}\\ &= \frac{1}{N_A+N_B}\Big(\frac{N_A}{1+\eta_1\tilde{\sigma}^2_{AE}P_A} + Tr\big((\mathbf{I}+\eta_1\tilde{\sigma}^2_{BE}P_B(\mathbf{I}+\frac{P_B}{N_B}\mathbf{H}^H_{AB}\mathbf{H}_{AB})^{-1})^{-1}\big)\Big)\end{aligned}\tag{5.16}$$

According to [56, p. 11], when $N_A, N_B \to \infty$ and $\frac{N_A}{N_B} \to \beta_1$ then we have the following converges almost surely

$$\begin{aligned}
&\frac{1}{N_B}Tr\big((\mathbf{I}+\eta_1\tilde{\sigma}^2_{BE}P_B(\mathbf{I}+\frac{P_B}{N_B}\mathbf{H}^H_{AB}\mathbf{H}_{AB})^{-1})^{-1}\big)\\
&=\int_0^\infty \frac{1}{1+\frac{\eta_1\tilde{\sigma}^2_{BE}P_B}{1+\sigma^2P_Bx}}dF^{N_B}_{\frac{1}{\sigma^2N_B}\mathbf{H}^H_{AB}\mathbf{H}_{AB}}(x)\\
&=\int_0^\infty 1-\frac{\eta_1\tilde{\sigma}^2_{BE}P_B}{1+\sigma^2P_Bx+\eta_1\tilde{\sigma}^2_{BE}P_B}dF^{N_B}_{\frac{1}{\sigma^2N_B}\mathbf{H}^H_{AB}\mathbf{H}_{AB}}(x)\\
&=1-\frac{\eta_1\tilde{\sigma}^2_{BE}P_B}{1+\eta_1\tilde{\sigma}^2_{BE}P_B}\int_0^\infty\frac{1}{1+\frac{\sigma^2P_B}{\eta_1\tilde{\sigma}^2_{BE}P_B+1}x}dF^{N_B}_{\frac{1}{\sigma^2N_B}\mathbf{H}^H_{AB}\mathbf{H}_{AB}}(x)\\
&=1-\frac{\eta_1\tilde{\sigma}^2_{BE}P_B}{1+\eta_1\tilde{\sigma}^2_{BE}P_B}(1-\frac{\mathcal{F}(\frac{\sigma^2P_B}{\eta_1\tilde{\sigma}^2_{BE}P_B+1},\beta_1)}{4\beta_1\frac{\sigma^2P_B}{\eta_1\tilde{\sigma}^2_{BE}P_B+1}})
\end{aligned}\tag{5.17}$$

where

$$\mathcal{F}(x,\beta)=\left(\sqrt{x(1+\sqrt{\beta})^2+1}-\sqrt{x(1-\sqrt{\beta})^2+1}\right)^2.\tag{5.18}$$

Now with (5.13)(5.16)(5.17)(5.18) we can compute $\eta_1$. In terms of the term $\mathcal{E}\big\{\log_2|\mathbf{I}+\mathbf{H}_E\mathbf{T}_0\mathbf{H}_E^H|\big\}$ in (5.5), as $N_E, (N_A+N_B)\to\infty$ with $\frac{N_A+N_B}{N_E}\to\beta_0$, we have the asymptotic form

$$\mathcal{E}\big\{\frac{1}{N_E}\log_2|\mathbf{I}+\mathbf{H}_E\mathbf{T}_0\mathbf{H}_E^H|\big\}\overset{m.s.}{\to}\beta_0\mathcal{V}_{\mathbf{T}_0}(\eta_0)-\log_2\eta_0+(\eta_0-1)\log_2 e\tag{5.19}$$

where $\mathcal{V}_{\mathbf{T}_0}(\eta_0)=\frac{1}{N_A+N_B}\big(N_A\log_2(1+\eta_0\tilde{\sigma}^2_{AE}P_A)+N_B\log_2(1+\eta_0\tilde{\sigma}^2_{BE}P_B)\big)$, $\eta_{\mathbf{T}_0}=\frac{1}{N_A+N_B}\big(\frac{N_A}{1+\eta_0\tilde{\sigma}^2_{AE}P_A}+\frac{N_B}{1+\eta_0\tilde{\sigma}^2_{BE}P_B}\big)$ and $1-\eta_0=\beta_0(1-\eta_{\mathbf{T}_0})$.

Combined with (5.5)(5.12)(5.14)(5.19), as $N_E, N_A, N_B\to\infty$ with $\frac{N_A+N_B}{N_E}\to\beta_0$ and $\frac{N_A}{N_B}\to\beta_1$, we have the asymptotic form of (5.8) as

$$\begin{aligned}
\mathbf{\Theta}_A=&N_A\log_2(\frac{1+\eta_1\tilde{\sigma}^2_{AE}P_A}{1+\eta_0\tilde{\sigma}^2_{AE}P_A})+N_B\log_2(\frac{1+\eta_1\tilde{\sigma}^2_{BE}P_B}{1+\eta_0\tilde{\sigma}^2_{BE}P_B})+N_A\log_2(1+\frac{\bar{\eta}_1\sigma^2P_B}{1+\eta_1\tilde{\sigma}^2_{BE}P_B})\\
&-N_B(\log_2\bar{\eta}_1-(\bar{\eta}_1-1)\log_2 e)+N_E(\log_2\frac{\eta_0}{\eta_1}+(\eta_1-\eta_0)\log_2 e)
\end{aligned}\tag{5.20}$$

Regarding to $\mathcal{R}_{B,s}$, under same conditions, it has the asymptotic form as

$$\begin{aligned}\mathbf{\Theta}_B = N_A \log_2(\frac{1+\eta_2\tilde{\sigma}_{AE}^2 P_A}{1+\eta_0\tilde{\sigma}_{AE}^2 P_A}) + N_B \log_2(\frac{1+\eta_2\tilde{\sigma}_{BE}^2 P_B}{1+\eta_0\tilde{\sigma}_{BE}^2 P_B}) + N_B \log_2(1+\frac{\bar{\eta}_2\sigma^2 P_A}{1+\eta_2\tilde{\sigma}_{AE}^2 P_A}) \\ - N_A(\log_2 \bar{\eta}_2 - (\bar{\eta}_2 - 1)\log_2 e) + N_E(\log_2 \frac{\eta_0}{\eta_2} + (\eta_2 - \eta_0)\log_2 e)\end{aligned} \tag{5.21}$$

where $\{\eta_2, \bar{\eta}_2\}$ are defined similar to $\{\eta_1, \bar{\eta}_1\}$ up to some changes. In Fig. 5.1 we show that the asymptotic form converges fast even when the antenna number is small.

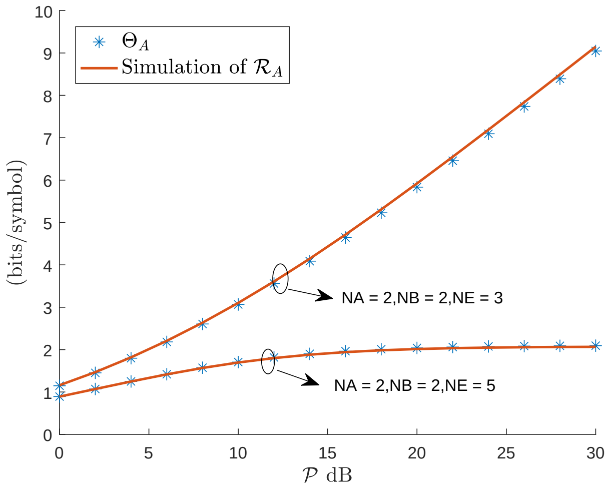

Figure 5.1: Comparison between $\mathbf{\Theta}_A$ and simulation of $\mathcal{R}_{A,s}$ in small antenna number

Then combined with (5.2)(5.20)(5.21), we have the following optimization problem

$$\max_{\alpha\in[\frac{\max\{N_A,N_B\}}{K_c},1]} \mathcal{R}_{key,asym} = \frac{1}{K_c} I(\hat{\mathbf{h}}_A; \hat{\mathbf{h}}_B) + (1-\alpha)(\mathbf{\Theta}_A + \mathbf{\Theta}_B) \tag{5.22}$$

where optimal $\alpha$ can be obtained by linear search.

Based on (5.20)(5.21), assuming $P_A = P_B = P$ and recall $\tilde{\sigma}_{AE}^2 = \frac{\sigma_{AE}^2 N_E}{N_A}$, $\tilde{\sigma}_{BE}^2 = \frac{\sigma_{BE}^2 N_E}{N_B}$, we have the following propositions:

**Proposition 14** *When $N_E \gg N_A + N_B$, $N_E \gg \frac{1}{P\sigma_i^2}$, $i = \{AE, BE\}$, we have $\mathbf{\Theta}_A = 0$, $\mathbf{\Theta}_B = 0$.*

**Proof.** When $N_E \gg N_A + N_B$, we have $\beta_0 \ll 1$. From the definition we know that the $\eta$-transform is between the range of $[0, 1]$. Then based on (5.13) correspondingly we have $\eta_1 = 1$ and based on $1 - \eta_0 = \beta_0(1 - \eta_{\mathbf{T}_0})$ we also have $\eta_0 = 1$. From (5.15) we have $\eta_{\bar{\mathbf{T}}_1} = \frac{1}{1+\bar{\eta}_1 \frac{\sigma^2 P}{1+\eta_1 \tilde{\sigma}_{BE}^2 P}}$ and the stated conditions and $\eta_1 = 1$ we have $\eta_{\bar{\mathbf{T}}_1} = 1$. Then based on $1 - \bar{\eta}_1 = \beta_1(1 - \eta_{\bar{\mathbf{T}}_1})$ we have $\bar{\eta}_1 \to 1$. Since all $\{\eta_0, \eta_1, \bar{\eta}_1\}$ are tends to 1, we have $\mathbf{\Theta}_A = 0$. Same proof can be applied to $\mathbf{\Theta}_B$. ■

From proposition 14 we know that when Eve's antenna number is large that $N_E \gg N_A + N_B$ and $N_E \gg \frac{1}{P\sigma_i^2}$, which means $N_E$ receiving antenna gain can compensate the low power, than the secret key rate from the second phase in (5.2) will be zero. Correspondingly, all the coherence time should be allocated to channel estimation.

## 5.5 Numerical Results

In this section, we show that simulation results of total achievable secret key rate $\mathcal{R}_{key,asym}$ with different $\alpha$: (1) $\alpha_{opt}$ is the solution from (5.22); (2) $\alpha_{fix} = \frac{\max\{N_A, N_B\}}{K_c}$ is setting the channel training time to the max antenna number which is the minimum number to estimate all degree of the channel and (3) $\alpha_{train-only} = 1$ means channel training only and no phase two for secret key generation. We set $N_A = N_B = 10$, $\sigma^2 = \sigma_{AE}^2 = \sigma_{BE}^2 = 1$ and $P_A = P_B = P$. In Fig. 5.2 we show $\mathcal{R}_{key,asym}$ with $N_E = 10$, $\alpha_{opt}$ performs similar to $\alpha_{fix}$ which means phase two is more important for secret key generation compared the phase one as coherence time increases. In Fig. 5.3, we show $\mathcal{R}_{key}$ with $N_E = 20$ and the results show that $\alpha_{opt}$ has better performance than the other trivial choices. In Fig. 5.4, we show $\mathcal{R}_{key,asym}$ with $N_E = 50$ and the results show that $\alpha_{opt}$ overlaps with $\alpha_{train-only}$ which is expected from proposition 14.

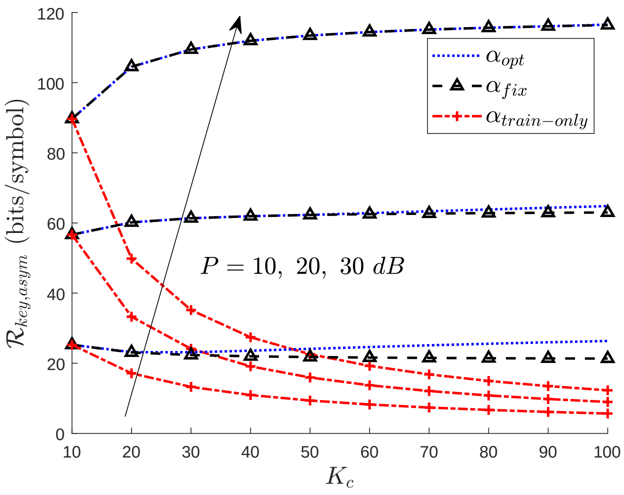

Figure 5.2: $\mathcal{R}_{key,asym}$ with $N_E = 10$

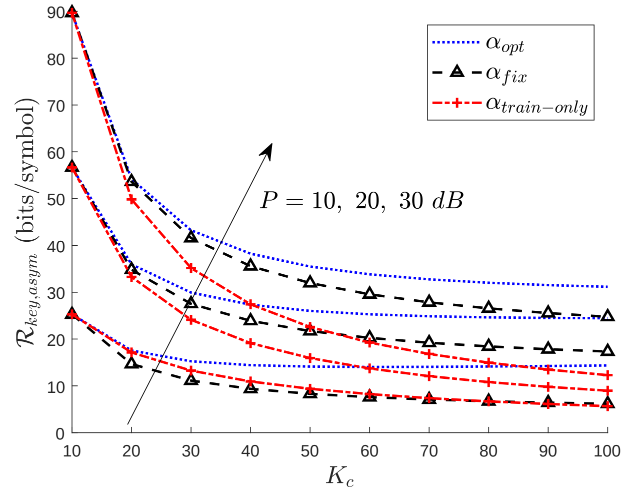

Figure 5.3: $\mathcal{R}_{key,asym}$ with $N_E = 20$

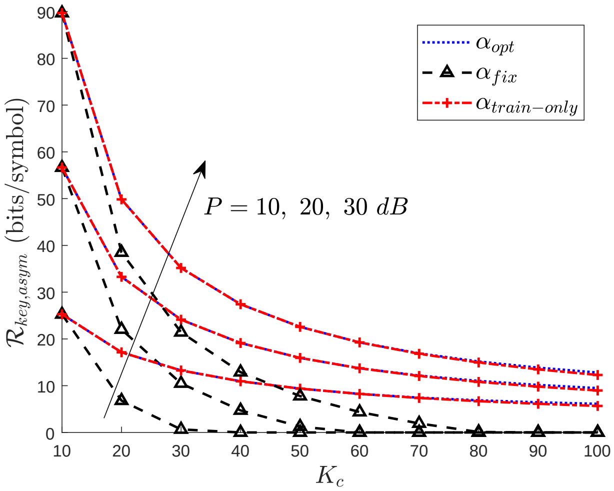

Figure 5.4: $\mathcal{R}_{key,asym}$ with $N_E = 50$

## 5.6 Conclusion

In this chapter we study the achievable secret key rate of the two-phase scheme. The results of secure degree of freedom of the secret key rate generated from the second phase show the advantage of using full-duplex transceivers compared to half-duplex. From the coherence time allocation we can know that when Eve's antennas are much more than the transmitting antennas from the users, allocating most of the coherence time for channel training is optimum for secret key generation.

## 5.7 Derivation of the Achievable Secret Key Rate

Assuming the secret key information generation process will engage $K$ coherence blocks. Before agreeing on the key, Alice and Bob have signals $\{\hat{\mathbf{h}}_A^K, \mathscr{X}_A^K, \mathscr{Y}_A^K\}$

and $\{\hat{\mathbf{h}}_B^K, \mathscr{X}_B^K, \mathscr{Y}_B^K\}$, where $\hat{\mathbf{h}}_A^K = \{\hat{\mathbf{h}}_A^{(1)}, \ldots, \hat{\mathbf{h}}_A^{(K)}\}$, $\mathscr{X}_A^K = \{\mathscr{X}_A^{(1)}, \ldots, \mathscr{X}_A^{(K)}\}$, $\mathscr{Y}_A^K = \{\mathscr{Y}_A^{(1)}, \ldots, \mathscr{Y}_A^{(K)}\}$, $\mathscr{X}_A^{(i)} = [(\mathbf{x}_A^{(i)}(1))^T, \ldots, (\mathbf{x}_A^{(i)}(K_2))^T]^T$ and $\mathscr{Y}_A^{(i)} = [(\mathbf{y}_A^{(i)}(1))^T, \ldots, (\mathbf{y}_A^{(i)}(K_2))^T]^T$. $\{\hat{\mathbf{h}}_B^K, \mathscr{X}_B^K, \mathscr{Y}_B^K\}$ is defined in a same manner. In order to use coding theorem, we need to first discretize the signals and the techniques we adapt is similar to [43, 3.4.1][36, V]

1. From the channel estimation, define $\mathcal{I}_1 = \{-j\Delta_1, -(j-1)\Delta_1, \ldots, (j-1)\Delta_1, j\Delta_1\}$, $\Delta_1 = \frac{1}{\sqrt{j}}$ and find out the elements in $\mathcal{I}_1$ that is closest to the real and imaginary part of the element in $\hat{\mathbf{h}}_A^{(i)}$ respectively. We denote the discrete vector as $[\hat{\mathbf{h}}_A^{(i)}]_j$ and $[\hat{\mathbf{h}}_B^{(i)}]_j$ is defined in a same manner.

2. Define $\mathcal{I}_2 = \{-k\Delta_2, -(k-1)\Delta_2, \ldots, (k-1)\Delta_2, k\Delta_2\}$, $\Delta_2 = \frac{1}{\sqrt{k}}$ and find the elements in $\mathcal{I}_2$ that is closest to the real and imaginary part of element in $\mathbf{x}_A^{(i)}(t)$ respectively. Denote the discrete vector as $[\mathbf{x}_A^{(i)}(t)]_k$ and its element as $[\mathbf{x}_{A,n}^{(i)}(t)]_k$, then the quantization should satisfy $|[\mathbf{x}_{A,n}^{(i)}(t)]_k| \leq |\mathbf{x}_{A,n}^{(i)}(t)|$ and $Tr\Bigg(\mathcal{E}\big([\mathbf{x}_A^{(i)}(t)]_k[\mathbf{x}_A^{(i)}(t)]_k^H\big)\Bigg) \leq Tr\Bigg(\mathcal{E}\big(\mathbf{x}_A^{(i)}(t)(\mathbf{x}_A^{(i)}(t))^H\big)\Bigg) \leq P_A(t)$. Define $[\mathbf{x}_B^{(i)}(t)]_k$ in a similar manner.

3. Let $\mathbf{y}_{A,k}^{(i)}(t) = \mathbf{H}_{AB}^{(i)}[\mathbf{x}_A^{(i)}(t)]_k + \mathbf{n}_A^{(i)}(t)$ be the output corresponding to the input $[\mathbf{x}_A^{(i)}(t)]_k$. Define $\mathcal{I}_3 = \{-l\Delta_3, -(l-1)\Delta_3, \ldots, (l-1)\Delta_3, l\Delta_3\}$, $\Delta_3 = \frac{1}{\sqrt{l}}$ and find the elements in $\mathcal{I}_3$ that is closest to the real and imaginary part of element in $\mathbf{y}_{A,k}^{(i)}(t)$ respectively. Denote the discrete vector as $[\mathbf{y}_{A,k}^{(i)}(t)]_l$ and define $[\mathbf{y}_{B,k}^{(i)}(t)]_l$ in a similar way respect to $[\mathbf{x}_B^{(i)}(t)]_k$.

Also we define the output $\mathbf{y}_{E,k}^{(i)}(t) = \mathbf{H}_{EA}^{(i)}[\mathbf{x}_A^{(i)}(t)]_k(t) + \mathbf{H}_{EB}^{(i)}[\mathbf{x}_B^{(i)}(t)]_k + \mathbf{n}_E^{(i)}(t)$ and $\mathscr{Y}_{E,k}^{(i)} = [(\mathbf{y}_{E,k}^{(i)}(1))^T, \ldots, (\mathbf{y}_{E,k}^{(i)}(K_2))^T]^T$. With the signal over $K$ coherence block, define $[\hat{\mathbf{h}}_A^K]_j = \{[\hat{\mathbf{h}}_A^{(1)}]_j, \ldots, [\hat{\mathbf{h}}_A^{(K)}]_j\}$, $[\hat{\mathbf{h}}_B^K]_j = \{[\hat{\mathbf{h}}_B^{(1)}]_j, \ldots, [\hat{\mathbf{h}}_B^{(K)}]_j\}$, $[\mathscr{X}_A^K]_k = \{[\mathscr{X}_A^{(1)}]_k, \ldots, [\mathscr{X}_A^{(K)}]_k\}$,

$[\mathscr{X}_B^K]_k = \{[\mathscr{X}_B^{(1)}]_k, \dots, [\mathscr{X}_B^{(K)}]_k\}$, $[\mathscr{Y}_{A,k}^K]_l = \{[\mathscr{Y}_{A,k}^{(1)}]_l, \dots, [\mathscr{Y}_{A,k}^{(K)}]_l\}$, $[\mathscr{Y}_{B,k}^K]_l = \{[\mathscr{Y}_{B,k}^{(1)}]_l, \dots, [\mathscr{Y}_{B,k}^{(K)}]_l\}$ and $\mathscr{Y}_{E,k}^K = \{\mathscr{Y}_{E,k}^{(1)}, \dots, \mathscr{Y}_{E,k}^{(K)}\}$

With above discrete signals, with large number of coherence blocks, a achievable rate given by [36, Proposition 1] is

$$\mathcal{R}_{key,\Delta} = \frac{1}{K_c} I([\hat{\mathbf{h}}_A]_j; [\hat{\mathbf{h}}_B]_j) + \frac{K_2}{K_c}(\mathcal{R}_{A,\Delta} + \mathcal{R}_{B,\Delta}) \tag{5.23}$$

where

$$\mathcal{R}_{A,\Delta} = I([\mathbf{y}_{A,k}]_l; [\mathbf{x}_B]_k, [\hat{\mathbf{h}}_A]_j) - I([\mathbf{y}_{A,k}]_l; \mathbf{y}_{E,k}, \mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) \tag{5.24a}$$

$$\mathcal{R}_{B,\Delta} = I([\mathbf{y}_{B,k}]_l; [\mathbf{x}_A]_k, [\hat{\mathbf{h}}_B]_j) - I([\mathbf{y}_{B,k}]_l; \mathbf{y}_{E,k}, \mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) \tag{5.24b}$$

From [43, p. 23] we know

$$\lim_{j\to\infty} I([\hat{\mathbf{h}}_A]_j; [\hat{\mathbf{h}}_B]_j) = I(\hat{\mathbf{h}}_A; \hat{\mathbf{h}}_B) \tag{5.25}$$

$$\lim_{j,l\to\infty} I([\mathbf{y}_{A,k}]_l; [\mathbf{x}_B]_k, [\hat{\mathbf{h}}_A]_j) = I(\mathbf{y}_{A,k}; [\mathbf{x}_B]_k, \hat{\mathbf{h}}_A) \tag{5.26}$$

$$\lim_{j,l\to\infty} I([\mathbf{y}_{B,k}]_l; [\mathbf{x}_A]_k, [\hat{\mathbf{h}}_B]_j) = I(\mathbf{y}_{B,k}; [\mathbf{x}_A]_k, \hat{\mathbf{h}}_B) \tag{5.27}$$

$$\lim_{l\to\infty} I([\mathbf{y}_{A,k}]_l; \mathbf{y}_{E,k}, \mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) = I(\mathbf{y}_{A,k}; \mathbf{y}_{E,k}, \mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) \tag{5.28}$$

$$\lim_{l\to\infty} I([\mathbf{y}_{B,k}]_l; \mathbf{y}_{E,k}, \mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) = I(\mathbf{y}_{B,k}; \mathbf{y}_{E,k}, \mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) \tag{5.29}$$

Now consider

$$\begin{aligned} & I(\mathbf{y}_{A,k}; [\mathbf{x}_B]_k, \hat{\mathbf{h}}_A) - I(\mathbf{y}_{A,k}; \mathbf{y}_{E,k}, \mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) \\ & = h(\mathbf{y}_{A,k}|\mathbf{y}_{E,k}, \mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) - h(\mathbf{y}_{A,k}|[\mathbf{x}_B]_k, \hat{\mathbf{h}}_A) \end{aligned} \tag{5.30}$$

From chapter 4 we know that when the channels are spatially uncorrelated, the optimal pilot becomes $\mathbf{P}_A = [\sqrt{\frac{\alpha K_c P_A}{N_A}}\mathbf{I}, \mathbf{0}]\mathbf{V}$ and $\mathbf{P}_B = [\sqrt{\frac{\alpha K_c P_B}{N_B}}\mathbf{I}, \mathbf{0}]\mathbf{V}$. Decompose the channel as $\mathbf{h} = \hat{\mathbf{h}} + \tilde{\mathbf{h}}$ while $\hat{\mathbf{h}}$ is LMMSE channel estimation and $\tilde{\mathbf{h}}$ is the channel estimation error. Since $\mathbf{h} \sim$

$\mathcal{CN}(0, \sigma^2\mathbf{I})$, then we have $\tilde{\mathbf{h}}_A \sim \mathcal{CN}(0, \frac{\sigma^2}{1+\sigma^2\alpha K_c P_B/N_B}\mathbf{I})$ and $\tilde{\mathbf{h}}_B \sim \mathcal{CN}(0, \frac{\sigma^2}{1+\sigma^2\alpha K_c P_A/N_A}\mathbf{I})$. Therefore, because $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} = (\mathbf{x}^T \otimes \mathbf{I})vec(\mathbf{H}) + \mathbf{n}$, we can bound $h(\mathbf{y}_{A,k}|[\mathbf{x}_B]_k, \hat{\mathbf{h}}_A)$ as

$$\begin{aligned} h(\mathbf{y}_{A,k}|[\mathbf{x}_B]_k, \hat{\mathbf{h}}_A) &= h(([\mathbf{x}_B]_k^T \otimes \mathbf{I})\mathbf{h} + \mathbf{n}_A|\hat{\mathbf{h}}_A, [\mathbf{x}_B]_k) \\ &= h(([\mathbf{x}_B]_k^T \otimes \mathbf{I})\tilde{\mathbf{h}}_A + \mathbf{n}_A|[\mathbf{x}_B]_k) \\ &= \mathcal{E}_{[\mathbf{x}_B]_k}\big(\log_2(\pi e)^{N_A}|(\frac{\sigma^2\|[\mathbf{x}_B]_k\|^2}{1+\sigma^2\alpha K_c P_B/N_B} + 1)\mathbf{I}|\big) \\ &\le N_A \log_2(\pi e)(1 + \frac{\sigma^2 P_B}{1+\sigma^2\alpha K_c P_B/N_B}) \end{aligned} \tag{5.31}$$

The last inequality in (5.31) is based on Jasen inequity and the last upper bound is based on $\mathcal{E}(\|[\mathbf{x}_B]_k\|^2) \le P_B$. In terms of $h(\mathbf{y}_{A,k}|\mathbf{y}_{BE,k}, \mathbf{h}_{BE}, \mathbf{h})$, we have

$$\begin{aligned} &\lim_{k\to\infty} h(\mathbf{y}_{A,k}|\mathbf{y}_{E,k}, \mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) \\ &= \lim_{k\to\infty} h(\mathbf{y}_{A,k}, \mathbf{y}_{E,k}|\mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) - \lim_{k\to\infty} h(\mathbf{y}_{E,k}|\mathbf{h}_{AE}, \mathbf{h}_{BE}) \end{aligned} \tag{5.32}$$

For the first term in (5.32), based on [43, p. 77] we have

$$\lim_{k\to\infty} \inf h(\mathbf{y}_{A,k}, \mathbf{y}_{E,k}|\mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) \ge h(\mathbf{y}_A, \mathbf{y}_E|\mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) \tag{5.33}$$

Define covariance matrix $\mathbf{Q}_{A,k} = \mathcal{E}\big([\mathbf{x}_A]_k[\mathbf{x}_A]_k^H\big)$ and $\mathbf{Q}_{B,k} = \mathcal{E}\big([\mathbf{x}_B]_k[\mathbf{x}_B]_k^H\big)$. Define a perturbation matrices $\mathbf{P}_A$ and $\mathbf{P}_B$ such that $\mathbf{Q}_{A,k} = \tilde{\mathbf{Q}}_A + \mathbf{P}_A$ and $\mathbf{Q}_{B,k} = \mathbf{Q}_B + \mathbf{P}_B$. Then the second term in (5.32) can be written as

$$
\begin{aligned}
&\lim_{k\to\infty} h(\mathbf{y}_{E,k}|\mathbf{h}_{AE}, \mathbf{h}_{BE}) \\
&\le \lim_{k\to\infty} \mathcal{E}\big(\log_2(\pi e)^{N_E}|\mathbf{H}_{EA}\mathbf{Q}_{A,k}\mathbf{H}_{EA}^H + \mathbf{H}_{EB}\mathbf{Q}_{B,k}\mathbf{H}_{EB}^H + \mathbf{I}|\big) \\
&= \lim_{k\to\infty} \mathcal{E}\big(\log_2(\pi e)^{N_E}|\mathbf{H}_{EA}(\mathbf{Q}_A + \mathbf{P}_A)\mathbf{H}_{EA}^H + \mathbf{H}_{EB}(\mathbf{Q}_B + \mathbf{P}_B)\mathbf{H}_{EB}^H + \mathbf{I}|\big) \\
&= \mathcal{E}\big(\log_2(\pi e)^{N_E}|\mathbf{H}_{EA}\mathbf{Q}_A\mathbf{H}_{EA}^H + \mathbf{H}_{EB}\mathbf{Q}_B\mathbf{H}_{EB}^H + \mathbf{I}|\big)
\end{aligned}
\tag{5.34}
$$

where the first inequality is based on Gaussian distribution maximize the differential entropy and the last equality is because $\lim_{k\to\infty}\mathbf{P}_A = \mathbf{0}$ and $\lim_{k\to\infty}\mathbf{P}_B = \mathbf{0}$.

Combine (5.28)(5.30) - (5.34), with fine quantization we can lower bound (5.24a) as

$$
\begin{aligned}
&\lim_{k\to\infty} I([\mathbf{y}_{A,k}]_l; [\mathbf{x}_B]_k, [\hat{\mathbf{h}}_A]_j) - I([\mathbf{y}_{A,k}]_l; \mathbf{y}_{E,k}, \mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) \\
&\ge h(\mathbf{y}_A, \mathbf{y}_E|\mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) - \mathcal{E}\big(\log_2(\pi e)^{N_E}|\mathbf{H}_{EA}\mathbf{Q}_A\mathbf{H}_{EA}^H + \mathbf{H}_{EB}\mathbf{Q}_B\mathbf{H}_{EB}^H + N_A\mathbf{I}|\big) \\
&\quad - N_A\log_2(\pi e)(1 + \frac{\sigma^2 P_B}{1 + \sigma^2\alpha K_c P_B/N_B}) \\
&= \mathcal{E}\big(\log_2|\mathbf{I} + \mathbf{H}_{AB}\mathbf{Q}_B\mathbf{H}_{AB}^H \\
&\quad - \mathbf{H}_{AB}\mathbf{Q}_B\mathbf{H}_{EB}^H(\mathbf{H}_{EA}\mathbf{Q}_A\mathbf{H}_{EA}^H + \mathbf{H}_{EB}\mathbf{Q}_B\mathbf{H}_{EB}^H + \mathbf{I})^{-1}\mathbf{H}_{EB}\mathbf{Q}_B\mathbf{H}_{AB}^H|\big) \\
&\quad - N_A\log_2(1 + \frac{\sigma^2 P_B}{1 + \sigma^2\alpha K_c P_B/N_B}) \\
&\triangleq \mathcal{R}_{A,s} - \mathcal{R}_{A,p}
\end{aligned}
\tag{5.35}
$$

where the equality in (5.35) is based on Schur-complement. Define $\mathcal{R}_{A,s} = \mathcal{E}\big(\log_2|\mathbf{I} + \mathbf{H}_{AB}\mathbf{Q}_B\mathbf{H}_{AB}^H - \mathbf{H}_{AB}\mathbf{Q}_B\mathbf{H}_{EB}^H(\mathbf{H}_{EA}\mathbf{Q}_A\mathbf{H}_{EA}^H + \mathbf{H}_{EB}\mathbf{Q}_B\mathbf{H}_{EB}^H + \mathbf{I})^{-1}\mathbf{H}_{EB}\mathbf{Q}_B\mathbf{H}_{AB}^H|\big)$ and $\mathcal{R}_{A,p} = N_A\log_2(1 + \frac{\sigma^2 P_B}{1+\sigma^2\alpha K_c P_B/N_B})$. Similarly with fine quantization we can have the lower bound (5.24b) as

$$
\lim_{k\to\infty} I([\mathbf{y}_{B,k}]_l; [\mathbf{x}_A]_k, [\hat{\mathbf{h}}_B]_j) - I([\mathbf{y}_{B,k}]_l; \mathbf{y}_{E,k}, \mathbf{h}_{AE}, \mathbf{h}_{BE}, \mathbf{h}) \ge \mathcal{R}_{B,s} - \mathcal{R}_{B,p}
\tag{5.36}
$$

where $\mathcal{R}_{B,s} = \mathcal{E}\big(\log_2|\mathbf{I}+\mathbf{H}_{AB}^T\mathbf{Q}_A\mathbf{H}_{AB}^* - \mathbf{H}_{AB}^T\mathbf{Q}_A\mathbf{H}_{EA}^H(\mathbf{H}_{EA}\mathbf{Q}_A\mathbf{H}_{EA}^H + \mathbf{H}_{EB}\mathbf{Q}_B\mathbf{H}_{EB}^H + \mathbf{I})^{-1}\mathbf{H}_{EA}\mathbf{Q}_A\mathbf{H}_{AB}^*|\big)$ and $\mathcal{R}_{B,p} = N_B\log_2(1+\frac{\sigma^2 P_A}{1+\sigma^2\alpha K_c P_A/N_A})$.

## 5.8 Random Matrix Asymptotic Property

**Definition 15** *The $\eta$-transform of $X$ with parameter $z$ is defined as*

$$\eta_X(z) = \mathbb{E}_X\{\frac{1}{1+zX}\} \tag{5.37}$$

*The Shannon transform of $X$ with parameter $z$ is defined as*

$$\mathcal{V}_X(z) = \mathbb{E}_X\{\log_2(1+zX)\} \tag{5.38}$$

*$z$ is a non-negative real number. If $X$ refer to a symmetric matrix, we have $\eta_X(z) = \eta_{\lambda(X)}(z)$ and $\mathcal{V}_X(z) = \mathcal{V}_{\lambda(X)}(z)$ where $\lambda$ is the eigenvalue of $X$.*

Then we have the following lemma:

**Lemma 16** *[56] Let $\mathbf{H}$ be an $N\times K$ matrix whose entries are i.i.d complex random variables with variance $\frac{1}{N}$. Let $\mathbf{T}$ be a $K\times K$ Hermitian non-negative random matrix which is independent of $\mathbf{H}$, whose empirical eigenvalue distribution converges almost surely to a nonrandom limit. Then the empirical eigenvalue distribution of $\mathbf{HTH}^H$ converges almost surely, as $K,\ N\to\infty$ with $\frac{K}{N}\to\beta$, to a distribution whose $\eta-$transform (denoted as $\eta$) with parameter $z$ satisfies*

$$1-\eta = \beta(1-\eta_{\mathbf{T}}(z\eta)) \tag{5.39}$$

*The corresponding Shannon transform satisfies*

$$\mathcal{V}_{\mathbf{HTH}^H}(z) = \beta \mathcal{V}_{\mathbf{T}}(z\eta) - \log_2 \eta + (\eta - 1)\log_2 e \qquad (5.40)$$

*where $z$ is a non-negative real number.*

# Chapter 6

# Conclusions

In this work, we investigate the techniques that improve physical layer security in wireless networks. We show that using full-duplex radio can enhance both secure data transmission and secret key generation.

In chapter 2, we develop a fast power allocation algorithm for a three-node multi-subcarrier network. With considering residual self-interference, our model is more practical than the prior works which consider perfect self-interference cancellation. Another unique feature of our work is that we consider both power and rate constraints in maximizing the secrecy capacity.

In chapter 3, we provide lower and upper bounds on the secure degrees of freedom (SDoF) of one-way and two-way wiretap channel model subject to ANECE requirements. Those bounds show that when the channel use for transmitting information symbol is less than the transmitting antenna number, using ANECE can provide the SDoF which equals to the DoF of channel capacity between the users. Such result has not been discovered in the literature and it is significant for understanding the property of ANECE.

In chapter 4, we present optimal designs of the pilot signals subject to ANECE requirement based on two criteria for optimality: 1) minimizing the sum of mean squared errors (MSE) of the minimum-mean-squared-error (MMSE) channel estimation at each and every user, and 2) maximizing the sum of the pair-wise mutual information (MI) between the signals excited by the pilots and observed by all users. The novelty of our works includes: 1) Closed-form optimal pilots are presented under a symmetric and isotropic condition; and 2) Algorithms for computing the optimal pilots for any other choices of the above parameters. The algorithm for optimal MMSE channel estimation is an extension of [37] from two users to more than two users. The algorithm for maximum MI extends [38] from two users to more than two users. These extensions are significant contributions while they are subject to the ANECE requirement.

In chapter 5, we analyze the achievable secret key rate for a two-phase key generation scheme. We consider a full-duplex MIMO system which is an extension to SISO system from the prior works. Through the SDoF analysis we show the advantage of using full-duplex and by having the expression of asymptotic secret key rate we derive an efficient algorithm for coherence time allocation between the two phases to maximize the secret key rate.

For the future study, other sophisticated math tools, i.e. [57], can be utilized to develop tighter bounds on SDoF of the system with ANECE (compared to the bounds in chapter 3, 5). New secret key generation scheme based on [58] and extending [58] to millimeter-wave system will also be interesting research topics.

# Bibliography

[1] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, “A survey on wireless security: Technical challenges, recent advances, and future trends,” *Proceedings of the IEEE*, vol. 104, no. 9, Sep 2016.

[2] Y. Wu, A. Khisti, C. Xiao, G. Caire, K.-K. Wong, and X. Gao, “A Survey of Physical Layer Security Techniques for 5G Wireless Networks and Challenges Ahead,” *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 679–695, Apr 2018.

[3] M. Bloch and J. Barros, *Physical-Layer Security.* Cambridge Press, 2011.

[4] A. D. Wyner, “The wire-tap channel,” *Bell Labs Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[5] I. Csiszar and J. Korner, “Broadcast channels with confidential messages,” *IEEE Trans. Inf. Theory*, vol. 24, no. 3, pp. 339–348, May 1978.

[6] U. M. Maurer, “Secret Key Agreement by Public Discussion from Common Information,” *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 461–470, 1993.

[7] 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.

[8] ——, “Common randomness in information theory and cryptography. Part II: CR Capacity,” *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1121–1132, 1993.

[9] L. Chen, Q. Zhu, and Y. Hua, “Fast computation for secure communication with full-duplex radio,” in *2016 IEEE 17th Int. Work. Signal Process. Adv. Wirel. Commun.* IEEE, jul 2016, pp. 1–5.

[10] L. Chen, Q. Zhu, W. Meng, and Y. Hua, “Fast power allocation for secure communication with full-duplex radio,” *IEEE Transactions on Signal Processing*, vol. 65, no. 14, pp. 3846–3861, 2017.

[11] R. Sohrabi, Q. Zhu, and Y. Hua, “Secrecy Analyses of a Full-Duplex MIMOME Network,” *Submitted to IEEE Trans. Signal Process.*, 2019.

[12] Q. Zhu and Y. Hua, "Optimal Pilots for Maximal Capacity of Secret Key Generation," *IEEE Globecom*, 2019.

[13] Q. Zhu, S. Wu, and Y. Hua, "Optimal Pilots for Anti-Eavesdropping Channel Estimation," *Submitted to IEEE Trans. Signal Process.*, 2019.

[14] W. Li, M. Ghogho, B. Chen, and C. Xiong, "Secure Communication via Sending Artificial Noise by the Receiver: Outage Secrecy Capacity/Region Analysis," *IEEE Commun. Lett.*, vol. 16, no. 10, pp. 1628–1631, Oct 2012.

[15] L. Li, Z. Chen, D. Zhang, and J. Fang, "A Full-Duplex Bob in the MIMO Gaussian Wiretap Channel: Scheme and Performance," *Signal Process. Lett. IEEE*, vol. 23, no. 1, pp. 107–111, Jan 2016.

[16] Y. Zhou, Z. Z. Xiang, Y. Zhu, and Z. Xue, "Application of Full-Duplex Wireless Technique into Secure MIMO Communication: Achievable Secrecy Rate based Optimization," *Signal Process. Lett. IEEE*, vol. 21, no. 7, pp. 804–808, Jul 2014.

[17] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of Loopback Self-Interference in Full-Duplex MIMO Relays," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5983–5993, Dec 2011.

[18] Y. Hua, P. Liang, Y. Ma, A. C. Cirik, and Q. Gao, "A Method for Broadband Full-Duplex MIMO Radio," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 793–796, Dec 2012.

[19] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-Driven Characterization of Full-Duplex Wireless Systems," *IEEE Trans. Wirel. Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec 2012.

[20] S. Hong, J. Brand, J. I. Choi, M. Jain, J. Mehlman, S. Katti, and P. Levis, "Applications of self-interference cancellation in 5G and beyond," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 114–121, Feb 2014.

[21] Y. Hua, Y. Ma, A. Gholian, Y. Li, A. C. Cirik, and P. Liang, "Radio self-interference cancellation by transmit beamforming, all-analog cancellation and blind digital tuning," *Signal Processing*, vol. 108, pp. 322–340, 2015.

[22] F. Oggier and B. Hassibi, "The Secrecy Capacity of the MIMO Wiretap Channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4961–4972, Aug 2011.

[23] A. Khisti and G. W. Wornell, "Secure Transmission With Multiple Antennas I: The MISOME Wiretap Channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 3088–3104, Nov 2010.

[24] ——, "Secure Transmission With Multiple Antennas Part II: The MIMOME Wiretap Channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5515–5532, Nov 2010.

[25] S. Goel and R. Negi, "Guaranteeing Secrecy using Artificial Noise," *IEEE Trans. Wirel. Comm.*, vol. 7, no. 6, pp. 2180–2189, Jun 2008.

[26] P. H. Lin, S. H. Lai, S. C. Lin, and H. J. Su, “On Secrecy Rate of the Generalized Artificial-Noise Assisted Secure Beamforming for Wiretap Channels,” *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1728–1740, Sep 2013.

[27] X. Tang, R. Liu, P. Spasojevic, and H. V. Poor, “Interference Assisted Secret Communication,” *IEEE Trans. Inf. Theory*, vol. 57, no. 5, pp. 3153–3167, May 2011.

[28] L. Lai and H. El Gamal, “The Relay-Eavesdropper Channel: Cooperation for Secrecy,” *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4005–4019, Sep 2008.

[29] Y. Liu, J. Li, and A. P. Petropulu, “Destination Assisted Cooperative Jamming for Wireless Physical-Layer Security,” *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 4, pp. 682–694, Apr 2013.

[30] G. Zheng, I. Krikidis, J. Li, A. P. Petropulu, and B. Ottersten, “Improving Physical Layer Secrecy Using Full-Duplex Jamming Receivers,” *IEEE Trans. Signal Process.*, vol. 61, no. 20, pp. 4962–4974, Oct 2013.

[31] 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.

[32] M. Kobayashi, P. Piantanida, S. Yang, and S. Shamai, “On the secrecy degrees of freedom of the multiantenna block fading wiretap channels,” *IEEE Trans. Inf. Forensics Secur.*, vol. 6, no. 3 PART 1, pp. 703–711, 2011.

[33] T. Y. Liu, P. Mukherjee, S. Ulukus, S. C. Lin, and Y. W. Peter Hong, “Secure degrees of freedom of MIMO Rayleigh block fading wiretap channels with no CSI anywhere,” *IEEE Trans. Wirel. Commun.*, vol. 14, no. 5, pp. 2655–2669, 2015.

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

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] Z. Li, R. Yates, and W. Trappe, “Secrecy Capacity of Independent Parallel Channels,” in *Secur. Wirel. Commun. Phys. Layer.* Boston, MA: Springer US, 2010, pp. 1–18.

[40] A. Beck and L. Tetruashvili, “On the Convergence of Block Coordinate Descent Type Methods,” *SIAM J. Optim.*, vol. 23, no. 4, pp. 2037–2060, 2013.

[41] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge university press, 2004.

[42] C. Fleury, “Sequential convex programming for structural optimization problems,” in *Optim. large Struct. Syst.* Springer, 1993, pp. 531–553.

[43] A. A. El Gamal and Y.-H. Kim, *Network information theory.* Cambridge Univ Press, 2011.

[44] R. A. Horn and C. R. Johnson, *Matrix analysis.* Cambridge University Press, 2012.

[45] O. Oyman, R. Nabar, H. Bolcskei, and A. Paulraj, “Characterizing the statistical properties of mutual information in mimo channels,” *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2784–2795, Nov 2003.

[46] A. Grant, “Rayleigh Fading Multi-Antenna Channels,” *EURASIP J. Adv. Signal Process.*, vol. 2002, no. 3, p. 260208, Dec 2002.

[47] E. Tekin and A. Yener, “The general Gaussian multiple-access and two-way wiretap channels: Achievable rates and cooperative jamming,” *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2735–2751, 2008.

[48] S. Jin, X. Gao, and X. You, “On the ergodic capacity of rank-1 Ricean-fading MIMO channels,” *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 502–517, 2007.

[49] 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.

[50] Q. Zhu and Y. Hua, “Optimal Pilots for Maximal Capacity of Secret Key Generation,” *Submitt. to 2019 IEEE Globecom*, pp. 1–6, 2019.

[51] 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.

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

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

[54] M. Andersson, A. Khisti, and M. Skoglund, “Secret-key agreement over a non-coherent block-fading MIMO wiretap channel,” *2012 IEEE Inf. Theory Work.(ITW) 2012*, pp. 153–157, 2012.

[55] T. F. Wong, M. Bloch, and J. M. Shea, “Secret sharing over fast-fading mimo wiretap channels,” *Eurasip J. Wirel. Commun. Netw.*, vol. 2009, 2009.

[56] A. M. Tulino and S. Verd, *Random Matrix Theory and Wireless Communications*, 2004, vol. 1, no. 1.

[57] L. Zheng and D. N. C. Tse, “Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel,” *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, 2002.

[58] Y. Hua, “Perfect Unconditional Secrecy for Network Security,” *Submitt. to IEEE Trans. Signal Process.*, 2019.