Secure Degrees of Freedom of MIMO Rayleigh Block Fading Wiretap Channels with No CSI Anywhere

Ta-Yuan Liu, Pritam Mukherjee, Sennur Ulukus, Shih-Chun Lin,
and Y.-W. Peter Hong

Abstract

We consider the block Rayleigh fading multiple-input multiple-output (MIMO) wiretap channel with no prior channel state information (CSI) available at any of the terminals. The channel gains remain constant within a coherence interval of T symbols, and then change to another independent realization in the next coherence interval. The transmitter, the legitimate receiver and the eavesdropper have n_t, n_r and n_e antennas, respectively. We determine the exact secure degrees of freedom (s.d.o.f.) of this system when $T \geq 2\min(n_t, n_r)$. We show that, in this case, the s.d.o.f. is exactly equal to $(\min(n_t, n_r) - n_e)^+(T - \min(n_t, n_r))/T$. The first term in this expression can be interpreted as the eavesdropper with n_e antennas taking away n_e antennas from both the transmitter and the legitimate receiver. The second term can be interpreted as a fraction of the s.d.o.f. being lost due to the lack of CSI at the legitimate receiver. In particular, the fraction loss, $\min(n_t, n_r)/T$, can be interpreted as the fraction of channel uses dedicated to training the legitimate receiver for it to learn its own CSI. We prove that this s.d.o.f. can be achieved by employing a constant norm channel input, which can be viewed as a generalization of discrete signalling to multiple dimensions.

I. Introduction

We consider the wiretap channel where a legitimate transmitter wishes to have information-theoretically secure communication with a legitimate receiver in the presence of an eavesdropper. The wiretap channel was introduced by Shannon [1] for the case of noiseless channels,

T.-Y. Liu and Y.-W. Peter Hong (emails: tyliu@erdos.ee.nthu.edu.tw and ywhong@ee.nthu.edu.tw) are with the Institute of Communications Engineering, National Tsing Hua University, Hsinchu, Taiwan 30013. P. Mukherjee and S. Ulukus (emails: pritamm@umd.edu and ulukus@umd.edu) are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742. S.-C. Lin (E-mail:sclin@mail.ntust.edu.tw.) is with the Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan 10607. This work was supported in part by the National Science Council, Taiwan, under grants NSC 100-2628-E-007-025-MY3, and US National Science Foundation Grants CNS 09-64632, CCF 09-64645, CCF 10-18185 and CNS 11-47811, and was presented in part at IEEE ICC, Sydney, Australia, June 2014.

where it was shown that secure keys and one-time-pad encryption were necessary for secure communications. The noisy wiretap channel was introduced by Wyner, who determined the capacity-equivocation region for the degraded case [2]. Csiszár and Körner generalized his result to arbitrary, not necessarily degraded, wiretap channels [3]. Leung-Yan-Cheong and Hellman determined the capacity-equivocation region of the Gaussian wiretap channel and showed that Gaussian signalling is optimal [4]. The secure degrees of freedom (s.d.o.f.) of the scalar Gaussian wiretap channel is zero.

The multiple-input multiple-output (MIMO) wiretap channel where the legitimate entities and the eavesdropper have multiple antennas was considered for the 2-2-1 case in [5] and the general case in [6]–[8]. These references determined the exact secrecy capacity of the MIMO wiretap channel for the case of full channel state information (CSI) at all terminals, and showed that no channel prefixing is necessary and Gaussian signalling is optimal. It can be deduced from these works that the s.d.o.f. of the MIMO wiretap channel with full CSI is $\min((n_t - n_e)^+, n_r)$, where n_t, n_r and n_e are the number of antennas at the transmitter, the legitimate receiver, and the eavesdropper, respectively, and $(x)^+ = \max(x, 0)$.

The fading wiretap channel with a single antenna at all terminals, where all parties have perfect CSI of all links, was considered in [9]–[12]. Modeling the fading wiretap under full CSI as a bank of independent parallel channels, these references showed that independent Gaussian signalling in all parallel channels, together with water-filling of the total power over these channels, is optimal. Reference [13] considered the single antenna wiretap channel where the transmitter has the legitimate receiver's CSI but no eavesdropper CSI under the assumption of infinite coherence times for channel fading, and showed that Gaussian signalling is optimal in this case. Reference [14] considered the same model under a fast fading condition (single symbol coherence time), and showed that M-QAM signalling or Gaussian signalling with added Gaussian artificial noise may outperform plain Gaussian signalling. In the single antenna fading channel, under all CSI conditions, the s.d.o.f. is zero, since it is zero under perfect CSI.

Using multiple antennas at the legitimate users however, non-zero s.d.o.f. may be achieved even under partial CSI conditions. Reference [15] showed that in a MIMO wiretap channel with perfect CSI at the receivers, but only a statistical CSI at the transmitter, under a fast fading Rayleigh channel, the s.d.o.f. of the system is $(\min(n_t, n_r) - n_e)^+$. Note that this may be less than the s.d.o.f. achievable under perfect CSI, which is $\min((n_t - n_e)^+, n_r)$. A comparison of these two s.d.o.f. may be interpreted as the eavesdropper taking away n_e antennas only from the transmitter in the case of perfect CSI [5]–[8], but n_e antennas from both the transmitter and the legitimate receiver in the case of partial CSI [15]. More strongly, reference [16] considered the case of an arbitrarily varying eavesdropper in a MIMO wiretap channel and showed that the same s.d.o.f. of $(\min(n_t, n_r) - n_e)^+$ can be achieved in this case. In [16], the CSI of the legitimate receiver is assumed known at the transmitter, however, nothing is known about the eavesdropper CSI, not even its probability distribution. This is an exceptionally strong modeling of the eavesdropper, where secrecy must be guaranteed for every realization of the eavesdropper channel; in a way, the eavesdropper may be thought to be controlling its channel adversarially.

All of the above work considered that some (either perfect or partial) CSI is available at some of the terminals. In practice, typically, the way CSI becomes available at the terminals is via the receivers measuring it and feeding it back to the transmitters. It is reasonable to assume that no CSI is known at the outset before the start of the communication. One must then take into consideration the cost of acquiring the CSI. In addition, the assumption of perfect CSI is an idealization; in reality, the terminals may only have an estimate of the channel in a delayed manner as discussed in [17]–[19]. Further, in most cases, eavesdropper CSI will not be available at the transmitter, because she will not feed her measurement back, and even if she does, she will not be truthful. Thus, it is more practical to assume that no CSI is available at any terminal a priori. Recently, reference [20] studied the case where no CSI is available at any terminal and the coherence time of the Rayleigh fading channel is one symbol duration. Reference [20] determined the exact secrecy capacity in this case and showed that discrete signalling is optimal. As in all other single antenna cases, the s.d.o.f. in [20] is zero. It can be shown that, even when multiple antennas are added, s.d.o.f. in the case of fast fading in [20] is still zero.

In this paper, we consider the MIMO wiretap channel under block Rayleigh fading, where the channel gains of both the legitimate receiver and the eavesdropper remain fixed for a coherence interval of T symbols, and then change to another independent realization in the next coherence interval. This models a Rayleigh fading wireless communication channel with a coherence time of T symbol durations. We consider the case where neither the transmitter nor the receivers have any CSI. This can be considered as an extension of [20] to the case of multiple antennas and larger (than one) coherence times. A similar channel model without any secrecy constraints was considered in [21], [22], where in [21] the structure of the optimal input distribution was found, and in [22] the degrees of freedom (d.o.f.) was determined to be $m(1 - m/T)$ where $m = \min(n_t, n_r, \lfloor T/2 \rfloor)$. Our work can also be considered as a wiretap version of [21], [22].

We show that when the coherence time T satisfies $T \geq 2\min(n_t, n_r)$, the s.d.o.f. of this system is exactly $(\min(n_t, n_r) - n_e)^+(T - \min(n_t, n_r))/T$. Compared to the MIMO wiretap channel results in [15], [16], where the legitimate receiver knows its channel gain, the s.d.o.f. in our case is exactly the same as those in [15], [16] except for a factor of $(T - \min(n_t, n_r))/T$. Intuitively, at high signal-to-noise ratio (SNR), the legitimate receiver needs $\min(n_t, n_r)$ channel uses out of T channel uses to learn its channel. Therefore, the factor $(T - \min(n_t, n_r))/T$ intuitively accounts for the number of channel uses lost for estimating the channel at the legitimate receiver. As in the cases of [15], [16], due to no CSI at the transmitter, the eavesdropper takes away n_e antennas from both the transmitter and the receiver, i.e., n_e is subtracted from $\min(n_t, n_r)$, as opposed to being subtracted only from n_t as in the case of full CSI at the transmitter [5]–[8]. In comparison to the case without any secrecy constraints in [21], [22], here we have a subtraction of n_e from the first term in the d.o.f. due to the presence of the eavesdropper.

Finally, it is interesting to note that one cannot achieve a positive s.d.o.f with either a long coherence time in a single antenna system [13] or with multiple antennas in a very short ($T = 1$) coherence time channel [20]; however, with some moderate coherence ($T \geq 2\min(n_t, n_r)$) and use of multiple antennas, it is possible to achieve positive s.d.o.f., as we show in this paper.

II. System Model and Problem Formulation

Consider a wiretap channel that consists of a transmitter with n_t antennas, a legitimate receiver with n_r antennas, and an eavesdropper with n_e antennas. The channel between the transmitter and the legitimate receiver is denoted by matrix $\mathbf{H} \in \mathcal{C}^{n_r \times n_t}$ and the channel between the transmitter and the eavesdropper is denoted by matrix $\mathbf{G} \in \mathcal{C}^{n_e \times n_t}$. The channels are Rayleigh fading, i.e., entries of the channel matrices are independent and identically distributed (i.i.d.) complex Gaussian random variables with zero-mean and unit-variance denoted by $\mathcal{CN}(0, 1)$. The channels are block fading, i.e., the channel coefficients remain constant throughout a coherence interval T and change independently across different intervals according to the same distribution.

Let $\mathbf{X} \in \mathcal{C}^{n_t \times T}$ denote the signal transmitted by the transmitter during a coherence interval. The transmitted signal is subject to an average power constraint as,

$$\frac{1}{T}\mathbb{E}\left[\mathrm{tr}(\mathbf{X}\mathbf{X}^\dagger)\right] \leq P, \tag{1}$$

where $\mathrm{tr}(\cdot)$ denotes the trace function. The received signals at the legitimate receiver and the eavesdropper are

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}_r, \tag{2}$$

$$\mathbf{Z} = \mathbf{G}\mathbf{X} + \mathbf{N}_e, \tag{3}$$

respectively, where $\mathbf{N}_r \in \mathcal{C}^{n_r \times T}$ and $\mathbf{N}_e \in \mathcal{C}^{n_e \times T}$ are their respective additive Gaussian noise terms. The entries of $\mathbf{N}_r$ and $\mathbf{N}_e$ are i.i.d. with distributions $\mathcal{CN}(0, \sigma_r^2)$ and $\mathcal{CN}(0, \sigma_e^2)$, respectively. The CSI, i.e., the realizations of $\mathbf{H}$ and $\mathbf{G}$, are not known to any of the terminals.

A $(2^{nR}, n)$ code consists of an encoder f_n at the transmitter that maps each secret message, say $W \in \mathcal{W} \triangleq \{1, \ldots, 2^{nR}\}$ into a length-n codeword and a decoder g_n at the legitimate receiver that maps its received signal into a message estimate $\hat{W} \in \mathcal{W}$. Each codeword is transmitted over multiple coherence intervals [21] and n is chosen as a multiple of T.

A secrecy rate R is said to be achievable if there exists an encoder f_n and a decoder g_n such that the probability of error at the legitimate receiver $\mathbb{P}\left[W \neq \hat{W}\right]$ goes to zero and the average equivocation at the eavesdropper measured by $\frac{1}{n}H(W|\mathbf{Z}^n)$ approaches $\frac{1}{n}H(W)$, as the codeword length $n \to \infty$, where $\mathbf{Z}^n$ denotes the signal received at the eavesdropper over n channel uses. The secrecy capacity C_s is the supremum of all such achievable secrecy rates. From [3], the secrecy capacity of the MIMO wiretap channel with no CSI at any terminal is

$$C_s = \frac{1}{T}\max_{V,\mathbf{X}} \; I(V;\mathbf{Y}) - I(V;\mathbf{Z}), \tag{4}$$

where V is an auxiliary random variable that satisfies the Markov chain $V \to \mathbf{X} \to \mathbf{Y}, \mathbf{Z}$. Determining the optimal joint distribution of $(V, \mathbf{X})$ and the resulting exact secrecy capacity expression is challenging, instead, in this paper, we focus on determining the s.d.o.f. which is defined as,

$$D_s = \lim_{P \to \infty} \frac{C_s}{\log P}. \tag{5}$$

The s.d.o.f. characterizes how the secrecy capacity scales with $\log(P)$ for large P, i.e., it is the pre-log factor of the secrecy capacity at high SNR.

III. Summary of the Main Results

In this section, we first summarize our main results; the proofs will be provided in the following sections. The results are encapsulated in the following lemmas and theorem.

Lemma 1 *For the MIMO wiretap channel in* (2)-(3)*, with no CSI at any terminal,*

$$D_s = 0, \quad \textit{if } n_r \leq n_e. \tag{6}$$

This implies a negative result that when the eavesdropper has more antennas than the legitimate user, i.e., $n_r \leq n_e$, the s.d.o.f. D_s is always zero. No matter how long the coherence time T is and how many transmitter antennas the system has, the secrecy capacity does not scale with the SNR. However, we show in the following lemmas that a positive s.d.o.f. can be achieved, for $n_r > n_e$ and $T \geq 2\min(n_t, n_r)$.

Lemma 2 *When $n_r > n_e$, $n_r \leq n_t$, and $T \geq 2n_r$, the s.d.o.f. is given by*

$$D_s = (n_r - n_e)\left(\frac{T - n_r}{T}\right). \tag{7}$$

Lemma 3 *When $n_r > n_e$, $n_r > n_t$, and $T \geq 2n_t$, the s.d.o.f. is given by*

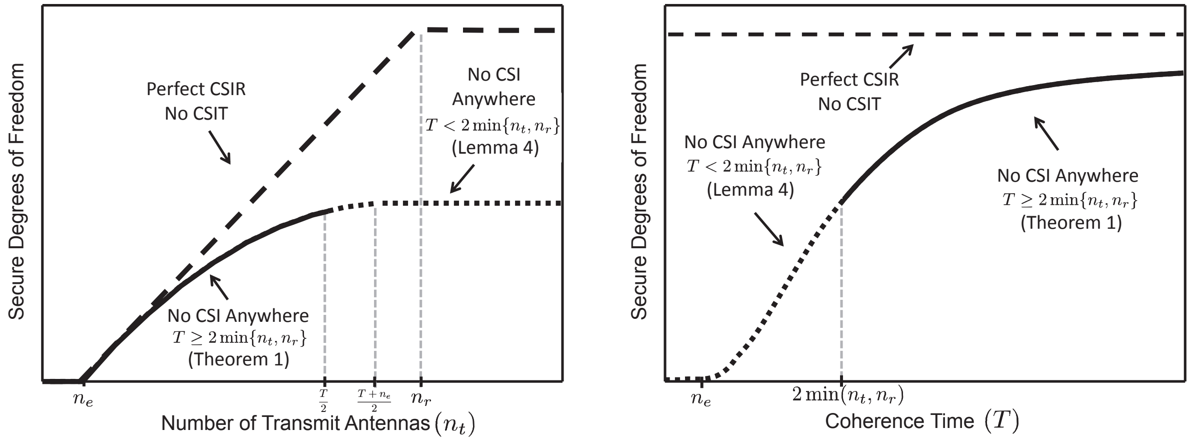
$$D_s = (n_t - n_e)^+\left(\frac{T - n_t}{T}\right). \tag{8}$$

Lemma 2 considers the case where the transmitter has more antennas than the receiver, whereas Lemma 3 considers the opposite case. Note that, in the latter case, a positivity operator $(\cdot)^+$ is required since n_t may be less than n_e. We combine the above three lemmas to obtain the following main result of our paper.

Theorem 1 *For the MIMO wiretap channel in* (2)-(3)*, with no CSI at any terminal, when $T \geq 2\min(n_t, n_r)$, the s.d.o.f. is given by*

$$D_s = (\min(n_t, n_r) - n_e)^+\left(\frac{T - \min(n_t, n_r)}{T}\right). \tag{9}$$

Note that when no secrecy constraint is considered, i.e., $n_e = 0$, the s.d.o.f. in Theorem 1 reduces to the d.o.f. of the noncoherent MIMO Rayleigh block fading channel [22]. Our s.d.o.f. is affected by two factors: the first factor $(\min(n_t, n_r) - n_e)^+$ is the s.d.o.f of the case where perfect CSI is available at the receivers [15] (i.e., where there is no cost due to lack of channel knowledge



(a) (b)

Fig. 1. Illustrations of the s.d.o.f. derived in Theorem 1 and Lemma 4. (a) The s.d.o.f. versus the number of transmit antennas n_t for $n_r > n_e$ and $2n_r > T + n_e$. (b) The s.d.o.f. versus coherence time T for $\min(n_t, n_r) > n_e$.

at the receiver), and the second factor $(1 - \min(n_t, n_r)/T)$ reflects the loss in efficiency due to the lack of knowledge of the CSI at the legitimate receiver. One can view the ratio $\min(n_t, n_r)/T$ as the cost of channel estimation at the legitimate receiver from the point of view of a training based scheme. Note that, even though Lemmas 2 and 3 (and, thus, Theorem 1) hold only for the case where $T \geq 2\min(n_t, n_t)$, the signalling scheme adopted in their achievability proofs can also be used to derive an achievable s.d.o.f. for the case where $T < 2\min(n_t, n_r)$, as given in the following lemma.

Lemma 4 *For arbitrary coherence time T, the s.d.o.f. satisfies*

$$D_s \geq (K - n_e)^+ \left(\frac{T - K}{T}\right) \tag{10}$$

where $K = \min(n_t, n_r, (T + n_e)/2)$.

Lemma 4 shows that, for given coherence time T and the number of eavesdropper's antennas n_e, the achievable s.d.o.f. for the case where $T < 2\min(n_t, n_r)$ increases with the number of legitimate antennas n_t until it reaches n_r or $(T + n_e)/2$. Even though more antennas at the transmitter and the legitimate receiver provides more dimensions for communication, it also implies that more resource is needed to cope with the lack of CSI at the receiver, which is reflected in the term $(T - K)/T$. More details can be found in Section VII.

The above main results can be visualized in Fig. 1. Here, we show the s.d.o.f. of the case with no CSI anywhere (i.e., Theorem 1 and Lemma 4) and compare with that of the case with perfect CSIR but no CSIT (or statistical CSIT) [15]. In Fig. 1(a), the number of receive antennas at the legitimate receiver and the eavesdropper (i.e., n_r and n_e, respectively) and the coherence time T are fixed and are chosen such that $n_r > n_e$ and $2n_r > T + n_e$. By varying the number of transmit antennas n_t, Theorem 1 shows that the s.d.o.f. with no CSI anywhere is zero when $n_t \leq n_e$, but increases with the number of transmit antennas when $n_e < n_t \leq T/2$. The increase is nonlinear as opposed to the case with perfect CSIR. However, as n_t increases beyond $T/2$, Theorem 1 no longer applies and the achievable s.d.o.f. in Lemma 4 is plotted instead (in dotted line). We can see that the achievable s.d.o.f. continues to increase with n_t when $T/2 < n_t < (T + n_e)/2$ and saturates when $n_t \geq (T + n_e)/2$. In Fig. 1(b), we show the s.d.o.f. versus coherence time T for the case where $\min(n_t, n_r) > n_e$. For $T \geq 2\min(n_t, n_r)$, the s.d.o.f. is given by Theorem 1 and is shown to approach that of the case with perfect CSIR as T increases. This is due to the fact that, when coherence time is sufficiently large, the impact due to lack of CSI can be neglected. Similarly, when $T < 2\min(n_t, n_r)$, the achievable s.d.o.f. in Lemma 4 is plotted instead.

IV. Proof of Lemma 1

To prove Lemma 1, we will in fact prove the following stronger result for this case:

$$C_s \leq \left[n_e \log\left(1 + \frac{P}{\sigma_r^2}\right) - n_e \log\left(1 + \frac{P}{\sigma_e^2}\right)\right]^+. \tag{11}$$

In order to derive the upper bound (11) on the secrecy capacity, we first note that for a fixed n_e, the secrecy capacity of the MIMO wiretap channel with $n_r = n_e$ is always greater than or equal to that of the case with $n_r < n_e$. Hence, it suffices to upper bound the secrecy capacity of the system with $n_r = n_e$, which we will call the *enhanced wiretap channel*.

For the enhanced wiretap channel, if $\sigma_r^2 \geq \sigma_e^2$, it is clear that the legitimate receiver is stochastically degraded with respect to the eavesdropper. Hence, the secrecy capacity in this case is zero. However, if $\sigma_r^2 < \sigma_e^2$, using the two conditions $n_r = n_e$ and $\sigma_r^2 < \sigma_e^2$, we can construct a physically degraded wiretap channel whose marginal distributions are identical to those of (2)-(3). The received signals of the equivalent degraded wiretap channel are

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}_r, \tag{12}$$

$$\mathbf{Z}' = \mathbf{H}\mathbf{X} + \mathbf{N}_r + \mathbf{N}'_e = \mathbf{Y} + \mathbf{N}'_e, \tag{13}$$

where the entries of $\mathbf{N}'_e \in \mathcal{C}^{n_e \times T}$ are i.i.d. Gaussian with zero-mean and variance $\sigma_e^2 - \sigma_r^2$, and $\mathbf{N}'_e$ is independent of $\mathbf{X}$, $\mathbf{H}$, and $\mathbf{N}_r$. Since the secrecy capacity depends only on the conditional marginal probabilities $p(\mathbf{Y}|\mathbf{X})$ and $p(\mathbf{Z}|\mathbf{X})$, and $\mathbf{H}$ and $\mathbf{G}$ are statistically the same, the physically degraded channel in (12)-(13) has the same secrecy capacity as the original stochastically degraded channel in (2)-(3). Due to the degradedness of the equivalent model in (12)-(13), we know, from [2], [3], that $V = \mathbf{X}$ is optimal (i.e., no channel prefixing is needed) and, thus, the secrecy capacity of the equivalent degraded wiretap channel is

$$C_s = \frac{1}{T} \max_{p_{\mathbf{X}} \in S_{p_{\mathbf{X}}}} \; I(\mathbf{X};\mathbf{Y}) - I(\mathbf{X};\mathbf{Z}'), \tag{14}$$

where $S_{p_{\mathbf{X}}}$ denotes the set of all input distributions which satisfy the power constraint in (1).

To derive an upper bound we first rewrite (14) as

$$T \cdot C_s = \max_{p_{\mathbf{X}} \in S_{p_{\mathbf{X}}}} h(\mathbf{Y}) - h(\mathbf{Z}') - h(\mathbf{Y}|\mathbf{X}) + h(\mathbf{Z}'|\mathbf{X}). \tag{15}$$

Now we note that if $n_r = n_e$ and $\sigma_r^2 < \sigma_e^2$, we have the following inequality for the wiretap channel in (12)-(13),

$$h(\mathbf{Y}|\mathbf{X}) - h(\mathbf{Z}'|\mathbf{X}) \geq h(\mathbf{Y}|\mathbf{X},\mathbf{H}) - h(\mathbf{Z}'|\mathbf{X},\mathbf{H}) \tag{16}$$

This is a vector generalization of [20, eqn. (12)], and can be proved by observing that (16) holds if and only if

$$I(\mathbf{Y};\mathbf{H}|\mathbf{X}) \geq I(\mathbf{Z}';\mathbf{H}|\mathbf{X}), \tag{17}$$

which is true since,

$$I(\mathbf{Y};\mathbf{H}|\mathbf{X}=\tilde{\mathbf{X}}) = n_r \log\left|\mathbf{I}_T + \frac{\tilde{\mathbf{X}}^\dagger\tilde{\mathbf{X}}}{\sigma_r^2}\right| \geq n_e \log\left|\mathbf{I}_T + \frac{\tilde{\mathbf{X}}^\dagger\tilde{\mathbf{X}}}{\sigma_e^2}\right| = I(\mathbf{Z}';\mathbf{H}|\mathbf{X}=\tilde{\mathbf{X}}) \tag{18}$$

where $\tilde{\mathbf{X}}$ denotes a realization of the random matrix $\mathbf{X}$. In deriving (18), we used the fact that $n_r = n_e$, $\sigma_r^2 < \sigma_e^2$ and that given $\mathbf{X}$, $\mathbf{Y}$ and $\mathbf{H}$ are jointly Gaussian and so are $\mathbf{Z}$ and $\mathbf{H}$. If $\mathbf{Y}_i$

denotes the ith row of $\mathbf{Y}$, then $\mathbf{Y}_i$ is a Gaussian vector independent of $\mathbf{Y}_j$, for all $i \neq j$, and the covariance matrix of $\mathbf{Y}_i$ is $\tilde{\mathbf{X}}^\dagger \tilde{\mathbf{X}} + \sigma_r^2 \mathbf{I}_T$, for all i.

Using (16) and (13) in (15), we obtain

$$\begin{aligned} T \cdot C_s &\leq \max_{p_{\mathbf{X}} \in S_{p_{\mathbf{X}}}} h(\mathbf{Y}) - h(\mathbf{Y} + \mathbf{N}'_e) - h(\mathbf{Y}|\mathbf{X}, \mathbf{H}) + h(\mathbf{Z}'|\mathbf{X}, \mathbf{H}) && (19) \\ &= \max_{p_{\mathbf{X}} \in S_{p_{\mathbf{X}}}} h(\mathbf{Y}) - h(\mathbf{Y} + \mathbf{N}'_e) + n_e T \log\left(\frac{\sigma_e^2}{\sigma_r^2}\right) && (20) \\ &\leq \max_{p_{\mathbf{X}} \in S_{p_{\mathbf{X}}}} h(\mathbf{Y}) - n_e T \log\left(e^{\frac{1}{n_e T} h(\mathbf{Y})} + e^{\frac{1}{n_e T} h(\mathbf{N}'_e)}\right) + n_e T \log\left(\frac{\sigma_e^2}{\sigma_r^2}\right) && (21) \\ &\leq h(\mathbf{Y}_G) - n_e T \log\left(e^{\frac{1}{n_e T} h(\mathbf{Y}_G)} + e^{\frac{1}{n_e T} h(\mathbf{N}'_e)}\right) + n_e T \log\left(\frac{\sigma_e^2}{\sigma_r^2}\right) && (22) \\ &= T n_e \log\left(1 + \frac{P}{\sigma_r^2}\right) - T n_e \log\left(1 + \frac{P}{\sigma_e^2}\right) && (23) \end{aligned}$$

where (21) follows from the entropy power inequality [23], and (22) follows from the fact that the right hand side of (21) is monotonically increasing in $h(\mathbf{Y})$, and that for a fixed total power constraint $\mathbb{E}[\mathrm{tr}(\mathbf{Y}\mathbf{Y}^\dagger)] \leq (P + \sigma_r^2) n_e T$, a Gaussian matrix $\mathbf{Y}_G \in \mathcal{C}^{n_e \times T}$ with entries that are i.i.d. Gaussian with zero-mean and variance $P + \sigma_r^2$ maximizes the differential entropy. This gives us the desired result in (11), completing the proof of Lemma 1.

V. Proof of Lemma 2

The proof of Lemma 2 is a bit more involved than that of Lemma 1. For clarity, we outline the steps of the proof here and leave the details to Appendix A. We first prove the converse and then provide a scheme that achieves the s.d.o.f. upper bound.

A. Converse Proof of Lemma 2

To find an upper bound for the s.d.o.f. D_s, we only need to consider the case where $\sigma_r^2 < \sigma_e^2$, since, with all other channel parameters remaining the same, the wiretap channel in (2)-(3) with $\sigma_r^2 < \sigma_e^2$ yields a larger secrecy capacity than that with $\sigma_r^2 \geq \sigma_e^2$. Under the assumption $\sigma_r^2 < \sigma_e^2$, we can once again construct a degraded equivalent channel (as we did in (12)-(13) for $n_r = n_e$), without changing C_s by selecting n_e row vectors from n_r rows of the legitimate channel matrix $\mathbf{H}$ to form a statistically marginally identical eavesdropper channel. For any fixed partition $\mathrm{p}_1 \cup \mathrm{p}_2 = \{1, \ldots, n_r\}$ where $|\mathrm{p}_1| = n_e$ and $\mathrm{p}_2 = \{1, \ldots, n_r\} \setminus \mathrm{p}_1$, we construct a degraded equivalent channel for (2)-(3) as follows

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}_r, \tag{24}$$

$$\mathbf{Z}_{\mathrm{p}_1} = \mathbf{H}_{\mathrm{p}_1}\mathbf{X} + \mathbf{N}_{r,\mathrm{p}_1} + \mathbf{N}'_e = \mathbf{Y}_{\mathrm{p}_1} + \mathbf{N}'_e, \tag{25}$$

where $\mathbf{H}_{\mathrm{p}_1}, \mathbf{N}_{r,\mathrm{p}_1}$ and $\mathbf{Y}_{\mathrm{p}_1}$ denote the collection of row vectors with indices belonging to p_1 from $\mathbf{H}, \mathbf{N}_r$ and $\mathbf{Y}$, respectively, and $\mathbf{Z}_{\mathrm{p}_1}$ denotes the equivalent eavesdropper's received signal constructed from $\mathbf{Y}_{\mathrm{p}_1}$. For any partition $(\mathrm{p}_1, \mathrm{p}_2)$, as in the proof of Lemma 1, the secrecy capacity of the degraded wiretap channel in (24)-(25) is

$$C_s = \frac{1}{T} \max_{p_{\mathbf{X}} \in S_{p_{\mathbf{X}}}} \; I(\mathbf{X}; \mathbf{Y}) - I(\mathbf{X}; \mathbf{Z}_{\mathrm{p}_1}). \tag{26}$$

From above, the optimization problem in (4) is transformed to a simpler problem which needs to be optimized only with respect to $\mathbf{X}$ as in (26). However, it is still hard to find the optimal input distribution $p_{\mathbf{X}}$. Instead, we characterize the optimal input structure with respect to (26) for the equivalent degraded channel given in (24)-(25). This helps us restrict possible input distributions and simplifies the problem. Interestingly, we show in the sequel that, due to the degradedness of the equivalent wiretap channel in (24)-(25) and the concavity of the secrecy rate in the input distribution for degraded channels [20], the optimal input structure in (26) is the same as the optimal input structure in the channel without secrecy constraints in [21].

Recall that a random matrix $\mathbf{M} \in \mathcal{C}^{N \times T}$ where $T \geq N$ is *isotropically distributed* (i.d.) if $p(\mathbf{M}) = p(\mathbf{M}\mathbf{U})$, for all deterministic $T \times T$ unitary matrices $\mathbf{U}$. The optimal input structure for the equivalent degraded wiretap channel in (24)-(25) is characterized in the following lemma.

Lemma 5 *When $n_r > n_e$ and $\sigma_r^2 < \sigma_e^2$, for the equivalent channel in* (24)-(25)*, the optimal input distribution that maximizes C_s in* (26) *has the structure*

$$\mathbf{X} = \mathbf{\Lambda}\mathbf{\Theta}, \tag{27}$$

if $T \geq n_t$, where $\mathbf{\Lambda}$ is an $n_t \times T$ diagonal random matrix with real and non-negative diagonal elements, and $\mathbf{\Theta}$ is a $T \times T$ i.d. unitary matrix which is independent of $\mathbf{\Lambda}$.

We provide a proof for Lemma 5 in Appendix A.

Although we cannot completely characterize the optimal $\mathbf{X}$, the result in Lemma 5 suffices to derive a useful upper bound for D_s. We can rewrite the secrecy capacity given in (26) and upper bound it as

$$T \cdot C_s = \max_{p_{\mathbf{X}} \in S^*_{p_{\mathbf{X}}}} I(\mathbf{X};\mathbf{Y}) - I(\mathbf{X};\mathbf{Z}_{\mathrm{p}_1}) \tag{28}$$

$$= \max_{p_{\mathbf{X}} \in S^*_{p_{\mathbf{X}}}} h(\mathbf{Y}_{\mathrm{p}_1}) + h(\mathbf{Y}_{\mathrm{p}_2}|\mathbf{Y}_{\mathrm{p}_1}) - h(\mathbf{Y}|\mathbf{X}) - h(\mathbf{Y}_{\mathrm{p}_1} + \mathbf{N}'_e) + h(\mathbf{Z}_{\mathrm{p}_1}|\mathbf{X}) \tag{29}$$

$$\leq \max_{p_{\mathbf{X}} \in S^*_{p_{\mathbf{X}}}} h(\mathbf{Y}_{\mathrm{p}_2}|\mathbf{Y}_{\mathrm{p}_1}) - h(\mathbf{Y}|\mathbf{X}) + h(\mathbf{Z}_{\mathrm{p}_1}|\mathbf{X}), \tag{30}$$

where $S^*_{p_{\mathbf{X}}}$ in (28) denotes the set of all input distributions having the optimal structure described in Lemma 5 and satisfying the power constraint in (1), matrix $\mathbf{Y}_{\mathrm{p}_2}$ in (29) is the collection of row vectors of $\mathbf{Y}$ with indices belonging to $\mathrm{p}_2 = \{1, \dots, n_r\} \backslash \mathrm{p}_1$, and the inequality (30) follows from $h(\mathbf{Y}_{\mathrm{p}_1}) \leq h(\mathbf{Y}_{\mathrm{p}_1} + \mathbf{N}'_e)$.

Now continuing from (30), we derive the desired upper bound in three steps.

Step 1: We derive an upper bound for $h(\mathbf{Y}_{\mathrm{p}_2}|\mathbf{Y}_{\mathrm{p}_1})$ in terms of $h(\mathbf{Y})$ so that we can focus only on $h(\mathbf{Y})$ later. This upper bound can be derived by using the following lemma.

Lemma 6 *Given an $m \times T$ random matrix $\mathbf{M}$ with differential entropy $h(\mathbf{M})$, for all $n \in \{1, \dots, m\}$, there must exist a partition $(\mathrm{p}_1, \mathrm{p}_2)$ where $\mathrm{p}_1 \cup \mathrm{p}_2 = \{1, \dots, m\}$, $|\mathrm{p}_1| = n$, and $|\mathrm{p}_2| = m - n$ such that*

$$h(\mathbf{M}_{\mathrm{p}_2}|\mathbf{M}_{\mathrm{p}_1}) \leq \frac{m-n}{m} h(\mathbf{M}), \tag{31}$$

where $\mathbf{M}_{\mathrm{p}_1}$ and $\mathbf{M}_{\mathrm{p}_2}$ denote the collection of row vectors of $\mathbf{M}$ with indices belonging to p_1 and p_2, respectively.

We provide a proof for Lemma 6 in Appendix B.

Now, from Lemma 6 and (30), we have

$$T \cdot C_s \leq \max_{p_{\mathbf{X}} \in S^*_{p_{\mathbf{X}}}} \frac{n_r - n_e}{n_r} h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}) + h(\mathbf{Z}_{\mathrm{p}_1}|\mathbf{X}), \tag{32}$$

which follows from the fact that, for any partition $(\mathrm{p}_1, \mathrm{p}_2)$, (30) is a valid upper bound. A similar inequality for (31) has been derived in [17] under the entropy symmetric condition which is not required in Lemma 6. However, it is necessary to note that the result in [17] is also applicable here since our problem coincidentally satisfies the entropy symmetric condition as well.

Step 2: We derive an upper bound for $h(\mathbf{Y})$ in (32), as given in the following lemma.

Lemma 7 *With the distribution of the channel input* $\mathbf{X}$ *satisfying the optimal structure in Lemma 5, the corresponding differential entropy of the legitimate receiver signal* $\mathbf{Y}$ *in* (12) *can be upper bounded as*

$$\max_{p_{\mathbf{X}} \in S^*_{p_{\mathbf{X}}}} h(\mathbf{Y}) \leq n_r^2 \log P + (T - n_r)\mathbb{E}\left[\log \det \mathbf{Y}\mathbf{Y}^\dagger\right] + o(\log P), \tag{33}$$

where $\lim_{P\to\infty} o(\log P)/\log P = 0$.

We provide a proof for Lemma 7 in Appendix C.

Note that given the input signal $\mathbf{X}$, each row vector of $\mathbf{Y}$ and $\mathbf{Z}_{\mathrm{p}_1}$ are i.i.d. Gaussian vectors, under the optimal input structure imposed by Lemma 5, the conditional differential entropy $h(\mathbf{Y}|\mathbf{X})$ and $h(\mathbf{Z}_{\mathrm{p}_1}|\mathbf{X})$ in (32) can be explicitly computed as

$$h(\mathbf{Y}|\mathbf{X}) = n_r \sum_{i=1}^{n_t} \mathbb{E}\left[\log \pi e(||\mathbf{X}_i||^2 + \sigma_r^2)\right] + n_r(T - n_t) \log \pi e \sigma_r^2, \tag{34}$$

$$h(\mathbf{Z}_{\mathrm{p}_1}|\mathbf{X}) = n_e \sum_{i=1}^{n_t} \mathbb{E}\left[\log \pi e(||\mathbf{X}_i||^2 + \sigma_e^2)\right] + n_e(T - n_t) \log \pi e \sigma_e^2, \tag{35}$$

where $\mathbf{X}_i$ is the ith row of the given input signal $\mathbf{X}$.

Now, by Lemma 7 and (32)-(35), we can further upper bound the secrecy capacity in (32) as

$$\begin{aligned} T \cdot C_s \leq \max_{p_{\mathbf{X}} \in S^*_{p_{\mathbf{X}}}} & \frac{n_r - n_e}{n_r}(T - n_r)\mathbb{E}\left[\log \det \mathbf{Y}\mathbf{Y}^\dagger\right] - (n_r - n_e)\sum_{i=1}^{n_t} \mathbb{E}\left[\log(||\mathbf{X}_i||^2 + \sigma_r^2)\right] \\ & + n_e \sum_{i=1}^{n_t} \mathbb{E}\left[\log\left(\frac{||\mathbf{X}_i||^2 + \sigma_e^2}{||\mathbf{X}_i||^2 + \sigma_r^2}\right)\right] + (n_r - n_e)n_r \log P + o(\log P). \end{aligned} \tag{36}$$

Furthermore, by using the fact that $\log(1 + x) \leq x$, it follows that

$$\mathbb{E}\left[\log\left(\frac{||\mathbf{X}_i||^2 + \sigma_e^2}{||\mathbf{X}_i||^2 + \sigma_r^2}\right)\right] \leq \mathbb{E}\left[\frac{\sigma_e^2 - \sigma_r^2}{||\mathbf{X}_i||^2 + \sigma_r^2}\right] \leq \frac{\sigma_e^2 - \sigma_r^2}{\sigma_r^2}, \tag{37}$$

where the right hand side of (37) is a constant independent of P. Therefore, by (36) and (37), we can upper bound the secrecy capacity $T \cdot C_s$ as

$$\begin{aligned} T \cdot C_s \leq \max_{p_{\mathbf{X}} \in S^*_{p_{\mathbf{X}}}} (n_r - n_e) \Bigg(& \frac{(T - 2n_r)}{n_r}\mathbb{E}\left[\log \det \mathbf{Y}\mathbf{Y}^\dagger\right] + \mathbb{E}\left[\log \det \mathbf{Y}\mathbf{Y}^\dagger\right] \\ & - \sum_{i=1}^{n_t} \mathbb{E}\left[\log(||\mathbf{X}_i||^2 + \sigma_r^2)\right] \Bigg) + (n_r - n_e)n_r \log P + o(\log P). \end{aligned} \tag{38}$$

By the assumptions $T \geq 2n_r$ and $n_r > n_e$, we obtain a further upper bound for (38) by developing upper bounds separately for $\mathbb{E}\left[\log\det \mathbf{Y}\mathbf{Y}^\dagger\right]$ and $\mathbb{E}\left[\log\det \mathbf{Y}\mathbf{Y}^\dagger\right] - \sum_{i=1}^{n_t} \mathbb{E}\left[\log(||\mathbf{X}_i||^2 + \sigma_r^2)\right]$, respectively. This is the task of step 3.

Step 3: We derive upper bounds for the two terms $\mathbb{E}\left[\log\det \mathbf{Y}\mathbf{Y}^\dagger\right]$ and $\mathbb{E}\left[\log\det \mathbf{Y}\mathbf{Y}^\dagger\right] - \sum_{i=1}^{n_t} \mathbb{E}\left[\log(||\mathbf{X}_i||^2 + \sigma_r^2)\right]$ in (38) separately using the following two lemmas.

Lemma 8 *With the distribution of the channel input* $\mathbf{X}$ *satisfying the optimal structure in Lemma 5, and with* $n_t \geq n_r$*, the legitimate received signal* $\mathbf{Y}$ *in* (12) *satisfies*

$$\max_{p_{\mathbf{X}} \in S^*_{p_{\mathbf{X}}}} \mathbb{E}\left[\log\det \mathbf{Y}\mathbf{Y}^\dagger\right] \leq n_r \log P + o(\log P), \tag{39}$$

where $\lim_{P\to\infty} o(\log P)/\log P = 0$.

Lemma 9 *With the distribution of the channel input* $\mathbf{X}$ *satisfying the optimal structure in Lemma 5, and with* $n_t \geq n_r$*, the legitimate received signal* $\mathbf{Y}$ *in* (12) *satisfies*

$$\max_{p_{\mathbf{X}} \in S^*_{p_{\mathbf{X}}}} \mathbb{E}\left[\log\det \mathbf{Y}\mathbf{Y}^\dagger\right] - \sum_{i=1}^{n_t} \mathbb{E}\left[\log(||\mathbf{X}_i||^2 + \sigma_r^2)\right] \leq k, \tag{40}$$

where k *is a constant which is independent of* P.

We provide proofs for Lemmas 8 and 9 in Appendices D and E, respectively. It should be mentioned that here we focus on the setting where $n_t \geq n_r$, and the random channel matrix $\mathbf{H}$ is not full column rank. Thus, the results of [22] cannot be directly applied to prove Lemmas 8 and 9. More discussion on this can be found at the end of Appendix E. In addition, in [22], where the conventional MIMO channel with no eavesdroppers was examined, a key step in proving the converse of the d.o.f. was the upper-bounding of the mutual information $I(\mathbf{X};\mathbf{Y})$. However, when proving the s.d.o.f. of wiretap channels, one instead needs to derive upper bounds for the difference in mutual information $I(\mathbf{X};\mathbf{Y}) - I(\mathbf{X};\mathbf{Z})$. The results in [22] do not apply in this case and, thus, new upper-bounding techniques are developed here.

Finally, using Lemmas 8 and 9 in (38), we obtain the desired upper bound on the s.d.o.f. as

$$D_s \leq (n_r - n_e)\left(\frac{T - n_r}{T}\right), \tag{41}$$

which completes the converse part of Lemma 2.

B. Achievability Proof of Lemma 2

Here, we show that a constant norm channel input [21], [22] transmitted on n_r antennas can achieve the s.d.o.f. upper bound given in (41). Specifically, let the channel input $\mathbf{X}_c$ be constant norm over n_r transmitter antennas and zero over the rest of the $n_t - n_r$ antennas, i.e.,

$$\mathbf{X}_c = \begin{bmatrix} \sqrt{\frac{PT}{n_r}}\mathbf{I}_{n_r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{\Theta}, \tag{42}$$

where $\mathbf{\Theta}$ is an $T \times T$ i.d. unitary matrix and $\mathbf{I}_{n_r}$ denotes the identity matrix with n_r dimension. We can lower bound the achievable secrecy rate R_s as follows:

$$\begin{aligned} T \cdot R_s &= I(\mathbf{X}_c; \mathbf{Y}) - I(\mathbf{X}_c; \mathbf{Z}) &(43)\\ &= h(\mathbf{Y}) - h(\mathbf{Z}) - h(\mathbf{Y}|\mathbf{X}_c) + h(\mathbf{Z}|\mathbf{X}_c) &(44)\\ &= h(\mathbf{Y}) - h(\mathbf{Z}) - n_r \sum_{i=1}^{n_r} \log\left(\frac{PT}{n_r} + \sigma_r^2\right) - n_r(T - n_t)\log \pi e \sigma_r^2 \\ &\quad + n_e \sum_{i=1}^{n_r} \log\left(\frac{PT}{n_r} + \sigma_e^2\right) + n_e(T - n_t)\log \pi e \sigma_e^2 &(45)\\ &\geq h(\mathbf{Y}) - h(\mathbf{Z}) - (n_r - n_e)n_r \log P + o(\log P), &(46) \end{aligned}$$

where (45) follows by applying (42) into (34) and (35), respectively.

Since $\mathbb{E}[\mathrm{tr}(\mathbf{Z}\mathbf{Z}^\dagger)] \leq (P + \sigma_e^2)n_e T$, the differential entropy $h(\mathbf{Z})$ of $\mathbf{Z}$ can be upper bounded by the differential entropy of an i.i.d. Gaussian matrix as

$$h(\mathbf{Z}) \leq n_e T \log\left(\pi e (P + \sigma_e^2)\right) = n_e T \log P + o(\log P), \tag{47}$$

and, the differential entropy $h(\mathbf{Y})$ can be lower bounded as

$$\begin{aligned} h(\mathbf{Y}) &\geq h(\mathbf{H}\mathbf{X}_c) &(48)\\ &= h(\mathbf{C}_{\mathbf{H}\mathbf{X}_c}) + \log|G(n_r, T)| + (T - n_r)\mathbb{E}\left[\log\det \mathbf{H}\mathbf{X}_c\mathbf{X}_c^\dagger\mathbf{H}^\dagger\right] &(49)\\ &= h\left(\sqrt{\frac{PT}{n_r}}\mathbf{H}_a\right) + \log|G(n_r, T)| + (T - n_r)\mathbb{E}\left[\log\det \frac{PT}{n_r}\mathbf{H}_a\mathbf{H}_a^\dagger\right] &(50)\\ &= n_r^2 \log \pi e \frac{PT}{n_r} + \log|G(n_r, T)| + (T - n_r)\log\left(\frac{PT}{n_r}\right)^{n_r} + (T - n_r)\mathbb{E}\left[\log\det \mathbf{H}_a\mathbf{H}_a^\dagger\right] &(51)\\ &= n_r T \log P + o(\log P), &(52) \end{aligned}$$

where $\mathbf{H}_a$ denotes the collection of the first n_r columns of matrix $\mathbf{H}$ and $|G(n_r, T)|$ is the volume of the Grassmann Manifold $G(n_r, T)$ (c.f. Appendix C), which is a finite constant. Note that (49) is obtained by applying Lemma 15 in Appendix C. A similar derivation, for the case where $n_t = n_r$, can be found in the proof of Lemma 15 in [22]. Therefore, from (46)-(47), and (52), we have the following lower bound on the secrecy rate

$$T \cdot R_s \geq (n_r - n_e)(T - n_r) \log P + o(\log P), \tag{53}$$

which implies that

$$D_s \geq (n_r - n_e) \left(\frac{T - n_r}{T} \right). \tag{54}$$

Together with the upper bound in (41), we conclude that, the exact s.d.o.f. for the case $n_t \geq n_r > n_e$ and $T \geq 2n_r$, is

$$D_s = (n_r - n_e) \left(\frac{T - n_r}{T} \right), \tag{55}$$

which completes the proof of Lemma 2.

As a final remark, we note that when $n_t \geq n_r$ and $T \geq 2n_r$, we can use only n_r transmitter antennas to achieve the optimal s.d.o.f in (55). Having more than n_r transmit antennas gives us no improvements, at least, as far as the s.d.o.f. is concerned.

VI. Proof of Lemma 3

The proof is based on the key observation that, when $n_t < n_r$, the receiver can use only n_t of its antennas without losing any s.d.o.f. That is, for a fixed n_t, the s.d.o.f. in the case where $n_t < n_r$ is, in fact, equal to the s.d.o.f. in the case with $n_r = n_t$. This fact is shown in the following lemma.

Lemma 10 *For the MIMO legitimate channel* (2)*, if $n_t < n_r$, for any input signal $\mathbf{X}$ satisfying the power constraint in* (1)*, we have*

$$I(\mathbf{X}; \mathbf{Y}) - I(\mathbf{X}; \mathbf{Y}_{n_t}) \leq o(\log P), \tag{56}$$

where $\mathbf{Y}_{n_t}$ denotes the collection of arbitrary n_t row vectors of the received signal matrix $\mathbf{Y}$.

We provide a proof for Lemma 10 in Appendix F.

To derive an upper bound for the s.d.o.f., we first focus on the case $\sigma_r^2 < \sigma_e^2$. When $n_r > n_e$ and $\sigma_r^2 < \sigma_e^2$, we can construct the same equivalent degraded channel in (24)-(25), in which case, the secrecy capacity can be written as in (26). The only difference here is that now the number of transmitter antennas is less than the number of legitimate receiver antennas, i.e., $n_t < n_r$. If we denote $C_s^{n_t<n_r}$ as the secrecy capacity of the wiretap channel in (2)-(3) with $\sigma_r^2 < \sigma_e^2$ and $n_t < n_r$, and $\mathbf{X}^*$ as the corresponding optimal input, we have

$$\begin{aligned} T \cdot C_s^{n_t<n_r} &= I(\mathbf{X}^*; \mathbf{Y}) - I(\mathbf{X}^*; \mathbf{Z}) & (57)\\ &\leq I(\mathbf{X}^*; \mathbf{Y}_{n_t}) - I(\mathbf{X}^*; \mathbf{Z}) + o(\log P) & (58)\\ &\leq \max_{p_{\mathbf{X}} \in S_{p_{\mathbf{X}}}} \left[I(\mathbf{X}; \mathbf{Y}_{n_t}) - I(\mathbf{X}; \mathbf{Z}) \right] + o(\log P) & (59)\\ &= T \cdot C_s^{n_t=n_r} + o(\log P), & (60) \end{aligned}$$

where (58) follows from Lemma 10, and $C_s^{n_t=n_r}$ in (60) is the secrecy capacity of the wiretap channel in (2)-(3) with smaller n_r as $n_t = n_r$. It is worthwhile to note that, in [22], a result similar to (56) but with more restrictions was given as

$$\max_{p_{\mathbf{X}} \in S_{p_{\mathbf{X}}}} I(\mathbf{X}; \mathbf{Y}) - \max_{p_{\mathbf{X}} \in S_{p_{\mathbf{X}}}} I(\mathbf{X}; \mathbf{Y}_{n_t}) \leq o(\log P). \tag{61}$$

The inequality in (61) is useful to prove the results in [22] for conventional MIMO channels, but is not sufficient for the proofs in MIMO wiretap channels. This is because, in (61), an upper bound was obtained when the input distribution is the one that maximizes $I(\mathbf{X}; \mathbf{Y})$. However, to obtain the inequality in (59), the input distribution must be one that maximizes the difference $I(\mathbf{X}; \mathbf{Y}_{n_t}) - I(\mathbf{X}; \mathbf{Z})$ instead. Thus, the more general result in Lemma 10 is required.

We already know the s.d.o.f. when $n_r = n_t$ and $T \geq 2n_r$. For $n_r = n_t > n_e$, the s.d.o.f. D_s is given by (55) from Lemma 2, and for $n_r = n_t \leq n_e$, $D_s = 0$ from Lemma 1. Thus, when $n_t < n_r$ and $T \geq 2n_t$, we get the required upper bound as

$$D_s \leq (n_t - n_e)^+ \left(\frac{T - n_t}{T} \right). \tag{62}$$

Furthermore, in the case where $\sigma_r^2 \geq \sigma_e^2$, one can upper bound the secrecy capacity by increasing σ_e^2 in the eavesdropper's channel. Thus, the upper bound (62) is still valid.

The achievability of the above upper bound follows by using a constant norm channel input over n_t transmitter antennas as described in Section V-B. However, at the legitimate receiver, only n_t receiver antennas are needed and we can ignore the remaining $(n_r - n_t)$ row vectors of the received signal matrix $\mathbf{Y}$ while decoding at high SNR. These matching converse and achievability results complete the proof of Lemma 3.

VII. Secure Degrees of Freedom for Short Coherence Time Systems

In this section, we provide some insights for the s.d.o.f. of short coherence time systems (i.e., systems with $T < 2\min\{n_t, n_r\}$). Recall that, when $n_r \leq n_e$, we know from Lemma 1 that the s.d.o.f. is zero regardless of the coherence time T. However, when $n_r > n_e$, our results in Theorem 1 hold only for the case with sufficiently large coherence time, i.e., for $T \geq 2\min(n_t, n_r)$. To study the s.d.o.f of short coherence time systems, we first consider a special case where $T = 1$ (i.e., fast fading channels). We show, in the following lemma, that the s.d.o.f. of the MIMO fast fading wiretap channel is zero regardless of how many antennas the terminals have. This is a generalization of [20] to the case of multiple antennas.

Lemma 11 *For the MIMO wiretap channel in* (2)-(3)*, with no CSI at any terminal and* $T = 1$*, we have* $D_s = 0$*.*

Proof: Here, we focus only on the case where $n_r > n_e$ since the s.d.o.f. is zero when $n_r \leq n_e$ (c.f., Lemma 1). Specifically, let us first consider the case where $n_r > n_e$ and $\sigma_r^2 \leq \sigma_e^2$. In this case, the MIMO wiretap channel can be converted to an equivalent degraded wiretap channel as in (24)-(25), similar to what was done in Section V, and its secrecy capacity can be written as in (26). Following the result in [21], we note in the following lemma that, when $n_t \geq T$, which is always the case when $T = 1$, the secrecy capacity of the equivalent degraded wiretap channel can be achieved by using only $n_t = T$ transmit antennas.

Lemma 12 *Suppose that* $n_t > T$ *and that the* $n_t \times T$ *input signal* $\mathbf{X}$ *with distribution* $p_{\mathbf{X}}$ *generates mutual informations* $I(\mathbf{X}; \mathbf{Y})$ *and* $I(\mathbf{X}; \mathbf{Z}_{\mathrm{p}_1})$ *on the main and the eavesdropper channels, respectively, described in* (24)-(25)*. Then, there exists an* $T \times T$ *input signal* $\mathbf{X}'$*, i.e., an input signal that utilizes only* T *transmit antennas, that generates the same mutual informations, i.e.,* $I(\mathbf{X}; \mathbf{Y}) = I(\mathbf{X}'; \mathbf{Y})$ *and* $I(\mathbf{X}; \mathbf{Y}) = I(\mathbf{X}'; \mathbf{Z}_{\mathrm{p}_1})$*.*

This lemma is a straightforward extension of [21, Theorem 1] and, thus, its proof is omitted here. The main idea is that both conditional probability density functions $p(\mathbf{Y}|\mathbf{X})$ and $p(\mathbf{Z}_{\mathrm{p}_1}|\mathbf{X})$

depend on $\mathbf{X}$ only through $\mathbf{X}^\dagger\mathbf{X}$. Hence, for any $n_t \times T$ input matrix $\mathbf{X}$, we can obtain the same mutual informations by using a $T \times T$ input matrix $\mathbf{X}'$ such that $\mathbf{X}'^\dagger\mathbf{X}' = \mathbf{X}^\dagger\mathbf{X}$.

It follows from Lemma 12 that, when $T = 1$, the secrecy capacity is the same as the secrecy capacity with a single transmit antenna only, i.e., $n_t = 1$. Moreover, by Lemma 10, we know that having more receive antennas than transmit antennas, i.e., having $n_r > n_t$, does not improve the s.d.o.f. Thus, when $n_r > n_e$, $\sigma_r^2 < \sigma_e^2$ and $T = 1$, the s.d.o.f. of the MIMO wiretap channel is the same as the secrecy capacity of the SISO case, i.e., $n_t = n_r = 1$, which is zero, as shown in [20]. Finally, since the secrecy capacity for the case with $\sigma_r^2 \le \sigma_e^2$ is greater than that with $\sigma_r^2 > \sigma_e^2$, we conclude that the s.d.o.f. of the fast fading wiretap channel, i.e., the case with $T = 1$, is zero regardless of the number of antennas at the terminals. ■

For the general case, the exact s.d.o.f. is unknown. However, as stated in Lemma 4, we can show that, by using constant norm input, the achievable s.d.o.f. can be given by

$$(K - n_e)^+ \left(\frac{T - K}{T}\right), \tag{63}$$

where $K \triangleq \min(n_t, n_r, (T + n_e)/2)$. To show this, we first note from Lemma 12 that, when $T < n_t$, the secrecy capacity can be achieved by using only T out of the n_t transmit antennas. That is, no further improvement in secrecy capacity can be obtained by using all n_t antennas. Therefore, in the following, we focus only on the case where $n_t \le T$.

Suppose that the constant norm input is applied over $m \le \min(n_t, n_r)$ transmitter antennas (c.f. in Section V-B) and let the channel input signal be written as

$$\mathbf{X} = \begin{bmatrix} \sqrt{\frac{PT}{m}}\mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{\Theta}. \tag{64}$$

Moreover, let us also assume that only m antennas are used at the legitimate receiver to receive the signal. Then, by following the same arguments as in Section V-B, we can show that the s.d.o.f. achieved by using (64) is given by

$$(m - n_e)^+ \left(\frac{T - m}{T}\right). \tag{65}$$

Note that (65) is a quadratic function that increases with m when $n_e < m < (T + n_e)/2$ and reaches its maximum value at the point $m = (T + n_e)/2$. Thus, together with the condition $m \le \min(n_t, n_r)$, the number of transmit and receive antennas that should be used to transmit the constant norm input signal is $m = K$ and the resulting achievable s.d.o.f. is given by (63). However, whether or not (63) is the maximum achievable s.d.o.f. is still an open problem.

VIII. Conclusion

We considered the Rayleigh block fading wiretap channel with no a priori CSI at any of the terminals. We constructed a degraded equivalent channel, and determined its secrecy capacity. We determined the exact s.d.o.f. of this channel model, when $T \geq 2\min(n_t, n_r)$, to be $(\min(n_t, n_r) - n_e)^+(T - \min(n_t, n_r))/T$. When $\min(n_t, n_r) \leq n_e$, the s.d.o.f. is zero no matter how long the coherence time T is; an example of this is the scalar wiretap channel where $n_t = n_r = n_e = 1$. When $T = 1$, the s.d.o.f. is zero no matter how many antennas the transmitter and the legitimate receiver may have. We showed in this paper that when we have some moderate channel coherence together with multiple antennas at the legitimate entities, we can have non-zero s.d.o.f. The needed condition for this is that the legitimate entities have more antennas than the eavesdropper.

Appendix A
Proof of Lemma 5

We first introduce the following two lemmas which will be useful for the proof of Lemma 5. These lemmas are straightforward extensions of Lemmas 1 and 3 of [21].

Lemma 13 *Suppose that the input signal $\mathbf{X}$ with distribution $p_{\mathbf{X}}$ generates mutual informations $I(\mathbf{X};\mathbf{Y})$ and $I(\mathbf{X};\mathbf{Z}_{\mathrm{p}1})$ on the main and the eavesdropper channels described in* (24)-(25). *For any $m \times m$ deterministic unitary matrix $\mathbf{V}$ and $T \times T$ deterministic unitary matrix $\mathbf{U}$, the input signal $\mathbf{V}^\dagger\mathbf{X}\mathbf{U}$ generates the same mutual informations, i.e., $I(\mathbf{X};\mathbf{Y}) = I(\mathbf{V}^\dagger\mathbf{X}\mathbf{U};\mathbf{Y})$ and $I(\mathbf{X};\mathbf{Z}_{\mathrm{p}1}) = I(\mathbf{V}^\dagger\mathbf{X}\mathbf{U};\mathbf{Z}_{\mathrm{p}1})$.*

Lemma 14 *Suppose that the input signal $\mathbf{X}$ with singular value decomposition $\mathbf{X} = \mathbf{\Psi}^\dagger\mathbf{\Lambda}\mathbf{\Phi}$ generates the mutual informations $I(\mathbf{X};\mathbf{Y})$ and $I(\mathbf{X};\mathbf{Z}_{\mathrm{p}1})$ on the main and the eavesdropper channels described in* (24)-(25). *Then, the input signal $\mathbf{X}' = \mathbf{\Lambda}\mathbf{\Phi}$ also generates the same mutual informations, i.e., $I(\mathbf{X};\mathbf{Y}) = I(\mathbf{X}';\mathbf{Y})$ and $I(\mathbf{X};\mathbf{Z}_{\mathrm{p}1}) = I(\mathbf{X}';\mathbf{Z}_{\mathrm{p}1})$.*

Note that, the above lemmas hold separately for $I(\mathbf{X};\mathbf{Y})$ and $I(\mathbf{X};\mathbf{Z}_{\mathrm{p}1})$ irrespective of the degradedness relation, and relies only on the fact that their respective channels, i.e., $p_{\mathbf{Y}|\mathbf{X}}$ and

$p_{\mathbf{Z}_{\mathrm{p}_1}|\mathbf{X}}$, are Gaussian. Using the above two lemmas, we will show that, for any input $\mathbf{X}$ with distribution $p_{\mathbf{X}}$, there exists input signal $\mathbf{X}^*$, satisfying the structure in (27) in Lemma 5, that achieves a higher secrecy rate for the degraded MIMO wiretap channel given in (24)-(25), i.e.,

$$\frac{1}{T}\left(I(\mathbf{X};\mathbf{Y}) - I(\mathbf{X};\mathbf{Z}_{\mathrm{p}_1})\right) \leq \frac{1}{T}\left(I(\mathbf{X}^*;\mathbf{Y}) - I(\mathbf{X}^*;\mathbf{Z}_{\mathrm{p}_1})\right). \tag{66}$$

To do this, we define the secrecy rate function for (24)-(25) as $R_s(p_{\mathbf{X}}) = \frac{1}{T}(I(\mathbf{X};\mathbf{Y}) - I(\mathbf{X};\mathbf{Z}_{\mathrm{p}_1}))$ where $p_{\mathbf{X}}$ denotes the probability distribution of $\mathbf{X}$. Note that R_s is a concave function of the input probability distribution [24] since the eavesdropper's channel (25) is degraded with respect to the main channel (24). Then, for any input signal $\mathbf{X} = \mathbf{\Psi}^\dagger\mathbf{\Lambda}\mathbf{\Phi}$, we can let $\mathbf{X}' = \mathbf{\Lambda}\mathbf{\Phi}$ and $\mathbf{X}^* = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Theta}'$ with $\mathbf{\Theta}'$ being a $T \times T$ i.d. unitary matrix which is independent of $\mathbf{X}$ (independent of $(\mathbf{\Psi}, \mathbf{\Lambda}, \mathbf{\Phi})$), and upper bound the secrecy rate with the input $\mathbf{X}$ as follows

$$R_s(p_{\mathbf{X}}) = R_s(p_{\mathbf{X}'}) \tag{67}$$

$$= R_s(p_{\mathbf{X}^*|\mathbf{\Theta}'=\tilde{\mathbf{\Theta}}'}) \tag{68}$$

$$= \int R_s(p_{\mathbf{X}^*|\tilde{\mathbf{\Theta}}'})\, dF(\tilde{\mathbf{\Theta}}') \tag{69}$$

$$\leq R_s\left(\int p_{\mathbf{X}^*|\tilde{\mathbf{\Theta}}'}\, dF(\tilde{\mathbf{\Theta}}')\right) \tag{70}$$

$$= R_s(p_{\mathbf{X}^*}), \tag{71}$$

where (67) follows from Lemma 14 and the equality in (68) follows from Lemma 13 with $\mathbf{V}$ being the identity matrix and $\mathbf{U}$ being the given realization $\tilde{\mathbf{\Theta}}'$, and (70) follows from Jensen's inequality since R_s is concave.

Now let $\mathbf{\Theta} = \mathbf{\Phi}\mathbf{\Theta}'$ such that $\mathbf{X}^* = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Theta}' = \mathbf{\Lambda}\mathbf{\Theta}$. The rest of the proof is to show that $\mathbf{\Theta}$ is also an i.d. unitary matrix and independent of $\mathbf{\Lambda}$. First, we have

$$p_{\mathbf{\Theta}}(\tilde{\mathbf{\Theta}}) = \int p_{\mathbf{\Theta}|\mathbf{\Phi}}(\tilde{\mathbf{\Theta}}|\tilde{\mathbf{\Phi}})\, dF(\tilde{\mathbf{\Phi}}) \tag{72}$$

$$= \int p_{\mathbf{\Theta}'|\mathbf{\Phi}}(\tilde{\mathbf{\Phi}}^{-1}\tilde{\mathbf{\Theta}}|\tilde{\mathbf{\Phi}})\, dF(\tilde{\mathbf{\Phi}}) \tag{73}$$

$$= \int p_{\mathbf{\Theta}'}(\tilde{\mathbf{\Phi}}^{-1}\tilde{\mathbf{\Theta}})\, dF(\tilde{\mathbf{\Phi}}) \tag{74}$$

$$= \int p_{\mathbf{\Theta}'}(\tilde{\mathbf{\Theta}})\, dF(\tilde{\mathbf{\Phi}}) \tag{75}$$

$$= p_{\mathbf{\Theta}'}(\tilde{\mathbf{\Theta}}), \tag{76}$$

where (74) comes from the fact that $\mathbf{\Theta}'$ is independent of $(\mathbf{\Psi}, \mathbf{\Lambda}, \mathbf{\Phi})$ and (75) comes from the fact that $\mathbf{\Theta}'$ is i.d. unitary [21]. Therefore, we show that $\mathbf{\Theta}$ has the same distribution as $\mathbf{\Theta}'$, and then $\mathbf{\Theta}$ is also i.d. unitary. For the independence between $\mathbf{\Theta}$ and $\mathbf{\Lambda}$, we have

$$\begin{aligned}
p_{\mathbf{\Theta}|\mathbf{\Lambda}}(\tilde{\mathbf{\Theta}}|\tilde{\mathbf{\Lambda}}) &= \int p_{\mathbf{\Theta}|\mathbf{\Lambda},\mathbf{\Phi}}(\tilde{\mathbf{\Theta}}|\tilde{\mathbf{\Lambda}}, \tilde{\mathbf{\Phi}}) p_{\mathbf{\Phi}|\mathbf{\Lambda}}(\tilde{\mathbf{\Phi}}|\tilde{\mathbf{\Lambda}}) d\tilde{\mathbf{\Phi}} && (77)\\
&= \int p_{\mathbf{\Theta}'|\mathbf{\Lambda},\mathbf{\Phi}}(\tilde{\mathbf{\Phi}}^{-1}\tilde{\mathbf{\Theta}}|\tilde{\mathbf{\Lambda}}, \tilde{\mathbf{\Phi}}) p_{\mathbf{\Phi}|\mathbf{\Lambda}}(\tilde{\mathbf{\Phi}}|\tilde{\mathbf{\Lambda}}) d\tilde{\mathbf{\Phi}} && (78)\\
&= \int p_{\mathbf{\Theta}'}(\tilde{\mathbf{\Phi}}^{-1}\tilde{\mathbf{\Theta}}) p_{\mathbf{\Phi}|\mathbf{\Lambda}}(\tilde{\mathbf{\Phi}}|\tilde{\mathbf{\Lambda}}) d\tilde{\mathbf{\Phi}} && (79)\\
&= \int p_{\mathbf{\Theta}'}(\tilde{\mathbf{\Theta}}) p_{\mathbf{\Phi}|\mathbf{\Lambda}}(\tilde{\mathbf{\Phi}}|\tilde{\mathbf{\Lambda}}) d\tilde{\mathbf{\Phi}} && (80)\\
&= p_{\mathbf{\Theta}'}(\tilde{\mathbf{\Theta}}) && (81)\\
&= p_{\mathbf{\Theta}}(\tilde{\mathbf{\Theta}}), && (82)
\end{aligned}$$

where (79) and (80) follow, respectively, from the derivations for (74) and (75), and (82) comes from the fact that $\mathbf{\Theta}$ and $\mathbf{\Theta}'$ have the same distribution. By (76) and (82), we conclude that $\mathbf{\Theta}$ is an i.d. unitary matrix and is independent of $\mathbf{\Lambda}$, completing the proof of Lemma 5.

Appendix B
Proof of Lemma 6

To show (31), we first define a function $\Pi : S \to S$ with $S = \{1, \ldots, m\}$ as follows to order the row vectors of $\mathbf{M}$

$$\begin{aligned}
\Pi(1) &= \arg \max_{i \in S} h(\mathbf{M}_i), && (83)\\
\Pi(2) &= \arg \max_{i \in S \setminus \{\Pi(1)\}} h(\mathbf{M}_i | \mathbf{M}_{\Pi(1)}), && (84)\\
\Pi(3) &= \arg \max_{i \in S \setminus \{\Pi(1), \Pi(2)\}} h(\mathbf{M}_i | \mathbf{M}_{\Pi(1)}, \mathbf{M}_{\Pi(2)}), && (85)\\
&\vdots && (86)\\
\Pi(m-1) &= \arg \max_{i \in S \setminus \{\bigcup_{j=1}^{m-2} \Pi(j)\}} h\left(\mathbf{M}_i \middle| \bigcup_{j=1}^{m-2} \mathbf{M}_{\Pi(j)}\right), && (87)\\
\Pi(m) &= \arg \max_{i \in S \setminus \{\bigcup_{j=1}^{m-1} \Pi(j)\}} h\left(\mathbf{M}_i \middle| \bigcup_{j=1}^{m-1} \mathbf{M}_{\Pi(j)}\right), && (88)
\end{aligned}$$

where $\mathbf{M}_i$ denotes the ith row vector of $\mathbf{M}$. Note that if we order the row vectors by this function Π, we have

$$h\left(\mathbf{M}_{\Pi(k)}\Big|\bigcup_{j=1}^{k-1}\mathbf{M}_{\Pi(j)}\right) \geq h\left(\mathbf{M}_{\Pi(k+1)}\Big|\bigcup_{j=1}^{k-1}\mathbf{M}_{\Pi(j)}\right) \tag{89}$$

$$\geq h\left(\mathbf{M}_{\Pi(k+1)}\Big|\bigcup_{j=1}^{k}\mathbf{M}_{\Pi(j)}\right), \quad \forall k \in \{1,\ldots,m-1\}, \tag{90}$$

where (89) comes from the definition of Π, and (90) is due to the fact that conditioning on one more $\mathbf{M}_{\Pi(k)}$ will reduce the differential entropy. The inequality in (90) implies that the conditional differential entropy $h\left(\mathbf{M}_{\Pi(k)}\Big|\bigcup_{j=1}^{k-1}\mathbf{M}_{\Pi(j)}\right)$ is non-increasing with respect to the index k.

Now, for any given number $n < m$, we can select $\mathrm{p}_1 = \{\bigcup_{j=1}^{n}\Pi(j)\}$ and $\mathrm{p}_2 = \{\bigcup_{j=n+1}^{m}\Pi(j)\}$ which form a partition of $S = \{1,\ldots,m\}$. We have

$$h(\mathbf{M}_{\mathrm{p}_2}|\mathbf{M}_{\mathrm{p}_1}) = \sum_{k=n+1}^{m} h\left(\mathbf{M}_{\Pi(k)}\Big|\bigcup_{j=1}^{k-1}\mathbf{M}_{\Pi(j)}\right) \tag{91}$$

$$\leq (m-n)h\left(\mathbf{M}_{\Pi(n+1)}\Big|\bigcup_{j=1}^{n}\mathbf{M}_{\Pi(j)}\right) \tag{92}$$

$$\leq (m-n)\frac{1}{n}\sum_{k=1}^{n} h\left(\mathbf{M}_{\Pi(k)}\Big|\bigcup_{j=1}^{k-1}\mathbf{M}_{\Pi(j)}\right) \tag{93}$$

$$= \frac{(m-n)}{n}h(\mathbf{M}_{\mathrm{p}_1}), \tag{94}$$

where (91) comes from the chain rule of differential entropy; and both (92) and (93) follow from (90). More specifically, (92) follows from the fact that the largest term inside the summation of (91) is the conditional differential entropy with index $k = n+1$ and (93) follows from the fact that the conditional differential entropy with index $k = n+1$ is smaller than each term inside the summation of (93). Finally, by adding $(m-n)h(\mathbf{M}_{\mathrm{p}_2}|\mathbf{M}_{\mathrm{p}_1})/n$ to both sides of (94), we obtain

$$\frac{m}{n}h(\mathbf{M}_{\mathrm{p}_2}|\mathbf{M}_{\mathrm{p}_1}) \leq \frac{m-n}{n}(h(\mathbf{M}_{\mathrm{p}_1}) + h(\mathbf{M}_{\mathrm{p}_2}|\mathbf{M}_{\mathrm{p}_1})), \tag{95}$$

which results in (31), completing the proof of Lemma 6.

APPENDIX C
PROOF OF LEMMA 7

Before showing the proof of Lemma 7, we first introduce some background from [22]. A $n \times T$ matrix $\mathbf{M}$ where $T \geq n$, can be represented by a change of coordinate system as

$$\mathbf{M} \rightarrow (\mathbf{\Omega_M}, \mathbf{C_M}), \tag{96}$$

where the subspace $\mathbf{\Omega_M}$ is generated by its own row vectors $\mathbf{M}_i s$, $\forall i \in \{1, \ldots, n\}$, and the $n \times n$ matrix $\mathbf{C_M}$ represents each row vector $\mathbf{M}_i$ with respect to an orthonormal basis of $\mathbf{\Omega_M}$. The mapping in (96) changes the coordinate system of matrix $\mathbf{M}$ from $\mathcal{C}^{n\times T}$ to $G(T, n) \times \mathcal{C}^{n\times n}$ where $G(T, n)$ is a Grassmann manifold with $n(T - n)$ d.o.f. Now we can state a result given in [22] as follows.

Lemma 15 *For a random matrix $\mathbf{M} \in \mathcal{C}^{n\times T}$, $T \geq n$, which is i.d., the differential entropy of $\mathbf{M}$ can be written as*

$$h(\mathbf{M}) = h(\mathbf{C_M}) + \log |G(T, n)| + (T - n)\mathbb{E}\left[\log \det \mathbf{M}\mathbf{M}^\dagger\right], \tag{97}$$

where the $n \times n$ matrix $\mathbf{C_M}$ and the Grassmann manifold $G(T, n)$ are defined following (96)*.*

To prove Lemma 7, note that due to Lemma 5, the received signal $\mathbf{Y}$ is i.d. Then, from Lemma 15, we have

$$\begin{aligned} h(\mathbf{Y}) &= h(\mathbf{C_Y}) + \log |G(T, n_r)| + (T - n_r)\mathbb{E}\left[\log \det \mathbf{Y}\mathbf{Y}^\dagger\right] \\ &\leq n_r^2 \log \pi e \left(\frac{(P + \sigma_r^2)T}{n_r}\right) + |G(T, n_r)| + (T - n_r)\mathbb{E}\left[\log \det \mathbf{Y}\mathbf{Y}^\dagger\right], \end{aligned} \tag{98}$$

where the inequality comes from upper bounding $h(\mathbf{C_Y})$ by assuming that each element of $\mathbf{C_Y}$ is i.i.d. Gaussian with variance $(P+\sigma_r^2)T/n_r$. Now, note that the volume of Grassmann manifold $G(T, n_r)$ is a finite constant which is independent of P as

$$|G(T, n)| = \frac{\prod_{i=T-n+1}^{T} \frac{2\pi^2}{(i-1)!}}{\prod_{i=1}^{n} \frac{2\pi^2}{(i-1)!}}. \tag{99}$$

Combining (98) and (99) completes the proof of Lemma 7.

APPENDIX D
PROOF OF LEMMA 8

First, we introduce a few useful lemmas from [25].

Lemma 16 *If* $\mathrm{M}_a \in \mathcal{C}^{m\times m}$ *and* $\mathrm{M}_b \in \mathcal{C}^{m\times m}$ *are both Hermitian matrices, then*

$$\sum_{i=1}^{m}(\lambda_i(\mathbf{M}_b) - \lambda_i(\mathbf{M}_a))^2 \leq ||\mathbf{M}_b - \mathbf{M}_a||_2^2, \tag{100}$$

where $\lambda_i(A)$ *denotes the* i*th largest eigenvalue of matrix* A *and* $||\cdot||_2$ *is the Frobenius norm.*

Lemma 17 *Let* M_a *and* M_b *be both* $m\times m$ *Hermitian matrices. Assume that* $\mathrm{M}_b - \mathrm{M}_a$ *is positive semi-definite. We have*

$$\lambda_i(\mathrm{M}_b) \geq \lambda_i(\mathrm{M}_a), \quad \forall i \in \{1,\dots,m\}. \tag{101}$$

Lemma 18 *For any matrix* $\mathrm{M}_a \in \mathcal{C}^{m\times n}$ *and* $\mathrm{M}_b \in \mathcal{C}^{n\times m}$ *where* $m \geq n$*,*

$$\lambda_i(\mathrm{M}_a\mathrm{M}_b) = \lambda_i(\mathrm{M}_b\mathrm{M}_a), \quad \forall i \in \{1,\dots,n\}. \tag{102}$$

From Lemmas 16 and 17, we can infer the following result.

Corollary 1 *Let* M_a *and* M_b *be both* $m\times m$ *Hermitian matrices and assume* $\mathrm{M}_b - \mathrm{M}_a$ *is positive semi-definite. Then,*

$$\lambda_i(\mathbf{M}_b) - \lambda_i(\mathbf{M}_a) \leq ||\mathbf{M}_b - \mathbf{M}_a||_2, \quad \forall i \in \{1,\dots,m\}. \tag{103}$$

To derive the upper bound on $\mathbb{E}\left[\log\det \mathbf{Y}\mathbf{Y}^\dagger\right]$ given in Lemma 8, we first recall that, given $\mathbf{X} = \tilde{\mathbf{X}}$, where $\tilde{\mathbf{X}}$ denotes a realization of $\mathbf{X}$, the row vectors of $\mathbf{Y}$ are i.i.d. Gaussian vectors with covariance matrix $\tilde{\mathbf{X}}^\dagger\tilde{\mathbf{X}} + \sigma_r^2\mathbf{I}_T$. Let $\mathbf{Q}_{n_r,T} \in \mathcal{C}^{n_r\times T}$ be a random matrix whose elements are i.i.d. complex Gaussian variables with zero-mean and unit-variance. Then, we have

$$(\mathbf{Y}|\mathbf{X} = \tilde{\mathbf{X}}) \overset{d}{\sim} \mathbf{Q}_{n_r,T}(\tilde{\mathbf{X}}^\dagger\tilde{\mathbf{X}} + \sigma_r^2\mathbf{I}_T)^{1/2}, \tag{104}$$

where $\mathbf{A} \overset{d}{\sim} \mathrm{B}$ denotes $\mathbf{A}$ has the same distribution as B. Therefore, we have

$$(\mathbf{Y}\mathbf{Y}^\dagger|\mathbf{X} = \tilde{\mathbf{X}}) \overset{d}{\sim} \mathbf{Q}_{n_r,T}(\tilde{\mathbf{X}}^\dagger\tilde{\mathbf{X}} + \sigma_r^2\mathbf{I}_T)\mathbf{Q}_{n_r,T}^\dagger. \tag{105}$$

From Lemma 5, we know that the optimal input can be written as $\mathbf{X} = \mathbf{\Lambda}\mathbf{\Theta}$ where $\mathbf{\Lambda} \in \mathcal{C}^{n_t \times T}$ is a diagonal random matrix and $\mathbf{\Theta} \in \mathcal{C}^{T \times T}$ is an i.d. unitary matrix. Hence, by taking $\tilde{\mathbf{X}} = \tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Theta}}$, we can rewrite (105) as follows

$$(\mathbf{Y}\mathbf{Y}^\dagger | \mathbf{X} = \tilde{\mathbf{X}}) \overset{d}{\sim} \mathbf{Q}_{n_r,T}\left(\tilde{\mathbf{\Theta}}^\dagger\tilde{\mathbf{\Lambda}}^\dagger\tilde{\mathbf{\Lambda}}\tilde{\mathbf{\Theta}} + \sigma_r^2\mathbf{I}_T\right)\mathbf{Q}_{n_r,T}^\dagger \tag{106}$$

$$\overset{d}{\sim} \mathbf{Q}_{n_r,T}\tilde{\mathbf{\Theta}}^\dagger\left(\text{diag}(||\tilde{\mathbf{X}}_1||^2, ..., ||\tilde{\mathbf{X}}_{n_t}||^2, 0, ..., 0) + \sigma_r^2\mathbf{I}_T\right)\tilde{\mathbf{\Theta}}\mathbf{Q}_{n_r,T}^\dagger \tag{107}$$

$$\overset{d}{\sim} \mathbf{Q}_a(\tilde{\mathbf{\Lambda}}_x^2 + \sigma_r^2\mathbf{I}_{n_t})\mathbf{Q}_a^\dagger + \sigma_r^2\mathbf{Q}_b\mathbf{Q}_b^\dagger, \tag{108}$$

where $\tilde{\mathbf{\Lambda}}_x \triangleq \text{diag}(||\tilde{\mathbf{X}}_1||, ||\tilde{\mathbf{X}}_2||, ..., ||\tilde{\mathbf{X}}_{n_t}||)$. In the above, (108) comes from the fact that the i.i.d. Gaussian random matrix $\mathbf{Q}_{n_r,T}$ is i.d. with $[\mathbf{Q}_a|\mathbf{Q}_b] = \mathbf{Q}_{n_r,T}$, where $\mathbf{Q}_a \in \mathcal{C}^{n_r \times n_t}$ contains the first n_t columns of $\mathbf{Q}_{n_r,T}$ and $\mathbf{Q}_b \in \mathcal{C}^{n_r \times (T-n_t)}$ contains the remaining $(T - n_t)$ columns of $\mathbf{Q}_{n_r,T}$. To simplify the notation, let

$$\mathbf{B} = \mathbf{Q}_a(\tilde{\mathbf{\Lambda}}_x^2 + \sigma_r^2\mathbf{I}_{n_t})\mathbf{Q}_a^\dagger + \sigma_r^2\mathbf{Q}_b\mathbf{Q}_b^\dagger, \tag{109}$$

$$\mathbf{A} = \mathbf{Q}_a(\tilde{\mathbf{\Lambda}}_x^2 + \sigma_r^2\mathbf{I}_{n_t})\mathbf{Q}_a^\dagger, \tag{110}$$

and we have

$$\mathbb{E}\left[\log\det\mathbf{Y}\mathbf{Y}^\dagger\right] = \mathbb{E}_{\mathbf{X}}\mathbb{E}\left[\log\det\mathbf{Y}\mathbf{Y}^\dagger | \mathbf{X} = \tilde{\mathbf{X}}\right] \tag{111}$$

$$= \mathbb{E}_{\mathbf{X}}\mathbb{E}\left[\log\prod_{i=1}^{n_r}\lambda_i(\mathbf{B})\right] \tag{112}$$

$$\leq \mathbb{E}_{\mathbf{X}}\sum_{i=1}^{n_r}\mathbb{E}_{\mathbf{Q}_a,\mathbf{Q}_b}\left[\log(\lambda_i(\mathbf{A}) + \sigma_r^2||\mathbf{Q}_b\mathbf{Q}_b^\dagger||_2)\right], \tag{113}$$

where (113) comes from applying Corollary 1 to (112). Then, by using Jensen's inequality on $\mathbf{Q}_b$, and the definition of $\mathbf{A}$ in (110), we get the following for the right hand side of (113),

$$\mathbb{E}_{\mathbf{X}}\sum_{i=1}^{n_r}\mathbb{E}_{\mathbf{Q}_a,\mathbf{Q}_b}\left[\log(\lambda_i(\mathbf{A}) + \sigma_r^2||\mathbf{Q}_b\mathbf{Q}_b^\dagger||_2)\right]$$

$$\leq \mathbb{E}_{\mathbf{X}}\sum_{i=1}^{n_r}\mathbb{E}_{\mathbf{Q}_a}\left[\log(\lambda_i(\mathbf{Q}_a(\mathbf{\Lambda}_x^2 + \sigma_r^2\mathbf{I}_{n_t})\mathbf{Q}_a^\dagger) + \sigma_r^2k_1)\right] \tag{114}$$

$$= \mathbb{E}_{\mathbf{X}}\mathbb{E}_{\mathbf{Q}_a}\left[\log\det\left(\mathbf{Q}_a(\mathbf{\Lambda}_x^2 + \sigma_r^2\mathbf{I}_{n_t})\mathbf{Q}_a^\dagger + \sigma_r^2k_1\mathbf{I}_{n_r}\right)\right] \tag{115}$$

$$= \mathbb{E}_{\mathbf{Q}_a}\mathbb{E}_{\mathbf{X}}\left[\log\det\left(\mathbf{Q}_a(\mathbf{\Lambda}_x^2 + \sigma_r^2\mathbf{I}_{n_t})\mathbf{Q}_a^\dagger + \sigma_r^2k_1\mathbf{I}_{n_r}\right)\right] \tag{116}$$

$$\leq \mathbb{E}_{\mathbf{Q}_a}\left[\log\det\left(\mathbf{Q}_a(\mathbb{E}_{\mathbf{X}}[\mathbf{\Lambda}_x^2] + \sigma_r^2\mathbf{I}_{n_t})\mathbf{Q}_a^\dagger + \sigma_r^2k_1\mathbf{I}_{n_r}\right)\right], \tag{117}$$

where $\mathbf{\Lambda}_x = \mathrm{diag}(||\mathbf{X}_1||, ||\mathbf{X}_2||, ..., ||\mathbf{X}_{n_t}||)$, and $k_1 = \mathbb{E}[||\mathbf{Q}_b\mathbf{Q}_b^\dagger||_2]$ in (114) is a finite constant independent of P, the exchange of expectation over $\mathbf{X}$ and $\mathbf{Q}_a$ in (116) follows from the fact that $\mathbf{X}$ and $\mathbf{Q}_a$ are independent, and (117) comes from applying Jensen's inequality on $\mathbf{X}$.

Note that the right-hand-side (RHS) of (117) is a concave function of $\mathbb{E}_\mathbf{X}\left[\mathbf{\Lambda}_x^2\right]$ and, since the distribution of $\mathbf{Q}_a$ is invariant to the permutation of its rows, it can be shown that the RHS of (117) is also a symmetric function with respect to the diagonal entries of $\mathbb{E}_\mathbf{X}\left[\mathbf{\Lambda}_x^2\right]$, where a symmetric function is defined as a function that is invariant to permutations of its input variables. These properties imply, from [26], that (117) is a Schur-concave function with respect to the diagonal entries of $\mathbb{E}_\mathbf{X}\left[\mathbf{\Lambda}_x^2\right]$. Recall the definition of Schur-concave functions as follows.

Definition 1 *A function $f : \mathcal{R}^n \to \mathcal{R}$ is said to be a Schur-concave function if, for any $\mathbf{x}$ and $\mathbf{y}$ such that $\mathbf{x} \prec \mathbf{y}$ (i.e., $\mathbf{x}$ is majorized by $\mathbf{y}$) [26], we have $f(\mathbf{x}) \geq f(\mathbf{y})$.*

Note that $\det(\cdot)$ is matrix nondecreasing on the set of positive semi-definite matrices [27, Section 3.6.1] (i.e., $\det(\mathbf{M}_a) \geq \det(\mathbf{M}_b)$, for all $\mathbf{M}_a, \mathbf{M}_b$, and $\mathbf{M}_a - \mathbf{M}_b$ that are positive semi-definite). This implies that, under the power constraint $\mathrm{tr}(\mathbb{E}_\mathbf{X}[\mathbf{\Lambda}_x^2]) \leq PT$, the RHS of (117) can be upper-bounded by taking $\mathbb{E}_\mathbf{X}[\mathbf{\Lambda}_x^2]$ that satisfies $\mathrm{tr}(\mathbb{E}_\mathbf{X}[\mathbf{\Lambda}_x^2]) = PT$. Moreover, since the diagonal entries of $(PT/n_t)\mathbf{I}_{n_t}$ form a vector that is majorized by all vectors summing up to PT, it follows by the property of Schur-concave functions that the RHS of (117) can be further upper bounded by choosing $\mathbb{E}_\mathbf{X}[\mathbf{\Lambda}_x^2] = (PT/n_t)\mathbf{I}_{n_t}$, Hence, we have

$$\begin{aligned}
\mathbb{E}\left[\log\det\mathbf{Y}\mathbf{Y}^\dagger\right] \leq & \mathbb{E}_{\mathbf{Q}_a}\left[\log\det\left(\mathbf{Q}_a\left(\frac{PT}{n_t}\mathbf{I}_{n_t} + \sigma_r^2\mathbf{I}_{n_t}\right)\mathbf{Q}_a^\dagger + \sigma_r^2 k_1\mathbf{I}_{n_r}\right)\right] & (118)\\
= & \mathbb{E}_{\mathbf{Q}_a}\left[\log\det\left(\left(\frac{PT}{n_t} + \sigma_r^2\right)\mathbf{Q}_a\mathbf{Q}_a^\dagger + \sigma_r^2 k_1\mathbf{I}_{n_r}\right)\right] & (119)\\
= & \mathbb{E}_{\mathbf{Q}_a}\left[\log\det\left(\left(\frac{PT}{n_t} + \sigma_r^2\right)\mathbf{I}_{n_r} + \sigma_r^2 k_1(\mathbf{Q}_a\mathbf{Q}_a^\dagger)^{-1}\right)\right] + k_2 & (120)\\
\leq & \mathbb{E}_{\mathbf{Q}_a}\left[\log\det\left(\left(\frac{PT}{n_t} + \sigma_r^2\right)\mathbf{I}_{n_r} + \sigma_r^2 k_1||(\mathbf{Q}_a\mathbf{Q}_a^\dagger)^{-1}||_2\mathbf{I}_{n_r}\right)\right] + k_2 & (121)\\
\leq & \log\det\left(\left(\frac{PT}{n_t} + \sigma_r^2\right)\mathbf{I}_{n_r} + \sigma_r^2 k_3\mathbf{I}_{n_r}\right) + k_2 & (122)\\
= & n_r\log\left(\frac{PT}{n_t} + \sigma_r^2(k_3 + 1)\right) + k_2, & (123)
\end{aligned}$$

where $k_2 = \mathbb{E}\left[\log\det\mathbf{Q}_a\mathbf{Q}_a^\dagger\right]$ in (120), and (121) follows from Corollary 1 as in the steps for deriving (113) from (112), because $\sigma_r^2 k_1(\mathbf{Q}_a\mathbf{Q}_a^\dagger)^{-1}$ is also Hermitian and positive semi-definite;

$k_3 = \mathbb{E}\left[k_1||(\mathbf{Q}_a\mathbf{Q}_a^\dagger)^{-1}||_2\right]$ in (122) and (122) comes from Jensen's inequality. Note that both k_2 and k_3 are finite constants independent of P. From (123), it follows that

$$\max_{p_x \in S^*_{P_x}} \mathbb{E}\left[\log\det \mathbf{Y}\mathbf{Y}^\dagger\right] \leq n_r \log P + o(\log P), \tag{124}$$

which concludes the proof of Lemma 8.

Note that, in [22] where the conventional MIMO channel was examined, it was sufficient to consider only the case where $n_t = n_r$ since, in the absence of eavesdroppers, increasing the number of transmit antennas does not improve the capacity in the high SNR regime. In this case, one can more easily rewrite (117) as

$$\mathbb{E}_{\mathbf{Q}_a}\left[\log\det\left((\mathbb{E}_{\mathbf{X}}[\mathbf{\Lambda}_x^2] + \sigma_r^2\mathbf{I}_{n_t}) + \sigma_r^2 k_1(\mathbf{Q}_a^\dagger\mathbf{Q}_a)^{-1}\right)\right] + \mathbb{E}_{\mathbf{Q}_a}\left[\log\det\mathbf{Q}_a^\dagger\mathbf{Q}_a\right] + \log\left(\sigma_r^2 k_1\right)^{n_r - n_t},$$

due to the invertibility of $\mathbf{Q}_a^\dagger\mathbf{Q}_a$. The remaining derivations are similar to that below (120). In wiretap channels, the problem does not reduce to the case where $n_t = n_r$ and, thus, a new upper-bounding technique was needed to cope with the singularity of $\mathbf{Q}_a^\dagger\mathbf{Q}_a$ when $n_t > n_r$.

Appendix E
Proof of Lemma 9

To prove this lemma, we start from (116) in the proof of Lemma 8 which is

$$\mathbb{E}\left[\log\det\mathbf{Y}\mathbf{Y}^\dagger\right] \leq \mathbb{E}_{\mathbf{Q}_a}\mathbb{E}_{\mathbf{X}}\left[\log\det\left(\mathbf{Q}_a(\mathbf{\Lambda}_x^2 + \sigma_r^2\mathbf{I}_{n_t})\mathbf{Q}_a^\dagger + \sigma_r^2 k_1 \mathbf{I}_{n_r}\right)\right]. \tag{125}$$

We rewrite the right hand side of the above inequality by replacing the determinant by the product of eigenvalues as

$$\mathbb{E}_{\mathbf{Q}_a}\sum_{i=1}^{n_r}\mathbb{E}_{\mathbf{X}}\left[\log\lambda_i\left(\mathbf{Q}_a(\mathbf{\Lambda}_x^2 + \sigma_r^2\mathbf{I}_{n_t})\mathbf{Q}_a^\dagger\right) + \sigma_r^2 k_1\right]. \tag{126}$$

Let $\mathbf{D}_x \triangleq \mathbf{\Lambda}_x^2 + \sigma_r^2\mathbf{I}_{n_t}$. Note that $\mathbf{D}_x$ is a real diagonal matrix. From Lemma 18, we know that

$$\lambda_i(\mathbf{Q}_a\mathbf{D}_x\mathbf{Q}_a^\dagger) = \lambda_i(\mathbf{D}_x^{1/2}\mathbf{Q}_a^\dagger\mathbf{Q}_a\mathbf{D}_x^{1/2}), \quad \forall i \in \{1, \ldots, n_r\}. \tag{127}$$

Let $\mathbf{Q}_{n_t,n_t} = \left[\mathbf{Q}_a^T|\mathbf{Q}_c^T\right]^T$, where $\mathbf{Q}_c \in \mathcal{C}^{(n_t - n_r)\times n_t}$ such that $\mathbf{Q}_{n_t,n_t}$ is a matrix with entries that are i.i.d. Gaussian with zero-mean and unit-variance. Then, we have

$$\mathbf{D}_x^{1/2}\mathbf{Q}_{n_t,n_t}^\dagger\mathbf{Q}_{n_t,n_t}\mathbf{D}_x^{1/2} = \mathbf{D}_x^{1/2}\mathbf{Q}_a^\dagger\mathbf{Q}_a\mathbf{D}_x^{1/2} + \mathbf{D}_x^{1/2}\mathbf{Q}_c^\dagger\mathbf{Q}_c\mathbf{D}_x^{1/2}. \tag{128}$$

Then, by applying the above, we can further upper bound (125) and (126) as

$$\mathbb{E}\left[\log\det \mathbf{Y}\mathbf{Y}^\dagger\right] \leq \mathbb{E}_{\mathbf{Q}_a}\sum_{i=1}^{n_r}\mathbb{E}_{\mathbf{X}}\left[\log\left(\lambda_i\left(\mathbf{Q}_a(\mathbf{\Lambda}_x^2+\sigma_r^2\mathbf{I}_{n_t})\mathbf{Q}_a^\dagger\right)+\sigma_r^2k_1\right)\right] \tag{129}$$

$$= \mathbb{E}_{\mathbf{Q}_a}\sum_{i=1}^{n_r}\mathbb{E}_{\mathbf{X}}\left[\log\left(\lambda_i\left(\mathbf{D}_x^{1/2}\mathbf{Q}_a^\dagger\mathbf{Q}_a\mathbf{D}_x^{1/2}\right)+\sigma_r^2k_1\right)\right] \tag{130}$$

$$\leq \mathbb{E}_{\mathbf{Q}_{n_t,n_t}}\sum_{i=1}^{n_r}\mathbb{E}_{\mathbf{X}}\left[\log\left(\lambda_i\left(\mathbf{D}_x^{1/2}\mathbf{Q}_{n_t,n_t}^\dagger\mathbf{Q}_{n_t,n_t}\mathbf{D}_x^{1/2}\right)+\sigma_r^2k_1\right)\right], \tag{131}$$

where (131) comes from (128) and Lemma 17, since $\mathbf{D}_x^{1/2}\mathbf{Q}_c^\dagger\mathbf{Q}_c\mathbf{D}_x^{1/2}$ is Hermitian and positive semi-definite.

Note that (131) only sums over the largest n_r eigenvalues of $\mathbf{D}_x^{1/2}\mathbf{Q}_{n_t,n_t}^\dagger\mathbf{Q}_{n_t,n_t}\mathbf{D}_x^{1/2}$. To further upper bound the right hand side of (131), one can add more terms corresponding to the rest of the (n_t-n_r) eigenvalues of $\mathbf{D}_x^{1/2}\mathbf{Q}_{n_t,n_t}^\dagger\mathbf{Q}_{n_t,n_t}\mathbf{D}_x^{1/2}$ as

$$\mathbb{E}_{\mathbf{Q}_{n_t,n_t}}\sum_{i=1}^{n_t}\mathbb{E}_{\mathbf{X}}\left[\log\left(\lambda_i\left(\mathbf{D}_x^{1/2}\mathbf{Q}_{n_t,n_t}^\dagger\mathbf{Q}_{n_t,n_t}\mathbf{D}_x^{1/2}\right)+1+\sigma_r^2k_1\right)\right]. \tag{132}$$

Note that we have added an additional 1 inside the logarithm of each term in the above summation to ensure that

$$\log\left(\lambda_i\left(\mathbf{D}_x^{1/2}\mathbf{Q}_{n_t,n_t}^\dagger\mathbf{Q}_{n_t,n_t}\mathbf{D}_x^{1/2}\right)+1+\sigma_r^2k_1\right)>0, \qquad \forall i\in\{n_t-n_r+1,\ldots,n_t\}, \tag{133}$$

because $k_1\sigma_r^2>0$ and each eigenvalue is non-negative. Moreover, by rewriting (132), we have

$$\mathbb{E}\left[\log\det \mathbf{Y}\mathbf{Y}^\dagger\right] \leq \mathbb{E}_{\mathbf{Q}_{n_t,n_t}}\sum_{i=1}^{n_t}\mathbb{E}_{\mathbf{X}}\left[\log\left(\lambda_i\left(\mathbf{D}_x\mathbf{Q}_{n_t,n_t}^\dagger\mathbf{Q}_{n_t,n_t}\right)+1+\sigma_r^2k_1\right)\right] \tag{134}$$

$$= \mathbb{E}_{\mathbf{Q}_{n_t,n_t}}\mathbb{E}_{\mathbf{X}}\left[\log\det\left(\mathbf{D}_x\mathbf{Q}_{n_t,n_t}^\dagger\mathbf{Q}_{n_t,n_t}+(1+\sigma_r^2k_1)\mathbf{I}_{n_t}\right)\right] \tag{135}$$

$$= \mathbb{E}_{\mathbf{Q}_{n_t,n_t}}\mathbb{E}_{\mathbf{X}}\left[\log\det\left(\mathbf{D}_x+(1+\sigma_r^2k_1)(\mathbf{Q}_{n_t,n_t}^\dagger\mathbf{Q}_{n_t,n_t})^{-1}\right)\right]+k_4 \tag{136}$$

$$\leq \mathbb{E}_{\mathbf{X}}\left[\log\det\left(\mathbf{D}_x+(1+\sigma_r^2k_1)\mathbb{E}_{\mathbf{Q}_{n_t,n_t}}\left[||(\mathbf{Q}_{n_t,n_t}^\dagger\mathbf{Q}_{n_t,n_t})^{-1}||_2\right]\mathbf{I}_{n_t}\right)\right]+k_4 \tag{137}$$

$$\leq \mathbb{E}_{\mathbf{X}}\left[\log\det\left(\mathbf{\Lambda}_x^2+k_5\mathbf{I}_{n_t}\right)\right]+k_4 \tag{138}$$

$$= \sum_{i=1}^{n_t}\mathbb{E}\left[\log\left(||\mathbf{X}_i||^2+k_5\right)\right]+k_4, \tag{139}$$

where $k_4=\mathbb{E}\left[\log\det\mathbf{Q}_{n_t,n_t}^\dagger\mathbf{Q}_{n_t,n_t}\right]$ and $k_5=\sigma_r^2+(1+\sigma_r^2k_1)\mathbb{E}\left[||(\mathbf{Q}_{n_t,n_t}^\dagger\mathbf{Q}_{n_t,n_t})^{-1}||_2\right]$, and the derivation of (137) from (136) follows similarly to that of (113) from (112). Finally, from the upper bound of $\mathbb{E}\left[\log\det \mathbf{Y}\mathbf{Y}^{\dagger}\right]$ in (139), we have

$$\begin{aligned}
&\mathbb{E}\left[\log\det \mathbf{Y}\mathbf{Y}^{\dagger}\right] - \sum_{i=1}^{n_t} \mathbb{E}\left[\log(||\mathbf{X}_i||^2 + \sigma_r^2)\right] \\
&\qquad \le \sum_{i=1}^{n_t} \mathbb{E}\left[\log\left(||\mathbf{X}_i||^2 + k_5\right)\right] + k_4 - \sum_{i=1}^{n_t} \mathbb{E}\left[\log(||\mathbf{X}_i||^2 + \sigma_r^2)\right] && (140) \\
&\qquad = \sum_{i=1}^{n_t} \mathbb{E}\left[\log\left(\frac{||\mathbf{X}_i||^2 + k_5}{||\mathbf{X}_i||^2 + \sigma_r^2}\right)\right] + k_4 && (141) \\
&\qquad \le \sum_{i=1}^{n_t} \mathbb{E}\left[\left(\frac{k_5 - \sigma_r^2}{||\mathbf{X}_i||^2 + \sigma_r^2}\right)\right] + k_4 && (142) \\
&\qquad \le n_t \left(\frac{k_5 - \sigma_r^2}{\sigma_r^2}\right) + k_4, && (143)
\end{aligned}$$

where (142) comes from $k_5 > \sigma_r^2$ by definition and $\log(1+x) < x$ when $x \ge 0$. Since k_4 in (136) and k_5 in (138) are finite constants independent of P, this completes the proof of Lemma 9.

Note here that, similar to Lemma 8, one cannot directly apply the results in [22] to prove Lemma 9. This is because the results in [22] rely on the invertibility of $\mathbf{Q}_a^{\dagger}\mathbf{Q}_a$, as mentioned below (124). However, this property may not hold here and, thus, new techniques were needed to upper bound $\mathbb{E}\left[\log\det \mathbf{Y}\mathbf{Y}^{\dagger}\right]$ as presented above.

Appendix F
Proof of Lemma 10

Let $\mathbf{Y}_{n_t}$, $\mathbf{H}_{n_t}$, and $\mathbf{N}_{r,n_t}$ be matrices formed by taking n_t rows arbitrarily from $\mathbf{Y}$, $\mathbf{H}$, and $\mathbf{N}_r$, respectively, and let $\mathbf{Y}_{n_r-n_t}$, $\mathbf{H}_{n_r-n_t}$, and $\mathbf{N}_{r,n_r-n_t}$ be the remaining $(n_r - n_t)$ rows of $\mathbf{Y}$, $\mathbf{H}$, and $\mathbf{N}_r$, respectively. Thus, we have

$$\mathbf{Y}_{n_t} = \mathbf{H}_{n_t}\mathbf{X} + \mathbf{N}_{r,n_t}, \tag{144}$$

$$\mathbf{Y}_{n_r-n_t} = \mathbf{H}_{n_r-n_t}\mathbf{X} + \mathbf{N}_{r,n_r-n_t}. \tag{145}$$

First, note that, if $n_t < n_r$, we can represent the channel matrix $\mathbf{H}_{n_r-n_t}$ in terms of linear combinations of row vectors of $\mathbf{H}_{n_t}$. Thus, we have

$$\mathbf{Y}_{n_r-n_t} = C(\mathbf{H})\mathbf{Y}_{n_t} + \mathbf{N}_{r,n_r-n_t} - C(\mathbf{H})\mathbf{N}_{r,n_t}, \tag{146}$$

where $C(\mathbf{H})$ is the linear combination matrix such that $C(\mathbf{H})\mathbf{H}_{n_t} = \mathbf{H}_{n_r-n_t}$. From (146), the following Markov relation holds

$$\mathbf{X} \rightarrow (\mathbf{Y}_{n_t}, C(\mathbf{H}), \mathbf{N}_{r,n_r-n_t} - C(\mathbf{H})\mathbf{N}_{r,n_t}) \rightarrow (\mathbf{Y}_{n_t}, \mathbf{Y}_{n_r-n_t}) \rightarrow \mathbf{Y}. \tag{147}$$

Hence, from the data processing inequality, we have

$$\begin{aligned}
I(\mathbf{X};\mathbf{Y}) - I(\mathbf{X};\mathbf{Y}_{n_t}) &\leq I(\mathbf{X};\mathbf{Y}_{n_t}, C(\mathbf{H}), \mathbf{N}_{r,n_r-n_t} - C(\mathbf{H})\mathbf{N}_{r,n_t}) - I(\mathbf{X};\mathbf{Y}_{n_t}) && (148)\\
&= I(\mathbf{X}; C(\mathbf{H}), \mathbf{N}_{r,n_r-n_t} - C(\mathbf{H})\mathbf{N}_{r,n_t}|\mathbf{Y}_{n_t}) && (149)\\
&= h(C(\mathbf{H}), \mathbf{N}_{r,n_r-n_t} - C(\mathbf{H})\mathbf{N}_{r,n_t}|\mathbf{Y}_{n_t})\\
&\quad - h(C(\mathbf{H}), \mathbf{N}_{r,n_r-n_t} - C(\mathbf{H})\mathbf{N}_{r,n_t}|\mathbf{Y}_{n_t}, \mathbf{X}). && (150)
\end{aligned}$$

The first term in (150) can be upper bounded by

$$h(C(\mathbf{H}), \mathbf{N}_{r,n_r-n_t} - C(\mathbf{H})\mathbf{N}_{r,n_t}|\mathbf{Y}_{n_t}) \leq h(C(\mathbf{H}), \mathbf{N}_{r,n_r-n_t} - C(\mathbf{H})\mathbf{N}_{r,n_t}) = o(\log P), \tag{151}$$

and the second term in (150) is lower bounded by

$$\begin{aligned}
h(C(\mathbf{H}),&\mathbf{N}_{r,n_r-n_t} - C(\mathbf{H})\mathbf{N}_{r,n_t}|\mathbf{Y}_{n_t}, \mathbf{X})\\
&\geq h(C(\mathbf{H}), \mathbf{N}_{r,n_r-n_t} - C(\mathbf{H})\mathbf{N}_{r,n_t}|\mathbf{Y}_{n_t}, \mathbf{X}, \mathbf{H}_{n_t}) && (152)\\
&= h(C(\mathbf{H}), \mathbf{N}_{r,n_r-n_t} - C(\mathbf{H})\mathbf{N}_{r,n_t}|\mathbf{H}_{n_t}, \mathbf{X}, \mathbf{N}_{r,n_t}) && (153)\\
&= h(C(\mathbf{H}), \mathbf{N}_{r,n_r-n_t} - C(\mathbf{H})\mathbf{N}_{r,n_t}|\mathbf{H}_{n_t}, \mathbf{N}_{r,n_t}) && (154)\\
&= h(C(\mathbf{H})|\mathbf{H}_{n_t}, \mathbf{N}_{r,n_t}) + h(\mathbf{N}_{r,n_r-n_t} - C(\mathbf{H})\mathbf{N}_{r,n_t}|\mathbf{H}_{n_t}, \mathbf{N}_{r,n_t}, C(\mathbf{H})) && (155)\\
&= h(C(\mathbf{H})|\mathbf{H}_{n_t}, \mathbf{N}_{r,n_t}) + h(\mathbf{N}_{r,n_r-n_t}) = o(\log P). && (156)
\end{aligned}$$

The result in (56) of Lemma 10 then follows from (150), (151), and (156).

References

[1] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal*, vol. 28, pp. 656–715, Oct. 1949.

[2] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, pp. 1355–1387, Oct. 1975.

[3] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, pp. 339–348, May 1978.

[4] S. Leung-Yan-Cheong and M. Hellman, "The Gaussian wiretap channel," *IEEE Transactions on Information Theory*, vol. 24, no. 4, pp. 451–456, Jul. 1978.

[5] S. Shafiee, N. Liu, and S. Ulukus, "Towards the secrecy capacity of the Gaussian MIMO wire-tap channel: The 2-2-1 channel," *IEEE Transactions on Information Theory*, vol. 55, no. 9, pp. 4033–4039, Sep. 2009.

[6] A. Khisti and G. Wornell, "Secure transmission with multiple antennas II: The MIMOME wiretap channel," *IEEE Transactions on Information Theory*, vol. 56, pp. 5515–5532, Nov. 2010.

[7] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4961–4972, Aug. 2011.

[8] T. Liu and S. Shamai, "A note on the secrecy capacity of the multi-antenna wiretap channel," *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2547–2553, Jun. 2009.

[9] Y. Liang and H. V. Poor, "Secure communication over fading channels," in Proc. of *44th annual Allerton Conference on Communication, Control, and Computing*, September 2006.

[10] Z. Li, R. D. Yates, and W. Trappe, "Secrecy capacity of independent parallel channels," in Proc. of *44th annual Allerton Conference on Communication, Control, and Computing*, Sep 2006.

[11] Y. Liang, H. V. Poor, and S. Shamai, "Secure communication over fading channels," *IEEE Transactions on Information Theory*, vol. 54, no. 6, pp. 2470–2492, Jun. 2008.

[12] Z. Li, R. D. Yates, and W. Trappe, "Secrecy capacity of independent parallel channels," in *Securing Wireless Communications at the Physical Layer*, R. Liu and W. Trappe, Eds. Springer US, 2010, pp. 1–18.

[13] P. K. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4687–4698, Oct. 2008.

[14] Z. Li, R. D. Yates, and W. Trappe, "Achieving secret communication for fast rayleigh fading channels," *IEEE Transactions on Wireless Communications*, vol. 9, no. 9, pp. 2792–2799, Sep. 2010.

[15] S.-C. Lin and C.-L. Lin, "On secrecy capacity of fast fading MIMOME wiretap channels with statistical CSIT," *IEEE Transactions on Wireless Communications*, vol. 13, no. 6, pp. 3293 –3306, 2014.

[16] X. He and A. Yener, "MIMO wiretap channels with arbitrarily varying eavesdropper channel states," *IEEE Transactions on Information Theory*, submitted, Jul. 2010. Available at [arXiv:1007.4801].

[17] S. Yang, M. Kobayashi, P. Piantanida, and S. Shamai, "Secrecy degrees of freedom of MIMO broadcast channels with delayed CSIT," *IEEE Transactions on Information Theory*, vol. 59, no. 9, pp. 5244–5256, 2013.

[18] L. Czap, V. M. Prabhakaran, S. Diggavi, and C. Fragouli, "Broadcasting private messages securely," in *IEEE International Symposium on Information Theory*, Jul. 2012.

[19] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Quantized vs. analog feedback for the MIMO broadcast channel: A comparison between zero forcing based achievable rates," in *IEEE International Symposium on Information Theory*, 2007.

[20] P. Mukherjee and S. Ulukus, "Fading wiretap channel with no CSI anywhere," in *IEEE International Symposium on Information Theory*, Jul. 2013.

[21] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139 –157, Jan. 1999.

[22] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Wiley-Interscience, 2006.

[24] M. van Dijk, "On a special class of broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 712 –714, Mar. 1997.

[25] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[26] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities : Theory of Majorization and Its Applications*, 2nd ed., ser. Springer series in statistics. Springer, 2011.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.