Secure Transmission With Multiple Antennas—Part II: The MIMOME Wiretap Channel

Ashish Khisti, *Member, IEEE*, and Gregory W. Wornell, *Fellow, IEEE*

***Abstract*—The capacity of the Gaussian wiretap channel model is analyzed when there are multiple antennas at the sender, intended receiver and eavesdropper. The associated channel matrices are fixed and known to all the terminals. A computable characterization of the secrecy capacity is established as the saddle point solution to a minimax problem. The converse is based on a Sato-type argument used in other broadcast settings, and the coding theorem is based on Gaussian wiretap codebooks.**

At high signal-to-noise ratio (SNR), the secrecy capacity is shown to be attained by simultaneously diagonalizing the channel matrices via the generalized singular value decomposition, and independently coding across the resulting parallel channels. The associated capacity is expressed in terms of the corresponding generalized singular values. It is shown that a semi-blind "masked" multi-input multi-output (MIMO) transmission strategy that sends information along directions in which there is gain to the intended receiver, and synthetic noise along directions in which there is not, can be arbitrarily far from capacity in this regime.

Necessary and sufficient conditions for the secrecy capacity to be zero are provided, which simplify in the limit of many antennas when the entries of the channel matrices are independent and identically distributed. The resulting scaling laws establish that to prevent secure communication, the eavesdropper needs three times as many antennas as the sender and intended receiver have jointly, and that the optimum division of antennas between sender and intended receiver is in the ratio of 2:1.

***Index Terms*—Broadcast channel, cryptography, MIMO wiretap channel, multiple antennas, secrecy capacity.**

I. Introduction

MULTIPLE antennas are a valuable resource in wireless communication. Over the last several years, there has been extensive activity in exploring the design, analysis, and implementation of wireless systems with multiple antennas, emphasizing their role in improving robustness and throughput. In this work, we develop aspects of the emerging role of multiple antennas in providing communication security at the physical layer.

Manuscript received August 13, 2008, revised July 01, 2010. Date of current version October 20, 2010. This work was supported in part by the National Science Foundation under Grant CCF-0515109. The material in this paper was presented in part at the Allerton Conference on Communications, Control, and Signal Processing, which was held in Monticello, IL, September 2007.

A. Khisti was with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. He is now with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON Canada M5S 3G4 (e-mail: akhisti@comm.utoronto.ca).

G. W. Wornell is with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: gww@mit.edu).

Communicated by H. Yamamoto, Associate Editor for Shannon Theory.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2010.2068852

The wiretap channel [1] is an information-theoretic model for physical-layer security. In the model, there are three terminals—a sender, an intended receiver, and an 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 broadcast channel is discrete and memoryless is developed in [2]. An explicit solution for the scalar Gaussian case is obtained in [3], where the optimality of Gaussian codebooks is established.

In this paper, we consider the case where there are multiple antennas at each of the three terminals, referring to it as the multi-input, multi-output, multi-eavesdropper (MIMOME) channel. In our model, the channel matrices are fixed and known to all three terminals. While the eavesdropper's channel being known to both the sender and the receiver in the problem formulation is a strong assumption, we remark in advance that the solution provides ultimate limits on secure transmission with multiple antennas, and thus serves as a starting point for other formulations. Further discussion of the modeling assumptions is provided in the companion paper [4] and the compound extension has been recently treated in [5].

The problem of evaluating the secrecy capacity of channels with multiple antennas has attracted increasing attention in recent years. As a starting point, for Gaussian models in which the channel matrices of intended receiver and eavesdropper are square and diagonal, the results in [6]–[9], which consider secure transmission over fading channels, can be applied. In particular, for this special case of independent parallel Gaussian subchannels, it follows that using independent Gaussian wiretap codebooks across the subchannels achieves capacity.

More generally, the MIMOME channel is a nondegraded broadcast channel to which the Csiszár–Körner capacity expression [2] applies in principle. However, computing the capacity directly from [2] appears difficult, as observed in, e.g., [10]–[13].

To the best of our knowledge, the first computable upper bound for the secrecy capacity of the Gaussian multi-antenna wiretap channel appears in [4], [14], which is used to establish the secrecy capacity in the special (MISOME) case that the intended receiver has a single antenna. This approach involves revealing the output of the eavesdropper's channel to the legitimate receiver to create a fictitious degraded broadcast channel, and results in a minimax expression for the upper bound, analogous to the technique of Sato [15] used to upper bound the sum-capacity of the multi-antenna broadcast channel; see, e.g., [16].

In [4], [14], this minimax upper bound is used to obtain a closed-form expression for the secrecy capacity in the MISOME

case. In addition, a number of insights are developed into the behavior of the secrecy capacity. In the high signal-to-noise ratio (SNR) regime, the simple masked beamforming scheme developed in [11] is shown to be near optimal. Also, the scaling behavior of the secrecy capacity in the limit of many antennas is studied.

We note that this upper bounding approach has been independently conceived by Ulukus *et al.* [17] and further applied to the case of two transmit antennas, two receive antennas, and a single eavesdropper antenna [18]. Subsequently, this minimax upper bound was shown to be tight for the MIMOME case in [19] and, independently, [20] (see also [21]). Both treatments start from the minimax upper bound of [4] and work with the optimality conditions to establish that the saddle value is achievable with the standard Gaussian wiretap code construction [2].

In some of the most recent work, [22] provides an alternative derivation of the MIMOME secrecy capacity using an approach based on channel-enhancement techniques introduced in [23]. The two approaches shed complementary insights into the problem. The minimax upper bounding approach in [19], [20] provides a computable characterization for the capacity expression and identifies a hidden convexity in optimizing the Csiszár–Körner expression with Gaussian inputs, whereas the channel enhancement approach does not. On the other hand the latter approach establishes the capacity given any covariance constraint on the input distribution, not just the sum-power constraint to which the minimax upper bounding approach has been limited. Yet another proof of the secrecy capacity appears in [37].

Finally, the diversity-multiplexing tradeoff of the multi-antenna wiretap channel has been recently studied in [24].

An outline of the paper is as follows. Section II summarizes some notational conventions for the paper. Section III describes the basic channel and system model, as well as a canonical decomposition of the channel in terms of its generalized singular values, which is used in some of the asymptotic analysis. Section IV summarizes the main results of the paper, and Sections V–VII provide the corresponding analysis. In particular, Section V develops the minimax characterization of the secrecy capacity, Section VI develops the high SNR analysis in terms of the generalized singular values, and Section VII develops the conditions under which the secrecy capacity is zero in the limit of many antennas. Finally, Section VIII contains some concluding remarks.

II. Notation

In terms of fonts, bold upper and lower case characters are used for matrices and vectors, respectively. Random variables are distinguished from their realizations by the use of san-serif fonts for the former and regular serifed fonts for the latter. Sets are denoted using caligraphic fonts. We generally reserve the symbols $I(\cdot)$ for mutual information, and $h(\cdot)$ for differential entropy, and all logarithms are base-2 unless otherwise indicated. In addition, $\mathcal{CN}(\mathbf{0},\mathbf{K})$ denotes a circularly-symmetrix complex-valued Gaussian random vector with covariance matrix $\mathbf{K}$.

The set of all n-dimensional complex-valued vectors is denoted by $\mathbb{C}^n$, and the set of $m \times n$-dimensional matrices is denoted using $\mathbb{C}^{m\times n}$. In addition, $\mathbf{I}$ denotes the identity matrix and $\mathbf{0}$ denotes the zero matrix. When the dimensions of these matrices is not clear from context, we will explicily indicate their size via subscripts; e.g., $\mathbf{0}_{n\times m}$ denotes an $n\times m$ zero matrix, $\mathbf{0}_n$ denotes a vector of zeros of length n, and $\mathbf{I}_n$ denotes an $n \times n$ identity matrix. We further use the notation $[\cdot]_{i:j}$ for $j \geq i$ to denote the subvector of its vector argument corresponding to indices $i, i+1, \ldots, j$. Likewise, $[\cdot]_{i:j,k:l}$ denotes the submatrix formed from rows i through j and columns k through l of its matrix argument.

Matrix transposition is denoted using the superscript $^{\mathrm{T}}$, the Hermitian (i.e., conjugate) transpose of a matrix is denoted using the superscript †, the Moore–Penrose pseudo-inverse is denoted by ‡, and the projection matrix onto the null space is denoted by $^{\sharp}$. In addition, $\mathrm{Null}\,(\cdot)$, $\mathrm{rank}\,(\cdot)$, and $\sigma_{\max}(\cdot)$ denote the null space, rank, and largest singular value, respectively, of their matrix arguments. Moreover, we say a matrix has full column-rank if its rank is equal to the number of columns, and the notation $\mathbf{A} \succ \mathbf{0}$ means that $\mathbf{A}$ is positive definite, with $\mathbf{A} \succeq \mathbf{0}$ likewise denoting positive semidefiniteness.

In other notation, $\dim(\cdot)$ denotes the dimension of its subspace argument, $\mathrm{span}\,(\cdot)$ denotes the subspace spanned by the collection of vectors that are its argument, $^{\perp}$ denotes the orthogonal complement of a subspace. Moreover, $\|\cdot\|$ denotes the usual Euclidean norm of a vector argument, $\mathrm{tr}\,(\cdot)$ and $\det(\cdot)$ denote the trace and determinant of a matrix, respectively, and $\mathrm{diag}\,(\cdot)$ denotes a diagonal matrix whose diagonal elements are given by its argument.

Finally, we use $\overset{\text{a.s.}}{=}$ and $\xrightarrow{\text{a.s.}}$ to denote almost-sure equality and convergence, respectively, and additionally use standard order notation. Specifically, $O(\epsilon)$ and $o(\epsilon)$ denote terms such that $O(\epsilon)/\epsilon < \infty$ and $o(\epsilon)/\epsilon \to 0$, respectively, in the associated limit, so that, e.g., $o(1)$ represents a vanishing term.

III. Channeland System Model

Using n_{t}, n_{r}, and n_{e} to denote the number of antennas at the sender, intended receiver, and eavesdropper, respectively, the received signals at the intended receiver and eavesdropper in the channel model of interest are, respectively

$$\begin{aligned}\mathsf{y}_{\mathrm{r}}(t) &= \mathbf{H}_{\mathrm{r}}\mathsf{x}(t) + \mathsf{z}_{\mathrm{r}}(t)\\ \mathsf{y}_{\mathrm{e}}(t) &= \mathbf{H}_{\mathrm{e}}\mathsf{x}(t) + \mathsf{z}_{\mathrm{e}}(t)\end{aligned}, \qquad t = 1,2,\ldots,n \tag{1}$$

where $\mathsf{x}(t)$ is the transmitted signal, 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 complex channel gain matrices, and where $\mathsf{z}_{\mathrm{r}}(t)$ and $\mathsf{z}_{\mathrm{e}}(t)$ are each independent and identically distributed (i.i.d.) noises whose samples are $\mathcal{CN}(\mathbf{0},\mathbf{I})$ random variables. The channel matrices are constant (over the transmission interval) and known to all the three terminals. Moreover, the channel input satisfies the power constraint

$$E\left[\frac{1}{n}\sum_{t=1}^{n}\|\mathsf{x}(t)\|^2\right] \leq P.$$

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

A. Channel Decomposition

For some of our analysis, it will be convenient to exploit the generalized singular value decomposition (GSVD) [25], [26] of the channel (1). To develop this decomposition, we first define the subspaces

$$\mathcal{S}_{\rm r} = \mathrm{Null}\,(\mathbf{H}_{\rm r})^{\perp} \cap \mathrm{Null}\,(\mathbf{H}_{\rm e}) \tag{2a}$$
$$\mathcal{S}_{\rm r,e} = \mathrm{Null}\,(\mathbf{H}_{\rm r})^{\perp} \cap \mathrm{Null}\,(\mathbf{H}_{\rm e})^{\perp} \tag{2b}$$
$$\mathcal{S}_{\rm e} = \mathrm{Null}\,(\mathbf{H}_{\rm r}) \cap \mathrm{Null}\,(\mathbf{H}_{\rm e})^{\perp} \tag{2c}$$
$$\mathcal{S}_{n} = \mathrm{Null}\,(\mathbf{H}_{\rm r}) \cap \mathrm{Null}\,(\mathbf{H}_{\rm e}) \tag{2d}$$

corresponding to classes of inputs that have nonzero gain to, respectively, the intended receiver only, both intended receiver and eavesdropper, the eavesdropper only, and neither. Letting

$$k \triangleq \mathrm{rank}\,(\mathbf{H}) \tag{3}$$

with

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{\rm r} \\ \mathbf{H}_{\rm e} \end{bmatrix} \tag{4}$$

it follows that $\dim(\mathcal{S}_n) = n_{\rm t} - k$. Moreover, we use the notation

$$p \triangleq \dim(\mathcal{S}_{\rm r}) \qquad \text{and} \qquad s \triangleq \dim(\mathcal{S}_{\rm r,e}) \tag{5}$$

from which it follows that $\dim(\mathcal{S}_{\rm e}) = k - p - s$.

Using this notation, our channel decomposition is as follows.

Definition 1: The GSVD of $(\mathbf{H}_{\rm r}, \mathbf{H}_{\rm e})$ takes the form

$$\mathbf{H}_{\rm r} = \mathbf{\Psi}_{\rm r}\mathbf{\Sigma}_{\rm r}\,[\,\mathbf{\Omega}^{-1} \quad \mathbf{0}_{k\times(n_{\rm t}-k)}\,]\,\mathbf{\Psi}_{\rm t}^{\dagger} \tag{6a}$$
$$\mathbf{H}_{\rm e} = \mathbf{\Psi}_{\rm e}\mathbf{\Sigma}_{\rm e}\,[\,\mathbf{\Omega}^{-1} \quad \mathbf{0}_{k\times(n_{\rm t}-k)}\,]\,\mathbf{\Psi}_{\rm t}^{\dagger} \tag{6b}$$

where $\mathbf{\Psi}_{\rm r} \in \mathbb{C}^{n_{\rm r}\times n_{\rm r}}$, $\mathbf{\Psi}_{\rm e} \in \mathbb{C}^{n_{\rm e}\times n_{\rm e}}$ and $\mathbf{\Psi}_{\rm t} \in \mathbb{C}^{n_{\rm t}\times n_{\rm t}}$ are unitary, where $\mathbf{\Omega} \in \mathbb{C}^{k\times k}$ is lower triangular and nonsingular, and where

$$\mathbf{\Sigma}_{\rm r} = \begin{array}{c} \\ n_{\rm r}-p-s \\ s \\ p \end{array} \begin{array}{c} \begin{array}{ccc} k-p-s & s & p \end{array} \\ \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\rm r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \end{array} \tag{7a}$$

$$\mathbf{\Sigma}_{\rm e} = \begin{array}{c} \\ k-p-s \\ s \\ n_{\rm e}+p-k \end{array} \begin{array}{c} \begin{array}{ccc} k-p-s & s & p \end{array} \\ \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_{\rm e} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \end{array} \tag{7b}$$

are diagonal with

$$\mathbf{D}_{\rm r} = \mathrm{diag}\,(r_1, \ldots, r_s), \quad \mathbf{D}_{\rm e} = \mathrm{diag}\,(e_1, \ldots, e_s) \tag{8}$$

the diagonal entries of which are real and strictly positive. The associated generalized singular values are

$$\sigma_i \triangleq \frac{r_i}{e_i}, \quad i = 1, 2, \ldots, s. \tag{9}$$

For convenience, we choose the (otherwise arbitrary) indexing so that $\sigma_1 \le \sigma_2 \le \cdots \le \sigma_s$.

IV. Summary of Main Results

In this section we summarize the main results in this paper. The analysis is provided in Sections V–VII.

A. MIMOME Secrecy Capacity

A characterization of the secrecy capacity of the MIMOME channel is as follows.

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

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

where

$$R_{+}(\mathbf{K}_P, \mathbf{K}_{\mathbf{\Phi}}) = I(\mathsf{x}; \mathsf{y}_{\rm r} | \mathsf{y}_{\rm e}) \tag{11}$$

with $\mathsf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_P)$ and

$$\mathcal{K}_P \triangleq \left\{ \mathbf{K}_P : \mathbf{K}_P \succeq \mathbf{0}, \quad \mathrm{tr}\,(\mathbf{K}_P) \le P \right\} \tag{12}$$

and where

$$\mathsf{z} \triangleq \begin{bmatrix} \mathsf{z}_{\rm r} \\ \mathsf{z}_{\rm e} \end{bmatrix} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_{\mathbf{\Phi}}) \tag{13}$$

with[1]

$$\mathcal{K}_{\mathbf{\Phi}} \triangleq \left\{ \mathbf{K}_{\mathbf{\Phi}} : \mathbf{K}_{\mathbf{\Phi}} = \begin{bmatrix} \mathbf{I}_{n_{\rm r}} & \bar{\mathbf{\Phi}} \\ \bar{\mathbf{\Phi}}^{\dagger} & \mathbf{I}_{n_{\rm e}} \end{bmatrix}, \ \mathbf{K}_{\mathbf{\Phi}} \succeq \mathbf{0} \right\}. \tag{14}$$

Furthermore, the minimax problem of (10) is convex-concave with saddle point solution $(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}})$, via which the secrecy capacity can be expressed in the form

$$C = R_{-}(\bar{\mathbf{K}}_P) \triangleq \log \frac{\det(\mathbf{I} + \mathbf{H}_{\rm r}\bar{\mathbf{K}}_P\mathbf{H}_{\rm r}^{\dagger})}{\det(\mathbf{I} + \mathbf{H}_{\rm e}\bar{\mathbf{K}}_P\mathbf{H}_{\rm e}^{\dagger})}. \tag{15}$$

Finally, $C = 0$ if and only if

$$\mathbf{H}_{\rm r} = \bar{\mathbf{\Theta}}\mathbf{H}_{\rm e} \tag{16}$$

where

$$\bar{\mathbf{\Theta}} \triangleq \mathbf{\Theta}(\bar{\mathbf{K}}_P), \qquad \mathbf{\Theta}(\mathbf{K}_P) \triangleq \mathbf{\Theta}(\mathbf{K}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) \tag{17}$$

with

$$\mathbf{\Theta}(\mathbf{K}_P, \mathbf{K}_{\mathbf{\Phi}}) \triangleq (\mathbf{H}_{\rm r}\mathbf{K}_P\mathbf{H}_{\rm e}^{\dagger} + \mathbf{\Phi})(\mathbf{I} + \mathbf{H}_{\rm e}\mathbf{K}_P\mathbf{H}_{\rm e}^{\dagger})^{-1} \tag{18}$$

denoting the coefficient in the linear minimum mean-square error (MMSE) estimate of $\mathbf{y}_{\rm r}$ from $\mathbf{y}_{\rm e}$,

Several remarks are worthwhile. First, our result can be related to the Csiszár-Körner characterization of the secrecy capacity for a nondegraded discrete memoryless broadcast channel $p_{y_{\rm r}, y_{\rm e}|x}$ in the form [2]

$$C = \max_{p_u, p_{x|u}} I(u; y_{\rm r}) - I(u; y_{\rm e}) \tag{19}$$

[1]The constraint $\mathbf{K}_{\mathbf{\Phi}} \succeq \mathbf{0}$ is equivalently expressed as the requirement that $\sigma_{\max}(\mathbf{\Phi}) \le 1$, as we will exploit.

where u is an auxiliary random variable (over some alphabet with bounded cardinality) that satisfies the Markov constraint $u \leftrightarrow x \leftrightarrow (y_{\rm r}, y_{\rm e})$. As [2] remarks, the secrecy capacity (19) can be extended to incorporate continuous-valued inputs of the type of interest in the present paper. With such an extension, Theorem 1, and in particular (15), can be interpreted as (indirectly) establishing a suitable Gaussian wiretap code for achieving capacity.[2] Specifically, via the chain rule,

$$I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) = \big[I(\mathbf{x};\mathbf{y}_{\rm r}) - I(\mathbf{x};\mathbf{y}_{\rm e})\big] + I(\mathbf{x};\mathbf{y}_{\rm e}|\mathbf{y}_{\rm r})$$

where the last term on the right-hand side is zero when $\bar{\mathbf{\Phi}} = \bar{\tilde{\mathbf{\Phi}}}$, and thus we have the following immediate corollary.

Corollary 1: The secrecy capacity of the MIMOME wiretap channel is achieved by a wiretap coding scheme in which $\mathsf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_P)$ with $\mathbf{K}_P = \bar{\mathbf{K}}_P$, and $\mathsf{x} = \mathsf{u}$.

From this perspective, our result can also be interpreted as a convex reformulation of the nonconvex optimization (19). Indeed, even after knowing that both an optimizing u is Gaussian and $\mathsf{x} = \mathsf{u}$ is sufficient—which itself is nontrivial—determining the optimal covariance via

$$\bar{\mathbf{K}}_P \in \arg\max_{\mathbf{K}_P \in \mathcal{K}_P} \log \frac{\det(\mathbf{I} + \mathbf{H}_{\rm r}\mathbf{K}_P\mathbf{H}_{\rm r}^\dagger)}{\det(\mathbf{I} + \mathbf{H}_{\rm e}\mathbf{K}_P\mathbf{H}_{\rm e}^\dagger)} \tag{20}$$

with $\mathcal{K}_P$ as defined in (12), is a nonconvex problem.[3] And even if one verifies that $\bar{\mathbf{K}}_P$ satisfies the Karush–Kuhn–Tucker (KKT) conditions associated with (20), these necessary conditions only establish local optimality, i.e., that $\bar{\mathbf{K}}_P$ is a stationary point of the associated objective function. By contrast, (10) establishes that the (global) solution to (20) is obtained as the solution to a convex problem, as well as establishing the optimality of a Gaussian input distribution.

Second, additional insights are obtained from the structure of the saddle point solution $(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}})$. In particular, using $\bar{\mathbf{\Phi}}$ to denote the optimal cross-covariance, i.e., [cf. (14)]

$$\bar{\mathbf{K}}_{\mathbf{\Phi}} \triangleq \mathbf{K}_{\bar{\mathbf{\Phi}}} = \begin{bmatrix} \mathbf{I}_{n_{\rm r}} & \bar{\mathbf{\Phi}} \\ \bar{\mathbf{\Phi}}^\dagger & \mathbf{I}_{n_{\rm e}} \end{bmatrix} \tag{21}$$

we establish in the course of our development of Theorem 1 the following key property.

Property 1: The saddle point solution $(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}})$ to the MIMOME wiretap channel capacity (10) satisfies

$$\bar{\mathbf{\Phi}}^\dagger \mathbf{H}_{\rm r} \bar{\mathbf{S}} = \mathbf{H}_{\rm e} \bar{\mathbf{S}}, \ \forall \, \text{full column-rank } \bar{\mathbf{S}} \text{ s.t. } \bar{\mathbf{S}}\bar{\mathbf{S}}^\dagger = \bar{\mathbf{K}}_P \tag{22}$$

provided $\mathbf{H}_{\rm r} \neq \bar{\mathbf{\Theta}}\mathbf{H}_{\rm e}$ (i.e., provided $C \neq 0$).

It follows from (22) that the effective channel to the eavesdropper is a degraded version of that to the intended receiver. Indeed, the intended receiver can simulate the eavesdropper channel by adding noise. Specifically, it generates

$$\mathsf{y}'_{\rm e} = \bar{\mathbf{\Phi}}^\dagger \mathsf{y}_{\rm r} + \mathsf{w},$$

where the added noise $\mathsf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I} - \bar{\mathbf{\Phi}}^\dagger\bar{\mathbf{\Phi}})$ is independent of $\mathsf{y}_{\rm r}$, so, using (1), (22), and the notation $\mathsf{x} = \bar{\mathbf{S}}\mathsf{x}'$ with $\mathsf{x}' \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, we have

$$\mathsf{y}'_{\rm e} = \bar{\mathbf{\Phi}}^\dagger \mathbf{H}_{\rm r}\bar{\mathbf{S}}\mathsf{x}' + \bar{\mathbf{\Phi}}^\dagger \mathsf{z}_{\rm r} + \mathsf{w} = \mathbf{H}_{\rm e}\bar{\mathbf{S}}\mathsf{x}' + \mathsf{z}'_{\rm e} = \mathbf{H}_{\rm e}\mathsf{x} + \mathsf{z}'_{\rm e}$$

where $\mathsf{z}'_{\rm e} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. In essence, the optimal signal design for transmission is such that no information is transmitted along any direction where the eavesdropper observes a stronger signal than the legitimate receiver. A key consequence is that a genie-aided system in which $\mathsf{y}_{\rm e}$ is provided to the receiver, which would otherwise provide only an upper bound on capacity in general, does not increase the capacity of the channel in this case, a feature that is ultimately central to our analysis.

Finally, the condition (16) corresponding to when the secrecy capacity is zero has a natural physical interpretation. In particular, under this condition, the effective channel to the intended receiver is a degraded version of that to the eavesdropper. Indeed, the eavesdropper can simulate the intended receiver by adding noise. Specifically, it generates

$$\mathsf{y}'_{\rm r} = \bar{\mathbf{\Theta}}\mathsf{y}_{\rm e} + \mathsf{w},$$

where the added noise $\mathsf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I} - \bar{\mathbf{\Phi}}\bar{\mathbf{\Phi}}^\dagger)$ is independent of $\mathsf{y}_{\rm r}$, so, using (1) we have

$$\mathsf{y}'_{\rm r} = \bar{\mathbf{\Theta}}\mathbf{H}_{\rm e}\mathsf{x} + \bar{\mathbf{\Theta}}\mathsf{z}_{\rm r} + \mathsf{w} = \mathbf{H}_{\rm r}\mathsf{x} + \mathsf{z}'_{\rm r},$$

where $\mathsf{z}'_{\rm r} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ since

$$\bar{\mathbf{\Theta}} = \bar{\mathbf{\Phi}} \quad \text{if} \quad \mathbf{H}_{\rm r} = \bar{\mathbf{\Theta}}\mathbf{H}_{\rm e} \tag{23}$$

which follows from (17) with (18).

B. Secrecy Capacity in the High-SNR Regime

In the high-SNR limit (i.e., $P \to \infty$), the secrecy capacity (10) is naturally described in terms of the GSVD of the channel (1) as defined in (6). The GSVD simultaneously diagonalizes the $\mathbf{H}_{\rm r}$ and $\mathbf{H}_{\rm e}$, yielding an equivalent parallel channel model for the problem. As such, a capacity-approaching scheme in the high-SNR regime involves using for transmission (with a wiretap code) only those subchannels for which the gain to the intended receiver is larger, and the following convenient expression for the capacity (10) results.

Theorem 2: Let $\sigma_1 \leq \sigma_2 \leq \cdots \leq \sigma_s$ be the generalized singular values of $(\mathbf{H}_{\rm r}, \mathbf{H}_{\rm e})$. Then as $P \to \infty$, the secrecy capacity of the MIMOME wiretap channel (1) takes the asymptotic form

$$C(P) = C_0(P) + \sum_{j:\sigma_j \geq 1} \log \sigma_j^2 - o(1) \tag{24}$$

where

$$C_0(P) = \begin{cases} \log\det\Big(\mathbf{I} + \frac{P}{p}\mathbf{H}_{\rm r}\mathbf{H}_{\rm e}^\sharp\mathbf{H}_{\rm r}^\dagger\Big), & \text{rank}\,(\mathbf{H}_{\rm e}) < n_{\rm t}, \\ 0, & \text{rank}\,(\mathbf{H}_{\rm e}) = n_{\rm t}, \end{cases} \tag{25}$$

[2]Each candidate (u, x) in (19) corresponds to a particular coding scheme based on binning, which we generically refer as a "wiretap code," which achieves rate $I(u; y_{\rm r}) - I(u; y_{\rm e})$.

[3]Note that in the high-SNR regime, (20) reduces to

$$\max_{\mathbf{K} \in \mathcal{K}_\infty} \log \frac{\det(\mathbf{H}_{\rm r}\mathbf{K}\mathbf{H}_{\rm r}^\dagger)}{\det(\mathbf{H}_{\rm e}\mathbf{K}\mathbf{H}_{\rm e}^\dagger)},$$

which is the well-studied multiple-discriminant function in multivariate statistics; see, e.g., [27].

with p and s as given in (5), and with $\mathbf{H}_{\mathrm{e}}^{\sharp}$ denoting the projection matrix onto $\mathrm{Null}\,(\mathbf{H}_{\mathrm{e}})$.

Note that a simple and intuitive transmission scheme for the MIMOME channel would involve simultaneously and isotropically transmitting information in $\mathrm{Null}\,(\mathbf{H}_{\mathrm{r}})^{\perp}$, where there is gain to the intended receiver, and (synthetic) noise in $\mathrm{Null}\,(\mathbf{H}_{\mathrm{r}})$, which does not affect the intended receiver but does reduce the quality of the eavesdroppers received signal.[4] This "masked" multi-input, multi-output (MIMO) transmission scheme is the natural generalization of the masked beamforming proposed in [11] for the MISOME wiretap channel. For the MISOME channel, such an approach is near optimal, as shown in [4]. However, we now show that such a masked multi-input multi-output (MIMO) scheme can be quite far from optimal on the MIMOME channel.

For convenience, we restrict our attention to the case in which $n_{\mathrm{r}} \leq n_{\mathrm{t}} \leq n_{\mathrm{e}}$ and $\mathbf{H}_{\mathrm{r}}$ and $\mathbf{H}_{\mathrm{e}}$ are full rank—i.e., $\mathrm{rank}\,(\mathbf{H}_{\mathrm{r}}) = n_{\mathrm{r}}$ and $\mathrm{rank}\,(\mathbf{H}_{\mathrm{e}}) = n_{\mathrm{t}}$—and thus $k = n_{\mathrm{t}}$, $p = 0$, and $s = n_{\mathrm{r}}$ in the GSVD.

The masked MIMO scheme is naturally viewed as a wiretap coding scheme in which a particular (rather than optimal) choice for $(\mathbf{x}, \mathbf{u})$ is imposed in (19). In particular, first we choose $\mathbf{u}$ to correspond to (information-bearing) codewords in a randomly generated codebook, i.e.,

$$\mathbf{u} = (b_1, \ldots, b_{n_{\mathrm{r}}}) \tag{26a}$$

where the elements are generated in an i.i.d. manner according to $\mathcal{CN}(0, P_{\mathrm{t}})$ with

$$P_{\mathrm{t}} \triangleq \frac{P}{n_{\mathrm{t}}}. \tag{26b}$$

Additionally, we let $b_{n_{\mathrm{r}}+1}, \ldots, b_{n_{\mathrm{t}}}$ be randomly generated (synthetic) noise, i.e., independent $\mathcal{CN}(0, P_{\mathrm{t}})$ random variables.

Next, we choose the transmission $\mathbf{x}$ according to

$$\mathbf{x} = \sum_{j=1}^{n_{\mathrm{t}}} b_j \mathbf{v}_j \tag{26c}$$

where the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{n_{\mathrm{t}}}$ are chosen as follows. Let

$$\mathbf{H}_{\mathrm{r}} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}_{\mathrm{r}}^{\dagger} \tag{27}$$

be the compact singular value decomposition (SVD) of $\mathbf{H}_{\mathrm{r}}$. Since $\mathrm{rank}\,(\mathbf{H}_{\mathrm{r}}) = n_{\mathrm{r}}$, this means that $\mathbf{U}$ is $n_{\mathrm{r}} \times n_{\mathrm{r}}$ and unitary, $\boldsymbol{\Delta}$ is $n_{\mathrm{r}} \times n_{\mathrm{r}}$ and diagonal with positive diagonal elements, and $\mathbf{V}_{\mathrm{r}}$ is $n_{\mathrm{t}} \times n_{\mathrm{r}}$ with orthogonal columns. Then we choose $\mathbf{v}_1, \ldots, \mathbf{v}_{n_{\mathrm{r}}}$ in (26c) as the columns of $\mathbf{V}_{\mathrm{r}}$, i.e.,

$$\mathbf{V}_{\mathrm{r}} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_{n_{\mathrm{r}}} \end{bmatrix},$$

and (freely) choose

$$\mathbf{V}_n \triangleq \begin{bmatrix} \mathbf{v}_{n_{\mathrm{r}}+1} & \cdots & \mathbf{v}_{n_{\mathrm{t}}} \end{bmatrix} \tag{28}$$

a basis for the null space of $\mathbf{H}_{\mathrm{r}}$, so that $\begin{bmatrix} \mathbf{V}_{\mathrm{r}} & \mathbf{V}_n \end{bmatrix}$ is unitary.

[4]Note that the scheme is semi-blind: the transmitter does not need to know $\mathbf{H}_{\mathrm{e}}$ to construct the required subspaces, but does need to know $\mathbf{H}_{\mathrm{e}}$ in order to choose the communication rate.

As we will establish, substituting these parameters in the argument of (19) yields the achievable rate

$$R_{\mathrm{SN}}(P) = \log\det\left[\left(P_{\mathrm{t}}\mathbf{I} + \boldsymbol{\Delta}^{-2}\right)\left(\mathbf{H}_{\mathrm{r}}(\mathbf{I} + P_{\mathrm{t}}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{H}_{\mathrm{r}}^{\dagger}\right)\right] \tag{29}$$

which in the high-SNR regime reduces to

$$\lim_{P\to\infty} R_{\mathrm{SN}}(P) = \log\det\left(\mathbf{H}_{\mathrm{r}}(\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{H}_{\mathrm{r}}^{\dagger}\right) = \sum_{j=1}^{n_{\mathrm{r}}} \log\sigma_j^2 \tag{30}$$

where the second equality comes from expanding $\mathbf{H}_{\mathrm{r}}$ and $\mathbf{H}_{\mathrm{e}}$ via (6), with $\sigma_1, \sigma_2, \ldots$ denoting the generalized singular values (9). Comparing (30) and (24), we see that the asymptotic gap to capacity is

$$\lim_{P\to\infty}\left[C(P) - R_{\mathrm{SN}}(P)\right] = \sum_{j:\sigma_j<1} \log\frac{1}{\sigma_j^2},$$

which, evidently, can be arbitrarily large when there are small singular values.

In concluding this section, we emphasize that only in the high-SNR regime do the generalized singular values of $(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}})$ completely characterize the capacity-achieving and masked MIMO coding schemes.

C. MIMOME Channel Scaling Laws

By using sufficiently many antennas, the eavesdropper can drive to secrecy capacity to zero. In such a regime, the eavesdropper would be able to decode a nonvanishing fraction of any sent message—even when the sender and receiver fully exploit knowledge of $\mathbf{H}_{\mathrm{e}}$. In general, this threshold depends on the numbers of antennas at the transmitter and intended receiver, as well as on the particular channels to intended receiver and eavesdropper. One characterization of this threshold is given by (16) in Theorem 1. An equivalent characterization that is more useful in the development of scaling laws, is as follows.

Claim 1: The secrecy capacity of the MIMOME channel is zero if and only if

$$\sigma_{\max}(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}) \triangleq \sup_{\mathbf{v}\in\mathbb{C}^{n_{\mathrm{t}}}} \frac{\|\mathbf{H}_{\mathrm{r}}\mathbf{v}\|}{\|\mathbf{H}_{\mathrm{e}}\mathbf{v}\|} \leq 1. \tag{31}$$

where $\sigma_{\max}(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}})$ denotes the channel's largest generalized singular value.

When the coefficients of the channels are drawn at random, and the numbers of antennas are large, the threshold becomes independent of the channel realization. The following result characterizes this scaling behavior.

Corollary 2: Suppose that $\mathbf{H}_{\mathrm{r}}$ and $\mathbf{H}_{\mathrm{e}}$ have i.i.d. $\mathcal{CN}(0,1)$ entries that are fixed for the entire period of transmission, and known to all the terminals. Then when $n_{\mathrm{r}}, n_{\mathrm{e}}, n_{\mathrm{t}} \to \infty$ such that $\gamma \triangleq n_{\mathrm{r}}/n_{\mathrm{e}}$ and $\beta \triangleq n_{\mathrm{t}}/n_{\mathrm{e}}$ are fixed constants, the secrecy capacity satisfies $C(\mathbf{H}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}) \xrightarrow{\text{a.s.}} 0$ if and only if

$$0 \leq \beta \leq \frac{1}{2} \quad \text{and} \quad \gamma \leq (1 - \sqrt{2\beta})^2. \tag{32}$$

Fig. 1 depicts the zero-capacity region (32). In this plot, the solid curve describes the relative number of antennas an eavesdropper needs to prevent secure communication, as a function

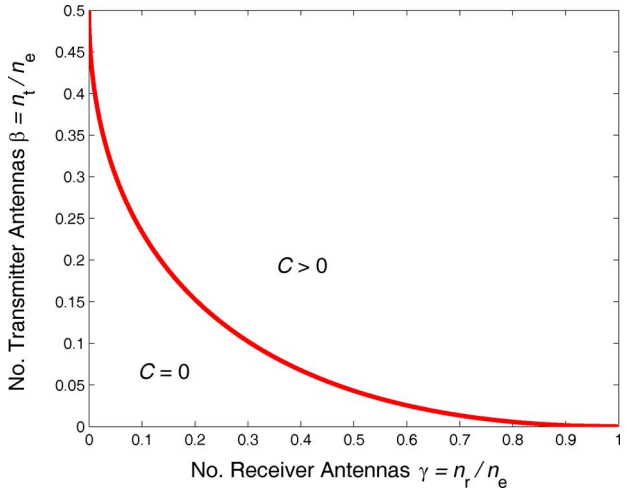

Fig. 1. The efficient frontier of secure communication region as a function of the number of antennas at the transmitter and intended receiver (relative to the number at the eavesdropper), in the limit of many antennas. The capacity is zero for any point below the curve, i.e., whenever the eavesdropper has sufficiently many antennas.

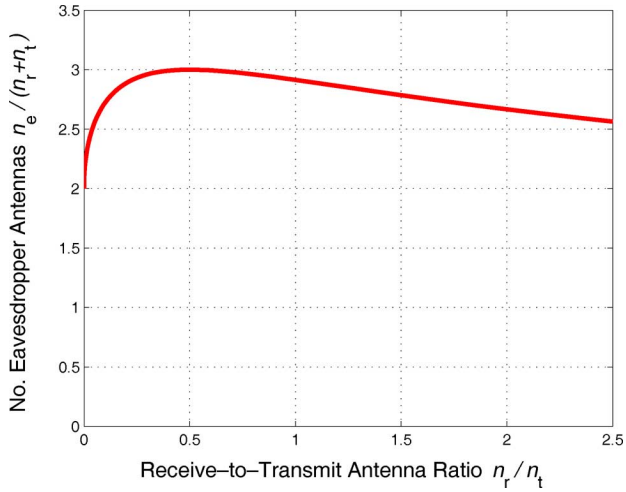

Fig. 2. The minimum (relative) number of eavesdropper antennas required to drive the secrecy capacity to zero, as a function of the antenna allocation between transmitter and intended receiver, in the limit of many antennas.

of the antenna resources available at the transmitter and intended receiver. The related scaling law developed for the MISOME case [4] corresponds to the vertical intercept of this plot: $C \xrightarrow{\text{a.s.}} 0$ when $\beta \leq 1/2$, i.e., when the eavesdropper has at least twice the number of antennas as the sender. Note, too, that the single transmit antenna (SIMOME) case corresponds to the horizontal intercept; in this case we see that $C \xrightarrow{\text{a.s.}} 0$ when $\gamma \leq 1$, i.e., when the eavesdropper has more antennas than the intended receiver.

We can further use such scaling analysis to determine the best asymptotic allocation of a (large) fixed number of antennas T between transmitter and intended receiver in the presence of an an eavesdropper. In particular, the optimum allocation is

$$(\beta_*, \gamma_*) = \underset{\substack{\{(\beta,\gamma):0\leq\beta\leq 1/2,\\ 0\leq\gamma\leq(1-\sqrt{2\beta})^2\}}}{\arg\min} (\beta+\gamma) = \left(\frac{2}{9}, \frac{1}{9}\right) \quad (33)$$

as is easily verified. Thus, the allocation that best thwarts the eavesdropper is $n_{\mathrm{r}}/n_{\mathrm{t}} = 1/2$, which requires the eavesdropper to use $3T$ antennas to prevent secure communication.

It is worth remarking that the objective function in (33) is rather insensitive to deviations from the optimal antenna allocation, as Fig. 2 demonstrates. If fact, even if we were to allocate equal numbers of antennas to the sender and the receiver, the eavesdropper would still need $(3/2+\sqrt{2})T \approx 2.9142\,T$ antennas to drive the secrecy capacity to zero.

V. MIMOME Secrecy Capacity Analysis

In this section we prove Theorem 1 . Our proof involves two main parts. We first recognize the right-hand side of (10) as an upper bound on the secrecy capacity, then exploit properties of the saddle point solution to establish

$$R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) = R_-(\bar{\mathbf{K}}_P) \quad (34)$$

where $R_-(\bar{\mathbf{K}}_P)$ is the lower bound (achievable rate) given in (15).

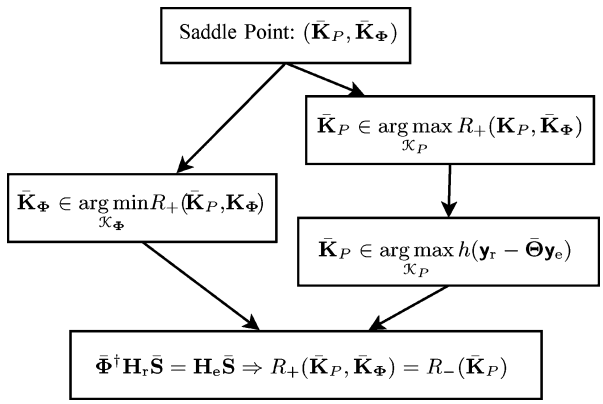

Fig. 3. Key steps in the proof of Theorem 1. First, the existence of a saddle point $(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}})$ is established, then the KKT conditions associated with the minimax expressions are used to simplify the saddle value to show that it matches the lower bound.

We begin by stating our upper bound, which is a trivial generalization of that established in [4].

Lemma 1 ([4]): An upper bound on the secrecy capacity of the MIMOME channel (1) is given by

$$C(P) \leq R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) = \min_{\mathbf{K}_{\mathbf{\Phi}}\in\mathcal{K}_{\mathbf{\Phi}}} \max_{\mathbf{K}_P\in\mathcal{K}_P} R_+(\mathbf{K}_P, \mathbf{K}_{\mathbf{\Phi}}) \quad (35)$$

where

$$R_+(\mathbf{K}_P, \mathbf{K}_{\mathbf{\Phi}}) \triangleq I(\mathbf{x}; \mathbf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) \quad (36)$$

with $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_P)$, and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_{\mathbf{\Phi}})$, and the domain sets $\mathcal{K}_P$ and $\mathcal{K}_{\mathbf{\Phi}}$ are defined via (12) and (14) respectively.

It remains to establish that this upper bound expression satisfies (34), which we do in the remainder of this section. We divide the proof into several steps, as depicted in Fig. 3.

Furthermore, we remark in advance that the analysis throughout is slightly simpler when $\mathbf{K}_{\mathbf{\Phi}} \succ \mathbf{0}$. Accordingly, in the following sections we focus on this nonsingular case and defer analysis for the singular case to appendices as it arises in our development. The key to analysis of the singular case is replacing the observations $\mathbf{y}_{\mathrm{r}}$ with reduced but equivalent

observations. In particular, we will make use of the following claim, a proof of which is provided in Appendix I.

Claim 2: Let the singular value decomposition of $\mathbf{\Phi}$ be expressed the form

$$\mathbf{\Phi} = [\,\mathbf{U}_1 \quad \mathbf{U}_2\,] \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Delta} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^\dagger \\ \mathbf{V}_2^\dagger \end{bmatrix}, \quad \sigma_{\max}(\mathbf{\Delta}) < 1. \tag{37}$$

Then if $p_{\mathbf{x}}$ is such that $I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) < \infty$, we have

$$I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) = I(\mathbf{x};\tilde{\mathbf{y}}_{\rm r}|\mathbf{y}_{\rm e}) \tag{38}$$

where

$$\tilde{\mathbf{y}}_{\rm r} \triangleq \mathbf{U}_2^\dagger \mathbf{y}_{\rm r} = \tilde{\mathbf{H}}_{\rm r}\mathbf{x} + \tilde{\mathbf{z}}_{\rm r} \tag{39}$$

with

$$\tilde{\mathbf{H}}_{\rm r} \triangleq \mathbf{U}_2^\dagger \mathbf{H}_{\rm r} \quad \text{and} \quad \tilde{\mathbf{z}}_{\rm r} \triangleq \mathbf{U}_2^\dagger \mathbf{z}_{\rm r}. \tag{40}$$

Symmetrically, if $p_{\mathbf{x}}$ is such that $I(\mathbf{x};\mathbf{y}_{\rm e}|\mathbf{y}_{\rm r}) < \infty$, we have

$$I(\mathbf{x};\mathbf{y}_{\rm e}|\mathbf{y}_{\rm r}) = I(\mathbf{x};\tilde{\mathbf{y}}_{\rm e}|\mathbf{y}_{\rm r}) \tag{41}$$

where

$$\tilde{\mathbf{y}}_{\rm e} \triangleq \mathbf{V}_2^\dagger \mathbf{y}_{\rm e} = \tilde{\mathbf{H}}_{\rm e}\mathbf{x} + \tilde{\mathbf{z}}_{\rm e} \tag{42}$$

with

$$\tilde{\mathbf{H}}_{\rm e} \triangleq \mathbf{V}_2^\dagger \mathbf{H}_{\rm e} \quad \text{and} \quad \tilde{\mathbf{z}}_{\rm e} \triangleq \mathbf{V}_2^\dagger \mathbf{z}_{\rm e}. \tag{43}$$

Finally, for any $p_{\mathbf{x}}$ we have that $I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) = \infty$ if and only if

$$\mathbf{T}\mathbf{K}_P\mathbf{T}^\dagger \neq \mathbf{0} \tag{44}$$

where $\mathbf{K}_P$ is the covariance associated with $p_{\mathbf{x}}$, and where

$$\mathbf{T} \triangleq \mathbf{U}_1^\dagger \mathbf{H}_{\rm r} - \mathbf{V}_1^\dagger \mathbf{H}_{\rm e}. \tag{45}$$

Note that when (38) holds, the equivalent model holds and

$$\tilde{\mathbf{\Phi}} \triangleq E[\tilde{\mathbf{z}}_{\rm r}\mathbf{z}_{\rm e}^\dagger] = \mathbf{U}_2^\dagger \mathbf{\Phi} \tag{46}$$

is the equivalent noise cross-covariance.

A. Existence of a Saddle Point Solution

We first show that the minimax upper bound is a convex-concave problem with a (finite) saddle point solution.

Lemma 2: The upper bound (35) has a saddle point solution, i.e., there exists $(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) \in \mathcal{K}_P \times \mathcal{K}_{\mathbf{\Phi}}$ such that

$$R_+(\mathbf{K}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) \leq R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) \leq R_+(\bar{\mathbf{K}}_P, \mathbf{K}_{\mathbf{\Phi}}) \tag{47}$$

holds for each $(\mathbf{K}_P, \mathbf{K}_{\mathbf{\Phi}}) \in \mathcal{K}_P \times \mathcal{K}_{\mathbf{\Phi}}$. Moreover, the saddle value is finite, i.e.

$$R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) < \infty. \tag{48}$$

Proof: Since the constraint sets $\mathcal{K}_P$ and $\mathcal{K}_{\mathbf{\Phi}}$ are convex and compact, from a special case of Sion's minimax theorem [28] it suffices to show that

$$R_+(\,\mathbf{K}_P, \cdot) \text{ is convex on } \mathcal{K}_{\mathbf{\Phi}} \text{ for each } \mathbf{K}_P \in \mathcal{K}_P \tag{P1}$$

$$R_+(\cdot,\, \mathbf{K}_{\mathbf{\Phi}}) \text{ is concave on } \mathcal{K}_P \text{ for each } \mathbf{K}_{\mathbf{\Phi}} \in \mathcal{K}_{\mathbf{\Phi}}. \tag{P2}$$

To first establish (P1), we begin by writing

$$I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) = I(\mathbf{x};\mathbf{y}_{\rm r},\mathbf{y}_{\rm e}) - I(\mathbf{x};\mathbf{y}_{\rm e}) \tag{49}$$

and observe that the second term in (49) is fixed for each $\mathbf{K}_{\mathbf{\Phi}} \in \mathcal{K}_{\mathbf{\Phi}}$. Thus it suffices to show that with $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_P)$, the first term in (49) is convex in $\mathbf{K}_{\mathbf{\Phi}}$. This is established in, e.g., [29, Lemma II-3, p. 3076].

We next establish (P2). With slight abuse of notation, we define $R_+(p_{\mathbf{x}}, \mathbf{K}_{\mathbf{\Phi}}) = I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e})$ with $\mathbf{x} \sim p_{\mathbf{x}}$ and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_{\mathbf{\Phi}})$. By contrast, our original notation $R_+(\mathbf{Q}, \mathbf{K}_{\mathbf{\Phi}})$ corresponds to the special case of $R_+(p_{\mathbf{x}}, \mathbf{K}_{\mathbf{\Phi}})$ in which $p_{\mathbf{x}} = \mathcal{CN}(\mathbf{0}, \mathbf{Q})$. Let $p_{\mathbf{x}}^0 = \mathcal{CN}(\mathbf{0}, \mathbf{Q}_0)$, $p_{\mathbf{x}}^1 = \mathcal{CN}(\mathbf{0}, \mathbf{Q}_1)$, $p_{\mathbf{x}}^\theta = \theta p_{\mathbf{x}}^1 + (1-\theta)p_{\mathbf{x}}^0$, and $\mathbf{Q}^\theta = (1-\theta)\mathbf{Q}_0 + \theta\mathbf{Q}_1$, for some $\theta \in [0,1]$. Then the required concavity follows from

$$\begin{aligned} R_+(\mathbf{Q}^\theta, \mathbf{K}_{\mathbf{\Phi}}) &= R_+(\mathcal{CN}(\mathbf{0}, \mathbf{Q}^\theta), \mathbf{K}_{\mathbf{\Phi}}) \\ &\geq R_+(p_{\mathbf{x}}^\theta, \mathbf{K}_{\mathbf{\Phi}}) \qquad (50) \\ &\geq (1-\theta)R_+(p_{\mathbf{x}}^0, \mathbf{K}_{\mathbf{\Phi}}) + \theta R_+(p_{\mathbf{x}}^1, \mathbf{K}_{\mathbf{\Phi}}) \\ &= (1-\theta)R_+(\mathbf{Q}_0, \mathbf{K}_{\mathbf{\Phi}}) + \theta R_+(\mathbf{Q}_1, \mathbf{K}_{\mathbf{\Phi}}) \qquad (51) \end{aligned}$$

where (50) follows from the fact that a Gaussian distribution maximizes $R_+(p_{\mathbf{x}}^\theta, \mathbf{K}_{\mathbf{\Phi}})$ among all distributions with a given covariance, which we discuss below, and where (51) follows from the fact that $I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e})$ is concave in $p_{\mathbf{x}}$ for each fixed $p_{\mathbf{y}_{\rm r},\mathbf{y}_{\rm e}|\mathbf{x}}$; see, e.g., [8, App. I].

Verifying (50) is straightforward when $\mathbf{K}_{\mathbf{\Phi}}$ is nonsingular, i.e., $\|\mathbf{\Phi}\|_2 < 1$. Specifically, with

$$\begin{aligned} &\mathbf{\Lambda}(\mathbf{K}_P, \mathbf{K}_{\mathbf{\Phi}}) \\ &\triangleq \mathbf{I} + \mathbf{H}_{\rm r}\mathbf{K}_P\mathbf{H}_{\rm r}^\dagger \\ &\quad - (\mathbf{\Phi} + \mathbf{H}_{\rm r}\mathbf{K}_P\mathbf{H}_{\rm e}^\dagger)(\mathbf{I} + \mathbf{H}_{\rm e}\mathbf{K}_P\mathbf{H}_{\rm e}^\dagger)^{-1}(\mathbf{\Phi}^\dagger + \mathbf{H}_{\rm e}\mathbf{K}_P\mathbf{H}_{\rm r}^\dagger) \end{aligned} \tag{52}$$

denoting the error covariance associated with the linear MMSE estimate $\mathbf{\Theta}(\mathbf{K}_P, \mathbf{K}_{\mathbf{\Phi}})\mathbf{y}_{\rm e}$ of $\mathbf{y}_{\rm r}$ from $\mathbf{y}_{\rm e}$, a simple generalization of [4, Lemma 2] yields

$$\begin{aligned} I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) &= h(\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) - h(\mathbf{z}_{\rm r}|\mathbf{z}_{\rm e}) \qquad (53) \\ &= h(\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) - \log\det \pi e(\mathbf{I} - \mathbf{\Phi}\mathbf{\Phi}^\dagger) \\ &\leq \log\det \mathbf{\Lambda}(\mathbf{K}_P, \mathbf{K}_{\mathbf{\Phi}}) - \log\det(\mathbf{I} - \mathbf{\Phi}\mathbf{\Phi}^\dagger) \qquad (54) \end{aligned}$$

where the last inequality is satisfied with equality if $p_{\mathbf{x}} = \mathcal{CN}(\mathbf{0}, \mathbf{K}_P)$. When $\mathbf{K}_{\mathbf{\Phi}}$ is singular, (53)–(54) is not well-defined, so some straightforward modifications to the approach are required; these we detail in Appendix II.

Finally, to verify (48), it suffices to note that

$$R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) \leq R_+(\bar{\mathbf{K}}_P, \mathbf{I}) \leq I(\mathbf{x};\mathbf{y}_{\rm e},\mathbf{y}_{\rm r}) < \infty$$

where the second inequality follows from the chain rule $I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) = I(\mathbf{x};\mathbf{y}_{\rm e},\mathbf{y}_{\rm r}) - I(\mathbf{x};\mathbf{y}_{\rm e})$, and where the last inequality follows from the fact that $\mathrm{cov}\,(\mathbf{z}) = \mathbf{I}$. ■

B. Property of the Saddle Point

To simplify evaluation of the associated saddle value, we now develop the Property 1. For notational convenience, we define $\bar{\mathbf{\Lambda}}$ via [cf. (52)]

$$\bar{\mathbf{\Lambda}} \triangleq \mathbf{\Lambda}(\bar{\mathbf{K}}_P), \qquad \mathbf{\Lambda}(\mathbf{K}_P) \triangleq \mathbf{\Lambda}(\mathbf{K}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}). \tag{55}$$

The required property is obtained by combining the following two lemmas.

Lemma 3: A saddle point solution $(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}})$ to (35) satisfies

$$(\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})\bar{\mathbf{K}}_P(\bar{\mathbf{\Phi}}^{\dagger}\mathbf{H}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}})^{\dagger} = \mathbf{0}. \tag{56}$$

Lemma 4: A saddle point solution $(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}})$ to (35) is such that

$$(\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})\bar{\mathbf{S}} \text{ has a full column-rank} \tag{57}$$

provided $\mathbf{H}_{\mathrm{r}} \neq \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}}$, where $\bar{\mathbf{S}}$ is a full column-rank matrix such that $\bar{\mathbf{S}}\bar{\mathbf{S}}^{\dagger} = \bar{\mathbf{K}}_P$.

In particular, combining (56) and (57) we immediately obtain (22), since for a full column-rank matrix $\mathbf{M}$, $\mathbf{Ma} = \mathbf{0}$ if and only if $\mathbf{a} = \mathbf{0}$.

In the remainder of the section, we prove the two lemmas.

Proof of Lemma 3: Here we consider the simpler case when $\bar{\mathbf{K}}_{\mathbf{\Phi}} \succ \mathbf{0}$; the extension of the proof to the case when $\bar{\mathbf{K}}_{\mathbf{\Phi}}$ is singular is provided in Appendix III.

We begin by noting that the second inequality in (47) implies

$$\bar{\mathbf{K}}_{\mathbf{\Phi}} \in \arg\min_{\mathbf{K}_{\mathbf{\Phi}} \in \mathcal{K}_{\mathbf{\Phi}}} R_{+}(\bar{\mathbf{K}}_P, \mathbf{K}_{\mathbf{\Phi}}). \tag{58}$$

The Lagrangian associated with the minimization (58) is

$$\mathcal{L}_{\mathbf{\Phi}}(\mathbf{K}_{\mathbf{\Phi}}, \mathbf{\Upsilon}) = R_{+}(\bar{\mathbf{K}}_P, \mathbf{K}_{\mathbf{\Phi}}) + \mathrm{tr}\,(\mathbf{\Upsilon}\mathbf{K}_{\mathbf{\Phi}}) \tag{59}$$

where the dual variable

$$\mathbf{\Upsilon} = \begin{array}{c} \\ n_{\mathrm{r}} \\ n_{\mathrm{e}} \end{array} \begin{array}{c} \begin{array}{cc} n_{\mathrm{r}} & n_{\mathrm{e}} \end{array} \\ \begin{bmatrix} \mathbf{\Upsilon}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Upsilon}_2 \end{bmatrix} \end{array} \tag{60}$$

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 KKT conditions yield

$$\begin{aligned} &\nabla_{\mathbf{K}_{\mathbf{\Phi}}} \mathcal{L}_{\mathbf{\Phi}}(\mathbf{K}_{\mathbf{\Phi}}, \mathbf{\Upsilon})\big|_{\mathbf{K}_{\mathbf{\Phi}} = \bar{\mathbf{K}}_{\mathbf{\Phi}}} \\ &\qquad = \nabla_{\mathbf{K}_{\mathbf{\Phi}}} R_{+}(\bar{\mathbf{K}}_P, \mathbf{K}_{\mathbf{\Phi}})\big|_{\mathbf{K}_{\mathbf{\Phi}} = \bar{\mathbf{K}}_{\mathbf{\Phi}}} + \mathbf{\Upsilon} = \mathbf{0}. \end{aligned} \tag{61}$$

Substituting

$$\begin{aligned} &\nabla_{\mathbf{K}_{\mathbf{\Phi}}} R_{+}(\bar{\mathbf{K}}_P, \mathbf{K}_{\mathbf{\Phi}})\big|_{\mathbf{K}_{\mathbf{\Phi}} = \bar{\mathbf{K}}_{\mathbf{\Phi}}} \\ &\quad = \nabla_{\mathbf{K}_{\mathbf{\Phi}}} \left[\log\det(\mathbf{K}_{\mathbf{\Phi}} + \mathbf{H}\bar{\mathbf{K}}_P\mathbf{H}^{\dagger}) - \log\det(\mathbf{K}_{\mathbf{\Phi}})\right]\big|_{\mathbf{K}_{\mathbf{\Phi}} = \bar{\mathbf{K}}_{\mathbf{\Phi}}} \\ &\quad = (\bar{\mathbf{K}}_{\mathbf{\Phi}} + \mathbf{H}\bar{\mathbf{K}}_P\mathbf{H}^{\dagger})^{-1} - \bar{\mathbf{K}}_{\mathbf{\Phi}}^{-1} \end{aligned} \tag{62}$$

with (4) into (61) and simplifying, we obtain,

$$\mathbf{H}\bar{\mathbf{K}}_P\mathbf{H}^{\dagger} = \bar{\mathbf{K}}_{\mathbf{\Phi}}\mathbf{\Upsilon}(\bar{\mathbf{K}}_{\mathbf{\Phi}} + \mathbf{H}\bar{\mathbf{K}}_P\mathbf{H}^{\dagger}). \tag{63}$$

To complete the proof requires a straightforward manipulation of (63) to obtain (56). Specifically, substituting for $\bar{\mathbf{K}}_{\mathbf{\Phi}}$ from (21) and $\mathbf{H}$ from (4) into (63), and carrying out the associated block matrix multiplication yields

$$\mathbf{H}_{\mathrm{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{r}}^{\dagger} = \mathbf{\Upsilon}_1(\mathbf{I} + \mathbf{H}_{\mathrm{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{r}}^{\dagger}) + \bar{\mathbf{\Phi}}\mathbf{\Upsilon}_2(\bar{\mathbf{\Phi}}^{\dagger} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{r}}^{\dagger}) \tag{64}$$

$$\mathbf{H}_{\mathrm{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger} = \mathbf{\Upsilon}_1(\bar{\mathbf{\Phi}} + \mathbf{H}_{\mathrm{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger}) + \bar{\mathbf{\Phi}}\mathbf{\Upsilon}_2(\mathbf{I} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger}) \tag{65}$$

$$\mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{r}}^{\dagger} = \bar{\mathbf{\Phi}}^{\dagger}\mathbf{\Upsilon}_1(\mathbf{I} + \mathbf{H}_{\mathrm{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{r}}^{\dagger}) + \mathbf{\Upsilon}_2(\bar{\mathbf{\Phi}}^{\dagger} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{r}}^{\dagger}) \tag{66}$$

$$\mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger} = \bar{\mathbf{\Phi}}^{\dagger}\mathbf{\Upsilon}_1(\bar{\mathbf{\Phi}} + \mathbf{H}_{\mathrm{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger}) + \mathbf{\Upsilon}_2(\mathbf{I} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger}). \tag{67}$$

Eliminating $\mathbf{\Upsilon}_1$ from (64) and (66), we obtain

$$(\bar{\mathbf{\Phi}}^{\dagger}\mathbf{H}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}})\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{r}}^{\dagger} = (\bar{\mathbf{\Phi}}^{\dagger}\bar{\mathbf{\Phi}} - \mathbf{I})\mathbf{\Upsilon}_2(\bar{\mathbf{\Phi}}^{\dagger} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{r}}^{\dagger}) \tag{68}$$

and eliminating $\mathbf{\Upsilon}_1$ from (65) and (67), we obtain

$$(\bar{\mathbf{\Phi}}^{\dagger}\mathbf{H}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}})\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger} = (\bar{\mathbf{\Phi}}^{\dagger}\bar{\mathbf{\Phi}} - \mathbf{I})\mathbf{\Upsilon}_2(\mathbf{I} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger}). \tag{69}$$

Finally, eliminating $\mathbf{\Upsilon}_2$ from (68) and (69), we obtain

$$\begin{aligned} &(\bar{\mathbf{\Phi}}^{\dagger}\mathbf{H}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}})\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{r}}^{\dagger} \\ &\quad = (\bar{\mathbf{\Phi}}^{\dagger}\mathbf{H}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}})\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger}(\mathbf{I} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger})^{-1}(\bar{\mathbf{\Phi}}^{\dagger} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{r}}^{\dagger}) \\ &\quad = (\bar{\mathbf{\Phi}}^{\dagger}\mathbf{H}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}})\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger}\bar{\mathbf{\Theta}}^{\dagger} \end{aligned} \tag{70}$$

which reduces to (56) as desired. ■

In preparation for proving Lemma 4, we establish the following key proposition, whose proof is provided in Appendix IV.

Proposition 1: When $\mathsf{x} \sim \mathcal{CN}(\mathbf{0}, \mathbf{K}_P)$ and $\mathsf{z} \sim \mathcal{CN}(\mathbf{0}, \bar{\mathbf{K}}_{\mathbf{\Phi}})$ with $\bar{\mathbf{K}}_{\mathbf{\Phi}} \succ \mathbf{0}$ in the model (1), we have[5]

$$\arg\max_{\mathbf{K}_P \in \mathcal{K}_P} h(\mathbf{y}_{\mathrm{r}} - \mathbf{\Theta}(\mathbf{K}_P)\mathbf{y}_{\mathrm{e}}) = \arg\max_{\mathbf{K}_P \in \mathcal{K}_P} h(\mathbf{y}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{y}_{\mathrm{e}}) \tag{71}$$

where $\bar{\mathbf{\Theta}}$ and $\mathbf{\Theta}(\mathbf{K}_P)$ are as defined in (17) with (18).

Proof of Lemma 4: Again, here we consider the simpler case when $\bar{\mathbf{K}}_{\mathbf{\Phi}}$ is nonsingular; a proof for the case when $\bar{\mathbf{K}}_{\mathbf{\Phi}}$ is singular is provided in Appendix V.

We begin by noting that

$$\bar{\mathbf{K}}_P \in \arg\max_{\mathbf{K}_P \in \mathcal{K}_P} R_{+}(\mathbf{K}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) \tag{72}$$

$$= \arg\max_{\mathbf{K}_P \in \mathcal{K}_P} h(\mathbf{y}_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}}) \tag{73}$$

$$= \arg\max_{\mathbf{K}_P \in \mathcal{K}_P} h(\mathbf{y}_{\mathrm{r}} - \mathbf{\Theta}(\mathbf{K}_P)\mathbf{y}_{\mathrm{e}}) \tag{74}$$

$$= \arg\max_{\mathbf{K}_P \in \mathcal{K}_P} h(\mathbf{y}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{y}_{\mathrm{e}}) \tag{75}$$

$$= \arg\max_{\mathbf{K}_P \in \mathcal{K}_P} \log\det(\mathbf{I} + \bar{\mathbf{H}}_{\mathrm{eff}}\mathbf{K}_P\bar{\mathbf{H}}_{\mathrm{eff}}^{\dagger}) \tag{76}$$

where (72) follows from the first inequality in (47), where (73) follows from the fact that $\bar{\mathbf{K}}_{\mathbf{\Phi}} \succ \mathbf{0}$, where (75) follows from Proposition 1, and where in (76) we have the effective channel[6]

$$\bar{\mathbf{H}}_{\mathrm{eff}} \triangleq \bar{\mathbf{J}}^{-1/2}(\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}}) \tag{77a}$$

with

$$\begin{aligned} \bar{\mathbf{J}} &\triangleq \mathbf{I} + \bar{\mathbf{\Theta}}\bar{\mathbf{\Theta}}^{\dagger} - \bar{\mathbf{\Theta}}\bar{\mathbf{\Phi}}^{\dagger} - \bar{\mathbf{\Phi}}\bar{\mathbf{\Theta}}^{\dagger} \\ &= (\mathbf{I} - \bar{\mathbf{\Phi}}\bar{\mathbf{\Phi}}^{\dagger}) + (\bar{\mathbf{\Theta}} - \bar{\mathbf{\Phi}})(\bar{\mathbf{\Theta}} - \bar{\mathbf{\Phi}})^{\dagger} \end{aligned} \tag{77b}$$

which is nonsingular since $\bar{\mathbf{K}}_{\mathbf{\Phi}} \succ \mathbf{0}$.

[5]Note that the maximum on the left-hand side is in general a lower bound on the maximum on the right-hand side.

[6]As an aside, note that (76) provides the interpretation of $\bar{\mathbf{K}}_P$ as an optimal input covariance for a MIMO channel with matrix $\bar{\mathbf{H}}_{\mathrm{eff}}$ and unit-variance white Gaussian noise.

Finally, because $\bar{\mathbf{J}} \succ \mathbf{0}$, showing (57) is equivalent to showing that that $\bar{\mathbf{H}}_{\text{eff}}\bar{\mathbf{S}}$ has full column-rank, which we establish in the sequel to conclude the proof. First, we express $\bar{\mathbf{H}}_{\text{eff}}$ in terms of its singular value decomposition

$$\bar{\mathbf{H}}_{\text{eff}} = \mathbf{A}\bar{\mathbf{\Sigma}}_{\text{eff}}\mathbf{B}^{\dagger} \tag{78}$$

i.e., $\mathbf{A}$ and $\mathbf{B}$ are unitary matrices, and

$$\bar{\mathbf{\Sigma}}_{\text{eff}} = \begin{array}{c} \\ \nu \\ n_{\text{r}}-\nu \end{array} \begin{array}{c} \begin{array}{cc} \nu & n_{\text{t}}-\nu \end{array} \\ \begin{bmatrix} \bar{\mathbf{\Sigma}}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \end{array} \tag{79}$$

where $\nu \triangleq \text{rank}\,(\bar{\mathbf{H}}_{\text{eff}}) > 0$ and $\bar{\mathbf{\Sigma}}_0$ is diagonal with strictly positive entries. We establish that $\bar{\mathbf{H}}_{\text{eff}}\bar{\mathbf{S}}$ has full column-rank by showing that the columns of $\bar{\mathbf{S}}$ are spanned by the first ν columns of $\mathbf{B}$, i.e.,

$$\bar{\mathbf{F}} \triangleq \mathbf{B}^{\dagger}\,\bar{\mathbf{K}}_P\,\mathbf{B} = \begin{array}{c} \\ \nu \\ n_{\text{t}}-\nu \end{array} \begin{array}{c} \begin{array}{cc} \nu & n_{\text{t}}-\nu \end{array} \\ \begin{bmatrix} \bar{\mathbf{F}}_0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \end{array} \tag{80}$$

for some $\bar{\mathbf{F}}_0 \succeq \mathbf{0}$.

To this end, substituting (78) into (76), we obtain

$$\begin{aligned} \bar{\mathbf{K}}_P &= \arg\max_{\mathbf{K}_P \in \mathcal{K}_P} \log\det(\mathbf{I} + \mathbf{A}\bar{\mathbf{\Sigma}}_{\text{eff}}\mathbf{B}^{\dagger}\mathbf{K}_P\mathbf{B}\bar{\mathbf{\Sigma}}_{\text{eff}}^{\dagger}\mathbf{A}^{\dagger}) \\ &= \arg\max_{\mathbf{K}_P \in \mathcal{K}_P} \log\det(\mathbf{I} + \bar{\mathbf{\Sigma}}_{\text{eff}}\mathbf{B}^{\dagger}\mathbf{K}_P\mathbf{B}\bar{\mathbf{\Sigma}}_{\text{eff}}^{\dagger}) \end{aligned} \tag{81}$$

Now $\mathbf{K}_P \in \mathcal{K}_P$ if and only if $\mathbf{F} = \mathbf{B}^{\dagger}\mathbf{K}_P\mathbf{B} \in \mathcal{K}_P$, so (81) implies that

$$\begin{aligned} \bar{\mathbf{F}} \in \arg\max_{\mathbf{F} \in \mathcal{K}_P} &\log\det(\mathbf{I} + \bar{\mathbf{\Sigma}}_{\text{eff}}\mathbf{F}\bar{\mathbf{\Sigma}}_{\text{eff}}^{\dagger}) \\ &= \arg\max_{\mathbf{F} \in \mathcal{K}_P} \log\det(\mathbf{I} + \bar{\mathbf{\Sigma}}_0\mathbf{F}_0\bar{\mathbf{\Sigma}}_0^{\dagger}) \end{aligned} \tag{82}$$

with $\mathbf{F}$ expressed in terms of the block notation

$$\mathbf{F} = \begin{array}{c} \\ \nu \\ n_{\text{t}}-\nu \end{array} \begin{array}{c} \begin{array}{cc} \nu & n_{\text{t}}-\nu \end{array} \\ \begin{bmatrix} \mathbf{F}_0 & \mathbf{F}_1 \\ \mathbf{F}_1^{\dagger} & \mathbf{F}_2 \end{bmatrix} \end{array} \tag{83}$$

and where (82) follows from (79).

Finally, it follows that $\bar{\mathbf{F}}_1$ and $\bar{\mathbf{F}}_2$, the $\mathbf{F}_1$ and $\mathbf{F}_2$ in (83) when $\mathbf{F} = \bar{\mathbf{F}}$, are both $\mathbf{0}$. Indeed, if $\bar{\mathbf{F}}_2 \neq 0$, then $\text{tr}\,(\bar{\mathbf{F}}_2) > 0$. This would contradict the optimality in (82): since the objective function only depends on $\bar{\mathbf{F}}_0$, one could strictly increase the objective function by increasing the trace of $\bar{\mathbf{F}}_0$ and decreasing the trace of $\bar{\mathbf{F}}_2$. Finally, since $\bar{\mathbf{F}} \succeq \mathbf{0}$ and $\bar{\mathbf{F}}_2 = 0$, it follows that $\bar{\mathbf{F}}_1 = 0$. ■

C. Evaluation of the Saddle Value: Proof of Theorem 1

The conditions in Lemmas 3 and 4 can be used in turn to establish the tightness of the upper bound (35).

Lemma 5: The saddle value $R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}})$ in (35) can be expressed as

$$R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) = \begin{cases} R_-(\bar{\mathbf{K}}_P), & \mathbf{H}_{\text{r}} \neq \bar{\mathbf{\Theta}}\mathbf{H}_{\text{e}} \\ 0, & \text{otherwise} \end{cases} \tag{84}$$

where $R_-(\bar{\mathbf{K}}_P)$ is as given in (15).

The proof of Theorem 1 is a direct consequence of Lemma 5. If $R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) = 0$, the capacity is zero, otherwise $R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) = R_-(\bar{\mathbf{K}}_P)$, and the latter expression is an achievable rate as can be seen by setting $p_u = p_{\mathbf{x}} = \mathcal{CN}(\mathbf{0}, \bar{\mathbf{K}}_P)$ in the argument of (19).

Thus, to conclude the section it remains only to prove our lemma.

Proof of Lemma 5: Here we consider the case when when $\bar{\mathbf{K}}_{\mathbf{\Phi}} \succ \mathbf{0}$, i.e., $\|\mathbf{\Phi}\|_2 < 1$; the proof for the case when $\bar{\mathbf{K}}_{\mathbf{\Phi}}$ is singular is provided in Appendix VI.

To obtain (84) when $\mathbf{H}_{\text{r}} \neq \bar{\mathbf{\Theta}}\mathbf{H}_{\text{e}}$, we begin by writing the gap between upper and lower bounds as

$$\begin{aligned} &R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) - R_-(\bar{\mathbf{K}}_P) \\ &\quad = I(\mathbf{x};\mathbf{y}_{\text{r}}|\mathbf{y}_{\text{e}}) - [I(\mathbf{x};\mathbf{y}_{\text{r}}) - I(\mathbf{x};\mathbf{y}_{\text{e}})] \\ &\quad = I(\mathbf{x};\mathbf{y}_{\text{e}}|\mathbf{y}_{\text{r}}) \\ &\quad = h(\mathbf{y}_{\text{e}}|\mathbf{y}_{\text{r}}) - h(\mathbf{z}_{\text{e}}|\mathbf{z}_{\text{r}}) \end{aligned} \tag{85}$$

then note that this gap is zero since

$$\begin{aligned} &h(\mathbf{y}_{\text{e}}|\mathbf{y}_{\text{r}}) \\ &= \log\det \pi e\mathbf{\Lambda}_{\text{b}} && (86) \\ &= \log\det \pi e(\mathbf{I}+\mathbf{H}_{\text{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\text{e}}^{\dagger} - \bar{\mathbf{\Phi}}^{\dagger}(\mathbf{I}+\mathbf{H}_{\text{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\text{r}}^{\dagger})\bar{\mathbf{\Phi}}) && (87) \\ &= \log\det \pi e(\mathbf{I}-\bar{\mathbf{\Phi}}^{\dagger}\bar{\mathbf{\Phi}}) \\ &= h(\mathbf{z}_{\text{e}}|\mathbf{z}_{\text{r}}) && (88) \end{aligned}$$

where in (86)

$$\begin{aligned} \mathbf{\Lambda}_{\text{b}} = &\mathbf{I} + \mathbf{H}_{\text{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\text{e}}^{\dagger} \\ &- (\bar{\mathbf{\Phi}}^{\dagger}+\mathbf{H}_{\text{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\text{r}}^{\dagger})(\mathbf{I}+\mathbf{H}_{\text{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\text{r}}^{\dagger})^{-1}(\bar{\mathbf{\Phi}}+\mathbf{H}_{\text{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\text{e}}^{\dagger})) \end{aligned} \tag{89}$$

is the "backward" error covariance associated with the linear MMSE estimate of $\mathbf{y}_{\text{e}}$ from $\mathbf{y}_{\text{r}}$, and where to obtain each of (87) and (88) we have used (22) of Property 1.

To obtain (84) when $\mathbf{H}_{\text{r}} = \bar{\mathbf{\Theta}}\mathbf{H}_{\text{e}}$, we note that

$$\begin{aligned} R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) &= I(\mathbf{x};\mathbf{y}_{\text{r}}|\mathbf{y}_{\text{e}}) && (90) \\ &= h(\mathbf{y}_{\text{r}}|\mathbf{y}_{\text{e}}) - h(\mathbf{z}_{\text{r}}|\mathbf{z}_{\text{e}}) \\ &= h(\mathbf{y}_{\text{r}} - \bar{\mathbf{\Theta}}\mathbf{y}_{\text{e}}) - h(\mathbf{z}_{\text{r}} - \bar{\mathbf{\Phi}}\mathbf{z}_{\text{e}}) && (91) \\ &= h(\mathbf{z}_{\text{r}} - \bar{\mathbf{\Theta}}\mathbf{z}_{\text{e}}) - h(\mathbf{z}_{\text{r}} - \bar{\mathbf{\Phi}}\mathbf{z}_{\text{e}}) && (92) \\ &= 0 && (93) \end{aligned}$$

where (91) follows from the fact that $\bar{\mathbf{\Theta}}$ in (17) is the coefficient in the MMSE estimate of $\mathbf{y}_{\text{r}}$ from $\mathbf{y}_{\text{e}}$, and $\bar{\mathbf{\Phi}}$ is the coefficient in the MMSE estimate of $\mathbf{z}_{\text{r}}$ from $\mathbf{z}_{\text{e}}$, where (92) follows via the relation $\mathbf{H}_{\text{r}} = \bar{\mathbf{\Theta}}\mathbf{H}_{\text{e}}$, so that $\mathbf{y}_{\text{r}} - \bar{\mathbf{\Theta}}\mathbf{y}_{\text{e}} = \mathbf{z}_{\text{r}} - \bar{\mathbf{\Theta}}\mathbf{z}_{\text{e}}$, and where (93) follows from (23). ■

VI. Capacity Analysis in the High-SNR Regime

We begin with a convenient upper bound that is used in our converse argument, then exploit the GSVD in developing the coding scheme for our achievability argument. We separately consider the cases where $\mathbf{H}_{\text{e}}$ does and does not have full column-rank.

Lemma 6: For all choices of $\boldsymbol{\Theta} \in \mathbb{C}^{n_\mathrm{r} \times n_\mathrm{t}}$ and $\boldsymbol{\Phi} \in \mathbb{C}^{n_\mathrm{r} \times n_\mathrm{e}}$ such that $\|\boldsymbol{\Phi}\|_2 \leq 1$, the secrecy capacity (35) of the channel (1) is upper bounded by

$$C(P) \leq \max_{\bar{\mathbf{K}}_P \in \mathcal{K}_P} R_{++}(\mathbf{K}_P, \boldsymbol{\Theta}, \boldsymbol{\Phi}) \tag{94a}$$

where

$$\begin{aligned} &R_{++}(\mathbf{K}_P, \boldsymbol{\Theta}, \boldsymbol{\Phi}) \\ &\quad \triangleq h(\mathbf{y}_\mathrm{r} - \boldsymbol{\Theta}\mathbf{y}_\mathrm{e}) - \log\det \pi e(\mathbf{I} - \boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger) \\ &\quad = \log \frac{\det(\hat{\mathbf{H}}\mathbf{K}_P\hat{\mathbf{H}}^\dagger + \mathbf{I} + \boldsymbol{\Theta}\boldsymbol{\Theta}^\dagger - \boldsymbol{\Theta}\boldsymbol{\Phi}^\dagger - \boldsymbol{\Phi}\boldsymbol{\Theta}^\dagger)}{\det(\mathbf{I} - \boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger)} \end{aligned} \tag{94b}$$

with

$$\hat{\mathbf{H}} = \mathbf{H}_\mathrm{r} - \boldsymbol{\Theta}\mathbf{H}_\mathrm{e}. \tag{94c}$$

Proof: First note that the objective function $R_+(\mathbf{K}_P, \mathbf{K}_{\boldsymbol{\Phi}})$ in (11) can be expressed in the form

$$\begin{aligned} R_+(\mathbf{K}_P, \mathbf{K}_{\boldsymbol{\Phi}}) &= I(\mathbf{x}; \mathbf{y}_\mathrm{r} | \mathbf{y}_\mathrm{e}) \\ &= h(\mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e}) - h(\mathbf{z}_\mathrm{r}|\mathbf{z}_\mathrm{e}) \\ &= h(\mathbf{y}_\mathrm{r}|\mathbf{y}_\mathrm{e}) - \log\det \pi e(\mathbf{I} - \boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger) \\ &= \min_{\boldsymbol{\Theta}} h(\mathbf{y}_\mathrm{r} - \boldsymbol{\Theta}\mathbf{y}_\mathrm{e}) - \log\det \pi e(\mathbf{I} - \boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger) \\ &= \min_{\boldsymbol{\Theta}} R_{++}(\mathbf{K}_P, \boldsymbol{\Theta}, \boldsymbol{\Phi}) \end{aligned} \tag{95}$$

Hence

$$\begin{aligned} R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\boldsymbol{\Phi}}) &= \min_{\mathcal{K}_{\boldsymbol{\Phi}}} \max_{\mathcal{K}_P} R_+(\mathbf{K}_P, \mathbf{K}_{\boldsymbol{\Phi}}) && (96) \\ &= \min_{\mathcal{K}_{\boldsymbol{\Phi}}} \max_{\mathcal{K}_P} \min_{\boldsymbol{\Theta}} R_{++}(\mathbf{K}_P, \boldsymbol{\Theta}, \bar{\boldsymbol{\Phi}}) && (97) \\ &\leq \min_{\mathcal{K}_{\boldsymbol{\Phi}}} \min_{\boldsymbol{\Theta}} \max_{\mathcal{K}_P} R_{++}(\mathbf{K}_P, \boldsymbol{\Theta}, \boldsymbol{\Phi}) && (98) \\ &= \min_{\boldsymbol{\Phi}: \|\boldsymbol{\Phi}\|_2 \leq 1} \min_{\boldsymbol{\Theta}} \max_{\mathbf{K}_P \in \mathcal{K}_P} R_{++}(\mathbf{K}_P, \boldsymbol{\Theta}, \boldsymbol{\Phi}) && (99) \end{aligned}$$

where to obtain (96) we have used (35), where to obtain (97) we have used (95), and where to obtain (98) we have used that a minimax quantity upper bounds a corresponding maximin quantity.

Finally, we further upper bound (99) by making arbitrary choices for $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$, yielding (94). ■

A. GSVD Properties

The following properties of the GSVD in Definition 1 are useful in our analysis.

First, the GSVD simultaneously diagonalizes the channels in our model (1). In particular, applying (6) we obtain

$$\begin{aligned} \tilde{\mathbf{y}}_\mathrm{r}(t) &= \tilde{\boldsymbol{\Sigma}}_\mathrm{r}\tilde{\mathbf{x}}(t) + \tilde{\mathbf{z}}_\mathrm{r}(t) \\ \tilde{\mathbf{y}}_\mathrm{e}(t) &= \tilde{\boldsymbol{\Sigma}}_\mathrm{e}\tilde{\mathbf{x}}(t) + \tilde{\mathbf{z}}_\mathrm{e}(t) \end{aligned} \tag{100}$$

where

$$\tilde{\boldsymbol{\Sigma}}_\mathrm{r} = \begin{array}{c} \\ s \\ p \end{array} \begin{array}{c} \begin{array}{ccc} k-p-s & s & p \end{array} \\ \begin{bmatrix} \mathbf{0} & \mathbf{D}_\mathrm{r} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} \end{array}$$

$$\tilde{\boldsymbol{\Sigma}}_\mathrm{e} = \begin{array}{c} \\ k-p-s \\ s \end{array} \begin{array}{c} \begin{array}{ccc} k-p-s & s & p \end{array} \\ \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_\mathrm{e} & \mathbf{0} \end{bmatrix} \end{array}.$$

and

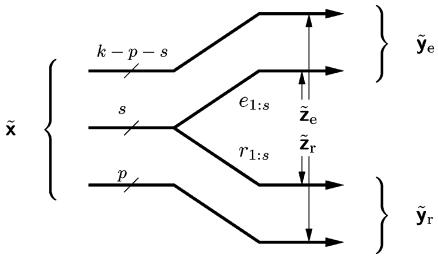

Fig. 4. Equivalent parallel channel model obtained via GSVD.

$$\begin{aligned} \tilde{\mathbf{x}}(t) &= \boldsymbol{\Omega}^{-1}\left[\boldsymbol{\Psi}_\mathrm{t}^\dagger \mathbf{x}(t)\right]_{1:k} \\ \tilde{\mathbf{y}}_\mathrm{r}(t) &= \left[\boldsymbol{\Psi}_\mathrm{r}^\dagger \mathbf{y}_\mathrm{r}(t)\right]_{n_\mathrm{r}-p-s+1:n_\mathrm{r}} \\ \tilde{\mathbf{y}}_\mathrm{e}(t) &= \left[\boldsymbol{\Psi}_\mathrm{e}^\dagger \mathbf{y}_\mathrm{e}(t)\right]_{1:k-p} \\ \tilde{\mathbf{z}}_\mathrm{r}(t) &= \left[\boldsymbol{\Psi}_\mathrm{r}^\dagger \mathbf{z}_\mathrm{r}(t)\right]_{n_\mathrm{r}-p-s+1:n_\mathrm{r}} \\ \tilde{\mathbf{z}}_\mathrm{e}(t) &= \left[\boldsymbol{\Psi}_\mathrm{e}^\dagger \mathbf{z}_\mathrm{e}(t)\right]_{1:k-p}. \end{aligned}$$

The corresponding equivalent channel is as depicted in Fig. 4.

Second, the GSVD yields a characterization of the null space of $\mathbf{H}_\mathrm{e}$. In particular,

$$\mathrm{Null}\,(\mathbf{H}_\mathrm{e}) = \mathcal{S}_\mathrm{r} \cup \mathcal{S}_n \tag{101}$$

where, expressing $\boldsymbol{\Psi}_\mathrm{t}$ as defined in (6) in terms of its columns $\boldsymbol{\psi}_i$, $i = 1, \ldots, n_\mathrm{t}$, viz.,

$$\boldsymbol{\Psi}_\mathrm{t} = [\,\boldsymbol{\psi}_1 \quad \cdots \quad \boldsymbol{\psi}_{n_\mathrm{t}}\,]$$

we have [cf. (2a), (2d)]

$$\mathcal{S}_\mathrm{r} = \mathrm{span}\,(\boldsymbol{\psi}_{k-p+1}, \ldots, \boldsymbol{\psi}_k) \tag{102a}$$

$$\mathcal{S}_n = \mathrm{span}\,(\boldsymbol{\psi}_{k+1}, \ldots, \boldsymbol{\psi}_{n_\mathrm{t}}). \tag{102b}$$

We first verify (102). To establish (102b), it suffices to note that

$$\mathbf{H}_\mathrm{r}\boldsymbol{\psi}_j = \mathbf{H}_\mathrm{e}\boldsymbol{\psi}_j = \mathbf{0}, \quad j = k+1, \ldots, n_\mathrm{t}$$

which can be readily verified from (6).

To establish (102a), we show for all $j \in \{k-p+1, \ldots, k\}$ that $\mathbf{H}_\mathrm{e}\boldsymbol{\psi}_j = \mathbf{0}$ and that the $\{\mathbf{H}_\mathrm{r}\boldsymbol{\psi}_j\}$ are linearly independent. It suffices to show that the last p columns of $\boldsymbol{\Sigma}_\mathrm{r}\boldsymbol{\Omega}^{-1}$ are linearly independent and the last p columns of $\boldsymbol{\Sigma}_\mathrm{e}\boldsymbol{\Omega}^{-1}$ are zero. To this end, note that since $\boldsymbol{\Omega}^{-1}$ in (6) is a lower triangular matrix, it can be expressed in the form

$$\boldsymbol{\Omega}^{-1} = \begin{array}{c} \\ k-p-s \\ s \\ p \end{array} \begin{array}{c} \begin{array}{ccc} k-p-s & s & p \end{array} \\ \begin{bmatrix} \boldsymbol{\Omega}_1^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{T}_{21} & \boldsymbol{\Omega}_2^{-1} & \mathbf{0} \\ \mathbf{T}_{31} & \mathbf{T}_{32} & \boldsymbol{\Omega}_3^{-1} \end{bmatrix} \end{array}. \tag{103}$$

By direct block left-multiplication of (103) with (7a) and (7b), we have

$$\boldsymbol{\Sigma}_\mathrm{r}\boldsymbol{\Omega}^{-1} = \begin{array}{c} \\ k-p-s \\ s \\ p \end{array} \begin{array}{c} \begin{array}{ccc} k-p-s & s & p \end{array} \\ \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{D}_\mathrm{r}\mathbf{T}_{21} & \mathbf{D}_\mathrm{r}\boldsymbol{\Omega}_2^{-1} & \mathbf{0} \\ \mathbf{T}_{31} & \mathbf{T}_{32} & \boldsymbol{\Omega}_3^{-1} \end{bmatrix} \end{array} \tag{104a}$$

$$\boldsymbol{\Sigma}_\mathrm{e}\boldsymbol{\Omega}^{-1} = \begin{array}{c} \\ k-p-s \\ s \\ p \end{array} \begin{array}{c} \begin{array}{ccc} k-p-s & s & p \end{array} \\ \begin{bmatrix} \boldsymbol{\Omega}_1^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{D}_\mathrm{e}\mathbf{T}_{21} & \mathbf{D}_\mathrm{e}\boldsymbol{\Omega}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \end{array}. \tag{104b}$$

Since $\mathbf{\Omega}_3$ is invertible (since $\mathbf{\Omega}$ is nonsingular), the last p columns of $\mathbf{\Sigma}_\mathrm{r}\mathbf{\Omega}^{-1}$ are linearly independent and the last p columns of $\mathbf{\Sigma}_\mathrm{e}\mathbf{\Omega}^{-1}$ are zero, establishing (102a).

To characterize $\mathrm{Null}\,(\mathbf{H}_\mathrm{e})$, we use (102a) and (102b) with (101) to obtain

$$\mathrm{Null}\,(\mathbf{H}_\mathrm{e}) = \mathrm{span}\,(\boldsymbol{\psi}_{k-p+1}, \ldots, \boldsymbol{\psi}_{n_\mathrm{t}})$$

from which we obtain that

$$\mathbf{H}_\mathrm{e}^\sharp = \mathbf{\Psi}_\mathrm{ne}\mathbf{\Psi}_\mathrm{ne}^\dagger \tag{105}$$

is the projection matrix onto $\mathrm{Null}\,(\mathbf{H}_\mathrm{e})$, where

$$\mathbf{\Psi}_\mathrm{ne} = [\,\boldsymbol{\psi}_{k-p+1} \quad \cdots \quad \boldsymbol{\psi}_{n_\mathrm{t}}\,]. \tag{106}$$

In turn, using (106) and (104a) in (6a) we obtain

$$\mathbf{H}_\mathrm{r}\mathbf{\Psi}_\mathrm{ne} = \mathbf{\Psi}_\mathrm{r} \begin{Bmatrix} & \begin{matrix} p & n_\mathrm{t}-k \end{matrix} \\ \begin{matrix} n_\mathrm{r}-p \\ p \end{matrix} & \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{\Omega}_3^{-1} & \mathbf{0} \end{bmatrix} \end{Bmatrix}$$

whence

$$\mathbf{H}_\mathrm{r}\mathbf{H}_\mathrm{e}^\sharp\mathbf{H}_\mathrm{r}^\dagger = \mathbf{\Psi}_\mathrm{r} \begin{Bmatrix} & \begin{matrix} n_\mathrm{r}-p & p \end{matrix} \\ \begin{matrix} n_\mathrm{r}-p \\ p \end{matrix} & \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega}_3^{-1}\mathbf{\Omega}_3^{-\dagger} \end{bmatrix} \end{Bmatrix} \mathbf{\Psi}_\mathrm{r}^\dagger. \tag{107}$$

Third, the GSVD can be more simply described when the matrix $\mathbf{H}_\mathrm{e}$ has a full column-rank. To see this, first note from (3) and (5) that

$$k = n_\mathrm{t} \quad \text{and} \quad p = 0 \tag{108}$$

respectively, and thus (6) specializes to

$$\mathbf{\Psi}_\mathrm{r}^\dagger\mathbf{H}_\mathrm{r}\mathbf{\Psi}_\mathrm{t}\mathbf{\Omega} = \mathbf{\Sigma}_\mathrm{r}, \quad \mathbf{\Psi}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e}\mathbf{\Psi}_\mathrm{t}\mathbf{\Omega} = \mathbf{\Sigma}_\mathrm{e} \tag{109a}$$

with [cf. (7)]

$$\mathbf{\Sigma}_\mathrm{r} = \begin{matrix} n_\mathrm{r}-s \\ s \end{matrix} \begin{bmatrix} \overset{n_\mathrm{t}-s}{\mathbf{0}} & \overset{s}{\mathbf{0}} \\ \mathbf{0} & \mathbf{D}_\mathrm{r} \end{bmatrix}, \quad \mathbf{\Sigma}_\mathrm{e} = \begin{matrix} n_\mathrm{t}-s \\ s \\ n_\mathrm{e}-n_\mathrm{t} \end{matrix} \begin{bmatrix} \overset{n_\mathrm{t}-s}{\mathbf{I}} & \overset{s}{\mathbf{0}} \\ \mathbf{0} & \mathbf{D}_\mathrm{e} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \tag{109b}$$

and $\mathbf{D}_\mathrm{r}$ and $\mathbf{D}_\mathrm{e}$ as in (8). Hence, it follows from (109) that

$$\mathbf{H}_\mathrm{e}^\ddagger \triangleq \mathbf{\Psi}_\mathrm{t}\mathbf{\Omega} \begin{Bmatrix} & \begin{matrix} n_\mathrm{t}-s & s & n_\mathrm{e}-n_\mathrm{t} \end{matrix} \\ \begin{matrix} n_\mathrm{t}-s \\ s \end{matrix} & \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_\mathrm{e}^{-1} & \mathbf{0} \end{bmatrix} \end{Bmatrix} \mathbf{\Psi}_\mathrm{e}^\dagger \tag{110}$$

satisfies $\mathbf{H}_\mathrm{e}^\ddagger\mathbf{H}_\mathrm{e} = \mathbf{I}$ and thus is the Moore-Penrose pseudo-inverse of $\mathbf{H}_\mathrm{e}$. Finally, from (109) and (110) we obtain

$$\mathbf{H}_\mathrm{r}\mathbf{H}_\mathrm{e}^\ddagger = \mathbf{\Psi}_\mathrm{r} \begin{Bmatrix} & \begin{matrix} n_\mathrm{t}-s & s & n_\mathrm{e}-n_\mathrm{t} \end{matrix} \\ \begin{matrix} n_\mathrm{r}-s \\ s \end{matrix} & \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_\mathrm{r}\mathbf{D}_\mathrm{e}^{-1} & \mathbf{0} \end{bmatrix} \end{Bmatrix} \mathbf{\Psi}_\mathrm{e}^\dagger$$

from which we see that the generalized singular values of $(\mathbf{H}_\mathrm{r}, \mathbf{H}_\mathrm{e})$ in (9) are also the (ordinary) singular values of $\mathbf{H}_\mathrm{r}\mathbf{H}_\mathrm{e}^\ddagger$.

We now turn to our secrecy capacity analysis in the high-SNR regime. There are two cases, which we consider separately.

B. *Case I:* $\mathrm{rank}\,(\mathbf{H}_\mathrm{e}) = n_\mathrm{t}$

In this case, we use that (108) holds and so the GSVD is given by (109), and thus $\dim\mathcal{S}_\mathrm{r,e} = s$, $\dim\mathcal{S}_\mathrm{e} = n_\mathrm{t} - s$, and $\dim\mathcal{S}_\mathrm{r} = \dim\mathcal{S}_n = 0$.

Achievability: In the equivalent parallel channel model of Fig. 4, there are s subchannels that go to the intended receiver (and also to the eavesdropper, with different gains), which correspond to $\mathcal{S}_\mathrm{r,e}$. Of these s subchannels, we use only the subset for which the gains to the intended receiver are stronger than those to the eavesdropper, and with these our communication scheme uses Gaussian wiretap codebooks.

In particular, we transmit

$$\mathsf{x} = \mathbf{\Psi}_\mathrm{t}\mathbf{\Omega}\begin{bmatrix} \mathbf{0}_{n_\mathrm{t}-s} \\ \mathsf{u} \end{bmatrix}, \quad \mathsf{u} = [0, \ldots, 0, u_\nu, u_{\nu+1}, \ldots, u_s] \tag{111}$$

where ν is the smallest integer such that $\sigma_j > 1$, and where the nonzero elements of u are i.i.d. $\mathcal{CN}(0, \alpha P)$ with $\alpha = 1/(n_\mathrm{t}\sigma_{\max}(\mathbf{\Omega}))$ so that the transmitted power is at most P. Using (111) and (109) in (1), the observations at the intended receiver and eavesdropper, respectively, take the form

$$\mathsf{y}_\mathrm{r} = \mathbf{\Psi}_\mathrm{r}\begin{bmatrix} \mathbf{0}_{n_\mathrm{t}-s} \\ \mathbf{D}_\mathrm{r}\mathsf{u} \end{bmatrix} + \mathsf{z}_\mathrm{r}, \quad \mathsf{y}_\mathrm{e} = \mathbf{\Psi}_\mathrm{e}\begin{bmatrix} \mathbf{0}_{n_\mathrm{t}-s} \\ \mathbf{D}_\mathrm{e}\mathsf{u} \\ \mathbf{0}_{n_\mathrm{e}-n_\mathrm{t}} \end{bmatrix} + \mathsf{z}_\mathrm{e}.$$

In turn, via (19), the (secrecy) rate achievable with this system is

$$\begin{aligned} R &= I(\mathsf{u};\mathsf{y}_\mathrm{r}) - I(\mathsf{u};\mathsf{y}_\mathrm{e}) \\ &= \sum_{j=\nu}^{n_\mathrm{t}} \log\frac{1+\alpha P r_j^2}{1+\alpha P e_j^2} \\ &= \sum_{j:\sigma_j>1} \log\sigma_j^2 - o(1) \end{aligned}$$

as required. ■

Converse: It suffices to use Lemma 6 with the choices

$$\mathbf{\Theta} = \mathbf{H}_\mathrm{r}\mathbf{H}_\mathrm{e}^\ddagger, \quad \mathbf{\Phi} = \mathbf{\Psi}_\mathrm{r} \begin{Bmatrix} & \begin{matrix} n_\mathrm{t}-s & s & n_\mathrm{e}-n_\mathrm{t} \end{matrix} \\ \begin{matrix} n_\mathrm{r}-s \\ s \end{matrix} & \begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Xi} & \mathbf{0} \end{bmatrix} \end{Bmatrix} \mathbf{\Psi}_\mathrm{e}^\dagger \tag{112a}$$

where

$$\mathbf{\Xi} = \mathrm{diag}\,(\xi_1, \xi_2, \ldots, \xi_s), \qquad \xi_i = \min\left(\sigma_i, \frac{1}{\sigma_i}\right) \tag{112b}$$

and where $\mathbf{H}_\mathrm{e}^\ddagger$ is the pseudo-inverse defined in (110). With these choices of parameters, (94c) evaluates to $\hat{\mathbf{H}} = \mathbf{0}$, so we can ignore the maximization over $\mathbf{K}_P$ in (94a). Simplifying (94) for our choice of parameters yields

$$\begin{aligned} R_{++} &\le \log\frac{\det(\mathbf{I} + (\mathbf{D}_\mathrm{r}\mathbf{D}_\mathrm{e}^{-1})^2 - 2\mathbf{D}_\mathrm{r}\mathbf{D}_\mathrm{e}^{-1}\mathbf{\Xi})}{\det(\mathbf{I} - \mathbf{\Xi}^2)} \\ &= \sum_{j:\sigma_j>1} \log\sigma_j^2 \end{aligned} \tag{113}$$

which establishes our result. ■

C. Case II: $\mathrm{rank}\,(\mathbf{H}_\mathrm{e}) < n_t$

In this case, we use the general form of the GSVD as given by (6), so now $\dim \mathcal{S}_\mathrm{r} = p > 0$ and $\dim \mathcal{S}_\mathrm{r,e} = s > 0$.

Achievability: In the equivalent parallel channel model of Fig. 4, there are p subchannels that go only to the intended receiver, corresponding to $\mathcal{S}_\mathrm{r}$, and s subchannels that go to both the intended receiver and eavesdropper (with different gains), corresponding to $\mathcal{S}_\mathrm{r,e}$. Our communication scheme uses both sets of subchannels independently with Gaussian (wiretap) codebooks.

In particular, we transmit

$$\mathsf{x} = \mathbf{\Psi}_\mathrm{t} \begin{bmatrix} \mathbf{0}_{k-p-s} \\ \mathbf{\Omega}_2 \mathsf{u} \\ \mathsf{v} \\ \mathbf{0}_{n_\mathrm{t}-k} \end{bmatrix} \tag{114}$$

where v and u are the length-p and length-s auxiliary random vectors associated with communication over $\mathcal{S}_\mathrm{r}$ and $\mathcal{S}_\mathrm{r,e}$, respectively. The elements of v are i.i.d. $\mathcal{CN}(0, (P - \sqrt{P})/p)$, corresponding to allocating power $P - \sqrt{P}$ to $\mathcal{S}_\mathrm{r}$. For $\mathcal{S}_\mathrm{r,e}$, we use only the subset of channels for which the gains to the intended receiver are stronger than those to the eavesdropper, so $\mathsf{u} = [0, \ldots, 0, u_\nu, \ldots, u_s]^\mathrm{T}$, where ν is the smallest integer such that $\sigma_j > 1$, and where the nonzero elements are i.i.d. $\mathcal{CN}(0, \alpha\sqrt{P})$, independent of v, with $\alpha = 1/(n_\mathrm{t}\sigma_{\max}(\mathbf{\Omega}_2))$ so that the power allocated to $\mathcal{S}_\mathrm{r,e}$ is at most $\sqrt{P}$.

With x as in (114), the observations at the intended receiver and eavesdropper, respectively, take the form

$$\mathsf{y}_\mathrm{r} = \mathbf{\Psi}_\mathrm{r} \begin{bmatrix} \mathbf{0}_{n_\mathrm{r}-p-s} \\ \mathbf{D}_\mathrm{r}\mathsf{u} \\ \mathbf{T}_{32}\mathbf{\Omega}_2^{-1}\mathsf{u} + \mathbf{\Omega}_3^{-1}\mathsf{v} \end{bmatrix} + \mathsf{z}_\mathrm{r}, \tag{115a}$$

$$\mathsf{y}_\mathrm{e} = \mathbf{\Psi}_\mathrm{e} \begin{bmatrix} \mathbf{0}_{k-p-s} \\ \mathbf{D}_\mathrm{e}\mathsf{u} \\ \mathbf{0}_{n_\mathrm{e}+p-k} \end{bmatrix} + \mathsf{z}_\mathrm{e}. \tag{115b}$$

Via (19), the system (115) achieves (secrecy) rate

$$\begin{aligned} R &= I(\mathsf{u}, \mathsf{v}; \mathsf{y}_\mathrm{r}) - I(\mathsf{u}, \mathsf{v}; \mathsf{y}_\mathrm{e}) \\ &= I(\mathsf{u}; \mathsf{y}_\mathrm{r}) - I(\mathsf{u}; \mathsf{y}_\mathrm{e}) + I(\mathsf{v}; \mathsf{y}_\mathrm{r}|\mathsf{u}) \end{aligned} \tag{116}$$

where (116) follows from the fact that v is independent of $(\mathsf{y}_\mathrm{e}, \mathsf{u})$, as (115b) reflects.

Evaluating the terms in (116), we obtain

$$\begin{aligned} I(\mathsf{u}; \mathsf{y}_\mathrm{r}) - I(\mathsf{u}; \mathsf{y}_\mathrm{e}) &= \sum_{j=\nu}^{n_\mathrm{t}} \log \frac{1 + \alpha\sqrt{P} r_j^2}{1 + \alpha\sqrt{P} e_j^2} \\ &= \sum_{j:\sigma_j > 1} \log \sigma_j^2 - o(1) \end{aligned} \tag{117}$$

and

$$\begin{aligned} I(\mathsf{v}; \mathsf{y}_\mathrm{r}|\mathsf{u}) &= \log\det\left(\mathbf{I} + \frac{P - \sqrt{P}}{p}\mathbf{\Omega}_3^{-1}\mathbf{\Omega}_3^{-\dagger}\right) \\ &= \log\det\left(\mathbf{I} + \frac{P}{p}\mathbf{\Omega}_3^{-1}\mathbf{\Omega}_3^{-\dagger}\right) - o(1) \quad (118) \\ &= \log\det\left(\mathbf{I} + \frac{P}{p}\mathbf{H}_\mathrm{r}\mathbf{H}_\mathrm{e}^\sharp\mathbf{H}_\mathrm{r}^\dagger\right) - o(1) \quad (119) \end{aligned}$$

where (118) follows from the continuity of $\log\det(\cdot)$, and where (119) follows from (107). Substituting (117) and (119) into (116) yields our desired result. ∎

Converse: To establish the converse, we use Lemma 6 with the choices

$$\mathbf{\Theta} = \mathbf{\Psi}_\mathrm{r} \begin{Bmatrix} \\ n_\mathrm{r}-s-p \\ s \\ p \end{Bmatrix} \overset{\begin{matrix} k-s-p & s & n_\mathrm{e}+p-k \end{matrix}}{\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_\mathrm{r}\mathbf{D}_\mathrm{e}^{-1} & \mathbf{0} \\ \mathbf{F}_{31} & \mathbf{F}_{32} & \mathbf{0} \end{bmatrix}} \Bigg\} \mathbf{\Psi}_\mathrm{e}^\dagger \tag{120}$$

and

$$\mathbf{\Phi} = \mathbf{\Psi}_\mathrm{r} \begin{Bmatrix} \\ n_\mathrm{r}-s-p \\ s \\ p \end{Bmatrix} \overset{\begin{matrix} k-s-p & s & n_\mathrm{e}+p-k \end{matrix}}{\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Xi} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}} \Bigg\} \mathbf{\Psi}_\mathrm{e}^\dagger \tag{121}$$

where $\mathbf{\Xi}$ is as defined in (112b), and where we choose

$$\begin{aligned} \mathbf{F}_{32} &= \mathbf{T}_{32}\mathbf{\Omega}_2\mathbf{D}_\mathrm{e}^{-1} \\ \mathbf{F}_{31} &= (\mathbf{T}_{31} - \mathbf{F}_{32}\mathbf{D}_\mathrm{e}\mathbf{T}_{21})\mathbf{\Omega}_1 \end{aligned}$$

with $\mathbf{T}_{21}$, $\mathbf{T}_{31}$, and $\mathbf{T}_{32}$ as defined in (103), so that

$$\begin{aligned} &\mathbf{H}_\mathrm{r} - \mathbf{\Theta}\mathbf{H}_\mathrm{e} \\ &= \mathbf{\Psi}_\mathrm{r}\big([\,\mathbf{\Sigma}_\mathrm{r}\mathbf{\Omega}^{-1} \quad \mathbf{0}_{n_\mathrm{r}\times n_\mathrm{t}-k}\,] \\ &\quad - \mathbf{\Psi}_\mathrm{r}^\dagger\mathbf{\Theta}\mathbf{\Psi}_\mathrm{e}\,[\,\mathbf{\Sigma}_\mathrm{e}\mathbf{\Omega}^{-1} \quad \mathbf{0}_{n_\mathrm{e}\times n_\mathrm{t}-k}\,]\big)\mathbf{\Psi}_\mathrm{t}^\dagger \\ &= \mathbf{\Psi}_\mathrm{r} \begin{Bmatrix} \\ n_\mathrm{r}-s-p \\ s \\ p \end{Bmatrix} \overset{\begin{matrix} k-p-s & s & p & n_\mathrm{t}-k \end{matrix}}{\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Omega}_3^{-1} & \mathbf{0} \end{bmatrix}} \Bigg\} \mathbf{\Psi}_\mathrm{t}^\dagger. \end{aligned}$$

The upper bound expression (94) can now be simplified as follows:

$$\begin{aligned} \hat{\mathbf{H}}\mathbf{K}_P\hat{\mathbf{H}}^\dagger &= (\mathbf{H}_\mathrm{r} - \mathbf{\Theta}\mathbf{H}_\mathrm{e})\mathbf{K}_P(\mathbf{H}_\mathrm{r} - \mathbf{\Theta}\mathbf{H}_\mathrm{e})^\dagger \\ &= \mathbf{\Psi}_\mathrm{r} \begin{Bmatrix} \\ n_\mathrm{r}-p-s \\ s \\ p \end{Bmatrix} \overset{\begin{matrix} n_\mathrm{r}-p-s & s & p \end{matrix}}{\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Omega}_3^{-1}\mathbf{Q}\mathbf{\Omega}_3^{-\dagger} \end{bmatrix}} \Bigg\} \mathbf{\Psi}_\mathrm{r}^\dagger \end{aligned} \tag{122}$$

where $\mathbf{Q}$ is related to $\mathbf{K}_P$ via

$$\mathbf{\Psi}_\mathrm{t}^\dagger\mathbf{K}_P\mathbf{\Psi}_t = \begin{matrix} \\ k-p \\ p \\ n_\mathrm{t}-k \end{matrix} \overset{\begin{matrix} k-p & p & n_\mathrm{t}-k \end{matrix}}{\begin{bmatrix} & & \\ & \mathbf{Q} & \\ & & \end{bmatrix}}$$

and satisfies $\mathrm{tr}\,(\mathbf{Q}) \le P$. From (122), (121) and (120), we have that the numerator in the right-hand side of (94b) simplifies to (123), shown at the bottom of the next page.

In turn, using (123) and the Fischer inequality (which generalizes Hadamard's inequality) for positive semidefinite matrices [30], we obtain

$$\begin{aligned} &\log\det\left(\mathbf{I} + \hat{\mathbf{H}}\mathbf{\Theta}\hat{\mathbf{H}}^\dagger + \mathbf{\Theta}\mathbf{\Theta}^\dagger - \mathbf{\Theta}\mathbf{\Phi}^\dagger - \mathbf{\Phi}\mathbf{\Theta}^\dagger\right) \\ &\quad \le \log\det(\mathbf{I} + (\mathbf{D}_\mathrm{r}\mathbf{D}_\mathrm{e}^{-1})^2 - 2\mathbf{D}_\mathrm{r}\mathbf{D}_\mathrm{e}^{-1}\mathbf{\Xi}) \\ &\qquad + \log\det(\mathbf{I} + \mathbf{F}_{31}\mathbf{F}_{31}^\dagger + \mathbf{F}_{32}\mathbf{F}_{32}^\dagger + \mathbf{\Omega}_3^{-1}\mathbf{Q}\mathbf{\Omega}_3^{-\dagger}) \end{aligned}$$

which when used with (94) yields

$$C(P) \le \log \frac{\det(\mathbf{I}+(\mathbf{D}_{\mathrm{r}}\mathbf{D}_{\mathrm{e}}^{-1})^2-2\mathbf{D}_{\mathrm{r}}\mathbf{D}_{\mathrm{e}}^{-1}\boldsymbol{\Xi})}{\det(\mathbf{I}-\boldsymbol{\Xi}^2)} + \max_{\substack{\mathbf{Q}\succeq\mathbf{0}:\\ \mathrm{tr}(\mathbf{Q})\le P}} \log\det(\mathbf{I}+\mathbf{F}_{31}\mathbf{F}_{31}^{\dagger}+\mathbf{F}_{32}\mathbf{F}_{32}^{\dagger}+\boldsymbol{\Omega}_3^{-1}\mathbf{Q}\boldsymbol{\Omega}_3^{-\dagger})$$

the first term of which is identical to (113). Thus, it remains only to establish that

$$\max_{\substack{\mathbf{Q}\succeq\mathbf{0}:\\ \mathrm{tr}(\mathbf{Q})\le P}} \log\det(\mathbf{I}+\mathbf{F}_{31}\mathbf{F}_{31}^{\dagger}+\mathbf{F}_{32}\mathbf{F}_{32}^{\dagger}+\boldsymbol{\Omega}_3^{-1}\mathbf{Q}\boldsymbol{\Omega}_3^{-\dagger}) \le \log\det\left(\mathbf{I}+\frac{P}{p}\mathbf{H}_{\mathrm{r}}\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{H}_{\mathrm{r}}^{\dagger}\right)+o(1). \quad (124)$$

To obtain (124), let

$$\gamma=\sigma_{\max}(\mathbf{F}_{31}\mathbf{F}_{31}^{\dagger}+\mathbf{F}_{32}\mathbf{F}_{32}^{\dagger}) \quad (125)$$

denote the largest singular value of the matrix $\mathbf{F}_{31}\mathbf{F}_{31}^{\dagger}+\mathbf{F}_{32}\mathbf{F}_{32}^{\dagger}$. Since $\log\det(\cdot)$ is increasing on the cone of positive semidefinite matrices, we have

$$\begin{aligned}
&\max_{\substack{\mathbf{Q}\succeq\mathbf{0}:\\ \mathrm{tr}(\mathbf{Q})\le P}} \log\det(\mathbf{I}+\mathbf{F}_{31}\mathbf{F}_{31}^{\dagger}+\mathbf{F}_{32}\mathbf{F}_{32}^{\dagger}+\boldsymbol{\Omega}_3^{-1}\mathbf{Q}\boldsymbol{\Omega}_3^{-\dagger}) \\
&\quad\le \max_{\substack{\mathbf{Q}\succeq\mathbf{0}:\\ \mathrm{tr}(\mathbf{Q})\le P}} \log\det((1+\gamma)\mathbf{I}+\boldsymbol{\Omega}_3^{-1}\mathbf{Q}\boldsymbol{\Omega}_3^{-\dagger}) \quad (126)\\
&\quad= \log\det\left((1+\gamma)\mathbf{I}+\frac{P}{p}\boldsymbol{\Omega}_3^{-1}\boldsymbol{\Omega}_3^{-\dagger}\right)+o(1) \quad (127)\\
&\quad= \log\det\left(\mathbf{I}+\frac{P}{p}\boldsymbol{\Omega}_3^{-1}\boldsymbol{\Omega}_3^{-\dagger}\right)+o(1)\\
&\quad= \log\det\left(\mathbf{I}+\frac{P}{p}\mathbf{H}_{\mathrm{r}}\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{H}_{\mathrm{r}}^{\dagger}\right)+o(1) \quad (128)
\end{aligned}$$

where (126) follows from the fact that $\gamma\mathbf{I}-\mathbf{F}_{31}\mathbf{F}_{31}^{\dagger}-\mathbf{F}_{32}\mathbf{F}_{32}^{\dagger}\succeq\mathbf{0}$, and (127) follows from the fact that water-filling provides a vanishingly small gain over flat power allocation when the channel matrix has a full rank (see, e.g., [31]), and (128) follows from (107).

D. Analysis of the Masked MIMO Transmission Scheme

To establish (29), we focus on the two terms in the argument of (19), obtaining

$$I(\mathbf{u};\mathbf{y}_{\mathrm{r}})=\log\det(\mathbf{I}+P_{\mathrm{t}}\mathbf{H}_{\mathrm{r}}\mathbf{H}_{\mathrm{r}}^{\dagger})=\log\det(\mathbf{I}+P_{\mathrm{t}}\boldsymbol{\Delta}^2) \quad (129)$$

where we have used (27) to obtain the second equality, and

$$I(\mathbf{u};\mathbf{y}_{\mathrm{e}})=h(\mathbf{y}_{\mathrm{e}})-h(\mathbf{y}_{\mathrm{e}}|\mathbf{u}) \quad (130)$$

with

$$h(\mathbf{y}_{\mathrm{e}})=\log\det(\mathbf{I}+P_{\mathrm{t}}\mathbf{H}_{\mathrm{e}}\mathbf{H}_{\mathrm{e}}^{\dagger}) \quad (131)$$

and

$$\begin{aligned}
h(\mathbf{y}_{\mathrm{e}}|\mathbf{u}) &= \log\det(\mathbf{I}+P_{\mathrm{t}}\mathbf{H}_{\mathrm{e}}\mathbf{V}_n\mathbf{V}_n^{\dagger}\mathbf{H}_{\mathrm{e}}^{\dagger})\\
&= \log\det(\mathbf{I}+P_{\mathrm{t}}\mathbf{H}_{\mathrm{e}}(\mathbf{I}-\mathbf{V}_{\mathrm{r}}\mathbf{V}_{\mathrm{r}}^{\dagger})\mathbf{H}_{\mathrm{e}}^{\dagger}) \quad (132)\\
&= \log\det(\mathbf{I}+P_{\mathrm{t}}(\mathbf{I}-\mathbf{V}_{\mathrm{r}}\mathbf{V}_{\mathrm{r}}^{\dagger})\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}) \quad (133)
\end{aligned}$$

where to obtain (132) we have used that $\mathbf{V}_{\mathrm{r}}\mathbf{V}_{\mathrm{r}}^{\dagger}+\mathbf{V}_n\mathbf{V}_n^{\dagger}=\mathbf{I}$ since $[\,\mathbf{V}_{\mathrm{r}}\quad\mathbf{V}_n\,]$ is unitary, and where to obtain (133) we have used that $\det(\mathbf{I}+\mathbf{AB})=\det(\mathbf{I}+\mathbf{BA})$ for any $\mathbf{A}$ and $\mathbf{B}$ of compatible dimensions.

In turn, substituting (131) and (133) into (130) we obtain, with some algebra:

$$\begin{aligned}
I(\mathbf{u};\mathbf{y}_{\mathrm{e}}) &= -\log\det(\mathbf{I}-P_{\mathrm{t}}(\mathbf{I}+P_{\mathrm{t}}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}(\mathbf{V}_{\mathrm{r}}\mathbf{V}_{\mathrm{r}}^{\dagger}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}))\\
&= -\log\det(\mathbf{I}-P_{\mathrm{t}}\mathbf{V}_{\mathrm{r}}^{\dagger}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}(\mathbf{I}+P_{\mathrm{t}}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{V}_{\mathrm{r}}))\\
&= -\log\det(\mathbf{V}_{\mathrm{r}}^{\dagger}(\mathbf{I}+P_{\mathrm{t}}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{V}_{\mathrm{r}}). \quad (134)
\end{aligned}$$

Finally, using (129) and (134) in the argument of (19), and again using (27), we obtain [cf. (29)]

$$\begin{aligned}
&R_{\mathrm{SN}}(P)\\
&\quad=\log\det(\mathbf{I}+P_{\mathrm{t}}\boldsymbol{\Delta}^2)+\log\det(\mathbf{V}_{\mathrm{r}}^{\dagger}(\mathbf{I}+P_{\mathrm{t}}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{V}_{\mathrm{r}})\\
&\quad=\log\det(P_{\mathrm{t}}\mathbf{I}+\boldsymbol{\Delta}^{-2})\\
&\qquad+\log\det(\mathbf{U}\boldsymbol{\Delta}\mathbf{V}_{\mathrm{r}}^{\dagger}(\mathbf{I}+P_{\mathrm{t}}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{V}_{\mathrm{r}}\boldsymbol{\Delta}\mathbf{U}^{\dagger})\\
&\quad=\log\det(P_{\mathrm{t}}\mathbf{I}+\boldsymbol{\Delta}^{-2})+\log\det(\mathbf{H}_{\mathrm{r}}(\mathbf{I}+P_{\mathrm{t}}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{H}_{\mathrm{r}}^{\dagger})
\end{aligned}$$

as required.

Finally, to establish the first equality in (30), we take the limit $P_{\mathrm{t}}\to\infty$ in (29). In particular, we have

$$\begin{aligned}
&R_{\mathrm{SN}}(P)\\
&=\log\det(\mathbf{I}+P_{\mathrm{t}}^{-1}\boldsymbol{\Delta}^{-2})\\
&\quad+\log\det(\mathbf{H}_{\mathrm{r}}(P_{\mathrm{t}}^{-1}\mathbf{I}+\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{H}_{\mathrm{r}}^{\dagger}),\\
&=O(P_{\mathrm{t}}^{-1})+\log\det(\mathbf{H}_{\mathrm{r}}((\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}+O(P_{\mathrm{t}}^{-1}))\mathbf{H}_{\mathrm{r}}^{\dagger}) \quad (135)\\
&=\log\det(\mathbf{H}_{\mathrm{r}}(\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{H}_{\mathrm{r}}^{\dagger})\\
&\quad+\log\det(\mathbf{I}+(\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1/2}O(P_{\mathrm{t}}^{-1})(\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-\dagger/2})\\
&=\log\det(\mathbf{H}_{\mathrm{r}}(\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{H}_{\mathrm{r}}^{\dagger})+O(P_{\mathrm{t}}^{-1}) \quad (136)
\end{aligned}$$

$$\hat{\mathbf{H}}\boldsymbol{\Theta}\hat{\mathbf{H}}^{\dagger}+\mathbf{I}+\boldsymbol{\Theta}\boldsymbol{\Theta}^{\dagger}-\boldsymbol{\Theta}\boldsymbol{\Phi}^{\dagger}-\boldsymbol{\Phi}\boldsymbol{\Theta}^{\dagger} = \boldsymbol{\Psi}_{\mathrm{r}}\begin{Bmatrix} n_{\mathrm{r}}-s-p & s & p\\ n_{\mathrm{r}}-s-p & \mathbf{I} & \mathbf{0} & \mathbf{0}\\ s & \mathbf{0} & \mathbf{I}+(\mathbf{D}_{\mathrm{r}}\mathbf{D}_{\mathrm{e}}^{-1})^2-2\mathbf{D}_{\mathrm{r}}\mathbf{D}_{\mathrm{e}}^{-1}\boldsymbol{\Xi} & (\mathbf{D}_{\mathrm{r}}\mathbf{D}_{\mathrm{e}}^{-1}-\boldsymbol{\Xi})\mathbf{F}_{32}^{\dagger}\\ p & \mathbf{0} & \mathbf{F}_{32}(\mathbf{D}_{\mathrm{r}}\mathbf{D}_{\mathrm{e}}^{-1}-\boldsymbol{\Xi}) & \mathbf{I}+\mathbf{F}_{31}\mathbf{F}_{31}^{\dagger}+\mathbf{F}_{32}\mathbf{F}_{32}^{\dagger}+\boldsymbol{\Omega}_3^{-1}\mathbf{Q}\boldsymbol{\Omega}_3^{-\dagger} \end{Bmatrix}\end{Bmatrix}\boldsymbol{\Psi}_{\mathrm{r}}^{\dagger}. \quad (123)$$

where to obtain (135) we have used that $(\epsilon\mathbf{I}+\mathbf{M})^{-1}=\mathbf{M}^{-1}+O(\epsilon)$ as $\epsilon\to 0$ for any invertible $\mathbf{M}$ [32], and where we have also used that $\log\det(\mathbf{I}+\mathbf{W})$ is continuous in the entries of $\mathbf{W}$.

VII. MIMOME Channel Scaling Laws

We first verify Claim 1, then use it to establish Corollary 2.

Proof of Claim 1: Clearly, $\sigma_{\max}(\mathbf{H}_{\rm r},\mathbf{H}_{\rm e})=\infty$ when [cf. (2a)] $\mathcal{S}_{\rm r}\neq\varnothing$. Otherwise, it is known (see, e.g., [33]) that $\sigma_{\max}(\cdot)$ is the largest generalized singular value of $(\mathbf{H}_{\rm r},\mathbf{H}_{\rm e})$ as defined in (9).

To establish that the secrecy capacity is zero whenever $\sigma_{\max}(\mathbf{H}_{\rm r},\mathbf{H}_{\rm e})\leq 1$, it suffices to consider the high-SNR secrecy capacity (24) when $\mathbf{H}_{\rm e}$ has full column-rank, which is clearly zero whenever $\sigma_{\max}\leq 1$.

When $\sigma_{\max}(\mathbf{H}_{\rm r},\mathbf{H}_{\rm e})>1$, there exists a vector $\mathbf{v}$ such that $\|\mathbf{H}_{\rm r}\mathbf{v}\|>\|\mathbf{H}_{\rm e}\mathbf{v}\|$. Then, choosing $\mathbf{x}=\mathbf{u}\sim\mathcal{CN}(\mathbf{0},P\mathbf{v}\mathbf{v}^\dagger)$ in the argument of (19) yields a strictly positive rate $R(P)$, so $C(P)\geq R(P)>0$ for all $P>0$. ∎

Combining Claim 1 and Fact 1 below, which is established in [34, p. 642], yields Corollary 2.

Fact 1 ([34], [35]): Suppose that $\mathbf{H}_{\rm r}$ and $\mathbf{H}_{\rm e}$ have i.i.d. $\mathcal{CN}(0,1)$ entries. Let $n_{\rm r},n_{\rm e},n_{\rm t}\to\infty$, while keeping $n_{\rm r}/n_{\rm e}=\gamma$ and $n_{\rm t}/n_{\rm e}=\beta$ fixed. Then if $\beta<1$,

$$\sigma_{\max}(\mathbf{H}_{\rm r},\mathbf{H}_{\rm e})\xrightarrow{\text{a.s.}}\gamma\left[\frac{1+\sqrt{1-(1-\beta)\left(1-\frac{\beta}{\gamma}\right)}}{1-\beta}\right]^2. \tag{137}$$

VIII. Concluding Remarks

This paper resolves several open questions regarding secure transmission with multiple antennas. First, it establishes the existence of a computable expression for the secrecy capacity of the MIMOME channel. Second, it establishes that a Gaussian input distribution optimizes the secrecy capacity expression of Csiszár and Körner for the MIMOME channel, and thus that capacity is achieved by Gaussian wiretap codes. Third, it establishes the optimum covariance structure for the input, exploiting hidden convexity in the problem. Nevertheless, many questions remain that are worth exploring. As one example, it remains to be determined whether such developments based on Sato's bounding techniques be extended beyond sum-power constraints, as the channel enhancement based approach of [22] can.

In addition, our analysis highlights the useful role that the GSVD plays both in calculating the capacity of the MIMOME channel in the high-SNR regime, and in designing codes for approaching this capacity. At the same time, we observed that a simple, semi-blind masked MIMO scheme can be arbtrarily far from capacity. However, for the special case of the MISOME channel, [4] shows that the corresponding masked beamforming scheme achieve rates close to capacity at high SNR. Thus, it remains to be determined whether there are better and/or more natural generalizations of the masked beamforming scheme for the general MIMOME channel. This warrants further investigation.

More generally, semi-blind schemes have the property that they require only partial knowledge of the channel to the eavesdropper. Much remains to be explored about what secrecy rates are achievable with such partial information. One recent work in this area [5] illustrates the use of interference alignment techniques for the compound extension of the multi-antenna wiretap channel. Another recent work [36], studies a constant-capacity compound wiretap channel model which again captures the constraint that the transmitter only knows the capacity (or an upper bound on the capacity) of the channel to the eavesdropper. Further insights may arise from considering other multiple eavesdropper scenarios with limited or no collusion.

Finally, we characterize when an eaversdropper can prevent secure communication, i.e., drive the secrecy capacity to zero. Our scaling laws on antenna requirements and their optimal distribution in limit of many antennas provide convenient rules of thumb for system designers, as the results become independent of the channel matrices in this limit. However, it remains to quantify for what numbers of antennas these asymptotic results become meaningful predictors of system behavior. As such, this represents yet another useful direction for further research.

Appendix I
Proof of Claim 2

To begin:

$$\begin{aligned} I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) &= I(\mathbf{x};\mathbf{U}_1^\dagger\mathbf{y}_{\rm r},\mathbf{U}_2^\dagger\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) && (138)\\ &= I(\mathbf{x};\tilde{\mathbf{y}}_{\rm r},\mathbf{U}_1^\dagger\mathbf{y}_{\rm r}-\mathbf{V}_1^\dagger\mathbf{y}_{\rm e}|\mathbf{y}_{\rm e})\\ &= I(\mathbf{x};\tilde{\mathbf{y}}_{\rm r},\mathbf{T}\mathbf{x}|\mathbf{y}_{\rm e}) && (139)\end{aligned}$$

where (138) follows from the fact that $[\,\mathbf{U}_1\quad\mathbf{U}_2\,]$ is unitary, and where (139) follows from substituting for $\mathbf{y}_{\rm r}$ and $\mathbf{y}_{\rm e}$ from (1), using (45), from and the fact that

$$\mathbf{U}_1^\dagger\mathbf{z}_{\rm r}\overset{\text{a.s.}}{=}\mathbf{V}_1^\dagger\mathbf{z}_{\rm e} \tag{140}$$

since

$$\operatorname{cov}(\mathbf{U}_1^\dagger\mathbf{z}_{\rm r},\mathbf{V}_1^\dagger\mathbf{z}_{\rm e})=E[\mathbf{U}_1^\dagger\mathbf{z}_{\rm r}\mathbf{z}_{\rm e}^\dagger\mathbf{V}_1]=\mathbf{U}_1^\dagger\mathbf{\Phi}\mathbf{V}_1=\mathbf{I}.$$

Now when $I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e})<\infty$, we have from (139) that $\mathbf{T}\mathbf{x}=\mathbf{0}$, so $I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e})=I(\mathbf{x};\tilde{\mathbf{y}}_{\rm r}|\mathbf{y}_{\rm e})$, establishing (38).

Similarly,

$$\begin{aligned} I(\mathbf{x};\mathbf{y}_{\rm e}|\mathbf{y}_{\rm r}) &= I(\mathbf{x};\mathbf{V}_1^\dagger\mathbf{y}_{\rm e},\mathbf{V}_2^\dagger\mathbf{y}_{\rm e}|\mathbf{y}_{\rm r}) && (141)\\ &= I(\mathbf{x};\tilde{\mathbf{y}}_{\rm e},\mathbf{V}_1^\dagger\mathbf{y}_{\rm e}-\mathbf{U}_1^\dagger\mathbf{y}_{\rm r}|\mathbf{y}_{\rm r})\\ &= I(\mathbf{x};\tilde{\mathbf{y}}_{\rm e},\mathbf{T}\mathbf{x}|\mathbf{y}_{\rm r}) && (142)\end{aligned}$$

where we have used that $[\,\mathbf{V}_1\quad\mathbf{V}_2\,]$ is unitary to obtain (141) and (140) to obtain (142). When $I(\mathbf{x};\mathbf{y}_{\rm e}|\mathbf{y}_{\rm r})<\infty$, we have from (142) that $\mathbf{T}\mathbf{x}=\mathbf{0}$, so $I(\mathbf{x};\mathbf{y}_{\rm e}|\mathbf{y}_{\rm r})=I(\mathbf{x};\tilde{\mathbf{y}}_{\rm e}|\mathbf{y}_{\rm r})$, establishing (41).

To verify the "only if" statement of the last part of the claim, when $I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e})=\infty$, we expand (139) via the chain rule to obtain

$$I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e})=I(\mathbf{x};\tilde{\mathbf{y}}_{\rm r}|\mathbf{y}_{\rm e})+I(\mathbf{x};\mathbf{T}\mathbf{x}|\tilde{\mathbf{y}}_{\rm r},\mathbf{y}_{\rm e}) \tag{143}$$

and note that if $\mathbf{T}\mathbf{x}\overset{\text{a.s.}}{=}\mathbf{0}$ then the second term on the right-hand side of (143) is zero. But the first term on the right-hand side is finite, so $\operatorname{cov}(\mathbf{T}\mathbf{x})\neq\mathbf{0}$, i.e., (44), holds.

To verify the "if" statement of the last part of the claim, we use the chain rule to write

$$\begin{aligned} I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) &= I(\mathbf{x};\tilde{\mathbf{y}}_{\rm r},\mathbf{T}\mathbf{x}|\mathbf{y}_{\rm e}) \\ &\geq I(\mathbf{x};\mathbf{T}\mathbf{x}|\mathbf{y}_{\rm e}) \\ &\geq I(\mathbf{x};\mathbf{T}\mathbf{x}) - I(\mathbf{x};\mathbf{y}_{\rm e}) \end{aligned} \tag{144}$$

and note that the first term in (144) is infinite when $\mathrm{cov}\,(\mathbf{T}\mathbf{x}) \neq \mathbf{0}$, while the second term is finite. ■

Appendix II
Optimizing $R_+(p_{\mathbf{x}},\mathbf{K}_{\mathbf{\Phi}})$ Over $p_{\mathbf{x}}$ With Singular $\mathbf{K}_{\mathbf{\Phi}}$

To establish that $I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e})$ with $\mathbf{z}\sim\mathcal{CN}(\mathbf{0},\mathbf{K}_{\mathbf{\Phi}})$ for singular $\mathbf{K}_{\mathbf{\Phi}}$ is maximized subject to the constraint $\mathrm{cov}\,(\mathbf{x}) = \mathbf{K}_P$ when $\mathbf{x}$ is Gaussian (hence, justifying (50) in this case), we exploit Claim 2.

In particular, if for all $p_{\mathbf{x}}$ meeting the covariance constraint we have $I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) < \infty$, then we can use (38), expanding and bounding $I(\mathbf{x};\tilde{\mathbf{y}}_{\rm r}|\mathbf{y}_{\rm e})$ in the same manner as (53)–(54), with $\tilde{\mathbf{y}}_{\rm r}$, $\tilde{\mathbf{z}}_{\rm r}$, $\tilde{\mathbf{\Lambda}} \triangleq \mathbf{U}_2^{\dagger}\mathbf{\Lambda}\mathbf{U}_2$ (the error covariance in the MMSE estimate of $\tilde{\mathbf{y}}_{\rm r}$ from $\mathbf{y}_{\rm e}$), and $\tilde{\mathbf{\Phi}} = \mathbf{U}_2^{\dagger}\mathbf{\Phi}$ [cf. (46)] replacing $\mathbf{y}_{\rm r}$, $\mathbf{z}_{\rm r}$, $\mathbf{\Lambda}$, and $\mathbf{\Phi}$, respectively. Specifically, we obtain that

$$I(\mathbf{x};\tilde{\mathbf{y}}_{\rm r}|\mathbf{y}_{\rm e}) = h(\tilde{\mathbf{y}}_{\rm r}|\mathbf{y}_{\rm e}) - h(\tilde{\mathbf{z}}_{\rm r}|\mathbf{z}_{\rm e}) \tag{145}$$

is maximized when $\mathbf{x}$ is Gaussian.

If, instead, there exists a $p_{\mathbf{x}}$ satisfying the covariance constraint such that $I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) = \infty$, then by the "only if" part of the last statement of Claim 2 we have that (44) holds. But by the "if" part of the same statement we know that $I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) = \infty$ for any $p_{\mathbf{x}}$ such that (44) holds, and in particular we may choose $p_{\mathbf{x}}$ to be Gaussian. ■

Appendix III
Proof of Lemma 3 for Singular $\bar{\mathbf{K}}_{\mathbf{\Phi}}$

We begin with the following.

Claim 3: There exists a matrix $\tilde{\mathbf{H}}$ such that the combined channel matrix (4) can be expressed in the form

$$\mathbf{H} = \mathbf{W}\tilde{\mathbf{H}} \tag{146}$$

where

$$\bar{\mathbf{K}}_{\mathbf{\Phi}} = \mathbf{W}\bar{\mathbf{\Xi}}\mathbf{W}^{\dagger} \tag{147}$$

is the compact singular value decomposition of $\bar{\mathbf{K}}_{\mathbf{\Phi}}$, i.e., where $\mathbf{W}$ has orthogonal columns ($\mathbf{W}^{\dagger}\mathbf{W} = \mathbf{I}$), and the diagonal matrix $\bar{\mathbf{\Xi}}$ has strictly positive diagonal entries.

Hence, the column space of $\mathbf{H}$ is a subspace of the column space of $\mathbf{W}$.

Proof: We establish our result by contradiction. Suppose the claim were false. Then clearly $I(\mathbf{x};\mathbf{y}_{\rm r},\mathbf{y}_{\rm e}) = \infty$ when we choose $\mathbf{x} = t\boldsymbol{v}$ where $\boldsymbol{v} \in \mathrm{Null}\,(\mathbf{W})$ and $\mathrm{var}\,t > 0$, which implies that

$$R_+(\mathbf{K}_P,\bar{\mathbf{K}}_{\mathbf{\Phi}}) = I(\mathbf{x};\mathbf{y}_{\rm r}|\mathbf{y}_{\rm e}) = I(\mathbf{x};\mathbf{y}_{\rm r},\mathbf{y}_{\rm e}) - I(\mathbf{x};\mathbf{y}_{\rm e}) = \infty$$

since $I(\mathbf{x};\mathbf{y}_{\rm e}) < \infty$ as $\mathrm{cov}\,(\mathbf{z}_{\rm e}) = \mathbf{I}$ is nonsingular. Hence

$$R_+(\bar{\mathbf{K}}_P,\bar{\mathbf{K}}_{\mathbf{\Phi}}) = \max_{\mathbf{K}_P\in\mathcal{K}_P} R_+(\mathbf{K}_P,\bar{\mathbf{K}}_{\mathbf{\Phi}}) = \infty. \tag{148}$$

But from (48) in Lemma 2 we know $R_+(\bar{\mathbf{K}}_P,\bar{\mathbf{K}}_{\mathbf{\Phi}}) < \infty$, which contradicts (148) and hence (146) must hold. ■

Using Claim 3, we see that in this case the original channel (1) with $\mathrm{cov}\,(\mathbf{z}) = \mathbf{K}_{\mathbf{\Phi}}$ can be replaced with the equivalent combined channel

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\mathbf{x} + \tilde{\mathbf{z}} \tag{149}$$

where

$$\tilde{\mathbf{y}} \triangleq \mathbf{W}^{\dagger}\begin{bmatrix}\mathbf{y}_{\rm r}\\ \mathbf{y}_{\rm e}\end{bmatrix}, \qquad \tilde{\mathbf{z}} \triangleq \mathbf{W}^{\dagger}\mathbf{z}$$

with $\mathrm{cov}\,(\mathbf{z}) = \mathbf{\Xi}$. Hence, we can write

$$R_+(\bar{\mathbf{K}}_P,\bar{\mathbf{K}}_{\mathbf{\Phi}}) = I(\mathbf{x};\mathbf{y}_{\rm r},\mathbf{y}_{\rm e}) - I(\mathbf{x};\mathbf{y}_{\rm e})$$

where

$$I(\mathbf{x};\mathbf{y}_{\rm r},\mathbf{y}_{\rm e}) = I(\mathbf{x};\tilde{\mathbf{y}}) = \log\frac{\det(\mathbf{\Xi}+\tilde{\mathbf{H}}\bar{\mathbf{K}}_P\tilde{\mathbf{H}}^{\dagger})}{\det(\mathbf{\Xi})} \tag{150}$$

and

$$I(\mathbf{x};\mathbf{y}_{\rm e}) = \log\det(\mathbf{I}+\mathbf{H}_{\rm e}\bar{\mathbf{K}}_P\mathbf{H}_{\rm e}^{\dagger}). \tag{151}$$

But from the saddle point property it follows that $\bar{\mathbf{\Xi}}$ can be expressed as

$$\bar{\mathbf{\Xi}} = \underset{\{\mathbf{\Xi}:\mathbf{W}\mathbf{\Xi}\mathbf{W}^{\dagger}\in\mathbf{K}_{\mathbf{\Phi}}\}}{\arg\min} \log\frac{\det(\mathbf{\Xi}+\tilde{\mathbf{H}}\bar{\mathbf{K}}_P\tilde{\mathbf{H}}^{\dagger})}{\det(\mathbf{\Xi})}. \tag{152}$$

In turn, the KKT conditions associated with the optimization (152) are

$$\bar{\mathbf{\Xi}}^{-1} - (\bar{\mathbf{\Xi}}+\tilde{\mathbf{H}}\bar{\mathbf{K}}_P\tilde{\mathbf{H}}^{\dagger})^{-1} = \mathbf{W}^{\dagger}\mathbf{\Upsilon}\mathbf{W}$$

or, equivalently

$$\tilde{\mathbf{H}}\bar{\mathbf{K}}_P\tilde{\mathbf{H}}^{\dagger} = \bar{\mathbf{\Xi}}\mathbf{W}^{\dagger}\mathbf{\Upsilon}\mathbf{W}(\bar{\mathbf{\Xi}}+\tilde{\mathbf{H}}\bar{\mathbf{K}}_P\tilde{\mathbf{H}}^{\dagger}) \tag{153}$$

where the dual variable $\mathbf{\Upsilon}$ is of the same block diagonal form as in the nonsingular case, viz., (60). Multiplying the left- and right-hand sides of (153) by $\mathbf{W}$ and $\mathbf{W}^{\dagger}$, respectively, and using (146) and (147) we obtain (63). Thus, the remainder of the proof uses the arguments following (63) in the proof for the nonsingular case to establish the desired result. ■

Appendix IV
Proof of Proposition 1

Consider first the right-hand side of (71). Since $h(\mathbf{y}_{\rm r}-\bar{\mathbf{\Theta}}\mathbf{y}_{\rm e})$ is concave in $\mathbf{K}_P \in \mathcal{K}_P$ and differentiable over $\mathcal{K}_P$, the KKT conditions associated with the Lagrangian

$$\begin{aligned}&\mathcal{L}_{\mathbf{\Theta}}(\mathbf{K}_P,\lambda,\mathbf{\Psi}) \\ &\quad = h(\mathbf{y}_{\rm r}-\bar{\mathbf{\Theta}}\mathbf{y}_{\rm e}) + \mathrm{tr}(\mathbf{\Psi}\mathbf{K}_P) - \lambda(\mathrm{tr}\,(\mathbf{K}_P) - P)\end{aligned} \tag{154}$$

are both necessary and sufficient, i.e., $\mathbf{K}_P$ is a solution to the right-hand side of (71) if and only if there exists a $\lambda \geq 0$ and $\mathbf{\Psi} \succeq \mathbf{0}$ such that

$$\begin{aligned}(\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})^{\dagger}\mathbf{\Gamma}(\mathbf{K}_P, \bar{\mathbf{K}}_P)^{-1}(\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}}) + \mathbf{\Psi} &= \lambda \mathbf{I}, \\ \mathrm{tr}\,(\mathbf{\Psi}\mathbf{K}_P) = 0, \quad \text{and} \quad \lambda(\mathrm{tr}\,(\mathbf{K}_P) - P) &= 0 \end{aligned} \tag{155}$$

where

$$\begin{aligned}&\mathbf{\Gamma}(\mathbf{K}_P, \bar{\mathbf{K}}_P) \\ &\triangleq \mathrm{cov}\,(\mathbf{y}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{y}_{\mathrm{e}}) \\ &= \mathbf{I} + \bar{\mathbf{\Theta}}\bar{\mathbf{\Theta}}^{\dagger} - \bar{\mathbf{\Theta}}\bar{\mathbf{\Phi}}^{\dagger} - \bar{\mathbf{\Phi}}\bar{\mathbf{\Theta}}^{\dagger} \\ &\quad + (\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})\mathbf{K}_P(\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})^{\dagger}. \end{aligned} \tag{156}$$

Considering next the left-hand side of (71), to which $\bar{\mathbf{K}}_P$ is a solution, we have, from the associated KKT conditions, that there exists $\lambda' \geq 0$ and $\mathbf{\Psi}' \succeq \mathbf{0}$ such that

$$\begin{aligned}\nabla_{\mathbf{K}_P} h(\mathbf{y}_{\mathrm{r}} - \mathbf{\Theta}(\mathbf{K}_P)\mathbf{y}_{\mathrm{e}})\big|_{\mathbf{K}_P = \bar{\mathbf{K}}_P} + \mathbf{\Psi}' &= \lambda' \mathbf{I} \\ \mathrm{tr}\,(\mathbf{\Psi}'\bar{\mathbf{K}}_P) = 0, \quad \text{and} \quad \lambda'(\mathrm{tr}\,(\bar{\mathbf{K}}_P) - P) &= 0 \end{aligned} \tag{157}$$

where $\mathbf{\Theta}(\mathbf{K}_P)$ is as defined in (18).

Thus, it remains to show that (155) and (157) are identical when $\mathbf{K}_P = \bar{\mathbf{K}}_P$. Focusing on the first equation in (157), we have

$$\begin{aligned}&\nabla_{\mathbf{K}_P} h(\mathbf{y}_{\mathrm{r}} - \mathbf{\Theta}(\mathbf{K}_P)\mathbf{y}_{\mathrm{e}})\big|_{\mathbf{K}_P = \bar{\mathbf{K}}_P} \\ &\quad = \nabla_{\mathbf{K}_P} h(\mathbf{y}_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}})\big|_{\mathbf{K}_P = \bar{\mathbf{K}}_P} \\ &\quad = \nabla_{\mathbf{K}_P} \{h(\mathbf{y}_{\mathrm{r}}, \mathbf{y}_{\mathrm{e}}) - h(\mathbf{y}_{\mathrm{e}})\}\big|_{\mathbf{K}_P = \bar{\mathbf{K}}_P} \\ &\quad = \mathbf{H}^{\dagger}(\mathbf{H}\bar{\mathbf{K}}_P\mathbf{H}^{\dagger} + \bar{\mathbf{K}}_{\mathbf{\Phi}})^{-1}\mathbf{H} - \mathbf{H}_{\mathrm{e}}^{\dagger}(\mathbf{I} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger})^{-1}\mathbf{H}_{\mathrm{e}}. \end{aligned} \tag{158}$$

In turn, substituting for $\mathbf{H}$ and $\bar{\mathbf{K}}_{\mathbf{\Phi}}$ from (4) and (21), and using (17), the first matrix inverse in (158) can be expressed in the form

$$\begin{aligned}&(\mathbf{H}\bar{\mathbf{K}}_P\mathbf{H}^{\dagger} + \bar{\mathbf{K}}_{\mathbf{\Phi}})^{-1} \\ &\quad = \begin{bmatrix} \mathbf{I} + \mathbf{H}_{\mathrm{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{r}}^{\dagger} & \bar{\mathbf{\Phi}} + \mathbf{H}_{\mathrm{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger} \\ \bar{\mathbf{\Phi}}^{\dagger} + \mathbf{H}_{\mathrm{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger} & \mathbf{I} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger} \end{bmatrix}^{-1} \\ &\quad = \begin{bmatrix} \mathbf{\Lambda}(\bar{\mathbf{K}}_P)^{-1} & -\mathbf{\Lambda}(\bar{\mathbf{K}}_P)^{-1}\bar{\mathbf{\Theta}} \\ -\bar{\mathbf{\Theta}}^{\dagger}\mathbf{\Lambda}(\bar{\mathbf{K}}_P)^{-1} & (\mathbf{I} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}})^{-1} + \bar{\mathbf{\Theta}}^{\dagger}\mathbf{\Lambda}(\bar{\mathbf{K}}_P)^{-1}\bar{\mathbf{\Theta}} \end{bmatrix} \end{aligned} \tag{159}$$

where $\mathbf{\Lambda}(\mathbf{K}_P)$ is as defined in (55), and where we have used the matrix inversion lemma (see, e.g., [32]). Substituting (159) into (158), and using the notation (55), yields, after some simplification:

$$\begin{aligned}&\nabla_{\mathbf{K}_P} h(\mathbf{y}_{\mathrm{r}} - \mathbf{\Theta}(\mathbf{K}_P)\mathbf{y}_{\mathrm{e}})\big|_{\mathbf{K}_P = \bar{\mathbf{K}}_P} \\ &\quad = \mathbf{H}^{\dagger}(\bar{\mathbf{K}}_{\mathbf{\Phi}} + \mathbf{H}\bar{\mathbf{K}}_P\mathbf{H}^{\dagger})^{-1}\mathbf{H} - \mathbf{H}_{\mathrm{e}}^{\dagger}(\mathbf{I} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger})^{-1}\mathbf{H}_{\mathrm{e}} \\ &\quad = (\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})^{\dagger}\bar{\mathbf{\Lambda}}^{-1}(\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}}). \end{aligned} \tag{160}$$

Comparing (160) with the first equation in (155), we see that it remains only to show that $\mathbf{\Gamma}(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_P) = \bar{\mathbf{\Lambda}}$, which is verified as follows. First, $\bar{\mathbf{\Theta}}\mathbf{y}_{\mathrm{e}}$ is the MMSE estimate of $\mathbf{y}_{\mathrm{r}}$ from $\mathbf{y}_{\mathrm{e}}$ when $\mathbf{K}_P = \bar{\mathbf{K}}_P$, and $\mathbf{\Gamma}(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_P) = \mathrm{cov}\,(\mathbf{y}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{y}_{\mathrm{e}}) = \mathrm{cov}\,(\mathbf{y}_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}})$ is the error covariance associated with the estimate. But by definition [cf. (55)] $\bar{\mathbf{\Lambda}} = \mathrm{cov}\,(\mathbf{y}_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}})$ is also the error covariance associated with the MMSE estimate when $\mathbf{K}_P = \bar{\mathbf{K}}_P$, so the conclusion follows. ■

Appendix V
Proof of Lemma 4 for Singular $\bar{\mathbf{K}}_{\mathbf{\Phi}}$

First, note that via (47) with (48), we have that $R_+(\mathbf{K}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) = I(\mathbf{x}; \mathbf{y}_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}}) < \infty$ for all $\mathbf{K}_P \in \mathcal{K}_P$. Hence, via (38) of Claim 2 we have

$$R_+(\mathbf{K}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) = I(\mathbf{x}; \tilde{\mathbf{y}}_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}}), \quad \forall \mathbf{K}_P \in \mathcal{K}_P \tag{161}$$

with the equivalent observations $\tilde{\mathbf{y}}_{\mathrm{r}}$ as given by (39) with (40). Moreover, the noise cross-covariance $\tilde{\mathbf{\Phi}} = \mathbf{U}_2^{\dagger}\bar{\mathbf{\Phi}}$ [cf. (46)] in the equivalent channel model has all its singular values strictly less than unity, i.e., the associated $\mathbf{K}_{\tilde{\mathbf{\Phi}}}$ is nonsingular.

Thus, we can apply to this equivalent model the arguments of the proof of Lemma 4 for the nonsingular case. In particular, from (72) onwards we replace $\mathbf{y}_{\mathrm{r}}$ with $\tilde{\mathbf{y}}_{\mathrm{r}}$, we replace $\mathbf{\Theta}(\mathbf{K}_P)$ and $\bar{\mathbf{\Theta}}$ with, respectively, [cf. (18), (17)]

$$\tilde{\mathbf{\Theta}}(\mathbf{K}_P) \triangleq (\tilde{\mathbf{H}}_{\mathrm{r}}\mathbf{K}_P\mathbf{H}_{\mathrm{e}}^{\dagger} + \tilde{\mathbf{\Phi}})(\mathbf{I} + \mathbf{H}_{\mathrm{e}}\mathbf{K}_P\mathbf{H}_{\mathrm{e}}^{\dagger})^{-1} = \mathbf{U}_2^{\dagger}\mathbf{\Theta}(\mathbf{K}_P) \tag{162}$$

and

$$\breve{\mathbf{\Theta}} \triangleq \tilde{\mathbf{\Theta}}(\bar{\mathbf{K}}_P) = \mathbf{U}_2^{\dagger}\bar{\mathbf{\Theta}} \tag{163}$$

which is the coefficient in the MMSE estimate of $\tilde{\mathbf{y}}_{\mathrm{r}}$ from $\mathbf{y}_{\mathrm{e}}$, and we replace $\bar{\mathbf{H}}_{\mathrm{eff}}$ and $\bar{\mathbf{J}}$ with, respectively, [cf. (77)]

$$\tilde{\mathbf{H}}_{\mathrm{eff}} \triangleq \tilde{\mathbf{J}}^{-1/2}(\tilde{\mathbf{H}}_{\mathrm{r}} - \breve{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})$$

and

$$\tilde{\mathbf{J}} \triangleq (\mathbf{I} - \tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}^{\dagger}) + (\breve{\mathbf{\Theta}} - \tilde{\mathbf{\Phi}})(\breve{\mathbf{\Theta}} - \tilde{\mathbf{\Phi}})^{\dagger} = \mathbf{U}_2^{\dagger}\bar{\mathbf{J}}\mathbf{U}_2$$

noting that $\tilde{\mathbf{J}} \succ \mathbf{0}$ since $\mathbf{K}_{\tilde{\mathbf{\Phi}}} \succ \mathbf{0}$. With these changes, and with the SVD

$$\tilde{\mathbf{H}}_{\mathrm{eff}} = \tilde{\mathbf{A}}\tilde{\mathbf{\Sigma}}_{\mathrm{eff}}\tilde{\mathbf{B}}^{\dagger}$$

replacing (78), the arguments apply and it follows that $(\tilde{\mathbf{H}}_{\mathrm{r}} - \breve{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})\mathbf{S}$ has a full column rank. Since

$$(\tilde{\mathbf{H}}_{\mathrm{r}} - \breve{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})\mathbf{S} = \mathbf{U}_2^{\dagger}(\mathbf{H}_{\mathrm{r}} - \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}})\mathbf{S}$$

it then follows that $(\mathbf{H}_{\mathrm{r}} - \mathbf{\Theta}\mathbf{H}_{\mathrm{e}})\mathbf{S}$ has a full column rank. ■

Appendix VI
Proof of Lemma 5 for Singular $\bar{\mathbf{K}}_{\mathbf{\Phi}}$

Consider first the case in which $\mathbf{H}_{\mathrm{r}} \neq \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}}$, and note that

$$R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) - R_-(\bar{\mathbf{K}}_P) = I(\mathbf{x}; \mathbf{y}_{\mathrm{e}}|\mathbf{y}_{\mathrm{r}}) < \infty$$

where the equality is reproduced from (85), and where the inequality follows from (48) and that $R_-(\bar{\mathbf{K}}_P) \geq 0$. Hence, applying (41) from Claim 2, we have

$$R_+(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) - R_-(\bar{\mathbf{K}}_P) = I(\mathbf{x}; \check{\mathbf{y}}_{\mathrm{e}}|\mathbf{y}_{\mathrm{r}})$$

with the equivalent observations $\tilde{\mathbf{y}}_{\mathrm{e}}$ as given by (42) with (43). Moreover, the noise cross-covariance

$$\breve{\mathbf{\Phi}} \triangleq E[\mathbf{z}_{\mathrm{r}}\tilde{\mathbf{z}}_{\mathrm{e}}^{\dagger}] = \bar{\mathbf{\Phi}}\mathbf{V}_2 \tag{164}$$

in the equivalent channel model has all its singular values strictly less than unity, i.e., the associated $\mathbf{K}_{\breve{\mathbf{\Phi}}}$ is nonsingular.

Thus, we can apply to this equivalent model the corresponding arguments of the proof of Lemma 5 for the nonsingular case. In particular, from (85) onwards we replace $\mathbf{y}_{\mathrm{e}}$ and $\mathbf{z}_{\mathrm{e}}$ with, respectively, $\tilde{\mathbf{y}}_{\mathrm{e}}$ and $\tilde{\mathbf{z}}_{\mathrm{e}}$, we replace $\mathbf{\Lambda}_{\mathrm{b}}$ with

$$\begin{aligned}\tilde{\mathbf{\Lambda}}_{\mathrm{b}} = &\\ &\mathbf{I} + \tilde{\mathbf{H}}_{\mathrm{e}}\bar{\mathbf{K}}_P\tilde{\mathbf{H}}_{\mathrm{e}}^{\dagger}\\ &- (\breve{\mathbf{\Phi}}^{\dagger} + \tilde{\mathbf{H}}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{r}}^{\dagger})(\mathbf{I} + \mathbf{H}_{\mathrm{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{r}}^{\dagger})^{-1}(\breve{\mathbf{\Phi}} + \mathbf{H}_{\mathrm{r}}\bar{\mathbf{K}}_P\tilde{\mathbf{H}}_{\mathrm{e}}^{\dagger})\\ &= \mathbf{V}_2^{\dagger}\mathbf{\Lambda}_{\mathrm{b}}\mathbf{V}_2\end{aligned}$$

which is the backward error covariance associated with the linear MMSE estimate of $\tilde{\mathbf{y}}_{\mathrm{e}}$ from $\mathbf{y}_{\mathrm{r}}$, and we replace the use of (22) in (87) and (88) with its form for the equivalent channel, viz., for all full column-rank $\bar{\mathbf{S}}$ such that $\bar{\mathbf{S}}\bar{\mathbf{S}}^{\dagger} = \bar{\mathbf{K}}_P$:

$$\breve{\mathbf{\Phi}}^{\dagger}\mathbf{H}_{\mathrm{r}}\bar{\mathbf{S}} = \mathbf{V}_2^{\dagger}\bar{\mathbf{\Phi}}^{\dagger}\mathbf{H}_{\mathrm{r}}\bar{\mathbf{S}} = \mathbf{V}_2^{\dagger}\mathbf{H}_{\mathrm{e}}\bar{\mathbf{S}} = \tilde{\mathbf{H}}_{\mathrm{e}}\bar{\mathbf{S}}$$

where to obtain the first equality we have used (164), where to obtain the second equality we have used Property 1, and where to obtain the third equality we have used (43).

Finally, consider the case in which $\mathbf{H}_{\mathrm{r}} = \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}}$. Since (48) holds, so does (38) of Claim 2, and thus

$$R_{+}(\bar{\mathbf{K}}_P, \bar{\mathbf{K}}_{\mathbf{\Phi}}) = I(\mathbf{x}; \tilde{\mathbf{y}}_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}}) \tag{165}$$

with the equivalent observations $\tilde{\mathbf{y}}_{\mathrm{r}}$ as given by (39) with (40).

Thus, we can apply to this equivalent model the corresponding arguments of the proof of Lemma 5 for the nonsingular case. In particular (and as in Appendix V), from (90) onwards (165) implies we replace $\mathbf{y}_{\mathrm{r}}$ and $\mathbf{z}_{\mathrm{r}}$ with, respectively, $\tilde{\mathbf{y}}_{\mathrm{r}}$ and $\tilde{\mathbf{z}}_{\mathrm{r}}$, we replace $\bar{\mathbf{\Phi}}$ with [cf. (46)] $\tilde{\bar{\mathbf{\Phi}}} = \mathbf{U}_2^{\dagger}\bar{\mathbf{\Phi}}$, the coefficient in the MMSE estimate of $\tilde{\mathbf{z}}_{\mathrm{r}}$ from $\mathbf{z}_{\mathrm{e}}$, and we replace $\bar{\mathbf{\Theta}}$ with [cf. (162),(163)]

$$\breve{\mathbf{\Theta}} = (\tilde{\mathbf{H}}_{\mathrm{r}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger} + \tilde{\bar{\mathbf{\Phi}}})(\mathbf{I} + \mathbf{H}_{\mathrm{e}}\bar{\mathbf{K}}_P\mathbf{H}_{\mathrm{e}}^{\dagger})^{-1} = \mathbf{U}_2^{\dagger}\bar{\mathbf{\Theta}} \tag{166}$$

the coefficient in the MMSE estimate of $\tilde{\mathbf{y}}_{\mathrm{r}}$ from $\mathbf{y}_{\mathrm{e}}$.

Note that in obtaining the counterpart of (92) we use that $\tilde{\mathbf{y}}_{\mathrm{r}} - \breve{\mathbf{\Theta}}\mathbf{y}_{\mathrm{e}} = \tilde{\mathbf{z}}_{\mathrm{r}} - \breve{\mathbf{\Theta}}\mathbf{z}_{\mathrm{e}}$ since

$$\tilde{\mathbf{H}}_{\mathrm{r}} = \mathbf{U}_2^{\dagger}\mathbf{H}_{\mathrm{r}} = \mathbf{U}_2^{\dagger}\bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}} = \breve{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}} \tag{167}$$

where the first equality follows from (40), the second equality follows from the assumption $\mathbf{H}_{\mathrm{r}} = \bar{\mathbf{\Theta}}\mathbf{H}_{\mathrm{e}}$, and the third equality from (166). Moreover, in obtaining the counterpart of (93) we use that $\breve{\mathbf{\Theta}} = \tilde{\bar{\mathbf{\Phi}}}$ when (167) holds. ■

ACKNOWLEDGMENT

The authors thank A. Wiesel for interesting discussions and help with numerical optimization of the saddle point expression in Theorem 1.

REFERENCES

[1] A. D. Wyner, "The wiretap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–1387, 1975.
[2] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, pp. 339–348, 1978.
[3] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wiretap channel," *IEEE Trans. Inf. Theory*, vol. 24, pp. 451–456, 1978.
[4] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas—Part I: The MISOME wiretap channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3088–3104, Jul. 2009.
[5] A. Khisti, "Interference alignment for the compound multi-antenna wiretap channel," *IEEE Trans. Inf. Theory*, 2010, submitted for publication.
[6] Y. Liang, H. V. Poor, and S. Shamai, "Secure communication over fading channels," *IEEE Trans. Inform. Theory*, vol. 54, pp. 2470–2492, Jun. 2008.
[7] Z. L. R. Yates and W. Trappe, "Secrecy capacity of independent parallel channels," in *Proc. Allerton Conf. Commun., Contr., Computing*, Monticello, IL, Sep. 2006.
[8] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure broadcasting over fading channels," *IEEE Trans. Inf. Theory*, vol. 54, pp. 2453–2469, Jun. 2008.
[9] P. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inf. Theory*, vol. 54, pp. 4687–4698, Oct. 2008.
[10] P. Parada and R. Blahut, "Secrecy capacity of SIMO and slow fading channels," in *Proc. Int. Symp. Inform. Theory*, Adelaide, Australia, Sep. 2005, pp. 2152–2155.
[11] R. Negi and S. Goel, "Guaranteeing secrecy using artificial noise," *IEEE Trans. Wireless Commun.*, vol. 7, no. 6, pp. 2180–2189, Jun. 2008.
[12] Z. Li, W. Trappe, and R. Yates, "Secret communication via multi-antenna transmission," in *Proc. Conf. Inform. Sci., Syst. (CISS)*, Baltimore, MD, Mar. 2007.
[13] S. Shaifee and S. Ulukus, "Achievable rates in Gaussian MISO channels with secrecy constraints," in *Proc. Int. Symp. Inform. Theory*, Nice, France, Jun. 2007.
[14] A. Khisti, G. W. Wornell, A. Wiesel, and Y. Eldar, "On the Gaussian MIMO wiretap channel," in *Proc. Int. Symp. Inform. Theory*, Nice, France, Jun. 2007.
[15] H. Sato, "An outer bound on the capacity region of broadcast channels," *IEEE Trans. Inf. Theory*, vol. 24, pp. 374–377, May 1978.
[16] W. Yu, "Uplink-downlink duality via minimax duality," *IEEE Trans. Inf. Theory*, vol. 52, pp. 361–374, Feb. 2006.
[17] S. Ulukus, Personal Communication 2007.
[18] 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. Inf. Theory*, vol. 55, pp. 4033–4039, Sep. 2009.
[19] A. Khisti and G. W. Wornell, "The MIMOME channel," in *Proc. Allerton Conf. Commun., Contr., Computing*, Monticello, IL, Sep. 2007.
[20] F. E. Oggier and B. Hassibi, "The secrecy capacity of the mimo wiretap channel," *CoRR*, vol. abs/0710.1920, 2007.
[21] F. Oggier and B. Hassibi, "The secrecy capacity of the 2×2 MIMO wiretap channel," in *Proc. Allerton Conf. Commun., Contr., Computing*, Monticello, IL, Sep. 2007.
[22] T. Liu and S. Shamai, "A note on the secrecy capacity of the multiple-antenna wiretap channel," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2547–2553, Jun. 2009.
[23] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, pp. 3936–3964, Sep. 2006.
[24] M. Yuksel and E. Erkip, "Diversity-multiplexing tradeoff for the multi-antenna wire-tap channel," in *Proc. Conf. Inform. Sci., Syst. (CISS)*, Princeton, NJ, Mar. 2008.
[25] C. Paige and M. A. Saunders, "Towards a generalized singular value decomposition," *SIAM J. Numer. Anal.*, vol. 18, no. 3, pp. 398–405, Jun. 1981.
[26] C. F. Van Loan, "Generalizing the singular value decomposition," *SIAM J. Numer. Anal.*, vol. 13, no. 1, pp. 76–83, 1976.
[27] S. Wilks, *Mathematical Statistics*. New York: Wiley, 1962.
[28] M. Sion, "On general minimax theorems," *Pac. J. Math.*, vol. 8, no. 1, pp. 171–176, 1958.
[29] S. N. Diggavi and T. M. Cover, "The worst additive noise under a covariance constraint," *IEEE Trans. Inform. Theory*, vol. IT-47, pp. 3072–3081, 2001.

[30] R. B. Bapat and T. E. S. Raghavan, *Non-Negative Matrices and Applications*. Cambridge, U.K: Cambridge Univ. Press, 1997.
[31] E. Martinian, Waterfilling Gains O(1/SNR) at High SNR 2004 [Online]. Available: http://allegro.mit.edu/pubs/posted/journal/2004-martinian-unpublished.pdf, unpublished
[32] K. Petersen and M. Pedersen, *The Matrix Cookbook*, Sep. 2007.
[33] G. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
[34] J. W. Silverstein, "The limiting eigenvalue distribution of a multivariate f matrix," *SIAM J. Math. Anal.*, vol. 16, no. 3, pp. 641–646, May 1985.
[35] Z. D. Bai and J. W. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *Ann. Prob.*, vol. 26, no. 1, pp. 316–345, 1998.
[36] V. Chandar, "Sparse Graph Codes for Compression, Sensing, and Secrecy," Ph.D., Mass. Inst. Technol., Cambridge, 2010.
[37] R. Bustin, R. Liu, H. V. Poor, and S. Shamai, "MMSE approach to the secrecy capacity of the MIMO Gaussian Wiretap Channel," *EURASIP J. Wireless Commun., Special Issue on Wireless Physical Layer Security*, Mar. 2009.

Ashish Khisti (S'01–M'08) received the B.A.Sc degree from the Engineering Science program at the University of Toronto, Toronto, ON, Canada in 2002, and the M.S. and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 2004 and 2008, respectively.

Since 2009, he has been an Assistant Professor in the Department of Electrical and Computer Engineering, University of Toronto. His research interests are in the area of information and coding theories and their applications to wireless communication systems, multimedia communication systems, inference, and security.

Prof. Khisti is a recipient of a Hewlett-Packard doctoral fellowship, a National Science Engineering Research Council (NSERC) postgraduate scholarship, and the Lucent Global Science Scholars award. He was awarded the Harold L. Hazen teaching award by the EECS Department at MIT as well as the Morris Joseph Levin EECS Masterworks Award for his masters thesis presentation "Coding Techniques for Multicasting."

Gregory W. Wornell (S'83–M'91–SM'00–F'04) received the B.A.Sc. degree from the University of British Columbia, Vancouver, BC, Canada, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, all in electrical engineering and computer science, in 1985, 1987, and 1991, respectively.

Since 1991, he has been on the faculty at MIT, where he is a Professor of electrical engineering and computer science, Co-Director of the Center for Wireless Networking, and Chair of Graduate Area I (Systems, Communication, Control, and Signal Processing) within the department's doctoral program. He has held visiting appointments at the former AT&T Bell Laboratories, Murray Hill, NJ; the University of California, Berkeley, and Hewlett-Packard Laboratories, Palo Alto, CA. His research interests and publications span the areas of signal processing, digital communication, and information theory, and include algorithms and architectures for wireless and sensor networks, broadband systems, and multimedia environments.

Prof. Wornell has been involved in the Signal Processing and Information Theory Societies of the IEEE in a variety of capacities and maintains a number of close industrial relationships and activities. He has won a number of awards for both his research and teaching.