The MIMOME Channel

Ashish Khisti and Gregory Wornell

Abstract— **The MIMOME channel is a Gaussian wiretap channel in which the sender, receiver, and eavesdropper all have multiple antennas. We characterize the secrecy capacity as the saddle-value of a minimax problem. Among other implications, our result establishes that a Gaussian distribution maximizes the secrecy capacity characterization of Csiszár and Körner when applied to the MIMOME channel. We also determine a necessary and sufficient condition for the secrecy capacity to be zero. Large antenna array analysis of this condition reveals several useful insights into the conditions under which secure communication is possible.**

I. Introduction

Multiple antennas are a valuable resource in wireless communications. Recently there has been a significant activity in exploring both the theoretical and practical aspects of wireless systems with multiple antennas. In this work we explore the role of multiple antennas for physical layer security, which is an emerging area of interest.

The wiretap channel [1] is an information theoretic model for physical layer security. The setup has three terminals — one sender, one receiver and one eavesdropper. The goal is to exploit the structure of the underlying broadcast channel to transmit a message reliably to the intended receiver, while leaking asymptotically no information to the eavesdropper. A single letter characterization of the secrecy capacity, when the underlying channel is a discrete memoryless broadcast channel, has been obtained by Csiszár and Körner [2]. An explicit solution for the scalar Gaussian case is obtained in [3].

In this paper we consider the case where all the three terminals have multiple antennas and naturally refer to it as multiple input, multiple output, multiple eavesdropper (MIMOME) channel. In this setup we assume that the channel matrices are fixed and known to all the three terminals. While the assumption that the eavesdropper's channel is known to both the sender and the receiver is obviously a strong assumption, we remark in advance that our solution provides ultimate limits on secure transmission with multiple antennas and could be a starting point for other formulations where the eavesdropper's channel may not be known to the sender and the receiver.

The main result of this paper is a characterization of the secrecy capacity of the MIMOME channel as the saddle value of a minimax problem. Our approach does not rely on the Csiszár and Körner capacity expression, but instead is based on the technique used in characterizing the sum rate of the MIMO broadcast channel (see, e.g., [4] and its references). We first develop a minimax expression that upper bounds the secrecy capacity and subsequently establish the tightness of this bound for the MIMOME channel.

The case where the channel matrices of intended receiver and eavesdropper are square and diagonal follows from the results in [5]–[8] that consider secure transmission over fading channels. The difficulty of optimizing the Csiszár and Körner expression for the general case has been reported in [9]–[11] and achievable rates have been investigated. The approach used in the present paper has been used in our earlier work [12], [13] to establish the secrecy capacity for two special cases: the case when the intended receiver has a single antenna (MISOME case) and the MIMOME secrecy capacity in the high SNR regime. This upper bounding approach was independently conceived by Ulukus et. al. [14] and further applied to the 2x2x1 case [15]. Finally, a related approach for the MIMOME channel, is developed independently in [16]. Also it is interesting to note that this upper bounding approach has been empirically observed to be tight for the problem of broadcasting two private messages to two receivers when each receiver has a single antenna [17]. For this setup a single letter characterization is not known for the discrete memoryless case [18], [19]

II. Channel Model

We denote the number of antennas at the sender, the receiver and the eavesdropper by n_t, n_r and n_e respectively.

$$\begin{aligned}\mathbf{y}_\mathrm{r}(t) &= \mathbf{H}_\mathrm{r}\mathbf{x}(t) + \mathbf{z}_\mathrm{r}(t)\\ \mathbf{y}_\mathrm{e}(t) &= \mathbf{H}_\mathrm{e}\mathbf{x}(t) + \mathbf{z}_\mathrm{e}(t),\end{aligned} \tag{1}$$

where $\mathbf{H}_\mathrm{r} \in \mathbb{C}^{n_\mathrm{r}\times n_\mathrm{t}}$ and $\mathbf{H}_\mathrm{e} \in \mathbb{C}^{n_\mathrm{e}\times n_\mathrm{t}}$ are channel matrices associated with the receiver and the eavesdropper. The channel matrices are fixed for the entire transmission period and known to all the three terminals. The additive noise $\mathbf{z}_\mathrm{r}(t)$ and $\mathbf{z}_\mathrm{e}(t)$ are circularly-symmetric and complex-valued Gaussian random variables. The input satisfies a power constraint $E\left[\frac{1}{n}\sum_{t=1}^{n}||\mathbf{x}(t)||^2\right] \le P$.

A rate R is achievable if there exists a sequence of length n codes, such that the error probability at the intended receiver and $\frac{1}{n}I(w;\mathbf{y}_\mathrm{e}^n)$ both approach zero as $n\to\infty$. The secrecy capacity is the supremum of all achievable rates.

III. MIMOME Secrecy Capacity

Our main result is the following characterization of the secrecy capacity of the MIMOME wiretap channel.

Theorem 1: The secrecy capacity of the MIMOME wiretap channel is

$$C = \min_{\mathbf{K}_\mathbf{\Phi}\in\mathcal{K}_\mathbf{\Phi}} \max_{\mathbf{K}_\mathrm{P}\in\mathcal{K}_\mathrm{P}} R_+(\mathbf{K}_\mathrm{P},\mathbf{K}_\mathbf{\Phi}), \tag{2}$$

This work was supported in part by NSF under Grant No. CCF-0515109.
The authors are with the Dept. EECS, MIT, Cambridge, MA, 02139. Email:{khisti,gww}@mit.edu

where $R_+(\mathbf{K}_\mathrm{P}, \mathbf{K}_{\mathbf{\Phi}}) = I(\mathsf{x}; \mathsf{y}_\mathrm{r} \mid \mathsf{y}_\mathrm{e})$ with $\mathsf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_\mathrm{P})$ and

$$\mathcal{K}_\mathrm{P} \triangleq \left\{ \mathbf{K}_\mathrm{P} \;\middle|\; \mathbf{K}_\mathrm{P} \succeq \mathbf{0}, \quad \mathrm{tr}(\mathbf{K}_\mathrm{P}) \le P \right\}, \tag{3}$$

and where $[\mathbf{z}_\mathrm{r}^\dagger, \mathbf{z}_\mathrm{e}^\dagger]^\dagger \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_{\mathbf{\Phi}})$, with

$$\begin{aligned} \mathcal{K}_{\mathbf{\Phi}} &\triangleq \left\{ \mathbf{K}_{\mathbf{\Phi}} \;\middle|\; \mathbf{K}_{\mathbf{\Phi}} = \begin{bmatrix} \mathbf{I}_{n_\mathrm{r}} & \mathbf{\Phi} \\ \mathbf{\Phi}^\dagger & \mathbf{I}_{n_\mathrm{e}} \end{bmatrix}, \quad \mathbf{K}_{\mathbf{\Phi}} \succeq \mathbf{0} \right\} \\ &= \left\{ \mathbf{K}_{\mathbf{\Phi}} \;\middle|\; \mathbf{K}_{\mathbf{\Phi}} = \begin{bmatrix} \mathbf{I}_{n_\mathrm{r}} & \mathbf{\Phi} \\ \mathbf{\Phi}^\dagger & \mathbf{I}_{n_\mathrm{e}} \end{bmatrix}, \quad \sigma_{\max}(\mathbf{\Phi}) \le 1 \right\}. \end{aligned} \tag{4}$$

Furthermore,[1] the minimax problem in (2) has a saddle point solution $(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)$ and the secrecy capacity can also be expressed as,

$$C = R_+(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) = \log \frac{\det(\mathbf{I} + \mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger)}{\det(\mathbf{I} + \mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger)}. \tag{5}$$

A. Connection with Csiszár and Körner Capacity

A characterization of the secrecy capacity for the non-degraded discrete memoryless broadcast channel $p_{\mathsf{y}_\mathrm{r},\mathsf{y}_\mathrm{e}|\mathsf{x}}$ is provided by Csiszár and Körner [2],

$$C = \max_{p_\mathsf{u}, p_{\mathsf{x}|\mathsf{u}}} I(\mathsf{u}; \mathsf{y}_\mathrm{r}) - I(\mathsf{u}; \mathsf{y}_\mathrm{e}), \tag{6}$$

where u is an auxiliary random variable (over a certain alphabet with bounded cardinality) that satisfies $\mathsf{u} \to \mathsf{x} \to (\mathsf{y}_\mathrm{r}, \mathsf{y}_\mathrm{e})$. As remarked in [2], the secrecy capacity (6) can be extended in principle to incorporate continuous-valued inputs. However, directly identifying the optimal u for the MIMOME case is not straightforward.

Theorem 1 indirectly establishes an optimal choice of u in (6). Suppose that $(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)$ is a saddle point solution to the minimax problem in (2). From (5) we have

$$R_+(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) = R_-(\bar{\mathbf{K}}_\mathrm{P}), \tag{7}$$

where

$$R_-(\bar{\mathbf{K}}_\mathrm{P}) \triangleq \log \frac{\det(\mathbf{I} + \mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger)}{\det(\mathbf{I} + \mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger)}$$

is the achievable rate obtained by evaluating (6) for $\mathsf{u} = \mathsf{x} \sim \mathcal{CN}(\mathbf{0}, \bar{\mathbf{K}}_\mathrm{P})$. This choice of $p_\mathsf{u}, p_{\mathsf{x}|\mathsf{u}}$ thus maximizes (6). Furthermore note that

$$\bar{\mathbf{K}}_\mathrm{P} \in \arg\max_{\mathbf{K}_\mathrm{P} \in \mathcal{K}_\mathrm{P}} \log \frac{\det(\mathbf{I} + \mathbf{H}_\mathrm{r}\mathbf{K}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger)}{\det(\mathbf{I} + \mathbf{H}_\mathrm{e}\mathbf{K}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger)} \tag{8}$$

where the set $\mathcal{K}_\mathrm{P}$ is defined in (3). Unlike the minimax problem (2) the maximization problem (8) is not a convex optimization problem since the objective function is not a concave function of $\mathbf{K}_\mathrm{P}$. Even if one verifies that $\bar{\mathbf{K}}_\mathrm{P}$ satisfies the optimality conditions associated with (8), this will only establish that $\bar{\mathbf{K}}_\mathrm{P}$ is a locally optimal solution. The capacity expression (2) provides a convex reformulation of (8) and establishes that $\bar{\mathbf{K}}_\mathrm{P}$ is a globally optimal solution in (8).[2]

B. Structure of the optimal solution

As we establish in Section IV-D, if $(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)$ is a saddle point solution to the minimax problem, if $\mathbf{S}$ is any matrix that has a full column rank matrix and satisfies $\bar{\mathbf{K}}_\mathrm{P} = \mathbf{S}\mathbf{S}^\dagger$ and if $\bar{\mathbf{\Phi}}$ is the cross-covariance matrix between the noise random variables (c.f. (4)), then

$$\mathbf{H}_\mathrm{e}\mathbf{S} = \bar{\mathbf{\Phi}}^\dagger \mathbf{H}_\mathrm{r}\mathbf{S}. \tag{9}$$

The condition in (9) admits an intuitive interpretation. From (4) $\bar{\mathbf{\Phi}}$ is a contraction matrix i.e., all its singular values are less than or equal to unity. The column space of $\mathbf{S}$ is the subspace in which the sender transmits information. So (9) states that no information is transmitted along any direction where the eavesdropper observes a stronger signal than the intended receiver. The effective channel of the eavesdropper, $\mathbf{H}_\mathrm{e}\mathbf{S}$, is a degraded version of the effective channel of the intended receiver, $\mathbf{H}_\mathrm{r}\mathbf{S}$.

IV. Proof of Theorem 1

Our proof involves two main parts. First we show that the right hand side in (2) is an upper bound on the secrecy capacity. Then we examine the optimality conditions associated with the saddle point solution to establish (7), which completes the proof since

$$C \le R_+(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) = R_-(\bar{\mathbf{K}}_\mathrm{P}) \le C.$$

That the right hand side in (2) is an upper bound on the secrecy capacity has already been established:

Lemma 1 ([12], [13]): An upper bound on the secrecy capacity for the MIMOME channel is

$$C \le \min_{\mathbf{K}_{\mathbf{\Phi}} \in \mathcal{K}_{\mathbf{\Phi}}} \max_{\mathbf{K}_\mathrm{P} \in \mathcal{K}_\mathrm{P}} R_+(\mathbf{K}_\mathrm{P}, \mathbf{K}_{\mathbf{\Phi}}), \tag{10}$$

where $\mathcal{K}_\mathrm{P}$ and $\mathcal{K}_{\mathbf{\Phi}}$ are defined in (3) and (4) respectively.

Hence it suffices to establish (7), which we do in the remainder of this section. We divide the proof into several steps, which are outlined in Fig. 1.

A. Existence of the Saddle Point

Our first step is to show that for the minimax problem in (2), a saddle point solution exists, i.e., there exists a point $(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)$ with $\bar{\mathbf{K}}_\mathrm{P} \in \mathcal{K}_\mathrm{P}$ and $\bar{\mathbf{K}}_\Phi \in \mathcal{K}_{\mathbf{\Phi}}$, such that for any $\mathbf{K}_\mathrm{P} \in \mathcal{K}_\mathrm{P}$ and $\mathbf{K}_{\mathbf{\Phi}} \in \mathcal{K}_{\mathbf{\Phi}}$, we have that

$$R_+(\mathbf{K}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) \le R_+(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) \le R_+(\bar{\mathbf{K}}_\mathrm{P}, \mathbf{K}_{\mathbf{\Phi}}). \tag{11}$$

Towards this end, we show the following convexity properties of the objective function.

Claim 1: For any fixed $\mathbf{K}_\mathrm{P} \in \mathcal{K}_\mathrm{P}$, the function $R_+(\mathbf{K}_\mathrm{P}, \mathbf{K}_{\mathbf{\Phi}})$ is convex in $\mathbf{K}_{\mathbf{\Phi}}$. For any fixed $\mathbf{K}_{\mathbf{\Phi}} \in \mathcal{K}_{\mathbf{\Phi}}$, the function $R_+(\mathbf{K}_\mathrm{P}, \mathbf{K}_{\mathbf{\Phi}})$ is concave in $\mathbf{K}_\mathrm{P}$.

[1] In the remainder of this paper, $\mathbf{I}$ denotes an identity matrix and $\mathbf{0}$ denotes the matrix with all zeros. The dimensions of these matrices will be suppressed and will be clear from the context. Also we use the superscript † to denote the hermitian conjugate of a matrix.

[2] The "high SNR" case of this problem i.e., $\max_{\mathbf{K} \in \mathcal{K}_\infty} \log \frac{\det(\mathbf{H}_\mathrm{r}\mathbf{K}\mathbf{H}_\mathrm{r}^\dagger)}{\det(\mathbf{H}_\mathrm{e}\mathbf{K}\mathbf{H}_\mathrm{e}^\dagger)}$ is known as the multiple-discriminant-function in multivariate statistics and is well-studied; see, e.g., [20].

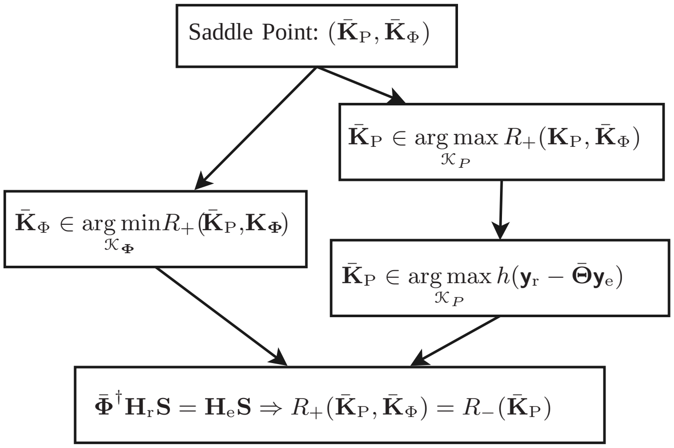

Fig. 1. Key steps in the Proof of Theorem 1. In section IV-A we establish that the minimax problem has a saddle point $(\bar{\mathbf{K}}_{\rm P}, \bar{\mathbf{K}}_{\Phi})$. Section IV-B obtains a condition satisfied by $(\bar{\mathbf{K}}_{\rm P}, \bar{\mathbf{K}}_{\Phi})$ via the KKT conditions associated with the noise covariance, while section IV-C obtains another condition that $(\bar{\mathbf{K}}_{\rm P}, \bar{\mathbf{K}}_{\Phi})$ satisfy, by first showing that $\bar{\mathbf{K}}_{\rm P}$ is also an optimal covariance of another MIMO channel. Combining these two conditions we show in Section IV-D that the upper and lower bounds coincide.

Proof: Recall that $R_+(\mathbf{K}_{\rm P}, \mathbf{K}_{\mathbf{\Phi}}) = I(\mathbf{x}; \mathbf{y}_{\rm r}, \mathbf{y}_{\rm e}) - I(\mathbf{x}; \mathbf{y}_{\rm e})$, with $\mathbf{x} \sim \mathcal{CN}(0, \mathbf{K}_{\rm P})$ and $[\mathbf{z}_{\rm r}^\dagger \ \mathbf{z}_{\rm e}^\dagger]^\dagger \sim \mathcal{CN}(0, \mathbf{K}_{\mathbf{\Phi}})$. For the convexity in $\mathbf{K}_{\mathbf{\Phi}}$, note that, $I(\mathbf{x}; \mathbf{y}_{\rm e})$ does not depend on $\mathbf{K}_{\mathbf{\Phi}}$, and $I(\mathbf{x}; \mathbf{y}_{\rm r}, \mathbf{y}_{\rm e})$ is known (see e.g., [21]) to be convex in $\mathbf{K}_{\mathbf{\Phi}}$. For the concavity in $\mathbf{K}_{\rm P}$, note that when $\mathbf{K}_{\mathbf{\Phi}} \succ \mathbf{0}$, we can express

$$R_+(\mathbf{K}_{\rm P}, \mathbf{K}_{\mathbf{\Phi}}) = \log\det \mathbf{\Lambda}(\mathbf{K}_{\rm P}) - \log\det \mathbf{K}_{\mathbf{\Phi}}, \qquad (12)$$

where

$$\mathbf{\Lambda}(\mathbf{K}_{\rm P}) \triangleq \mathbf{I} + \mathbf{H}_{\rm r}\mathbf{K}_{\rm P}\mathbf{H}_{\rm r}^\dagger - (\mathbf{\Phi} + \mathbf{H}_{\rm r}\mathbf{K}_{\rm P}\mathbf{H}_{\rm e}^\dagger)(\mathbf{I} + \mathbf{H}_{\rm e}\mathbf{K}_{\rm P}\mathbf{H}_{\rm e}^\dagger)^{-1}(\mathbf{\Phi}^\dagger + \mathbf{H}_{\rm e}\mathbf{K}_{\rm P}\mathbf{H}_{\rm r}^\dagger) \qquad (13)$$

is the Schur compliment of the matrix

$$\begin{bmatrix} \mathbf{I} + \mathbf{H}_{\rm r}\mathbf{K}_{\rm P}\mathbf{H}_{\rm r}^\dagger & \mathbf{\Phi} + \mathbf{H}_{\rm r}\mathbf{K}_{\rm P}\mathbf{H}_{\rm e}^\dagger \\ \mathbf{\Phi}^\dagger + \mathbf{H}_{\rm e}\mathbf{K}_{\rm P}\mathbf{H}_{\rm r}^\dagger & \mathbf{I} + \mathbf{H}_{\rm e}\mathbf{K}_{\rm P}\mathbf{H}_{\rm e}^\dagger \end{bmatrix}. \qquad (14)$$

Since the Schur complement is jointly concave in the constituent matrices [22, page 21, Corollary 1.5.3], which in turn are linear in $\mathbf{K}_{\rm P}$, it follows that $\mathbf{\Lambda}(\mathbf{K}_{\rm P})$ is concave in $\mathbf{K}_{\rm P}$ and hence from the composition theorem we have that $R_+(\mathbf{K}_{\rm P}, \mathbf{K}_{\mathbf{\Phi}})$ is concave[3] in $\mathbf{K}_{\rm P}$. The case when $\mathbf{K}_{\mathbf{\Phi}}$ is singular, can be handled via the singular value decomposition of $\mathbf{\Phi}$, and will be treated in the full paper. ■

Notice that both the domain sets $\mathcal{K}_{\rm P}$ and $\mathcal{K}_{\mathbf{\Phi}}$ are convex and compact, hence the existence of a saddle point solution $(\bar{\mathbf{K}}_{\rm P}, \bar{\mathbf{K}}_{\Phi})$ is established via the minimax theorem [23].

In the sequel, we define $\bar{\mathbf{\Phi}}$ via

$$\bar{\mathbf{K}}_{\Phi} = \begin{bmatrix} \mathbf{I}_{n_{\rm r}} & \bar{\mathbf{\Phi}} \\ \bar{\mathbf{\Phi}}^\dagger & \mathbf{I}_{n_{\rm e}} \end{bmatrix}. \qquad (15)$$

[3]The concavitiy result can also be established via [25, pg. 506, Theorem 16.9.1], by observing that $\mathbf{I} + \mathbf{H}_{\rm e}\mathbf{K}_{\rm P}\mathbf{H}_{\rm e}^\dagger$ is a minor of the matrix in (14).

B. Least favorable noise condition

From (11), we have that

$$\bar{\mathbf{K}}_{\Phi} \in \arg\min_{\mathbf{K}_{\mathbf{\Phi}} \in \mathcal{K}_{\mathbf{\Phi}}} R_+(\bar{\mathbf{K}}_{\rm P}, \mathbf{K}_{\mathbf{\Phi}}). \qquad (16)$$

The optimality conditions associated with (16) yield the following.

Lemma 2: Suppose that $(\bar{\mathbf{K}}_{\rm P}, \bar{\mathbf{K}}_{\Phi})$ is a saddle point solution to the minimax problem in (2). Then

$$(\mathbf{H}_{\rm r} - \bar{\mathbf{\Theta}}\mathbf{H}_{\rm e})\bar{\mathbf{K}}_{\rm P}(\bar{\mathbf{\Phi}}^\dagger\mathbf{H}_{\rm r} - \mathbf{H}_{\rm e})^\dagger = \mathbf{0}. \qquad (17)$$

where $\bar{\mathbf{\Phi}}$ is as defined via (15) and

$$\bar{\mathbf{\Theta}} = (\mathbf{H}_{\rm r}\bar{\mathbf{K}}_{\rm P}\mathbf{H}_{\rm e}^\dagger + \bar{\mathbf{\Phi}})(\mathbf{I} + \mathbf{H}_{\rm e}\bar{\mathbf{K}}_{\rm P}\mathbf{H}_{\rm e}^\dagger)^{-1}. \qquad (18)$$

We will see subsequently, that (17) has a useful structure, which can be combined with the optimality condition associated with $\bar{\mathbf{K}}_{\rm P}$. The proof is most direct when the noise covariance $\bar{\mathbf{K}}_{\Phi}$ at the saddle point is non-singular. Hence we will establish (17) in this special case first and then consider the case when $\bar{\mathbf{K}}_{\Phi}$ is singular.

1) $\bar{\mathbf{K}}_{\Phi}$ is non-singular.: The Lagrangian associated with the minimization (16) is

$$\mathcal{L}_{\mathbf{\Phi}}(\mathbf{K}_{\mathbf{\Phi}}, \mathbf{\Upsilon}) = R_+(\bar{\mathbf{K}}_{\rm P}, \mathbf{K}_{\mathbf{\Phi}}) + \mathrm{tr}(\mathbf{\Upsilon}\mathbf{K}_{\mathbf{\Phi}}), \qquad (19)$$

where the dual variable

$$\mathbf{\Upsilon} = \begin{matrix} & n_{\rm r} & n_{\rm e} \\ n_{\rm r} & \mathbf{\Upsilon}_1 & \mathbf{0} \\ n_{\rm e} & \mathbf{0} & \mathbf{\Upsilon}_2 \end{matrix} \qquad (20)$$

is a block diagonal matrix corresponding to the constraint that the noise covariance $\mathbf{K}_{\mathbf{\Phi}}$ must have identity matrices on its diagonal. The associated Karush-Kuhn-Tucker (KKT) conditions yield

$$\nabla_{\mathbf{K}_{\mathbf{\Phi}}} R_+(\bar{\mathbf{K}}_{\rm P}, \mathbf{K}_{\mathbf{\Phi}})\big|_{\bar{\mathbf{K}}_{\Phi}} + \mathbf{\Upsilon} = \mathbf{0},$$

where

$$\nabla_{\mathbf{K}_{\mathbf{\Phi}}} R_+(\bar{\mathbf{K}}_{\rm P}, \mathbf{K}_{\mathbf{\Phi}})\big|_{\bar{\mathbf{K}}_{\Phi}} \qquad (21)$$
$$= \nabla_{\mathbf{K}_{\mathbf{\Phi}}} \left[\log\det(\mathbf{K}_{\mathbf{\Phi}} + \mathbf{H}_{\rm t}\bar{\mathbf{K}}_{\rm P}\mathbf{H}_{\rm t}^\dagger) - \log\det(\mathbf{K}_{\mathbf{\Phi}})\right]\Big|_{\bar{\mathbf{K}}_{\Phi}}$$
$$= (\bar{\mathbf{K}}_{\Phi} + \mathbf{H}_{\rm t}\bar{\mathbf{K}}_{\rm P}\mathbf{H}_{\rm t}^\dagger)^{-1} - \bar{\mathbf{K}}_{\Phi}^{-1}, \qquad (22)$$

with the convenient notation

$$\mathbf{H}_{\rm t} = \begin{bmatrix} \mathbf{H}_{\rm r} \\ \mathbf{H}_{\rm e} \end{bmatrix}, \qquad (23)$$

which in turn implies that

$$\mathbf{H}_{\rm t}\bar{\mathbf{K}}_{\rm P}\mathbf{H}_{\rm t}^\dagger = \bar{\mathbf{K}}_{\Phi}\mathbf{\Upsilon}(\bar{\mathbf{K}}_{\Phi} + \mathbf{H}_{\rm t}\bar{\mathbf{K}}_{\rm P}\mathbf{H}_{\rm t}^\dagger). \qquad (24)$$

The relation in (17) follows from (24) through a straightforward computation that exploits the block diagonal structure of $\mathbf{\Upsilon}$, which we provide in Appendix I.

2) $\bar{\mathbf{K}}_\Phi$ is singular: When the noise covariance $\bar{\mathbf{K}}_\Phi$ is singular, as we now show, (24) still holds. Note that this will complete the proof, since the steps in Appendix I that simplify (24) do not require that $\bar{\mathbf{K}}_\Phi$ be non-singular.

In the singular case we define another optimization problem whose optimality conditions yield (24). An analogous approach has been taken earlier by Yu [4] for dealing with singular noise for the MIMO broadcast channel.

Suppose that

$$\bar{\mathbf{K}}_\Phi = \mathbf{W}\bar{\mathbf{\Omega}}\mathbf{W}^\dagger, \tag{25}$$

where $\mathbf{W}$ is a matrix with orthogonal columns, i.e., $\mathbf{W}^\dagger\mathbf{W} = \mathbf{I}$ and $\bar{\mathbf{\Omega}}$ is a non-singular matrix. We first note that it must also be the case that

$$\mathbf{H}_\mathrm{t} = \mathbf{W}\mathbf{G}, \tag{26}$$

i.e., the column space of $\mathbf{H}_\mathrm{t}$ is a subspace of the column space of $\mathbf{W}$. If this were not the case, by receiving a signal in the null space of $\mathbf{W}$, one can obtain arbitrarily high rate, i.e.,

$$\max_{\mathbf{K}_\mathrm{P}\in\mathcal{K}_\mathrm{P}} R_+(\mathbf{K}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) = \infty, \tag{27}$$

which contradicts that $\bar{\mathbf{K}}_\Phi$ is a saddle point solution.

Now observe that $\bar{\mathbf{\Omega}}$ in (25) is a solution to the following minimization problem,

$$\begin{aligned} &\min_{\mathbf{\Omega}\in\mathcal{K}_\Omega} R_\Omega(\mathbf{\Omega}), \\ &R_\Omega(\mathbf{\Omega}) = \log\frac{\det(\mathbf{G}\bar{\mathbf{K}}_\mathrm{P}\mathbf{G}^\dagger + \mathbf{\Omega})}{\det(\mathbf{\Omega})}, \\ &\mathcal{K}_\Omega = \left\{\mathbf{\Omega}\;\middle|\; \mathbf{W}\mathbf{\Omega}\mathbf{W}^\dagger = \begin{bmatrix}\mathbf{I}_{n_r} & \mathbf{\Phi}\\ \mathbf{\Phi}^\dagger & \mathbf{I}_{n_e}\end{bmatrix} \succeq 0\right\}. \end{aligned} \tag{28}$$

Indeed $\bar{\mathbf{\Omega}}$ is a feasible point for (28). Also with $\mathbf{z}_\Omega \sim \mathcal{CN}(\mathbf{0}, \Omega)$, one can show that

$$R_\Omega(\mathbf{\Omega}) = R_+(\bar{\mathbf{K}}_\mathrm{P}, \mathbf{W}\Omega\mathbf{W}^\dagger) + \log\det(\mathbf{I} + \mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger), \tag{29}$$

from which the optimality of $\bar{\mathbf{\Omega}}$ readily follows. The optimality conditions associated with the minimization problem (28) give

$$\begin{aligned} &\bar{\mathbf{\Omega}}^{-1} - (\mathbf{G}\bar{\mathbf{K}}_\mathrm{P}\mathbf{G}^\dagger + \bar{\mathbf{\Omega}})^{-1} = \mathbf{W}^\dagger\mathbf{\Upsilon}\mathbf{W}, \\ &\Rightarrow \mathbf{G}\bar{\mathbf{K}}_\mathrm{P}\mathbf{G}^\dagger = \bar{\mathbf{\Omega}}\mathbf{W}^\dagger\mathbf{\Upsilon}\mathbf{W}(\bar{\mathbf{\Omega}} + \mathbf{G}\bar{\mathbf{K}}_\mathrm{P}\mathbf{G}^\dagger) \end{aligned} \tag{30}$$

where $\mathbf{\Upsilon}$ has the block diagonal form in (20). Multiplying the left and right and side of (30) with $\mathbf{W}$ and $\mathbf{W}^\dagger$ respectively and using (25) and (26) we have that

$$\mathbf{H}_\mathrm{t}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{t}^\dagger = \bar{\mathbf{K}}_\Phi\mathbf{\Upsilon}(\bar{\mathbf{K}}_\Phi + \mathbf{H}_\mathrm{t}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{t}^\dagger), \tag{31}$$

which coincides with (24).

C. Optimal Input Covariance Property

Given that $(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)$ is a saddle point solution in (2) we have from (11) that

$$\bar{\mathbf{K}}_\mathrm{P} \in \arg\max_{\mathbf{K}_\mathrm{P}\in\mathcal{K}_\mathrm{P}} R_+(\mathbf{K}_\mathrm{P}, \bar{\mathbf{K}}_\Phi). \tag{32}$$

We show that (32) in turn implies the following property.

Lemma 3: Suppose that $\bar{\mathbf{K}}_\mathrm{P} = \mathbf{S}\mathbf{S}^\dagger$, where $\mathbf{S}$ has a full column rank. Then provided $(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e}) \neq \mathbf{0}$, the matrix

$$\mathbf{M} = (\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})\mathbf{S} \tag{33}$$

has a full column rank, where $\bar{\mathbf{\Phi}}$ and $\bar{\mathbf{\Theta}}$ are defined via (15) and (18), respectively.

The rest of this subsection is devoted to the proof of Lemma 3, and accordingly we assume that the saddle point solution $(\bar{\mathbf{K}}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)$ satisfies $(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e}) \neq \mathbf{0}$. As with Lemma 2, the proof is most direct when $\bar{\mathbf{K}}_\Phi$ is non-singular. Hence we will treat this case first and consider the case when $\bar{\mathbf{K}}_\Phi$ is singular subsequently.

1) $\bar{\mathbf{K}}_\Phi$ is non-singular: In this case, we can write the optimality condition (32) as

$$\begin{aligned} \bar{\mathbf{K}}_\mathrm{P} &\in \arg\max_{\mathbf{K}_\mathrm{P}\in\mathcal{K}_\mathrm{P}} R_+(\mathbf{K}_\mathrm{P}, \bar{\mathbf{K}}_\Phi) \\ &= \arg\max_{\mathbf{K}_\mathrm{P}\in\mathcal{K}_\mathrm{P}} h(\mathbf{y}_\mathrm{r} \mid \mathbf{y}_\mathrm{e}) \\ &= \arg\max_{\mathbf{K}_\mathrm{P}\in\mathcal{K}_\mathrm{P}} h\left(\mathbf{y}_\mathrm{r} - \Theta(\mathbf{K}_\mathrm{P})\mathbf{y}_\mathrm{e}\right), \end{aligned} \tag{34}$$

where $\Theta(\mathbf{K}_\mathrm{P}) = (\mathbf{H}_\mathrm{r}\mathbf{K}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger + \bar{\mathbf{\Phi}})(\mathbf{H}_\mathrm{e}\mathbf{K}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger + \mathbf{I})^{-1}$ is the linear minimum mean squared estimation coefficient of $\mathbf{y}_\mathrm{r}$ given $\mathbf{y}_\mathrm{e}$. Instead of directly working with the optimality conditions associated with (34) we reformulate the problem as below.

Claim 2: Suppose that $\bar{\mathbf{K}}_\Phi \succ \mathbf{0}$ and define

$$\mathcal{H}(\mathbf{K}_\mathrm{P}) \triangleq h(\mathbf{y}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{y}_\mathrm{e}) = \log\det(\mathbf{\Gamma}(\mathbf{K}_\mathrm{P})), \tag{35}$$

where

$$\begin{aligned} \mathbf{\Gamma}(\mathbf{K}_\mathrm{P}) \triangleq \mathbf{I} + \bar{\mathbf{\Theta}}\bar{\mathbf{\Theta}}^\dagger - \bar{\mathbf{\Theta}}\bar{\mathbf{\Phi}}^\dagger - \bar{\mathbf{\Phi}}\bar{\mathbf{\Theta}}^\dagger + \\ (\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})\mathbf{K}_\mathrm{P}(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})^\dagger. \end{aligned} \tag{36}$$

Then,

$$\bar{\mathbf{K}}_\mathrm{P} \in \arg\max_{\mathbf{K}_\mathrm{P}\in\mathcal{K}_\mathrm{P}} \mathcal{H}(\mathbf{K}_\mathrm{P}). \tag{37}$$

Remark 1: The objective function in (37) is similar to the one in (34), but with $\bar{\mathbf{\Theta}}$ fixed, i.e., the variables $\bar{\mathbf{\Theta}}$ and $\mathbf{K}_\mathrm{P}$ are decoupled in (37). This key step enables us to work with the simpler objective function in (37) and complete the proof.

Proof: To establish (37) note that since $\mathcal{H}(\cdot)$ is a concave function in $\mathbf{K}_\mathrm{P} \in \mathcal{K}_\mathrm{P}$ and differentiable over $\mathcal{K}_\mathrm{P}$, the optimality conditions associated with the Lagrangian

$$\mathcal{L}_\Theta(\mathbf{K}_\mathrm{P}, \lambda, \mathbf{\Psi}) = \mathcal{H}(\mathbf{K}_\mathrm{P}) + \mathrm{tr}(\mathbf{\Psi}\mathbf{K}_\mathrm{P}) - \lambda(\mathrm{tr}(\mathbf{K}_\mathrm{P}) - P), \tag{38}$$

are both necessary and sufficient. Thus $\mathbf{K}_\mathrm{P}$ is an optimal solution to (37) if and only if there exists a $\lambda \geq 0$ and $\mathbf{\Psi} \succeq 0$ such that

$$\begin{aligned} &(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})^\dagger[\mathbf{\Gamma}(\mathbf{K}_\mathrm{P})]^{-1}(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e}) + \mathbf{\Psi} = \lambda\mathbf{I}, \\ &\mathrm{tr}(\mathbf{\Psi}\mathbf{K}_\mathrm{P}) = 0, \quad \lambda(\mathrm{tr}(\mathbf{K}_\mathrm{P}) - P) = 0, \end{aligned} \tag{39}$$

where $\mathbf{\Gamma}(\cdot)$ is defined in (36). These parameters for $\bar{\mathbf{K}}_\mathrm{P}$ are obtained from the optimality conditions associated with (32).

Since $R_+(\mathbf{K}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)$ is differentiable at each $\mathbf{K}_P \in \mathcal{K}_\mathrm{P}$ whenever $\bar{\mathbf{K}}_\Phi \succ \mathbf{0}$, $\bar{\mathbf{K}}_\mathrm{P}$ satisfies the associated KKT conditions — there exists a $\lambda_0 \geq 0$ and $\mathbf{\Psi}_0 \succeq 0$ such that

$$\begin{aligned}\nabla_{\mathbf{K}_\mathrm{P}} R(\mathbf{K}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)\Big|_{\bar{\mathbf{K}}_\mathrm{P}} + \mathbf{\Psi}_0 = \lambda_0 \mathbf{I} \\ \lambda_0(\mathrm{tr}(\bar{\mathbf{K}}_\mathrm{P}) - P) = 0, \quad \mathrm{tr}(\mathbf{\Psi}_0 \bar{\mathbf{K}}_\mathrm{P}) = 0.\end{aligned} \tag{40}$$

We show in Appendix II that

$$\nabla_{\mathbf{K}_\mathrm{P}} R(\mathbf{K}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)\Big|_{\bar{\mathbf{K}}_\mathrm{P}} = (\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})^\dagger [\mathbf{\Lambda}(\bar{\mathbf{K}}_\mathrm{P})]^{-1} (\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e}), \tag{41}$$

where $\mathbf{\Lambda}(\cdot)$, defined in (13), satisfies satisfies[4] $\mathbf{\Lambda}(\bar{\mathbf{K}}_\mathrm{P}) = \mathbf{\Gamma}(\bar{\mathbf{K}}_\mathrm{P})$. Hence the first condition in (40) reduces to

$$(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})^\dagger [\mathbf{\Gamma}(\bar{\mathbf{K}}_\mathrm{P})]^{-1} (\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e}) + \mathbf{\Psi}_0 = \lambda_0 \mathbf{I}. \tag{42}$$

Comparing (40) and (42) with (39), we note that $(\bar{\mathbf{K}}_\mathrm{P}, \lambda_0, \mathbf{\Psi}_0)$ satisfy the conditions in (39), thus establishing (37). ■

Claim 3: Suppose that $\bar{\mathbf{K}}_\Phi \succ \mathbf{0}$ and $\hat{\mathbf{K}}_\mathrm{P}$ be any optimal solution to

$$\hat{\mathbf{K}}_\mathrm{P} \in \arg\max_{\mathcal{K}_\mathrm{P}} \mathcal{H}(\mathbf{K}_\mathrm{P}). \tag{43}$$

Suppose that $\mathbf{S}_\mathrm{P}$ is a matrix with a full column rank such that

$$\hat{\mathbf{K}}_\mathrm{P} = \mathbf{S}_\mathrm{P}\mathbf{S}_\mathrm{P}^\dagger \tag{44}$$

then $(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})\mathbf{S}_\mathrm{P}$ has a full column rank.

Note that the claim in Lemma 3 follows from Claim 2 and Claim 3. It remains to prove Claim 3.

Proof: The proof is based on the so called water-filling principle [25]. From (43), we have

$$\hat{\mathbf{K}}_\mathrm{P} = \arg\max_{\mathbf{K}_\mathrm{P} \in \mathcal{K}_\mathrm{P}} \log\det(\mathbf{I} + \mathbf{J}^{-\frac{1}{2}} (\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e}) \mathbf{K}_\mathrm{P} (\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})^\dagger \mathbf{J}^{-\frac{1}{2}}), \tag{45}$$

where $\mathbf{J} \triangleq \mathbf{I} + \bar{\mathbf{\Theta}}\bar{\mathbf{\Theta}}^\dagger - \bar{\mathbf{\Theta}}\bar{\mathbf{\Phi}}^\dagger - \bar{\mathbf{\Phi}}\bar{\mathbf{\Theta}}^\dagger \succ \mathbf{0}$, i.e., $\hat{\mathbf{K}}_\mathrm{P}$ is an optimal input covariance for a MIMO channel with white noise and matrix $\mathbf{H}_\mathrm{eff} \triangleq \mathbf{J}^{-\frac{1}{2}}(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})$. We can now consider the usual water-filling properties associated with $\hat{\mathbf{K}}_\mathrm{P}$ to establish that $(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})\mathbf{S}_\mathrm{P}$ has a full column rank.

Let $\mathrm{rank}(\mathbf{H}_\mathrm{eff}) = \nu$ and let us denote the non-zero singular values (in non-increasing order) by $\sigma_1, \sigma_2, \ldots, \sigma_\nu$. Let $\mathbf{\Sigma}_0 = \mathrm{diag}(\sigma_1, \ldots, \sigma_\nu)$, and

$$\mathbf{\Sigma} = \begin{array}{c} \\ \nu \\ n_\mathrm{r}-\nu \end{array} \begin{array}{c} \begin{array}{cc} \nu & n_\mathrm{t}-\nu \end{array} \\ \begin{bmatrix} \mathbf{\Sigma}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \end{array}, \tag{46}$$

be such that

$$\mathbf{H}_\mathrm{eff} = \mathbf{A}\mathbf{\Sigma}\mathbf{B}^\dagger = \mathbf{A}_1\mathbf{\Sigma}_0\mathbf{B}_1^\dagger, \tag{47}$$

[4] To verify this relation, note that $\mathbf{\Gamma}(\mathbf{K}_\mathrm{P})$ is the variance of $\mathbf{y}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{y}_\mathrm{e}$. When $\mathbf{K}_\mathrm{P} = \bar{\mathbf{K}}_\mathrm{P}$, note that $\bar{\mathbf{\Theta}}\mathbf{y}_\mathrm{e}$ is the MMSE estimate of $\mathbf{y}_\mathrm{r}$ given $\mathbf{y}_\mathrm{e}$ and $\mathbf{\Gamma}(\mathbf{K}_\mathrm{P})$ is the associated MMSE estimation error.

is the singular value decomposition of $\mathbf{H}_\mathrm{eff}$ where $\mathbf{A}$ and $\mathbf{B}$ are unitary matrices in $\mathbb{C}^{n_\mathrm{r} \times n_\mathrm{r}}$ and $\mathbb{C}^{n_\mathrm{t} \times n_\mathrm{t}}$ and

$$\mathbf{A} = \begin{array}{c} \begin{array}{cc} \nu & n_\mathrm{r}-\nu \end{array} \\ \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \end{array}, \quad \mathbf{B} = \begin{array}{c} \begin{array}{cc} \nu & n_\mathrm{t}-\nu \end{array} \\ \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2 \end{bmatrix} \end{array}. \tag{48}$$

From (45) we have that

$$\begin{aligned}\hat{\mathbf{K}}_\mathrm{P} &\in \arg\max_{\mathcal{K}_P} \log\det(\mathbf{I} + \mathbf{H}_\mathrm{eff}\mathbf{K}_\mathrm{P}\mathbf{H}_\mathrm{eff}^\dagger) \\ &= \arg\max_{\mathcal{K}_P} \log\det(\mathbf{I} + \mathbf{A}\mathbf{\Sigma}\mathbf{B}^\dagger\mathbf{K}_\mathrm{P}\mathbf{B}\mathbf{\Sigma}^\dagger\mathbf{A}^\dagger) \\ &= \arg\max_{\mathcal{K}_P} \log\det(\mathbf{I} + \mathbf{B}^\dagger\mathbf{K}_\mathrm{P}\mathbf{B}\mathbf{\Sigma}^\dagger\mathbf{\Sigma})\end{aligned} \tag{49}$$

Since $\mathbf{B}$ is unitary, we have that $\mathbf{B}^\dagger\mathbf{K}_\mathrm{P}\mathbf{B} \in \mathcal{K}_P$ and hence it follows from (49) that

$$\mathbf{F} \triangleq \mathbf{B}^\dagger\hat{\mathbf{K}}_\mathrm{P}\mathbf{B} \in \arg\max_{\mathcal{K}_P} \log\det(\mathbf{I} + \mathbf{K}_\mathrm{P}\mathbf{\Sigma}^\dagger\mathbf{\Sigma}). \tag{50}$$

We now show that any such $\mathbf{F}$ is diagonal and $\mathbf{F}_{ii} = 0$ for $i > \nu$. From the Hadamard inequality [25, Section 16.8], we have that

$$\log\det(\mathbf{I} + \mathbf{F}\mathbf{\Sigma}^\dagger\mathbf{\Sigma}) \leq \sum_{i=1}^{n_\mathrm{t}} \log(1 + \mathbf{F}_{ii}\sigma_i^2) = \sum_{i=1}^{\nu} \log(1 + \mathbf{F}_{ii}\sigma_i^2), \tag{51}$$

with equality if and only if the matrix $\mathbf{F}\mathbf{\Sigma}^\dagger\mathbf{\Sigma}$ is a diagonal matrix. We now show that any optimal $\mathbf{F}$ in (50) has the form

$$\mathbf{F} = \begin{array}{c} \\ \nu \\ n_\mathrm{t}-\nu \end{array} \begin{array}{c} \begin{array}{cc} \nu & n_\mathrm{t}-\nu \end{array} \\ \begin{bmatrix} \mathbf{F}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \end{array} \tag{52}$$

where $\mathbf{F}_0$ is a diagonal matrix. Clearly any optimal $\mathbf{F}$ attains the upper bound in (51), hence it follows that (1) $\sum_{i=1}^{\nu} \mathbf{F}_{ii} = P$, and $\mathbf{F}_{ii} = 0$ for $i > \nu$ and (2) $\mathbf{F}\mathbf{\Sigma}^\dagger\mathbf{\Sigma}$ is a diagonal matrix. The first condition, together with the fact that $\mathbf{F} \succeq \mathbf{0}$ imples that the lower diagonal matrix in (52) is zero, while the second condition implies that the off-diagonal matrices in (52) are zero and that $\mathbf{F}_0$ is diagonal.

From (50), we have that

$$\hat{\mathbf{K}}_\mathrm{P} = \mathbf{B}\mathbf{F}\mathbf{B}^\dagger = \mathbf{B}_1\mathbf{F}_0\mathbf{B}_1^\dagger \tag{53}$$

and hence for any $\mathbf{S}_\mathrm{P}$ that has a full column rank and satisfies (44), we have

$$\mathrm{col}(\mathbf{S}_\mathrm{P}) \subseteq \mathrm{col}(\mathbf{B}_1) = \mathrm{Null}^\perp(\mathbf{H}_\mathrm{eff}) = \mathrm{Null}^\perp(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e}),$$

which implies that $(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})\mathbf{S}_\mathrm{P}$ has a full column rank. ■

2) $\bar{\mathbf{K}}_\Phi$ is singular: The case when $\bar{\mathbf{K}}_\Phi$ is singular can be handled by considering an appropriately reduced channel matrix. In this case $\bar{\mathbf{\Phi}}$ has $d \geq 1$ singular values equal to unity and hence we can express its SVD as

$$\bar{\mathbf{\Phi}} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Delta} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^\dagger \\ \mathbf{V}_2^\dagger \end{bmatrix} \tag{54}$$

where $\sigma_\mathrm{max}(\mathbf{\Delta}) < 1$.

First we obtain some conditions that are satisfied when the saddle point noise covariance is singular.

Claim 4: Suppose that $(\bar{\mathbf{K}}_{\rm P}, \bar{\mathbf{K}}_{\Phi})$ is a saddle point solution to the minimax problem in (2) and the singular value decomposition of $\bar{\boldsymbol{\Phi}}$ is given as in (54). Then we have that

$$\mathbf{U}_1^\dagger \mathbf{z}_{\rm r} \overset{\rm a.s.}{=} \mathbf{V}_1^\dagger \mathbf{z}_{\rm e} \tag{55a}$$

$$\mathbf{U}_1^\dagger \mathbf{H}_{\rm r}, = \mathbf{V}_1^\dagger \mathbf{H}_{\rm e}, \tag{55b}$$

$$R_+(\mathbf{K}_{\rm P}, \bar{\mathbf{K}}_{\Phi}) = I(\mathbf{x}; \mathbf{U}_2^\dagger \mathbf{y}_{\rm r} \mid \mathbf{y}_{\rm e}), \quad \forall\, \mathbf{K}_{\rm P} \in \mathcal{K}_{\rm P}. \tag{55c}$$

Proof: To establish (55a), we simply note that

$$E[\mathbf{U}_1^\dagger \mathbf{z}_{\rm r} \mathbf{z}_{\rm e}^\dagger \mathbf{V}_1] = \mathbf{U}_1^\dagger \bar{\boldsymbol{\Phi}} \mathbf{V}_1 = \mathbf{I},$$

i.e., the Gaussian random variables $\mathbf{U}_1^\dagger \mathbf{z}_{\rm r}$ and $\mathbf{V}_1^\dagger \mathbf{z}_{\rm e}$ are perfectly correlated. Next note that

$$\begin{aligned} R_+(\mathbf{K}_{\rm P}, \bar{\mathbf{K}}_{\Phi}) &= I(\mathbf{x}; \mathbf{y}_{\rm r} | \mathbf{y}_{\rm e}) \\ &= I(\mathbf{x}; \mathbf{U}_1^\dagger \mathbf{y}_{\rm r}, \mathbf{U}_2^\dagger \mathbf{y}_{\rm r} | \mathbf{y}_{\rm e}) \\ &= I(\mathbf{x}; \mathbf{U}_2^\dagger \mathbf{y}_{\rm r}, \mathbf{U}_1^\dagger \mathbf{y}_{\rm r} - \mathbf{V}_1^\dagger \mathbf{y}_{\rm e} | \mathbf{y}_{\rm e}) \\ &= I(\mathbf{x}; \mathbf{U}_2^\dagger \mathbf{y}_{\rm r}, \mathbf{U}_1^\dagger \mathbf{H}_{\rm r} \mathbf{x} - \mathbf{V}_1^\dagger \mathbf{H}_{\rm e} \mathbf{x} | \mathbf{y}_{\rm e}). \end{aligned}$$

Since $\bar{\mathbf{K}}_{\Phi}$ is a saddle point solution, we must have $\max_{\mathbf{K}_{\rm P}} R_+(\mathbf{K}_{\rm P}, \bar{\mathbf{K}}_{\Phi}) < \infty$ and hence $\mathbf{U}_1^\dagger \mathbf{H}_{\rm r} = \mathbf{V}_1^\dagger \mathbf{H}_{\rm e}$, and $R_+(\bar{\mathbf{K}}_{\rm P}, \bar{\mathbf{K}}_{\Phi}) = I(\mathbf{x}; \mathbf{U}_2^\dagger \mathbf{y}_{\rm r} \mid \mathbf{y}_{\rm e})$, establishing (55b) and (55c). ■

Thus with $\hat{\mathbf{H}}_{\rm r} = \mathbf{U}_2^\dagger \mathbf{H}_{\rm r}$, and $\hat{\mathbf{z}}_{\rm r} = \mathbf{U}_2^\dagger \mathbf{z}_{\rm r}$ and

$$\hat{\mathbf{y}}_{\rm r} = \mathbf{U}_2^\dagger \mathbf{y}_{\rm r} = \hat{\mathbf{H}}_{\rm r} \mathbf{x} + \hat{\mathbf{z}}_{\rm r}, \tag{56}$$

we have from (55c), that

$$\bar{\mathbf{K}}_{\rm P} \in \arg\max_{\mathcal{K}_{\rm P}} I(\mathbf{x}; \hat{\mathbf{y}}_{\rm r} \mid \mathbf{y}_{\rm e}). \tag{57}$$

Since $\hat{\boldsymbol{\Phi}} = E[\hat{\mathbf{z}}_{\rm r} \mathbf{z}_{\rm e}^\dagger] \prec \mathbf{I}$, it follows from (57) and Claim 2 that

$$\bar{\mathbf{K}}_{\rm P} \in \arg\max_{\mathcal{K}_P} \hat{\mathcal{H}}(\mathbf{K}_{\rm P}) \tag{58}$$

where

$$\begin{aligned} &\hat{\mathcal{H}}(\mathbf{K}_{\rm P}) = h(\hat{\mathbf{y}}_{\rm r} - \hat{\boldsymbol{\Theta}} \mathbf{y}_{\rm e}), \\ &\qquad \hat{\boldsymbol{\Theta}} = \mathbf{U}_2^\dagger (\mathbf{H}_{\rm r} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm e}^\dagger + \bar{\boldsymbol{\Phi}})(\mathbf{I} + \mathbf{H}_{\rm e} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm e}^\dagger)^{-1}. \end{aligned}$$

Along the lines of Claim 3 we then have that

$$(\hat{\mathbf{H}}_{\rm r} - \hat{\boldsymbol{\Theta}} \mathbf{H}_{\rm e}) \mathbf{S} = \mathbf{U}_2^\dagger (\mathbf{H}_{\rm r} - \bar{\boldsymbol{\Theta}} \mathbf{H}_{\rm e}) \mathbf{S}$$

has a full column rank, which in turn establishes that $(\mathbf{H}_{\rm r} - \bar{\boldsymbol{\Theta}} \mathbf{H}_{\rm e}) \mathbf{S}$ has a full column rank.

D. Saddle Value

We use the results from Lemma 2 and Lemma 3 to establish (7). To invoke Lemma 3, we will first assume that the saddle point solution $(\bar{\mathbf{K}}_{\rm P}, \bar{\mathbf{K}}_{\Phi})$ is such that $\mathbf{H}_{\rm r} - \bar{\boldsymbol{\Theta}} \mathbf{H}_{\rm e} \neq \mathbf{0}$ and treat the case $\mathbf{H}_{\rm r} - \bar{\boldsymbol{\Theta}} \mathbf{H}_{\rm e} = \mathbf{0}$ subsequently. Note that from Lemma 2 we have that

$$(\mathbf{H}_{\rm r} - \bar{\boldsymbol{\Theta}} \mathbf{H}_{\rm e}) \mathbf{S} \mathbf{S}^\dagger (\bar{\boldsymbol{\Phi}}^\dagger \mathbf{H}_{\rm r} - \mathbf{H}_{\rm e})^\dagger = \mathbf{0}, \tag{59}$$

and since $\mathbf{M} = (\mathbf{H}_{\rm r} - \bar{\boldsymbol{\Theta}} \mathbf{H}_{\rm e}) \mathbf{S}$ has a full column rank, (59) reduces to

$$\bar{\boldsymbol{\Phi}}^\dagger \mathbf{H}_{\rm r} \mathbf{S} = \mathbf{H}_{\rm e} \mathbf{S}. \tag{60}$$

The difference between the upper and lower bounds is given by

$$\begin{aligned} \boldsymbol{\Delta} R &= R_+(\bar{\mathbf{K}}_{\rm P}, \bar{\mathbf{K}}_{\Phi}) - R_-(\bar{\mathbf{K}}_{\rm P}) \\ &= I(\mathbf{x}; \mathbf{y}_{\rm r} \mid \mathbf{y}_{\rm e}) - [I(\mathbf{x}; \mathbf{y}_{\rm r}) - I(\mathbf{x}; \mathbf{y}_{\rm e})] \\ &= I(\mathbf{x}; \mathbf{y}_{\rm e} \mid \mathbf{y}_{\rm r}). \end{aligned} \tag{61}$$

If $\bar{\mathbf{K}}_{\Phi} \succ \mathbf{0}$, then $I(\mathbf{x}; \mathbf{y}_{\rm e} \mid \mathbf{y}_{\rm r}) = h(\mathbf{y}_{\rm e} \mid \mathbf{y}_{\rm r}) - h(\mathbf{z}_{\rm e} \mid \mathbf{z}_{\rm r})$ and

$$\begin{aligned} &h(\mathbf{y}_{\rm e} \mid \mathbf{y}_{\rm r}) \\ &= \log\det(\mathbf{I} + \mathbf{H}_{\rm e} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm e}^\dagger - \\ &\qquad (\mathbf{H}_{\rm e} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm r}^\dagger + \bar{\boldsymbol{\Phi}}^\dagger)(\mathbf{H}_{\rm r} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm r}^\dagger + \mathbf{I})^{-1}(\mathbf{H}_{\rm r} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm e}^\dagger + \bar{\boldsymbol{\Phi}})) \\ &= \log\det(\mathbf{I} + \mathbf{H}_{\rm e} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm e}^\dagger - \bar{\boldsymbol{\Phi}}^\dagger (\mathbf{H}_{\rm r} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm r}^\dagger + \mathbf{I}) \bar{\boldsymbol{\Phi}}) \\ &= \log\det(\mathbf{I} - \bar{\boldsymbol{\Phi}}^\dagger \bar{\boldsymbol{\Phi}}) = h(\mathbf{z}_{\rm e} \mid \mathbf{z}_{\rm r}), \end{aligned} \tag{62}$$

where we have used the relation (60) in simplifying (62). This shows that the difference ΔR in (61) is zero, thus establishing (7) whenever $\bar{\mathbf{K}}_{\Phi}$ is non-singular.

To establish the result when $\bar{\mathbf{K}}_{\Phi}$ is singular, note that from (55a) and (55b) in Claim 4,

$$\begin{aligned} \Delta R &= I(\mathbf{x}; \mathbf{y}_{\rm e} \mid \mathbf{y}_{\rm r}), \\ &= I(\mathbf{x}; \mathbf{V}_2^\dagger \mathbf{y}_{\rm e} \mid \mathbf{y}_{\rm r}), \end{aligned} \tag{63}$$

which is zero as shown below.

$$\begin{aligned} &h(\mathbf{V}_2^\dagger \mathbf{y}_{\rm e} \mid \mathbf{y}_{\rm r}) \\ &= \log\det(\mathbf{I} + \mathbf{V}_2^\dagger \mathbf{H}_{\rm e} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm e}^\dagger \mathbf{V}_2 - (\mathbf{V}_2^\dagger \mathbf{H}_{\rm e} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm r}^\dagger + \boldsymbol{\Delta}^\dagger \mathbf{U}_2^\dagger) \\ &\qquad (\mathbf{I} + \mathbf{H}_{\rm r} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm r}^\dagger)^{-1} (\mathbf{H}_{\rm r} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm e}^\dagger \mathbf{V}_2 + \mathbf{U}_2 \boldsymbol{\Delta})) \quad (64) \\ &= \log\det(\mathbf{I} + \boldsymbol{\Delta}^\dagger \mathbf{U}_2^\dagger \mathbf{H}_{\rm r} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm r}^\dagger \mathbf{U}_2 \boldsymbol{\Delta} \\ &\qquad - \boldsymbol{\Delta}^\dagger \mathbf{U}_2^\dagger (\mathbf{I} + \mathbf{H}_{\rm r} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm r}^\dagger) \mathbf{U}_2 \boldsymbol{\Delta}) \\ &= \log\det(\mathbf{I} - \boldsymbol{\Delta}^\dagger \boldsymbol{\Delta}) \\ &= h(\mathbf{V}_2^\dagger \mathbf{z}_{\rm e} \mid \mathbf{U}_2^\dagger \mathbf{z}_{\rm r}) = h(\mathbf{V}_2^\dagger \mathbf{z}_{\rm e} \mid \mathbf{z}_{\rm r}), \end{aligned} \tag{65}$$

where we have used from (60) that

$$\mathbf{V}_2^\dagger \bar{\boldsymbol{\Phi}}^\dagger \mathbf{H}_{\rm r} \mathbf{S} = \mathbf{V}_2^\dagger \mathbf{H}_{\rm e} \mathbf{S} \Rightarrow \boldsymbol{\Delta}^\dagger \mathbf{U}_2^\dagger \mathbf{H}_{\rm r} \mathbf{S} = \mathbf{V}_2^\dagger \mathbf{H}_{\rm e} \mathbf{S},$$

in simplifying (64) and the equality in (65) follows from the fact that $\mathbf{U}_1^\dagger \mathbf{z}_{\rm r}$ is independent of $(\mathbf{U}_2^\dagger \mathbf{z}_{\rm r}, \mathbf{V}_2^\dagger \mathbf{z}_{\rm e})$. This establishes (7) when $\bar{\mathbf{K}}_{\Phi}$ is singular.

It remains to consider the case when the saddle point solution $(\bar{\mathbf{K}}_{\rm P}, \bar{\mathbf{K}}_{\Phi})$ is such that

$$\bar{\boldsymbol{\Theta}} \mathbf{H}_{\rm e} = \mathbf{H}_{\rm r}. \tag{66}$$

In this case, we show that the saddle value and hence the capacity is zero. From (18), $\bar{\boldsymbol{\Theta}} = (\bar{\boldsymbol{\Phi}} + \mathbf{H}_{\rm r} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm e}^\dagger)(\mathbf{I} + \mathbf{H}_{\rm e} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm e}^\dagger)^{-1}$, hence we have

$$\bar{\boldsymbol{\Theta}} + \bar{\boldsymbol{\Theta}} \mathbf{H}_{\rm e} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm e}^\dagger = \bar{\boldsymbol{\Phi}} + \mathbf{H}_{\rm r} \bar{\mathbf{K}}_{\rm P} \mathbf{H}_{\rm e}^\dagger. \tag{67}$$

Substituting (66) in (67), we have that $\bar{\boldsymbol{\Phi}} = \bar{\boldsymbol{\Theta}}$, and using this relation it can be verified that $R_+(\bar{\mathbf{K}}_{\rm P}, \bar{\mathbf{K}}_{\Phi}) = \mathbf{0}$. This completes the proof of Theorem 1.

V. Zero-Capacity Condition and Scaling Laws

The conditions on $\mathbf{H}_\mathrm{r}$ and $\mathbf{H}_\mathrm{e}$ for which the secrecy capacity is zero have a simple form.

Lemma 4: The secrecy capacity of the MIMOME channel is zero if and only if

$$\sigma_{\max}(\mathbf{H}_\mathrm{r}, \mathbf{H}_\mathrm{e}) \triangleq \sup_{\mathbf{v}\in\mathbb{C}^{n_\mathrm{t}}} \frac{||\mathbf{H}_\mathrm{r}\mathbf{v}||}{||\mathbf{H}_\mathrm{e}\mathbf{v}||} \leq 1. \quad (68)$$

We omit the proof of this condition due to space constraints. The quantity $\sigma_{\max}(\mathbf{H}_\mathrm{r}, \mathbf{H}_\mathrm{e})$ is the largest generalized singular value of the channel matrices [26]. Analysis of the zero-capacity condition in the limit of large number of antennas provides several useful insights we develop below.

For our analysis, we use the following convergence property of the largest generalized singular value for Gaussian matrices.

Fact 1 ([27], [28]): Suppose that $\mathbf{H}_\mathrm{r}$ and $\mathbf{H}_\mathrm{e}$ have i.i.d. $\mathcal{CN}(0,1)$ entries. Let $n_\mathrm{r}, n_\mathrm{e}, n_\mathrm{t} \to \infty$, while keeping $n_\mathrm{r}/n_\mathrm{e} = \gamma$ and $n_\mathrm{t}/n_\mathrm{e} = \beta$ fixed. If $\beta < 1$, then the largest generalized singular value of $(\mathsf{H}_\mathrm{r}, \mathsf{H}_\mathrm{e})$ converges almost surely to

$$\sigma_{\max}(\mathsf{H}_\mathrm{r}, \mathsf{H}_\mathrm{e}) \overset{\text{a.s.}}{\to} \gamma \left[\frac{1+\sqrt{1-(1-\beta)\left(1-\frac{\beta}{\gamma}\right)}}{1-\beta}\right]^2. \quad (69)$$

By combining Lemma 4 and Fact 1, one can deduce the following condition for the zero-capacity condition.

Corollary 1: Suppose that H_r and H_e have i.i.d. $CN(0,1)$ entries. Suppose that $n_\mathrm{r}, n_\mathrm{e}, n_\mathrm{t} \to \infty$, while keeping $n_\mathrm{r}/n_\mathrm{e} = \gamma$ and $n_\mathrm{t}/n_\mathrm{e} = \beta$ fixed. The secrecy capacity[5] $C(\mathsf{H}_\mathrm{r}, \mathsf{H}_\mathrm{e})$ converges almost surely to zero if and only if $0 \leq \beta \leq 1/2$, $0 \leq \gamma \leq 1$, and

$$\gamma \leq (1-\sqrt{2\beta})^2. \quad (70)$$

Figs. 2 and 3 provide further insight into the asymptotic analysis for the capacity achieving scheme. In Fig. 2, we show the values of (β, γ) where the secrecy rate is zero. If the eavesdropper increases its antennas at a sufficiently high rate so that the point (β, γ) lies below the solid curve, then secrecy capacity is zero. The MISOME case corresponds to the vertical intercept of this plot. The secrecy capacity is zero, if $\beta \leq 1/2$, i.e., the eavesdropper has at least twice the number of antennas as the sender. The single transmit antenna (SIMOME) case corresponds to the horizontal intercept. In this case the secrecy capacity is zero if $\gamma \leq 1$, i.e., the eavesdropper has more antennas than the receiver.

In Fig. 3, we consider the scenario where a total of $T \gg 1$ antennas are divided between the sender and the receiver. The horizontal axis plots the ratio $n_\mathrm{r}/n_\mathrm{t}$, while the vertical axis plots the minimum number of antennas at the eavesdropper (normalized by T) for the secrecy capacity to be zero. We note that the optimal allocation of antennas, that maximizes the number of eavesdropper antennas happens at $n_\mathrm{r}/n_\mathrm{t} = 1/2$. This can be explicitly obtained from the following minimization

$$\begin{aligned}&\text{minimize } \beta+\gamma \\ &\text{subject to, } \gamma \geq (1-\sqrt{2\beta})^2,\ \beta \geq 0,\ \gamma \geq 0.\end{aligned} \quad (71)$$

The optimal solution can be easily verified to be $(\beta^*, \gamma^*) = (2/9, 1/9)$. In this case, the eavesdropper needs $\approx 3T$ antennas for the secrecy capacity to be zero. We remark that the objective function in (71) is not sensitive to variations in the optimal solution. If fact even if we allocate equal number of antennas to the sender and the receiver, the eavesdropper needs $\frac{(3+2\sqrt{2})}{2}T \approx 2.9142 \times T$ antennas for the secrecy capacity to be zero.

Acknowledgement

Ami Wiesel provided a numerical optimizer to evaluate the saddle point expression in Theorem 1.

Appendix I
Least Favorable Noise Property

Substituting for $\bar{\mathbf{K}}_\Phi$ and $\mathbf{H}_\mathrm{t}$ in (24) and carrying out the block matrix multiplication gives

$$\begin{aligned}
\mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger &= \mathbf{\Upsilon}_1(\mathbf{I}+\mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger) + \bar{\mathbf{\Phi}}\mathbf{\Upsilon}_2(\bar{\mathbf{\Phi}}^\dagger + \mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger)\\
\mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger &= \mathbf{\Upsilon}_1(\bar{\mathbf{\Phi}}+\mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger) + \bar{\mathbf{\Phi}}\mathbf{\Upsilon}_2(\mathbf{I} + \mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger)\\
\mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger &= \bar{\mathbf{\Phi}}^\dagger\mathbf{\Upsilon}_1(\mathbf{I}+\mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger) + \mathbf{\Upsilon}_2(\bar{\mathbf{\Phi}}^\dagger + \mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger)\\
\mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger &= \bar{\mathbf{\Phi}}^\dagger\mathbf{\Upsilon}_1(\bar{\mathbf{\Phi}}+\mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger) + \mathbf{\Upsilon}_2(\mathbf{I} + \mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger).
\end{aligned} \quad (72)$$

Eliminating $\mathbf{\Upsilon}_1$ from the first and third equation above, we have

$$(\bar{\mathbf{\Phi}}^\dagger\mathbf{H}_\mathrm{r} - \mathbf{H}_\mathrm{e})\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger = (\bar{\mathbf{\Phi}}^\dagger\bar{\mathbf{\Phi}} - \mathbf{I})\mathbf{\Upsilon}_2(\bar{\mathbf{\Phi}}^\dagger + \mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger). \quad (73)$$

Similarly eliminating $\mathbf{\Upsilon}_1$ from the second and fourth equations in (72) we have

$$(\bar{\mathbf{\Phi}}^\dagger\mathbf{H}_\mathrm{r} - \mathbf{H}_\mathrm{e})\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger = (\bar{\mathbf{\Phi}}^\dagger\bar{\mathbf{\Phi}} - \mathbf{I})\mathbf{\Upsilon}_2(\mathbf{I} + \mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger). \quad (74)$$

Finally, eliminating $\mathbf{\Upsilon}_2$ from (73) and (74) we obtain (17).

Appendix II
KKT Condition

First note that,

$$\begin{aligned}&\nabla_{\mathbf{K}_\mathrm{P}} R_+(\mathbf{K}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)\\ &= \mathbf{H}_\mathrm{t}^\dagger(\mathbf{H}_\mathrm{t}\mathbf{K}_\mathrm{P}\mathbf{H}_\mathrm{t}^\dagger + \bar{\mathbf{K}}_\Phi)^{-1}\mathbf{H}_\mathrm{t} - \mathbf{H}_\mathrm{e}^\dagger(\mathbf{I}+\mathbf{H}_\mathrm{e}\mathbf{K}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger)^{-1}\mathbf{H}_\mathrm{e}.\end{aligned} \quad (75)$$

Substituting for $\mathbf{H}_\mathrm{t}$ and $\bar{\mathbf{K}}_\Phi$ from (23) and (15),

$$\begin{aligned}&(\bar{\mathbf{K}}_\Phi + \mathbf{H}_\mathrm{t}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{t}^\dagger)^{-1}\\
&= \begin{bmatrix}\mathbf{I}+\mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{r}^\dagger & \bar{\mathbf{\Phi}}+\mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger\\ \bar{\mathbf{\Phi}}^\dagger + \mathbf{H}_\mathrm{r}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger & \mathbf{I}+\mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger\end{bmatrix}^{-1}\\
&= \begin{bmatrix}\mathbf{\Lambda}^{-1} & -\mathbf{\Lambda}^{-1}\bar{\mathbf{\Theta}}\\ -\bar{\mathbf{\Theta}}^\dagger\mathbf{\Lambda}^{-1} & (\mathbf{I}+\mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e})^{-1}+\bar{\mathbf{\Theta}}^\dagger\mathbf{\Lambda}^{-1}\bar{\mathbf{\Theta}}\end{bmatrix}^{-1},\end{aligned}$$

[5] We assume that the channels are sampled once, then stay fixed for the entire period of transmission, and are revealed to all the terminals.

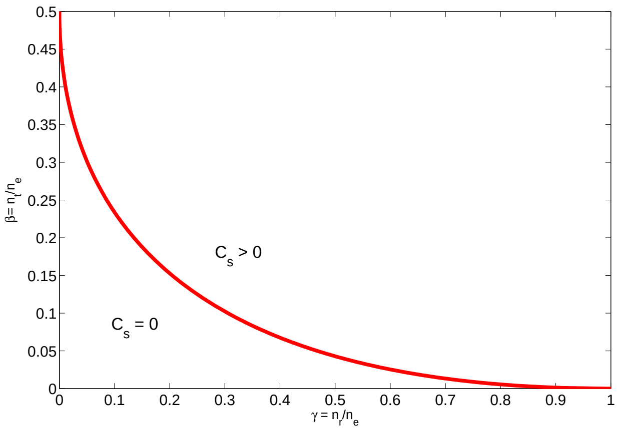

Fig. 2. Zero-capacity condition in the (β, γ) plane. The capacity is zero for any point below the curve, i.e., the eavesdropper has sufficiently many antennas to get non-vanishing fraction of the message, even when the sender and receiver fully exploit the knowledge of $\mathbf{H}_\mathrm{e}$.

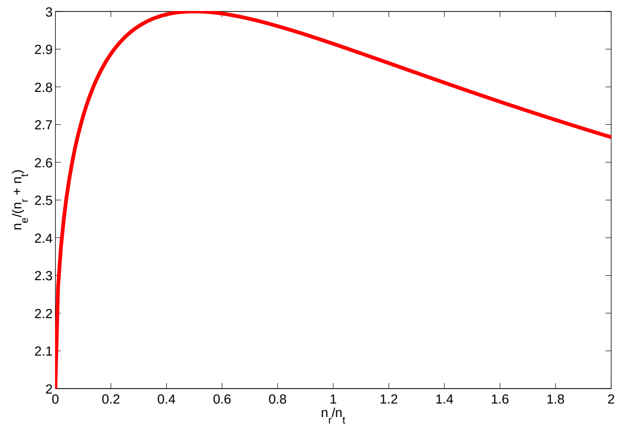

Fig. 3. The minimum number of eavesdropping antennas per sender plus receiver antenna for the secrecy capacity to be zero, plotted as a function of $n_\mathrm{r}/n_\mathrm{t}$.

where we have used the matrix inversion lemma (e.g., [29]), and $\mathbf{\Lambda} \triangleq \mathbf{\Lambda}(\bar{\mathbf{K}}_\mathrm{P})$ is defined in (13), and $\bar{\mathbf{\Theta}}$ is as defined in (18). Substituting into (75) and simplifying gives

$$\begin{aligned}
&\nabla_{\mathbf{K}_\mathrm{P}} R_+(\mathbf{K}_\mathrm{P}, \bar{\mathbf{K}}_\Phi)\Big|_{\bar{\mathbf{K}}_\mathrm{P}} \\
=&\mathbf{H}_\mathrm{t}^\dagger(\bar{\mathbf{K}}_\Phi + \mathbf{H}_\mathrm{t}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{t}^\dagger)^{-1}\mathbf{H}_\mathrm{t} - \mathbf{H}_\mathrm{e}^\dagger(\mathbf{I} + \mathbf{H}_\mathrm{e}\bar{\mathbf{K}}_\mathrm{P}\mathbf{H}_\mathrm{e}^\dagger)^{-1}\mathbf{H}_\mathrm{e} \\
&= (\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})^\dagger[\mathbf{\Lambda}(\bar{\mathbf{K}}_\mathrm{P})]^{-1}(\mathbf{H}_\mathrm{r} - \bar{\mathbf{\Theta}}\mathbf{H}_\mathrm{e})
\end{aligned}$$

as required.

References

[1] A. D. Wyner, "The wiretap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–87, 1975.

[2] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inform. Theory*, vol. 24, pp. 339–348, 1978.

[3] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wiretap channel," *IEEE Trans. Inform. Theory*, vol. 24, pp. 451–56, 1978.

[4] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Inform. Theory*, vol. 52, pp. 361–374, Feb. 2006.

[5] Y. Liang, H. V. Poor, and S. Shamai, "Secure communication over fading channels," *IEEE Trans. Inform. Theory*, submitted.

[6] Z. Li, R. Yates, and W. Trappe, "Secrecy capacity of independent parallel channels," in *Proc. Allerton Conf. Commun., Contr., Computing*, 2006.

[7] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure Broadcasting," *Submitted to IEEE Trans. Inform. Theory, Special Issue on Information Theoretic Security*, Feb. 2007.

[8] P. Gopala, L. Lai, and H. E. Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inform. Theory*, submitted, 2006.

[9] R. Negi and S. Goel, "Secret communication using artificial noise," in *Proc. Vehic. Tech. Conf.*, 2005.

[10] Z. Li, W. Trappe, and R. Yates, "Secret communication via multi-antenna transmission," in *Forty-First Annual Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, Mar. 2007.

[11] S. Shaifee and S. Ulukus, "Achievable rates in Gaussian MISO channels with secrecy constraints," in *Proc. Int. Symp. Inform. Theory*, June 2007.

[12] A. Khisti, G. W. Wornell, A. Wiesel, and Y. Eldar, "On the Gaussian MIMO wiretap channel," in *Proc. Int. Symp. Inform. Theory*, Nice, 2007.

[13] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas: The MISOME wiretap channel," *Submitted Aug. 2007, IEEE Trans. Inform. Theory, available online, http://arxiv.org/abs/0708.4219*.

[14] S. Ulukus, "Personal communication." 2007.

[15] S. Shafiee, N. Liu, and S. Ulukus, "Towards the secrecy capacity of the Gaussian MIMO wire-tap channel: The 2-2-1 channel," *IEEE Trans. Inform. Theory*, sept, submitted 2007.

[16] O. Frederique and B. Hassibi, "The secrecy capacity of the 2x2 MIMO wiretap channel," in *Proc. 45th Allerton Conf. on Communication, Control and Computing*, Montecillo, IL, 2007.

[17] R. Liu and V. Poor, "Multiple antenna secure broadcast over wireless networks," *http://arxiv.org/abs/0705.1183*, 2007.

[18] R. Liu, I. Maric, P. Spasojevic, and R. D. Yates, "Discrete memoryless interference and broadcast channels with confidential messages: Secrecy capacity regions," *IEEE Trans. Inform. Theory*, Feb. 2007, submitted, http://arxiv.org/abs/cs/0702099.

[19] N. Cai, "Private capacity of broadcast channels," *General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science*, vol. 4123, 2006.

[20] S. Wilks, *Mathematical Statistics*. John Wiley, 1962.

[21] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inform. Theory*, vol. IT-47, no. 7, pp. 3072–3081, 2001.

[22] R. Bhatia, *Positive Definite Matrices*. Princeton Press, 2007.

[23] D. P. Bertsekas, A. Nedic, and A. Ozdaglar, *Convex Analysis and Optimization*. Athena Scientific, 2003.

[24] R. A. Horn and I. Olkin, "When does $A^*A = B^*B$ and why does one want to know?" *The American Mathematical Monthly,*, vol. 103, pp. 470–482, 1996.

[25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.

[26] G. Golub and C. F. V. Loan, *Matrix Computations (3rd ed)*. Johns Hopkins University Press, 1996.

[27] J. W. Silverstein, "The limiting eigenvalue distribution of a multivariate F- matrix," *SIAM Journal on Mathematical Analysis*, vol. 16, pp. 641–646, 1985.

[28] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional random matrices," *Annals of Probability*, vol. 26, pp. 316–345, 1998.

[29] K. Petersen and M. Pedersen, "The Matrix Cookbook," September, 2007.