Secret Key Generation from MIMO Channel With or Without Reciprocity

1st Yingbo Hua
Department of Electrical and Computer Engineering
University of California at Riverside
Riverside, California, USA
yhua@ece.ucr.edu

2nd Ahmed Maksud
Department of Electrical and Computer Engineering
University of California at Riverside
Riverside, California, USA
amaks002@ucr.edu

***Abstract*—This paper presents a generalized channel probing (GCP) method and a generalized pre-processing (GPP) method as two consecutive frontend steps for secret key generation from a MIMO channel between two multi-antenna legitimate nodes against a multi-antenna eavesdropper (Eve). The degree of freedom (DoF) of the secret key capacity (SKC) of GCP/GPP are highlighted and discussed. If the number of antennas on Eve is larger than or equal to the larger number of the antennas on the two legitimate nodes, the SKC-DoF of GCP/GPP within each coherence period equals its minimum, which is either zero for non-reciprocal channel or the product of the numbers of antennas on the two legitimate nodes for reciprocal channel. Otherwise, the SKC-DoF of GCP/GPP increases with the number of random transmissions in GCP within each coherence period regardless of the channel reciprocity. A computational algorithm required for GPP is also discussed, and its performance illustrated via simulations.**

***Index Terms*—Secret key generation, secret key capacity, degree of freedom, channel probing, pre-processing.**

I. Introduction

Future networks such as Internet-of-Things will continue to increase its massive scale involving many billions of nodes. There will be increased levels of challenges for security problems including authenticity, confidentiality and integrity. All these security problems can be alleviated if there is a strong secret key between each pair of legitimate nodes (Alice and Bob). This paper addresses how to generate a secret key from a MIMO channel between Alice and Bob.

There have been many prior works on secret key generation (SKG) from wireless channels such as [1] and [2]. The basic concept behind these works is that if a wireless channel (including all relevant transceivers) between Alice and Bob is reciprocal then Alice and Bob can each send a public pilot so that they can each obtain a consistent estimate of the reciprocal channel gain. With the consistent estimates that are highly correlated with each other, they can perform quantization, information reconciliation and privacy amplification to generate a final secret key. But the degree of freedom (DoF) of the secret key capacity (SKC) of such a method for a $n_A \times n_B$ MIMO channel can be shown to be $n_A n_B$ (in bits per doubling of power per coherence time and band) as long as the eavesdropper (Eve) in the neighborhood has at least one antenna. This means that the maximum achievable secret key rate of such a method in bits per coherence period is always fixed for a given power regardless of coherence bandwidth and duration.

Recently there have been attempts such as [3] and [4] to increase the secret key rate in bits per second per subchannel beyond the constraint imposed by coherence time. In [3], the authors considered a SISO channel between Alice and Bob, and proposed that Alice and Bob each transmits a sequence of random pilots in hope that the secret key rate would increase with the length of the two random sequences. But we can show that the SKC-DoF of their scheme is only one as long as Eve is present. In [4], the authors considered a $n_A \times 1$ MISO channel from Alice to Bob, and proposed that Alice transmits a sequence of random symbols per subchannel via a randomly chosen antenna while Bob sends a public pilot in return. They also proposed that within each coherence period, Alice and Bob repeat the above transmissions multiple times (more than n_A times) each via a randomly chosen antenna at Alice. We can also show that the SKC-DoF of the scheme in [4] is only n_A (invariant to the lengths of the transmitted random sequences) even if Eve has just a single antenna.

In this paper, we look deeper into the frontend of the entire process of SKG. We will focus on the frontend steps of SKG, i.e., channel probing and pre-processing, as shown in Fig. 1, which is in great contrast to the conventional framework such as in [5] and [11] where channel probing and pre-processing have not received sufficient attention. More specifically, we consider a single-subcarrier (with no loss of generality for applications in orthogonal frequency division multiplexing systems) $n_A \times n_B$ MIMO channel from Alice to Bob in the presence of an eavesdropper (Eve) with n_E antennas. We first present a generalized channel probing (GCP) method where Alice and Bob each transmits a sequence of random vectors within each coherence period. We will show that under a Gaussian assumption of all the random vectors and channel parameters, the SKC-DoF of the GCP method increases with the number of transmitted random vectors from Alice if

This work was supported in part by the U.S. Department of Defense under W911NF-20-2-0267. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

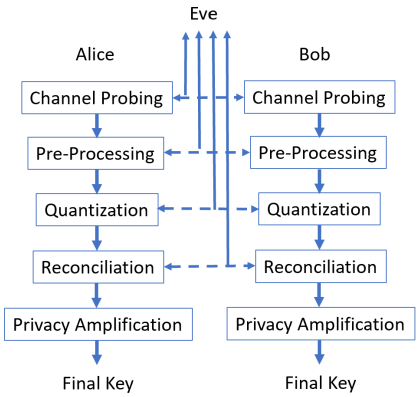

Fig. 1. An enhanced framework for SKG where channel probing and pre-processing as two frontend steps are stressed. The dashed lines generally leak information to Eve

$n_A > n_E$ and/or increases with the number of transmitted random vectors from Bob if $n_B > n_E$. If $n_E \geq \max(n_A, n_B)$, the SKC-DoF of GCP equals its minimum $n_A n_B \delta$ where $\delta = 1$ if the MIMO channel is reciprocal and $\delta = 0$ if the MIMO channel is not reciprocal.

It is useful to compare the above result with the DoF of the secrecy capacity of directly transmitting a secret over a $n_A \times n_B$ MIMO channel. The latter as shown in [7] is zero if $n_E \geq \min(n_A, n_B)$. This highlights a significant difference between "secret key capacity" and "secrecy capacity" for MIMO channels. It is important to stress that secret key capacity is achieved with additional public communication, e.g., see [13] and [14].

In this paper, we will also present a generalized pre-processing (GPP) method that allows Alice and Bob to generate a pair of highly correlated secret vectors that are ready to be quantized, which is then followed by information reconciliation and privacy amplification to produce a final secret key. Furthermore, we show that the GPP method preserves the SKC-DoF. In other words, the public communications used for GPP does not change the SKC-DoF.

Unlike all prior methods for SKG that depend on the availability of channel reciprocity even if Eve has just a single antenna, the methods shown in this paper can still yield a high secret key rate from a non-reciprocal channel provided that Alice or Bob has (or both have) more antennas than Eve.

In the rest of this paper, we will provide more details of the GCP and GPP methods. Further details other than simulation results can be found in [16].

II. Generalized Channel Probing

We consider a flat-fading or single-subcarrier MIMO channel between Alice and Bob in the presence of Eve, They have n_A, n_B and n_E antennas respectively. To extract a secret key from the MIMO channel, we consider the channel probing method where Alice transmits a random matrix $\mathbf{X}_A = [\mathbf{x}_{A,1}, \cdots, \mathbf{x}_{A,m_A}]$ using n_A antennas and $m_A \geq n_A$ time slots, and then Bob transmits another random matrix $\mathbf{X}_B = [\mathbf{x}_{B,1}, \cdots, \mathbf{x}_{B,m_B}]$ using n_B antennas and $m_B \geq n_B$ time slots. Alice does not know $\mathbf{X}_B$, and Bob does not know $\mathbf{X}_A$. (This channel probing method is a generalization of that using random pilots as proposed in [3] and [6] for SISO channel.) Consequently, the signals received by Alice and Bob can be written as

$$\mathbf{Y}_A = \mathbf{H}_{A,B}\mathbf{X}_B + \mathbf{W}_A, \tag{1}$$

$$\mathbf{Y}_B = \mathbf{H}_{B,A}\mathbf{X}_A + \mathbf{W}_B. \tag{2}$$

Here $\mathbf{H}_{A,B}$ is the channel matrix from Bob to Alice, $\mathbf{H}_{B,A}$ is the channel matrix from Alice to Bob, and $\mathbf{W}_A$ and $\mathbf{W}_B$ are the noise matrices. Correspondingly, the signals received by Eve are

$$\mathbf{Y}_{E,A} = \mathbf{G}_A\mathbf{X}_A + \mathbf{W}_{E,A}, \tag{3}$$

$$\mathbf{Y}_{E,B} = \mathbf{G}_B\mathbf{X}_B + \mathbf{W}_{E,B} \tag{4}$$

where the notations are defined in an obvious way. At the conclusion of the channel probing, the complete data sets available at Alice, Bob and Eve are respectively $\mathcal{X} = \{\mathbf{X}_A, \mathbf{Y}_A\}$, $\mathcal{Y} = \{\mathbf{X}_B, \mathbf{Y}_B\}$ and $\mathcal{Z} = \{\mathbf{Y}_{E,A}, \mathbf{Y}_{E,B}\}$.

The capacity C_S of the secret key, in bits per independent realization of $\{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$, that Alice and Bob can generate is known to be bounded as follows:

***Lemma** 1:* $C_L \doteq \max(C_A, C_B) \leq C_S \leq \min(C_0, C_Z) \leq C_Z \doteq C_U$ with $C_0 = I(\mathcal{X};\mathcal{Y})$, $C_Z = I(\mathcal{X};\mathcal{Y}|\mathcal{Z})$, $C_A = C_0 - I(\mathcal{X};\mathcal{Z})$ and $C_B = C_0 - I(\mathcal{Y};\mathcal{Z})$.

This lemma follows from the results shown in [13] and [14] for discrete $\{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$, and hence follows from the generalized definition of mutual information shown in [15] for continuous $\{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$.

To analyse C_S, we assume the following. All entries in $\mathbf{X}_A$ and $\mathbf{X}_B$ are independent and identically distributed (i.i.d.) circular complex Gaussian random variables with zero mean and variance P, i.e., with the probability density function (PDF) $\mathcal{CN}(0, P)$. All entries in $\mathbf{W}_A$, $\mathbf{W}_B$, $\mathbf{W}_{E,A}$ and $\mathbf{W}_{E,B}$ are i.i.d. $\mathcal{CN}(0, 1)$. All entries in $\mathbf{G}_A$ and $\mathbf{G}_B$ are i.i.d. $\mathcal{CN}(0, 1)$. All entries in *each* of $\mathbf{H}_{A,B}$ and $\mathbf{H}_{B,A}$ are i.i.d. $\mathcal{CN}(0, 1)$. But $vec(\mathbf{H}_{A,B})$ and $vec(\mathbf{H}_{B,A}^T)$ are jointly Gaussian with the correlation matrix $\rho\mathbf{I}_{n_A n_B}$. If $|\rho| = 1$, the channel between Alice and Bob is said to be (perfectly) reciprocal. If $|\rho| < 1$, the channel is said to be not reciprocal. Unless already mentioned otherwise, the above matrices are independent of each other.

With the above assumptions, a complete characterization of C_L and C_U seems hard. It seems still hard to just find an expression useful for numerical computation of C_L and C_U. Part of the reason seems to be the difficulty to obtain the PDF of such terms like $\mathbf{H}_{A,B}\mathbf{X}_B$ where both matrices are Gaussian.

To void the above mentioned difficulty, we will only consider the DoF of C_L and C_U. Note that if a function $f(P)$ can be written as $f(P) \approx d \log_2 P + c$ as $P \to \infty$ where d and c are invariant to P, then d is said to be the DoF of $f(P)$ (relative to $\log_2 P$).

The following is proven in [16]:

***Theorem** 1: $DoF(C_L) = DoF(C_S) = DoF(C_U) = DoF(C_Z)$ with*

$$DoF(C_S) = a_{A,B} + a_{B,A} + b_{A,B} + b_{B,A} - 2n_An_B + n_An_B\delta_{|\rho|-1} \quad (5)$$

where $a_{A,B} = \min(n_B, (n_A - n_E)^+)m_A$, $b_{A,B} = \min(n_B, (n_B + n_E - n_A)^+)n_A$, $\delta_{|\rho|-1} = 0$ if $|\rho| < 1$, and $\delta_{|\rho|-1} = 1$ if $|\rho| = 1$. Furthermore, $DoF(C_B) = DoF(C_S)$ if $n_A \geq n_B$, and $DoF(C_A) = DoF(C_S)$ if $n_A \leq n_B$.

This theorem can be simplified in the following cases of n_E:

1) For $n_E \geq \max(n_A, n_B)$,

$$DoF(C_S) = n_An_B\delta_{|\rho|-1} \quad (6)$$

which is zero if $|\rho| < 1$, or n_An_B if $|\rho| = 1$. In [3], the case of $n_A = n_B = 1$ was considered. The above result shows that their channel probing scheme with $m_A \geq 1$ and/or $m_B \geq 1$ has the same DoF (which is one) as using $m_A = m_B = 1$.

2) For $n_B \leq n_E < n_A$,

$$DoF(C_S) = \min(n_B, (n_A - n_E)^+)m_A + (n_B + n_E - n_A)^+n_A - n_An_B + n_An_B\delta_{|\rho|-1} \quad (7)$$

which increases as m_A ($\geq n_A$) increases, but is invariant to m_B ($\geq n_B$). Also in this case, the channel reciprocity is not very crucial for a large DoF. The above result is very useful for the situation where a base station with a large number of antennas is used to establish a secret key with a mobile node with a small number of antennas. Note that the scheme in [4] for the case of $n_A > n_B = 1$ has its SKC-DoF equal to n_A as long as $n_E \geq 1$. In other words, the SKC-DoF of the scheme in [4] does not benefit from the situation where $n_E < n_A$.

3) For $n_E < n_B \leq n_A$,

$$DoF(C_S) = \min(n_B, (n_A - n_E)^+)m_A + \min(n_A, (n_B - n_E)^+)m_B + n_An_B\delta_{|\rho|-1} \quad (8)$$

which increases as either m_A or m_B increases. The first term corresponds to the transmission from Alice to Bob while the second term corresponds to the transmission from Bob to Alice.

A. Comparison to Wiretap Channel Model

Theorem 1 is based on what is called source model for physical layer security [14]. In [7], a MIMO wiretap-channel (WTC) model is considered where secret information is directly transmitted over the channel without additional public communications. Using the notations defined in this paper, the main conclusion from [7] is that the DoF of the secrecy capacity $C_{S,WTC}$ (also called secure DoF) for direct transmission over the $n_A \times n_B$ MIMO channel against Eve with n_E antennas in bits per channel coherent period of total T sampling intervals is

$$DoF(C_{S,WTC}) = (\min(n_A, n_B) - n_E)^+(T - \min(n_A, n_B)) \quad (9)$$

provided $T \geq 2\min(n_A, n_B)$. We see that $DoF(C_{S,WTC})$ does not benefit from a possible reciprocity of the channel, and $DoF(C_{S,WTC})$ vanishes as soon as $n_E \geq \min(n_A, n_B)$ (as opposed to $n_E \geq \max(n_A, n_B)$). Unlike $DoF(C_{S,WTC})$, $DoF(C_S) = n_An_B$ if $n_E \geq \max(n_A, n_B)$ and $|\rho| = 1$, and $DoF(C_S)$ increases with m_A for $n_B \leq n_E < n_A$ as shown in (7). Furthermore, for the case of $n_E < \min(n_A, n_B)$, we can let $n_A \geq n_B$ and $T = m_A + m_B$, and then it follows from (8) and (9) that

$$DoF(C_S) - DoF(C_{S,WTC}) = (\min[n_B, (n_A - n_E)] - (n_B - n_E))m_A + (n_B - n_E)n_B + n_An_B\delta_{|\rho|-1}. \quad (10)$$

This difference $DoF(C_S) - DoF(C_{S,WTC})$ is strictly positive and also increases with m_A subject to $n_A > n_B$.

III. Generalized Pre-Processing

Given $\mathcal{X}$ and $\mathcal{Y}$ at Alice and Bob respectively, they now need to produce a pair of highly correlated secret vectors $\mathbf{v}_A$ and $\mathbf{v}_B$. Assuming $n_A \geq n_B$, we consider the following pre-processing method. Let Bob generate an $n_B \times (m_A + m_B - n_B)$ random matrix $\mathbf{U} = [\mathbf{U}_0, \mathbf{U}_1, \mathbf{U}_2]$ where $\mathbf{U}_0$, $\mathbf{U}_1$ and $\mathbf{U}_2$ have n_B, $m_A - n_B$ and $m_B - n_B$ columns respectively. Then, Bob uses another channel to transmit the following matrices to Alice:

$$\mathbf{X}'_B = \mathbf{X}_B + [\mathbf{U}_0, \mathbf{U}_2], \quad (11)$$

$$\mathbf{Y}'_B = \mathbf{Y}_B + [\mathbf{U}_0, \mathbf{U}_1]. \quad (12)$$

We will assume that the second channel is public and hence both Alice and Eve receive $\mathbf{X}'_B$ and $\mathbf{Y}'_B$. Since Bob knows $\mathbf{U}$, if Alice can obtain a good estimate $\hat{\mathbf{U}}$ of $\mathbf{U}$, then Alice and Bob would have a pair of highly correlated secret vectors, i.e., $\mathbf{v}_A = vec(\hat{\mathbf{U}})$ and $\mathbf{v}_B = vec(\mathbf{U})$. These two vectors can be further processed by quantization using such methods as the coset based method, [8], the guard-band based method [2] and the continuous encryption based method [9], [10]. The two bit streams at Alice and Bob after the quantization can be further processed by methods such as in [11] and [12] for information reconciliation and privacy amplification to produce the final secret key.

Let $\mathcal{X}' = \{\mathcal{X}, \mathbf{X}'_B, \mathbf{Y}'_B\}$, $\mathcal{Y}' = \{\mathcal{Y}, \mathbf{U}\}$ and $\mathcal{Z}' = \{\mathcal{Z}, \mathbf{X}'_B, \mathbf{Y}'_B\}$. Assume that all entries in $\mathbf{U}$ are i.i.d. $\mathcal{CN}(0, P)$. The following is proved in [16]:

***Theorem** 2: If $|\rho| = 1$ and $n_A \geq n_B$, the secret key capacity C'_S based on $\{\mathcal{X}', \mathcal{Y}', \mathcal{Z}'\}$ has the same DoF as C_S based on $\{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$, i.e.,*

$$DoF(C'_S) = DoF(C_S). \quad (13)$$

This theorem says that the leakage to Eve due to $\{\mathbf{X}'_B, \mathbf{Y}'_B\}$ does not change the DoF of the secret key capacity from that given by (5). This generalized pre-processing method is inspired by a conceptual approach shown in section 4.2.1 in [14] where Bob transmits publicly the modulo sum of a uniform random variable $\mathcal{U}$ and a discrete $\mathcal{Y}$ (both belonging to a common finite set). By doing so, the lower bound C_B

on C_S is achieved. But an application of that approach for $\{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$ obtained via the generalized channel probing would require a specific coding scheme, which is not yet available.

Furthermore, if the MIMO channel is not reciprocal, i.e., $|\rho| < 1$, the following Corollary is proved in [16]:

***Corollary** 1:* If $|\rho| < 1$ and $n_A \geq n_B$, and we let $\mathbf{U}_0$ and the first $n_A - n_B$ columns of $\mathbf{U}_1$ be public, then the secret key capacity C_S' based on $\{\mathcal{X}', \mathcal{Y}', \mathcal{Z}'\}$ has the same DoF as C_S based on $\{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$, i.e., $DoF(C_S') = DoF(C_S)$.

It is important to notice that without the reciprocity of the MIMO channel, $DoS(C_S)$ does not vanish unless $n_E \geq \max(n_A, n_B)$. See the discussion of Theorem 1.

IV. Estimation of Secret Vector at Alice

Assuming $n_A \geq n_B$, the implementation of the GPP method requires Alice to obtain a good estimate of $\mathbf{U}$ from her knowledge of $\mathcal{X}' = \{\mathbf{X}_A, \mathbf{Y}_A, \mathbf{X}_B', \mathbf{Y}_B'\}$. We will consider separately the cases of reciprocal channel and non-reciprocal channel.

A. The case of $|\rho| = 1$

For the reciprocal channel case, the key equations that Alice needs to exploit are

$$\mathbf{Y}_A = \mathbf{H}(\mathbf{X}_B' - [\mathbf{U}_0, \mathbf{U}_2]) + \mathbf{W}_A, \tag{14}$$

$$\mathbf{Y}_B' = \mathbf{H}^T\mathbf{X}_A + [\mathbf{U}_0, \mathbf{U}_1] + \mathbf{W}_B, \tag{15}$$

where the unknowns are $\mathbf{H}$ and $\mathbf{U} = [\mathbf{U}_0, \mathbf{U}_1, \mathbf{U}_2]$. Also notice that (14) is nonlinear.

To show more insights into (14) and (15), we let $\mathbf{Y}_A = [\mathbf{Y}_{A,\alpha}, \mathbf{Y}_{A,\beta}]$ with $\mathbf{Y}_{A,\alpha}$ consisting of the first n_B columns of $\mathbf{Y}_A$ and $\mathbf{Y}_{A,\beta}$ consisting of all other columns of $\mathbf{Y}_A$. We will use the subscripts $_\alpha$ and $_\beta$ to indicate the same partitions for all relevant matrices. Then we know

$$\mathbf{Y}_{A,\alpha} = \mathbf{H}[\mathbf{X}_{B,\alpha}' - \mathbf{U}_0] + \mathbf{W}_{A,\alpha}, \tag{16}$$

$$\mathbf{Y}_{A,\beta} = \mathbf{H}[\mathbf{X}_{B,\beta}' - \mathbf{U}_2] + \mathbf{W}_{A,\beta}, \tag{17}$$

$$\mathbf{Y}_{B,\alpha}' = \mathbf{H}^T\mathbf{X}_{A,\alpha} + \mathbf{U}_0 + \mathbf{W}_{B,\alpha}, \tag{18}$$

$$\mathbf{Y}_{B,\beta}' = \mathbf{H}^T\mathbf{X}_{A,\beta} + \mathbf{U}_1 + \mathbf{W}_{B,\beta}. \tag{19}$$

If $\hat{\mathbf{H}}$ is given, then the least-square (LS) estimates of $\mathbf{U}_0$, $\mathbf{U}_1$ and $\mathbf{U}_2$ are as follows:

$$\hat{\mathbf{U}}_0 = (\hat{\mathbf{H}}^H\hat{\mathbf{H}} + \mathbf{I}_{n_B})^{-1}(-\hat{\mathbf{H}}^H\Delta\mathbf{Y}_{A,\alpha} + \Delta\mathbf{Y}_{B,\alpha}'), \tag{20}$$

$$\hat{\mathbf{U}}_1 = \mathbf{Y}_{B,\beta}' - \hat{\mathbf{H}}^T\mathbf{X}_{A,\beta}, \tag{21}$$

$$\hat{\mathbf{U}}_2 = -(\hat{\mathbf{H}}^H\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}^H\Delta\mathbf{Y}_{A,\beta}, \tag{22}$$

with $\Delta\mathbf{Y}_{A,\alpha} = \mathbf{Y}_{A,\alpha} - \hat{\mathbf{H}}\mathbf{X}_{B,\alpha}'$, $\Delta\mathbf{Y}_{B,\alpha}' = \mathbf{Y}_{B,\alpha}' - \hat{\mathbf{H}}^T\mathbf{X}_{A,\alpha}$ and $\Delta\mathbf{Y}_{A,\beta} = \mathbf{Y}_{A,\beta} - \hat{\mathbf{H}}\mathbf{X}_{B,\beta}'$. Note that (20) is the LS solution of $\mathbf{U}_0$ to (16) and (18), or equivalently,

$$\hat{\mathbf{U}}_0 = arg\min_{\mathbf{U}_0} J_0 \tag{23}$$

with

$$J_0 = \left\| \begin{bmatrix} \Delta\mathbf{Y}_{A,\alpha} \\ \Delta\mathbf{Y}_{B,\alpha}' \end{bmatrix} - \begin{bmatrix} -\hat{\mathbf{H}} \\ \mathbf{I}_{n_B} \end{bmatrix} \mathbf{U}_0 \right\|_F^2. \tag{24}$$

Here $\|\mathbf{M}\|_F^2 \doteq Tr(\mathbf{M}\mathbf{M}^H)$ for any matrix $\mathbf{M}$. We will also write $J_0 = \|\mathbf{Y}_0 - \mathbf{H}_0\mathbf{U}_0\|_F^2$ with $\mathbf{Y}_0$ and $\mathbf{H}_0$ defined in an obvious way.

If $\hat{\mathbf{U}}_0$ is a consistent estimate of $\mathbf{U}_0$, then a consistent estimate of $\mathbf{H}$ follows from (16), i.e.,

$$\hat{\mathbf{H}} = \mathbf{Y}_{A,\alpha}[\mathbf{X}_{B,\alpha}' - \hat{\mathbf{U}}_0]^{-1} \tag{25}$$

which then leads to consistent estimates of $\mathbf{U}_1$ and $\mathbf{U}_2$ via (21) and (22).

To find a consistent estimate $\hat{\mathbf{U}}_0$, we can use (25) in (18), which yields

$$(\mathbf{X}_{B,\alpha}' - \hat{\mathbf{U}}_0)^T(\mathbf{Y}_{B,\alpha}' - \hat{\mathbf{U}}_0) = \mathbf{Y}_{A,\alpha}^T\mathbf{X}_{A,\alpha}. \tag{26}$$

This is an $n_B \times n_B$ quadratic matrix equation of the $n_B \times n_B$ unknown matrix $\hat{\mathbf{U}}_0$, which in general have multiple (but no more than $2^{n_B^2}$) solutions. One of the solutions in the absence of noise is the desired solution $\mathbf{U}_0$.

Every solution to (26) can be written as $\hat{\mathbf{U}}_0 = \mathbf{U}_0 - \Delta\hat{\mathbf{U}}_0$. Then (26) implies $(\mathbf{X}_{B,\alpha} + \Delta\hat{\mathbf{U}}_0)^T(\mathbf{Y}_{B,\alpha} + \Delta\hat{\mathbf{U}}_0) = \mathbf{Y}_{A,\alpha}^T\mathbf{X}_{A,\alpha}$. Clearly, every nonzero $\Delta\hat{\mathbf{U}}_0$ in the absence of noise is independent of $\mathbf{U}_0$. For example, if $n_B = 1$, then $\Delta\hat{\mathbf{U}}_0 = -\mathbf{X}_{B,\alpha} - \mathbf{Y}_{B,\alpha}$. Furthermore, one can verify that the corresponding estimates of $\mathbf{U}_1$ and $\mathbf{U}_2$ from (21) and (22) can be also written as $\hat{\mathbf{U}}_1 = \mathbf{U}_1 - \Delta\hat{\mathbf{U}}_1$ and $\hat{\mathbf{U}}_2 = \mathbf{U}_2 - \Delta\hat{\mathbf{U}}_2$ where $\Delta\hat{\mathbf{U}}_1$ and $\Delta\hat{\mathbf{U}}_2$ in the absence of noise are also independent of $\mathbf{U}$. Therefore, among all solutions to (26) in the absence of noise, the desired solution has the minimum variance. Provided that the number $n_U = n_B(m_A + m_B - n_B)$ of entries in $\mathbf{U}$ is large, the desired solution to (26) can be detected by choosing the one corresponding to the smallest $\frac{1}{Pn_U}\|\hat{\mathbf{U}}\|_F^2$ which approaches to one for large n_U.

To show an algorithm to solve (26), we can write $\mathbf{T}_1 = \mathbf{X}_{B,\alpha}' - \hat{\mathbf{U}}_0$ and $\mathbf{T}_2 = \mathbf{Y}_{B,\alpha}' - \hat{\mathbf{U}}_0$. Then (26) is equivalent to

$$\begin{cases} \mathbf{T}_1^T\mathbf{T}_2 = \mathbf{Z}, \\ \mathbf{T}_1 - \mathbf{T}_2 = \mathbf{Y}. \end{cases} \tag{27}$$

with $\mathbf{Y} = \mathbf{X}_{B,\alpha}' - \mathbf{Y}_{B,\alpha}'$ and $\mathbf{Z} = \mathbf{Y}_{A,\alpha}^T\mathbf{X}_{A,\alpha} = \mathbf{X}_{B,\alpha}^T\mathbf{Y}_{B,\alpha}$. For a random initial guess $\hat{\mathbf{U}}_0^{(0)}$ of $\mathbf{U}_0$, let $\mathbf{T}_1^{(0)} = \mathbf{X}_{B,\alpha}' - \hat{\mathbf{U}}_0^{(0)}$. Then for each $\mathbf{T}_1^{(i)}$ with $i \geq 0$, we first compute

$$\begin{cases} \mathbf{T}_2^{(i)'} = \mathbf{T}_1^{(i)^{-T}}\mathbf{Z}, \\ \mathbf{T}_2^{(i)''} = \mathbf{T}_1^{(i)} - \mathbf{Y}, \end{cases} \tag{28}$$

which are two possible solutions of $\mathbf{T}_2$ based on the two equations in (27). We then update the estimate of $\mathbf{T}_2$ by taking the average: $\mathbf{T}_2^{(i)} = \frac{1}{2}(\mathbf{T}_2^{(i)'} + \mathbf{T}_2^{(i)''})$. With the updated $\mathbf{T}_2$, we can renew $\mathbf{T}_1$ in two different ways, i.e., via the nonlinear equation in (28) or the linear equation in (28). (These two choices often lead to two different solutions upon convergence.) We repeat the above process until $\|\mathbf{T}_2^{(i)'} - \mathbf{T}_2^{(i)''}\|$ is sufficiently small. Upon convergence, an estimate of $\mathbf{U}_0$ is $\hat{\mathbf{U}}_0 = \mathbf{X}_{B,\alpha}' - \hat{\mathbf{T}}_1$.

After a good initial estimate of $\mathbf{H}$ is found, the maximum likelihood (ML) estimates of all unknowns (i.e., $\mathbf{H}$, $\mathbf{U}_0$, $\mathbf{U}_1$

and $\mathbf{U}_2$) can be found by minimizing the following cost function:

$$J = \|\mathbf{Y}_A - \mathbf{H}(\mathbf{X}'_B - [\mathbf{U}_0, \mathbf{U}_2])\|_F^2 + \|\mathbf{Y}'_B - \mathbf{H}^T\mathbf{X}_A - [\mathbf{U}_0, \mathbf{U}_1]\|_F^2. \quad (29)$$

A gradient method for updating the estimate of $\mathbf{H}$ is

$$\hat{\mathbf{H}}^{(k+1)} = \hat{\mathbf{H}}^{(k)} - \eta \left.\frac{\partial J}{\partial \mathbf{H}}\right|_k \quad (30)$$

where k denotes the k-th iteration, and η is a step size which can be optimized by backtracking such as Armijo's algorithm. Furthermore, one can verify from (29) that

$$\frac{\partial J}{\partial \mathbf{H}} = -2(\mathbf{Y}_A - \mathbf{H}(\mathbf{X}'_B - [\mathbf{U}_0, \mathbf{U}_2]))(\mathbf{X}'_B - [\mathbf{U}_0, \mathbf{U}_2])^H - 2[(\mathbf{Y}'_B - \mathbf{H}^T\mathbf{X}_A - [\mathbf{U}_0, \mathbf{U}_1])\mathbf{X}_A^H]^T \quad (31)$$

where $\mathbf{H}$, $\mathbf{U}_0$, $\mathbf{U}_1$ and $\mathbf{U}_2$ need to be replaced by their best estimates at every iteration.

B. The case of $|\rho| < 1$

For this non-reciprocal channel case, we will treat $\mathbf{H}_{A,B}$ and $\mathbf{H}_{B,A}$ as two independent matrices. Also note that in this case, $\mathbf{U}_0$ and the first $n_A - n_B$ columns of $\mathbf{U}_1$ are public.

Alice can now compute an initial consistent estimate of $\mathbf{H}_{A,B}$ based on (16) as follows:

$$\hat{\mathbf{H}}_{A,B} = \mathbf{Y}_{A,\alpha}(\mathbf{X}'_{B,\alpha} - \mathbf{U}_0)^{-1}. \quad (32)$$

With any $\hat{\mathbf{H}}_{A,B}$, the ML estimate of $\mathbf{U}_2$ is the LS solution of (17), i.e.,

$$\hat{\mathbf{U}}_2 = \mathbf{X}'_{B,\beta} - (\hat{\mathbf{H}}_{A,B}^H\hat{\mathbf{H}}_{A,B})^{-1}\hat{\mathbf{H}}_{A,B}^H\mathbf{Y}_{A,\beta}. \quad (33)$$

The ML estimate of $\mathbf{H}_{A,B}$ (and hence $\mathbf{U}_2$) can be found by a gradient search of the LS solution of (16) and (17) with $\mathbf{H} = \mathbf{H}_{A,B}$, i.e.,

$$\hat{\mathbf{H}}_{A,B,k+1} = \hat{\mathbf{H}}_{A,B,k} - \eta \left.\frac{\partial J_1}{\partial \mathbf{H}_{A,B}}\right|_k \quad (34)$$

where J_1 is the first term in (29), and $\frac{\partial J_1}{\partial \mathbf{H}_{A,B}}$ is the first term in (31) with $\mathbf{H} = \mathbf{H}_{A,B}$. For ML estimation of $\mathbf{H}_{B,A}$ and the unknowns in $\mathbf{U}_1$, let $\mathbf{Y}'_{B,\gamma}$ and $\mathbf{X}_{A,\gamma}$ be each the first n_A columns of $\mathbf{Y}'_B$ and $\mathbf{X}_A$ respectively, $\mathbf{U}_\gamma$ be the first n_A columns of $[\mathbf{U}_0, \mathbf{U}_1]$, and $\mathbf{U}_\tau$, $\mathbf{Y}'_{B,\tau}$ and $\mathbf{X}_{A,\tau}$ be each the last $m_A - n_A$ columns of $\mathbf{U}_1$, $\mathbf{Y}'_B$ and $\mathbf{X}_A$ respectively. Then the ML estimates of $\mathbf{H}_{B,A}$ and $\mathbf{U}_\tau$ are given by the LS solution to (18) and (19) with $\mathbf{H}^T = \mathbf{H}_{B,A}$, i.e.,

$$[\hat{\mathbf{H}}_{B,A}, \hat{\mathbf{U}}_\tau] = [\mathbf{T}_1, \mathbf{Y}'_{B,\tau}] \begin{bmatrix} \mathbf{X}_A\mathbf{X}_A^H & \mathbf{X}_{A,\tau} \\ \mathbf{X}_{A,\tau}^H & \mathbf{I}_{m_A - n_A} \end{bmatrix}^{-1} \quad (35)$$

with $\mathbf{T}_1 = (\mathbf{Y}'_{B,\gamma} - \mathbf{U}_\gamma)\mathbf{X}_{A,\gamma}^H + \mathbf{Y}'_{B,\tau}\mathbf{X}_{A,\tau}^H$.

Note that unlike the case where a reciprocal channel is fully exploited, the complexity of the above method is much lower. Furthermore, if we know that $n_A > n_E \geq n_B$, then the optimal choice of m_B in terms of SKC-DoF can be chosen to be n_B. In this case, $\mathbf{U}_2$ is empty, and the estimation of $\mathbf{H}_{A,B}$ is no longer needed. In other words, if $m_B = n_B$, Alice only needs the optimal estimate of the $n_B \times (m_A - n_A)$ matrix $\mathbf{U}_\tau$ as given in (35), which can be further written (using block matrix inversion) as

$$\hat{\mathbf{U}}_\tau = \mathbf{T}_1\mathbf{T}_2 + \mathbf{Y}'_{B,\tau}\mathbf{T}_3 \quad (36)$$

with $\mathbf{T}_2 = -(\mathbf{X}_{A,\gamma}\mathbf{X}_{A,\gamma}^H)^{-1}\mathbf{X}_{A,\tau}$ and $\mathbf{T}_3 = \mathbf{I}_{m_A - n_A} + \mathbf{X}_{A,\tau}^H(\mathbf{X}_{A,\gamma}\mathbf{X}_{A,\gamma}^H)^{-1}\mathbf{X}_{A,\tau}$. Alice and Bob can then use the pair of secret vectors $vec(\hat{\mathbf{U}}_\tau)$ and $vec(\mathbf{U}_\tau)$, respectively, to generate the final secret key.

V. Simulation

For simulation, we will normalize all signals such that the entries of $\mathbf{U}$, $\mathbf{X}_A$ and $\mathbf{X}_B$ all have the unit variance while the entries of $\mathbf{W}_A$ and $\mathbf{W}_B$ all have the variance equal to $\frac{1}{P}$.

A. The case of $|\rho| = 1$

For the reciprocal channel case, we considered $n_A = 8$, $n_B = 2$, $m_A = 128$ and $m_B = 4$. The algorithm for this case relies on an initial estimate of $\mathbf{U}_0$ from multiple solutions to (26). For $n_B = 2$, our proposed algorithm based on (27) can yield up to four solutions for $\hat{\mathbf{U}}_0$. Recall that for each $\hat{\mathbf{U}}_0$, there are corresponding $\hat{\mathbf{U}}_1$ and $\hat{\mathbf{U}}_2$ from (21) and (22). Among the multiple choices of $\hat{\mathbf{U}} = [\hat{\mathbf{U}}_0, \hat{\mathbf{U}}_1, \hat{\mathbf{U}}_2]$, we will pick the one which has the smallest value of $\|\hat{\mathbf{U}}\|_F^2$. This is the initial estimate of $\mathbf{U}$ before the use of gradient search shown in (30). An improved estimate of $\mathbf{U}$ follows after the gradient search.

We have run the above algorithm for each of 1000 independent realizations of $\mathbf{U}$, $\mathbf{X}_A$, $\mathbf{X}_B$, $\mathbf{H}$, $\mathbf{W}_A$ and $\mathbf{W}_B$ at different values of $\mathtt{SNR(dB)} \doteq 10\log_{10} P$. Figs. 2, 3 and 4 show the distributions of the element-wise squared errors (in log scale) of the entries in $\hat{\mathbf{U}}_0$, $\hat{\mathbf{U}}_1$ and $\hat{\mathbf{U}}_2$ respectively. As expected, the gradient search reduces the error distributions. The MSE values are heavily influenced by the larger errors. We also see that the median errors on $\hat{\mathbf{U}}_0$, $\hat{\mathbf{U}}_1$ and $\hat{\mathbf{U}}_2$ are somewhat different.

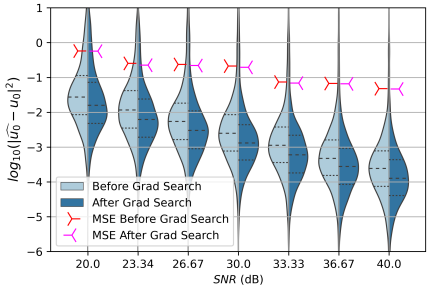

Fig. 2. Distributions ("violin" plots) of squared estimation errors in the elements of $\hat{\mathbf{U}}_0$. The lightly shaded curves facing the left and the heavily shaded curves facing the right are histograms.

Furthermore, we see that the errors in the "upper tails" shown in Figs. 2, 3 and 4 are relatively large although they represent a small percentage. These large errors are mostly due to the situations where a desired solution to (26) was not obtained by the algorithm based on (27) subject to a given number (chosen up to 5 in simulation) of random

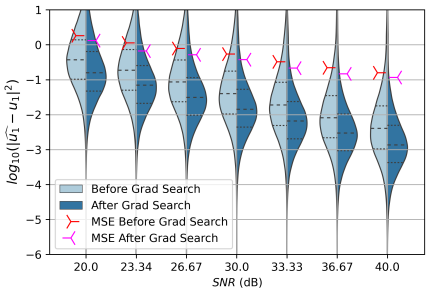

Fig. 3. Distributions of squared estimation errors in the elements of $\hat{\mathbf{U}}_1$.

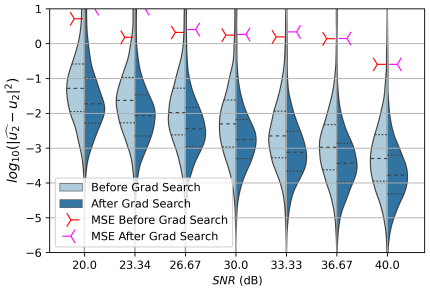

Fig. 4. Distributions of squared estimation errors in the elements of $\hat{\mathbf{U}}_2$.

initializations of $\hat{\mathbf{U}}_0$. In order to show the performances that are not overly impacted by such situations, we also considered a thresholding that only accepted the initial estimates of $\mathbf{U}$ which satisfy $\frac{1}{n_U}\|\hat{\mathbf{U}}\|_F^2 < 1.5$. Note that in the absence of noise, $\lim_{n_U\to\infty}\frac{1}{n_U}\|\mathbf{U}\|_F^2 = 1$. The percentages of the accepted realizations is shown in Fig. 5. An effect of the thresholding on the estimated $\mathbf{U}_1$ (for example) can be seen by comparing Fig. 6 with Fig. 3.

B. For the case of $|\rho| < 1$

In this case, we also considered $n_A = 8$, $n_B = 2$, $m_A = 128$ and $m_B = 4$, and chose independent $\mathbf{H}_{A,B}$ and $\mathbf{H}_{B,A}$. The performance of the algorithm (36) is illustrated in Fig. 7

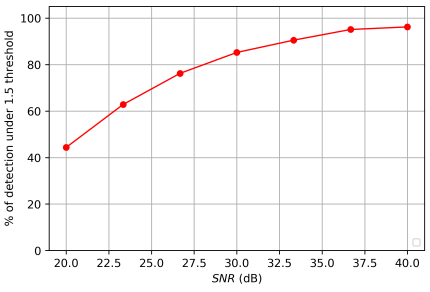

Fig. 5. Percentages of accepted realizations after thresholding.

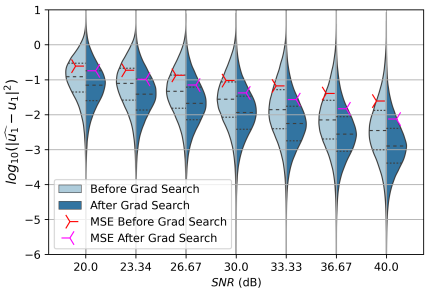

Fig. 6. Distributions of squared estimation errors in the elements of $\hat{\mathbf{U}}_1$ with thresholding.

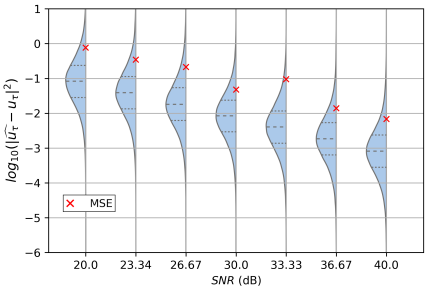

Fig. 7. Distributions of squared estimation errors in the elements of $\hat{\mathbf{U}}_\tau$.

REFERENCES

[1] R. Wilson, D. Tse, and R. A. Scholtz, "Channel identification: secret sharing using reciprocity in ultrawideband channels," IEEE Trans. Inf. Forensics Secur., vol. 2, no. 3, pp. 364–375, Sep. 2007.

[2] J. W. Wallace and R. K. Sharma, "Automatic secret keys from reciprocal MIMO wireless channels: measurement and analysis," IEEE Transactions on Information Forensics and Security, vol. 5, no. 3, pp. 381-392, Sept. 2010.

[3] N. Aldaghri and H. Mahdavifar, "Physical layer secret key generation in static environments", IEEE Transactions on Information Forensics and Security, Vol. 15, pp. 2692-2705, Feb. 2020.

[4] G. Li, H. Yang, J. Zhang, H. Liu, and A. Hu, "Fast and secure key generation with channel obfuscation in slowly varying environments," IEEE INFOCOM 2022, May 2022.

[5] J. Zhang, G. Li, A. Marshall, A. Hu, and L. Hanzo, "A new frontier for IoT security emerging from three decades of key generation relying on wireless channels," IEEE Access, vol. 8, pp. 138406–138446, 2020.

[6] D. E. Simmons, N. Bhargav, J. P. Coon, S. L. Cotton, "Physical layer security over OFDM-based links: conjugate-and-return," 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), Glasgow, UK.

[7] T.-Y. Liu, P. Mukherjee, S. Ulukus, S.-C. Lin, and Y.-W. P. Hong, "Secure degrees of freedom of MIMO Rayleigh block fading wiretap channels with no CSI anywhere," IEEE Trans. Wireless Commun., vol. 14, no. 5, pp. 2655–2669, May 2015.

[8] S. S. Pradhan and K. Ramchandran, "Distributed source coding using syndromes (DISCUS): Design and construction," IEEE Trans. Inf. Theory, vol. 49, no. 3, pp. 626–643, Mar. 2003.

[9] A. Maksud and Y. Hua, "Secret key generation by continuous encryption before quantization," IEEE Signal Processing Letters, Vol. 29, pp. 1497-1501, June 2022.

[10] Y. Hua and A. Maksud, "Continuous encryption functions for security over networks," Signal Processing, Volume 203, pp. 1-15, Feb. 2023.

[11] C. Huth, R. Guillaume, T. Strohm, P. Duplys, and I. A. Samuel, "Information reconciliation schemes in physical-layer security: A survey," Comput. Netw., vol. 109, pp. 84-104, 2016.

[12] Y. Dodis, B. Kanukurthi, J. Katz, L. Reyzin, and A. Smith, "Robust fuzzy extractors and authenticated key agreement from close secrets," IEEE Transactions on Information Theory, Vol. 58, No. 9., pp. 6207-6222, Sept 2012.

[13] U. M. Maurer, "Secret key agreement by public discussion from common information," IEEE Transactions on Information Theory, Vol. 39, No. 3, pp. 733-742, May 1993.

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

[15] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd., Wiley, 2006.

[16] Y. Hua, "Generalized channel probing and generalized pre-processing for secret key generation," IEEE Transactions on Signal Processing, submitted Nov 2022. Available upon request.