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

# Secure Transmission With Multiple Antennas I: The MISOME Wiretap Channel

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

***Abstract*—The role of multiple antennas for secure communication is investigated within the framework of Wyner's wiretap channel. We characterize the secrecy capacity in terms of generalized eigenvalues when the sender and eavesdropper have multiple antennas, the intended receiver has a single antenna, and the channel matrices are fixed and known to all the terminals, and show that a beamforming strategy is capacity-achieving. In addition, we study a masked beamforming scheme that radiates power isotropically in all directions and show that it attains near-optimal performance in the high SNR regime. Insights into the scaling behavior of the capacity in the large antenna regime as well as extensions to ergodic fading channels are also provided.**

***Index Terms*—Artificial noise, broadcast channel, cryptography, generalized eigenvalues, masked beamforming, MIMO systems, multiple antennas, secrecy capacity, secure space-time codes, wiretap channel.**

## I. Introduction

MULTIPLE-ELEMENT antenna arrays are finding growing use in wireless communication networks. Much research to date has focused on the role of such arrays in enhancing the throughput and robustness for wireless communication systems. By contrast, this paper focuses on the role of such arrays in a less explored aspect of wireless systems, enhancing security. Specifically, we develop and optimize physical layer techniques for using multiple antennas to protect digital transmissions from potential eavesdroppers, and analyze the resulting performance characteristics.

A natural framework for protecting information at the physical layer is the so-called wiretap channel introduced by Wyner [1] and associated notion of secrecy capacity. In the basic wiretap channel, there are three terminals, one sender, one receiver, and one eavesdropper. Wyner's original treatment established the secrecy capacity for the case where the underlying broadcast channel between the sender and the receiver and eavesdropper is a degraded one. Subsequent work generalized this result to nondegraded discrete memoryless broadcast channels [2], and applied it to the basic Gaussian channel [3].

Motivated by emerging wireless communication applications, there is growing interest in extending the basic Gaussian wiretap channel to the case when the terminals have multiple antennas; see, e.g., [4]–[12] and the references therein. While in principle the secrecy capacity for such nondegraded broadcast channels is developed in [2] by Csiszár and Körner, the solution is in terms of an optimized auxiliary random variable and has been prohibitively difficult to explicitly evaluate. Thus, such characterizations of the solution have not proved particularly useful in practice.

In this paper, we investigate practical characterizations for the specific scenario in which the sender and eavesdropper have multiple antennas, but the intended receiver has a single antenna. We refer to this configuration as the multi-input, single-output, multi-eavesdropper (MISOME) case. It is worth emphasizing that the multiple eavesdropper antennas can correspond to a physical multiple-element antenna array at a single eavesdropper, a collection of geographically dispersed but perfectly colluding single-antenna eavesdroppers, or related variations. We note that the case where there are multiple non-colluding eavesdroppers has been recently been studied as a *compound* wire-tap channel problem [13], [14]. Furthermore, in the companion paper [15], we treat the multi-input, multi-output, multi-eavesdropper (MIMOME) case.

We first develop the secrecy capacity when the complex channel gains are fixed and known to all the terminals. A novel aspect of our derivation is our approach to (tightly) upper bounding the secrecy capacity for the wiretap channel. Our result thus indirectly establishes the optimum choice of auxiliary random variable in the secrecy capacity expression of [2], addressing an open problem.

The capacity achieving scheme requires that the sender and the intended receiver have knowledge of the eavesdropper's channel (and thus number of antennas as well), which is often not practical. We also analyze a *masked beamforming* scheme described in [4], [5]. This scheme is "semi-blind", i.e., the eavesdropper's channel knowledge is not used in choosing the transmit directions but used in selecting the rate.

In addition we study the secrecy capacity in two regimes: the high SNR regime and the limit of many antennas. In the high SNR regime, we show that the masked beamforming scheme achieves a rate that is close to the capacity achieving scheme. In the limit of many antennas we observe a that a critical threshold

Manuscript received October 09, 2007. revised November 23, 2009. Current version published June 16, 2010. This work was supported in part by the National Science Foundation under Grant CCF-0515109. The material in this paper was presented at International Symposium on Information Theory (ISIT), Nice, France, July 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 M5S3G4 Canada (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.2010.2048445

for the ratio of eavesdropping antennas to transmitting antennas. Above this threshold, the secrecy capacity equals zero (almost surely), whereas below this threshold the secrecy capacity remains positive. The masked beamforming scheme also exhibits a similar threshold.

Our results extend to the case of time-varying channels. We focus on the case of fast (ergodic, Rayleigh) fading, where the message is transmitted over a block that is long compared to the coherence time of the fading. In our model the state of the channel to the receiver is known by all three parties (sender, receiver, and eavesdropper), but the state of the channel to the eavesdropper is known only to the eavesdropper. Building on techniques developed for the single transmitter antenna wiretap problems [8], [9], we develop upper and lower bounds on the secrecy capacity both for finitely many antennas and in the large antenna limit.

As a final comment, we note that the idea of protecting information at the physical layer (rather than the application layer) is not a conventional approach in contemporary cryptography. Indeed, the common architecture today has the lower network layers focus on providing a noiseless public bit-pipe and the higher network layers focus on enabling privacy via the exchange and distribution of encryption keys among legitimate parties prior to the commencement of communication. As discussed in [7], [9], [16], and [17], for many emerging applications, existing key distribution methods are difficult to exploit effectively. In such cases, physical-layer mechanisms such as those developed in this paper constitute a potentially attractive alternative approach to providing transmission security.

The organization of the paper is as follows. Section II summarizes some convenient notation used in the paper and some mathematical preliminaries. Section III describes the channel and system model of interest. Section IV states all the main results of the paper. The proofs of our results appear in subsequent sections and the more technical details are provided in the Appendices. Section V provides an alternate upper bound while Section VI provides the secrecy capacity. Our analysis of the masked beamforming scheme is provided in Section VII while the scaling laws of the secrecy capacity and the masked beamforming scheme are provided in Section VIII. The extension to ergodic fading channels with only intended receiver's channel state information is treated in Sections IX and X contains some concluding remarks.

## II. Preliminaries

### A. Notation

Bold upper and lower case characters are used for matrices and vectors, respectively. Random variables are distinguished from realizations by the use of san-serif fonts for the former and seriffed fonts for the latter. And we generally reserve the symbols $I$ for mutual information, $H$ for entropy, and $h$ for differential entropy. All logarithms are base-2 unless otherwise indicated.

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 $\mathbb{C}^{m\times n}$. Matrix transposition is denoted using the superscript $^{\mathrm{T}}$, and the Hermitian (i.e., conjugate) transpose of a matrix is denoted using the superscript $^{\dagger}$. Moreover, $\mathrm{Null}(\cdot)$ denotes the null space of its matrix argument, and $\mathrm{tr}(\cdot)$ and $\det(\cdot)$ denote the trace and determinant of a matrix, respectively. The notation $\mathbf{A} \succeq 0$ means that $\mathbf{A}$ is a positive semidefinite matrix and we reserve the symbol $\mathbf{I}$ to denote the identity matrix, whose dimensions will be clear from the context.

A sequence of length $n$ is either denoted by $\{x(t)\}_{t=1}^{n}$ or sometimes more succinctly as $x^n$; in addition, we sometimes need notation the $x_i^j$ for a sequence $x_i, x_{i+1}, \ldots, x_j$.

Finally, $\mathcal{CN}(0, \mathbf{K})$ denotes a zero-mean circularly-symmetric complex Gaussian distribution with covariance $\mathbf{K}$, and we use the notation $\{\cdot\}^+ \triangleq \max(0, \cdot)$ throughout the paper.

### B. Generalized Eigenvalues

Many of our results arise out of generalized eigenvalue analysis. We summarize the properties of generalized eigenvalues and eigenvectors we require in the sequel. For more extensive developments of the topic, see, e.g., [18] and [19].

*Definition 1 (Generalized Eigenvalues):* For a Hermitian matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ and positive definite[1] matrix $\mathbf{B} \in \mathbb{C}^{n\times n}$, we refer to $(\lambda, \boldsymbol{\psi})$ as a generalized eigenvalue-eigenvector pair of $(\mathbf{A}, \mathbf{B})$ if $(\lambda, \boldsymbol{\psi})$ satisfy

$$\mathbf{A}\boldsymbol{\psi} = \lambda \mathbf{B}\boldsymbol{\psi}. \tag{1}$$

Since $\mathbf{B}$ in Definition 1 is invertible, first note that generalized eigenvalues and eigenvectors can be readily expressed in terms of regular ones. Specifically:

*Fact 1:* The generalized eigenvalues and eigenvectors of the pair $(\mathbf{A}, \mathbf{B})$ are the regular eigenvalues and eigenvectors of the matrix $\mathbf{B}^{-1}\mathbf{A}$.

Other characterizations reveal more useful properties for our development. For example, we have the following:

*Fact 2 (Variational Characterization):* The generalized eigenvectors of $(\mathbf{A}, \mathbf{B})$ are the stationary point solution to a particular Rayleigh quotient. Specifically, the largest generalized eigenvalue is the maximum of the Rayleigh quotient[2]

$$\lambda_{\max}(\mathbf{A}, \mathbf{B}) = \max_{\boldsymbol{\psi}\in\mathbb{C}^n} \frac{\boldsymbol{\psi}^{\dagger}\mathbf{A}\boldsymbol{\psi}}{\boldsymbol{\psi}^{\dagger}\mathbf{B}\boldsymbol{\psi}} \tag{2}$$

and the optimum is attained by the eigenvector corresponding to $\lambda_{\max}(\mathbf{A}, \mathbf{B})$.

The case when $\mathbf{A}$ has rank one is of special interest to us. In this case, the generalized eigenvalue admits a particularly simple expression:

*Fact 3 (Quadratic Form):* When $\mathbf{A}$ in Definition 1 has rank one, i.e., $\mathbf{A} = \mathbf{a}\mathbf{a}^{\dagger}$ for some $\mathbf{a} \in \mathbb{C}^n$, then

$$\lambda_{\max}(\mathbf{a}\mathbf{a}^{\dagger}, \mathbf{B}) = \mathbf{a}^{\dagger}\mathbf{B}^{-1}\mathbf{a}. \tag{3}$$

[1]When $\mathbf{B}$ is singular, we replace $\lambda$ with a pair $(\alpha, \beta)$ that satisfies $\beta\mathbf{A}\boldsymbol{\psi} = \alpha\mathbf{B}\boldsymbol{\psi}$. A solution for which $\alpha \neq 0$ and $\beta = 0$ corresponds to an infinite eigenvector. Generalized eigenvalues and eigenvectors also arise in simultaneous diagonalization of $(\mathbf{A}, \mathbf{B})$ [18].

[2]Throughout the paper, we use $\lambda_{\max}$ to denote the largest eigenvalue. Whether this is a regular or generalized eigenvalue will be clear from context, and when there is a need to be explicit, the relevant matrix or matrices will be indicated as arguments. The Rayleigh quotient is defined as the argument of the maxima in (2).

## III. Channel and System Model

The MISOME channel and system model is as follows. We use $n_{\mathrm{t}}$ and $n_{\mathrm{e}}$ to denote the number of sender and eavesdropper antennas, respectively; the (intended) receiver has a single antenna. The signals observed at the receiver and eavesdropper, respectively, are, for $t = 1, 2, \ldots,$

$$\begin{aligned} \mathsf{y}_{\mathrm{r}}(t) &= \mathbf{h}_{\mathrm{r}}^{\dagger}\mathbf{x}(t) + \mathsf{z}_{\mathrm{r}}(t) \\ \mathbf{y}_{\mathrm{e}}(t) &= \mathbf{H}_{\mathrm{e}}\mathbf{x}(t) + \mathbf{z}_{\mathrm{e}}(t) \end{aligned} \tag{4}$$

where $\mathbf{x}(t) \in \mathbb{C}^{n_{\mathrm{t}}}$ is the transmitted signal vector, $\mathbf{h}_{\mathrm{r}} \in \mathbb{C}^{n_{\mathrm{t}}}$ and $\mathbf{H}_{\mathrm{e}} \in \mathbb{C}^{n_{\mathrm{e}} \times n_{\mathrm{t}}}$ are complex channel gains, and $\mathsf{z}_{\mathrm{r}}(t)$ and $\mathbf{z}_{\mathrm{e}}(t)$ are independent identically-distributed (i.i.d.) circularly-symmetric complex-valued Gaussian noises: $\mathsf{z}_{\mathrm{r}}(t) \sim \mathcal{CN}(0,1)$ and $\mathbf{z}_{\mathrm{e}}(t) \sim \mathcal{CN}(0, \mathbf{I})$. Moreover, the noises are independent, and the input satisfies an average power constraint of $P$, i.e.,

$$E\left[\frac{1}{n}\sum_{t=1}^{n} \|\mathbf{x}(t)\|^2\right] \leq P. \tag{5}$$

Finally, except when otherwise indicated, all channel gains are fixed throughout the entire transmission period, and are known to all the terminals.

Communication takes place at a rate $R$ in bits per channel use over a transmission interval of length $n$. Specifically, a $(2^{nR}, n)$ code for the channel consists of a message $\mathsf{w}$ uniformly distributed over the index set $\mathcal{W}_n = \{1, 2, \ldots, 2^{nR}\}$, an encoder $\mu_n : \mathcal{W}_n \to \mathbb{C}^{n_{\mathrm{t}} \times n}$ that maps the message $\mathsf{w}$ to the transmitted (vector) sequence $\{\mathbf{x}(t)\}_{t=1}^{n}$, and a decoding function $\nu_n : \mathbb{C}^n \to \mathcal{W}_n$ that maps the received sequence $\{y_{\mathrm{r}}(t)\}_{t=1}^{n}$ to a message estimate $\hat{\mathsf{w}}$. The error event is $\mathcal{E}_n = \{\nu_n(\mu_n(\mathsf{w})) \neq \mathsf{w}\}$, and the amount of information obtained by the eavesdropper from the transmission is measured via the equivocation $I(\mathsf{w}; \mathbf{y}_{\mathrm{e}}^n)$.

*Definition 2 (Secrecy Capacity):* A secrecy rate $R$ is achievable if there exists a sequence of $(2^{nR}, n)$ codes such that $\Pr(\mathcal{E}_n) \to 0$ and the equivocation term $\frac{1}{n}H(\mathsf{w}|\mathbf{y}_{\mathrm{e}}^n) \geq \frac{1}{n}H(\mathsf{w}) - \varepsilon_n$ where $\varepsilon_n$ approaches zero as $n \to \infty$. The *secrecy capacity* is the supremum of all achievable secrecy-rates.

Note that our notion of secrecy capacity follows [1]–[3] in requiring a vanishing *per-symbol* mutual information for the eavesdropper's channel (hence, the normalization by $n$ in Definition 2). Practically, this means that while the eavesdropper is unable to decode any fixed fraction of the message bits, it does not preclude the possibility of decoding a fixed number (but vanishing fraction) of the message bits.

Maurer and Wolf [20] (see also [21]) have observed that for discrete memoryless channels, the secrecy capacity is not reduced even when one imposes the stronger requirement that $I(\mathsf{w}; \mathbf{y}_{\mathrm{e}}^n) \to 0$ as $n \to \infty$. However, we remark in advance that it remains an open question whether a similar result holds for the Gaussian case of interest in this work.

## IV. Main Results

The MISOME wiretap channel is a nondegraded broadcast channel. In Csiszár and Körner [2], the secrecy capacity of the nondegraded discrete memoryless broadcast channel $p_{\mathsf{y}_{\mathrm{r}}, \mathbf{y}_{\mathrm{e}}|\mathbf{x}}$ is expressed in the form

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

where $\mathsf{u}$ is an auxiliary random variable over a certain alphabet that satisfies the Markov relation $\mathsf{u} \leftrightarrow \mathsf{x} \leftrightarrow (\mathsf{y}_{\mathrm{r}}, \mathsf{y}_{\mathrm{e}})$. Moreover, the secrecy capacity (6) readily extends to the continuous alphabet case with a power constraint, so it also gives a characterization of the MISOME channel capacity.

Rather than attempting to solve for the optimal choice of $\mathsf{u}$ and $p_{\mathsf{x}|\mathsf{u}}$ in (6) directly to evaluate this capacity,[3] we consider an indirect approach based on a useful upper bound as the converse, which we describe next. We note in advance that, as described in [10], our upper bound has the added benefit that it extends easily to the MIMOME case (i.e., when the receiver has multiple antennas).

### A. Upper Bound on Achievable Rates

A key result is the following upper bound, which we derive in Section V.

*Theorem 1:* An upper bound on the secrecy capacity for the MISOME channel model is

$$R_+ = \min_{\mathbf{K}_{\boldsymbol{\phi}} \in \mathcal{K}_{\boldsymbol{\phi}}} \max_{\mathbf{K}_P \in \mathcal{K}_P} R_+(\mathbf{K}_P, \mathbf{K}_{\boldsymbol{\phi}}) \tag{7}$$

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

$$\mathcal{K}_P \triangleq \{\mathbf{K}_P : \mathbf{K}_P \succeq 0, \quad \mathrm{tr}(\mathbf{K}_P) \leq P\} \tag{8}$$

and where

$$\begin{bmatrix} \mathsf{z}_{\mathrm{r}} \\ \mathbf{z}_{\mathrm{e}} \end{bmatrix} \sim \mathcal{CN}(0, \mathbf{K}_{\boldsymbol{\phi}}) \tag{9}$$

with

$$\begin{aligned} \mathcal{K}_{\boldsymbol{\phi}} &\triangleq \left\{\mathbf{K}_{\boldsymbol{\phi}} : \mathbf{K}_{\boldsymbol{\phi}} = \begin{bmatrix} 1 & \boldsymbol{\phi}^{\dagger} \\ \boldsymbol{\phi} & \mathbf{I} \end{bmatrix}, \quad \mathbf{K}_{\boldsymbol{\phi}} \succeq 0\right\} \\ &= \left\{\mathbf{K}_{\boldsymbol{\phi}} : \mathbf{K}_{\boldsymbol{\phi}} = \begin{bmatrix} 1 & \boldsymbol{\phi}^{\dagger} \\ \boldsymbol{\phi} & \mathbf{I} \end{bmatrix}, \quad \|\boldsymbol{\phi}\| \leq 1\right\}. \end{aligned} \tag{10}$$

To obtain this bound, we consider a genie-aided channel in which the eavesdropper observes $\mathbf{y}_{\mathrm{e}}$ but the receiver observes *both* $\mathsf{y}_{\mathrm{r}}$ and $\mathbf{y}_{\mathrm{e}}$. Such a channel clearly has a capacity larger than the original channel. Moreover, since it is a degraded broadcast channel, the secrecy capacity of the genie-aided channel can be easily derived and is given by (cf. [1]) $\max I(\mathbf{x}; \mathsf{y}_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}})$ where the maximum is over the choice of input distributions. As we will see, it is straightforward to establish that the maximizing input distribution is Gaussian (in contrast to the original channel). Next, while the secrecy capacity of the original channel depends only on the marginal distributions $p_{\mathsf{y}_{\mathrm{r}}|\mathbf{x}}$ and $p_{\mathbf{y}_{\mathrm{e}}|\mathbf{x}}$ (see, e.g., [2]), mutual information $I(\mathbf{x}; \mathsf{y}_{\mathrm{r}}|\mathbf{y}_{\mathrm{e}})$ for the genie-aided channel depends on the joint distribution $p_{\mathsf{y}_{\mathrm{r}}, \mathbf{y}_{\mathrm{e}}|\mathbf{x}}$. Accordingly we obtain the tightest such upper bound

[3]The direct approach is explored in, e.g., [11] and [12], where the difficulty of performing this optimization is reported even when restricting $p_{\mathsf{x}|\mathsf{u}}$ to be singular (a deterministic mapping) and/or the input distribution to be Gaussian.

by finding the joint distribution (having the required marginal distributions), whence (7).

The optimization (7) can be carried out analytically, yielding an explicit expression, as we now develop.

### B. MISOME Secrecy Capacity

The upper bound described in the preceding section is achievable, yielding the MISOME channel capacity. Specifically, we have the following theorem, which we prove in Section VI-A.

*Theorem 2:* The secrecy capacity of the channel (4) is

$$C(P) = \left\{\log \lambda_{\max}\left(\mathbf{I} + P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I} + P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}\right)\right\}^{+} \tag{11}$$

with $\lambda_{\max}$ denoting the largest generalized eigenvalue of its argument pair. Furthermore, the capacity is obtained by beamforming (i.e., signaling with rank one covariance) along the direction $\boldsymbol{\psi}_{\max}$ of the[4] generalized eigenvector corresponding to $\lambda_{\max}$ with an encoding of the message using a code for the scalar Gaussian wiretap channel.

We emphasize that the beamforming direction in Theorem 2 for achieving capacity will in general depend on all of the target receiver's channel $\mathbf{h}_{\mathrm{r}}$, the eavesdropper's channel $\mathbf{H}_{\mathrm{e}}$, and the SNR ($P$). In the high SNR regime, the MISOME capacity (11) exhibits one of two possible behaviors, corresponding to whether

$$\lim_{P\to\infty} C(P) = \left\{\log \lambda_{\max}\left(\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}\right)\right\}^{+} \tag{12}$$

is finite or infinite, which depends on whether or not $\mathbf{h}_{\mathrm{r}}$ has a component in the null space of $\mathbf{H}_{\mathrm{e}}$. Specifically, we have the following corollary, which we prove in Section VI-B.

*Corollary 1:* The high SNR asymptote of the secrecy capacity (11) takes the form

$$\lim_{P\to\infty} C(P) = \{\log \lambda_{\max}(\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\}^{+} < \infty \quad \text{if } \mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}} = \mathbf{0} \tag{13a}$$

$$\lim_{P\to\infty} [C(P) - \log P] = \log \|\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}}\|^2 \quad \text{if } \mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}} \neq \mathbf{0} \tag{13b}$$

where $\mathbf{H}_{\mathrm{e}}^{\sharp}$ denotes the projection matrix onto the null space of $\mathbf{H}_{\mathrm{e}}$, i.e.,

$$\mathbf{H}_{\mathrm{e}}^{\sharp} = \boldsymbol{\Psi}_{\mathrm{e}}\boldsymbol{\Psi}_{\mathrm{e}}^{\dagger},$$

where $\boldsymbol{\Psi}_{\mathrm{e}}$ is a matrix whose columns constitute an orthonormal basis for the null space of $\mathbf{H}_{\mathrm{e}}$.

This behavior can be understood rather intuitively. In particular, when $\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}} = \mathbf{0}$, as is typically the case when the eavesdropper uses enough antennas ($n_{\mathrm{e}} \geq n_{\mathrm{t}}$) or the intended receiver has an otherwise unfortunate channel, the secrecy capacity saturates to a constant as $\mathrm{SNR} \to \infty$. In essence, while more transmit power is advantageous to communication to the intended receiver, it is also advantageous to the eavesdropper, resulting in diminishing returns.

By contrast, when $\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}} \neq \mathbf{0}$, as is typically the case when, e.g., the eavesdropper uses insufficiently many antennas ($n_{\mathrm{e}} < n_{\mathrm{t}}$) unless the eavesdropper has an otherwise unfortunate channel, the transmitter is able to steer a null to the eavesdropper without simultaneously nulling the receiver, and, thus, capacity grows by 1 b/s/Hz with every 3 dB increase in transmit power as it would if there were no eavesdropper to contend with. The MISOME capacity (11) is also readily specialized to the low SNR regime, as we develop in Section VI-C, and takes the following form.

*Corollary 2:* The low SNR asymptote of the secrecy capacity is

$$\lim_{P\to 0} \frac{C(P)}{P} = \frac{1}{\ln 2}\{\lambda_{\max}(\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger} - \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\}^{+}. \tag{14}$$

In this low SNR regime, the direction of optimal beamforming vector approaches the (regular) eigenvector corresponding to the largest (regular) eigenvalue of $\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger} - \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}$. Note that the optimal direction is in general not along $\mathbf{h}_{\mathrm{r}}$. Thus, ignoring the eavesdropper is in general not an optimal strategy even at low SNR.

### C. Masked Beamforming

In our basic model, the channel gains are fixed and known to all the terminals. Our capacity-achieving scheme in Theorem 2 uses the knowledge of $\mathbf{H}_{\mathrm{e}}$ for selecting the beamforming direction. However, in many applications, it may be difficult to know the eavesdropper's channel. Accordingly, in this section we analyze a simple alternative scheme that uses only knowledge of $\mathbf{h}_{\mathrm{r}}$ in choosing the transmit directions, yet achieves near-optimal performance in the high SNR regime.

The scheme we analyze is a masked beamforming scheme described in [4], [5]. In this scheme, the transmitter signals isotropically (i.e., with a covariance that is a scaled identity matrix), and as such can be naturally viewed as a "secure space-time code." More specifically, it simultaneously transmits the message (encoded using a scalar Gaussian wiretap code) in the direction corresponding to the intended receiver's channel $\mathbf{h}_{\mathrm{r}}$ while transmitting synthesized spatio-temporal white noise in the orthogonal subspace (i.e., all other directions).

The performance of masked beamforming is given by the following proposition, which is proved in Section VII-A.

*Proposition 1 (Masked Beamforming Secrecy Rate):* A rate achievable by the masked beamforming scheme for the MISOME channel is

$$R_{\mathrm{MB}}(P) = \left\{\log \lambda_{\max}\left(\frac{P}{n_{\mathrm{t}}}\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I} + \frac{P}{n_{\mathrm{t}}}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}\right) + \log\left(1 + \frac{n_{\mathrm{t}}}{P\|\mathbf{h}_{\mathrm{r}}\|^2}\right)\right\}^{+}. \tag{15}$$

While the rate (15) is, in general, suboptimal, it asymptotically near-optimal in the following sense, as developed in Section VII-B.

*Theorem 3:* The rate $R_{\mathrm{MB}}(P)$ achievable by masked beamforming scheme for the MISOME case [cf. (15)] satisfies

$$\lim_{P\to\infty}\left[C\left(\frac{P}{n_{\mathrm{t}}}\right) - R_{\mathrm{MB}}(P)\right] = 0. \tag{16}$$

[4]If there is more than one generalized eigenvector for $\lambda_{\max}$, we choose any one of them.

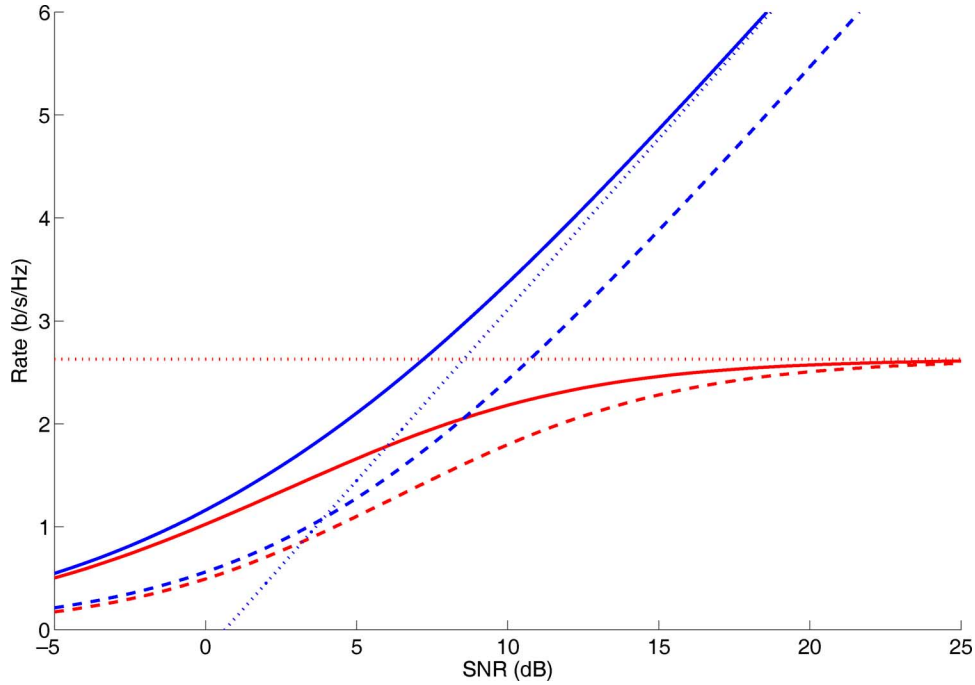


Fig. 1. Performance over an example MISOME channel with $n_{\mathrm{t}} = 2$ transmit antennas. The successively lower solid curves give the secrecy capacity for $n_{\mathrm{e}} = 1$ and $n_{\mathrm{e}} = 2$ eavesdropper antennas, respectively, and the dotted curves indicat the corresponding high-SNR asymptote. The dashed curves give the corresponding rates achievable by masked beamforming.

From the relation in (16) we note that, in the high SNR regime, the masked beamforming scheme achieves a rate of $C(P/n_{\mathrm{t}})$, where $n_{\mathrm{t}}$ is the number of transmit antennas. Combining (16) with (13), we see that the asymptotic masked beamforming loss is at most $\log n_{\mathrm{t}}$ b/s/Hz, or equivalently $10 \log_{10} n_{\mathrm{t}}$ dB in SNR. Specifically

$$\lim_{P \to \infty} [C(P) - R_{\mathrm{MB}}(P)] = \begin{cases} \log n_{\mathrm{t}}, & \mathbf{H}_{\mathrm{e}}^{\sharp} \mathbf{h}_{\mathrm{r}} \neq \mathbf{0} \\ 0, & \mathbf{H}_{\mathrm{e}}^{\sharp} \mathbf{h}_{\mathrm{r}} = \mathbf{0}. \end{cases} \tag{17}$$

That at least some loss (if vanishing) is associated with the masked beamforming scheme is expected, since the capacity-achieving scheme performs beamforming to concentrate the transmission along the optimal direction, whereas the masked beamforming scheme uses isotropic inputs. As one final comment, note that although the covariance structure of the masked beamforming transmission does not depend on the eavesdropper's channel, the rate of the base (scalar Gaussian wiretap) code does, as (15) reflects. In practice, the selection of this rate determines an insecurity zone around the sender, whereby the transmission is secure from eavesdroppers outside this zone, but insecure from ones inside.

### D. Example

In this section, we illustrate the preceding results for a typical MISOME channel. In our example, there are $n_{\mathrm{t}} = 2$ transmit antennas, and $n_{\mathrm{e}} = 2$ eavesdropper antennas. The channel to the receiver is

$$\mathbf{h}_{\mathrm{r}} = [0.0991 + j0.8676 \quad 1.0814 - j1.1281]^{\mathrm{T}}$$

while the channel to the eavesdropper is

$$\mathbf{H}_{\mathrm{e},1} = \begin{bmatrix} 0.3880 + j1.2024 & -0.9825 + j0.5914 \\ 0.4709 - j0.3073 & 0.6815 - j0.2125 \end{bmatrix} \tag{18}$$

where $j = \sqrt{-1}$.

Fig. 1 depicts communication rate as a function of SNR. The upper and lower solid curves depict the secrecy capacity (11) when the eavesdropper is using one or both its antennas, respectively.[5] As the curves reflect, when the eavesdropper has only a single antenna, the transmitter can securely communicate at any desired rate to its intended receiver by using enough power. However, by using both its antennas, the eavesdropper caps the rate at which the transmitter can communicate securely regardless of how much power it has available. Note that the lower and upper curves are representative of the cases where $\mathbf{H}_{\mathrm{e}}^{\sharp} \mathbf{h}_{\mathrm{r}}$ is, and is not $\mathbf{0}$, respectively.

Fig. 1 also shows other curves of interest. In particular, using dotted curves we superimpose the secrecy capacity high-SNR asymptotes as given by (13). As is apparent, these asymptotes can be quite accurate approximations even for moderate values of SNR. Finally, using dashed curves we show the rate (15) achievable by the masked beamforming coding scheme, which doesn't use knowledge of the eavesdropper channel. Consistent with (17), the loss in performance at high SNR approaches 3 dB when the eavesdropper uses only one of its antennas, and 0 when it uses both. Again, these are good estimates of the performance loss even at moderate SNR. Thus, the penalty for ignorance of the eavesdropper's channel can be quite small in practice.

[5]When a single eavesdropper antenna is in use, the relevant channel corresponds to the first row of (18).

### E. Scaling Laws in the Large System Limit

Our analysis in Section IV-B of the scaling behavior of capacity with SNR in the high SNR limit with a fixed number of antennas in the system yielded several useful insights into secure space-time coding systems. In this section, we develop equally valuable insights from a complementary scaling. In particular, we consider the scaling behavior of capacity with the number of antennas in the large system limit at a fixed SNR.

One convenient feature of such analysis is that for many large ensembles of channel gains, almost all randomly drawn realizations produce the same capacity asymptotes. For our analysis, we restrict our attention to an ensemble corresponding to Rayleigh fading in which $\mathbf{h}_{\mathrm{r}}$ and $\mathbf{H}_{\mathrm{e}}$ are independent, and each has i.i.d. $\mathcal{CN}(0,1)$ entries. The realization from the ensemble is known to all terminals prior to communication.

In anticipation of our analysis, we make the dependency of secrecy rates on the number of transmit and eavesdropper antennas explicit in our notation (but leave the dependency on the realization of $\mathbf{h}_{\mathrm{r}}$ and $\mathbf{H}_{\mathrm{e}}$ implicit). Specifically, we now use $C(P,n_{\mathrm{t}},n_{\mathrm{e}})$ to denote the secrecy capacity, and $R_{\mathrm{MB}}(P,n_{\mathrm{t}},n_{\mathrm{e}})$ to denote the rate of the masked beamforming scheme. With this notation, the scaled rates of interest are

$$\tilde{C}(\gamma,\beta)=\lim_{n_{\mathrm{t}}\to\infty} C\left(P{=}\gamma/n_{\mathrm{t}}, n_{\mathrm{t}}, n_{\mathrm{e}}{=}\beta n_{\mathrm{t}}\right) \tag{19a}$$

and

$$\tilde{R}_{\mathrm{MB}}(\gamma,\beta)=\lim_{n_{\mathrm{t}}\to\infty} R_{\mathrm{MB}}(P{=}\gamma, n_{\mathrm{t}}, n_{\mathrm{e}}{=}\beta n_{\mathrm{t}}). \tag{19b}$$

Our choice of scalings ensures that the $\tilde{C}(\gamma,\beta)$ and $\tilde{R}_{\mathrm{MB}}(\gamma,\beta)$ are not degenerate. In particular, note that the capacity scaling (19a) involves an SNR normalization. In particular, the transmitted power $P$ is reduced as the number of transmitter antennas $n_{\mathrm{t}}$ grows so as to keep the *received* SNR remains fixed (at specified value $\gamma$) independent of $n_{\mathrm{t}}$. However, the scaling (19b) is not SNR normalized in this way. This is because the masked beamforming already suffers a nominal factor of $n_{\mathrm{t}}$ SNR loss [cf. (16)] relative to a capacity-achieving system.

In what follows, we do not attempt an exact evaluation of the secrecy rates for our chosen scalings. Rather we find compact lower and upper bounds that are tight in the high SNR limit.

We begin with our lower bound, which is derived in Section VIII-B.

*Corollary 3 (Scaling Laws):* The asymptotic secrecy capacity satisfies

$$\tilde{C}(\gamma,\beta) \overset{\text{a.s.}}{\geq} \{\log \xi(\gamma,\beta)\}^{+} \tag{20}$$

where

$$\xi(\gamma,\beta)= \gamma-\frac{1}{4}\left[\sqrt{1{+}\gamma\left(1{+}\sqrt{\beta}\right)^2}-\sqrt{1{+}\gamma\left(1{-}\sqrt{\beta}\right)^2}\right]^2. \tag{21}$$

Furthermore, the same bound holds for the corresponding asymptotic masked beamforming rate, i.e.,

$$\tilde{R}_{\mathrm{MB}}(\gamma,\beta) \overset{\text{a.s.}}{\geq} \{\log \xi(\gamma,\beta)\}^{+}. \tag{22}$$

Since the secrecy rates increase monotonically with SNR, the infinite-SNR rates constitute a useful upper bound. As derived in Section VIII-C, this bound is as follows.

*Corollary 4:* The asymptotic secrecy capacity satisfies

$$\begin{aligned}\tilde{C}(\gamma,\beta) &\leq \lim_{n_{\mathrm{t}}\to\infty}\lim_{P\to\infty} C(P,n_{\mathrm{t}},\beta n_{\mathrm{t}}) \\ &\overset{\text{a.s.}}{=} \tilde{C}(\infty,\beta) \triangleq \begin{cases} 0, & \beta\geq 2 \\ -\log(\beta-1), & 1<\beta<2 \\ \infty, & \beta\leq 1. \end{cases}\end{aligned} \tag{23}$$

Furthermore, the right hand side of (23) is also an upper bound on $\tilde{R}_{\mathrm{MB}}(\gamma,\beta)$, i.e.,

$$\begin{aligned}\tilde{R}_{\mathrm{MB}}(\gamma,\beta) &\leq \lim_{n_{\mathrm{t}}\to\infty}\lim_{P\to\infty} R_{\mathrm{MB}}(P,n_{\mathrm{t}},\beta n_{\mathrm{t}}) \\ &\overset{\text{a.s.}}{=} \tilde{C}(\infty,\beta).\end{aligned} \tag{24}$$

Note that it is straightforward to verify that the lower bound (20) is tight at high SNR, i.e., that, for all $\beta$

$$\{\log \xi(\infty,\beta)\}^{+} = \tilde{C}(\infty,\beta). \tag{25}$$

The same argment confirms the corresponding behavior for masked beamforming.

Our lower and upper bounds of Corollary 3 and 4, respectively, are depicted in Fig. 2. In particular, we plot rate as a function of the antenna ratio $\beta$ for various values of the SNR $\gamma$.

As Fig. 2 reflects, there are essentially three main regions of behavior. For $\beta \leq 1$, the eavesdropper is effectively thwarted and any desired required rate is achieved. Second, for $1 \leq \beta < 2$ the eavesdropper has proportionally more antennas than the sender, and thus can cap the secure rate achievable to the receiver regardless of how much power the transmitter has available.

Finally for $\beta \geq 2$, the eavesdropper is able to entirely prevent secure communication (drive the secrecy capacity to zero) even if the transmitter has unlimited power available. Useful intuition for this phenomenon is obtained from consideration of the masked beamforming scheme, in which the sender transmits the signal of interest in the direction of $\mathbf{h}_{\mathrm{r}}$ and synthesized noise in the $n_{\mathrm{t}}-1$ directions orthogonal to $\mathbf{h}_{\mathrm{r}}$. With such a transmission, the intended receiver experiences a channel gain of $\|\mathbf{h}_{\mathrm{r}}\|^2 P/n_{\mathrm{t}}$. In the high SNR regime, the eavesdropper must cancel the synthesized noise, which requires at least $n_{\mathrm{t}}-1$ receive antennas. Moreover, after canceling the noise it must have the "beamforming gain" of $n_{\mathrm{t}}$ so its channel quality is of the same order as that of the intended receiver. This requires having at least $n_{\mathrm{t}}$ more antennas. Thus, at least $2n_{\mathrm{t}}-1$ antennas are required by the eavesdropper to guarantee successful interception of the transmission irrespective of the power used, which corresponds to $\beta \geq 2$ as $n_{\mathrm{t}} \to \infty$.

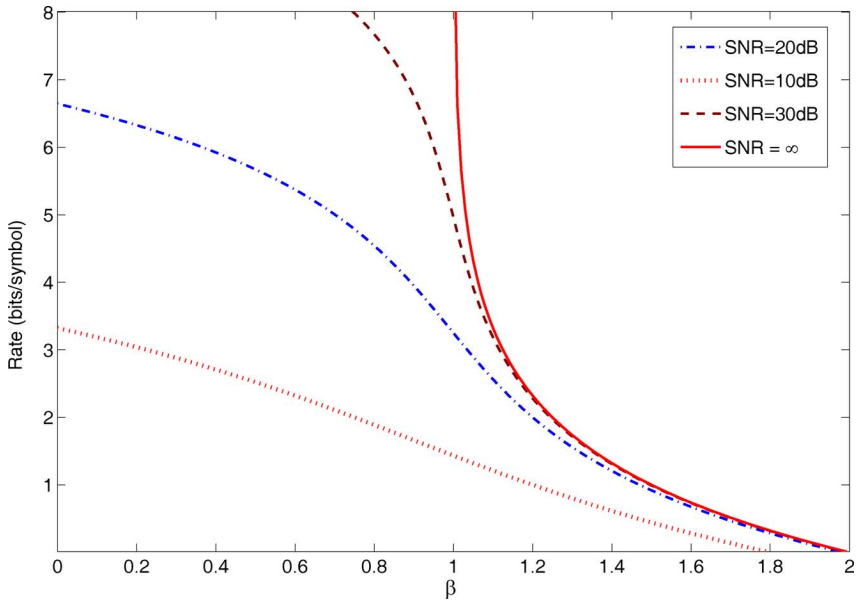


Fig. 2. Secrecy capacity bounds in the large system limit. The solid red curve is the high SNR secrecy capacity, which is an upper bound on the for finite SNR. The progressively lower dashed curves are lower bounds on the asymptotic secrecy capacity (and masked beamforming secrecy rate). The channel realizations are fixed but drawn at random according to Gaussian distribution.

### *F. Capacity Bounds in Fading*

Thus far, we have focused on the scenarios where the receiver and eavesdropper channels are fixed for the duration $n$ of the message transmission. In this section, we briefly turn our attention to the case of time-varying channels, specifically the case of fast fading where there are many channel fluctuations during the course of transmission. In particular, we consider a model in which $\mathbf{h}_{\mathrm{r}}(t)$ and $\mathbf{H}_{\mathrm{e}}(t)$ are temporally and spatially i.i.d. sequences that are independent of one another and have $\mathcal{CN}(0,1)$ elements, corresponding to Rayleigh fading.

In our model, $\mathbf{h}_{\mathrm{r}}(t)$ is known (in a causal manner) to all the three terminals, but only the eavesdropper has knowledge of $\mathbf{H}_{\mathrm{e}}(t)$. Accordingly, the channel model is, for $t = 1, 2, \ldots,$

$$\begin{aligned} \mathsf{y}_{\mathrm{r}}(t) &= \mathbf{h}_{\mathrm{r}}{}^{\dagger}(t)\mathbf{x}(t) + \mathsf{z}_{\mathrm{r}}(t) \\ \mathbf{y}_{\mathrm{e}}(t) &= \mathbf{H}_{\mathrm{e}}(t)\mathbf{x}(t) + \mathbf{z}_{\mathrm{e}}(t). \end{aligned} \tag{26}$$

The definition of the secrecy rate and capacity is as in Definition 2, with the exception that the equivocation $I(\mathsf{w}; \mathbf{y}_{\mathrm{e}}^n)$ is replaced with $I(\mathsf{w}; \mathbf{y}_{\mathrm{e}}^n, \mathbf{H}_{\mathrm{e}}{}^n | \mathbf{h}_{\mathrm{r}}{}^n)$, which takes into account the channel state information at the different terminals.

For this model, we have the following nontrivial upper and lower bounds on the secrecy capacity, which are developed in Section IX. The upper bound is developed via the same genie-aided channel analysis used in the proof of Theorem 2, but with modifications to account for the presence of fading. The lower bound is achieved by the adaptive version of masked beamforming described in [4].

*Theorem 4:* The secrecy capacity for the MISOME fast fading channel (26) is bounded by

$$C_{\mathrm{FF}}(P, n_{\mathrm{t}}, n_{\mathrm{e}}) \geq \max_{\rho(\cdot)\in\mathcal{P}_{\mathrm{FF}}} E[R_{\mathrm{FF},-}(\mathbf{h}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}, \rho(\cdot))] \tag{27a}$$

$$C_{\mathrm{FF}}(P, n_{\mathrm{t}}, n_{\mathrm{e}}) \leq \max_{\rho(\cdot)\in\mathcal{P}_{\mathrm{FF}}} E[R_{\mathrm{FF},+}(\mathbf{h}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}, \rho(\cdot))] \tag{27b}$$

where $\mathcal{P}_{\mathrm{FF}}$ is the set of all valid power allocations, i.e.,

$$\mathcal{P}_{\mathrm{FF}} = \left\{\rho(\cdot) \;\middle|\; \rho(\cdot) \geq 0,\ E[\rho(\mathbf{h}_{\mathrm{r}})] \leq P\right\} \tag{28}$$

and

$$\begin{aligned} R_{\mathrm{FF},-}(\mathbf{h}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}, \rho(\cdot)) \triangleq \\ \log\left(\frac{\rho(\mathbf{h}_{\mathrm{r}})}{n_{\mathrm{t}}}\mathbf{h}_{\mathrm{r}}^{\dagger}\left[\mathbf{I} + \frac{\rho(\mathbf{h}_{\mathrm{r}})}{n_{\mathrm{t}}}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}\right]^{-1}\mathbf{h}_{\mathrm{r}}\right) \\ + \log\left(1 + \frac{n_{\mathrm{t}}}{\rho(\mathbf{h}_{\mathrm{r}})\|\mathbf{h}_{\mathrm{r}}\|^2}\right). \end{aligned} \tag{29a}$$

$$\begin{aligned} R_{\mathrm{FF},+}(\mathbf{h}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}, \rho(\cdot)) \triangleq \\ \left\{\log\lambda_{\max}(\mathbf{I} + \rho(\mathbf{h}_{\mathrm{r}})\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I} + \rho(\mathbf{h}_{\mathrm{r}})\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\right\}^{+}. \end{aligned} \tag{29b}$$

In general, our upper and lower bounds do not coincide. Indeed, even in the case of single antennas at all terminals ($n_{\mathrm{t}} = n_{\mathrm{e}} = 1$), the secrecy capacity for the fading channel is unknown, except in the case of large coherence period [8]. However, based on our scaling analysis in Section IV-E, there is one regime in which the capacity can be calculated: in the limit of both high SNR and a large system. Indeed, since (22) and (23) hold for almost every channel realization, we have the following proposition, whose proof is provided in Section IX-C.

*Proposition 2:* The secrecy capacity of the fast fading channel satisfies

$$\lim_{n_{\mathrm{t}}\to\infty} C_{\mathrm{FF}}(P{=}\gamma, n_{\mathrm{t}}, n_{\mathrm{e}}{=}\beta n_{\mathrm{t}}) \geq \{\log\xi(\gamma,\beta)\}^{+} \tag{30}$$

where $\xi(\cdot,\cdot)$ is as defined in (21), and

$$\lim_{n_{\mathrm{t}}\to\infty} C_{\mathrm{FF}}(P{=}\gamma, n_{\mathrm{t}}, n_{\mathrm{e}}{=}\beta n_{\mathrm{t}}) \leq \tilde{C}(\infty,\beta) \tag{31}$$

with the $\tilde{C}(\infty,\beta)$ as given in (23).

Finally, via (25), we see that (30) and (31) converge as $\gamma \to \infty$.

This concludes our statement of the main results. The following sections are devoted to the proofs of these results and some further discussion.

## V. Upper Bound Derivation

In this section, we prove Theorem 1. We begin with the following lemma, which establishes that the capacity of genie-aided channel is an upper bound on the channel of interest. A proof is provided in Appendix I, and closely follows the general converse of Wyner [1], but differs in that the latter was for discrete channels and thus did not incorporate a power constraint.

*Lemma 1:* An upper bound on the secrecy capacity of the MISOME wiretap channel is

$$C \leq \max_{p_{\mathbf{x}} \in \mathcal{P}} I(\mathbf{x}; \mathsf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) \tag{32}$$

where $\mathcal{P}$ is the set of all probability distributions that satisfy $E[\|\mathbf{x}\|^2] \leq P$.

Among all such bounds, we can choose that corresponding to the noises $(\mathsf{z}_{\mathrm{r}}, \mathbf{z}_{\mathrm{e}})$ being jointly Gaussian (they are already constrained to be marginally Gaussian) with a covariance making the bound as small as possible. Then, provided the maximizing distribution in (32) is Gaussian, we can express the final bound in the form (7)

It thus remains only to show that the maximizing distribution is Gaussian.

*Lemma 2:* For each $\mathbf{K}_{\boldsymbol{\phi}} \in \mathcal{K}_{\boldsymbol{\phi}}$, the distribution $p_{\mathbf{x}}$ maximizing $I(\mathbf{x}; \mathsf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}})$ is Gaussian.

*Proof:* Since

$$I(\mathbf{x}; \mathsf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) = h(\mathsf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) - h(\mathsf{z}_{\mathrm{r}} | \mathbf{z}_{\mathrm{e}})$$

and the second term does not depend on $p_{\mathbf{x}}$, it suffices to establish that $h(\mathsf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}})$ is maximized when $\mathbf{x}$ is Gaussian.

To this end, let $\boldsymbol{\alpha}_{\mathrm{LMMSE}} \mathbf{y}_{\mathrm{e}}$ denote the linear minimum mean-square error (MMSE) estimator of $\mathsf{y}_{\mathrm{r}}$ from $\mathbf{y}_{\mathrm{e}}$, and $\lambda_{\mathrm{LMMSE}}$ the corresponding mean-square estimation error. Recall that

$$\boldsymbol{\alpha}_{\mathrm{LMMSE}} = (\mathbf{h}_{\mathrm{r}}^\dagger \mathbf{K}_P \mathbf{H}_{\mathrm{e}}^\dagger + \boldsymbol{\phi}^\dagger)(\mathbf{I} + \mathbf{H}_{\mathrm{e}} \mathbf{K}_P \mathbf{H}_{\mathrm{e}}^\dagger)^{-1} \tag{33}$$

$$\lambda_{\mathrm{LMMSE}} = 1 + \mathbf{h}_{\mathrm{r}}^\dagger \mathbf{K}_P \mathbf{h}_{\mathrm{r}} - (\mathbf{h}_{\mathrm{r}}^\dagger \mathbf{K}_P \mathbf{H}_{\mathrm{e}}^\dagger + \boldsymbol{\phi}^\dagger)(\mathbf{I} + \mathbf{H}_{\mathrm{e}} \mathbf{K}_P \mathbf{H}_{\mathrm{e}}^\dagger)^{-1} (\boldsymbol{\phi} + \mathbf{H}_{\mathrm{e}} \mathbf{K}_P \mathbf{h}_{\mathrm{r}}) \tag{34}$$

depend on the input and noise distributions only through their (joint) second-moment characterization, i.e.,

$$\mathbf{K}_P = \operatorname{cov} \mathbf{x}, \qquad \mathbf{K}_{\boldsymbol{\phi}} = \begin{bmatrix} 1 & \boldsymbol{\phi}^\dagger \\ \boldsymbol{\phi} & \mathbf{I} \end{bmatrix} = \operatorname{cov} \begin{bmatrix} \mathsf{z}_{\mathrm{r}} \\ \mathbf{z}_{\mathrm{e}} \end{bmatrix}. \tag{35}$$

Proceeding, we have

$$h(\mathsf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) = h(\mathsf{y}_{\mathrm{r}} - \boldsymbol{\alpha}_{\mathrm{LMMSE}} \mathbf{y}_{\mathrm{e}} | \mathbf{y}_{\mathrm{e}}) \tag{36}$$

$$\leq h(\mathsf{y}_{\mathrm{r}} - \boldsymbol{\alpha}_{\mathrm{LMMSE}} \mathbf{y}_{\mathrm{e}}) \tag{37}$$

$$\leq \log 2\pi e \lambda_{\mathrm{LMMSE}} \tag{38}$$

where (36) holds because adding a constant doesn't change entropy, (37) holds because conditioning only reduces differential entropy, and (38) is the maximum entropy bound on differential entropy expressed in terms of

$$\operatorname{var} \mathsf{e} = \lambda_{\mathrm{LMMSE}} \tag{39}$$

where $\mathsf{e}$ is the estimation error

$$\mathsf{e} = (\mathsf{y}_{\mathrm{r}} - \boldsymbol{\alpha}_{\mathrm{LMMSE}} \mathbf{y}_{\mathrm{e}}). \tag{40}$$

It remains only to verify that the above inequalities are tight for a Gaussian distribution. To see this, note that (37) holds with equality when $\mathbf{x}$ is Gaussian (and thus $(\mathsf{y}_{\mathrm{r}}, \mathbf{y}_{\mathrm{e}})$ are jointly Gaussian) since in this case $\mathsf{e}$ is the (unconstrained) MMSE estimation error and is, therefore, independent of the "data" $\mathbf{y}_{\mathrm{e}}$. Furthermore, note that in this case (38) holds with equality since the Gaussian distribution maximizes differential entropy subject to a variance constraint. ■

## VI. MISOME Secrecy Capacity Derivation

In this section, we derive the MISOME capacity and its high and low SNR asymptotes.

### A. Proof of Theorem 2

Achievability of (11) follows from evaluating (6) with the particular choices

$$\mathsf{u} \sim \mathcal{CN}(0, P), \quad \mathbf{x} = \boldsymbol{\psi}_{\max} \mathsf{u} \tag{41}$$

where $\boldsymbol{\psi}_{\max}$ is as defined in Theorem 2. With this choice of parameters

$$I(\mathsf{u}; \mathsf{y}_{\mathrm{r}}) - I(\mathsf{u}; \mathbf{y}_{\mathrm{e}})$$

$$= I(\mathbf{x}; \mathsf{y}_{\mathrm{r}}) - I(\mathbf{x}; \mathbf{y}_{\mathrm{e}}) \tag{42}$$

$$= \log\left(1 + P|\mathbf{h}_{\mathrm{r}}^\dagger \boldsymbol{\psi}_{\max}|^2\right) - \log\left(1 + P\|\mathbf{H}_{\mathrm{e}} \boldsymbol{\psi}_{\max}\|^2\right) \tag{43}$$

$$= \log \frac{\boldsymbol{\psi}_{\max}^\dagger (\mathbf{I} + P \mathbf{h}_{\mathrm{r}} \mathbf{h}_{\mathrm{r}}^\dagger) \boldsymbol{\psi}_{\max}}{\boldsymbol{\psi}_{\max}^\dagger (\mathbf{I} + P \mathbf{H}_{\mathrm{e}}^\dagger \mathbf{H}_{\mathrm{e}}) \boldsymbol{\psi}_{\max}}$$

$$= \log \lambda_{\max}(\mathbf{I} + P \mathbf{h}_{\mathrm{r}} \mathbf{h}_{\mathrm{r}}^\dagger, \mathbf{I} + P \mathbf{H}_{\mathrm{e}}^\dagger \mathbf{H}_{\mathrm{e}}) \tag{44}$$

where (42) follows from the fact that $\mathbf{x}$ is a deterministic function of $\mathsf{u}$, (43) follows from the choice of $\mathbf{x}$ and $\mathsf{u}$ in (41), and (44) follows from the variational characterization of generalized eigenvalues (2).

We next show a converse, that rates greater than (11) are not achievable using our upper bound. Specifically, we show that (11) corresponds to our upper bound expression (7) in Theorem 1.

It suffices to show that a particular choice of $\boldsymbol{\phi}$ that is admissible (i.e., such that $\mathbf{K}_{\boldsymbol{\phi}} \in \mathcal{K}_{\boldsymbol{\phi}}$) minimizes (7). We can do this by showing that

$$\max_{\mathbf{K}_P \in \mathcal{K}_P} R_+(\mathbf{K}_P, \mathbf{K}_{\boldsymbol{\phi}}) \tag{45}$$

with the chosen $\boldsymbol{\phi}$ corresponds to (11).

Since only the first term on the right hand side of

$$R_+(\mathbf{K}_P, \mathbf{K}_{\boldsymbol{\phi}}) = I(\mathbf{x}; \mathsf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) = h(\mathsf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) - h(\mathsf{z}_{\mathrm{r}} | \mathbf{z}_{\mathrm{e}})$$

depends on $\mathbf{K}_P$, we can restrict our attention to maximizing this first term with respect to $\mathbf{K}_P$.

Proceeding, exploiting that all variables are jointly Gaussian, we express this first term in the form of the optimization

$$\begin{aligned} h(\mathrm{y_r}|\mathbf{y_e}) &= \min_{\boldsymbol{\theta}\in\mathbb{C}^{n_e}} h(\mathrm{y_r}-\boldsymbol{\theta}^\dagger \mathbf{y_e}) \\ &= \min_{\boldsymbol{\theta}\in\mathbb{C}^{n_e}} h((\mathbf{h_r}-\mathbf{H_e}^\dagger\boldsymbol{\theta})^\dagger \mathbf{x} + z_r - \boldsymbol{\theta}^\dagger \mathrm{z_e}) \\ &= \min_{\boldsymbol{\theta}\in\mathbb{C}^{n_e}} \log\big[(\mathbf{h_r}-\mathbf{H_e}^\dagger\boldsymbol{\theta})^\dagger \mathbf{K}_P(\mathbf{h_r}-\mathbf{H_e}^\dagger\boldsymbol{\theta}) \\ &\quad + 1 + ||\boldsymbol{\theta}||^2 - 2\mathrm{Re}\{\boldsymbol{\theta}^\dagger\boldsymbol{\phi}\}\big] \end{aligned} \tag{46}$$

and bound its maximimum over $\mathbf{K}_P$ according to

$$\begin{aligned} &\max_{\mathbf{K}_P\in\mathcal{K}_P} h(\mathrm{y_r}|\mathbf{y_e}) \\ &\quad= \max_{\mathbf{K}_P\in\mathcal{K}_P} \min_{\boldsymbol{\theta}\in\mathbb{C}^{n_e}} \log\big[(\mathbf{h_r}-\mathbf{H_e}^\dagger\boldsymbol{\theta})^\dagger \mathbf{K}_P(\mathbf{h_r}-\mathbf{H_e}^\dagger\boldsymbol{\theta}) \\ &\qquad + 1 + ||\boldsymbol{\theta}||^2 - 2\,\mathrm{Re}\{\boldsymbol{\theta}^\dagger\boldsymbol{\phi}\}\big] \\ &\quad\le \min_{\boldsymbol{\theta}\in\mathbb{C}^{n_e}} \max_{\mathbf{K}_P\in\mathcal{K}_P} \log\big[(\mathbf{h_r}-\mathbf{H_e}^\dagger\boldsymbol{\theta})^\dagger \mathbf{K}_P(\mathbf{h_r}-\mathbf{H_e}^\dagger\boldsymbol{\theta}) \\ &\qquad + 1 + ||\boldsymbol{\theta}||^2 - 2\,\mathrm{Re}\{\boldsymbol{\theta}^\dagger\boldsymbol{\phi}\}\big] \\ &\quad= \min_{\boldsymbol{\theta}\in\mathbb{C}^{n_e}} \log\big[P||\mathbf{h_r}-\mathbf{H_e}^\dagger\boldsymbol{\theta}||^2 + 1 + ||\boldsymbol{\theta}||^2 - 2\,\mathrm{Re}\{\boldsymbol{\theta}^\dagger\boldsymbol{\phi}\}\big] \end{aligned} \tag{47}$$

where (47) follows[6] by observing that a rank one $\mathbf{K}_P$ maximizes the quadratic form $(\mathbf{h_r}-\mathbf{H_e}^\dagger\boldsymbol{\theta})^\dagger \mathbf{K}_P(\mathbf{h_r}-\mathbf{H_e}^\dagger\boldsymbol{\theta})$.

We now separately consider the cases $\lambda_{\max} > 1$ and $\lambda_{\max} \le 1$.

*Case:* $\lambda_{\max} > 1$*:* We show that the choice

$$\boldsymbol{\phi} = \frac{\mathbf{H_e}\boldsymbol{\psi}_{\max}}{\mathbf{h_r}^\dagger\boldsymbol{\psi}_{\max}} \tag{48}$$

in (45) yields (11), i.e., $\log\lambda_{\max}$.

We begin by noting that since $\lambda_{\max} > 1$, the variational characterization (2) establishes that $||\boldsymbol{\phi}|| < 1$ and thus $\mathbf{K}_{\boldsymbol{\phi}} \in \mathcal{K}_{\boldsymbol{\phi}}$ as defined in (10).

Then, provided that, with $\boldsymbol{\phi}$ as given in (48), the right hand side of (47) evaluates to

$$\begin{aligned} &\min_{\boldsymbol{\theta}\in\mathbb{C}^{n_e}} \log\big[P||\mathbf{h_r}-\mathbf{H_e}^\dagger\boldsymbol{\theta}||^2 + 1 + ||\boldsymbol{\theta}||^2 - 2\,\mathrm{Re}\{\boldsymbol{\theta}^\dagger\boldsymbol{\phi}\}\big] \\ &\qquad = \log\left(\lambda_{\max}\cdot(1-||\boldsymbol{\phi}||^2)\right) \end{aligned} \tag{49}$$

we have

$$\begin{aligned} R_+ &\le \max_{\mathbf{K}_P\in\mathcal{K}_P} R_+(\mathbf{K}_P, K_{\boldsymbol{\phi}}) \\ &= \max_{\mathbf{K}_P\in\mathcal{K}_P} h(\mathrm{y_r}|\mathbf{y_e}) - h(\mathrm{z_r}|\mathbf{z_e}) \\ &\le \log(\lambda_{\max}\cdot(1-||\boldsymbol{\phi}||^2)) - \log(1-||\boldsymbol{\phi}||^2) \\ &= \log(\lambda_{\max}) \end{aligned}$$

[6]The elegant derivation that rank-1 covariance matrix maximizes the differential entropy (46) was suggested to us by Y. C. Eldar and A. Wiesel. In the literature, this line of reasoning has been used in deriving an extremal characterization of the Schur complement of a matrix (see, e.g., [22, Chapter 20], [23]).

i.e., (11), as required. Verifying (49) with (48) is a straightforward computation, the details of which are provided in Appendix II.

*Case:* $\lambda_{\max} \le 1$*,* $\mathbf{H_e}$ *Full Column Rank:* We show that the choice

$$\boldsymbol{\phi} = \mathbf{H_e}(\mathbf{H_e}^\dagger\mathbf{H_e})^{-1}\mathbf{h_r} \tag{50}$$

in (45) yields (11), i.e., zero.

To verify that $||\boldsymbol{\phi}|| \le 1$, first note that since $\lambda_{\max} \le 1$, it follows from (2) that

$$\lambda_{\max}(\mathbf{I}+P\mathbf{h_r}\mathbf{h_r}^\dagger, \mathbf{I}+P\mathbf{H_e}^\dagger\mathbf{H_e}) \le 1 \Leftrightarrow \lambda_{\max}(\mathbf{h_r}\mathbf{h_r}^\dagger, \mathbf{H_e}^\dagger\mathbf{H_e}) \le 1 \tag{51}$$

so that for any choice of $\boldsymbol{\psi}$

$$\boldsymbol{\psi}^\dagger\mathbf{h_r}\mathbf{h_r}^\dagger\boldsymbol{\psi} \le \boldsymbol{\psi}^\dagger\mathbf{H_e}^\dagger\mathbf{H_e}\boldsymbol{\psi}. \tag{52}$$

Choosing $\boldsymbol{\psi} = (\mathbf{H_e}^\dagger\mathbf{H_e})^{-1}\mathbf{h_r}$ in (52) yields $||\boldsymbol{\phi}||^2 \le ||\boldsymbol{\phi}||$, i.e., $||\boldsymbol{\phi}|| \le 1$, as required.

Next, note that (47) is further upper bounded by choosing any particular choice of $\boldsymbol{\theta}$. Choosing $\boldsymbol{\theta} = \boldsymbol{\phi}$ yields

$$R_+ \le \log\left(\frac{P||\mathbf{h_r}-\mathbf{H_e}^\dagger\boldsymbol{\phi}||^2}{1-||\boldsymbol{\phi}||^2}+1\right) \tag{53}$$

which with the choice (50) for $\boldsymbol{\phi}$ is zero.

*Case:* $\lambda_{\max} \le 1$*,* $\mathbf{H_e}$ *Not Full Column Rank:* Consider a new MISOME channel with $n_t' < n_t$ transmit antennas, where $n_t'$ is the column rank of $\mathbf{H_e}$, where the intended receiver and eavesdropper channel gains are given by

$$\mathbf{g_r} = \mathbf{Q}^\dagger\mathbf{h_r}, \qquad \mathbf{G_e} = \mathbf{H_e}\mathbf{Q} \tag{54}$$

and where $\mathbf{Q}$ is a matrix whose columns constitute an orthogonal basis for the column space of $\mathbf{H_e}^\dagger$, so that in this new channel $\mathbf{G_e}$ has full rank. Then provided the new channel (54) has the same capacity as the original channel, it follows by the analysis of the previous case that the capacity of both channels is zero. Thus, it remains only to show the following.

*Claim 1:* The MISOME channel $(\mathbf{g_r}, \mathbf{G_e})$ corresponding to (54) has the same secrecy capacity as that corresponding to $(\mathbf{h_r}, \mathbf{H_e})$.

*Proof:* First we show that the new channel capacity is no larger than the original one. In particular, we have

$$\begin{aligned} &\lambda_{\max}(\mathbf{I}+P\mathbf{g_r}\mathbf{g_r}^\dagger, \mathbf{I}+P\mathbf{G_e}^\dagger\mathbf{G_e}) \\ &\qquad= \max_{\{\boldsymbol{\psi}':||\boldsymbol{\psi}'||=1\}} \left\{\frac{1+P|\mathbf{g_r}^\dagger\boldsymbol{\psi}'|^2}{1+P||\mathbf{G_e}\boldsymbol{\psi}'||^2}\right\} && (55) \\ &\qquad= \max_{\{\boldsymbol{\psi}':||\boldsymbol{\psi}'||=1\}} \frac{1+P|\mathbf{h_r}^\dagger\mathbf{Q}\boldsymbol{\psi}'|^2}{1+P||\mathbf{H_e}\mathbf{Q}\boldsymbol{\psi}'||^2} && (56) \\ &\qquad= \max_{\{\boldsymbol{\psi}:\boldsymbol{\psi}=\mathbf{Q}\boldsymbol{\psi}',||\boldsymbol{\psi}||=1\}} \frac{1+P|\mathbf{h_r}^\dagger\boldsymbol{\psi}|^2}{1+P||\mathbf{H_e}\boldsymbol{\psi}||^2} && (57) \\ &\qquad\le \max_{\{\boldsymbol{\psi}:||\boldsymbol{\psi}||=1\}} \left\{\frac{1+P|\mathbf{h_r}^\dagger\boldsymbol{\psi}|^2}{1+P||\mathbf{H_e}\boldsymbol{\psi}||^2}\right\} && (58) \\ &\qquad= \lambda_{\max}(\mathbf{I}+P\mathbf{h_r}\mathbf{h_r}^\dagger, \mathbf{I}+P\mathbf{H_e}^\dagger\mathbf{H_e}) && (59) \end{aligned}$$

where to obtain (55) we have used (2) for the new channel, to obtain (56) we have used (54), to obtain (57) we have used that $\mathbf{Q}^\dagger\mathbf{Q} = \mathbf{I}$, to obtain (58) we have used that we are maximizing over a larger set, and to obtain (59) we have used (2) for the original channel. Thus

$$\begin{aligned}\left\{\lambda_{\max}(\mathbf{I}+P\mathbf{g}_{\rm r}\mathbf{g}_{\rm r}^\dagger,\mathbf{I}+P\mathbf{G}_{\rm e}^\dagger\mathbf{G}_{\rm e})\right\}^+ \\ \le \left\{\lambda_{\max}(\mathbf{I}+P\mathbf{h}_{\rm r}\mathbf{h}_{\rm r}^\dagger,\mathbf{I}+P\mathbf{H}_{\rm e}^\dagger\mathbf{H}_{\rm e})\right\}^+ \quad (60)\end{aligned}$$

Next, we show the new channel capacity is no smaller than the original one. To begin, note that

$$\mathrm{Null}(\mathbf{H}_{\rm e}) \subseteq \mathrm{Null}(\mathbf{h}_{\rm r}^\dagger) \quad (61)$$

since if $\mathrm{Null}(\mathbf{H}_{\rm e}) \not\subseteq \mathrm{Null}(\mathbf{h}_{\rm r}^\dagger)$, then $\lambda_{\max}(\mathbf{h}_{\rm r}\mathbf{h}_{\rm r}^\dagger, \mathbf{H}_{\rm e}^\dagger\mathbf{H}_{\rm e}) = \infty$, which would violate (51).

Proceeding, every $\mathbf{x} \in \mathbb{C}^{n_{\rm t}}$ can we written as

$$\mathbf{x} = \mathbf{Q}\mathbf{x}' + \tilde{\mathbf{x}} \quad (62)$$

where $\mathbf{H}_{\rm e}\tilde{\mathbf{x}} = \mathbf{0}$ and thus, via (61), $\mathbf{h}_{\rm r}^\dagger\tilde{\mathbf{x}} = 0$ as well. Hence, we have that $\mathbf{h}_{\rm r}^\dagger\mathbf{x} = \mathbf{g}_{\rm r}^\dagger\mathbf{x}'$, $\mathbf{H}_{\rm e}\mathbf{x} = \mathbf{G}_{\rm e}\mathbf{x}'$, and $\|\mathbf{x}'\|^2 \le \|\mathbf{x}\|^2$, so any rate achieved by $p_{\mathbf{x}}$ on the channel $(\mathbf{h}_{\rm r}, \mathbf{H}_{\rm e})$ is also achieved by $p_{\mathbf{x}'}$ on the channel $(\mathbf{g}_{\rm r}, \mathbf{G}_{\rm e})$, with $p_{\mathbf{x}'}$ derived from $p_{\mathbf{x}}$ via (62), whence

$$\begin{aligned}\left\{\lambda_{\max}(\mathbf{I}+P\mathbf{g}_{\rm r}\mathbf{g}_{\rm r}^\dagger,\mathbf{I}+P\mathbf{G}_{\rm e}^\dagger\mathbf{G}_{\rm e})\right\}^+ \\ \ge \left\{\lambda_{\max}(\mathbf{I}+P\mathbf{h}_{\rm r}\mathbf{h}_{\rm r}^\dagger,\mathbf{I}+P\mathbf{H}_{\rm e}^\dagger\mathbf{H}_{\rm e})\right\}^+. \quad (63)\end{aligned}$$

Combining (63) and (60) establishes our claim. ■

### B. Proof of Corollary 1

We restrict our attention to the case $\lambda_{\max} > 1$ where the capacity is nonzero. In this case, since, via (2)

$$\lambda_{\max}(\mathbf{I}+P\mathbf{h}_{\rm r}\mathbf{h}_{\rm r}^\dagger,\mathbf{I}+P\mathbf{H}_{\rm e}^\dagger\mathbf{H}_{\rm e}) = \frac{1+P|\mathbf{h}_{\rm r}^\dagger\boldsymbol{\psi}_{\max}(P)|^2}{1+P\|\mathbf{H}_{\rm e}\boldsymbol{\psi}_{\max}(P)\|^2} > 1 \quad (64)$$

where

$$\boldsymbol{\psi}_{\max}(P) \triangleq \arg\max_{\{\boldsymbol{\psi}:\|\boldsymbol{\psi}\|=1\}} \frac{1+P|\mathbf{h}_{\rm r}^\dagger\boldsymbol{\psi}|^2}{1+P\|\mathbf{H}_{\rm e}\boldsymbol{\psi}\|^2}. \quad (65)$$

We have

$$|\mathbf{h}_{\rm r}^\dagger\boldsymbol{\psi}_{\max}(P)| > \|\mathbf{H}_{\rm e}\boldsymbol{\psi}_{\max}(P)\| \quad (66)$$

for all $P > 0$.

To obtain an upper bound note that, for all $P > 0$

$$\begin{aligned}\lambda_{\max}(\mathbf{I}+P\mathbf{h}_{\rm r}\mathbf{h}_{\rm r}^\dagger,\mathbf{I}+P\mathbf{H}_{\rm e}^\dagger\mathbf{H}_{\rm e}) \\ \le \frac{|\mathbf{h}_{\rm r}^\dagger\boldsymbol{\psi}_{\max}(P)|^2}{\|\mathbf{H}_{\rm e}\boldsymbol{\psi}_{\max}(P)\|^2} \quad (67) \\ \le \lambda_{\max}(\mathbf{h}_{\rm r}\mathbf{h}_{\rm r}^\dagger, \mathbf{H}_{\rm e}^\dagger\mathbf{H}_{\rm e}) \quad (68)\end{aligned}$$

where (67) follows from the Rayleigh quotient expansion (64) and the fact that, due to (66), the right hand side of (64) is increasing in $P$, and where (68) follows from (2). Thus, since the right hand side of (68) is independent of $P$, we have

$$\lim_{P\to\infty} \lambda_{\max}(\mathbf{I}+P\mathbf{h}_{\rm r}\mathbf{h}_{\rm r}^\dagger,\mathbf{I}+P\mathbf{H}_{\rm e}^\dagger\mathbf{H}_{\rm e}) \le \lambda_{\max}(\mathbf{h}_{\rm r}\mathbf{h}_{\rm r}^\dagger,\mathbf{H}_{\rm e}^\dagger\mathbf{H}_{\rm e}). \quad (69)$$

Next, defining

$$\boldsymbol{\psi}_{\max}(\infty) \triangleq \arg\max_{\boldsymbol{\psi}} \frac{|\mathbf{h}_{\rm r}^\dagger\boldsymbol{\psi}|^2}{\|\mathbf{H}_{\rm e}\boldsymbol{\psi}\|^2} \quad (70)$$

we have the lower bound

$$\begin{aligned}\lim_{P\to\infty} \lambda_{\max}(\mathbf{I}+P\mathbf{h}_{\rm r}\mathbf{h}_{\rm r}^\dagger,\mathbf{I}+P\mathbf{H}_{\rm e}^\dagger\mathbf{H}_{\rm e}) \\ \ge \lim_{P\to\infty} \frac{1/P+|\mathbf{h}_{\rm r}^\dagger\boldsymbol{\psi}_{\max}(\infty)|^2}{1/P+\|\mathbf{H}_{\rm e}\boldsymbol{\psi}_{\max}(\infty)\|^2} \quad (71) \\ = \lambda_{\max}(\mathbf{h}_{\rm r}\mathbf{h}_{\rm r}^\dagger, \mathbf{H}_{\rm e}^\dagger\mathbf{H}_{\rm e}) \quad (72)\end{aligned}$$

where (71) follows from (2) and (72) follows from (70).

Since (69) and (72) coincide, we obtain (12). Thus, to obtain the remainder of (13a), we need only verify the following.

*Claim 2:* The high SNR capacity is finite, i.e., $\lambda_{\max}(\mathbf{h}_{\rm r}\mathbf{h}_{\rm r}^\dagger, \mathbf{H}_{\rm e}^\dagger\mathbf{H}_{\rm e}) < \infty$, when $\mathbf{H}_{\rm e}^\sharp\mathbf{h}_{\rm r} = \mathbf{0}$.

*Proof:* We argue by contradiction. Suppose $\lambda_{\max}(\mathbf{h}_{\rm r}\mathbf{h}_{\rm r}^\dagger, \mathbf{H}_{\rm e}^\dagger\mathbf{H}_{\rm e}) = \infty$. Then there must exist a sequence $\boldsymbol{\psi}_k$ such that $\|\mathbf{H}_{\rm e}\boldsymbol{\psi}_k\| > 0$ for each $k = 1, 2, \ldots$, but $\|\mathbf{H}_{\rm e}\boldsymbol{\psi}_k\| \to 0$ as $k \to \infty$. But then the hypothesis cannot be true, because, as we now show, $|\mathbf{h}_{\rm r}^\dagger\boldsymbol{\psi}|^2/\|\mathbf{H}_{\rm e}\boldsymbol{\psi}\|^2$, when viewed as a function of $\boldsymbol{\psi}$, is bounded whenever the denominator is nonzero.

Let $\boldsymbol{\psi}$ be any vector such that $\|\mathbf{H}_{\rm e}\boldsymbol{\psi}\| \triangleq \delta > 0$. It suffices to show that

$$\frac{|\mathbf{h}_{\rm r}^\dagger\boldsymbol{\psi}|^2}{\|\mathbf{H}_{\rm e}\boldsymbol{\psi}\|^2} \le \frac{\|\mathbf{h}_{\rm r}\|^2}{\sigma^2} \quad (73)$$

where $\sigma^2$ is the smallest *nonzero* singular value of $\mathbf{H}_{\rm e}$. To verify (73), we first express $\boldsymbol{\psi}$ in the form

$$\boldsymbol{\psi} = c\boldsymbol{\psi}' + d\tilde{\boldsymbol{\psi}} \quad (74)$$

where $\boldsymbol{\psi}'$ and $\tilde{\boldsymbol{\psi}}$ are unit vectors, $c$ and $d$ are real and nonnegative, $d\tilde{\boldsymbol{\psi}}$ is the projection of $\boldsymbol{\psi}$ onto the null space of $\mathbf{H}_{\rm e}$, and $c\boldsymbol{\psi}'$ is the projection of $\boldsymbol{\psi}$ onto the orthogonal complement of this null space.

Next, we note that $\delta = \|\mathbf{H}_{\rm e}\boldsymbol{\psi}\| = c\|\mathbf{H}_{\rm e}\boldsymbol{\psi}'\| \ge c\sigma$, whence

$$c \le \frac{\delta}{\sigma} \quad (75)$$

but since $\mathbf{H}_{\rm e}^\sharp\mathbf{h}_{\rm r} = 0$ it follows that $\mathbf{h}_{\rm r}^\dagger\tilde{\boldsymbol{\psi}} = 0$, so

$$|\mathbf{h}_{\rm r}^\dagger\boldsymbol{\psi}|^2 = c^2|\mathbf{h}_{\rm r}^\dagger\boldsymbol{\psi}'|^2 \le c^2\|\mathbf{h}_{\rm r}\|^2 \le \frac{\delta^2}{\sigma^2}\|\mathbf{h}_{\rm r}\|^2 \quad (76)$$

where the first inequality follows from the Cauchy-Schwarz inequality, and the second inequality is a simple substitution from (75). Dividing through by $\|\mathbf{H}_{\rm e}\boldsymbol{\psi}\|^2 = \delta^2$ in (76) yields (73). ■

We now develop (13b) for the case where $\mathbf{H}_{\rm e}^\sharp\mathbf{h}_{\rm r} \neq \mathbf{0}$.

First, defining

$$\mathcal{S}_\infty = \{\boldsymbol{\psi} : \|\boldsymbol{\psi}\| = 1, \|\mathbf{H}_{\rm e}\boldsymbol{\psi}\| = 0\} \quad (77)$$

we obtain the lower bound

$$\begin{aligned}\frac{1}{P}\lambda_{\max}(\mathbf{I}+P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger},\mathbf{I}+P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})&\geq\max_{\boldsymbol{\psi}\in\mathcal{S}_{\infty}}\frac{1/P+|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}|^2}{1+P||\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}||^2}\\&=\max_{\boldsymbol{\psi}\in\mathcal{S}_{\infty}}\frac{1}{P}+|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}|^2 \quad (78)\\&=\frac{1}{P}+||\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}}||^2 \quad (79)\end{aligned}$$

where to obtain (79), we have used

$$\max_{\{\boldsymbol{\psi}:||\boldsymbol{\psi}||=1,\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}=0\}}\left|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}\right|^2=\left|\left|\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}}\right|\right|^2. \quad (80)$$

Next we develop an upper bound. We first establish the following.

*Claim 3:* If $\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}}\neq\mathbf{0}$ then there is a function $\varepsilon(P)$ such that $\varepsilon(P)\to 0$ as $P\to\infty$, and

$$||\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}_{\max}(P)||\leq\varepsilon(P).$$

*Proof:* We have

$$\begin{aligned}\frac{1+P||\mathbf{h}_{\mathrm{r}}||^2}{1+P||\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}_{\max}(P)||^2}&\geq\frac{1+P|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}(P)|^2}{1+P||\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}_{\max}(P)||^2} \quad (81)\\&\geq\max_{\{\boldsymbol{\psi}:\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}=\mathbf{0},||\boldsymbol{\psi}||=1\}}\frac{1+P|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}|^2}{1+P||\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}||^2} \quad (82)\\&=\max_{\{\boldsymbol{\psi}:\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}=\mathbf{0},||\boldsymbol{\psi}||=1\}}\left(1+P|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}|^2\right)\\&=1+P||\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}}||^2 \quad (83)\end{aligned}$$

where to obtain (81), we have used the Cauchy-Schwarz inequality $|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}(P)|^2\leq||\mathbf{h}_{\mathrm{r}}||^2$, to obtain (82), we have used (65), and to obtain (83), we have used (80).

Rearranging (83) then gives

$$||\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}_{\max}(P)||^2\leq\frac{1}{P}\left(\frac{1+P||\mathbf{h}_{\mathrm{r}}||^2}{1+P||\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}}||^2}-1\right)\triangleq\varepsilon^2(P).$$

as desired. ∎

Thus, with $\mathcal{S}_P=\{\boldsymbol{\psi}:||\boldsymbol{\psi}||=1,||\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}||\leq\varepsilon(P)\}$, we have

$$\begin{aligned}\frac{1}{P}\lambda_{\max}(\mathbf{I}+P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger},\mathbf{I}+P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})&=\max_{\boldsymbol{\psi}\in\mathcal{S}_P}\frac{1/P+|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}|^2}{1+P||\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}||^2} \quad (84)\\&\leq\max_{\boldsymbol{\psi}\in\mathcal{S}_P}\frac{1}{P}+|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}|^2 \quad (85)\end{aligned}$$

where (84) follows from (2) and Claim 3 that the maximizing $\boldsymbol{\psi}_{\max}$ lies in $\mathcal{S}_P$.

Now, as we will show

$$\max_{\boldsymbol{\psi}\in\mathcal{S}_P}|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}|^2\leq||\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}}||^2+\frac{\varepsilon^2(P)}{\sigma^2}||\mathbf{h}_{\mathrm{r}}||^2. \quad (86)$$

So using (86) in (85), we obtain

$$\begin{aligned}\frac{1}{P}\lambda_{\max}(\mathbf{I}+P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger},\mathbf{I}+P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})&\\\leq||\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}}||^2+\frac{\varepsilon^2(P)}{\sigma^2}||\mathbf{h}_{\mathrm{r}}||^2+\frac{1}{P}.& \quad (87)\end{aligned}$$

Finally, combining (87) and (79), we obtain

$$\lim_{P\to\infty}\frac{1}{P}\lambda_{\max}(\mathbf{I}+P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger},\mathbf{I}+P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})=||\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}}||^2$$

whence (13b).

Thus, it remains only to verify (86), which we do now.

We start by expressing $\boldsymbol{\psi}\in\mathcal{S}_P$ in the form [cf. (74)]

$$\boldsymbol{\psi}=c\boldsymbol{\psi}'+d\tilde{\boldsymbol{\psi}} \quad (88)$$

where $\boldsymbol{\psi}'$ and $\tilde{\boldsymbol{\psi}}$ are unit vectors, $c,d$ are real valued scalars in $[0,1]$, $d\tilde{\boldsymbol{\psi}}$ is the projection of $\boldsymbol{\psi}$ onto the null space of $\mathbf{H}_{\mathrm{e}}$, and $c\boldsymbol{\psi}'$ is the projection of $\boldsymbol{\psi}$ onto the orthogonal complement of this null space.

With these definitions, we have

$$\varepsilon(P)\geq||\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}||=c||\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}'||\geq c\sigma \quad (89)$$

since $\mathbf{H}_{\mathrm{e}}\tilde{\boldsymbol{\psi}}=\mathbf{0}$ and $||\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}'||\geq\sigma$. Finally

$$\begin{aligned}|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}|^2&=|d\mathbf{h}_{\mathrm{r}}^{\dagger}\tilde{\boldsymbol{\psi}}+c\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}'|^2 \quad (90)\\&=d^2|\mathbf{h}_{\mathrm{r}}^{\dagger}\tilde{\boldsymbol{\psi}}|^2+c^2|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}'|^2 \quad (91)\\&\leq|\mathbf{h}_{\mathrm{r}}^{\dagger}\tilde{\boldsymbol{\psi}}|^2+\frac{\varepsilon(P)^2}{\sigma^2}|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}'|^2 \quad (92)\\&\leq|\mathbf{h}_{\mathrm{r}}^{\dagger}\tilde{\boldsymbol{\psi}}|^2+\frac{\varepsilon(P)^2}{\sigma^2}||\mathbf{h}_{\mathrm{r}}||^2 \quad (93)\\&\leq||\mathbf{H}_{\mathrm{e}}^{\sharp}\mathbf{h}_{\mathrm{r}}||^2+\frac{\varepsilon(P)^2}{\sigma^2}||\mathbf{h}_{\mathrm{r}}||^2 \quad (94)\end{aligned}$$

where (90) follows from substituting (88), (91) follows from the fact that $\boldsymbol{\psi}'$, and $\tilde{\boldsymbol{\psi}}$ are orthogonal, (92) follows from using (89) to bound $c^2$, and (94) follows from the fact that $\mathbf{H}_{\mathrm{e}}\tilde{\boldsymbol{\psi}}=\mathbf{0}$ and (80).

### C. Proof of Corollary 2

We consider the limit $P\to 0$. In the following steps, the order notation $\mathcal{O}(P)$ means that $\mathcal{O}(P)/P\to 0$ as $P\to 0$

$$\begin{aligned}&\lambda_{\max}(\mathbf{I}+P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger},\mathbf{I}+P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}) \quad (95)\\&=\lambda_{\max}\left((\mathbf{I}+P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}(\mathbf{I}+P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger})\right) \quad (96)\\&=\lambda_{\max}\left((\mathbf{I}-P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}+\mathcal{O}(P))(\mathbf{I}+P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger})\right) \quad (97)\\&=\lambda_{\max}\left((\mathbf{I}-P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})(\mathbf{I}+P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger})\right)+\mathcal{O}(P) \quad (98)\\&=\lambda_{\max}\left(\mathbf{I}+P(\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}-\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\right)+\mathcal{O}(P) \quad (99)\\&=1+P\lambda_{\max}(\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}-\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})+\mathcal{O}(P) \quad (100)\end{aligned}$$

where (96) follows from the definition of generalized eigenvalue, (97) follows from the Taylor series expansion of $(\mathbf{I}+P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}$, where we have assumed that $P$ is sufficiently small so that all eigenvalues of $P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}$ are less than unity, (98) and (99) follow from the continuity of the eigenvalue function in its

arguments and (100) follows from the property of eigenvalue function that $\lambda(\mathbf{I}+\mathbf{A}) = 1+\lambda(\mathbf{A})$.

In turn, we have

$$\frac{C(P)}{P} = \frac{\log(1+P\lambda_{\max}(\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}-\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})+\mathcal{O}(P))}{P} \quad (101)$$

$$= \frac{\lambda_{\max}(\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger} - \mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})}{\ln 2} + \frac{\mathcal{O}(P)}{P} \quad (102)$$

where to obtain (101) we have used (100) in (11), and to obtain (102) we have used Taylor Series expansion of the $\ln(\cdot)$ function.

Finally, taking the limit $P \to 0$ in (102) yields (14) as desired.

## VII. Masked Beamforming Scheme Analysis

From Csiszár–Körner [2], secrecy rate $R = I(\mathsf{u};\mathsf{y}_{\mathrm{r}}) - I(\mathsf{u};\mathbf{y}_{\mathrm{e}})$ is achievable for any choice of $p_{\mathsf{u}}$ and $p_{\mathbf{x}|\mathsf{u}}$ that satisfy the power constraint $E[|\mathsf{x}|^2] \le P$. While a capacity-achieving scheme corresponds to maximizing this rate over the choice of $p_{\mathsf{u}}$ and $p_{\mathbf{x}|\mathsf{u}}$ [cf. (6)], the masked beamforming scheme corresponds to different (suboptimal) choice of these distributions. In particular, we choose

$$p_{\mathsf{u}} = \mathcal{CN}(0,\tilde{P}) \quad \text{and} \quad p_{\mathbf{x}|\mathsf{u}} = \mathcal{CN}(\mathsf{u}\tilde{\mathbf{h}}_{\mathrm{r}}, \tilde{P}(\mathbf{I}-\tilde{\mathbf{h}}_{\mathrm{r}}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}) \quad (103)$$

where we have chosen the convenient normalizations

$$\tilde{P} = \frac{P}{n_{\mathrm{t}}} \quad (104)$$

and

$$\tilde{\mathbf{h}}_{\mathrm{r}} = \frac{\mathbf{h}_{\mathrm{r}}}{\|\mathbf{h}_{\mathrm{r}}\|}. \quad (105)$$

In this form, the secrecy rate of masked beamforming is readily obtained, as we now show

### A. Proof of Proposition 1

With $p_{\mathsf{u}}$ and $p_{\mathbf{x}|\mathsf{u}}$ as in (103), we evaluate (6). To this end, first we have

$$I(\mathsf{u};\mathsf{y}_{\mathrm{r}}) = \log(1+\tilde{P}\|\mathbf{h}_{\mathrm{r}}\|^2). \quad (106)$$

Then, to evaluate $I(\mathsf{u};\mathbf{y}_{\mathrm{e}})$, note that

$$h(\mathbf{y}_{\mathrm{e}}) = \log\det(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}\mathbf{H}_{\mathrm{e}}^{\dagger})$$
$$h(\mathbf{y}_{\mathrm{e}}|\mathsf{u}) = \log\det(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}(\mathbf{I}-\tilde{\mathbf{h}}_{\mathrm{r}}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger})\mathbf{H}_{\mathrm{e}}^{\dagger})$$

so

$$\begin{aligned}
&I(\mathsf{u};\mathbf{y}_{\mathrm{e}})\\
&= h(\mathbf{y}_{\mathrm{e}}) - h(\mathbf{y}_{\mathrm{e}}|\boldsymbol{u})\\
&= \log\det(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}\mathbf{H}_{\mathrm{e}}^{\dagger}) - \log\det(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}(\mathbf{I}-\tilde{\mathbf{h}}_{\mathrm{r}}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger})\mathbf{H}_{\mathrm{e}}^{\dagger})\\
&= \log\det(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}) - \log\det(\mathbf{I}+\tilde{P}(\mathbf{I}-\tilde{\mathbf{h}}_{\mathrm{r}}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger})\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\\
&= \log\det(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\\
&\quad - \log\det(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}-\tilde{P}\tilde{\mathbf{h}}_{\mathrm{r}}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\\
&= -\log\det\left(\mathbf{I}-\tilde{P}\tilde{\mathbf{h}}_{\mathrm{r}}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\right)\\
&= -\log\left(1-\tilde{P}\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\tilde{\mathbf{h}}_{\mathrm{r}}\right)\\
&= -\log\left(\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\tilde{\mathbf{h}}_{\mathrm{r}}\right) \quad (107)
\end{aligned}$$

where we have repeatedly used the matrix identity $\det(\mathbf{I}+\mathbf{AB}) = \det(\mathbf{I}+\mathbf{BA})$ valid for any $\mathbf{A}$ and $\mathbf{B}$ with compatible dimensions.

Thus, combining (106) and (107), we obtain (15) as desired

$$\begin{aligned}
&R_{\mathrm{MB}}(P)\\
&= I(\mathsf{u};\mathsf{y}_{\mathrm{r}}) - I(\mathsf{u};\mathbf{y}_{\mathrm{e}})\\
&= \log(1+\tilde{P}\|\mathbf{h}_{\mathrm{r}}\|^2) + \log(\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\tilde{\mathbf{h}}_{\mathrm{r}})\\
&= \log\left(1+\frac{1}{\tilde{P}\|\mathbf{h}_{\mathrm{r}}\|^2}\right) + \log(\tilde{P}\mathbf{h}_{\mathrm{r}}^{\dagger}(\mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{h}_{\mathrm{r}})\\
&= \log\left(1+\frac{1}{\tilde{P}\|\mathbf{h}_{\mathrm{r}}\|^2}\right) + \log(\lambda_{\max}(\tilde{P}\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}))
\end{aligned}$$

where to obtain the last equality we have used the special form (3) for the largest generalized eigenvalue.

### B. Proof of Theorem 3

First, from Theorem 2 and Proposition 1, we have, with again $\tilde{P}$ as in (104) for convenience

$$C\left(\frac{P}{n_{\mathrm{t}}}\right) - R_{\mathrm{MB}}(P) \le \log\frac{\lambda_{\max}(\mathbf{I}+\tilde{P}\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})}{\lambda_{\max}(\tilde{P}\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})}. \quad (108)$$

Next, with $\boldsymbol{\psi}_{\max}$ denoting the generalized eigenvector corresponding to $\lambda_{\max}(\mathbf{I}+\tilde{P}\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})$, we have

$$\lambda_{\max}(\mathbf{I}+\tilde{P}\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}) = \frac{1+\tilde{P}|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}|^2}{1+\tilde{P}\|\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}_{\max}\|^2} \quad (109)$$

$$\lambda_{\max}(\tilde{P}\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger}, \mathbf{I}+\tilde{P}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}) \ge \frac{\tilde{P}|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}|^2}{1+\tilde{P}\|\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}_{\max}\|^2}. \quad (110)$$

$$(111)$$

Finally, substituting (109) and (110) into (108), we obtain

$$0 \le C\left(\frac{P}{n_{\mathrm{t}}}\right) - R_{\mathrm{MB}}(P) \le \log\left(1+\frac{n_{\mathrm{t}}}{P|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}|^2}\right) \quad (112)$$

the right hand side of which approaches zero as $P \to \infty$, whence (16) as desired.

## VIII. Scaling Laws Development

We begin by summarizing a few well-known results from random matrix theory that will be useful in our scaling laws; for further details, see, e.g., [24].

### A. Some Random Matrix Properties

Three basic facts will suffice for our purposes.

*Fact 4:* Suppose that $\mathsf{v}$ is a random length-$n$ complex vector with independent, zero-mean, variance-$1/n$ elements, and that $\mathsf{B}$ is a random $n \times n$ complex positive semidefinite matrix distributed independently of $\mathsf{v}$. Then if the spectrum of $\mathsf{B}$ converges, we have

$$\lim_{n\to\infty} \mathsf{v}^{\dagger}(\mathbf{I}+\gamma\mathsf{B})^{-1}\mathsf{v} \overset{\text{a.s.}}{=} \eta_{\mathsf{B}}(\gamma) \quad (113)$$

where $\eta_{\mathsf{B}}(\gamma)$ is the $\eta$-transform [24] of the matrix $\mathsf{B}$.

Of particular interest to us is the $\eta$-transform of a special class of matrices below.

*Fact 5:* Suppose that $\mathbf{H} \in \mathbb{C}^{K\times N}$ is random matrix whose entries are i.i.d. with variance $1/N$. As $K, N \to \infty$ with the ratio $K/N \triangleq \beta$ fixed, the $\eta$-transform of $\mathsf{B} = \mathbf{H}^\dagger\mathbf{H}$ is given by

$$\eta_{\mathbf{H}^\dagger\mathbf{H}}(\gamma) = \frac{\xi(\gamma,\beta)}{\gamma} \tag{114}$$

where $\xi(\cdot,\cdot)$ is as defined in (21).

The distribution of generalized eigenvalues of the pair $(\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}{}^\dagger, \mathbf{H}_\mathrm{e}{}^\dagger\mathbf{H}_\mathrm{e})$ is also known [25], [26]. For our purposes, the following is sufficient.

*Fact 6:* Suppose that $\mathbf{h}_\mathrm{r}$ and $\mathbf{H}_\mathrm{e}$ have i.i.d. $\mathcal{CN}(0,1)$ entries, and $n_\mathrm{e} > n_\mathrm{t}$. Then

$$\lambda_{\max}(\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}{}^\dagger, \mathbf{H}_\mathrm{e}{}^\dagger\mathbf{H}_\mathrm{e}) \sim \frac{2n_\mathrm{t}}{2n_\mathrm{e}-2n_\mathrm{t}+1} F_{2n_\mathrm{t},2n_\mathrm{e}-2n_\mathrm{t}+1} \tag{115}$$

where $F_{2n_\mathrm{t},2n_\mathrm{e}-2n_\mathrm{t}+1}$ is the F-distribution with $2n_\mathrm{t}$ and $2n_\mathrm{e}-2n_\mathrm{t}+1$ degrees of freedom, i.e.,

$$F_{2n_\mathrm{t},2n_\mathrm{e}-2n_\mathrm{t}+1} \stackrel{\mathrm{d}}{=} \frac{\mathsf{v}_1/(2n_\mathrm{t})}{\mathsf{v}_2/(2n_\mathrm{e}-2n_\mathrm{t}+1)} \tag{116}$$

where $\stackrel{\mathrm{d}}{=}$ denote equality in distribution, and where $\mathsf{v}_1$ and $\mathsf{v}_2$ are independent chi-squared random variables with $2n_\mathrm{t}$ and $2n_\mathrm{e}-2n_\mathrm{t}+1$ degrees of freedom, respectively.

Using Fact 6, it follows that with $\beta = n_\mathrm{e}/n_\mathrm{t}$ fixed

$$\lim_{n_\mathrm{t}\to\infty} \lambda_{\max}(\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}{}^\dagger, \mathbf{H}_\mathrm{e}{}^\dagger\mathbf{H}_\mathrm{e}) \stackrel{\mathrm{a.s.}}{=} \frac{1}{\beta-1}, \quad \text{when } \beta > 1. \tag{117}$$

Indeed, from the strong law of large numbers we have that the random variables $\mathsf{v}_1$ and $\mathsf{v}_2$ in (116) satisfy, for $\beta > 1$

$$\lim_{n_\mathrm{t}\to\infty} \frac{\mathsf{v}_1}{2n_\mathrm{t}} \stackrel{\mathrm{a.s.}}{=} 1, \quad \text{and} \quad \lim_{n_\mathrm{t}\to\infty} \frac{\mathsf{v}_2}{2n_\mathrm{t}(\beta-1)+1} \stackrel{\mathrm{a.s.}}{=} 1. \tag{118}$$

Combining (118) with (116) yields (117).

### B. Proof of Corollary 3

First, from Theorem 2, we have that

$$\begin{aligned} C(P,n_\mathrm{t},n_\mathrm{e}) &= \left\{\log\lambda_{\max}(\mathbf{I}+P\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}{}^\dagger, \mathbf{I}+P\mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e})\right\}^+ \\ &\geq \left\{\log\lambda_{\max}(P\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}{}^\dagger, \mathbf{I}+P\mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e})\right\}^+ \\ &= \left\{\log\left(P\mathbf{h}_\mathrm{r}{}^\dagger(\mathbf{I}+P\mathbf{H}_\mathrm{e}^\dagger\mathbf{H}_\mathrm{e})^{-1}\mathbf{h}_\mathrm{r}\right)\right\}^+ \end{aligned} \tag{119}$$

where (119) follows from the quadratic form representation (3) of the generalized eigenvalue.

Rewriting (119) using the notation

$$\tilde{\mathsf{h}}_\mathrm{r} = \frac{1}{\sqrt{n_\mathrm{t}}}\mathbf{h}_\mathrm{r}, \quad \text{and} \quad \tilde{\mathbf{H}}_\mathrm{e} = \frac{1}{\sqrt{n_\mathrm{t}}}\mathbf{H}_\mathrm{e} \tag{120}$$

we then obtain (20) as desired

$$\begin{aligned} \tilde{C}(\gamma,\beta) &= C(\gamma/n_\mathrm{t}, n_\mathrm{t}, \beta n_\mathrm{t}) \\ &\geq \left\{\log\left(\gamma\tilde{\mathsf{h}}_\mathrm{r}(\mathbf{I}+\gamma\tilde{\mathbf{H}}_\mathrm{e}^\dagger\tilde{\mathbf{H}}_\mathrm{e})^{-1}\tilde{\mathsf{h}}_\mathrm{r}\right)\right\}^+ \\ &\stackrel{\mathrm{a.s.}}{\longrightarrow} \{\log\xi(\gamma,\beta)\}^+ \quad \text{as } n_\mathrm{t}\to\infty \end{aligned} \tag{121}$$

where to obtain (121), we have applied (113) and (114).

The derivation of the scaling law (22) for the masked beamforming scheme is analogous. Indeed, from Proposition 1, we have

$$\begin{aligned} R_{\mathrm{MB}}(\gamma,n_\mathrm{t},\beta n_\mathrm{t}) &\geq \left\{\log\lambda_{\max}(\gamma\tilde{\mathsf{h}}_\mathrm{r}\tilde{\mathsf{h}}_\mathrm{r}^\dagger, \mathbf{I}+\gamma\tilde{\mathbf{H}}_\mathrm{e}^\dagger\tilde{\mathbf{H}}_\mathrm{e})\right\}^+ \\ &= \left\{\log\left(\gamma\tilde{\mathsf{h}}_\mathrm{r}^\dagger(\mathbf{I}+\gamma\tilde{\mathbf{H}}_\mathrm{e}^\dagger\tilde{\mathbf{H}}_\mathrm{e})^{-1}\tilde{\mathsf{h}}_\mathrm{r}\right)\right\}^+ \\ &\stackrel{\mathrm{a.s.}}{\longrightarrow} \{\log\xi(\gamma,\beta)\}^+ \quad \text{as } n_\mathrm{t}\to\infty \end{aligned}$$

where as above the last line comes from applying (113) and (114).

### C. Proof of Corollary 4

When $\beta < 1$ (i.e., $n_\mathrm{e} < n_\mathrm{t}$), we have $\mathbf{H}_\mathrm{e}^\sharp\mathbf{h}_\mathrm{r} \neq \mathbf{0}$ almost surely, so (13b) holds, i.e.,

$$\lim_{P\to\infty} C(P) = \infty \tag{122}$$

as (23) reflects.

When $\beta \geq 1$ (i.e., $n_\mathrm{e} > n_\mathrm{t}$) $\mathbf{H}_\mathrm{e}{}^\dagger\mathbf{H}_\mathrm{e}$ is nonsingular almost surely, (13a) holds, i.e.,

$$\lim_{P\to\infty} C(P) = \left\{\log\lambda(\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}{}^\dagger, \mathbf{H}_\mathrm{e}{}^\dagger\mathbf{H}_\mathrm{e})\right\}^+.$$

Taking the limit $n_\mathrm{e}, n_\mathrm{t} \to \infty$ with $n_\mathrm{e}/n_\mathrm{t} = \beta$ fixed, and using (117), we obtain

$$\lim_{n_\mathrm{t}\to\infty}\lim_{P\to\infty} C(P) = \{-\log(\beta-1)\}^+$$

as (23) asserts.

Furthermore, via (16), we have that

$$\lim_{P\to\infty} R_{\mathrm{MB}}(P) = \left\{\log\lambda(\mathbf{h}_\mathrm{r}\mathbf{h}_\mathrm{r}{}^\dagger, \mathbf{H}_\mathrm{e}{}^\dagger\mathbf{H}_\mathrm{e})\right\}^+ = \lim_{P\to\infty} C(P)$$

whence (24).

## IX. Fading Channel Analysis

We prove the lower and upper bounds of Theorem 4 separately.

### A. Proof of (27a)

By viewing the fading channel as a set of parallel channels indexed by the channel gain $\mathbf{h}_\mathrm{r}$ of the intended receiver[7] and the eavesdropper's observation as $(\mathbf{y}_\mathrm{e}, \mathbf{H}_\mathrm{e})$, the rate

$$R = I(\mathsf{u};\mathsf{y}_\mathrm{r} \mid \mathbf{h}_\mathrm{r}) - I(\mathsf{u};\mathbf{y}_\mathrm{e},\mathbf{H}_\mathrm{e} \mid \mathbf{h}_\mathrm{r}). \tag{123}$$

is achievable for any choice of $p_{\mathsf{u}|\mathbf{h}_\mathrm{r}}$ and $p_{\mathbf{x}|\mathsf{u},\mathbf{h}_\mathrm{r}}$ that satisfies the power constraint $E[\rho(\mathbf{h}_\mathrm{r})] \leq P$. We choose distributions corresponding to an adaptive version of masked beamforming, i.e., [cf. (103)]

$$p_{\mathsf{u}|\mathbf{h}_\mathrm{r}} = \mathcal{CN}(0,\tilde{\rho}(\mathbf{h}_\mathrm{r})), \quad p_{\mathbf{x}|\boldsymbol{u},\mathbf{h}_\mathrm{r}} = \mathcal{CN}\left(\mathsf{u}\tilde{\mathsf{h}}_\mathrm{r}, \tilde{\rho}(\mathbf{h}_\mathrm{r})(\mathbf{I}-\tilde{\mathsf{h}}_\mathrm{r}\tilde{\mathsf{h}}_\mathrm{r}^\dagger)\right) \tag{124}$$

[7]Since the fading coefficients are continuous valued, one has to discretize these coefficients before mapping to parallel channels. By choosing appropriately fine quantization levels one can approach the rate as closely as possible. See, e.g., [9] for a discussion.

where we have chosen the convenient normalizations [cf. (104) and (105)]

$$\tilde{\rho}(\mathbf{h}_{\mathrm{r}}) \triangleq \frac{\rho(\mathbf{h}_{\mathrm{r}})}{n_{\mathrm{t}}} \tag{125}$$

and

$$\tilde{\mathsf{h}}_{\mathrm{r}} = \frac{\mathbf{h}_{\mathrm{r}}}{||\mathbf{h}_{\mathrm{r}}||}. \tag{126}$$

Evaluating (123) with the distributions (124) yields (27a) with (29a)

$$\begin{aligned} I(\mathsf{u};\mathsf{y}_{\mathrm{r}} \mid \mathbf{h}_{\mathrm{r}}) - I(\mathsf{u};\mathbf{y}_{\mathrm{e}},\mathbf{H}_{\mathrm{e}} \mid \mathbf{h}_{\mathrm{r}}) & \quad (127)\\ = E[\log(1+\tilde{\rho}(\mathbf{h}_{\mathrm{r}})\|\mathbf{h}_{\mathrm{r}}\|^2)] & \\ + E[\log(\tilde{\mathbf{h}}_{\mathrm{r}}^{\dagger}(\mathbf{I}+\tilde{\rho}(\mathbf{h}_{\mathrm{r}})\mathbf{H}_{\mathrm{e}}{}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\tilde{\mathbf{h}}_{\mathrm{r}})] & \quad (128)\\ = E\left[\log\left(1+\frac{1}{\tilde{\rho}(\mathbf{h}_{\mathrm{r}})\|\mathbf{h}_{\mathrm{r}}\|^2}\right)\right] & \\ + E\left[\log\left(\tilde{\rho}(\mathbf{h}_{\mathrm{r}})\mathbf{h}_{\mathrm{r}}{}^{\dagger}(\mathbf{I}+\tilde{\rho}(\mathbf{h}_{\mathrm{r}})\mathbf{H}_{\mathrm{e}}{}^{\dagger}\mathbf{H}_{\mathrm{e}})^{-1}\mathbf{h}_{\mathrm{r}}\right)\right] & \quad (129) \end{aligned}$$

where the steps leading to (128) are analogous to those used in Section VII-A for the nonfading case and hence have been omitted.

### B. Proof of (27b)

Suppose that there is a sequence of $(2^{nR},n)$ codes such that for a sequence $\varepsilon_n$ (with $\varepsilon_n \to 0$ as $n \to \infty$)

$$\begin{aligned} \frac{1}{n}H(\mathsf{w}) - \frac{1}{n}H(\mathsf{w}|\mathbf{y}_{\mathrm{e}}^n,\mathbf{H}_{\mathrm{e}}{}^n,\mathbf{h}_{\mathrm{r}}{}^n) \le \varepsilon_n & \\ \Pr(\hat{\mathsf{w}} \neq \mathsf{w}) \le \varepsilon_n. & \quad (130) \end{aligned}$$

*1) An Auxiliary Channel:* We now introduce another channel for which the noise vaiables $\mathsf{z}_{\mathrm{r}}(t)$ and $\mathbf{z}_{\mathrm{e}}(t)$ are correlated, but the conditions in (130) still hold. Hence, any rate achievable on the original channel is also achievable on this new channel. In what follows, we will upper bound the rate achievable for this new channel instead of the original channel.

We begin by introducing some notation. Let

$$\rho_t(\mathbf{h}_{\mathrm{r}}^t) \triangleq E\left[||\mathbf{x}(t)||^2 \mid \mathbf{h}_{\mathrm{r}}{}^t = \mathbf{h}_{\mathrm{r}}^t\right] \tag{131}$$

denote the transmitted power at time $t$, when the channel realization of the intended receiver from time 1 to $t$ is $\mathbf{h}_{\mathrm{r}}^t$. Note that $\rho_t(\cdot)$ satisfies the long term average power constraint, i.e.,

$$E_{\mathbf{h}_{\mathrm{r}}{}^n}\left[\frac{1}{n}\sum_{t=1}^{n}\rho_t(\mathbf{h}_{\mathrm{r}}{}^t)\right] \le P. \tag{132}$$

Next, let, $p_{\mathbf{h}_{\mathrm{r}}}$ and $p_{\mathbf{H}_{\mathrm{e}}}$ denote the density functions of $\mathbf{h}_{\mathrm{r}}$ and $\mathbf{H}_{\mathrm{e}}$, respectively, and let $p_{\mathsf{z}_{\mathrm{r}}}$ and $p_{\mathbf{z}_{\mathrm{e}}}$ denote the density function of the noise random variables in our channel model (26). Observe that the constraints in (130) (and hence the capacity) depend only on the distributions $p_{\mathbf{z}_{\mathrm{e}}{}^n,\mathbf{h}_{\mathrm{r}}{}^n,\mathbf{H}_{\mathrm{e}}{}^n}(\mathbf{z}_e^n,\mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n)$ and $p_{\mathsf{z}_{\mathrm{r}}^n,\mathbf{h}_{\mathrm{r}}{}^n}(z_r^n,\mathbf{h}_{\mathrm{r}}^n)$. Furthermore since the channel model (26) is memoryless and $(\mathbf{h}_{\mathrm{r}},\mathbf{H}_{\mathrm{e}})$ are i.i.d. and mutually independent, we have

$$\begin{aligned} p_{\mathbf{z}_{\mathrm{e}}{}^n,\mathbf{h}_{\mathrm{r}}{}^n,\mathbf{H}_{\mathrm{e}}{}^n}(\mathbf{z}_e^n,\mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n) = & \\ \prod_{t=1}^{n} p_{\mathbf{z}_{\mathrm{e}}}(\mathbf{z}_{\mathrm{e}}(t))p_{\mathbf{h}_{\mathrm{r}}}(\mathbf{h}_{\mathrm{r}}(t))p_{\mathbf{H}_{\mathrm{e}}}(\mathbf{H}_{\mathrm{e}}(t)) & \quad (133) \end{aligned}$$

$$p_{\mathsf{z}_{\mathrm{r}}^n,\mathbf{h}_{\mathrm{r}}{}^n}(z_r^n,\mathbf{h}_{\mathrm{r}}^n) = \prod_{t=1}^{n} p_{\mathsf{z}_{\mathrm{r}}}(z_r(t))p_{\mathbf{h}_{\mathrm{r}}}(\mathbf{h}_{\mathrm{r}}(t)). \tag{134}$$

Let $\mathcal{P}_t$ denote the set of conditional-joint distributions $p_{\mathsf{z}_{\mathrm{r}}(t),\mathbf{z}_{\mathrm{e}}(t)|\mathbf{h}_{\mathrm{r}}{}^n,\mathbf{H}_{\mathrm{e}}{}^n}$ with fixed conditional-marginals, i.e.,

$$\begin{aligned} \mathcal{P}_t = \big\{ & p_{\mathsf{z}_{\mathrm{r}}(t),\mathbf{z}_{\mathrm{e}}(t)|\mathbf{h}_{\mathrm{r}}{}^n,\mathbf{H}_{\mathrm{e}}{}^n}(z_r,\mathbf{z}_e \mid \mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n) \mid \\ & p_{\mathsf{z}_{\mathrm{r}}(t)|\mathbf{h}_{\mathrm{r}}{}^n,\mathbf{H}_{\mathrm{e}}{}^n}(z_r \mid \mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n) = p_{\mathsf{z}_{\mathrm{r}}}(z_r) \\ & p_{\mathbf{z}_{\mathrm{e}}(t)|\mathbf{h}_{\mathrm{r}}{}^n,\mathbf{H}_{\mathrm{e}}{}^n}(\mathbf{z}_e \mid \mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n) = p_{\mathbf{z}_{\mathrm{e}}}(\mathbf{z}_e)\big\}. \end{aligned} \tag{135}$$

Suppose that for each $t = 1,2,\ldots,n$ we select a distribution $p_{\mathsf{z}_{\mathrm{r}}(t),\mathbf{z}_{\mathrm{e}}(t)|\mathbf{h}_{\mathrm{r}}{}^n,\mathbf{H}_{\mathrm{e}}{}^n} \in \mathcal{P}_t$ and consider a channel with distribution

$$\begin{aligned} p_{\mathsf{z}_{\mathrm{r}}^n,\mathbf{z}_{\mathrm{e}}{}^n,\mathbf{h}_{\mathrm{r}}{}^n,\mathbf{H}_{\mathrm{e}}{}^n}(z_r^n,\mathbf{z}_e^n,\mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n) = & \\ \prod_{t=1}^{n} p_{\mathsf{z}_{\mathrm{r}}(t),\mathbf{z}_{\mathrm{e}}(t)|\mathbf{h}_{\mathrm{r}}{}^n,\mathbf{H}_{\mathrm{e}}{}^n}(z_r(t),\mathbf{z}_e(t)|\mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n) & \\ \times p_{\mathbf{h}_{\mathrm{r}}}(\mathbf{h}_{\mathrm{r}}(t))p_{\mathbf{H}_{\mathrm{e}}}(\mathbf{H}_{\mathrm{e}}(t)). & \quad (136) \end{aligned}$$

This new channel distribution has noise variables $(\mathsf{z}_{\mathrm{r}}(t),\mathbf{z}_{\mathrm{e}}(t))$ correlated, where the correlation is possibly time-dependent, but from (135) and (136), note that $\mathsf{z}_{\mathrm{r}}^n$ and $\mathbf{z}_{\mathrm{e}}{}^n$ are marginally Gaussian and i.i.d., and satisfy (133) and (134). Hence, the conditions in (130) are satisfied for this channel and the rate $R$ is achievable.

In the sequel, we select $p_{\mathsf{z}_{\mathrm{r}}(t),\mathbf{z}_{\mathrm{e}}(t)|\mathbf{h}_{\mathrm{r}}{}^n,\mathbf{H}_{\mathrm{e}}{}^n}(z_r,\mathbf{z}_e \mid \mathbf{h}_{\mathrm{r}}^n,\mathbf{H}_{\mathrm{e}}^n)$ to be the worst case noise distribution for the Gaussian channel with gains $\mathbf{h}_{\mathrm{r}}(t)$, and, $\mathbf{H}_{\mathrm{e}}(t)$, and power of $\rho_t(\mathbf{h}_{\mathrm{r}}^t)$ in Theorem 2, i.e., if $\boldsymbol{\psi}_t$ is the eigenvector corresponding to the largest generalized eigenvalue $\lambda_{\max}(\mathbf{I}+\rho_t(\mathbf{h}_{\mathrm{r}}^t)\mathbf{h}_{\mathrm{r}}(t)\mathbf{h}_{\mathrm{r}}(t)^{\dagger},\mathbf{I}+\rho_t(\mathbf{h}_{\mathrm{r}}^t)\mathbf{H}_{\mathrm{e}}^{\dagger}(t)\mathbf{H}_{\mathrm{e}}(t))$,

$$\begin{aligned} p_{\mathsf{z}_{\mathrm{r}}(t),\mathbf{z}_{\mathrm{e}}(t)|\mathbf{h}_{\mathrm{r}}{}^n,\mathbf{H}_{\mathrm{e}}{}^n} = & \mathcal{CN}\left(0,\begin{bmatrix} 1 & \boldsymbol{\phi}_t^{\dagger} \\ \boldsymbol{\phi}_t & \mathbf{I} \end{bmatrix}\right), \text{ where} \\ \boldsymbol{\phi}_t = & \begin{cases} \frac{1}{\mathbf{h}_{\mathrm{r}}^{\dagger}(t)\boldsymbol{\psi}_t}(\mathbf{H}_{\mathrm{e}}(t)\boldsymbol{\psi}_t), & \lambda_{\max} \ge 1 \\ \mathbf{G}_{\mathrm{e}}(t)(\mathbf{G}_{\mathrm{e}}^{\dagger}(t)\mathbf{G}_{\mathrm{e}}(t))^{-1}\mathbf{g}_{\mathrm{r}}(t), & \lambda_{\max} < 1 \end{cases} \end{aligned} \tag{137}$$

and where $\mathbf{G}_{\mathrm{e}}(t)$ and $\mathbf{g}_{\mathrm{r}}(t)$ are related to $\mathbf{H}_{\mathrm{e}}(t)$ and $\mathbf{h}_{\mathrm{r}}(t)$ as in (54). Our choice of $p_{\mathsf{z}_{\mathrm{r}}(t),\mathbf{z}_{\mathrm{e}}(t)|\mathbf{h}_{\mathrm{r}}{}^n,\mathbf{H}_{\mathrm{e}}{}^n}$ is such that $(\mathsf{z}_{\mathrm{r}}(t),\mathbf{z}_{\mathrm{e}}(t))$ only depend on the $(\mathbf{H}_{\mathrm{e}}(t),\mathbf{h}_{\mathrm{r}}(t),\rho_t(\mathbf{h}_{\mathrm{r}}{}^t))$, i.e.,

$$(\mathbf{H}_{\mathrm{e}}{}^n,\mathbf{h}_{\mathrm{r}}{}^n) \to (\rho(\mathbf{h}_{\mathrm{r}}{}^t),\mathbf{h}_{\mathrm{r}}(t),\mathbf{H}_{\mathrm{e}}(t)) \to (\mathsf{z}_{\mathrm{r}}(t),\mathbf{z}_{\mathrm{e}}(t)) \tag{138}$$

forms a Markov chain.

*2) Upper Bound on the Auxiliary Channel:* We now upper bound the secrecy rate for the channel (136). Note that this also upper bounds the rate on the original channel.

From Fano's inequality, that there exists a sequence $\varepsilon'_n$ such that $\varepsilon'_n \to 0$ as $n \to \infty$, and

$$\frac{1}{n}H(\mathsf{w}|\mathsf{y}_{\mathrm{r}}^n, \mathbf{h}_{\mathrm{r}}{}^n) \le \varepsilon'_n$$

$$\begin{aligned} nR &= H(\mathsf{w}) = I(\mathsf{w}; \mathsf{y}_{\mathrm{r}}^n \mid \mathbf{h}_{\mathrm{r}}{}^n) + n\varepsilon'_n \\ &= I(\mathsf{w}; \mathsf{y}_{\mathrm{r}}^n \mid \mathbf{h}_{\mathrm{r}}{}^n) - I(\mathsf{w}; \mathbf{y}_{\mathrm{e}}^n, \mathbf{H}_{\mathrm{e}}{}^n \mid \mathbf{h}_{\mathrm{r}}{}^n) + n(\varepsilon_n + \varepsilon'_n) && (139) \\ &\le I(\mathsf{w}; \mathsf{y}_{\mathrm{r}}^n \mid \mathbf{h}_{\mathrm{r}}{}^n, \mathbf{H}_{\mathrm{e}}{}^n, \mathbf{y}_{\mathrm{e}}^n) + n(\varepsilon_n + \varepsilon'_n) \\ &\le I(\mathbf{x}^n; \mathsf{y}_{\mathrm{r}}^n \mid \mathbf{h}_{\mathrm{r}}{}^n, \mathbf{H}_{\mathrm{e}}{}^n, \mathbf{y}_{\mathrm{e}}^n) + n(\varepsilon_n + \varepsilon'_n) && (140) \\ &\le \sum_{t=1}^{n} I(\mathbf{x}(t); \mathsf{y}_{\mathrm{r}}(t) \mid \mathbf{H}_{\mathrm{e}}{}^n, \mathbf{h}_{\mathrm{r}}{}^n, \mathbf{y}_{\mathrm{e}}(t)) + n(\varepsilon_n + \varepsilon'_n) \\ & && (141) \end{aligned}$$

where (139) follows from the secrecy condition [cf. (130)], and (140) follows from the Markov relation $\mathsf{w} \leftrightarrow (\mathbf{x}^n, \mathbf{y}_{\mathrm{e}}^n, \mathbf{h}_{\mathrm{r}}{}^n, \mathbf{H}_{\mathrm{e}}{}^n) \leftrightarrow \mathsf{y}_{\mathrm{r}}^n$, and (141) holds because for the channel (136), we have

$$h(\mathsf{y}_{\mathrm{r}}^n|\mathbf{y}_{\mathrm{e}}^n, \mathbf{H}_{\mathrm{e}}{}^n, \mathbf{h}_{\mathrm{r}}{}^n, \mathbf{x}^n) = \sum_{t=1}^{n} h(\mathsf{y}_{\mathrm{r}}(t)|\mathbf{y}_{\mathrm{e}}(t), \mathbf{h}_{\mathrm{r}}{}^n, \mathbf{H}_{\mathrm{e}}{}^n, \mathbf{x}(t)).$$

We next upper bound the term $I(\mathbf{x}(t); \mathsf{y}_{\mathrm{r}}(t) \mid \mathsf{y}_{\mathrm{e}}(t), \mathbf{H}_{\mathrm{e}}{}^n, \mathbf{h}_{\mathrm{r}}{}^n)$ in (141) for each $t = 1, 2, \ldots, n$

$$\begin{aligned} & I(\mathbf{x}(t); \mathsf{y}_{\mathrm{r}}(t) \mid \mathsf{y}_{\mathrm{e}}(t), \mathbf{H}_{\mathrm{e}}{}^n, \mathbf{h}_{\mathrm{r}}{}^n) \\ & \le I(\mathbf{x}(t); \mathsf{y}_{\mathrm{r}}(t) \mid \mathsf{y}_{\mathrm{e}}(t), \mathbf{H}_{\mathrm{e}}(t), \mathbf{h}_{\mathrm{r}}(t), \rho_t(\mathbf{h}_{\mathrm{r}}{}^t)) && (142) \\ & \le E[\{\log \lambda_{\max}(\mathbf{I} + \rho_t(\mathbf{h}_{\mathrm{r}}{}^t)\mathbf{h}_{\mathrm{r}}(t)\mathbf{h}_{\mathrm{r}}{}^\dagger(t) \\ & \quad \mathbf{I} + \rho_t(\mathbf{h}_{\mathrm{r}}{}^t)\mathbf{H}_{\mathrm{e}}{}^\dagger(t)\mathbf{H}_{\mathrm{e}}(t))\}^+] && (143) \end{aligned}$$

where (142) follows from the fact that [cf. (138)]

$$(\mathbf{H}_{\mathrm{e}}{}^n, \mathbf{h}_{\mathrm{r}}{}^n) \to (\mathbf{x}(t), \rho_t(\mathbf{h}_{\mathrm{r}}{}^t), \mathbf{h}_{\mathrm{r}}(t), \mathbf{H}_{\mathrm{e}}(t)) \to (\mathsf{y}_{\mathrm{r}}(t), \mathbf{y}_{\mathrm{e}}(t))$$

forms a Markov chain and (143) follows since our choice of the noise distribution in (137) is the worst case noise in (7) for the Gaussian channel with gains $\mathbf{h}_{\mathrm{r}}(t)$, $\mathbf{H}_{\mathrm{e}}(t)$ and power $\rho_t(\mathbf{h}_{\mathrm{r}}^t)$; hence, the derivation in Theorem 2 applies.

Substituting (143) into (141), we have

$$\begin{aligned} & nR - n(\varepsilon_n + \varepsilon'_n) \\ & = \sum_{t=1}^{n} E_{\mathbf{H}_{\mathrm{e}}(t), \mathbf{h}_{\mathrm{r}}{}^t}\big[\{\log \lambda_{\max}(\mathbf{I} + \rho_t(\mathbf{h}_{\mathrm{r}}{}^t)\mathbf{h}_{\mathrm{r}}(t)\mathbf{h}_{\mathrm{r}}{}^\dagger(t) \\ & \quad \mathbf{I} + \rho_t(\mathbf{h}_{\mathrm{r}}{}^t)\mathbf{H}_{\mathrm{e}}{}^\dagger(t)\mathbf{H}_{\mathrm{e}}(t))\}^+\big] && (144) \\ & \le \sum_{t=1}^{n} E_{\mathbf{H}_{\mathrm{e}}(t), \mathbf{h}_{\mathrm{r}}(t)}\big[\{\log \lambda_{\max}(\mathbf{I} + E_{\mathbf{h}_{\mathrm{r}}{}^{t-1}}[\rho_t(\mathbf{h}_{\mathrm{r}}{}^t)]\mathbf{h}_{\mathrm{r}}(t)\mathbf{h}_{\mathrm{r}}{}^\dagger(t) \\ & \quad \mathbf{I} + E_{\mathbf{h}_{\mathrm{r}}{}^{t-1}}[\rho_t(\mathbf{h}_{\mathrm{r}}{}^t)]\mathbf{H}_{\mathrm{e}}{}^\dagger(t)\mathbf{H}_{\mathrm{e}}(t))\}^+\big] && (145) \\ & = \sum_{t=1}^{n} E_{\mathbf{H}_{\mathrm{e}}(t), \mathbf{h}_{\mathrm{r}}(t)}\big[\{\log \lambda_{\max}(\mathbf{I} + \hat{\rho}_t(\mathbf{h}_{\mathrm{r}}(t))\mathbf{h}_{\mathrm{r}}(t)\mathbf{h}_{\mathrm{r}}{}^\dagger(t) \\ & \quad \mathbf{I} + \hat{\rho}_t(\mathbf{h}_{\mathrm{r}}(t))\mathbf{H}_{\mathrm{e}}{}^\dagger(t)\mathbf{H}_{\mathrm{e}}(t))\}^+\big] && (146) \\ & = \sum_{t=1}^{n} E_{\mathbf{H}_{\mathrm{e}}, \mathbf{h}_{\mathrm{r}}}[\{\log \lambda_{\max}(\mathbf{I} + \hat{\rho}_t(\mathbf{h}_{\mathrm{r}})\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}{}^\dagger, \mathbf{I} + \hat{\rho}_t(\mathbf{h}_{\mathrm{r}})\mathbf{H}_{\mathrm{e}}{}^\dagger\mathbf{H}_{\mathrm{e}})\}^+] \\ & && (147) \\ & \le nE_{\mathbf{H}_{\mathrm{e}}, \mathbf{h}_{\mathrm{r}}}\big[\{\log \lambda_{\max}(\mathbf{I} + \sum_{t=1}^{n} \frac{1}{n}\hat{\rho}_t(\mathbf{h}_{\mathrm{r}})\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}{}^\dagger \\ & \quad \mathbf{I} + \sum_{t=1}^{n} \frac{1}{n}\hat{\rho}_t(\mathbf{h}_{\mathrm{r}})\mathbf{H}_{\mathrm{e}}{}^\dagger\mathbf{H}_{\mathrm{e}})\}^+\big] && (148) \\ & = nE_{\mathbf{H}_{\mathrm{e}}, \mathbf{h}_{\mathrm{r}}}[\{\log \lambda_{\max}(\mathbf{I} + \rho(\mathbf{h}_{\mathrm{r}})\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}{}^\dagger, \mathbf{I} + \rho(\mathbf{h}_{\mathrm{r}})\mathbf{H}_{\mathrm{e}}{}^\dagger\mathbf{H}_{\mathrm{e}})\}^+] \\ & && (149) \end{aligned}$$

where (145) and (148) follow from Jensen's inequality since $C(P) = \{\log \lambda_{\max}(\mathbf{I} + P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}{}^\dagger, \mathbf{I} + P\mathbf{H}_{\mathrm{e}}{}^\dagger\mathbf{H}_{\mathrm{e}})\}^+$ is a capacity and, therefore, concave in $P$, (146) follows by defining

$$\hat{\rho}_t(\mathbf{h}_{\mathrm{r}}) = E_{\mathbf{h}_{\mathrm{r}}{}^{t-1}}[\rho_t(\mathbf{h}_{\mathrm{r}}{}^t)]. \tag{150}$$

Equation (147) follows from the fact that the distribution of both $\mathbf{h}_{\mathrm{r}}$ and $\mathbf{H}_{\mathrm{e}}$ does not depend on $t$, and (149) follows by defining $\rho(\mathbf{h}_{\mathrm{r}}) = \frac{1}{n}\sum_{t=1}^{n} \hat{\rho}_t(\mathbf{h}_{\mathrm{r}})$.

To complete the proof, note that

$$\begin{aligned} E_{\mathbf{h}_{\mathrm{r}}}[\rho(\mathbf{h}_{\mathrm{r}})] &= \frac{1}{n}\sum_{t=1}^{n} E_{\mathbf{h}_{\mathrm{r}}}[\hat{\rho}_t(\mathbf{h}_{\mathrm{r}})] \\ &= \frac{1}{n}\sum_{t=1}^{n} E_{\mathbf{h}_{\mathrm{r}}{}^t}[\rho_t(\mathbf{h}_{\mathrm{r}}{}^t)] && (151) \\ &= \frac{1}{n}\sum_{t=1}^{n} E_{\mathbf{h}_{\mathrm{r}}{}^n}[\rho_t(\mathbf{h}_{\mathrm{r}}{}^t)] \le P && (152) \end{aligned}$$

where (151) follows from (150) and the fact that the channel gains are i.i.d., and (152) follows from (132).

### C. Proof of Proposition 2

The proof is immediate from Corollary 3, 4, and Theorem 4. For the lower bound, we only consider the case when $\log \xi(P, \beta) > 0$, since otherwise the rate is zero. We select $\rho(\mathbf{h}_{\mathrm{r}}) = P$ to be fixed for each $\mathbf{h}_{\mathrm{r}}$. Then we have from Corollary 3 that

$$R_{\mathrm{FF},-}(\mathbf{h}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}, P) \xrightarrow{\text{a.s.}} \log \xi(P, \beta).$$

Finally since almost-sure convergence implies convergence in expectation

$$\lim_{n_{\mathrm{t}} \to \infty} E[R_{\mathrm{FF},-}(\mathbf{h}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}, P)] = \log \xi(P, \beta)$$

which establishes the lower bound (30). For the upper bound, since

$$R_{\mathrm{FF},+}(\mathbf{h}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}, P) = \big\{\log \lambda_{\max}(\mathbf{I} + P\mathbf{h}_r\mathbf{h}_r^\dagger, \mathbf{I} + P\mathbf{H}_{\mathrm{e}}^\dagger\mathbf{H}_{\mathrm{e}})\big\}^+$$

we have from Theorem 4 that

$$\lim_{n_{\mathrm{t}} \to \infty} R_{\mathrm{FF},+}(\mathbf{h}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}, P) \overset{\text{a.s.}}{\le} \tilde{C}(\infty, \beta) \tag{153}$$

and hence

$$\begin{aligned} \lim_{n_{\mathrm{t}} \to \infty} C_{FF}(P = \gamma, n_{\mathrm{t}}, n_{\mathrm{e}} = \beta n_{\mathrm{t}}) &\le \lim_{n_{\mathrm{t}} \to \infty} E[R_{FF,+}(\mathbf{h}_{\mathrm{r}}, \mathbf{H}_{\mathrm{e}}, \gamma)] \\ &\le \tilde{C}(\infty, \beta) \end{aligned}$$

where we again use the fact that almost sure convergence implies convergence in expectation.

## X. Concluding Remarks

The present work characterizes the key performance characteristics and tradeoffs inherent in communication over the MISOME channel. Since the submission of the present paper, the capacity for the more general MIMOME case has also been obtained [15], [27]–[30]. However, extensions to time-varying fading channels for that case remain to be developed.

More generally, many recent architectures for wireless systems exploit the knowledge of the channel at the physical layer in order to increase the system throughput and reliability. Many of these systems have a side benefit of providing security. It is naturally of interest to quantify these gains and identify potential applications.

## Appendix I
## Proof of Lemma 1

Suppose there exists a sequence of $(2^{nR}, n)$ codes such that for every $\varepsilon > 0$, and $n$ sufficiently large, we have that

$$\Pr(\mathsf{w} \neq \hat{\mathsf{w}}) \leq \varepsilon \tag{154}$$

$$\frac{1}{n} I(\mathsf{w}; \mathbf{y}_{\mathrm{e}}^n) \leq \varepsilon \tag{155}$$

$$\frac{1}{n} \sum_{i=1}^{n} E[||\mathbf{x}(i)||^2] \leq P. \tag{156}$$

We first note that (154) implies, from Fano's inequality

$$\frac{1}{n} I(\mathsf{w}; \mathsf{y}_{\mathrm{r}}^n) \geq R - \varepsilon_{\mathrm{F}} \tag{157}$$

where $\varepsilon_{\mathrm{F}} \to 0$ as $\varepsilon \to 0$. Combining (155) and (157), we have for $\varepsilon' = \varepsilon + \varepsilon_{\mathrm{F}}$

$$\begin{aligned}
nR - n\varepsilon' &\leq I(\mathsf{w}; \mathsf{y}_{\mathrm{r}}^n) - I(\mathsf{w}; \mathbf{y}_{\mathrm{e}}^n) \\
&\leq I(\mathsf{w}; \mathsf{y}_{\mathrm{r}}^n, \mathbf{y}_{\mathrm{e}}^n) - I(\mathsf{w}; \mathbf{y}_{\mathrm{e}}^n) && (158)\\
&= I(\mathsf{w}; \mathsf{y}_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n) && (159)\\
&= h(\mathsf{y}_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n) - h(\mathsf{y}_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n, \mathsf{w}) \\
&\leq h(\mathsf{y}_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n) - h(\mathsf{y}_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n, \mathsf{w}, \mathbf{x}^n) && (160)\\
&= h(\mathsf{y}_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n) - h(\mathsf{y}_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n, \mathbf{x}^n) && (161)\\
&= h(\mathsf{y}_{\mathrm{r}}^n | \mathbf{y}_{\mathrm{e}}^n) - \sum_{t=1}^{n} h(\mathsf{y}_{\mathrm{r}}(t) | \mathbf{y}_{\mathrm{e}}(t), \mathbf{x}(t)) && (162)\\
&\leq \sum_{t=1}^{n} h(\mathsf{y}_{\mathrm{r}}(t) | \mathbf{y}_{\mathrm{e}}(t)) - \sum_{t=1}^{n} h(\mathsf{y}_{\mathrm{r}}(t) | \mathbf{y}_{\mathrm{e}}(t), \mathbf{x}(t)) \\
&= nI(\mathbf{x}; \mathsf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}, \mathsf{q}) && (163)\\
&\leq nI(\mathbf{x}; \mathsf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}}) && (164)
\end{aligned}$$

where (158) and (159) each follow from the chain of mutual information, (160) follows from the fact that conditioning cannot increase differential entropy, (161) follows from the Markov relation $\mathsf{w} \leftrightarrow (\mathbf{x}^n, \mathbf{y}_{\mathrm{e}}^n) \leftrightarrow \mathsf{y}_{\mathrm{r}}^n$, and (162) follows from the fact the channel is memoryless. Moreover, (163) is obtained by defining a time-sharing random variable $\mathsf{q}$ that takes values uniformly over the index set $\{1, 2, \ldots, n\}$ and defining $(\mathbf{x}, \mathsf{y}_{\mathrm{r}}, \mathbf{y}_{\mathrm{e}})$ to be the tuple of random variables that conditioned on $\mathsf{q} = t$, have the same joint distribution as $(\mathbf{x}(t), \mathsf{y}_{\mathrm{r}}(t), \mathbf{y}_{\mathrm{e}}(t))$. It then follows that for our choice of $\mathbf{x}$ and given (156), $E[||\mathbf{x}||^2] \leq P$. Finally, (164) follows from the fact that $I(\mathbf{x}; \mathsf{y}_{\mathrm{r}} | \mathbf{y}_{\mathrm{e}})$ is concave in $p_{\mathbf{x}}$ (see, e.g., [9, Appendix I] for a proof), so that Jensen's inequality can be applied.

## Appendix II
## Derivation of (49)

The argument of the logarithm on left hand side of (49) is convex in $\boldsymbol{\theta}$, so it is straightforward to verify that the minimizing $\boldsymbol{\theta}$ is

$$\boldsymbol{\theta} = (\mathbf{I} + P\mathbf{H}_{\mathrm{e}}\mathbf{H}_{\mathrm{e}}^{\dagger})^{-1}(P\mathbf{H}_{\mathrm{e}}\mathbf{h}_{\mathrm{r}} + \boldsymbol{\phi}). \tag{165}$$

In the sequel, we exploit that by the definition of generalized eigenvalues via (1)

$$(\mathbf{I} + P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger})\boldsymbol{\psi}_{\max} = \lambda_{\max}(\mathbf{I} + P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\boldsymbol{\psi}_{\max} \tag{166}$$

or, rearranging

$$\left(\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger} - \lambda_{\max}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}\right)\boldsymbol{\psi}_{\max} = \frac{(\lambda_{\max} - 1)}{P} \cdot \boldsymbol{\psi}_{\max}. \tag{167}$$

First we obtain a more convenient expression for $\boldsymbol{\theta}$, as follows:

$$\begin{aligned}
\boldsymbol{\theta} &= (\mathbf{I} + P\mathbf{H}_{\mathrm{e}}\mathbf{H}_{\mathrm{e}}^{\dagger})^{-1}\left(P\mathbf{H}_{\mathrm{e}}\mathbf{h}_{\mathrm{r}} + \frac{1}{\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}}\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}_{\max}\right) && (168)\\
&= (\mathbf{I} + P\mathbf{H}_{\mathrm{e}}\mathbf{H}_{\mathrm{e}}^{\dagger})^{-1}\frac{\mathbf{H}_{\mathrm{e}}(P\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger} + \mathbf{I})\boldsymbol{\psi}_{\max}}{\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}} \\
&= (\mathbf{I} + P\mathbf{H}_{\mathrm{e}}\mathbf{H}_{\mathrm{e}}^{\dagger})^{-1}\frac{\lambda_{\max}\mathbf{H}_{\mathrm{e}}(P\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}} + \mathbf{I})\boldsymbol{\psi}_{\max}}{\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}} && (169)\\
&= (\mathbf{I} + P\mathbf{H}_{\mathrm{e}}\mathbf{H}_{\mathrm{e}}^{\dagger})^{-1}\frac{\lambda_{\max} \cdot (P\mathbf{H}_{\mathrm{e}}\mathbf{H}_{\mathrm{e}}^{\dagger} + \mathbf{I})\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}_{\max}}{\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}} && (170)\\
&= \lambda_{\max}\boldsymbol{\phi} && (171)
\end{aligned}$$

where (168) follows from substituting (48) into (165), and (169) follows from substituting via (166).

Next, we have that

$$\begin{aligned}
\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^{\dagger}\boldsymbol{\theta} &= \mathbf{h}_{\mathrm{r}} - \frac{\lambda_{\max}}{\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}_{\max} && (172)\\
&= \frac{(\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^{\dagger} - \lambda_{\max}\mathbf{H}_{\mathrm{e}}^{\dagger}\mathbf{H}_{\mathrm{e}})\boldsymbol{\psi}_{\max}}{\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}} \\
&= \frac{(\lambda_{\max} - 1)\boldsymbol{\psi}_{\max}}{P\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}} && (173)\\
& && (174)
\end{aligned}$$

where (172) follows from substituting from (171) with (48), and (173) follows by substituting (167). Thus

$$P||\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^{\dagger}\boldsymbol{\theta}||^2 = (\lambda_{\max} - 1)\left[\frac{(\lambda_{\max} - 1)}{P|\mathbf{h}_{\mathrm{r}}^{\dagger}\boldsymbol{\psi}_{\max}|^2}\right]. \tag{175}$$

To simplify (175) further, we exploit that

$$1-\lambda_{\max}\|\boldsymbol{\phi}\|^2 = 1-\lambda_{\max}\frac{\boldsymbol{\psi}_{\max}^\dagger \mathbf{H}_{\mathrm{e}}^\dagger \mathbf{H}_{\mathrm{e}}\boldsymbol{\psi}_{\max}}{\boldsymbol{\psi}_{\max}^\dagger \mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^\dagger\boldsymbol{\psi}_{\max}} \tag{176}$$

$$= \frac{\boldsymbol{\psi}_{\max}^\dagger(\mathbf{h}_{\mathrm{r}}\mathbf{h}_{\mathrm{r}}^\dagger - \lambda_{\max}\mathbf{H}_{\mathrm{e}}^\dagger\mathbf{H}_{\mathrm{e}})\boldsymbol{\psi}_{\max}}{|\mathbf{h}_{\mathrm{r}}^\dagger\boldsymbol{\psi}_{\max}|^2}$$

$$= \frac{(\lambda_{\max}-1)}{P|\mathbf{h}_{\mathrm{r}}^\dagger\boldsymbol{\psi}_{\max}|^2} \tag{177}$$

where (176) follows by again substituting from (48), and (177) follows by again substituting from (167). In turn, replacing the term in brackets in (175) according to (177) then yields

$$P\|\mathbf{h}_{\mathrm{r}} - \mathbf{H}_{\mathrm{e}}^\dagger\boldsymbol{\theta}\|^2 = (\lambda_{\max}-1)(1-\lambda_{\max}\|\boldsymbol{\phi}\|^2). \tag{178}$$

Finally, substituting (178) then (171) into the left hand side of (49) yields, following some minor algebra, the right hand side as desired.

## ACKNOWLEDGMENT

The authors would like to thank Y. C. Eldar and A. Wiesel for providing an elegant justification that rank one covariance maximizes the upper bound in Theorem 1, which appears between (46)–(47).

## REFERENCES

[1] A. D. Wyner, "The wiretap channel," *Bell Syst. Tech. J.*, vol. 54, pp. 1355–87, 1975.
[2] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. IT-24, pp. 339–348, 1978.
[3] S. K. Leung-Yan-Cheong and M. E. Hellman, "The Gaussian wiretap channel," *IEEE Trans. Inf. Theory*, vol. IT-24, pp. 451–56, 1978.
[4] R. Negi and S. Goel, "Secret communication using artificial noise," in *Proc. Vehic. Tech. Conf.*, Dallas, TX, Sep. 2005, pp. 1906–1910.
[5] S. Goel and R. Negi, "Secret communication in presence of colluding eavesdroppers," in *Proc. IEEE Military Commun. Conf.*, Atlantic City, NJ, 2005, pp. 1501–1506.
[6] 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.
[7] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure broadcasting with multiuser diversity," presented at the Allerton Conf. Commun., Contr., Computing, Monticello, IL, Sep. 2006.
[8] P. Gopala, L. Lai, and H. El Gamal, "On the secrecy capacity of fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4687–4698, Oct. 2008.
[9] A. Khisti, A. Tchamkerten, and G. W. Wornell, "Secure broadcasting over fading channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2453–2469, Jun. 2008.
[10] A. Khisti, G. W. Wornell, A. Wiesel, and Y. Eldar, "On the Gaussian MIMO wiretap channel," presented at the Int. Symp. Inform. Theory, Nice, France, Jun. 2007.
[11] Z. Li, W. Trappe, and R. Yates, "Secret communication via multi-antenna transmission," presented at the Conf. Inform. Sci., Syst. (CISS), Baltimore, MD, Mar. 2007.
[12] S. Shaifee and S. Ulukus, "Achievable rates in Gaussian MISO channels with secrecy constraints," presented at the Int. Symp. Inform. Theory, Nice, France, Jun. 2007.
[13] Y. Liang, G. Kramer, V. Poor, and S. Shamai, "Compound wire-tap channels," presented at the Allerton Conf. Commun., Contr., Computing, Monticello, IL, 2007.
[14] T. Liu, V. Prabhakaran, and S. Vishwanath, "The secrecy capacity of a class of parallel Gaussian compound wiretap channels," presented at the Int. Symp. Inform. Theory, Toronto, ON, Canada, Jul. 2008.
[15] A. Khisti and G. W. Wornell, "Secure transmission with multiple antennas II: The MIMOME wiretap channel," *IEEE Trans. Inf. Theory*, to be published.
[16] M. Bloch, J. Barros, M. R. D. Rodrigues, and S. W. McLaughlin, "Wireless information-theoretic security," *IEEE Trans. Inf. Theory*, vol. 54, no. 6, pp. 2515–2534, Jun. 2008.
[17] J. Barros, Short Intensive Course on Physical-Layer Security: Theory and Practice Univ. Illinois, Urbana-Champaign, 2007 [Online]. Available: http://www.dcc.fc.up.pt/ barros/
[18] G. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.
[19] LAPACK Users' Guide,3rd ed. Aug. 1999 [Online]. Available: http://www.netlib.org/lapack/lug/lapack_lug.html
[20] U. M. Maurer and S. Wolf, "Information-theoretic key agreement: From weak to strong secrecy for free," presented at the EUROCRYPT, Bruges, Belgium, 2000.
[21] I. Csiszár, "Almost independence and secrecy capacity," *Probl. Inform. Transmission*, vol. 32, pp. 48–57, 1996.
[22] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.
[23] C. Li and R. Mathias, "Extremal characterizations of the Schur complement and resulting inequalities," *SIAM Rev.*, vol. 42, no. 2, pp. 233–246, 2000.
[24] A. M. Tulino and S. Verdu, "Random matrix theory and wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 1, Jun. 2004.
[25] M. Kang and M. S. Alouini, "Hotelling's generalized distribution and performance of 2D-RAKE receivers," *IEEE Trans. Inf. Theory*, vol. 49, no. 1, pp. 317–323, Jan. 2003.
[26] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. Hoboken, NJ: Wiley, 1982.
[27] 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, no. 9, pp. 4033–4039, Sep. 2009.
[28] A. Khisti and G. W. Wornell, "The MIMOME channel," presented at the Allerton Conf. Commun., Contr., Computing, Monticello, IL, Sep. 2007.
[29] F. E. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," presented at the CoRR, Oct. 10, 2007, abs/0710.1920.
[30] T. Liu and S. Shamai, "A note on the secrecy capacity of the multi-antenna wiretap channel," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2547–2553, 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, CA; 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