Capacity Bounds Via Duality With Applications to Multiple-Antenna Systems on Flat-Fading Channels

Amos Lapidoth, *Senior Member, IEEE,* and Stefan M. Moser, *Student Member, IEEE*

***Abstract*—A technique is proposed for the derivation of upper bounds on channel capacity. It is based on a dual expression for channel capacity where the maximization (of mutual information) over distributions on the channel input alphabet is replaced with a minimization (of average relative entropy) over distributions on the channel output alphabet. We also propose a technique for the analysis of the asymptotic capacity of cost-constrained channels. The technique is based on the observation that under fairly mild conditions capacity achieving input distributions "escape to infinity."**

The above techniques are applied to multiple-antenna flat-fading channels with memory where the realization of the fading process is unknown at the transmitter and unknown (or only partially known) at the receiver. It is demonstrated that, for high signal-to-noise ratio (SNR), the capacity of such channels typically grows only double-logarithmically in the SNR. To better understand this phenomenon and the rates at which it occurs, we introduce the fading number as the second-order term in the high-SNR asymptotic expansion of capacity, and derive estimates on its value for various systems. It is suggested that at rates that are significantly higher than the fading number, communication becomes extremely power inefficient, thus posing a practical limit on practically achievable rates.

Upper and lower bounds on the fading number are also presented. For single-input–single-output (SISO) systems the bounds coincide, thus yielding a complete characterization of the fading number for general stationary and ergodic fading processes. We also demonstrate that for memoryless multiple-input single-output (MISO) channels, the fading number is achievable using beam-forming, and we derive an expression for the optimal beam direction. This direction depends on the fading law and is, in general, not the direction that maximizes the SNR on the induced SISO channel. Using a new closed-form expression for the expectation of the logarithm of a noncentral chi-square distributed random variable we provide some closed-form expressions for the fading number of some systems with Gaussian fading, including SISO systems with circularly symmetric stationary and ergodic Gaussian fading. The fading number of the latter is determined by the fading mean, fading variance, and the mean squared error in predicting the present fading from its past; it is not directly related to the Doppler spread.

For the Rayleigh, Ricean, and multiple-antenna Rayleigh-fading channels we also present firm (nonasymptotic) upper and lower bounds on channel capacity. These bounds are asymptotically tight in the sense that their difference from capacity approaches zero at high SNR, and their ratio to capacity approaches one at low SNR.

***Index Terms*—Channel capacity, duality, fading channels, flat fading, high signal-to-noise ratio (SNR), multiple-antenna fading number, noncentral chi-square, Rayleigh fading, Ricean fading, upper bounds.**

I. INTRODUCTION

THE purpose of this paper is twofold: to propose a general technique for deriving upper bounds on channel capacity, and to use this technique in order to study multiple-antenna systems on flat-fading channels. To motivate the proposed technique consider the classical expression for the capacity C of a discrete memoryless channel (DMC) of law $W(y|x)$ over the finite input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$

$$C = \max_{Q \in \mathcal{P}(\mathcal{X})} I(Q; W) \tag{1}$$

where $\mathcal{P}(\mathcal{X})$ denotes the set of all probability measures on $\mathcal{X}$ and where $I(Q; W)$ denotes the mutual information between the channel terminals when the input is distributed according to the law Q. That is,

$$I(Q; W) = \sum_{x,\, y} Q(x) W(y|x) \log \frac{W(y|x)}{(QW)(y)} \tag{2}$$

where (QW) denotes the output distribution corresponding to the input law Q, i.e.,

$$(QW)(y) = \sum_{x' \in \mathcal{X}} Q(x') W(y|x'), \qquad y \in \mathcal{Y}. \tag{3}$$

While the optimization over input distributions complicates the exact computation of C, (1) leads to very natural lower bounds on C. Indeed, any input distribution $Q \in \mathcal{P}(\mathcal{X})$ leads to a lower bound

$$C \geq I(Q; W). \tag{4}$$

A good choice for Q in the above would be a distribution that is close to a capacity-achieving input distribution and that leads to a tractable expression for $I(Q; W)$.

This latter issue of tractability may not be so critical for DMCs, but it is quite important for channels over continuous alphabets. For such channels, the mutual information can be expanded in terms of differential entropies in two ways

$$\begin{aligned} I(Q; W) &= h(X) - h(X|Y) & (5)\\ &= h(Y) - h(Y|X). & (6) \end{aligned}$$

Manuscript received October 2, 2002; revised June 14, 2003. This research was conducted in part while A. Lapidoth was a resident at the Rockefeller Foundation Bellagio Study and Conference Center. The work of S. M. Moser was supported in part by the Swiss Federal Institute of Technology under Grant TH-23 02-2. The material in this paper was presented in part at the 2001 IEEE International Symposium on Information Theory, Washington, DC; the 2001 Workshop on Information Theory, Cairns, Australia; and the 2002 IEEE International Symposium on Information Theory, Lausanne, Switzerland.

The authors are with the Department of Information Technology and Electrical Engineering, Swiss Federal Institute of Technology in Zurich (ETHZ), CH-8092 Zurich, Switzerland.

Communicated by B. Hassibi, Associate Editor for Communications.
Digital Object Identifier 10.1109/TIT.2003.817449

To understand the difficulty in evaluating $I(Q; W)$ note that channels are typically modeled so that the output law $W(\cdot|x)$ corresponding to each input $x \in \mathcal{X}$ be a "nice" function, but that for a given input distribution $Q \in \mathcal{P}(\mathcal{X})$—even if "nice"—the posterior law on X given Y will typically be complicated to compute, let alone $h(X|Y)$. Thus, while a "nice" choice for Q will typically allow for an analytic calculation of $h(X)$, the calculation of $h(X|Y)$ will typically be complicated and (5) might not be tractable. Alternatively, if one tries to compute (6) then the nice law of $W(\cdot|x)$ and a nice choice for Q will typically allow one to compute $h(Y|X)$, but the computation of $h(Y)$, which is required for (6), might be difficult because the output law corresponding to a nice input and a nice channel need not be nice.

A dual expression for channel capacity is [1]

$$C = \min_{R \in \mathcal{P}(\mathcal{Y})} \max_{x \in \mathcal{X}} D\left(W(\cdot|x) \| R(\cdot)\right) \tag{7}$$

where $D(\cdot\|\cdot)$ denotes relative entropy so that

$$D\left(W(\cdot|x) \| R(\cdot)\right) = \sum_{y \in \mathcal{Y}} W(y|x) \log \frac{W(y|x)}{R(y)}. \tag{8}$$

Every choice of a distribution $R(\cdot)$ on the output $\mathcal{Y}$ thus leads to an upper bound on channel capacity

$$C \leq \max_{x \in \mathcal{X}} D\left(W(\cdot|x) \| R(\cdot)\right). \tag{9}$$

In fact, by considering the identity [2]

$$\sum_{x \in \mathcal{X}} Q(x) D\left(W(\cdot|x) \| R(\cdot)\right) = I(Q; W) + D\left((QW)(\cdot) \| R(\cdot)\right) \tag{10}$$

and by noting that relative entropy is nonnegative, we obtain the bound

$$\boxed{I(Q; W) \leq \sum_{x \in \mathcal{X}} Q(x) D\left(W(\cdot|x) \| R(\cdot)\right), \quad R \in \mathcal{P}(\mathcal{Y})} \tag{11}$$

which implies (9).

As noted above, any choice of a distribution $R(\cdot) \in \mathcal{P}(\mathcal{Y})$ on the output alphabet leads to an upper bound on channel capacity via (9). One should typically choose $R(\cdot)$ to be close to the capacity-achieving output distribution and so as to guarantee that (9) be tractable. This latter condition need not be so difficult to satisfy. Indeed, since the channel law $W(\cdot|x)$ is often modeled using a "nice" law, and since we are at liberty to choose $R(\cdot)$ to be nice, there is hope that $D(W(\cdot|x)\|R(\cdot))$ may be tractable and be a reasonable function of x that can be then maximized. While this latter maximization is unavoidable, it is at least over input symbols and not over distributions.

In this paper we shall extend (11), and hence also (9), to general alphabets and also demonstrate how to account for input constraints. Such constraints can be accounted for by modifying (9) by introducing Lagrange multipliers, as in [1], or by working with (11), as we have chosen to do.

We shall apply the proposed approach to the study of the capacity of multiple-antenna flat-fading channels where the transmitter and receiver—while cognizant of the fading probability law—have no knowledge (or, in the receiver case, only partial knowledge) of the realization of the fading matrix. Other channels to which the proposed approach has been successfully applied include finite-state channels with intersymbol interference memory [3], phase noise channels [4], and the Poisson channel [5]. For an extension of this technique to the analysis of error exponents see [6].

The fading model we address is described in Section III after a brief word about notation in Section II. The rest of the paper is organized as follows. In Section IV, we present our main results concerning the capacity of multiple-antenna fading channels. Subsequent sections are more technical. In Section V, we prove the extension of (11) to continuous alphabet channels. In Section VI, we provide some of the mathematical background that will be useful in the study of the capacity of the fading channel. This section can be glanced over in a first reading and referred to later as needed. Section VII concludes the paper with a brief summary and a discussion of some of the results.

II. Notation

We try to use upper case letters for random quantities and lower case letters for their realizations. This rule, however, is broken when dealing with deterministic matrices and some constants. To better differentiate between scalars, vectors, and matrices we have resorted to using different fonts for the different quantities. Upper case letters such as X are used to denote scalar random variables taking value in the reals $\mathbb{R}$ or in the complex plane $\mathbb{C}$. Their realizations are typically written in lower case, e.g., x. For random vectors we use bold face capitals, e.g., $\boldsymbol{X}$ and bold lower case for their realization, e.g., $\boldsymbol{x}$. Deterministic matrices are denoted by upper case letters but of a special font, e.g., H. For random matrices we use yet another font, e.g., $\mathbb{H}$. Scalars are typically denoted using Greek letters, but the energy per symbol is denoted by $\mathcal{E}_{\mathrm{s}}$.

The entries of matrices are denoted using superscripts so that $H^{(r,t)}$ denotes the (random) component of the random matrix $\mathbb{H}$ that lies in row- r and column-t. Note that our generic row index is r and the generic column index is t because we think of r as indexing the receive antennas and of t as indexing the transmit antennas. Consequently, the number of rows in the matrix will be often denoted by n_{R} and the number of columns by n_{T}. Subscripts will be typically reserved for time indexes. Thus, the fading matrix at time k will be denoted by $\mathbb{H}_k$.

We use $\|\cdot\|$ to denote the Euclidean norm of vectors or the Euclidean operator norm of matrices. That is,

$$\|\boldsymbol{x}\| = \sqrt{\sum_{t=1}^{n_{\mathrm{T}}} |x^{(t)}|^2}, \qquad \boldsymbol{x} \in \mathbb{C}^{n_{\mathrm{T}}} \tag{12}$$

$$\|\mathsf{A}\| = \max_{\|\hat{\boldsymbol{w}}\|=1} \|\mathsf{A}\hat{\boldsymbol{w}}\|. \tag{13}$$

Thus, $\|\mathsf{A}\|$ is the maximal singular value of the matrix A.

The Frobenius norm of matrices is denoted by $\|\cdot\|_{\mathrm{F}}$ and is given by the square root of the sum of the squared magnitudes of the elements of the matrix, i.e.,

$$\|\mathsf{A}\|_{\mathrm{F}} = \sqrt{\mathrm{tr}\,(\mathsf{A}^\dagger \mathsf{A})}. \tag{14}$$

Here $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, $(\cdot)^\dagger$ denotes Hermitian conjugation, and we shall use $(\cdot)^{\mathsf{T}}$ to denote the transpose (without conjugation) of a matrix. Note that for any matrix A

$$\|\mathsf{A}\| \leq \|\mathsf{A}\|_{\mathrm{F}}$$

as can be verified by upper-bounding the squared magnitude of each of the components of $\mathsf{A}\hat{\boldsymbol{w}}$ using the Cauchy–Schwarz inequality.

All rates specified in this paper are in nats per channel use. We use $\log(\cdot)$ to denote the natural logarithmic function, and set $\log^+(\cdot)$ to denote its positive part, i.e.,

$$\log^+(\xi) = \max\{0, \log(\xi)\}. \tag{15}$$

We shall denote the indicator function by $I\{\text{statement}\}$. It takes on the value 1 if the statement is true, and the value 0 if the statement is false.

We shall denote the mean-μ variance-σ^2 univariate real Gaussian distribution by $\mathcal{N}_{\mathbb{R}}(\mu, \sigma^2)$. Similarly, the mean-$\boldsymbol{\mu}$ covariance-K multivariate real Gaussian distribution will be denoted $\mathcal{N}_{\mathbb{R}}(\boldsymbol{\mu}, \mathsf{K})$. Analogously, a complex random variable X will be said to have a $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$ distribution if $X - \mu$ is a circularly symmetric Gaussian random variable of variance $\mathsf{E}[|X|^2] = \sigma^2$, i.e., if the real and imaginary parts of $X - \mu$ are independent $\mathcal{N}_{\mathbb{R}}(0, \sigma^2/2)$ random variables. Similarly, we shall write $\boldsymbol{X} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \mathsf{K})$ if $\boldsymbol{X} - \boldsymbol{\mu}$ is a circularly symmetric zero-mean Gaussian random vector of covariance matrix $\mathsf{E}[\boldsymbol{X}\boldsymbol{X}^\dagger] = \mathsf{K}$. The notation $\mathcal{N}(\boldsymbol{\mu}, \mathsf{K})$, without the subscript to indicate whether the distribution is complex or real, will indicate that the stated result holds in both cases.

In dealing with sequences of random variables we shall use a combination of superscripts and lower scripts to address consecutive subsets. Thus, if $X_1, X_2, \ldots,$ is a sequence of random variables, then X_k^n will designate the sequence $X_k, \ldots, X_n$.

III. The Channel Model

We consider a channel with n_{T} transmit antennas and n_{R} receive antennas whose time-k output $\boldsymbol{Y}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ is given by

$$\boldsymbol{Y}_k = \mathbb{H}_k \boldsymbol{x}_k + \boldsymbol{Z}_k = \sum_{t=1}^{n_{\mathrm{T}}} x_k^{(t)} \boldsymbol{H}_k^{(t)} + \boldsymbol{Z}_k. \tag{16}$$

Here, $\boldsymbol{x}_k = (x_k^{(1)}, \ldots, x_k^{(n_{\mathrm{T}})})^{\mathsf{T}} \in \mathbb{C}^{n_{\mathrm{T}}}$ denotes the time-k input vector; the random matrix $\mathbb{H}_k \in \mathbb{C}^{n_{\mathrm{R}} \times n_{\mathrm{T}}}$ denotes the time-k fading matrix of columns $\boldsymbol{H}_k^{(1)}, \ldots, \boldsymbol{H}_k^{(n_{\mathrm{T}})}$; and the random vector $\boldsymbol{Z}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ denotes the time-k additive noise vector.

Unless otherwise specified, we shall assume throughout that the random vectors $\{\boldsymbol{Z}_k\}$ are spatially and temporally white zero-mean circularly symmetric complex Gaussians, i.e., that $\{\boldsymbol{Z}_k\}$ are independent and identically distributed (i.i.d.) $\mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma^2 \mathsf{I})$ for some $\sigma^2 > 0$. Here I denotes the identity matrix. Similarly, we shall assume throughout that the matrix-valued fading process $\{\mathbb{H}_k\}$ is stationary and ergodic and independent of the vector-valued additive noise process $\{\boldsymbol{Z}_k\}$. We shall also assume a finite-energy fading gain, i.e.,

$$\mathsf{E}\left[\|\mathbb{H}_k\|_{\mathrm{F}}^2\right] < \infty \tag{17}$$

where $\|\mathbb{H}_k\|_{\mathrm{F}}$ denotes the Frobenius norm of the matrix $\mathbb{H}_k$, see (14).

We denote the capacity of this channel with average power $\mathcal{E}_{\mathrm{s}}$ by $C(\mathrm{SNR})$, where $\mathrm{SNR} = \mathcal{E}_{\mathrm{s}}/\sigma^2$ denotes the signal-to-noise ratio. Thus,

$$C(\mathrm{SNR}) = \lim_{n\to\infty} \sup \frac{1}{n} I(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n; \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n) \tag{18}$$

where the supremum is over all joint distributions on the input vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ under which

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[\|\boldsymbol{X}_k\|^2\right] \leq \mathcal{E}_{\mathrm{s}} \tag{19}$$

and where $I(\cdot;\cdot)$ denotes the mutual information functional.

We shall often focus on *memoryless fading* where the random matrices $\{\mathbb{H}_k\}$ are i.i.d. In this case, we shall drop the time-dependence index and write

$$\boldsymbol{Y} = \mathbb{H}\boldsymbol{x} + \boldsymbol{Z} = \sum_{t=1}^{n_{\mathrm{T}}} x^{(t)} \boldsymbol{H}^{(t)} + \boldsymbol{Z}. \tag{20}$$

Note that memoryless fading still allows for dependence among the components of the fading matrix at a given instant k. Thus, in (20) the components of $\mathbb{H}$ need not be independent of each other.

Since mutual information is concave in the input distribution, for memoryless fading we can replace (19) with the stricter constraint

$$\mathsf{E}\left[\boldsymbol{X}^\dagger \boldsymbol{X}\right] \leq \mathcal{E}_{\mathrm{s}}. \tag{21}$$

A special case of memoryless fading is *memoryless Gaussian fading*. In this case, the matrix $\mathbb{H}$ can be written as

$$\mathbb{H} = \mathsf{D} + \tilde{\mathbb{H}} \tag{22}$$

where the mean matrix D is a deterministic $n_{\mathrm{R}} \times n_{\mathrm{T}}$ complex matrix, and where the $n_{\mathrm{T}} \cdot n_{\mathrm{R}}$ components $\{\tilde{H}^{(r,t)}\}_{r,t}$ of $\tilde{\mathbb{H}}$ are zero-mean jointly circularly symmetric and jointly Gaussian complex random variables. To be even more explicit, we shall sometimes refer to memoryless fading of a law that is not necessarily Gaussian as *general memoryless fading*.

Some special cases of memoryless Gaussian fading include the following.

- *Rayleigh fading*, where $n_{\mathrm{R}} = n_{\mathrm{T}} = 1$, the mean matrix D is zero, and $\tilde{H}$ is a zero-mean unit-variance circularly symmetric complex Gaussian.
- *Multiple-antenna Rayleigh fading*, where $\mathsf{D} = \mathbf{0}$ and the $n_{\mathrm{R}} \cdot n_{\mathrm{T}}$ components of $\tilde{\mathbb{H}}$ are independent zero-mean unit-variance circularly symmetric complex Gaussians.

- *Ricean fading*, where $n_{\mathrm{R}} = n_{\mathrm{T}} = 1$, the mean matrix—which is now a scalar d called the "specular component"—is not necessarily zero, and $\tilde{H}$ is a zero-mean unit-variance circularly symmetric complex Gaussian.

IV. Capacity Results

Apart from Section IV-A, which presents some results that are applicable to general channels of output alphabet $\mathbb{C}^{n_{\mathrm{R}}}$, the results in this section are focused on fading channels. In Section IV-A, we use the continuous-alphabet version of (11), namely, (186) of Theorem 5.1, in conjunction with the family of output laws (204) to derive general upper bounds on the mutual information between the terminals of a channel of output alphabet $\mathbb{C}^{n_{\mathrm{R}}}$. In Section IV-B, we use these inequalities to demonstrate the inefficiency of high-SNR signaling on fading channels. Motivated by these results, we define in Section IV-C the fading number, which is the second-order term of the high-SNR expansion of channel capacity and which gives some indication of the rates above which channel capacity increases only double-logarithmically with the SNR. Section IV-D is devoted to the calculation of the fading number of memoryless fading and Section IV-E to the calculation of the fading number for fading with memory. This section is concluded in Section IV-F with a nonasymptotic capacity analysis of some specific memoryless Gaussian fading channels including the Rayleigh, Rice, and multiple-antenna Rayleigh fading channels.

A. A Specific Bound on Mutual Information

Once we extend the basic inequality (11) to general alphabets in Theorem 5.1 of Section V, we can apply it to channels whose output alphabet is $\mathbb{C}^{n_{\mathrm{R}}}$ by considering the output distributions $R(\cdot)$ whose densities (with respect to the Lebesgue measure on $\mathbb{C}^{n_{\mathrm{R}}}$) are given by

$$\frac{\Gamma(n_{\mathrm{R}})|\det \mathsf{A}|^2}{\pi^{n_{\mathrm{R}}}\beta^{\alpha}\Gamma(\alpha,\, \delta/\beta)} \cdot \left(\|\mathsf{A}\boldsymbol{y}\|^2 + \delta\right)^{(\alpha-1)} \cdot \|\mathsf{A}\boldsymbol{y}\|^{2(1-n_{\mathrm{R}})} e^{-(\|\mathsf{A}\boldsymbol{y}\|^2+\delta)/\beta}, \qquad \boldsymbol{y} \in \mathbb{C}^{n_{\mathrm{R}}} \tag{23}$$

where $\alpha,\, \beta > 0$, $\delta \geq 0$, and where A is any nonsingular deterministic $n_{\mathrm{R}} \times n_{\mathrm{R}}$ complex matrix. (See (204) and the discussion preceding it in Section VI-A for a discussion of this family of densities.) Here $\Gamma(\cdot)$ denotes the Gamma function (197) and $\Gamma(\cdot,\, \cdot)$ denotes the incomplete Gamma function (200).

With this choice of $R(\cdot)$ we have

$$\begin{aligned} &D\left(W(\cdot|\boldsymbol{x})\|R(\cdot)\right) \\ &\quad = -h(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}) + \log \pi^{n_{\mathrm{R}}} + \log \beta^{\alpha} \\ &\qquad + \log\Gamma(\alpha,\, \delta/\beta) - \log\Gamma(n_{\mathrm{R}}) - \log|\det \mathsf{A}|^2 \\ &\qquad + (n_{\mathrm{R}}-1)\mathsf{E}\left[\log\|\mathsf{A}\boldsymbol{Y}\|^2 | \boldsymbol{X}=\boldsymbol{x}\right] \\ &\qquad + (1-\alpha)\mathsf{E}\left[\log\left(\|\mathsf{A}\boldsymbol{Y}\|^2+\delta\right) | \boldsymbol{X}=\boldsymbol{x}\right] \\ &\qquad + \frac{1}{\beta}\,\mathsf{E}\left[\|\mathsf{A}\boldsymbol{Y}\|^2+\delta | \boldsymbol{X}=\boldsymbol{x}\right] \end{aligned} \tag{24}$$

so that by the basic inequality (11), (186) we obtain a general upper bound on the the mutual information for a channel whose output takes value in $\mathbb{C}^{n_{\mathrm{R}}}$

$$\boxed{\begin{aligned} &I(Q;\, W) \\ &\quad \leq -h_Q(\boldsymbol{Y}|\boldsymbol{X}) + \log \pi^{n_{\mathrm{R}}} + \log \beta^{\alpha} \\ &\qquad + \log\Gamma(\alpha,\, \delta/\beta) - \log\Gamma(n_{\mathrm{R}}) - \log|\det \mathsf{A}|^2 \\ &\qquad + (n_{\mathrm{R}}-1)\mathsf{E}_Q\left[\log\|\mathsf{A}\boldsymbol{Y}\|^2\right] \\ &\qquad + (1-\alpha)\mathsf{E}_Q\left[\log(\|\mathsf{A}\boldsymbol{Y}\|^2+\delta)\right] \\ &\qquad + \frac{1}{\beta}\,\mathsf{E}_Q\left[\|\mathsf{A}\boldsymbol{Y}\|^2+\delta\right], \\ &\qquad\quad \alpha,\, \beta > 0,\, \delta \geq 0,\, \det(\mathsf{A}) \neq 0 \end{aligned}} \tag{25}$$

where $h_Q(\boldsymbol{Y}|\boldsymbol{X}) = \int h(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x})\,\mathrm{d}Q(\boldsymbol{x})$ denotes the average conditional differential entropy when $\boldsymbol{X}$ is distributed according to the law Q. Notice that in (25) we have denoted a generic input to the channel by $\boldsymbol{X}$ because we have in mind that the input to the channel is a complex vector, but the result is more general.

The (typically suboptimal) choice of

$$\delta = 0, \qquad \beta = \frac{\mathsf{E}_Q\left[\|\mathsf{A}\boldsymbol{Y}\|^2\right]}{\alpha} \tag{26}$$

in (25) yields the simpler upper bound

$$\begin{aligned} I(Q;\, W) \leq{}& \log \pi^{n_{\mathrm{R}}} - \log\Gamma(n_{\mathrm{R}}) - \log|\det \mathsf{A}|^2 \\ &+ n_{\mathrm{R}}\mathsf{E}_Q\left[\log\|\mathsf{A}\boldsymbol{Y}\|^2\right] - h_Q(\boldsymbol{Y}|\boldsymbol{X}) \\ &+ \alpha\left(1 + \log \mathsf{E}_Q\left[\|\mathsf{A}\boldsymbol{Y}\|^2\right] - \mathsf{E}_Q\left[\log\|\mathsf{A}\boldsymbol{Y}\|^2\right]\right) \\ &+ \log\Gamma(\alpha) - \alpha\log\alpha, \qquad \alpha > 0,\, \det(\mathsf{A}) \neq 0. \end{aligned} \tag{27}$$

This upper bound is tight enough to obtain the first term in the high-SNR capacity expansion, but not quite tight enough for the finer analysis of the second term (which will be defined later as the fading number).

Note 4.1: It is interesting to note that for low-SNR fading channels, the crude bound (27) is tight. Indeed, if we further simplify it with the choice of $\alpha = n_{\mathrm{R}}$ and A satisfying $\mathsf{A}^{\dagger}\mathsf{A} = (\mathsf{E}[[\boldsymbol{Y}\boldsymbol{Y}^{\dagger}])^{-1}$ then (27) reduces to the max-entropy bound[1]

$$I(Q;\, W) \leq \log\left((\pi \cdot e)^{n_{\mathrm{R}}} \det \mathsf{E}\left[\boldsymbol{Y}\boldsymbol{Y}^{\dagger}\right]\right) - h(\boldsymbol{Y}|\boldsymbol{X}) \tag{28}$$

which is tight enough to obtain the slope of the capacity–energy curve at zero SNR. See Appendix I for details.

To use (25) to obtain upper bounds on channel capacity, one needs to upper bound the right-hand side (RHS) of (25) over all admissible input distributions. For some examples on how this may be carried out, please see Section IV-F. The analysis typically requires one to derive upper bounds on expressions of the form $\mathsf{E}_Q[g(X)]$ for some real function $g(\cdot)$ and for some unknown (capacity achieving) input distribution that is only known to satisfy some input constraint, e.g., $\mathsf{E}_Q[|X|^2] \leq \mathcal{E}_{\mathrm{s}}$. This is often performed using Jensen's inequality (if $g(\cdot)$ is concave),

[1] The fact that this choice reduces to the max-entropy bound is not surprising. Indeed, the choice $\alpha = n_{\mathrm{R}}$ reduces the Gamma distribution (199) to a central χ^2 distribution, thus reducing (203) to an i.i.d. multivariate Gaussian distribution so that (204) becomes a general multivariate Gaussian distribution.

or using the trivial upper bound $\sup_\xi g(\xi)$, when all else fails. Under additional support constraints (peak power) and/or additional moments constraints one may resort to results on Chebyshev Systems, see, e.g., [7]. Another ueful approach, which we demonstrate in Section IV-F, is the use of ideas related to stochastic ordering of distributions [8]; see also Section VI-B.

We may also apply (11) to channels $W(t|s)$ whose output T takes value in the set of nonnegative reals $\mathbb{R}^+$. We can choose the output distribution $R(\cdot)$ to be a regularized Gamma distribution (199), so that

$$\begin{aligned} &D(W(\cdot|s)\|R(\cdot)) \\ &\quad = -h(T|S=s) + \log\beta^\alpha + \log\Gamma(\alpha,\, \delta/\beta) \\ &\qquad + (1-\alpha)\mathsf{E}[\log(T+\delta)|S=s] + \frac{1}{\beta}\,\mathsf{E}[T+\delta|S=s]. \end{aligned} \tag{29}$$

Using the basic inequality (186), we obtain the general upper bound on channels whose output takes value in $\mathbb{R}^+$

$$\boxed{\begin{aligned} &I(Q;\,W) \\ &\quad \le -h_Q(T|S) + \log\beta^\alpha + \log\Gamma(\alpha,\,\delta/\beta) \\ &\qquad + (1-\alpha)\mathsf{E}_Q[\log(T+\delta)] + \frac{1}{\beta}\,\mathsf{E}_Q[T+\delta], \\ &\qquad\quad \alpha,\,\beta>0,\,\delta\ge 0. \end{aligned}} \tag{30}$$

This inequality has proven useful in [4], in the analysis of channels with phase noise, and in [5], in the study of the capacity of the discrete-time Poisson channel. In both cases, it was used with the (typically suboptimal) choice of

$$\delta = 0, \qquad \beta = \frac{\mathsf{E}_Q[T]}{\alpha} \tag{31}$$

which yields the simpler upper bound

$$\begin{aligned} I(Q;\,W) \le& -h_Q(T|S) + \mathsf{E}_Q[\log T] + \log\Gamma(\alpha) \\ &+ \alpha(1+\log\mathsf{E}_Q[T] - \mathsf{E}_Q[\log T]) - \alpha\log\alpha, \\ &\qquad\qquad \alpha > 0. \end{aligned} \tag{32}$$

In the Poisson case, where the output is discrete, it was applied to an information lossless smoothed version of the output.

B. Communication at High SNR Is Power Inefficient

We now turn to some asymptotic analysis of channel capacity at high SNR. Our first result here is that at high SNR capacity grows only double-logarithmically in the SNR, and in fact, the difference between channel capacity and $\log\log \mathrm{SNR}$ is bounded as the SNR tends to infinity. We shall state this result in a fairly general setting that also allows for the availability of some side information at the receiver (but not at the transmitter). To demonstrate the robustness of this result, we shall state it without requiring that the additive noise be spatially and temporally white Gaussian. We shall only require that it be stationary and ergodic, of finite energy, and of finite entropy rate.

Theorem 4.2: Consider a multiple-input multiple-output (MIMO) fading channel

$$\boldsymbol{Y} = \mathbb{H}_k \boldsymbol{x}_k + \boldsymbol{Z}'_k \tag{33}$$

with some receiver side information (possibly null) $\boldsymbol{S}_k$, where the fading process $\{\mathbb{H}_k\}$ and receiver side information $\{\boldsymbol{S}_k\}$ are jointly stationary and ergodic, and independent of the stationary and ergodic additive noise process $\{\boldsymbol{Z}'_k\}$. Assume further that the joint law of $(\{\mathbb{H}_k\},\,\{\boldsymbol{S}_k\},\,\{\boldsymbol{Z}'_k\})$ does not depend on the input sequence $\{\boldsymbol{x}_k\}$. Let $C(\mathcal{E}_\mathrm{s})$ denote the capacity of this channel under an average power constraint $\mathcal{E}_\mathrm{s}$ on the input, so that

$$C(\mathcal{E}_\mathrm{s}) = \lim_{n\to\infty} \frac{1}{n} \sup I\left(\boldsymbol{X}_1^n;\, \boldsymbol{Y}_1^n,\, \boldsymbol{S}_1^n\right) \tag{34}$$

where the supremum is over all input distributions on $\boldsymbol{X}_1^n$ satisfying

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[\|\boldsymbol{X}_k\|^2\right] \le \mathcal{E}_\mathrm{s}. \tag{35}$$

Assume that both $\{\mathbb{H}_k\}$ and $\{\boldsymbol{Z}'_k\}$ are of finite differential entropy rate

$$h(\{\mathbb{H}_k\}), \quad h(\{\boldsymbol{Z}'_k\}) > -\infty; \tag{36}$$

that both have finite second moments

$$\mathsf{E}\left[\|\mathbb{H}_k\|_\mathrm{F}^2\right], \quad \mathsf{E}\left[\|\boldsymbol{Z}'_k\|^2\right] < \infty; \tag{37}$$

and that the mutual information rate $I(\{\mathbb{H}_k\};\,\{\boldsymbol{S}_k\})$ is finite, i.e.,

$$\lim_{n\to\infty} \frac{1}{n}\, I\left(\mathbb{H}_1^n;\, \boldsymbol{S}_1^n\right) < \infty. \tag{38}$$

Then

$$\overline{\lim_{\mathcal{E}_\mathrm{s}\to\infty}} \{C(\mathcal{E}_\mathrm{s}) - \log\log\mathcal{E}_\mathrm{s}\} < \infty. \tag{39}$$

Proof: The proof of this theorem for memoryless fading, memoryless additive noise, and in the absence of receiver side information is given in Appendix II. It is based on an asymptotic analysis of the bound (27) with A chosen as the $n_\mathrm{R}\times n_\mathrm{R}$ identity matrix.

The more general case follows from the simpler case by Lemma 4.5 ahead. □

It is interesting to note that under the assumptions of Theorem 4.2, Gaussian input signals are highly suboptimal. In fact, such input signals achieve a mutual information that is *bounded* in the power $\mathcal{E}_\mathrm{s}$. This result was recently proved by Lapidoth and Shamai [9, Proposition 6.3.1] for single-antenna ($n_\mathrm{R} = n_\mathrm{T} = 1$) and Gaussian inputs. Here we generalize it to the MIMO case and any scale family of input distributions.

Theorem 4.3: Let the fading process $\{\mathbb{H}_k\}$ and additive noise process $\{\boldsymbol{Z}'_k\}$ satisfy the assumptions of Theorem 4.2, i.e., be independent stationary and ergodic processes satisfying (36) and (37). Let $\{\boldsymbol{X}_k\}$ be some stationary process (independent of the fading and additive noise) with $\mathsf{E}[\|\boldsymbol{X}_k\|^2] = 1$ and

$$\mathsf{E}[\log\|\boldsymbol{X}_k\|] > -\infty. \tag{40}$$

Then

$$\begin{aligned} \sup_{\mathcal{E}_\mathrm{s}>0} \lim_{n\to\infty} \frac{1}{n}\, I\Big(&\sqrt{\mathcal{E}_\mathrm{s}}\boldsymbol{X}_1,\,\ldots,\,\sqrt{\mathcal{E}_\mathrm{s}}\boldsymbol{X}_n; \\ &\sqrt{\mathcal{E}_\mathrm{s}}\mathbb{H}_1\boldsymbol{X}_1 + \boldsymbol{Z}'_1,\,\ldots,\,\sqrt{\mathcal{E}_\mathrm{s}}\mathbb{H}_n\boldsymbol{X}_n + \boldsymbol{Z}'_n\Big) < \infty. \end{aligned} \tag{41}$$

Proof: For a proof in the memoryless case see Appendix III. The more general case (even with some receiver side information) follows from Lemma 4.5 ahead. □

Note 4.4: Condition (40) is satisfied whenever $\boldsymbol{X}_k$ is a nondeterministic Gaussian vector. In fact, it is satisfied whenever some subset of the components of $\boldsymbol{X}_k$ has a finite joint differential entropy; see Lemma 6.7. (For deterministic inputs the claim is trivial.)

To better understand the role played by the channel memory and by the side information, the following lemma can be useful.

Lemma 4.5: Let $\boldsymbol{x}_k$, $\boldsymbol{Y}_k$, $\mathbb{H}_k$, $\boldsymbol{S}_k$, and $\boldsymbol{Z}'_k$ be as in Theorem 4.2. Then for any positive integer n

$$\frac{1}{n} I(\boldsymbol{X}_1^n; \boldsymbol{Y}_1^n, \boldsymbol{S}_1^n) \le \frac{1}{n} I(\boldsymbol{X}_1^n; \boldsymbol{Y}_1^n) + \frac{1}{n} I(\mathbb{H}_1^n; \boldsymbol{S}_1^n) \tag{42}$$

$$\le I(\boldsymbol{X}_1; \boldsymbol{Y}_1) + I(\mathbb{H}_n; \mathbb{H}_1^{n-1}) + I(\boldsymbol{Z}'_n; \boldsymbol{Z}'^{n-1}_1) + \frac{1}{n} I(\mathbb{H}_1^n; \boldsymbol{S}_1^n). \tag{43}$$

Proof: See Appendix IV. □

C. The Fading Number

Motivated by Theorem 4.2 we next define the fading number. Henceforth we shall always assume that the additive noise $\{\boldsymbol{Z}_k\}$ is spatially and temporally white Gaussian noise of covariance matrix $\sigma^2 \mathsf{I}$.

Definition 4.6: The fading number $\chi(\{\mathbb{H}_k\}|\{\boldsymbol{S}_k\})$ of a stationary and ergodic matrix valued fading process $\{\mathbb{H}_k\}$ in the presence of receiver side information $\{\boldsymbol{S}_k\}$ is defined as

$$\chi\left(\{\mathbb{H}_k\}|\{\boldsymbol{S}_k\}\right) = \varlimsup_{\mathcal{E}_s \uparrow \infty} \left\{ C(\mathcal{E}_s) - \log\log \frac{\mathcal{E}_s}{\sigma^2} \right\}. \tag{44}$$

Thus, whenever χ is finite and the limit in (44) exists

$$C(\mathcal{E}_s) = \log\left(1 + e^{\chi} \log\left(1 + \frac{\mathcal{E}_s}{\sigma^2}\right)\right) + o(1) \tag{45}$$

where the $o(1)$ term tends to zero as $\mathcal{E}_s$ tends to infinity. Note that as in (45) and hereafter we omit the argument of χ when it is clear from the context.

The fading number is thus the second term in the high-SNR expansion of channel capacity. Since an exact expression for channel capacity seems intractable, the approximation (44) may be useful for the understanding of the behavior of channel capacity at high SNR.

The fading number serves, however, an additional purpose. The design of communication systems that operate in the region where capacity grows only double-logarithmically in the SNR is extremely power inefficient. Thus, one would expect that system designers will try to avoid this region and design the systems for lower rates (e.g., by using more bandwidth). The fading number may give an indication of roughly how high need the rate be before one enters this high-SNR region. At rates that significantly[2] exceed the fading number, one should expect to square the SNR for every additional bit per channel use. In this sense, the fading number can be viewed as an indication of the practical limiting rate for power-efficient communication over the channel.

The following somewhat unintuitive observation is a consequence of the behavior of the $\log\log(\cdot)$ function under scaling

$$\lim_{\mathrm{SNR}\uparrow\infty} \{\log\log(\alpha \mathrm{SNR}) - \log\log \mathrm{SNR}\} = 0, \qquad 0 < \alpha \in \mathbb{R}. \tag{46}$$

It may simplify the computation of the fading number, especially for multiple-input single-output (MISO) and single-input multiple-output (SIMO) systems where the fading is spatially correlated.

Lemma 4.7: Consider a stationary and ergodic fading process $\{\mathbb{H}_k\}$ with n_{T} transmit antennas and n_{R} receive antennas. Let the $n_{\mathrm{T}} \times n_{\mathrm{T}}$ deterministic matrix F and the $n_{\mathrm{R}} \times n_{\mathrm{R}}$ deterministic matrix G be both nonsingular. Then

$$\chi\left(\{\mathsf{G}\mathbb{H}_k\mathsf{F}\}\right) = \chi\left(\{\mathbb{H}_k\}\right). \tag{47}$$

Proof: The proof of this lemma is given in Appendix V. It is based on the following intuitive ideas. The first is that the channel of fading $\mathbb{H}_k\mathsf{F}$ can be mimicked on the channel of fading $\mathbb{H}_k$ by replacing the input $\boldsymbol{X}$ with the input $\mathsf{F}\boldsymbol{X}$. In doing so, we might be boosting the input power and thus possibly violating the input constraint, but we note that the power boost is at most multiplicative (by $\|\mathsf{F}\|^2$) and is thus insignificant on a double-logarithmic scale. Similarly, the channel of fading $\mathbb{H}_k$ can be mimicked on the channel of fading $\mathbb{H}_k\mathsf{F}$ by multiplying the input $\boldsymbol{X}$ by F^{-1}—again, at a power boost that is at most multiplicative (by $\|\mathsf{F}^{-1}\|^2$).

The invariance with respect to multiplication on the left by G can be argued in a similar way by post-multiplying the channel output. This causes noise coloring and noise boosting, but this phenomenon can be shown to be insignificant on a $\log\log$ scale. □

D. On the Fading Number for Memoryless Fading

1) Trading Additive Noise for Input Constraints: The following theorem gives an equivalent expression for the fading number of memoryless fading channels. In this expression, the additive Gaussian noise is not present, but its place is taken by an additional constraint on the input, namely, that all inputs must be bounded away from zero.

Theorem 4.8: Consider the general memoryless fading channel (20) of fading matrix $\mathbb{H}$ and assume $0 < \mathsf{E}[\|\mathbb{H}\|_{\mathrm{F}}^2] < \infty$ and $h(\mathbb{H}) > -\infty$. Then the channel fading number $\chi(\mathbb{H})$ is given by

$$\chi(\mathbb{H}) = \varlimsup_{\tilde{\mathcal{E}}_s \uparrow \infty} \left\{ \sup_{\substack{\|\tilde{\boldsymbol{X}}\|^2 \ge \mathcal{E}_0 \\ \mathsf{E}[\|\tilde{\boldsymbol{X}}\|^2] \le \tilde{\mathcal{E}}_s}} I\left(\tilde{\boldsymbol{X}}; \mathbb{H}\tilde{\boldsymbol{X}}\right) - \log\log \frac{\tilde{\mathcal{E}}_s}{\mathcal{E}_0} \right\} \tag{48}$$

where $\mathcal{E}_0$ denotes any fixed nonzero energy, e.g., one unit of energy.[3]

[2]One should remember that for some channels (e.g., the i.i.d. Rayleigh fading channel—see (85) ahead), the fading number may be negative. Since zero bits can always be transmitted with zero power, we use the term "significantly exceed" rather than simply "exceed."

[3]The symbol $\mathcal{E}_0$ can be replaced everywhere with 1, but we have chosen not to do so in order to better keep track of units.

Moreover, the fading number χ can be achieved by input distributions $Q_{\mathcal{E}_s}$ that are bounded away from the origin in the sense that

$$Q_{\mathcal{E}_s}(\|\boldsymbol{X}\| \geq x_{\min}) = 1 \tag{49}$$

where

$$\lim_{\mathcal{E}_s \uparrow \infty} x_{\min} = \infty. \tag{50}$$

This theorem is proved in two steps. In the first step, we show that the RHS of (48) is a lower bound to $\chi(\mathbb{H})$, and in the second step we show that it is also an upper bound. The first step is the easier one. It is an immediate consequence of the following lemma.

Lemma 4.9: Let the random vector $\boldsymbol{X}$ take value in $\mathbb{C}^{n_T}$ and satisfy

$$\Pr\left(\|\boldsymbol{X}\|^2 \geq x_{\min}^2\right) = 1 \tag{51}$$

for some $x_{\min} > 0$. Let $\mathbb{H}$ be a random $n_R \times n_T$ matrix satisfying $h(\mathbb{H}) > -\infty$ and $\mathsf{E}[\|\mathbb{H}\|_F^2] < \infty$. Let $\boldsymbol{Z} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \sigma^2 \mathsf{I})$ and assume that $\boldsymbol{X}$, $\mathbb{H}$, and $\boldsymbol{Z}$ are independent. Then

$$I(\boldsymbol{X}; \mathbb{H}\boldsymbol{X} + \boldsymbol{Z}) \geq I(\boldsymbol{X}; \mathbb{H}\boldsymbol{X}) - \sup_{\|\hat{\boldsymbol{x}}\|=1} \left\{ h\left(\mathbb{H}\hat{\boldsymbol{x}} + \frac{\boldsymbol{Z}}{x_{\min}}\right) - h(\mathbb{H}\hat{\boldsymbol{x}}) \right\}. \tag{52}$$

Consequently, for any fixed positive energy $\mathcal{E}_0$

$$\sup_{\mathsf{E}[\|\boldsymbol{X}\|^2] \leq \mathcal{E}_s} \{I(\boldsymbol{X}; \mathbb{H}\boldsymbol{X} + \boldsymbol{Z})\} \geq \sup_{\substack{\|\tilde{\boldsymbol{X}}\|^2 \geq \mathcal{E}_0 \\ \mathsf{E}[\|\tilde{\boldsymbol{X}}\|^2] \leq \mathcal{E}_s/\rho}} \left\{ I\left(\tilde{\boldsymbol{X}}; \mathbb{H}\tilde{\boldsymbol{X}}\right) \right\} - \sup_{\|\hat{\boldsymbol{x}}\|=1} \left\{ h\left(\mathbb{H}\hat{\boldsymbol{x}} + \frac{\boldsymbol{Z}}{\sqrt{\rho \mathcal{E}_0}}\right) - h(\mathbb{H}\hat{\boldsymbol{x}}) \right\}, \qquad \rho > 0 \tag{53}$$

and

$$\chi(\mathbb{H}) \geq \varlimsup_{\tilde{\mathcal{E}}_s \uparrow \infty} \left\{ \sup_{\substack{\|\tilde{\boldsymbol{X}}\|^2 \geq \mathcal{E}_0 \\ \mathsf{E}[\|\tilde{\boldsymbol{X}}\|^2] \leq \tilde{\mathcal{E}}_s}} \left\{ I\left(\tilde{\boldsymbol{X}}; \mathbb{H}\tilde{\boldsymbol{X}}\right) \right\} - \log\log \frac{\tilde{\mathcal{E}}_s}{\mathcal{E}_0} \right\}. \tag{54}$$

Proof: Inequality (52) follows from the basic properties of differential entropy as follows:

$$\begin{aligned} I(\boldsymbol{X}; \mathbb{H}\boldsymbol{X} + \boldsymbol{Z}) &= h(\mathbb{H}\boldsymbol{X} + \boldsymbol{Z}) - h(\mathbb{H}\boldsymbol{X} + \boldsymbol{Z}|\boldsymbol{X}) \\ &\geq h(\mathbb{H}\boldsymbol{X}) - h(\mathbb{H}\boldsymbol{X} + \boldsymbol{Z}|\boldsymbol{X}) \\ &= I(\boldsymbol{X}; \mathbb{H}\boldsymbol{X}) - (h(\mathbb{H}\boldsymbol{X} + \boldsymbol{Z}|\boldsymbol{X}) - h(\mathbb{H}\boldsymbol{X}|\boldsymbol{X})) \\ &\geq I(\boldsymbol{X}; \mathbb{H}\boldsymbol{X}) - \sup_{\|\boldsymbol{x}\| \geq x_{\min}} \{h(\mathbb{H}\boldsymbol{X} + \boldsymbol{Z}|\boldsymbol{X} = \boldsymbol{x}) - h(\mathbb{H}\boldsymbol{X}|\boldsymbol{X} = \boldsymbol{x})\} \\ &= I(\boldsymbol{X}; \mathbb{H}\boldsymbol{X}) - \sup_{\|\hat{\boldsymbol{x}}\|=1} \left\{ h\left(\mathbb{H}\hat{\boldsymbol{x}} + \frac{\boldsymbol{Z}}{x_{\min}}\right) - h(\mathbb{H}\hat{\boldsymbol{x}}) \right\}, \\ &\qquad \Pr\left(\|\boldsymbol{X}\|^2 \geq x_{\min}^2\right) = 1. \end{aligned}$$

Inequality (53) follows from (52) by limiting the supremum of $I(\boldsymbol{X}; \mathbb{H}\boldsymbol{X} + \boldsymbol{Z})$ to random vectors $\boldsymbol{X}$ that are of the form $\boldsymbol{X} = \sqrt{\rho} \cdot \tilde{\boldsymbol{X}}$ for some $\tilde{\boldsymbol{X}}$ satisfying $\Pr(\|\tilde{\boldsymbol{X}}\|^2 \geq \mathcal{E}_0) = 1$ and $\mathsf{E}[\|\tilde{\boldsymbol{X}}\|^2] = \mathcal{E}_s/\rho$ as follows:

$$\begin{aligned} \sup_{\mathsf{E}[\|\boldsymbol{X}\|^2] \leq \mathcal{E}_s} I(\boldsymbol{X}; \mathbb{H}\boldsymbol{X} + \boldsymbol{Z}) &\geq \sup_{\substack{\|\tilde{\boldsymbol{X}}\|^2 \geq \mathcal{E}_0 \\ \mathsf{E}[\|\tilde{\boldsymbol{X}}\|^2] \leq \mathcal{E}_s/\rho}} I\left(\sqrt{\rho} \cdot \tilde{\boldsymbol{X}}; \sqrt{\rho} \cdot \mathbb{H}\tilde{\boldsymbol{X}} + \boldsymbol{Z}\right) && (55) \\ &\geq \sup_{\substack{\|\tilde{\boldsymbol{X}}\|^2 \geq \mathcal{E}_0 \\ \mathsf{E}[\|\tilde{\boldsymbol{X}}\|^2] \leq \mathcal{E}_s/\rho}} I\left(\tilde{\boldsymbol{X}}; \mathbb{H}\tilde{\boldsymbol{X}}\right) - \sup_{\|\hat{\boldsymbol{x}}\|=1} \left\{ h\left(\mathbb{H}\hat{\boldsymbol{x}} + \frac{\boldsymbol{Z}}{\sqrt{\rho \mathcal{E}_0}}\right) - h(\mathbb{H}\hat{\boldsymbol{x}}) \right\} && (56) \end{aligned}$$

where the second inequality follows by (52) because vectors $\boldsymbol{X}$ of this form satisfy $\Pr(\|\boldsymbol{X}\|^2 \geq \rho \mathcal{E}_0) = 1$.

To prove (54), we shall use (53) with ρ growing with $\mathcal{E}_s$ in a controlled way. Defining $\tilde{\mathcal{E}}_s = \mathcal{E}_s/\rho$ we have

$$\begin{aligned} \chi(\mathbb{H}) &= \varlimsup_{\mathcal{E}_s \to \infty} \left\{ \sup_{\mathsf{E}[\|\boldsymbol{X}\|^2] \leq \mathcal{E}_s} I(\boldsymbol{X}; \mathbb{H}\boldsymbol{X} + \boldsymbol{Z}) - \log\log \frac{\mathcal{E}_s}{\sigma^2} \right\} \\ &\geq \varlimsup_{\tilde{\mathcal{E}}_s \to \infty} \left\{ \sup_{\substack{\|\tilde{\boldsymbol{X}}\|^2 \geq \mathcal{E}_0 \\ \mathsf{E}[\|\tilde{\boldsymbol{X}}\|^2] \leq \tilde{\mathcal{E}}_s}} I\left(\tilde{\boldsymbol{X}}; \mathbb{H}\tilde{\boldsymbol{X}}\right) - \log\log \frac{\tilde{\mathcal{E}}_s}{\mathcal{E}_0} \right\} \\ &\quad - \varlimsup_{\mathcal{E}_s \to \infty} \left\{ \log\log \frac{\mathcal{E}_s}{\sigma^2} - \log\log \frac{\tilde{\mathcal{E}}_s}{\mathcal{E}_0} \right\} \\ &\quad - \varlimsup_{\mathcal{E}_s \to \infty} \sup_{\|\hat{\boldsymbol{x}}\|=1} \left\{ h\left(\mathbb{H}\hat{\boldsymbol{x}} + \frac{\boldsymbol{Z}}{\sqrt{\rho \mathcal{E}_0}}\right) - h(\mathbb{H}\hat{\boldsymbol{x}}) \right\}. \end{aligned}$$

To prove (54) it thus follows that it suffices to require that ρ tend to infinity so that by Lemma 6.11 the third term on the RHS of the inequality will tend to zero, and to additionally require that $\log \rho / \log \mathcal{E}_s$ tend to zero, so that the second term on the RHS of the inequality will tend to zero. An example of a choice that meets these two requirements is

$$\rho = \log \frac{\mathcal{E}_s}{\mathcal{E}_0}. \tag{57}$$

□

We next continue with the second step in the proof of Theorem 4.8. In this step, we show that the RHS of (48) is an upper bound to $\chi(\mathbb{H})$. We show that by trying to make light of the constraint $\|\tilde{\boldsymbol{X}}\|^2 \geq \mathcal{E}_0$. More specifically, we shall show that even in the presence of noise—let alone in its absence—this constraint does not preclude one from achieving the fading number.

We thus next show that even in the presence of additive temporally and spatially white Gaussian noise, the fading number $\chi(\mathbb{H})$ can be achieved using input distributions that satisfy the constraint $\|\tilde{\boldsymbol{X}}\|^2 \geq \mathcal{E}_0$. The technique we use to prove this claim may be of independent interest, and we, therefore, present this proof in a somewhat general setting. The proof hinges on the fact that the capacity of our channel can be achieved by input distributions that assign to any fixed compact set a probability that tends to zero as the SNR tends to infinity. This property turns out to hold for many cost constrained channels of interests, and we therefore define it in a fairly general setting. We need, however, the following preliminary standard definition.

Definition 4.10: Given a channel $W(\cdot|\cdot)$ over the input alphabet $\mathcal{X}$ and the output alphabet $\mathcal{Y}$ and given some nonnegative cost function $g\colon \mathcal{X} \to \mathbb{R}^+$, we define the capacity–cost function $C\colon \mathbb{R}^+ \to \mathbb{R}^+$ by

$$C(\Upsilon) = \sup_{\mathsf{E}_Q[g(X)] \le \Upsilon} I(Q; W), \qquad \Upsilon \ge 0. \tag{58}$$

We say that $C(\Upsilon)$ is achieved if the supremum in (58) is achieved.

We are now ready for the definition of capacity-achieving input distributions that escape to infinity. For an intuitive understanding of the following definition and some of its consequences, it is best to focus on the example where the channel inputs are vectors in Euclidean space and where the cost function $g(\cdot)$ is the squared Euclidean norm.

Definition 4.11: Let $C(\cdot)$ denote the capacity–cost function of a channel $W(\cdot|\cdot)$ over the input alphabet $\mathcal{X}$ and the output alphabet $\mathcal{Y}$ with the nonnegative cost function $g\colon \mathcal{X} \to \mathbb{R}^+$. Assume $C(\Upsilon) < \infty$, for any fixed Υ. We shall say that the capacity of this channel can be achieved by *input distributions that escape to infinity*, if for any $\Upsilon_0 \ge 0$ there exist input distributions $\{Q_\Upsilon\}_{\Upsilon \ge 0}$ satisfying $\mathsf{E}_{Q_\Upsilon}[g(X)] \le \Upsilon$ such that

$$\lim_{\Upsilon \uparrow \infty} \{C(\Upsilon) - I(Q_\Upsilon; W)\} = 0 \tag{59}$$

and

$$\lim_{\Upsilon \uparrow \infty} Q_\Upsilon\left(g(X) \le \Upsilon_0\right) = 0. \tag{60}$$

Intuition suggests that if capacity can be achieved using input distributions that assign an ever decreasing probability to a set $\mathcal{K}$, then at high SNR the capacity should not suffer appreciably from constraining the inputs to lie outside $\mathcal{K}$ almost surely. This intuition is made precise in the following theorem.

Theorem 4.12: Consider a channel of law $W(\cdot|\cdot)$ over the input alphabet $\mathcal{X}$ and the output alphabet $\mathcal{Y}$. Let $g\colon \mathcal{X} \to \mathbb{R}^+$ be some nonnegative cost function, and let $C(\cdot)$ denote the capacity–cost function associated with $W(\cdot|\cdot)$ and g. Assume $C(\Upsilon) < \infty$, for any fixed Υ. Fix some $\Upsilon_0 \ge 0$ and let $\mathcal{K} = \{x \in \mathcal{X}\colon g(x) \le \Upsilon_0\}$. Let $C_c(\Upsilon)$ denote the capacity–cost function when the inputs are additionally constrained to lie outside $\mathcal{K}$. Let $d(\Upsilon)$ be some mapping $d\colon \mathbb{R}^+ \to \mathbb{R}^+$ satisfying

$$\varlimsup_{\epsilon \downarrow 0} \varlimsup_{\Upsilon \uparrow \infty} |d((1+\epsilon)\Upsilon) - d(\Upsilon)| = 0 \tag{61}$$

for example[4]

$$d(\Upsilon) = \log(1 + \log(1+\Upsilon)), \quad \text{or} \quad d(\Upsilon) = \frac{1}{2}\log(1+\Upsilon). \tag{62}$$

Assume that, as the cost Υ tends to infinity, capacity-achieving input distributions escape to infinity. Then

$$\varliminf_{\Upsilon \uparrow \infty} \{C_c(\Upsilon) - d(\Upsilon)\} = \varlimsup_{\Upsilon \uparrow \infty} \{C(\Upsilon) - d(\Upsilon)\}. \tag{63}$$

Proof: Placing additional input constraints cannot increase capacity. Hence,

$$C_c(\Upsilon) \le C(\Upsilon)$$

so that the left-hand side (LHS) of (63) cannot exceed its RHS. We now proceed to prove the reverse inequality.

Let $\{\Upsilon_n\} \uparrow \infty$ and $\{Q_n\} \subset \mathcal{P}(\mathcal{X})$ satisfy

$$\mathsf{E}_{Q_n}[g(X)] \le \Upsilon_n \tag{64}$$

$$\lim_{n\to\infty} \{I(Q_n; W) - d(\Upsilon_n)\} = \varlimsup_{\Upsilon \uparrow \infty} \{C(\Upsilon) - d(\Upsilon)\} \tag{65}$$

and

$$\lim_{n\to\infty} Q_n(\mathcal{K}) = 0. \tag{66}$$

By (66) it follows that for all sufficiently large n the probability $Q_n(\mathcal{K}^c)$ of the set-complement $\mathcal{K}^c$ of $\mathcal{K}$ is strictly larger than zero and we can, therefore, define the conditional law $\tilde{Q}_n$ so that for any Borel set $\mathcal{A} \subset \mathcal{X}$ and any sufficiently large n

$$\tilde{Q}_n(\mathcal{A}) = \frac{Q_n(\mathcal{A} \cap \mathcal{K}^c)}{Q_n(\mathcal{K}^c)}. \tag{67}$$

Thus, under the prior Q_n, the probability measure $\tilde{Q}_n$ corresponds to the *a posteriori* distribution on the input conditional on $x \notin \mathcal{K}$.

Note that by the nonnegativity of the cost function it follows that the cost associated with $\tilde{Q}_n$ is not appreciably larger than the one associated with Q_n. Indeed, if we define

$$\tilde{\Upsilon}_n = \frac{\Upsilon_n}{Q_n(\mathcal{K}^c)} \tag{68}$$

then the cost associated with $\tilde{Q}_n$ satisfies

$$\begin{aligned} \mathsf{E}_{\tilde{Q}_n}[g(X)] &\le \frac{1}{Q_n(\mathcal{K}^c)} \mathsf{E}_{Q_n}[g(X)] \\ &\le \frac{\Upsilon_n}{Q_n(\mathcal{K}^c)} \\ &= \tilde{\Upsilon}_n. \end{aligned} \tag{69}$$

Let X be distributed according to Q_n, and let the binary-valued random variable E be defined by

$$E = \begin{cases} 0, & \text{if } X \in \mathcal{K} \\ 1, & \text{if } X \notin \mathcal{K}. \end{cases} \tag{70}$$

Note that the probability that E takes on the value 0 is $Q_n(\mathcal{K})$, which by (66) tends to zero with n. We now have

$$\begin{aligned} I(Q_n; W) &= I(X; Y) \\ &= I(X, E; Y) \\ &= I(E; Y) + I(X; Y|E) \\ &\le H(E) + Q_n(\mathcal{K}) I(X; Y|E=0) \\ &\quad + (1 - Q_n(\mathcal{K})) I(X; Y|E=1) \\ &= H_\mathrm{b}(Q_n(\mathcal{K})) + Q_n(\mathcal{K}) I(X; Y|E=0) \\ &\quad + (1 - Q_n(\mathcal{K})) I\left(\tilde{Q}_n; W\right) \\ &\le H_\mathrm{b}(Q_n(\mathcal{K})) + Q_n(\mathcal{K}) I(X; Y|E=0) \\ &\quad + C_c\left(\tilde{\Upsilon}_n\right) \end{aligned} \tag{71}$$

where $H_\mathrm{b}(\cdot)$ denotes the binary entropy function, i.e.,

$$H_\mathrm{b}(\xi) = \xi \log\frac{1}{\xi} + (1-\xi)\log\frac{1}{1-\xi}, \qquad 0 < \xi < 1. \tag{72}$$

By subtracting $d(\Upsilon_n)$ from both sides of (71) we obtain

$$\begin{aligned} &I(Q_n; W) - d(\Upsilon_n) \\ &\quad \le H_\mathrm{b}(Q_n(\mathcal{K})) + Q_n(\mathcal{K}) I(X; Y|E=0) \\ &\qquad + \left(d\left(\tilde{\Upsilon}_n\right) - d(\Upsilon_n)\right) + \left(C_c\left(\tilde{\Upsilon}_n\right) - d\left(\tilde{\Upsilon}_n\right)\right). \end{aligned} \tag{73}$$

We now consider the limiting behavior (as $n \to \infty$) of both sides of the inequality (73). Beginning with the LHS, we note that, by (65), it tends to $\varlimsup\{C(\Upsilon) - d(\Upsilon)\}$.

[4]In this paper, we shall only be interested in the case where $d(\Upsilon) = \log(1 + \log(1+\Upsilon))$ but the other example can be useful in other applications. See, for example, [4], [5].

We now consider the RHS of (73). By inspecting the behavior of the binary entropy function $H_b(\cdot)$ about zero, it follows from (66) that the first term tends to zero. Similarly, since $I(X; Y|E = 0)$ is bounded in Υ (because the capacity under the sole constraint that the input must lie in $\mathcal{K}$ is bounded by $C(\Upsilon_0) < \infty$), it follows again from (66) that the second term also tends to zero. The third term converges to zero by (61), (68), and (66). The only remaining term is the last one, which gives us the inequality

$$\overline{\lim_{\Upsilon \uparrow \infty}} \{C(\Upsilon) - d(\Upsilon)\} \leq \overline{\lim_{\tilde{\Upsilon} \uparrow \infty}} \left\{C_c\left(\tilde{\Upsilon}\right) - d\left(\tilde{\Upsilon}\right)\right\} \quad (74)$$

thus concluding the proof of the theorem.

It is interesting to note that an expansion dual to (71) is also useful in the study of the redundancy–capacity theorem of universal source coding. See (without costs) [10, Proof of Theorem 1]. □

For our multiple-antenna fading channel, the capacity cost function is clearly finite (if the noise variance is positive). Indeed, this is even the case when the receiver knows the realization of the fading matrix. Thus, by Theorem 4.12, if we could show that the capacity of our channel is attained by input distributions that escape to infinity, we would also deduce that—even in the presence of noise—the constraint $\|\tilde{\boldsymbol{X}}\|^2 \geq \mathcal{E}_0$ does not preclude one from achieving the fading number.

The proof of Theorem 4.8 will thus be concluded once we show that for our fading channel, capacity can be achieved using input distributions that escape to infinity. We next derive some general conditions that guarantee this property. Again, since such conditions may be useful in other contexts, we state the conditions in fairly general terms.

Theorem 4.13: Let the input alphabet $\mathcal{X}$ and the output alphabet $\mathcal{Y}$ of a channel $W(\cdot|\cdot)$ be separable metric spaces, and assume that for any Borel set $\mathcal{B} \subset \mathcal{Y}$, the mapping $x \mapsto W(\mathcal{B}|x)$ from $\mathcal{X}$ to [0, 1] is Borel measurable. Let the nonnegative cost function $g: \mathcal{X} \to \mathbb{R}^+$ be measurable, and let $C(\cdot)$ be the capacity–cost function for the channel $W(\cdot|\cdot)$ and the cost function $g(\cdot)$. Assume:

- For any fixed cost Υ the constrained capacity is finite, but as the cost tends to infinity, the capacity increases to infinity sublinearly

$$\lim_{\Upsilon \uparrow \infty} C(\Upsilon) = \infty, \qquad \lim_{\Upsilon \uparrow \infty} \frac{C(\Upsilon)}{\Upsilon} = 0. \quad (75)$$

- For any two input distributions of finite cost the directional derivative of the mutual information exists and is given by[5]

$$\lim_{\lambda \downarrow 0} \frac{1}{\lambda}\Big(I(\lambda Q_2 + (1-\lambda)Q_1; W) - I(Q_1; W)\Big) = \int D(W(\cdot|x)\|(Q_1 W)(\cdot))\,\mathrm{d}Q_2(x) - I(Q_1; W),$$
$$\mathsf{E}_{Q_1}[g(X)],\ \mathsf{E}_{Q_2}[g(X)] < \infty. \quad (76)$$

Then, capacity-achieving input distributions escape to infinity.

Proof: See Appendix VI. □

[5]It suffices that this hold for all probability measures Q_1 for which $\mathsf{E}_{Q_1}[g(X)]$ is sufficiently large.

To conclude the proof of Theorem 4.8 it now only remains to check that the assumptions of Theorem 4.8 imply that the fading channel satisfies the conditions of Theorem 4.13.

The condition $\mathsf{E}[\|\mathbb{H}\|_F^2] > 0$ implies that channel capacity is unbounded. Indeed, this condition guarantees that by spacing any finite number of symbols sufficiently apart, we can achieve an arbitrarily small uncoded probability of error. See [11, Section IV.B] for the details.

The condition $\mathsf{E}[\|\mathbb{H}\|_F^2] < \infty$ guarantees that the capacity can only grow sublinearly in the power. Indeed, the sublinear growth is guaranteed even if the receiver has knowledge of the fading matrix, because this condition guarantees that the power in $\mathbb{H}\boldsymbol{X}$ grows at most linearly in $\mathcal{E}_s$, so that the presence of the additive noise guarantees that capacity can grow at most logarithmically in $\mathcal{E}_s$. (The additional condition $h(\mathbb{H}) > -\infty$ guarantees, of course, an even slower increase in capacity, namely, a double-logarithmic one.)

Finally, the technical condition regarding the directional derivatives (76) can be verified as in [11, Appendix II.B (63)].

We will now briefly summarize the proof of Theorem 4.8:

Proof: We proved Theorem 4.8 in two steps. In the first, see Lemma 4.9, we proved that by restricting the minimum norm that the channel inputs may have, we can mimic the limiting behavior of a channel without noise. That is,

$$\chi(\mathbb{H}) \geq \overline{\lim_{\tilde{\mathcal{E}}_s \uparrow \infty}} \left\{ \sup_{\substack{\|\tilde{\boldsymbol{X}}\|^2 \geq \mathcal{E}_0 \\ \mathsf{E}[\|\tilde{\boldsymbol{X}}\|^2] \leq \tilde{\mathcal{E}}_s}} I\left(\tilde{\boldsymbol{X}}; \mathbb{H}\tilde{\boldsymbol{X}}\right) - \log\log\frac{\tilde{\mathcal{E}}_s}{\mathcal{E}_0} \right\}. \quad (77)$$

In the second step we showed that, even in the presence of noise—let alone in its absence—the fading number can be achieved with inputs that are additionally constrained to lie outside a fixed energy ball. This was shown by demonstrating that the capacity of the fading channel can be achieved by input distributions that escape to infinity, and by showing that for such channels, the high-SNR capacity asymptotics can be achieved even subject to an additional minimum cost constraint. The former claim was proved by proving general conditions for capacity-achieving input distributions to escape to infinity (see Theorem 4.13) and by verifying that the fading channel satisfies these conditions (see the discussion following the proof of Theorem 4.13). The latter claim about the capacity asymptotics of channel with capacity-achieving input distributions escaping to infinity was proved in Theorem 4.12. □

2) An Upper Bound on χ for Memoryless Fading: Having established Theorem 4.8 we can now upper-bound $I(\tilde{\boldsymbol{X}}; \mathbb{H}\tilde{\boldsymbol{X}})$ in (48) using the bound (25). In this way, we can obtain the following upper bound on the fading number of memoryless fading channels.

Theorem 4.14: Consider a memoryless fading channel (20) of fading matrix $\mathbb{H}$ satisfying $0 < \mathsf{E}[\|\mathbb{H}\|_F^2] < \infty$ and $h(\mathbb{H}) > -\infty$. Then the fading number $\chi(\mathbb{H})$ is upper-bounded by χ_u, where

$$\chi_u = \log \pi^{n_R} - \log\Gamma(n_R) + \inf_{\mathsf{A}} \sup_{\|\hat{\boldsymbol{x}}\|=1} \left\{ n_R \mathsf{E}\left[\log\|\mathsf{A}\mathbb{H}\hat{\boldsymbol{x}}\|^2\right] - h(\mathsf{A}\mathbb{H}\hat{\boldsymbol{x}}) \right\} \quad (78)$$

and where the infimum is over all nonsingular $n_{\rm R} \times n_{\rm R}$ complex matrices A.

Proof: See Appendix VII. □

Note 4.15: Using Jensen's inequality applied to the concave function $\log(\cdot)$ one can further upper-bound (78) by

$$\chi_{\rm u} \le \log \pi^{n_{\rm R}} - \log \Gamma(n_{\rm R}) + \inf_{\mathsf{A}} \sup_{\|\hat{\boldsymbol{x}}\|=1} \left\{ n_{\rm R} \log \mathsf{E}\left[\|\mathsf{A}\mathbb{H}\hat{\boldsymbol{x}}\|^2\right] - h(\mathsf{A}\mathbb{H}\hat{\boldsymbol{x}}) \right\}. \quad (79)$$

This bound is generally not tight, but it is often much simpler to compute than (78).

3) Memoryless Single-Input Single-Output (SISO) Systems: A lower bound to the fading number can be obtained from Theorem 4.8 by lower-bounding $I(\tilde{\boldsymbol{X}}; \mathbb{H}\tilde{\boldsymbol{X}})$ using specific input distributions. For example, in the SISO case, we can obtain a lower bound from Theorem 4.8 by considering an input $\tilde{X}$ that is circularly symmetric with $\log|\tilde{X}|^2$ being uniformly distributed between $\log \mathcal{E}_0$ and $\log \mathcal{E}_{\rm s}$. It turns out that the resulting lower bound on χ coincides with the upper bound. Thus, for SISO channels we have a complete characterization of the fading number. In fact, we can show that the fading number can be achieved even if the average power constraint $\mathsf{E}[|X|^2] \le \mathcal{E}_{\rm s}$ is replaced with the peak power constraint $|X|^2 \le \mathcal{E}_{\rm s}$.

Theorem 4.16: Consider a SISO memoryless fading channel with a complex fading variable H. Assume that $\mathsf{E}[|H|^2] < \infty$ and $h(H) > -\infty$. Then the limsup in (44) is also a liminf (i.e., the limit exists) and the fading number $\chi(H)$ is given by

$$\chi(H) = \log \pi + \mathsf{E}\left[\log |H|^2\right] - h(H). \quad (80)$$

Moreover, this fading number is achievable by circularly symmetric inputs X whose log magnitude $\log|X|$ is uniformly distributed over the interval $[\log x_{\min}, 1/2 \log \mathcal{E}_{\rm s}]$ for any $x_{\min}(\mathcal{E}_{\rm s})$ satisfying

$$\lim_{\mathcal{E}_{\rm s}\to\infty} x_{\min} = \infty$$

and

$$\lim_{\mathcal{E}_{\rm s}\to\infty} \frac{\log x_{\min}}{\log \mathcal{E}_{\rm s}} = 0.$$

Proof: The fact that the RHS of (80) is an upper bound on $\chi(H)$ follows from Theorem 4.14 by choosing the matrix A in (78) as the 1×1 identity matrix (i.e., the scalar 1).

To derive a lower bound on $\chi(H)$, we use Theorem 4.8 with the choice of $\tilde{X}$ being a circularly symmetric random variable such that

$$\log\left|\tilde{X}\right|^2 \sim \text{Uniform}[\log \mathcal{E}_0, \log \mathcal{E}_{\rm s}]. \quad (81)$$

Indeed, for this choice of $\tilde{X}$ we have

$$\begin{aligned}
I\left(\tilde{X}; H\tilde{X}\right) &= h\left(H\tilde{X}\right) - h\left(H\tilde{X} \,\middle|\, \tilde{X}\right) \\
&= h\left(H\tilde{X}\right) - \mathsf{E}\left[\log|\tilde{X}|^2\right] - h(H) \\
&\ge h\left(H\tilde{X} \,\middle|\, H\right) - \mathsf{E}\left[\log|\tilde{X}|^2\right] - h(H) \\
&= \mathsf{E}\left[\log|H|^2\right] + h\left(\tilde{X}\right) - \mathsf{E}\left[\log|\tilde{X}|^2\right] - h(H) \\
&= \log 2\pi + h\left(\left|\tilde{X}\right|\right) - \mathsf{E}\left[\log\left|\tilde{X}\right|\right] + \mathsf{E}\left[\log|H|^2\right] - h(H) \\
&= \log 2\pi + h\left(\log\left|\tilde{X}\right|\right) + \mathsf{E}\left[\log|H|^2\right] - h(H) \\
&= h\left(\log\left|\tilde{X}\right|^2\right) + \log\pi + \mathsf{E}\left[\log|H|^2\right] - h(H) \quad (82) \\
&= \log\log\frac{\mathcal{E}_{\rm s}}{\mathcal{E}_0} + \log\pi + \mathsf{E}\left[\log|H|^2\right] - h(H). \quad (83)
\end{aligned}$$

Here, the first equality follows from the definition of mutual information; the subsequent equality from the behavior under scaling of the differential entropy of *complex* random variables; the following inequality because conditioning cannot increase differential entropy; the following equality by (319) for the differential entropy of a circularly symmetric random variables; the subsequent equality by relating the differential entropy of a positive random variable to that of its logarithm as in Lemma 6.15; the following equality by the behavior of differential entropy under scaling of a *real* random variable; and the final equality by evaluating the differential entropy using (81).

The nature of the input distributions X that achieve $\chi(H)$ follows from (81) and Lemma 4.9. □

Note 4.17: It is interesting to note that the fading number is achievable by input distributions of a law that does not depend on the fading law. This observation can be useful in the analysis of fading channels with some side information available at the receiver and/or the transmitter; see Proposition 4.23.

Note 4.18: Note also that the achievability of the fading number with the above input distributions demonstrates that for SISO channels, the fading number does not depend on whether average or peak power constraints are imposed.

Corollary 4.19: For Ricean fading, i.e., memoryless SISO Gaussian fading, the fading number is given (in nats) by

$$\chi\left(\mathcal{N}_{\mathbb{C}}(d, 1)\right) = -1 + \log|d|^2 - \text{Ei}(-|d|^2) \quad (84)$$

where $\text{Ei}(\cdot)$ denotes the exponential integral function defined in (211). In the special case where $d = 0$, i.e., Rayleigh fading, the fading number is thus given by

$$\chi\left(\mathcal{N}_{\mathbb{C}}(0, 1)\right) = -1 - \gamma \quad (85)$$

where $\gamma \approx 0.577$ denotes Euler's constant.

Proof: Follows directly from Theorem 4.16 by evaluating the differential entropy of a complex Gaussian random variable and by evaluating the expectation of the logarithm of a noncentral chi-square random variable (209). □

With the aid of Lemma 4.9 we can also obtain an asymptotically tight firm lower bound to the capacity of the Ricean channel.

Corollary 4.20: Let $C(\mathcal{E}_{\rm s}/\sigma^2)$ denote the capacity of a memoryless SISO Ricean fading channel of fading law $\mathcal{N}_{\mathbb{C}}(d, 1)$, average power $\mathcal{E}_{\rm s}$, and additive Gaussian noise variance σ^2. Then

$$C(\mathcal{E}_{\rm s}/\sigma^2) \ge h_{\max}(\mathcal{E}_{\rm s}/\mathcal{E}_{\min}) + \chi\left(\mathcal{N}_{\mathbb{C}}(d, 1)\right) - \log\left(1 + \frac{\sigma^2}{\mathcal{E}_{\min}}\right), \qquad 0 < \mathcal{E}_{\min} < \mathcal{E}_{\rm s} \quad (86)$$

where $h_{\max}(\mathcal{E}_{\rm s}/\mathcal{E}_{\min})$ is the maximum differential entropy a real random variable W can have if it is to satisfy

$$W \geq \log \mathcal{E}_{\min} \quad \text{and} \quad \mathsf{E}[e^W] \leq \mathcal{E}_{\rm s} \tag{87}$$

namely

$$h_{\max}(\mathcal{E}_{\rm s}/\mathcal{E}_{\min}) = \log(-\mathrm{Ei}(-\zeta)) + \zeta \frac{\mathcal{E}_{\rm s}}{\mathcal{E}_{\min}} \tag{88}$$

where ζ is the solution to the equation

$$\frac{e^{-\zeta}}{-\mathrm{Ei}(-\zeta)} = \zeta \frac{\mathcal{E}_{\rm s}}{\mathcal{E}_{\min}}. \tag{89}$$

Note 4.21: A looser but simpler lower bound follows from the bound

$$h_{\max}(\mathcal{E}_{\rm s}/\mathcal{E}_{\min}) \geq \log\log \frac{\mathcal{E}_{\rm s}}{\mathcal{E}_{\min}} \tag{90}$$

which can be verified by considering the differential entropy $h(W)$ of a random variable W that is uniformly distributed over the interval $[\log \mathcal{E}_{\min}, \log \mathcal{E}_{\rm s}]$ and that thus satisfies the constraints (87).

Proof: Fix some $0 < \mathcal{E}_{\min} < \mathcal{E}_{\rm s}$ and consider a circularly symmetric random variable X such that

$$|X|^2 \geq \mathcal{E}_{\min} \tag{91}$$

$$\mathsf{E}[|X|^2] = \mathcal{E}_{\rm s} \tag{92}$$

$$h(\log |X|^2) = h_{\max}(\mathcal{E}_{\rm s}/\mathcal{E}_{\min}). \tag{93}$$

Then

$$\begin{aligned} &I(X; HX + Z) \\ &\quad \geq I(X; HX) - \sup_{|x|^2 = \mathcal{E}_{\min}} \{h(Hx + Z) - h(Hx)\} \\ &\quad = I(X; HX) - \left(\log\left(\pi e(\mathcal{E}_{\min} + \sigma^2)\right) - \log(\pi e \mathcal{E}_{\min})\right) \\ &\quad = I(X; HX) - \log\left(1 + \frac{\sigma^2}{\mathcal{E}_{\min}}\right) \\ &\quad \geq h\left(\log |X|^2\right) + \chi(\mathcal{N}_{\mathbb{C}}(d, 1)) - \log\left(1 + \frac{\sigma^2}{\mathcal{E}_{\min}}\right) \\ &\quad = h_{\max}(\mathcal{E}_{\rm s}/\mathcal{E}_{\min}) + \chi(\mathcal{N}_{\mathbb{C}}(d, 1)) - \log\left(1 + \frac{\sigma^2}{\mathcal{E}_{\min}}\right). \end{aligned}$$

Here, the first inequality follows from Lemma 4.9; the subsequent equality follows by the explicit evaluation of the differential entropy of the Gaussian distribution; the subsequent equality by direct calculation; the following inequality by lower bounding $h(HX)$ by $h(HX|H)$ as in the steps leading to (82); and the final equality by our choice of $\log |X|^2$ as having the max-entropy distribution.

The expression for $h_{\max}(\mathcal{E}_{\rm s}/\mathcal{E}_{\min})$ follows by noting that the density that achieves $h_{\max}(\mathcal{E}_{\rm s}/\mathcal{E}_{\min})$ is of the form [12, Theorem 11.1.1]

$$\frac{1}{c} \cdot e^{-\lambda e^w}, \qquad w \geq \log \mathcal{E}_{\min}. \tag{94}$$

□

Theorem 4.16 gives an exact expression for the RHS of (48) for SISO channels and demonstrates that the limsup is in fact a limit. While the theorem is stated for complex channels, it can also be used to obtain a useful relationship for real channels. This relation will be useful in the analysis of SIMO channels.

Corollary 4.22: Let $|A|$ be a nonnegative real random variable satisfying $\mathsf{E}[|A|^2] < \infty$ and $h(|A|) > -\infty$. Then

$$\lim_{\tilde{\mathcal{E}}_{\rm s} \uparrow \infty} \left\{ \sup_{\substack{|\tilde{X}|^2 \geq \mathcal{E}_0 \\ \mathsf{E}[|\tilde{X}|^2] \leq \tilde{\mathcal{E}}_{\rm s}}} I\left(\left|\tilde{X}\right|; |A| \cdot \left|\tilde{X}\right|\right) - \log\log \frac{\tilde{\mathcal{E}}_{\rm s}}{\mathcal{E}_0} \right\} = \mathsf{E}[\log |A|] - h(|A|) - \log 2. \tag{95}$$

Proof: This follows by Theorem 4.8 and Theorem 4.16 applied to the fading $H = |A| e^{j\Theta}$ where Θ is independent of $|A|$ and uniformly distributed over the interval $[-\pi, \pi)$. For this circularly symmetric law of H no information can be passed via the phase so that the limits in (95) and (48) agree. They are consequently both given by Theorem 4.16 as

$$\log \pi + \mathsf{E}\left[\log |H|^2\right] - h(H) = \mathsf{E}[\log |A|] - h(|A|) - \log 2 \tag{96}$$

where the relationship $h(H) = \log 2\pi + h(|A|) + \mathsf{E}[\log |A|]$ follows from Lemma 6.16, which relates the differential entropy of a complex random variable (in our case H) to the joint differential entropy of its magnitude and phase (in our case $|A|$ and Θ, which are independent); see (319)–(320). □

An alternative proof, which does not require embedding the real channel in a complex one, can be based on choosing inputs of energies with logarithms that are uniformly distributed on the interval $[\log \mathcal{E}_0, \log \mathcal{E}_{\rm s}]$ and then invoking Lemma 6.10.

4) Memoryless SISO Systems With Side-Information: Theorem 4.16 extends to situations where the receiver (but not transmitter) has some side information regarding the realization of the fading. This setting will be explored in greater detail in Theorem 4.41 but we send forward the following simple case, which turns out to be instrumental to the analysis of the more general case.

Proposition 4.23: Let H be some complex random variable satisfying $\mathsf{E}[|H|^2] < \infty$. Assume that the pair (H, S) is independent of the additive noise $Z \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$ and that the joint law of (H, S, Z) does not depend on the channel input X. Further assume $h(H|S) > -\infty$. Then

$$\lim_{\mathcal{E}_{\rm s}/\sigma^2 \uparrow \infty} \left\{ \sup_{\mathsf{E}[|X|^2] \leq \mathcal{E}_{\rm s}} I(X; HX + Z, S) - \log\log \frac{\mathcal{E}_{\rm s}}{\sigma^2} \right\} = \log \pi + \mathsf{E}\left[\log |H|^2\right] - h(H|S) \tag{97}$$

i.e.,

$$\chi(H|S) = \log \pi + \mathsf{E}\left[\log |H|^2\right] - h(H|S). \tag{98}$$

This fading number is achievable by input distributions of the form given in Theorem 4.16.

Note 4.24: Under fairly general conditions, (98) continues to be valid even if the state is also known to the transmitter. This is, for example, the case if S takes value in a finite set.

Note 4.25: It is interesting to compare the fading numbers in the presence of receiver side information (98) and in its absence

(80). The side information increases the fading number by the mutual information $I(H;\, S)$.

Proof: The proof relies heavily on the fact that the fading number of a SISO system can be achieved by input distributions that do not depend on the fading law. The proof of this proposition is given in Appendix VIII. Here we merely give a plausibility argument.

Choose X to be distributed according to the law specified in Theorem 4.16. Then, for any realization $S = s$, we have by Theorem 4.16 (applied to fading law $H|S = s$) that at high SNR

$$\begin{aligned} I(X;\, HX + Z, S = s) \\ &= I(X;\, HX + Z|S = s) \qquad (99)\\ &\approx \log\log\frac{\mathcal{E}_s}{\sigma^2} + \log\pi + \mathsf{E}\left[\log|H|^2|S = s\right] - h(H|S = s) \qquad (100)\end{aligned}$$

from which the result follows by taking expectations with respect to S.

The technical problem with this argument is in interchanging the order of taking the expectation with respect to S with the taking of the limit as $\mathcal{E}_s/\sigma^2$ tends to infinity. □

Corollary 4.26: Assume that $H \sim \mathcal{N}_{\mathbb{C}}(d,\, 1)$ and that the pair $(H,\, S)$ are jointly Gaussian and jointly circularly symmetric. Let $\epsilon^2_{\text{est}} > 0$ denote the mean squared error in estimation H from S. Then, the fading number in the presence at the receiver of the side information S is given by

$$\chi(H|S) = -1 + \log\frac{|d|^2}{\epsilon^2_{\text{est}}} - \text{Ei}(-|d|^2). \qquad (101)$$

Proof: Follows directly from Proposition 4.23 and Corollary 4.19 by noting that

$$I(H;\, S) = \log\frac{1}{\epsilon^2_{\text{est}}}. \qquad (102)$$

□

5) Memoryless MISO Systems and Beam Forming: A different extension of Theorem 4.16 is to MISO fading channels. With the aid of Theorem 4.16 and the upper bound of Theorem 4.14 we can now also compute the fading number of memoryless MISO fading channels. As a by-product we shall infer that—in the sense that it allows one to achieve the fading number—beam forming is asymptotically optimal. By beam forming we refer here to choosing some fixed deterministic unit vector $\hat{\boldsymbol{x}}$ and limiting the inputs to vectors in $\mathbb{C}^{n_T}$ that are colinear with it. Such an approach can greatly reduce the complexity of the code/decoder.

Theorem 4.27: Consider a MISO memoryless fading channel with row fading vector $\boldsymbol{H}^{\mathsf{T}}$, where $\boldsymbol{H}$ is random column vector in $\mathbb{C}^{n_T}$ satisfying $0 < \mathsf{E}[\|\boldsymbol{H}\|^2] < \infty$ and $h(\boldsymbol{H}) > -\infty$. Then the limsup in (44) is also a liminf (i.e., the limit exists) and the fading number $\chi(\boldsymbol{H}^{\mathsf{T}})$ is given by

$$\chi(\boldsymbol{H}^{\mathsf{T}}) = \sup_{\|\hat{\boldsymbol{x}}\|=1}\left\{\log\pi + \mathsf{E}\left[\log\left|\boldsymbol{H}^{\mathsf{T}}\hat{\boldsymbol{x}}\right|^2\right] - h(\boldsymbol{H}^{\mathsf{T}}\hat{\boldsymbol{x}})\right\}. \qquad (103)$$

Moreover, this fading number is achievable by inputs that can be expressed as the product of a deterministic unit vector in $\mathbb{C}^{n_T}$ and a circularly symmetric scalar complex random variable of a law specified in Theorem 4.16.

Proof: If the channel input is of the form $X \cdot \hat{\boldsymbol{x}}$ where $\hat{\boldsymbol{x}}$ is a deterministic unit vector and X is a scalar complex random variable satisfying $\mathsf{E}[|X|^2] < \mathcal{E}_s$, then the channel output Y is a scalar random variable that can be expressed as

$$Y = (\boldsymbol{H}^{\mathsf{T}}\hat{\boldsymbol{x}})X + Z$$

i.e., as the output of a SISO fading channel of fading $\boldsymbol{H}^{\mathsf{T}}\hat{\boldsymbol{x}}$ and hence (by Theorem 4.16) of fading number

$$\chi(\boldsymbol{H}^{\mathsf{T}}\hat{\boldsymbol{x}}) = \log\pi + \mathsf{E}\left[\log|\boldsymbol{H}^{\mathsf{T}}\hat{\boldsymbol{x}}|^2\right] - h(\boldsymbol{H}^{\mathsf{T}}\hat{\boldsymbol{x}}).$$

Since $\hat{\boldsymbol{x}}$ can be arbitrary, this demonstrates that the RHS of (103) is a lower bound to $\chi(\boldsymbol{H}^{\mathsf{T}})$. *A priori*, it is not clear that this bound is tight, since there could ostensibly be other inputs that are not of the form $X \cdot \hat{\boldsymbol{x}}$ and that give rise to higher mutual informations and perhaps also to higher fading numbers. This, however, is ruled out by the upper bound on the fading number $\chi(\boldsymbol{H}^{\mathsf{T}})$ of Theorem 4.14 (applied to the fading marix $\mathbb{H} = \boldsymbol{H}^{\mathsf{T}}$ with the matrix A chosen as the 1×1 identity matrix), which coincides with the RHS of (103). □

Corollary 4.28: Consider a memoryless Gaussian MISO fading channel where the fading matrix is a row vector $\boldsymbol{H}^{\mathsf{T}}$, where $\boldsymbol{H} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{d},\, \mathsf{K})$, $\det\mathsf{K} \neq 0$, $\boldsymbol{H} \in \mathbb{C}^{n_T}$. Then the fading number is given by

$$\chi(\boldsymbol{H}^{\mathsf{T}}) = -1 + \log d_*^2 - \text{Ei}(-d_*^2) \qquad (104)$$

where

$$d_* = \max_{\|\hat{\boldsymbol{x}}\|=1}\frac{|\mathsf{E}[\boldsymbol{H}^{\mathsf{T}}]\hat{\boldsymbol{x}}|}{\sqrt{\text{Var}(\boldsymbol{H}^{\mathsf{T}}\hat{\boldsymbol{x}})}}. \qquad (105)$$

Proof: Follows directly from Theorem 4.27 and Corollary 4.19 because for any (deterministic) beam direction $\hat{\boldsymbol{x}}$, the concatenation of the beam-forming mapping $\mathbb{C} \ni x \mapsto x \cdot \hat{\boldsymbol{x}} \in \mathbb{C}^{n_T}$ and the fading channel results in the mapping $x \mapsto \boldsymbol{H}^{\mathsf{T}}(x \cdot \hat{\boldsymbol{x}}) + Z$, which corresponds to a SISO Ricean channel. □

Note 4.29: In the above corollary, if the mean vector $\boldsymbol{d}$ is zero, then d_* of (105) is zero, and the fading number is that of a Rayleigh-fading channel, i.e., $-1 - \gamma$. It is achievable by beam forming with an arbitrarily chosen direction.

6) Memoryless SIMO Fading: For memoryless SIMO fading, the capacity-achieving input distribution is circularly symmetric. Indeed, since mutual information over such channels is invariant under deterministic rotation of the input distribution, the concavity of mutual information implies that there is no loss in optimality in considering only circularly symmetric input distributions. This is true also in the presence of side information. Consequently we have the following.

Proposition 4.30: Consider a memoryless fading system where the fading vector $\boldsymbol{H}$ takes value in $\mathbb{C}^{n_R}$ and satisfies $h(\boldsymbol{H}) > -\infty$ and $\mathsf{E}[\|\boldsymbol{H}\|^2] < \infty$. Assume that the additive noise vector $\boldsymbol{Z}$ has an $\mathcal{N}_{\mathbb{C}}(\boldsymbol{0},\, \sigma^2\mathsf{I})$ distribution. Then, the fading number $\chi(\boldsymbol{H})$ is given by

$$\chi(\boldsymbol{H}) = I\big(\Theta;\, \boldsymbol{H}e^{j\Theta}\big) + \mathsf{E}[\log\|\boldsymbol{H}\|] - h\Big(\|\boldsymbol{H}\|\,\Big|\,\hat{\boldsymbol{H}}e^{j\Theta}\Big) - \log 2 \qquad (106)$$

where Θ is independent of $\boldsymbol{H}$ and uniformly distributed over the interval $[-\pi, \pi)$, and $\hat{\boldsymbol{H}} = \boldsymbol{H}/\|\boldsymbol{H}\|$. Alternatively, it can be expressed as

$$\chi(\boldsymbol{H}) = h_\lambda\left(\hat{\boldsymbol{H}}e^{j\Theta}\right) - h(\boldsymbol{H}) + n_{\mathrm{R}}\mathsf{E}\left[\log\|\boldsymbol{H}\|^2\right] - \log 2 \quad (107)$$

where h_λ is the differential entropy on the sphere, see (323).

Proof: Let $X = |X|e^{j\Theta}$ have a circularly symmetric distribution so that $|X|$ and Θ are independent and Θ is uniformly distributed over $[-\pi, \pi)$. Then

$$\begin{aligned} I(X; \boldsymbol{H}X) &= I\left(X; \hat{\boldsymbol{H}}e^{j\Theta}\right) + I\left(X; \|\boldsymbol{H}\|\cdot|X| \,\middle|\, \hat{\boldsymbol{H}}e^{j\Theta}\right) \\ &= I\left(\Theta; \hat{\boldsymbol{H}}e^{j\Theta}\right) + I\left(|X|; \hat{\boldsymbol{H}}e^{j\Theta} \,\middle|\, \Theta\right) \\ &\quad + I\left(|X|; \|\boldsymbol{H}\|\cdot|X| \,\middle|\, \hat{\boldsymbol{H}}e^{j\Theta}\right) \\ &\quad + I\left(\Theta; \|\boldsymbol{H}\|\cdot|X| \,\middle|\, \hat{\boldsymbol{H}}e^{j\Theta}, |X|\right) \\ &= I\left(\Theta; \hat{\boldsymbol{H}}e^{j\Theta}\right) + I\left(|X|; \|\boldsymbol{H}\|\cdot|X| \,\middle|\, \hat{\boldsymbol{H}}e^{j\Theta}\right) \\ &\quad + I\left(\Theta; \|\boldsymbol{H}\| \,\middle|\, \hat{\boldsymbol{H}}e^{j\Theta}\right) \\ &= I\left(\Theta; \boldsymbol{H}e^{j\Theta}\right) + I\left(|X|; \|\boldsymbol{H}\|\cdot|X| \,\middle|\, \hat{\boldsymbol{H}}e^{j\Theta}\right). \end{aligned}$$

The result (106) now follows by analyzing the asymptotics of the second term on the right using a conditional version of Corollary 4.22, which follows from Proposition 4.23 in much the same way that Corollary 4.22 follows from Theorem 4.16.

To derive (107) from (106)

$$\begin{aligned} \chi &= I\left(\Theta; \boldsymbol{H}e^{j\Theta}\right) - h\left(\|\boldsymbol{H}\| \,\middle|\, \hat{\boldsymbol{H}}e^{j\Theta}\right) + \mathsf{E}[\log\|\boldsymbol{H}\|] - \log 2 \\ &= I(\Theta; \|\boldsymbol{H}\|) + I\left(\Theta; \hat{\boldsymbol{H}}e^{j\Theta} \,\middle|\, \|\boldsymbol{H}\|\right) - h\left(\|\boldsymbol{H}\| \,\middle|\, \hat{\boldsymbol{H}}e^{j\Theta}\right) \\ &\quad + \mathsf{E}[\log\|\boldsymbol{H}\|] - \log 2 \\ &= h_\lambda\left(\hat{\boldsymbol{H}}e^{j\Theta} \,\middle|\, \|\boldsymbol{H}\|\right) - h_\lambda\left(\hat{\boldsymbol{H}}e^{j\Theta} \,\middle|\, \Theta, \|\boldsymbol{H}\|\right) \\ &\quad - h\left(\|\boldsymbol{H}\| \,\middle|\, \hat{\boldsymbol{H}}e^{j\Theta}\right) + \mathsf{E}[\log\|\boldsymbol{H}\|] - \log 2 \\ &= h_\lambda\left(\hat{\boldsymbol{H}}e^{j\Theta} \,\middle|\, \|\boldsymbol{H}\|\right) - h\left(\|\boldsymbol{H}\| \,\middle|\, \hat{\boldsymbol{H}}e^{j\Theta}\right) - h_\lambda\left(\hat{\boldsymbol{H}} \,\middle|\, \|\boldsymbol{H}\|\right) \\ &\quad + \mathsf{E}[\log\|\boldsymbol{H}\|] - \log 2 \\ &= h_\lambda\left(\hat{\boldsymbol{H}}e^{j\Theta}\right) - h(\|\boldsymbol{H}\|) - h_\lambda\left(\hat{\boldsymbol{H}} \,\middle|\, \|\boldsymbol{H}\|\right) \\ &\quad + \mathsf{E}[\log\|\boldsymbol{H}\|] - \log 2 \\ &= h_\lambda\left(\hat{\boldsymbol{H}}e^{j\Theta}\right) - h\left(\|\boldsymbol{H}\|, \hat{\boldsymbol{H}}\right) + \mathsf{E}[\log\|\boldsymbol{H}\|] - \log 2 \\ &= h_\lambda\left(\hat{\boldsymbol{H}}e^{j\Theta}\right) - h(\boldsymbol{H}) + (2n_{\mathrm{R}} - 1)\mathsf{E}[\log\|\boldsymbol{H}\|] \\ &\quad + \mathsf{E}[\log\|\boldsymbol{H}\|] - \log 2 \\ &= h_\lambda\left(\hat{\boldsymbol{H}}e^{j\Theta}\right) - h(\boldsymbol{H}) + n_{\mathrm{R}}\mathsf{E}\left[\log\|\boldsymbol{H}\|^2\right] - \log 2 \end{aligned}$$

where we use Lemma 6.17 for the required change of coordinates. □

Note 4.31: This result extends immediately to the case where the receiver has some side information S such that $(\boldsymbol{H}, \boldsymbol{S})$ are independent of the input and additive noise. In that case, (107) should be replaced with

$$\chi(\boldsymbol{H}) = h_\lambda\left(\hat{\boldsymbol{H}}e^{j\Theta} \,\middle|\, \boldsymbol{S}\right) - h(\boldsymbol{H}|\boldsymbol{S}) + n_{\mathrm{R}}\mathsf{E}\left[\log\|\boldsymbol{H}\|^2\right] - \log 2 \quad (108)$$

thus demonstrating that the increase in the fading number may be smaller than $I(\boldsymbol{H}; \boldsymbol{S})$.

Corollary 4.32: Consider a zero-mean memoryless Gaussian SIMO fading channel of a nonsingular $n_{\mathrm{R}} \times n_{\mathrm{R}}$ covariance matrix K, i.e.,

$$\boldsymbol{H} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \mathsf{K}), \qquad \det\mathsf{K} \neq 0. \quad (109)$$

Then the fading number $\chi(\mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \mathsf{K}))$ is given by

$$\chi\left(\mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \mathsf{K})\right) = n_{\mathrm{R}}\psi(n_{\mathrm{R}}) - n_{\mathrm{R}} - \log\Gamma(n_{\mathrm{R}}), \qquad \det\mathsf{K} \neq 0. \quad (110)$$

where $\psi(\cdot)$ denotes Euler's psi-function (213).

Proof: By Lemma 4.7, the fading number is unchanged when $\boldsymbol{H}$ is pre-multiplied by a nonsingular matrix, so that we might as well consider the case where the covariance matrix K is the identity, and the components of $\boldsymbol{H}$ are, therefore, i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$. In this case, $I(\Theta; \boldsymbol{H}e^{j\Theta}) = 0$, and $\boldsymbol{H}$ is isotropically distributed[6] so that $h(\|\boldsymbol{H}\| \,|\, \hat{\boldsymbol{H}}e^{j\Theta}) = h(\|\boldsymbol{H}\|)$. Denoting by $c_{n_{\mathrm{R}}} = 2\pi^{n_{\mathrm{R}}}/\Gamma(n_{\mathrm{R}})$ the surface area of the n_{R}-dimensional sphere in $\mathbb{C}^{n_{\mathrm{R}}}$ we have from (106)

$$\begin{aligned} &\mathsf{E}[\log\|\boldsymbol{H}\|] - h(\|\boldsymbol{H}\|) - \log 2 \\ &\quad = \mathsf{E}[\log\|\boldsymbol{H}\|] - (h(\|\boldsymbol{H}\|) + \log c_{n_{\mathrm{R}}}) + \log\frac{c_{n_{\mathrm{R}}}}{2} \\ &\quad = n_{\mathrm{R}}\mathsf{E}\left[\log\|\boldsymbol{H}\|^2\right] - h(\boldsymbol{H}) + \log\frac{c_{n_{\mathrm{R}}}}{2} \\ &\quad = n_{\mathrm{R}}\mathsf{E}\left[\log\|\boldsymbol{H}\|^2\right] - n_{\mathrm{R}}\log\pi e + \log\frac{c_{n_{\mathrm{R}}}}{2} \\ &\quad = n_{\mathrm{R}}\psi(n_{\mathrm{R}}) - n_{\mathrm{R}} - \log\Gamma(n_{\mathrm{R}}). \end{aligned}$$

Here, the second equality follows by Lemma 6.17 because $\boldsymbol{H}$ is isotropically distributed, the subsequent equality by evaluating $h(\boldsymbol{H})$, and the final equality by the expression for the expected logarithm of a central chi-square random variable (209)–(212). □

Note 4.33: Using the approximation

$$\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n+\frac{1}{12n+1}} < n! < \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n+\frac{1}{12n}} \quad (111)$$

one can show that

$$\begin{aligned} \chi(\mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \mathsf{K})) &= n_{\mathrm{R}}\psi(n_{\mathrm{R}}) - n_{\mathrm{R}} - \log\Gamma(n_{\mathrm{R}}) \\ &= \frac{1}{2}\log\frac{n_{\mathrm{R}}}{2\pi} - \frac{1}{2} - \frac{1}{6n_{\mathrm{R}}} + \mathrm{O}\left(\frac{1}{n_{\mathrm{R}}^2}\right), \ \det\mathsf{K} \neq 0 \end{aligned} \quad (112)$$

which can be compared to the results of Sengupta and Mitra [13] who studied this scenario for K being the identity matrix and under the approximation $n_{\mathrm{R}} \gg 1$. The approximation they got using the Laplace integration method for $n_{\mathrm{R}} \gg 1$ is

$$C \approx \frac{1}{2}\log\frac{n_{\mathrm{R}}}{2\pi} + \log\log\mathrm{SNR}. \quad (113)$$

But for the constant $1/2$, the expressions (113) and (112) agree as $n_{\mathrm{R}} \to \infty$.

7) Memoryless MIMO Rotation Commutative Fading: A different extension of Theorem 4.16 is to MIMO fading matrices that are of a law with a particular kind of symmetry that we call "rotation commutative."

[6]For a definition of isotropic distributions see Definition 6.19.

Definition 4.34: We shall say that the law of a random $n \times n$ matrix $\mathbb{H}$ is *rotation commutative* if for any deterministic unitary $n \times n$ matrix V the law of $\mathsf{V}\mathbb{H}$ is identical to the law of $\mathbb{H}\mathsf{V}$.

For such laws we can extend Theorem 4.16 as follows.

Proposition 4.35: Consider a memoryless fading channel where the number of receive antennas and transmit antennas are equal ($n_\mathrm{R} = n_\mathrm{T}$) and where the fading matrix $\mathbb{H}$ is rotation commutative. Further assume that $h(\mathbb{H}) > -\infty$ and $\mathsf{E}[\|\mathbb{H}\|_\mathrm{F}^2] < \infty$. Then the limsup in (44) is also a liminf (i.e., the limit exists) and the fading number $\chi(\mathbb{H})$ is given by

$$\chi(\mathbb{H}) = \log \frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})} - \log 2 + n_\mathrm{R} \mathsf{E}\left[\log \|\mathbb{H}\hat{\boldsymbol{e}}\|^2\right] - h(\mathbb{H}\hat{\boldsymbol{e}}) \quad (114)$$

where $\hat{\boldsymbol{e}}$ is any deterministic unit vector in $\mathbb{C}^{n_\mathrm{T}}$ and $c_{n_\mathrm{R}} = 2\pi^{n_\mathrm{R}}/\Gamma(n_\mathrm{R})$ is the surface area of a unit sphere in $\mathbb{C}^{n_\mathrm{R}}$. Moreover, this fading number is achievable by inputs that can be expressed as the product of a uniformly distributed random vector on the unit n_T-sphere multiplied by an independent circularly symmetric scalar random variable of a law specified in Theorem 4.16.

Proof: This result can be viewed as a special case of an analogous result for channels that are "rotation commutative in the generalized sense," namely, Theorem 4.39. The proof is therefore omitted. □

Corollary 4.36: Consider memoryless Gaussian fading of the form $\mathbb{H} = d\mathsf{I} + \tilde{\mathbb{H}}$ where $n_\mathrm{R} = n_\mathrm{T} = m$, the matrix I denotes the identity matrix, $d \in \mathbb{C}$ is deterministic, and the components of $\tilde{\mathbb{H}}$ are i.i.d. $\mathcal{N}_\mathbb{C}(0, 1)$. Then the fading number is given by

$$\chi = m g_m(|d|^2) - m - \log \Gamma(m) \quad (115)$$

where the function $g_m(z)$ is defined in (210).

Proof: Follows from (114) by direct computation of the differential entropy of the multivariate Gaussian distribution and of the expectation of the logarithm of a noncentral chi-square distributed random variable with $2m$ degrees of freedom (209). □

Definition 4.37: We shall say that the $n_\mathrm{R} \times n_\mathrm{T}$ random matrix $\mathbb{H}$ is *rotation commutative in the generalized sense* if the following two conditions hold:

- for any deterministic unitary $n_\mathrm{T} \times n_\mathrm{T}$ matrix V_t, there exists an $n_\mathrm{R} \times n_\mathrm{R}$ deterministic unitary matrix V_r such that

$$\mathsf{V}_r \mathbb{H} \stackrel{\mathcal{L}}{=} \mathbb{H}\mathsf{V}_t \quad (116)$$

  where $\stackrel{\mathcal{L}}{=}$ stands for "equal in law";
- for any deterministic unitary $n_\mathrm{R} \times n_\mathrm{R}$ matrix V_r, there exists a deterministic unitary $n_\mathrm{T} \times n_\mathrm{T}$ matrix V_t such that (116) holds.

The following lemma lists some of the properties that will be useful for the analysis of the fading number of such matrices.

Lemma 4.38: Let $\mathbb{H}$ be rotation commutative in the generalized sense. Then the following two statements hold:

- if $\hat{\boldsymbol{X}} \in \mathbb{C}^{n_\mathrm{T}}$ is an isotropically distributed random vector that is independent of $\mathbb{H}$, then $\mathbb{H}\hat{\boldsymbol{X}} \in \mathbb{C}^{n_\mathrm{R}}$ is isotropically distributed;
- if $\hat{\boldsymbol{e}}, \hat{\boldsymbol{e}}' \in \mathbb{C}^{n_\mathrm{T}}$ are two deterministic unit vectors, then

$$\|\mathbb{H}\hat{\boldsymbol{e}}\| \stackrel{\mathcal{L}}{=} \|\mathbb{H}\hat{\boldsymbol{e}}'\|, \qquad \|\hat{\boldsymbol{e}}\| = \|\hat{\boldsymbol{e}}'\| = 1 \quad (117)$$
$$h(\mathbb{H}\hat{\boldsymbol{e}}) = h(\mathbb{H}\hat{\boldsymbol{e}}'), \qquad \|\hat{\boldsymbol{e}}\| = \|\hat{\boldsymbol{e}}'\| = 1. \quad (118)$$

Proof: We shall prove the first part of the lemma by showing that for any deterministic $n_\mathrm{R} \times n_\mathrm{R}$ matrix V_r the law of $\mathsf{V}_r\mathbb{H}\hat{\boldsymbol{X}}$ is identical to the law of $\mathbb{H}\hat{\boldsymbol{X}}$. To this end, let V_t be such that $\mathsf{V}_r\mathbb{H} \stackrel{\mathcal{L}}{=} \mathbb{H}\mathsf{V}_t$. Then

$$\begin{aligned}\mathsf{V}_r\mathbb{H}\hat{\boldsymbol{X}} &\stackrel{\mathcal{L}}{=} \mathbb{H}\mathsf{V}_t\hat{\boldsymbol{X}} \\ &\stackrel{\mathcal{L}}{=} \mathbb{H}\hat{\boldsymbol{X}}\end{aligned}$$

where the second equality in law follows because $\hat{\boldsymbol{X}}$ is isotropically distributed.

To prove the second claim, let $\mathsf{V}_{\hat{\boldsymbol{e}}\mapsto\hat{\boldsymbol{e}}'}$ be some deterministic unitary matrix satisfying $\mathsf{V}_{\hat{\boldsymbol{e}}\mapsto\hat{\boldsymbol{e}}'}\hat{\boldsymbol{e}} = \hat{\boldsymbol{e}}'$. Let U be a deterministic $n_\mathrm{T} \times n_\mathrm{T}$ unitary matrix such that $\mathsf{U}\mathbb{H} \stackrel{\mathcal{L}}{=} \mathbb{H}\mathsf{V}_{\hat{\boldsymbol{e}}\mapsto\hat{\boldsymbol{e}}'}$. Thus,

$$\begin{aligned}\mathbb{H}\hat{\boldsymbol{e}}' &= \mathbb{H}\mathsf{V}_{\hat{\boldsymbol{e}}\mapsto\hat{\boldsymbol{e}}'}\hat{\boldsymbol{e}} \\ &\stackrel{\mathcal{L}}{=} \mathsf{U}\mathbb{H}\hat{\boldsymbol{e}}.\end{aligned}$$

The lemma now follows by noting that both the L_2 norm of a random vector and its differential entropy are invariant with respect to deterministic unitary matrix multiplication. □

We are now ready to generalize Proposition 4.35 to fading matrices that are rotation commutative in the generalized sense.

Theorem 4.39: Consider a memoryless fading channel where the fading matrix $\mathbb{H}$ is rotation commutative in the generalized sense. Further assume $h(\mathbb{H}) > -\infty$ and $\mathsf{E}[\|\mathbb{H}\|_\mathrm{F}^2] < \infty$. Then the limsup in (44) is also a liminf (i.e., the limit exists) and the fading number $\chi(\mathbb{H})$ is given by

$$\chi(\mathbb{H}) = \log \frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})} - \log 2 + n_\mathrm{R}\mathsf{E}[\log \|\mathbb{H}\hat{\boldsymbol{e}}\|^2] - h(\mathbb{H}\hat{\boldsymbol{e}}) \quad (119)$$

where $\hat{\boldsymbol{e}}$ is any deterministic unit vector in $\mathbb{C}^{n_\mathrm{T}}$ and $c_{n_\mathrm{R}} = 2\pi^{n_\mathrm{R}}/\Gamma(n_\mathrm{R})$ is the surface area of a unit sphere in $\mathbb{C}^{n_\mathrm{R}}$. Moreover, this fading number is achievable by inputs that can be expressed as the product of a uniformly distributed random vector on the unit n_T-sphere multiplied by an independent circularly symmetric scalar random variable of a law specified in Theorem 4.16.

Proof: The fact that the RHS of (119) is an upper bound to $\chi(\mathbb{H})$ follows directly from Theorem 4.14 applied with A chosen as the identity matrix.

To derive a lower bound, let $\boldsymbol{X} = \hat{\boldsymbol{X}} \cdot \|\boldsymbol{X}\|$ be isotropically distributed with $\log \|\boldsymbol{X}\|^2$ uniformly distributed over the interval $[\log \mathcal{E}_0, \log \mathcal{E}_\mathrm{s}]$ and independent of the Haar distributed unit vector $\hat{\boldsymbol{X}} = \boldsymbol{X}/\|\boldsymbol{X}\|$. Let $\hat{\boldsymbol{e}}$ be an arbitrary unit vector in $\mathbb{C}^{n_\mathrm{T}}$. Using the chain rule we now have

$$I(\boldsymbol{X}; \mathbb{H}\boldsymbol{X}) = I(\|\boldsymbol{X}\|; \mathbb{H}\boldsymbol{X}) + I\left(\hat{\boldsymbol{X}}; \mathbb{H}\boldsymbol{X} \,\middle|\, \|\boldsymbol{X}\|\right). \quad (120)$$

The term $I(\|\boldsymbol{X}\|;\, \mathbb{H}\boldsymbol{X})$ can be written as $I(\|\boldsymbol{X}\|;\, (\mathbb{H}\hat{\boldsymbol{X}})\cdot\|\boldsymbol{X}\|)$. But $(\mathbb{H}\hat{\boldsymbol{X}})$ is independent of $\|\boldsymbol{X}\|$ and isotropically distributed, so that

$$\begin{aligned} I(\|\boldsymbol{X}\|;\, \mathbb{H}\boldsymbol{X}) &= I(\|\boldsymbol{X}\|;\, \|\mathbb{H}\hat{\boldsymbol{e}}\|\cdot\|\boldsymbol{X}\|) \\ &= \log\log\frac{\mathcal{E}_{\mathrm{s}}}{\mathcal{E}_0} + \mathsf{E}[\log\|\mathbb{H}\hat{\boldsymbol{e}}\|] \\ &\quad - h(\|\mathbb{H}\hat{\boldsymbol{e}}\|) - \log 2 + o(1) \end{aligned} \tag{121}$$

where the last equality follows by Corollary 4.22.

We now turn to the second term on the RHS of (120)

$$\begin{aligned} I\left(\hat{\boldsymbol{X}};\, \mathbb{H}\boldsymbol{X}\,\middle|\,\|\boldsymbol{X}\|\right) &= I\left(\hat{\boldsymbol{X}};\, \mathbb{H}\hat{\boldsymbol{X}}\cdot\|\boldsymbol{X}\|\,\middle|\,\|\boldsymbol{X}\|\right) \\ &= I\left(\hat{\boldsymbol{X}};\, \mathbb{H}\hat{\boldsymbol{X}}\right) \\ &= h(\mathbb{H}\hat{\boldsymbol{X}}) - h(\mathbb{H}\hat{\boldsymbol{X}}|\hat{\boldsymbol{X}}) \\ &= h(\|\mathbb{H}\hat{\boldsymbol{e}}\|) + \log c_{n_{\mathrm{R}}} \\ &\quad + (2n_{\mathrm{R}}-1)\mathsf{E}[\log\|\mathbb{H}\hat{\boldsymbol{e}}\|] - h(\mathbb{H}\hat{\boldsymbol{e}}) \end{aligned} \tag{122}$$

where the last equality follows by Lemma 6.17 because $\mathbb{H}\hat{\boldsymbol{X}}$ is isotropically distributed and because $\|\mathbb{H}\hat{\boldsymbol{X}}\| \stackrel{\mathcal{L}}{=} \|\mathbb{H}\hat{\boldsymbol{e}}\|$.

The theorem now follows from (120)–(122). □

8) Memoryless Gaussian MIMO Fading With Mean: We next briefly discuss the fading number $\chi(\mathbb{H})$ of an $n_{\mathrm{R}} \times n_{\mathrm{T}}$ random matrix $\mathbb{H}$ that is of the form

$$\mathbb{H} = \mathsf{D} + \tilde{\mathbb{H}} \tag{123}$$

where D is a deterministic $n_{\mathrm{R}} \times n_{\mathrm{T}}$ matrix and $\tilde{\mathbb{H}}$ is a random $n_{\mathrm{R}} \times n_{\mathrm{T}}$ matrix of i.i.d. $\mathcal{N}_{\mathbb{C}}(0,\,1)$ components. Note that by Lemma 4.7 and the singular value decomposition (SVD) it follows that for the purposes of computing the fading number $\chi(\mathbb{H})$ we may assume without loss of generality that D is "diagonal" (in the sense that $d^{(r,\,t)}$ is zero whenever $r \neq t$) and that the terms on the diagonal correspond to the singular values of the mean matrix.

The case $n_{\mathrm{R}} = n_{\mathrm{T}} = 1$ corresponds to SISO Ricean fading with the corresponding fading number given in Corollary 4.19. The case $n_{\mathrm{T}} > 1$, $n_{\mathrm{R}} = 1$ corresponds to MISO fading and is addressed in Corollary 4.28. And the case $n_{\mathrm{R}} = n_{\mathrm{T}}$ with $\mathsf{D} = d\cdot\mathsf{I}$ being a scalar matrix is addressed in Corollary 4.36.

For this general model we were unable to obtain an exact expression for the fading number. A trivial lower bound

$$\chi(\mathbb{H}) \geq g_1\left(\|\mathsf{D}\|^2\right) - 1 \tag{124}$$

where $\|\mathsf{D}\|$ denotes the matrix norm as defined in (13) and the function g_1 is defined in (210), can be derived by considering a beamforming transmission strategy with linear combining at the receiver.

The upper bound

$$\chi(\mathbb{H}) \leq n_{\mathrm{R}}\cdot g_{n_{\mathrm{R}}}\left(\|\mathsf{D}\|^2\right) - n_{\mathrm{R}} - \log\Gamma(n_{\mathrm{R}}) \tag{125}$$

follows from Theorem 4.14 applied with A being the identity matrix.

The upper bound, however, can be improved in many instances by optimizing over the $n_{\mathrm{R}} \times n_{\mathrm{R}}$ matrix A. This optimization is greatly simplified (albeit with some loss in tightness) using Jensen's inequality, which leads to (79). For the case at hand we obtain

$$\begin{aligned} \chi_{\mathrm{u}} \leq{}& -n_{\mathrm{R}} - \log\Gamma(n_{\mathrm{R}}) \\ &+ n_{\mathrm{R}}\log\inf_{\det(\mathsf{A}\mathsf{A}^\dagger)=1}\left\{\|\mathsf{A}\mathsf{D}\|^2 + \operatorname{tr}\left(\mathsf{A}\mathsf{A}^\dagger\right)\right\}. \end{aligned} \tag{126}$$

While the minimization over the matrix A can be performed analytically in accordance with the singular values of the mean matrix (e.g., by choosing A diagonal with some of its diagonal elements being proportional to the reciprocal of the corresponding singular values and the rest being constant), we present here a suboptimal choice that will lead to a bound that depends only on the maximal singular value. This choice will suffice to capture the dependence of the fading number on the number of receiver antennas n_{R} and the number of transmitter antennas n_{T}.

The suboptimal choice of A is a simple one. We choose it diagonal, with the diagonal elements taking on one of two values according to whether the corresponding singular value of the mean matrix is zero or not. Optimizing on the choice of the values that the diagonal elements of A may take we obtain the bound

$$\begin{aligned} \chi(\mathbb{H}) \leq{}& n_{\mathsf{D}}\log\left(1 + \frac{|d_{\max}|^2}{n_{\mathsf{D}}}\right) \\ &+ n_{\mathrm{R}}\log n_{\mathrm{R}} - n_{\mathrm{R}} - \log\Gamma(n_{\mathrm{R}}) \end{aligned} \tag{127}$$

where n_{D} denotes the rank of D and $d_{\max}$ the maximal singular value of D. Recalling that the maximal singular value is the operator norm of the matrix and upper-bounding the rank of D by $\min\{n_{\mathrm{R}},\, n_{\mathrm{T}}\}$ we obtain

$$\begin{aligned} \chi(\mathbb{H}) \leq{}& \min\{n_{\mathrm{R}},\, n_{\mathrm{T}}\}\log\left(1 + \frac{\|\mathsf{D}\|^2}{\min\{n_{\mathrm{R}},\, n_{\mathrm{T}}\}}\right) \\ &+ n_{\mathrm{R}}\log n_{\mathrm{R}} - n_{\mathrm{R}} - \log\Gamma(n_{\mathrm{R}}). \end{aligned} \tag{128}$$

E. *On the Fading Number of Fading With Memory*

From Lemma 4.5 we obtain immediately the following upper bound on the fading number.

Theorem 4.40: Let the side information $\{\boldsymbol{S}_k\}$ and fading process $\{\mathbb{H}_k\}$ satisfy the assumptions of Theorem 4.2, and let the additive noise be spatially and temporally Gaussian. Then

$$\chi(\{\mathbb{H}_k\}|\{\boldsymbol{S}_k\}) \leq \chi(\{\mathbb{H}_k\}) + I(\{\mathbb{H}_k\};\, \{\boldsymbol{S}_k\}) \tag{129}$$

$$\leq \chi_{\mathrm{i.i.d.}}(\mathbb{H}_1) + h(\mathbb{H}_1) - h(\{\mathbb{H}_k\}|\{\boldsymbol{S}_k\}) \tag{130}$$

where $\chi_{\mathrm{i.i.d.}}(\mathbb{H}_1)$ denotes the fading number in the memoryless fading case with equal marginal and no side information, and where

$$I(\{\mathbb{H}_k\};\, \{\boldsymbol{S}_k\}) \triangleq \lim_{n\uparrow\infty}\frac{1}{n}I(\mathbb{H}_1^n;\, \boldsymbol{S}_1^n) \tag{131}$$

$$h(\{\mathbb{H}_k\}|\{\boldsymbol{S}_k\}) \triangleq \lim_{n\uparrow\infty}\frac{1}{n}h(\mathbb{H}_1^n|\boldsymbol{S}_1^n). \tag{132}$$

For SISO systems, this bound is tight.

Theorem 4.41: Consider a SISO system where the side information $\{\boldsymbol{S}_k\}$ and fading process $\{H_k\}$ satisfy the assumptions of Theorem 4.2, and let the additive noise be spatially and temporally Gaussian noise. Then the limsup in (44) is also a liminf (i.e., the limit exists) and the fading number $\chi(\{H_k\}|\{S_k\})$ is given by

$$\chi(\{H_k\}|\{S_k\}) = \log\pi + \mathsf{E}\big[\log|H_1|^2\big] - h(\{H_k\}|\{S_k\}). \tag{133}$$

Moreover, this fading number is achievable by i.i.d. input distributions of marginals of the form specified in Theorem 4.16.

Proof: In view of Theorem 4.40, it suffices to demonstrate that the proposed fading number is achievable. Here we shall present the main ingredients of the proof and leave some of the technical details for Appendix IX.

Let $\{X_k\}$ be i.i.d. circularly symmetric random variables with

$$\log|X_k|^2 \sim \text{Uniform}[\log x_{\min}^2, \log\mathcal{E}_{\mathrm{s}}]. \tag{134}$$

Our proof will hinge on the fact that if $\log x_{\min}^2$ grows sublinearly in $\log\mathcal{E}_{\mathrm{s}}$ to infinity, then this input distribution achieves the fading number of any memoryless SISO channel with any side information, and on the fact that this input distribution allows us to "identify" the channel, in the sense that from past inputs and past outputs one can ever more accurately estimate past fading levels. The details follow.

Fix some (large) positive integer κ and use the chain rule and the nonnegativity of mutual information to obtain

$$\begin{aligned}\frac{1}{n}I(X^n; Y^n, \boldsymbol{S}^n) &= \frac{1}{n}\sum_{k=1}^{n} I\big(X_k; Y^n, \boldsymbol{S}^n|X^{k-1}\big)\\ &\geq \frac{1}{n}\sum_{k=\kappa+1}^{n-\kappa} I\big(X_k; Y^n, \boldsymbol{S}^n|X^{k-1}\big).\end{aligned}$$

We shall now obtain a firm bound on $I(X_k; Y^n, \boldsymbol{S}^n|X^{k-1})$ for $\kappa+1 \leq k \leq n-\kappa$. By letting $n\to\infty$ we shall deduce that this firm lower bound is also a lower bound on the limiting mutual information. Consider then some $\kappa+1 \leq k \leq n-\kappa$. Then because $\{X_k\}$ are i.i.d.

$$\begin{aligned}&I\big(X_k; Y^n, \boldsymbol{S}^n|X^{k-1}\big)\\ &\quad= I\big(X_k; Y^n, \boldsymbol{S}^n, X^{k-1}\big)\\ &\quad\geq I\Big(X_k; Y_k, Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa}\Big)\\ &\quad= I\Big(X_k; Y_k, Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, H_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa}\Big) - \epsilon\\ &\quad= I\Big(X_k; Y_k, H_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa}\Big) - \epsilon\\ &\quad= I\Big(X_{\kappa+1}; Y_{\kappa+1}, H_1^{\kappa}, \boldsymbol{S}_1^{2\kappa+1}\Big) - \epsilon\\ &\quad= I\Big(X_{\kappa+1}; Y_{\kappa+1}\big|H_1^{\kappa}, \boldsymbol{S}_1^{2\kappa+1}\Big) - \epsilon,\\ &\qquad\qquad\qquad \kappa+1 \leq k \leq n-\kappa\end{aligned} \tag{135}$$

where the equality before last follows from stationarity and where ϵ, which is given by

$$\begin{aligned}\epsilon = I\left(X_k; Y_k, Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, H_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa}\right)\\ -I\left(X_k; Y_k, Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa}\right)\end{aligned} \tag{136}$$

will be shown in Appendix IX to tend to zero as $x_{\min}^2 \uparrow \infty$. In fact, that is where we use the fact that the proposed input distribution allows us to "identify" the channel.

Returning to the RHS of (135) we can now view the term

$$I\left(X_{\kappa+1}; Y_{\kappa+1}\big|H_1^{\kappa}, \boldsymbol{S}_1^{2\kappa+1}\right)$$

as the mutual information across a memoryless fading channel in the presence of the side information

$$\left(H_1^{\kappa}, \boldsymbol{S}_1^{2\kappa+1}\right).$$

Thus, using Proposition 4.23, we obtain that the fading number

$$\log\pi + \mathsf{E}\left[\log|H_1|^2\right] - h\left(H_{\kappa+1}\big|H_1^{\kappa}, \boldsymbol{S}_1^{2\kappa+1}\right)$$

is achievable. The proof will now be concluded by showing that

$$\lim_{\kappa\to\infty} h\left(H_{\kappa+1}\big|H_1^{\kappa}, \boldsymbol{S}_1^{2\kappa+1}\right) \leq h\left(\{H_k\}|\{\boldsymbol{S}_k\}\right). \tag{137}$$

(There is, in fact, equality in the above, but we only need the inequality.) This follows from the inequality

$$\begin{aligned}\frac{1}{\kappa}h\left(H_1^{\kappa}|\boldsymbol{S}_1^{\kappa}\right) &= \frac{1}{\kappa}\sum_{k=1}^{\kappa} h\left(H_k|H_1^{k-1}, \boldsymbol{S}_1^{\kappa}\right)\\ &\geq h\left(H_{\kappa+1}|H_1^{\kappa}, \boldsymbol{S}_1^{2\kappa+1}\right)\end{aligned} \tag{138}$$

which holds by stationarity and because conditioning cannot increase differential entropy

$$\begin{aligned}h\big(H_k\big|H_1^{k-1}, \boldsymbol{S}_1^{\kappa}\big) &= h\Big(H_{\kappa+1}\big|H_{\kappa-k+2}^{\kappa}, \boldsymbol{S}_{\kappa-k+2}^{2\kappa-k+1}\Big)\\ &\geq h\Big(H_{\kappa+1}\big|H_1^{\kappa}, \boldsymbol{S}_1^{2\kappa+1}\Big), \quad 0 \leq k \leq \kappa+1.\end{aligned}$$

□

Corollary 4.42: Consider a SISO fading process $\{H_k\}$ such that for some specular component $d \in \mathbb{C}$ the process $\{H_k - d\}$ is a zero-mean unit-variance circularly symmetric stationary and ergodic complex Gaussian process whose spectrum is of continuous part $F'(\lambda)$, $-1/2 \leq \lambda \leq 1/2$. Then

$$\begin{aligned}\chi\left(\{H_k\}|\{S_k\}\right) = \log|d|^2 - \mathrm{Ei}(-|d|^2) - 1\\ + \log\frac{1}{\epsilon_{\mathrm{MSE}}^2} + I(\{H_k\}; \{\boldsymbol{S}_k\})\end{aligned} \tag{139}$$

where $\epsilon_{\mathrm{MSE}}^2 > 0$ denotes the minimum mean squared error in predicting the present fading from its past (assumed positive)

$$\epsilon_{\mathrm{MSE}}^2 = \exp\left\{\int_{-1/2}^{1/2} \log F'(\lambda)\,\mathrm{d}\lambda\right\} \tag{140}$$

and where the mutual information rate is assumed finite.

Proof: Follows from Theorem 4.41 by evaluating the expectation of the logarithm of a noncentral chi-square random variable as in (209) and by expressing the entropy rate of a Gaussian process in terms of the minimum mean squared error in estimating its present value from its past. □

F. Nonasymptotic Bounds

1) Rayleigh-Fading Channel: The memoryless SISO Rayleigh-fading channel corresponds to the general memoryless fading model (20) in the special case where the random matrix $\mathbb{H}$ is a scalar $\mathcal{N}_{\mathbb{C}}(0, 1)$ random variable. The capacity of this channel was studied in [14] and [11]. Taricco and Elia [14] derived a lower bound on capacity and also argued that at high SNR, capacity grows double-logarithmically in the SNR. Abou-Faycal *et al.* showed that for any given SNR, capacity is achieved by an input distribution of a finite number of mass points, and they were thus able to express capacity as a finite-dimensional (nonconcave) optimization problem over the locations and weights of the mass points. This allowed for an exact calculation of channel capacity at low SNR, but not at high SNR, where the number of mass points becomes large and the optimization problem, while finite dimensional, becomes intractable.

Here we shall use (25) in order to obtain the upper bound

$$C \le \inf_{\alpha,\,\beta'>0}\inf_{\delta'\ge 0}\Bigg\{-1+\alpha\log\beta'+\log\Gamma\left(\alpha,\,\frac{\delta'}{\beta'}\right) + \frac{1+\mathcal{E}_s/\sigma^2}{\beta'}+\frac{\delta'}{\beta'}+(1-\alpha) \cdot\left(\log\delta'-e^{\delta'}\cdot\mathrm{Ei}(-\delta')\right)\Bigg\}. \quad (141)$$

To this end, we first note that conditional on the input $X=x$, the channel output Y has an $\mathcal{N}_{\mathbb{C}}(0,\,|x|^2+\sigma^2)$ distribution, so that $|Y|^2$ has an exponential distribution of mean $|x|^2+\sigma^2$. Consequently

$$h(Y|X=x) = \log\left(\pi e(|x|^2+\sigma^2)\right) \quad (142)$$

$$\mathsf{E}[|Y|^2|X=x] = |x|^2+\sigma^2 \quad (143)$$

$$\mathsf{E}[\log(|Y|^2+\delta)|X=x] = \log\delta - e^{\frac{\delta}{|x|^2+\sigma^2}} \cdot\mathrm{Ei}\left(-\frac{\delta}{|x|^2+\sigma^2}\right) \quad (144)$$

where the last equality follows from [15, 4.337 (2)]. It thus follows from (25) that

$$\begin{aligned} I(Q;\,W) &\le -1+\alpha\log\frac{\beta}{\delta}+\log\Gamma\left(\alpha,\,\frac{\delta}{\beta}\right)+\frac{\mathcal{E}_s+\sigma^2+\delta}{\beta} \\ &\quad + \mathsf{E}\left[\log\frac{\delta}{|X|^2+\sigma^2}-(1-\alpha)e^{\frac{\delta}{|X|^2+\sigma^2}} \cdot\mathrm{Ei}\left(-\frac{\delta}{|X|^2+\sigma^2}\right)\right] \quad (145)\\ &\le -1+\alpha\log\frac{\beta}{\delta}+\log\Gamma\left(\alpha,\,\frac{\delta}{\beta}\right)+\frac{\mathcal{E}_s+\sigma^2+\delta}{\beta} \\ &\quad +\sup_x\left\{\log\frac{\delta}{|x|^2+\sigma^2}-(1-\alpha)e^{\frac{\delta}{|x|^2+\sigma^2}} \cdot\mathrm{Ei}\left(-\frac{\delta}{|x|^2+\sigma^2}\right)\right\} \quad (146)\\ &= -1+\alpha\log\frac{\beta}{\delta}+\log\Gamma\left(\alpha,\,\frac{\delta}{\beta}\right)+\frac{\mathcal{E}_s+\sigma^2+\delta}{\beta} \\ &\quad +\log\frac{\delta}{\sigma^2}-(1-\alpha)e^{\frac{\delta}{\sigma^2}}\cdot\mathrm{Ei}\left(-\frac{\delta}{\sigma^2}\right), \\ &\qquad \alpha,\,\beta>0,\ \delta\ge 0 \quad (147)\end{aligned}$$

where the final equality follows because for every $\alpha>0$ the function

$$\mathbb{R}^+ \ni w \mapsto \log w-(1-\alpha)e^{w}\cdot\mathrm{Ei}(-w)$$

is monotonically increasing in w. The inequality (141) now follows from (147) upon substituting

$$\delta'=\frac{\delta}{\sigma^2},\qquad \beta'=\frac{\beta}{\sigma^2}. \quad (148)$$

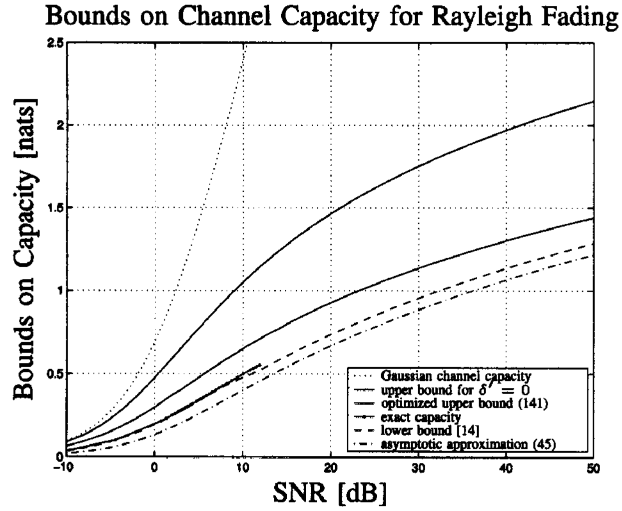

Fig. 1. Bounds on the capacity of a Rayleigh-fading channel. Depicted are the upper bound of (141)); the upper bound that results from the suboptimal choice[7] of $\delta'=0$; the lower bound of Taricco and Elia [14]; the exact expression from [11]; the approximation of (45), (85); and the capacity $\log(1+\mathrm{SNR})$ of a Gaussian channel of equal SNR.

Fig. 1 depicts the upper bound (141) on channel capacity. For reference we also plot the cruder but simpler upper bound[7] that results from choosing $\delta'=0$; the asymptotic approximation (45), (85); the lower bound of [14]; the exact expression of [11] in the region where it is amenable to numerical calculation; and the capacity $\log(1+\mathrm{SNR})$ of an additive white Gaussian noise channel of equal SNR.

It is interesting to note the dramatic difference between the high-SNR behavior of channel capacity in the absence of side information (85)

$$C=\log\log\frac{\mathcal{E}_s}{\sigma^2}-1-\gamma+o(1) \quad (149)$$

and in its presence (when perfect) [16]

$$C_{\mathrm{PSI}}=\log\frac{\mathcal{E}_s}{\sigma^2}-\gamma+o(1). \quad (150)$$

2) Multiple-Antenna Rayleigh-Fading Channel: Next, we consider a Rayleigh-fading channel with n_T transmit and n_R receive antennas, i.e., the channel (20) specialized to the case where $\mathbb{H}$ is a complex $n_R\times n_T$ random matrix of i.i.d. $\mathcal{N}_{\mathbb{C}}(0,\,1)$ components. We shall derive an upper bound on the capacity C whose difference from capacity will shrink to zero as SNR $\uparrow\infty$ and whose ratio to capacity will tend to one as SNR $\downarrow 0$. It is based on an application of (25) with $\mathsf{A}=\mathsf{I}$ and is given by

$$C\le\inf_{\alpha,\,\beta'>0}\inf_{\delta'\ge 0}\Bigg\{-n_R+n_R\psi(n_R)-\log\Gamma(n_R) +\alpha\log\beta'+\log\Gamma(\alpha,\delta'/\beta')-\alpha\psi(n_R)+\frac{n_R(1+\mathrm{SNR})}{\beta'} +\frac{\delta'}{\beta'}+\frac{1-\alpha}{n_R-1}\delta'\cdot I\{\alpha\le 1\}\Bigg\},\qquad n_R>1 \quad (151)$$

where SNR is defined as $\mathcal{E}_s/\sigma^2$.

[7]The bound resulting from this suboptimal choice coincides with the upper bound of Taricco and Elia. Note that this bound is not asymptotically tight at high SNR. It is off by 1 nat per channel use.

To derive (151) from (25) we begin by considering the term

$$(1-\alpha)\mathsf{E}_Q\left[\log(\|\boldsymbol{Y}\|^2+\delta)\right] = (1-\alpha)\mathsf{E}_Q\left[\log\|\boldsymbol{Y}\|^2\right] + (1-\alpha)\mathsf{E}\left[\log\left(1+\frac{\delta}{\|\boldsymbol{Y}\|^2}\right)\right]. \quad (152)$$

For $0<\alpha\le 1$, we upper-bound the second term on the RHS of (152) as follows:

$$\begin{aligned}(1-\alpha)\mathsf{E}_Q&\left[\log\left(1+\frac{\delta}{\|\boldsymbol{Y}\|^2}\right)\right]\\ &\le (1-\alpha)\mathsf{E}_Q\left[\left(\frac{\delta}{\|\boldsymbol{Y}\|^2}\right)\right]\\ &= (1-\alpha)\,\frac{\delta}{n_\mathrm{R}-1}\,\mathsf{E}_Q\left[\frac{1}{\|\boldsymbol{X}\|^2+\sigma^2}\right]\\ &\le (1-\alpha)\,\frac{\delta}{n_\mathrm{R}-1}\max_{\boldsymbol{x}}\frac{1}{\|\boldsymbol{x}\|^2+\sigma^2}\\ &= \frac{\delta(1-\alpha)}{\sigma^2(n_\mathrm{R}-1)}, \qquad 0<\alpha\le 1,\ n_\mathrm{R}>1 \qquad (153)\end{aligned}$$

and for $\alpha>1$, we upper-bound it by

$$(1-\alpha)\mathsf{E}_Q\left[\log\left(1+\frac{\delta}{\|\boldsymbol{Y}\|^2}\right)\right]\le 0, \qquad \alpha>1. \quad (154)$$

We compute the remaining terms in (25) by noting that, conditional on the input $\boldsymbol{X}=\boldsymbol{x}$, the channel output $\boldsymbol{Y}$ has an $\mathcal{N}_\mathbb{C}(\mathbf{0},(\|\boldsymbol{x}\|^2+\sigma^2)\mathsf{I})$ distribution and $\|\boldsymbol{Y}\|^2$ is central chi-square distributed. Consequently

$$h(\boldsymbol{Y}|\boldsymbol{X}=\boldsymbol{x}) = n_\mathrm{R}\log\pi + n_\mathrm{R} + n_\mathrm{R}\log(\|\boldsymbol{x}\|^2+\sigma^2) \quad (155)$$

$$\mathsf{E}[\|\boldsymbol{Y}\|^2|\boldsymbol{X}=\boldsymbol{x}] = n_\mathrm{R}(\|\boldsymbol{x}\|^2+\sigma^2) \quad (156)$$

$$\mathsf{E}[\log\|\boldsymbol{Y}\|^2|\boldsymbol{X}=\boldsymbol{x}] = \log(\|\boldsymbol{x}\|^2+\sigma^2)+\psi(n_\mathrm{R}) \quad (157)$$

where the last expression follows from the general expression (209) for the expected logarithm of a noncentral chi-square distribution in the special case where the noncentrality parameter is zero.

Using (152)–(157) and with our choice $\mathsf{A}=\mathsf{I}$ we now get from (25) that

$$\begin{aligned}I(Q;W) &\le -n_\mathrm{R} - n_\mathrm{R}\mathsf{E}_Q\left[\log(\|\boldsymbol{X}\|^2+\sigma^2)\right]\\ &\quad -\log\Gamma(n_\mathrm{R}) + \alpha\log\beta + \log\Gamma(\alpha,\delta/\beta)\\ &\quad + (n_\mathrm{R}-\alpha)\big(\mathsf{E}_Q[\log(\|\boldsymbol{X}\|^2+\sigma^2)] + \psi(n_\mathrm{R})\big)\\ &\quad + \frac{n_\mathrm{R}(\mathcal{E}_\mathrm{s}+\sigma^2)}{\beta} + \frac{\delta}{\beta}\\ &\quad + (1-\alpha)\mathsf{E}\left[\log\left(1+\frac{\delta}{\|\boldsymbol{Y}\|^2}\right)\right] \qquad (158)\\ &\le -n_\mathrm{R} + n_\mathrm{R}\psi(n_\mathrm{R}) - \log\Gamma(n_\mathrm{R})\\ &\quad + \alpha\log\beta + \log\Gamma(\alpha,\delta/\beta) - \alpha\log\sigma^2\\ &\quad - \alpha\psi(n_\mathrm{R}) + \frac{n_\mathrm{R}(\mathcal{E}_\mathrm{s}+\sigma^2)}{\beta} + \frac{\delta}{\beta}\\ &\quad + \frac{\delta(1-\alpha)}{\sigma^2(n_\mathrm{R}-1)}\cdot I\{\alpha\le 1\}, \qquad n_\mathrm{R}>1 \quad (159)\end{aligned}$$

where for the second inequality we upper-bounded

$$-\alpha\mathsf{E}_Q\left[\log(\|\boldsymbol{X}\|^2+\sigma^2)\right]\le -\alpha\log\sigma^2, \qquad \alpha>0.$$

Note that, for high SNR, the optimal values of α and δ tend to zero. Therefore, in spite of the rather crude bound (153), we will get the correct asymptotic behavior. Similarly, in the low-SNR regime, the optimal choice for α tends to n_R and the optimal value of δ tends to zero. Therefore, also in the low-SNR regime we get the correct asymptotic behavior. See also Note 4.1.

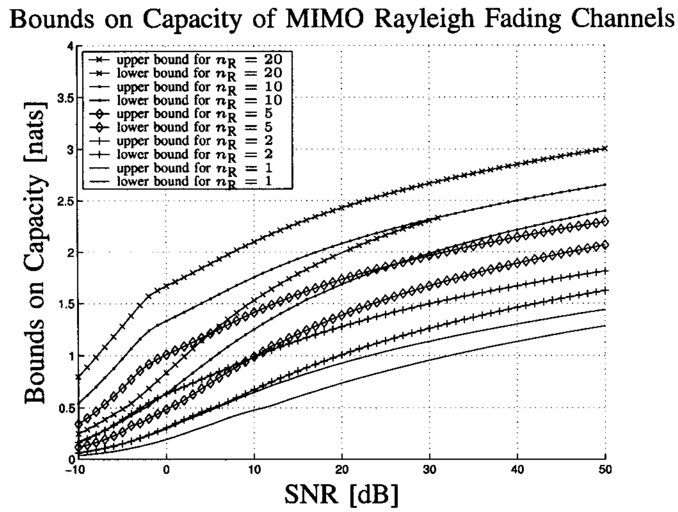

Fig. 2. The proposed upper bound (151) on the capacity of MIMO Rayleigh-fading channel for various numbers of receive antennas $n_\mathrm{R}>1$. The upper bound for $n_\mathrm{R}=1$ is taken from (141). For reference, a generalization[8] of the lower bound of [14] is also depicted.

The bound (151) now follows from (159) upon substituting

$$\delta' = \frac{\delta}{\sigma^2}, \qquad \beta' = \frac{\beta}{\sigma^2}. \quad (160)$$

Note that when further substituting

$$\eta = \frac{\delta'}{\beta'} \quad (161)$$

we can express the optimal value of β' in terms of α and η

$$\beta'^{*} = \begin{cases}\frac{n_\mathrm{R}-1}{2\eta(1-\alpha)}\left(\sqrt{\alpha^2+4\eta\,\frac{n_\mathrm{R}}{n_\mathrm{R}-1}\,(1-\alpha)(1+\mathrm{SNR})}-\alpha\right), & \\ \qquad 0<\alpha\le 1 & \\ \dfrac{n_\mathrm{R}(1+\mathrm{SNR})}{\alpha}, \quad 1<\alpha\le n_\mathrm{R}. & \end{cases} \quad (162)$$

Fig. 2 depicts the upper bound (151) for various values of the number of receive antennas n_R. The upper bound for $n_\mathrm{R}=1$ is taken from Section IV-F-1, (141). For reference, we also plot lower bounds that extend the bounds of [14][8] to the case $n_\mathrm{R}>1$. Note that the number of transmit antennas n_T does not influence channel capacity.

Again, it is of interest to compare the channel capacity in the absence of receiver side information to capacity in its presence. The latter was computed by Telatar [17]. Here we only consider the case where $n_\mathrm{T}=1$ so that $\mathbb{H}$ is a random vector, which we denote by $\boldsymbol{H}$, whose n_R components are i.i.d. $\mathcal{N}_\mathbb{C}(0,1)$. For this case, we have that the capacity in the presence of receiver side information is given by

$$\begin{aligned}C_\mathrm{PSI}(\mathrm{SNR}) &= \mathsf{E}\left[\log\left(1+\|\boldsymbol{H}\|^2\,\frac{\mathcal{E}_\mathrm{s}}{\sigma^2}\right)\right] \qquad (163)\\ &= \log\mathrm{SNR} + \mathsf{E}\left[\log\|\boldsymbol{H}\|^2\right] + o(1) \qquad (164)\\ &= \log\mathrm{SNR} + \psi(n_\mathrm{R}) + o(1), \qquad n_\mathrm{T}=1 \quad (165)\end{aligned}$$

[8]Taricco and Elia only consider single-antenna systems. The idea of considering point mass distributions of equal weights at a finite geometric series of locations can, however, be extended also to our scenario.

where the function $\psi(\cdot)$ is defined in (213), the $o(1)$ terms tends to zero as the SNR tends to infinity, and the calculation of $\mathsf{E}[\log \|\boldsymbol{H}\|^2]$ is based on (209).

3) Ricean Fading Channel: We next address the memoryless SISO Ricean fading channel, which corresponds to the channel (20) with the fading matrix being a $\mathcal{N}_{\mathbb{C}}(d, 1)$ scalar. Here, the mean $d \in \mathbb{C}$ is a deterministic constant that is often called the specular component. We shall upper-bound channel capacity using (25) and lower-bound it using Corollary 4.20 and the generalized mutual information (GMI) [9]. The former lower bound is useful at high SNR, whereas the latter is preferable at low SNR. We begin with the upper bound

$$C \leq \inf_{\substack{0<\alpha\leq 1 \\ \beta'>0,\, \delta'\geq 0}} \left\{ -1 + \alpha \log \beta' + \log \Gamma\left(\alpha, \frac{\delta'}{\beta'}\right) + \frac{\delta'}{\beta'} + \frac{1+\rho}{\beta'} + (1-\alpha)\left(\log \frac{|d|^2\rho}{\rho+|d|^2+1} - \mathrm{Ei}\left(-\frac{|d|^2\rho}{\rho+|d|^2+1}\right) + \log \delta' - e^{\delta'}\mathrm{Ei}(-\delta') + \gamma\right)\right\} \tag{166}$$

where γ denotes Euler's constant and where we introduced the output SNR

$$\rho = \frac{(|d|^2+1)\mathcal{E}_{\mathrm{s}}}{\sigma^2} \tag{167}$$

i.e., the ratio of received signal power to received noise power. This bound is shown in Fig. 3 for different values of the specular component d.

To derive this bound using (25) we note that conditional on $X = x$, the output Y has an $\mathcal{N}_{\mathbb{C}}(d\cdot x,\, |x|^2+\sigma^2)$ distribution so that

$$h(Y|X=x) = \log \pi + 1 + \log(|x|^2+\sigma^2) \tag{168}$$

$$\mathsf{E}[|Y|^2|X=x] = |x|^2 + \sigma^2 + |d|^2|x|^2. \tag{169}$$

The additional term we need for the computation of (25) is $\mathsf{E}[\log(|Y|^2+\delta)|X=x]$. It can be upper-bounded by

$$\begin{aligned} &\mathsf{E}[\log(|Y|^2+\delta)|X=x] \\ &\quad \leq \mathsf{E}[\log|Y|^2|X=x] + \sup_x \left\{\mathsf{E}[\log(|Y|^2+\delta)|X=x] - \mathsf{E}[\log|Y|^2|X=x]\right\} \\ &\quad = \mathsf{E}[\log|Y|^2|X=x] + \mathsf{E}\left[\log\left(1+\frac{\delta}{|Y|^2}\right)\middle| X=0\right] \\ &\quad = \mathsf{E}[\log|Y|^2|X=x] + \log\frac{\delta}{\sigma^2} - e^{\delta/\sigma^2}\mathrm{Ei}\left(-\frac{\delta}{\sigma^2}\right) + \gamma \end{aligned} \tag{170}$$

where, by (209)

$$\mathsf{E}[\log|Y|^2|X=x] = \log\left(\frac{|d|^2|x|^2}{|x|^2+\sigma^2}\right) - \mathrm{Ei}\left(-\frac{|d|^2|x|^2}{|x|^2+\sigma^2}\right) + \log(|x|^2+\sigma^2). \tag{171}$$

Here, the first equality follows by (215) because the function $\xi \mapsto \log(1+\xi) - \log(\xi)$ is monotonically decreasing and because the distribution of Y conditional on $X = x$ is stochastically larger than the distribution of Y conditional on $X = 0$. Indeed, by (219), the distribution $|\mathcal{N}_{\mathbb{C}}(dx,\, |x|^2+\sigma^2)|^2$ is stochastically larger than the distribution $|\mathcal{N}_{\mathbb{C}}(0,\, |x|^2+\sigma^2)|^2$, and, by a scaling argument, the latter is stochastically larger than the distribution $|\mathcal{N}_{\mathbb{C}}(0,\, \sigma^2)|^2$. The second equality follows by a direct calculation [15, 4.337 (2)].

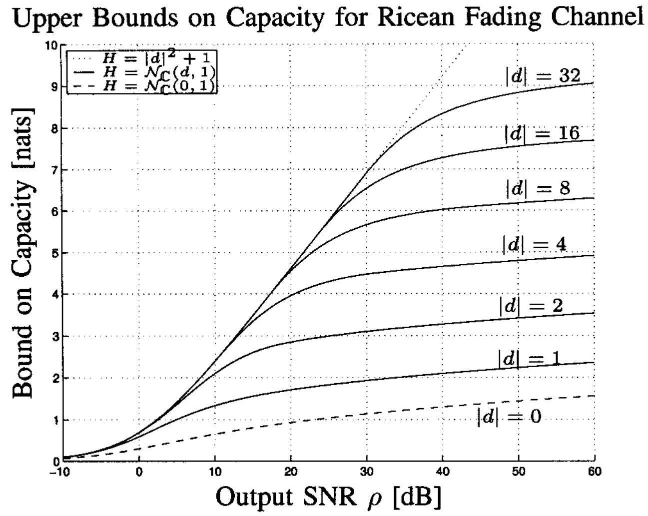

Fig. 3. The upper bound (166) on the capacity of a Ricean fading channel for different values of the specular component d. The dotted line depicts the capacity of a Gaussian channel of equal output SNR, namely, $\log(1+\rho)$.

Using (25) we thus obtain

$$\begin{aligned} I(Q;W) &\leq -1 + \alpha\log\beta + \log\Gamma(\alpha, \delta/\beta) - \alpha\mathsf{E}_Q\left[\log(|X|^2+\sigma^2)\right] + \frac{(1+|d|^2)\mathcal{E}_{\mathrm{s}}+\sigma^2}{\beta} + \frac{\delta}{\beta} \\ &\quad + (1-\alpha)\mathsf{E}_Q\left[\log\left(\frac{|d|^2|X|^2}{|X|^2+\sigma^2}\right) - \mathrm{Ei}\left(-\frac{|d|^2|X|^2}{|X|^2+\sigma^2}\right)\right] \\ &\quad + (1-\alpha)\left(\log\frac{\delta}{\sigma^2} - e^{\delta/\sigma^2}\mathrm{Ei}\left(-\frac{\delta}{\sigma^2}\right) + \gamma\right) \qquad (172) \\ &\leq -1 + \alpha\log\frac{\beta}{\sigma^2} + \log\Gamma\left(\alpha, \frac{\delta}{\beta}\right) + \frac{(1+|d|^2)\mathcal{E}_{\mathrm{s}}+\sigma^2}{\beta} \\ &\quad + \frac{\delta}{\beta} + (1-\alpha)\left(\log\left(\frac{|d|^2\mathcal{E}_{\mathrm{s}}}{\mathcal{E}_{\mathrm{s}}+\sigma^2}\right) - \mathrm{Ei}\left(-\frac{|d|^2\mathcal{E}_{\mathrm{s}}}{\mathcal{E}_{\mathrm{s}}+\sigma^2}\right) + \log\frac{\delta}{\sigma^2} - e^{\delta/\sigma^2}\mathrm{Ei}\left(-\frac{\delta}{\sigma^2}\right) + \gamma\right), \\ &\qquad 0<\alpha\leq 1,\ \beta>0,\ \delta\geq 0. \qquad (173) \end{aligned}$$

Here, the second inequality follows upon additionally restricting α so that $1-\alpha \geq 0$; upon applying Jensen's inequality to the concave function $g_1(\cdot)$ (see (210)); and upon upper-bounding $-\log(|x|^2+\sigma^2)$ by $-\log(\sigma^2)$. The inequality (166) now follows from (173) using the substitutions (148) and (167).

At the cost of some slackness at high SNR, the bound (166) can be simplified by choosing $\delta' = 0$ and $\beta' = (1+\rho)/\alpha$. This leads to the simplified bound

$$C \leq \inf_{0<\alpha\leq 1}\left\{-1 + \log\Gamma(\alpha) + \alpha(1+\log(1+\rho)) - \alpha\log\alpha + (1-\alpha)\left(\log\frac{|d|^2\rho}{\rho+|d|^2+1} - \mathrm{Ei}\left(-\frac{|d|^2\rho}{\rho+|d|^2+1}\right)\right)\right\}. \tag{174}$$

We now turn to lower-bound channel capacity. At low SNR, we consider the suboptimal signaling scheme where the input distribution is $\mathcal{N}_{\mathbb{C}}(0, \mathcal{E}_s)$ and where the receiver perfoms nearest neighbor decoding. The generalized mutual information (GMI) [9] is a lower bound on the achievable rates under these additional restrictions and is thus also a lower bound on the capacity without these restrictions. For the case at hand, the GMI I_{GMI} is given by

$$I_{\text{GMI}} = \log\left(1 + \frac{|d|^2 \mathcal{E}_s}{\sigma^2 + \mathcal{E}_s}\right) \tag{175}$$

$$= \log\left(1 + \frac{|d|^2 \rho}{\rho + |d|^2 + 1}\right). \tag{176}$$

Here, the first equality follows from [9, Corollary 3.0.1] (by substituting $Re^{j\Phi_k(1)} = d$), and the second equality follows from our definition of ρ as the output SNR; see (167).

Since the RHS of (175) is bounded in the SNR, it is apparent that this bound is quite useless at high SNR. This boundedness has nothing to do with the structure of the decoder. It is a direct consequence of using the suboptimal $\mathcal{N}_{\mathbb{C}}(0, \mathcal{E}_s)$ input distribution; see Theorem 4.3.

At high SNR, a better bound is the bound of Corollary 4.20. This bound is tight in the sense that at high SNR it achieves the fading number of the Ricean channel. It can be rewritten as

$$C \geq C_\ell = \sup_{0<\alpha\leq 1}\left\{\log(-\text{Ei}(-\alpha\beta)) + \beta - \log\left(1 + \frac{|d|^2+1}{\alpha\rho}\right)\right\} - 1 + \log|d|^2 - \text{Ei}(-|d|^2) \tag{177}$$

where β is the solution to the equation

$$\beta = \frac{e^{-\alpha\beta}}{-\text{Ei}(-\alpha\beta)}. \tag{178}$$

These two lower bounds can be combined to yield

$$C \geq \max\{I_{\text{GMI}}, C_\ell\}. \tag{179}$$

As a matter of fact, by a time-sharing argument, one can show that this lower bound can be improved to the convex hull of the maximum.

Again, the difference between the high-SNR channel capacity in the absence of side information

$$C = \log\log\frac{\mathcal{E}_s}{\sigma^2} - 1 + \log|d|^2 - \text{Ei}(-|d|^2) + o(1) \tag{180}$$

and in its presence is striking. The latter is given by

$$\begin{aligned} C_{\text{PSI}} &= \mathsf{E}\left[\log\left(1 + \frac{|H|^2 \mathcal{E}_s}{\sigma^2}\right)\right] \\ &= \mathsf{E}\left[\log\left(\frac{|H|^2 \mathcal{E}_s}{\sigma^2}\right)\right] + o(1) \\ &= \log\frac{(|d|^2+1)\mathcal{E}_s}{\sigma^2} - \log\frac{|d|^2+1}{|d|^2} - \text{Ei}(-|d|^2) + o(1) \end{aligned} \tag{181}$$

where the last equation follows from the expression for the expected logarithm of a noncentral chi-square random variable (209).

Fig. 4 shows the situation for a specular component $d = 8$. It depicts the upper bound (166) and the lower bound (179). For reference, we also plot the cruder but simpler upper bound that results from choosing $\delta' = 0$, see (174); the capacity $\log(1 + \text{SNR})$ of an additive white Gaussian noise channel of equal output SNR; and the fading number. It is seen that at rates that are significantly higher than the fading number, communication becomes extremely power inefficient.

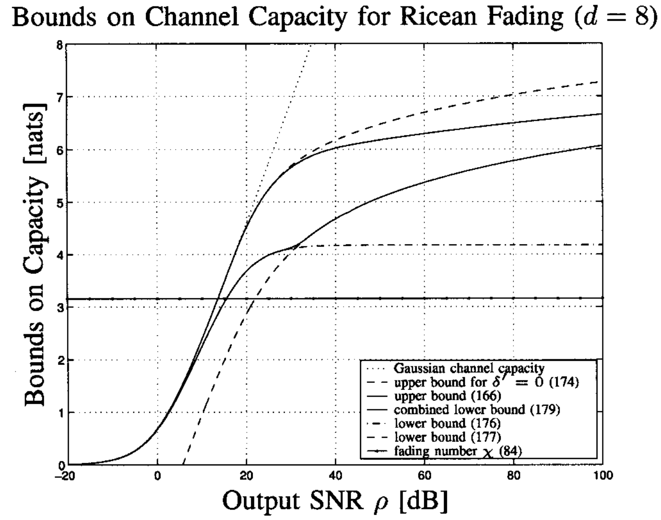

Fig. 4. Bounds on the capacity of a Ricean fading channel with specular component $d = 8$: the tighter upper bound is given in (166) and the simplified upper bound in (174); the lower bound is given in (179). For comparison, the channel capacity of a Gaussian channel is shown as dotted line. Note that on the abscissa ρ denotes the output SNR (167).

4) Multiple-Antenna Gaussian Fading Channel: We finally treat the more general case of a fading channel with n_T transmit and n_R receive antennas

$$\boldsymbol{Y} = \mathbb{H}\boldsymbol{x} + \mathsf{D}\boldsymbol{x} + \boldsymbol{Z}$$

where $\mathbb{H}$ is an $n_R \times n_T$ matrix with each entry i.i.d. $\sim \mathcal{N}_{\mathbb{C}}(0, 1)$, D is a constant $n_R \times n_T$ matrix, and $\boldsymbol{Z} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma^2 \cdot \mathsf{I}_{n_R})$.

The upper bound is given as

$$\begin{aligned} C \leq \inf_{0<\alpha\leq n_R}\Bigg\{ &-n_R - \log\Gamma(n_R) + (n_R - \alpha) g_{n_R}\left(\frac{\delta_{\max}^2 \mathcal{E}_s}{\mathcal{E}_s + \sigma^2}\right) \\ &+ \alpha\left(1 + \log\frac{n_R(\mathcal{E}_s + \sigma^2) + \delta_{\max}^2 \mathcal{E}_s}{\sigma^2}\right) \\ &+ \log\Gamma(\alpha) - \alpha\log\alpha\Bigg\} \end{aligned} \tag{182}$$

where $\delta_{\max}$ is the maximum singular value of D, and where the function $g_m(\cdot)$ is defined in Appendix X.

This bound is based on (27), using

$$h(\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x}) = n_R \log\left(\pi e(\|\boldsymbol{x}\|^2 + \sigma^2)\right) \tag{183}$$

$$\mathsf{E}[\|\boldsymbol{Y}\|^2 | \boldsymbol{X} = \boldsymbol{x}] = n_R(\|\boldsymbol{x}\|^2 + \sigma^2) + \|\mathsf{D}\boldsymbol{x}\|^2 \tag{184}$$

$$\mathsf{E}[\log\|\boldsymbol{Y}\|^2 | \boldsymbol{X} = \boldsymbol{x}] = \log(\|\boldsymbol{x}\|^2 + \sigma^2) + g_{n_R}\left(\frac{\|\mathsf{D}\boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^2 + \sigma^2}\right). \tag{185}$$

Further note that

$$\begin{aligned} \mathsf{E}_{Q^*}\left[\|\mathsf{D}\boldsymbol{X}\|^2\right] &= \mathsf{E}_{Q^*}\left[\text{tr}\left((\mathsf{D}\boldsymbol{X})(\mathsf{D}\boldsymbol{X})^\dagger\right)\right] \\ &= \text{tr}\left(\mathsf{D}\mathsf{K}_{\boldsymbol{X}}\mathsf{D}^\dagger\right) \\ &= \text{tr}\left(\mathsf{D}^\dagger\mathsf{D}\mathsf{K}_{\boldsymbol{X}}\right) \end{aligned}$$

or

$$\begin{aligned}\mathsf{E}_{Q^*}\left[\|\mathsf{D}\boldsymbol{X}\|^2\right] &\le \mathsf{E}_{Q^*}\left[\|\mathsf{D}\|^2 \cdot \|\boldsymbol{X}\|^2\right]\\ &= \mathsf{E}_{Q^*}\left[\delta_{\max}^2 \cdot \|\boldsymbol{X}\|^2\right]\\ &= \delta_{\max}^2 \cdot \mathcal{E}_{\mathrm{s}}\end{aligned}$$

where $\delta_{\max}$ is the maximum singular value of D.

V. The Basic Inequality

In this section, we extend (11) to channels over infinite alphabets. As noted earlier, the finite-alphabet version of this bound follows directly from the identity (10), which can be found, for example, in [2] and [1, Sec. 2.3 (3.7)]. In fact, the inequality (11) also appears in [18, Exercise 4.17], except that there the distribution $R(\cdot)$ is required to correspond to some input distribution, i.e., to be of the form $R(\cdot) = (\tilde{Q}W)(\cdot)$. This restriction complicates things a great deal when dealing with infinite alphabets, but is fortunately superfluous.

It should be noted that identity (10) plays a key role in the capacity-redundancy theorem of universal coding. See, for example, [10] and references therein. For the related infinite-alphabet universal source coding problem see [19].

Theorem 5.1: Let the input alphabet $\mathcal{X}$ and the output alphabet $\mathcal{Y}$ of a channel $W(\cdot|\cdot)$ be separable metric spaces, and assume that for any Borel set $\mathcal{B} \subset \mathcal{Y}$ the mapping $x \mapsto W(\mathcal{B}|x)$ from $\mathcal{X}$ to [0, 1] is Borel measurable. Let $Q(\cdot)$ be any probability measure on $\mathcal{X}$, and $R(\cdot)$ any probability measure on $\mathcal{Y}$. Then, the mutual information $I(Q;W)$ can be bounded by

$$\boxed{I(Q;W) \le \int D\left(W(\cdot|x)\|R(\cdot)\right) \mathrm{d}Q(x).} \tag{186}$$

Here, for any $x \in \mathcal{X}$, the term $D(W(\cdot|x)\|R(\cdot))$ denotes the relative entropy between the measure $W(\cdot|x)$ on $\mathcal{Y}$ and the measure $R(\cdot)$ on $\mathcal{Y}$, i.e.,

$$D\left(W(\cdot|x)\|R(\cdot)\right) = \begin{cases}\int \log \frac{\mathrm{d}W(\cdot|x)}{\mathrm{d}R(\cdot)}\,\mathrm{d}W(\cdot|x), & \text{if } W(\cdot|x) \ll R(\cdot)\\ +\infty, & \text{otherwise.}\end{cases}$$

Proof: To prove the theorem it suffices to consider the case where the output alphabet $\mathcal{Y}$ is finite. Indeed, we could treat the more general case by a limiting argument applied to successively finer output quantization. With successively finer sequence of partitions that generate the σ-algebra on $\mathcal{Y}$, the mutual information between the input and the quantized output will converge to the unquantized mutual information, and the RHS of (186) will converge by the monotonicity of the relative entropy with respect to partition refinements and the Monotone Convergence Theorem. See [20, Theorem (9.15) Part (i) on p. 261] for some of the needed supporting theorems. Henceforth, we shall therefore assume that the output alphabet $\mathcal{Y}$ is finite, i.e.,

$$|\mathcal{Y}| < \infty. \tag{187}$$

If $D(W(\cdot|x)\|R(\cdot))$ is not Q-a.s. finite, then, by the nonnegativity of relative entropy, the RHS of (186) is $+\infty$ and the claim is proved. We shall thus consider now the case where

$$D\left(W(\cdot|x)\|R(\cdot)\right) < \infty, \qquad Q\text{-a.s.} \tag{188}$$

i.e.,

$$R(y) = 0 \implies W(y|x) = 0 \qquad Q\text{-a.s.} \tag{189}$$

The measurability assumption on the channel allows us to define

$$(QW)(y) = \int W(y|x)\,\mathrm{d}Q(x) \tag{190}$$

which, in view of (189), demonstrates that

$$R(y) = 0 \implies (QW)(y) = 0. \tag{191}$$

Also, by (190)

$$(QW)(y) = 0 \implies W(y|x) = 0 \qquad Q\text{-a.s.} \tag{192}$$

Since $\mathcal{Y}$ is now assumed finite, we can rewrite the RHS of (186) as

$$\int D\left(W(\cdot|x)\|R(\cdot)\right)\,\mathrm{d}Q(x) = \int \sum_{y\in\mathcal{Y}} W(y|x)\log\frac{W(y|x)}{R(y)}\,\mathrm{d}Q(x)$$

where we define

$$0\log\frac{0}{\alpha} = 0, \qquad \alpha \ge 0 \tag{193}$$

and all the terms in the sum are Q-a.s. finite by (189).

We now note that

$$W(y|x)\log\frac{W(y|x)}{R(y)} = W(y|x)\log\frac{W(y|x)}{(QW)(y)} + W(y|x)\log\frac{(QW)(y)}{R(y)}, \qquad Q\text{-a.s.}$$

This follows from (191)–(193) whenever $W(y|x) = 0$ (whence all the terms are zero) and from the contrapositives of (189) and (192) whenever $W(y|x) > 0$. Consequently

$$\int D(W(\cdot|x)\|R(\cdot))\,\mathrm{d}Q(x) = \int\left(\sum_{y\in\mathcal{Y}} W(y|x)\log\frac{W(y|x)}{(QW)(y)} + \sum_{y\in\mathcal{Y}} W(y|x)\log\frac{(QW)(y)}{R(y)}\right)\mathrm{d}Q(x). \tag{194}$$

By [21, p. 1728 (2.10)]

$$\int \sum_{y\in\mathcal{Y}} W(y|x)\log\frac{W(y|x)}{(QW)(y)}\,\mathrm{d}Q(x) = I(Q;W).$$

It remains to show that the integral of the additional terms on the RHS of (194) is nonnegative. This follows because for each $y \in \mathcal{Y}$

$$\int W(y|x)\log\frac{(QW)(y)}{R(y)}\,\mathrm{d}Q(x) = (QW)(y)\log\frac{(QW)(y)}{R(y)}, \qquad y \in \mathcal{Y}$$

so that

$$\begin{aligned}\int \sum_{y\in\mathcal{Y}} W(y|x)\log\frac{(QW)(y)}{R(y)}\,\mathrm{d}Q(x) &= \sum_{y\in\mathcal{Y}} (QW)(y)\log\frac{(QW)(y)}{R(y)}\\ &= D((QW)(\cdot)\|R(\cdot))\\ &\ge 0.\end{aligned}$$

□

VI. Mathematical Preliminaries

In this section, we present some of the mathematical tools and results that are needed for the analysis of the capacity of flat-fading channels. This section may be glanced over in first reading and referred to as needed.

A. Some Useful Distributions

In this subsection, we introduce some of the distributions and special functions that appear in this work.

By a zero-mean unit-variance circularly symmetric Gaussian distribution denoted $\mathcal{N}_{\mathbb{C}}(0, 1)$ we shall refer to the distribution on the complex field $\mathbb{C}$ of density

$$\frac{1}{\pi} e^{-|z|^2}, \qquad z \in \mathbb{C}.$$

More generally, we let $\mathcal{N}_{\mathbb{C}}(\mu, \sigma^2)$ denote the distribution that results when an $\mathcal{N}_{\mathbb{C}}(0, 1)$ random variable is scaled by $\sigma \geq 0$ and shifted by $\mu \in \mathbb{C}$. In the multivariate case we write $\boldsymbol{X} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \mathsf{K})$ if $\boldsymbol{X} - \boldsymbol{\mu}$ is a circularly symmetric zero-mean multivariate Gaussian random vector, i.e., if $\boldsymbol{X} - \boldsymbol{\mu}$ can be expressed as the product of a deterministic matrix and a complex random vector whose components are i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$. In particular

$$\mathsf{E}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^{\mathsf{T}}] = 0; \quad \mathsf{E}[(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^{\dagger}] = \mathsf{K}$$

where $(\cdot)^{\mathsf{T}}$ denotes the transpose operation and $(\cdot)^{\dagger}$ denotes Hermitian conjugation. See [22], [17] and references therein for additional information on circularly symmetric Gaussian variables.

If $U_1, \ldots, U_m$ are i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 2)$ then the sum $\sum_{j=1}^m |U_j|^2$ has a central chi-square distribution with $2m$ degrees of freedom—a distribution that is typically denoted χ^2_{2m} and which has the density (over the nonnegative real line) [23, Ch. 18]

$$\frac{1}{2^m \Gamma(m)} x^{m-1} e^{-x/2}, \qquad x \geq 0. \tag{195}$$

This distribution is a special case ($\alpha = m$ and $\beta = 2$) of the Gamma distribution on the nonnegative real numbers, which is of density [23, Ch. 17]

$$\frac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha} \Gamma(\alpha)}, \qquad x \geq 0, \, \alpha, \, \beta > 0. \tag{196}$$

Here $\Gamma(\alpha)$ denotes the Gamma function and is given by

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} \, \mathrm{d}t, \qquad \alpha > 0. \tag{197}$$

In fact, $\frac{1}{2}\sum_{j=1}^m |U_j|^2$, which corresponds to the sum of the squared-magnitudes of m independent $\mathcal{N}_{\mathbb{C}}(0, 1)$ random variables, is a special case ($\alpha = m$) of the standard Gamma distribution, which corresponds to the Gamma distribution with $\beta = 1$, i.e.,

$$\frac{x^{\alpha-1} e^{-x}}{\Gamma(\alpha)}, \qquad x \geq 0, \, \alpha > 0. \tag{198}$$

See [23, Ch. 17] for additional details on the Gamma and central chi-square distributions.

For $\alpha < 1$, the density of the Gamma distribution has a singularity at the origin. This motivates us to define a *regularized Gamma distribution* on the nonnegative real line to be of the density

$$\frac{(x+\delta)^{\alpha-1} e^{-(x+\delta)/\beta}}{\beta^{\alpha} \Gamma(\alpha, \delta/\beta)}, \qquad x \geq 0, \, \alpha, \, \beta > 0, \, \delta \geq 0. \tag{199}$$

Here $\Gamma(\alpha, \xi)$ denotes the incomplete Gamma function and is given by

$$\Gamma(\alpha, \xi) = \int_{\xi}^{\infty} t^{\alpha-1} e^{-t} \, \mathrm{d}t, \qquad \alpha > 0, \, \xi \geq 0. \tag{200}$$

For $\delta = 0$, the regularized Gamma distribution (199) thus coincides with the Gamma distribution (196). For $\delta > 0$, however, the density of the regularized Gamma distribution is bounded for all values of $\alpha, \beta > 0$.

We next derive an isotropic distribution on $\mathbb{C}^{n_{\mathrm{R}}}$ under which $\|\boldsymbol{Y}\|^2$ has a regularized Gamma distribution. Here $\|\cdot\|$ denotes the Euclidean norm as in (12). We first recall that if a nonnegative real random variable W is of density $f_W(w)$ and if $V = \sqrt{W}$ then the density f_V of V is given by $f_V(v) = 2v f_W(v^2)$. Consequently, if $\|\boldsymbol{Y}\|^2$ is to have a regularized Gamma distribution (199) then $\|\boldsymbol{Y}\|$ should be of density

$$2\|\boldsymbol{y}\| \frac{(\|\boldsymbol{y}\|^2 + \delta)^{\alpha-1} e^{-(\|\boldsymbol{y}\|^2+\delta)/\beta}}{\beta^{\alpha} \Gamma(\alpha, \delta/\beta)}, \qquad \alpha, \, \beta > 0, \, \delta \geq 0. \tag{201}$$

Next we recall that the surface area of an n_{R}-dimensional complex sphere of radius r is

$$\frac{2\pi^{n_{\mathrm{R}}} r^{2n_{\mathrm{R}}-1}}{\Gamma(n_{\mathrm{R}})} \tag{202}$$

so that the density of the isotropic distribution on $\mathbb{C}^{n_{\mathrm{R}}}$ under which $\|\boldsymbol{Y}\|^2$ is of a regularized Gamma distribution is the ratio of (201) to (202), namely

$$\frac{\|\boldsymbol{y}\|^{2(1-n_{\mathrm{R}})} (\|\boldsymbol{y}\|^2 + \delta)^{(\alpha-1)} e^{-(\|\boldsymbol{y}\|^2+\delta)/\beta} \Gamma(n_{\mathrm{R}})}{\pi^{n_{\mathrm{R}}} \beta^{\alpha} \Gamma(\alpha, \delta/\beta)}. \tag{203}$$

A linear transformation on such isotropic distributions leads to the family of distributions on $\mathbb{C}^{n_{\mathrm{R}}}$ that will be of most interest to us. For any $\alpha, \beta > 0$, $\delta \geq 0$ and any nonsingular deterministic matrix $\mathsf{A} \in \mathbb{C}^{n_{\mathrm{R}} \times n_{\mathrm{R}}}$ the density on $\mathbb{C}^{n_{\mathrm{R}}}$ is

$$\frac{\|\mathsf{A}\boldsymbol{y}\|^{2(1-n_{\mathrm{R}})} (\|\mathsf{A}\boldsymbol{y}\|^2 + \delta)^{(\alpha-1)} e^{-(\|\mathsf{A}\boldsymbol{y}\|^2+\delta)/\beta} \Gamma(n_{\mathrm{R}}) |\det \mathsf{A}|^2}{\pi^{n_{\mathrm{R}}} \beta^{\alpha} \Gamma(\alpha, \delta/\beta)}. \tag{204}$$

For the tightness of the proposed bounds at low SNR, it will be important to note that this family of densities includes the family of all zero-mean circularly symmetric Gaussians on $\mathbb{C}^{n_{\mathrm{R}}}$ with nonsingular covariance matrices. Indeed, such Gaussians are obtained by setting, $\alpha = n_{\mathrm{R}}$, $\delta = 0$, and $\beta = 1$.

A nonnegative real random variable is said to have a *noncentral chi-square* distribution with n *degrees of freedom* and *noncentrality parameter* s^2 if it is distributed like

$$\sum_{j=1}^{n} (X_j + \mu_j)^2 \tag{205}$$

where $\{X_j\}_{j=1}^n$ are i.i.d. $\mathcal{N}_{\mathbb{R}}(0, 1)$ and $\{\mu_j\}_{j=1}^n$ satisfy

$$s^2 = \sum_{j=1}^{n} \mu_j^2. \tag{206}$$

(The distribution of (205) depends on the constants $\{\mu_j\}$ only via the sum of their squares.) The pdf of such a distribution is given by [24, Ch. 29]

$$\frac{1}{2}\left(\frac{x}{s^2}\right)^{\frac{n-2}{4}} e^{-\frac{s^2+x}{2}} I_{n/2-1}\left(s\sqrt{x}\right), \qquad x \geq 0. \tag{207}$$

Here $I_\nu(\cdot)$ denotes the modified Bessel function of the first kind of order $\nu \in \mathbb{R}$, i.e.,

$$I_\nu(x) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(\nu+k+1)}\left(\frac{x}{2}\right)^{\nu+2k}, \qquad x \geq 0 \tag{208}$$

(see [15, eq. 8, 445]).

If the number of degrees of freedom n is even, i.e., if $n = 2m$ for some nonnegative integer m, then the noncentral chi-square distribution can also be expressed as a sum of the squared-norms of *complex* Gaussians.

We send forward the following expression for the expected logarithm of scaled noncentral chi-square distributed random variables: Let $\{U_j\}_{j=1}^m$ be i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$, let $\{\mu_j\}_{j=1}^m$ be arbitrary complex constants, and let λ be a nonzero complex constant. Then

$$\mathsf{E}\left[\log\left(\sum_{j=1}^{m} |\lambda U_j + \mu_j|^2\right)\right] = \log(|\lambda|^2) + g_m\left(\frac{\sum_{j=1}^{m} |\mu_j|^2}{|\lambda|^2}\right) \tag{209}$$

where

$$g_m(x) = \log(x) - \mathrm{Ei}(-x) + \sum_{j=1}^{m-1}(-1)^j\left(e^{-x}(j-1)! - \frac{(m-1)!}{j(m-1-j)!}\right)x^{-j} \tag{210}$$

and $\mathrm{Ei}(\cdot)$ denotes the exponential integral function defined as

$$\mathrm{Ei}(-x) = -\int_x^\infty \frac{e^{-t}}{t}\,\mathrm{d}t, \qquad x > 0. \tag{211}$$

For future reference we note here that the function $g_m(\cdot)$ is a monotonically increasing concave function with

$$g_m(0) = \psi(m) \tag{212}$$

where $\psi(m)$ is given by

$$\psi(m) = -\gamma + \sum_{j=1}^{m-1}\frac{1}{j} \tag{213}$$

and γ denotes Euler's constant.

See Appendix X for a derivation of this expectation and the properties of $g_m(\cdot)$.

B. Stochastic Ordering

Here we recall some of the basic definitions related to stochastic ordering. Only the univariate real case will be addressed. For more on stochastic ordering please refer to [25], [26], [8] and references therein. The following definitions and statements can be found, for example, in [8, Sec. 1.A].

Definition 6.1: Let F_1 and F_2 be two distributions on the real line. We shall say that F_1 is stochastically larger than (or equal to) F_2 and write

$$F_1 \geq^{\mathrm{st}} F_2$$

if the following equivalent conditions hold.

- For any $-\infty < \zeta < +\infty$, the probability of the half ray $(\zeta, +\infty)$ under the law F_1 is at least as large as the probability of that ray under F_2, i.e.,
$$F_1((\zeta, +\infty)) \geq F_2((\zeta, +\infty)), \qquad -\infty < \zeta < +\infty. \tag{214}$$
- For any increasing functions ϕ
$$\int \phi(x)\,dF_1(x) \geq \int \phi(x)\,\mathrm{d}F_2(x), \qquad \phi \text{ is increasing} \tag{215}$$
whenever the expectations exist.
- There exists a probability space with two random variables X, Y such that X is F_1 distributed, Y is F_2 distributed, and $X \geq Y$, almost surely, i.e.,
$$X \sim F_1, \quad Y \sim F_2, \quad X \geq Y \text{ a.s.} \tag{216}$$

Slightly abusing notation we shall sometimes write

$$X \geq^{\mathrm{st}} Y$$

for two real random variables X, Y to indicate that the distribution functions F_X of X is stochastically larger than (or equal to) the distribution function F_Y of Y. With this notation it follows from (216) that

- if $X \geq^{\mathrm{st}} Y$ then for any increasing function ϕ
$$\phi(X) \geq^{\mathrm{st}} \phi(Y);$$
- also, if $X_1 \geq^{\mathrm{st}} Y_1$; $X_2 \geq^{\mathrm{st}} Y_2$; the pair (X_1, X_2) is independent; and the pair (Y_1, Y_2) is independent, then
$$X_1 + X_2 \geq^{\mathrm{st}} Y_1 + Y_2. \tag{217}$$

The main result we need is stated in the following lemma (see also [27, Sec. 4.2.2]):

Lemma 6.2: The following claims demonstrate stochastic ordering for some specific distributions.

a) Let the real random variable X have a continuous strictly unimodal symmetric density f_X, i.e.,
$$f_X(x) = f_X(-x), \qquad x \in \mathbb{R}$$
and
$$f_X(x') > f_X(x'') \;\Leftrightarrow\; |x'| < |x''|, \qquad x', x'' \in \mathbb{R}.$$
Let $\mu_1, \mu_2 \in \mathbb{R}$ be deterministic. Then
$$(X+\mu_1)^2 \geq^{\mathrm{st}} (X+\mu_2)^2, \qquad |\mu_1| \geq |\mu_2|. \tag{218}$$

b) Suppose $X_1 \sim \mathcal{N}(\mu_1, \sigma^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma^2)$. Then $|\mu_1| \geq |\mu_2|$ implies $|X_1|^2 \geq^{\mathrm{st}} |X_2|^2$, or stated a little more sloppily
$$|\mathcal{N}(\mu_1, \sigma^2)|^2 \geq^{\mathrm{st}} |\mathcal{N}(\mu_2, \sigma^2)|^2, \qquad |\mu_1| \geq |\mu_2|. \tag{219}$$

c) For any mean vector $\boldsymbol{\mu}$ and any covariance matrix K

$$\|\mathcal{N}(\boldsymbol{\mu}, \mathsf{K})\|^2 \geq^{\mathrm{st}} \|\mathcal{N}(\mathbf{0}, \mathsf{K})\|^2. \tag{220}$$

d) Let $\boldsymbol{W} \sim \mathcal{N}(\mathbf{0}, \mathsf{K})$ and let $\boldsymbol{V}$ be independent of $\boldsymbol{W}$. Then

$$\|\boldsymbol{W}+\boldsymbol{V}\|^2 \geq^{\mathrm{st}} \|\boldsymbol{W}\|^2.$$

e) Let K_1 and K_2 be two nonnegative definite matrices. Then

$$\|\mathcal{N}(\mathbf{0}, \mathsf{K}_1+\mathsf{K}_2)\|^2 \geq^{\mathrm{st}} \|\mathcal{N}(\mathbf{0}, \mathsf{K}_1)\|^2. \tag{221}$$

Proof:

a) The symmetry of the distribution of X allows us to assume, without loss of generality

$$\mu_1 \geq \mu_2 \geq 0.$$

For any $\zeta > 0$ we have

$$\begin{aligned}\Pr\left((X+\mu_1)^2 \geq \zeta^2\right) &= \Pr\left(X+\mu_1 \geq \zeta\right) + \Pr\left(X+\mu_1 \leq -\zeta\right)\\ &= \Pr\left(X \geq \zeta-\mu_1\right) + \Pr\left(X \geq \zeta+\mu_1\right)\end{aligned}$$

where we have used the symmetry of the distribution of X in the second equality. Similarly

$$\Pr\left((X+\mu_2)^2 \geq \zeta^2\right) = \Pr(X \geq \zeta-\mu_2) + \Pr(X \geq \zeta+\mu_2).$$

Consequently

$$\Pr\left((X+\mu_1)^2 \geq \zeta^2\right) - \Pr\left((X+\mu_2)^2 \geq \zeta^2\right) = \int_{\zeta-\mu_1}^{\zeta-\mu_2} f_X(x)\,\mathrm{d}x - \int_{\zeta+\mu_2}^{\zeta+\mu_1} f_X(x)\,\mathrm{d}x.$$

Both the integrals on the RHS of the above are over intervals of length $\mu_1-\mu_2$ but they differ in their centers. We shall now conclude the proof by showing that the first integral is greater or equal to the second because its center is closer to the origin.

To this end, define the function

$$g(\beta) = \int_{\beta-(\mu_1-\mu_2)/2}^{\beta+(\mu_1-\mu_2)/2} f_X(x)\,\mathrm{d}x, \qquad \beta \in \mathbb{R}.$$

The function $g(\beta)$ is thus symmetric and takes value in the interval $[0, 1]$. Expressing its derivative as

$$\frac{\mathrm{d}g}{\mathrm{d}\beta} = f_X\left(\beta+(\mu_1-\mu_2)/2\right) - f_X\left(\beta-(\mu_1-\mu_2)/2\right)$$

we conclude from the strict monotonicity of $f_X(|x|)$ that $\frac{\mathrm{d}g}{\mathrm{d}\beta}$ is zero at the origin, and negative whenever $\beta > 0$. Thus, $g(\beta)$ attains its maximum at the origin, and is monotonically decreasing for $\beta \geq 0$.

b) The real case follows immediately from Part a). For the complex case we note that $|\mathcal{N}_{\mathbb{C}}(\mu_1, \sigma^2)|^2$, being noncentral chi-square distributed, can be written as the sum of the squares of two real Gaussian random variables $\mathcal{N}_{\mathbb{R}}(0, \sigma^2/2)$ and $\mathcal{N}_{\mathbb{R}}(|\mu_1|, \sigma^2/2)$. Expressing $|\mathcal{N}_{\mathbb{C}}(\mu_2, \sigma^2)|^2$ similarly proves the claim by Part a) and (217).

c) Using a diagonalization argument we shall prove that Part c) follows from Part b) and (217). Let $\boldsymbol{W} \sim \mathcal{N}(\mathbf{0}, \mathsf{K})$. We shall demonstrate that $\|\boldsymbol{W}+\boldsymbol{\mu}\|^2 \geq^{\mathrm{st}} \|\boldsymbol{W}\|^2$. Let U be a unitary matrix that diagonalizes K so that

$$\mathsf{U}\mathsf{K}\mathsf{U}^\dagger = \mathrm{Diag}\{\lambda_j\}$$

where $\{\lambda_j\}$ are the (nonnegative) eigenvalues of K. Let $\tilde{\boldsymbol{W}} = \mathsf{U}\boldsymbol{W}$ so that the covariance matrix $\mathsf{K}_{\tilde{\boldsymbol{W}}}$ of $\tilde{\boldsymbol{W}}$ is $\mathrm{Diag}\{\lambda_j\}$ and so that the components $\{\tilde{W}_j\}$ of $\tilde{\boldsymbol{W}}$ are independent. Let $\tilde{\boldsymbol{\mu}} = \mathsf{U}\boldsymbol{\mu}$. Since U is unitary, i.e., norm-preserving

$$\begin{aligned}\|\boldsymbol{W}+\boldsymbol{\mu}\|^2 &= \|\mathsf{U}(\boldsymbol{W}+\boldsymbol{\mu})\|^2\\ &= \left\|\tilde{\boldsymbol{W}}+\tilde{\boldsymbol{\mu}}\right\|^2\\ &= \sum_j \left|\tilde{W}_j+\tilde{\mu}_j\right|^2\end{aligned} \tag{222}$$

where $\tilde{\mu}_j$ denotes the jth component of $\tilde{\boldsymbol{\mu}}$. Similarly

$$\begin{aligned}\|\boldsymbol{W}\|^2 &= \left\|\tilde{\boldsymbol{W}}\right\|^2\\ &= \sum_j \left|\tilde{W}_j\right|^2.\end{aligned} \tag{223}$$

But, by Part b) it follows that

$$\left|\tilde{W}_j+\tilde{\mu}_j\right|^2 \geq^{\mathrm{st}} \left|\tilde{W}_j\right|^2$$

and, consequently, by the independence of $\{\tilde{W}_j\}$ and by (217), it follows that

$$\sum_j \left|\tilde{W}_j+\tilde{\mu}_j\right|^2 \geq^{\mathrm{st}} \sum_j \left|\tilde{W}_j\right|^2$$

which concludes the proof by (222) and (223).

d) This follows from Part c) by a conditioning argument

$$\begin{aligned}\Pr\left(\|\boldsymbol{W}+\boldsymbol{V}\|^2 > \zeta\right) &= \int \Pr\left(\|\boldsymbol{W}+\boldsymbol{v}\|^2 > \zeta\right)\mathrm{d}F_{\boldsymbol{V}}(\boldsymbol{v})\\ &\geq \int \Pr\left(\|\boldsymbol{W}\|^2 > \zeta\right)\mathrm{d}F_{\boldsymbol{V}}(\boldsymbol{v})\\ &= \Pr\left(\|\boldsymbol{W}\|^2 > \zeta\right)\end{aligned}$$

where the inequality follows from Part c).

e) This follows from Part d) by choosing $\boldsymbol{W} \sim \mathcal{N}(\mathbf{0}, \mathsf{K}_1)$ and $\boldsymbol{V} \sim \mathcal{N}(\mathbf{0}, \mathsf{K}_2)$. □

C. The Rayleigh–Ritz Theorem

The name of Lord Rayleigh appears in this paper not only in reference to the fading distribution that is named after him, but also because the following proposition is based on the Rayleigh–Ritz characterization [28, Theorem 4.2.2] of the extremal eigenvalues of Hermitian matrices.

Proposition 6.3: Let $\mathbb{H}$ be an $n_{\mathrm{R}} \times n_{\mathrm{T}}$ complex random matrix all of whose components have a finite second moment, so that $\mathsf{E}[\|\mathbb{H}\|_{\mathrm{F}}^2] < \infty$. Denote by $\lambda_{\max}(\mathbb{H})$ ($\lambda_{\min}(\mathbb{H})$) the largest (resp., smallest) eigenvalue of the covariance matrix of the $n_{\mathrm{R}} \cdot n_{\mathrm{T}}$-random vector that results when the entries of $\mathbb{H}$ are stacked on top of each other. Let $\hat{\boldsymbol{x}} \in \mathbb{C}^{n_{\mathrm{T}}}$ be a deterministic unit-norm complex vector and denote by $\lambda_{\max}(\mathbb{H}\hat{\boldsymbol{x}})$ ($\lambda_{\min}(\mathbb{H}\hat{\boldsymbol{x}})$) the maximal (resp., minimal) eigenvalue of the covariance matrix of the random vector $\mathbb{H}\hat{\boldsymbol{x}}$. Then

$$\lambda_{\max}(\mathbb{H}\hat{\boldsymbol{x}}) \leq \lambda_{\max}(\mathbb{H}) \tag{224}$$

$$\lambda_{\min}(\mathbb{H}\hat{\boldsymbol{x}}) \geq \lambda_{\min}(\mathbb{H}). \tag{225}$$

Proof: The proofs of (224) and (225) are almost identical, so we shall only prove the latter. By the Rayleigh–Ritz theorem

[28, Theorem 4.2.2] the smallest eigenvalue of the covariance matrix of an n-dimensional random vector $\boldsymbol{W}$ can be expressed as

$$\min_{\sum_{j=1}^{n}|\alpha^{(j)}|^2=1} \mathsf{E}\left[\left|\sum_{j=1}^{n}\alpha^{(j)}W^{(j)}\right|^2\right].$$

Consequently, for any $\|\hat{\boldsymbol{x}}\|=1$, there exist coefficients $\{\alpha^{(r)}\}$ such that $\sum_{r=1}^{n_\mathrm{R}}|\alpha^{(r)}|^2=1$ and

$$\begin{aligned}\lambda_{\min}(\mathbb{H}\hat{\boldsymbol{x}}) &= \mathsf{E}\left[\left|\sum_{r=1}^{n_\mathrm{R}}\alpha^{(r)}(\mathbb{H}\hat{\boldsymbol{x}})^{(r)}\right|^2\right]\\ &= \mathsf{E}\left[\left|\sum_{r=1}^{n_\mathrm{R}}\sum_{t=1}^{n_\mathrm{T}}\alpha^{(r)}\hat{x}^{(t)}H^{(r,t)}\right|^2\right]\\ &\geq \lambda_{\min}(\mathbb{H})\end{aligned}$$

where the last inequality follows from the Rayleigh–Ritz characterization of $\lambda_{\min}(\mathbb{H})$, because the sum of squares of the $n_\mathrm{R}\cdot n_\mathrm{T}$ coefficients $\{(\alpha^{(r)}\cdot\hat{x}^{(t)})\}$ is one, but these coefficients do not necessarily achieve $\lambda_{\min}(\mathbb{H})$. □

D. Differential Entropy

In this subsection, we present some results on differential entropy. Our focus is on relationships between the differential entropy $h(\boldsymbol{X})$ of a random vector $\boldsymbol{X}$ and the expectation of the logarithm of its norm $\mathsf{E}[\log\|\boldsymbol{X}\|]$. We shall also need some results relating the differential entropy $h(\mathbb{H})$ of a random matrix $\mathbb{H}$ and the differential entropy $h(\mathbb{H}\boldsymbol{x})$ of the vector $\mathbb{H}\boldsymbol{x}$ that results when the matrix $\mathbb{H}$ is multiplied by a deterministic vector $\boldsymbol{x}$. Results on uniform continuity of differential entropy will also be presented.

1) Some Definitions: The differential entropy $h(\boldsymbol{X})$ of an n-dimensional real random vector $\boldsymbol{X}$ is defined if the density $f_{\boldsymbol{X}}(\boldsymbol{x})$ (with respect to the Lebesgue measure on $\mathbb{R}^n$) is defined and if at least one of the integrals

$$\begin{aligned}h^+(\boldsymbol{X}) &= \int_{\{\boldsymbol{x}\in\mathbb{R}^n:\,0<f_{\boldsymbol{X}}(\boldsymbol{x})<1\}} f_{\boldsymbol{X}}(\boldsymbol{x})\log\frac{1}{f_{\boldsymbol{X}}(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{x}\\ h^-(\boldsymbol{X}) &= \int_{\{\boldsymbol{x}\in\mathbb{R}^n:\,f_{\boldsymbol{X}}(\boldsymbol{x})>1\}} f_{\boldsymbol{X}}(\boldsymbol{x})\log f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\end{aligned}$$

is finite. In this case, $h(\boldsymbol{X})$ is defined as the difference between the two nonnegative integrals

$$h(\boldsymbol{X}) = h^+(\boldsymbol{X}) - h^-(\boldsymbol{X}) \tag{226}$$

where we use the rules $+\infty - a = +\infty$ and $a - \infty = -\infty$ for all $a\in\mathbb{R}$. This is written as

$$h(\boldsymbol{X}) = \int_{\mathbb{R}^n} f_{\boldsymbol{X}}(\boldsymbol{x})\log\frac{1}{f_{\boldsymbol{X}}(\boldsymbol{x})}\,\mathrm{d}\boldsymbol{x}. \tag{227}$$

The differential entropy of an n-dimensional *complex* random variable is defined as the differential entropy of the $2n$-dimensional real vector comprising of the real and imaginary parts of each of its components. Finally, the differential entropy $h(\mathbb{H})$ of a random matrix $\mathbb{H}$ is the differential entropy of the vector comprising of its entries.

2) Some Bounds and Integrability Conditions: We begin with some upper bounds on $h(\boldsymbol{X})$ and $h^+(\boldsymbol{X})$. The bound on $h(\boldsymbol{X})$ is standard. The bound on $h^+(\boldsymbol{X})$ is fairly crude, but it serves to show that a second-moment constraint not only guarantees that $h(\boldsymbol{X})$ is bounded, but also gives an upper bound on $h^+(\boldsymbol{X})$.

Lemma 6.4: Let the complex random vector $\boldsymbol{X}$ have a density $f_{\boldsymbol{X}}(\boldsymbol{x})$ with respect to the Lebesgue measure on $\mathbb{C}^n$, and assume that its norm is of finite second moment

$$\mathsf{E}\left[\|\boldsymbol{X}\|^2\right] < \infty. \tag{228}$$

Then

$$h(\boldsymbol{X}) \leq n\log\left(\pi e\mathsf{E}\left[\|\boldsymbol{X}\|^2\right]/n\right) \tag{229}$$

and

$$h^+(\boldsymbol{X}) \leq \frac{n+1}{e} + n\log^+\left(\pi e\mathsf{E}\left[\|\boldsymbol{X}\|^2\right]/n\right). \tag{230}$$

Proof: Inequality (229) is standard. Its proof relies on the fact that of all random vectors of given marginals, differential entropy is maximized by the one whose components are independent [12]; on the fact that of all complex random variables of a given second moment, differential entropy is maximized by the circularly symmetric Gaussian distribution; and by the concavity (in the variance) of the differential entropy of a circularly symmetric complex Gaussian.

We now proceed to prove inequality (230). To this end define

$$\mathcal{X}^+ = \{\boldsymbol{x}\in\mathbb{C}^n\colon f_{\boldsymbol{X}}(\boldsymbol{x})<1\} \tag{231}$$

and set

$$\alpha = \Pr(\boldsymbol{X}\in\mathcal{X}^+) = \Pr(f_{\boldsymbol{X}}(\boldsymbol{X})<1). \tag{232}$$

If $\alpha = 0$ then $h^+(\boldsymbol{X}) = 0$ and inequality (230) is satisfied. We now focus on the case $\alpha > 0$. In this case, we can express $h^+(\boldsymbol{X})$ as

$$h^+(\boldsymbol{X}) = \alpha\int_{\mathcal{X}^+}\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{\alpha}\log\frac{1}{f_{\boldsymbol{X}}(\boldsymbol{x})/\alpha}\,\mathrm{d}\boldsymbol{x} - \alpha\log\alpha. \tag{233}$$

Equation (233) thus relates $h^+(\boldsymbol{X})$ to α and to the differential entropy of the density on $\mathcal{X}^+$ given by

$$\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{\alpha}, \qquad \boldsymbol{x}\in\mathcal{X}^+. \tag{234}$$

This density on $\mathcal{X}^+$ is of second moment

$$\begin{aligned}\int_{\mathcal{X}^+}\|\boldsymbol{x}\|^2\frac{f_{\boldsymbol{X}}(\boldsymbol{x})}{\alpha}\,\mathrm{d}\boldsymbol{x} &\leq \frac{1}{\alpha}\int_{\mathbb{C}^n}\|\boldsymbol{x}\|^2 f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\ &= \alpha^{-1}\mathsf{E}\left[\|\boldsymbol{X}\|^2\right]\end{aligned}$$

so that by (229) and (233)

$$h^+(\boldsymbol{X}) \leq \alpha\cdot n\log\left(\pi e\mathsf{E}\left[\|\boldsymbol{X}\|^2\right]/n\right) - (n+1)\alpha\log\alpha. \tag{235}$$

Inequality (230) now follows by noting that $\alpha\log\alpha \geq -1/e$, for all $\alpha > 0$, and by noting that $\alpha\log x \leq \log^+ x$, for all $0\leq\alpha\leq 1$. □

We now prove a conditional version of Lemma 6.4.

Lemma 6.5: Let the random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$ satisfy $\mathsf{E}[\|\boldsymbol{X}\|^2] < \infty$ and $h(\boldsymbol{X}|\boldsymbol{Y}) > -\infty$. Let

$$h^-(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}) = \int_{\boldsymbol{x}:\, f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})>1} f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y}) \log f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{x} \tag{236}$$

$$h^-(\boldsymbol{X}|\boldsymbol{Y}) = \int_{\boldsymbol{y}} h^-(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \tag{237}$$

$$h^+(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}) = \int_{\boldsymbol{x}:\, 0<f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})<1} f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y}) \log \frac{1}{f_{\boldsymbol{X}|\boldsymbol{Y}}(\boldsymbol{x}|\boldsymbol{y})}\,\mathrm{d}\boldsymbol{x} \tag{238}$$

$$h^+(\boldsymbol{X}|\boldsymbol{Y}) = \int_{\boldsymbol{y}} h^+(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}) f_{\boldsymbol{Y}}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}. \tag{239}$$

Then the mappings $\boldsymbol{y} \mapsto h^-(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y})$ and $\boldsymbol{y} \mapsto h^+(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y})$ are nonnegative integrable mappings, i.e.,

$$h^-(\boldsymbol{X}|\boldsymbol{Y}),\, h^+(\boldsymbol{X}|\boldsymbol{Y}) < \infty. \tag{240}$$

In particular

$$h^-(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}),\, h^+(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}) < \infty, \qquad \boldsymbol{Y}\text{-a.s.} \tag{241}$$

Proof: We shall prove the results for $h^-(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y})$ and $h^-(\boldsymbol{X}|\boldsymbol{Y})$. The analogous proofs for h^+ are omitted. Using

$$h^-(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}) = h^+(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}) - h(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}) \tag{242}$$

and Lemma 6.4 we obtain

$$h^-(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}) \le \frac{n+1}{e} + n\log^+(\pi e \mathsf{E}[\|\boldsymbol{X}\|^2|\boldsymbol{Y}=\boldsymbol{y}]/n) - h(\boldsymbol{X}|\boldsymbol{Y}=\boldsymbol{y}). \tag{243}$$

Consequently

$$h^-(\boldsymbol{X}|\boldsymbol{Y}) \le \frac{n+1}{e} - h(\boldsymbol{X}|\boldsymbol{Y}) + \int_{\boldsymbol{y}:\, \mathsf{E}[\|\boldsymbol{X}\|^2|\boldsymbol{Y}=\boldsymbol{y}]\ge n/(\pi e)} f_{\boldsymbol{Y}}(\boldsymbol{y}) \cdot n\log\left(\frac{\pi e \mathsf{E}[\|\boldsymbol{X}\|^2|\boldsymbol{Y}=\boldsymbol{y}]}{n}\right)\mathrm{d}\boldsymbol{y}. \tag{244}$$

Using the upper bound $\log\xi < \xi$ and expanding the range of integration we obtain

$$h^-(\boldsymbol{X}|\boldsymbol{Y}) \le \frac{n+1}{e} - h(\boldsymbol{X}|\boldsymbol{Y}) + \pi e \mathsf{E}\left[\|\boldsymbol{X}\|^2\right] < \infty. \tag{245}$$

□

3) Random Matrices Operating on Deterministic Vectors: The next lemma will be used to exhibit a uniform lower bound on $h(\mathbb{H}\boldsymbol{x})$ in terms of $h(\mathbb{H})$. Using Lemma 6.4, an analogous conclusion about $h^-(\mathbb{H}\boldsymbol{x})$ can be drawn.

Lemma 6.6: Let the random $n_{\mathrm{R}} \times n_{\mathrm{T}}$ matrix $\mathbb{H}$ satisfy $h(\mathbb{H}) > -\infty$ and $\mathsf{E}[\|\mathbb{H}\|_{\mathrm{F}}^2] < \infty$. Then

$$\inf_{\|\hat{\boldsymbol{x}}\|=1} h(\mathbb{H}\hat{\boldsymbol{x}}) > -\infty \tag{246}$$

and

$$\sup_{\|\hat{\boldsymbol{x}}\|=1} h^-(\mathbb{H}\hat{\boldsymbol{x}}) < \infty. \tag{247}$$

Proof: Assume without loss of generality that $\mathbb{H}$ is of zero mean. Since $h(\mathbb{H}) > -\infty$, it follows that the $(n_{\mathrm{R}} \cdot n_{\mathrm{T}}) \times (n_{\mathrm{R}} \cdot n_{\mathrm{T}})$ covariance matrix of the $n_{\mathrm{R}} \cdot n_{\mathrm{T}}$ components of $\mathbb{H}$ is nonsingular. Let $\lambda_{\min}(\mathbb{H}) > 0$ denote the minimal eigenvalue of this covariance matrix. Let $\mathbb{H}_{\mathrm{G}}$ denote a zero-mean Gaussian matrix whose components have the same covariance matrix as the components of $\mathbb{H}$. The nonsingularity of the covariance matrix implies that

$$h(\mathbb{H}_{\mathrm{G}}) > -\infty. \tag{248}$$

Using the data processing inequality for relative entropy we obtain

$$D(\mathbb{H}\hat{\boldsymbol{x}}\|\mathbb{H}_{\mathrm{G}}\hat{\boldsymbol{x}}) \le D(\mathbb{H}\|\mathbb{H}_{\mathrm{G}}) \tag{249}$$

where the relative entropy between two random vectors or matrices is defined as the relative entropy between their corresponding distributions. It now follows from (249) that

$$h(\mathbb{H}_{\mathrm{G}}\hat{\boldsymbol{x}}) - h(\mathbb{H}\hat{\boldsymbol{x}}) \le h(\mathbb{H}_{\mathrm{G}}) - h(\mathbb{H}) \tag{250}$$

or

$$h(\mathbb{H}\hat{\boldsymbol{x}}) \ge h(\mathbb{H}_{\mathrm{G}}\hat{\boldsymbol{x}}) + h(\mathbb{H}) - h(\mathbb{H}_{\mathrm{G}}). \tag{251}$$

Any lower bound on $h(\mathbb{H}_{\mathrm{G}}\hat{\boldsymbol{x}})$ will now yield a bound on $h(\mathbb{H}\hat{\boldsymbol{x}})$ via (251). For example, using Proposition 6.3 we obtain that the smallest eigenvalue of the covariance matrix of $\mathbb{H}_{\mathrm{G}}\hat{\boldsymbol{x}}$ is no smaller than the smallest eigenvalue of the covariance matrix of the elements of $\mathbb{H}_{\mathrm{G}}$, namely, $\lambda_{\min}(\mathbb{H})$. We thus obtain

$$h(\mathbb{H}\hat{\boldsymbol{x}}) \ge n_{\mathrm{R}}\log(\pi e \lambda_{\min}(\mathbb{H})) + h(\mathbb{H}) - h(\mathbb{H}_{\mathrm{G}}) \tag{252}$$

which implies (246).

To prove (247) we note that the condition $\mathsf{E}[\|\mathbb{H}\|_{\mathrm{F}}^2] < +\infty$ implies

$$\sup_{\|\hat{\boldsymbol{x}}\|=1} \mathsf{E}[\|\mathbb{H}\hat{\boldsymbol{x}}\|^2] < +\infty \tag{253}$$

so that by Lemma 6.4

$$\sup_{\|\hat{\boldsymbol{x}}\|=1} h^+(\mathbb{H}\hat{\boldsymbol{x}}) < +\infty. \tag{254}$$

The proof of (247) now follows from (246) and (254) by expressing $h^-(\mathbb{H}\hat{\boldsymbol{x}})$ as $h^+(\mathbb{H}\hat{\boldsymbol{x}}) - h(\mathbb{H}\hat{\boldsymbol{x}})$. □

4) Differential Entropy and Expectations of Logarithms: The following lemma relates the differential entropy of a random vector to the expectation of the logarithm of its norm.

Lemma 6.7: Let $\boldsymbol{X}$ be an n-dimensional complex random vector of density $f_{\boldsymbol{X}}(\boldsymbol{x})$. Then the following relationships between differential entropy and the expected log-norm hold.

a) For any $0 < \delta \le 1$ and $0 < \alpha < n$

$$\mathsf{E}\left[\log\|\boldsymbol{X}+\boldsymbol{c}\|^{-1} \cdot I\{\|\boldsymbol{X}+\boldsymbol{c}\| \le \delta\}\right] \le \epsilon(n,\delta,\alpha) + \frac{1}{\alpha}\int_{\substack{\|\boldsymbol{x}+\boldsymbol{c}\|\le\delta \\ f_{\boldsymbol{X}}(\boldsymbol{x})\ge 1}} f_{\boldsymbol{X}}(\boldsymbol{x})\log f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \quad \boldsymbol{c}\in\mathbb{C}^n$$
$$\le \epsilon(n,\delta,\alpha) + \frac{1}{\alpha}\sup_{\boldsymbol{c}\in\mathbb{C}^n}\int_{\substack{\|\boldsymbol{x}+\boldsymbol{c}\|\le\delta \\ f_{\boldsymbol{X}}(\boldsymbol{x})\ge 1}} f_{\boldsymbol{X}}(\boldsymbol{x})\log f_{\boldsymbol{X}}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

where the term $\epsilon(n,\,\delta,\,\alpha)$ tends to zero as $\delta \downarrow 0$.

b) Consequently, if in addition $h^-(\boldsymbol{X}) < \infty$, then

$$\lim_{\delta \downarrow 0} \sup_{\boldsymbol{c} \in \mathbb{C}^n} \mathsf{E}\left[\log \|\boldsymbol{X} + \boldsymbol{c}\|^{-1} \cdot I\{\|\boldsymbol{X} + \boldsymbol{c}\| \leq \delta\}\right] = 0. \quad (255)$$

c) Moreover, for any $0 < \delta \leq 1$ and $0 < \alpha < n$ we have the uniform bound

$$\sup_{\boldsymbol{c} \in \mathbb{C}^n} \mathsf{E}\left[\log \|\boldsymbol{X} + \boldsymbol{c}\|^{-1} \cdot I\{\|\boldsymbol{X} + \boldsymbol{c}\| \leq \delta\}\right] \leq \epsilon(n, \delta, \alpha) + \frac{1}{\alpha} h^-(\boldsymbol{X}). \quad (256)$$

d) If $h^-(\boldsymbol{X}) < \infty$, then for any $0 < \alpha < n$ there exists some finite number $\Delta(n, \alpha)$ (not depending on the law of $\boldsymbol{X}$) such that

$$\mathsf{E}[\log \|\boldsymbol{X}\|] \geq -\frac{1}{\alpha} h^-(\boldsymbol{X}) - \Delta(n, \alpha). \quad (257)$$

e) Let $\boldsymbol{Y}$ be an n-dimensional random vector. Assume that the differential entropy of some subset of its elements is defined and is greater than $-\infty$. Then

$$\mathsf{E}[\log \|\boldsymbol{Y}\|] > -\infty. \quad (258)$$

f) Let the random matrix $\mathbb{H}$ satisfy $h(\mathbb{H}) > -\infty$ and $\mathsf{E}[\|\mathbb{H}\|_{\mathrm{F}}^2] < +\infty$. Then

$$\inf_{\|\hat{\boldsymbol{x}}\|=1} \mathsf{E}[\log \|\mathbb{H}\hat{\boldsymbol{x}}\|] > -\infty. \quad (259)$$

g) Let the random vectors $\boldsymbol{Z}$ and $\boldsymbol{X}$ be independent and of the same dimension. Assume that $h^-(\boldsymbol{Z}) < \infty$. Then

$$\mathsf{E}[\log \|\boldsymbol{Z} + \boldsymbol{X}\|] > -\infty. \quad (260)$$

Proof: To prove Part a) we express

$$\mathsf{E}[\log \|\boldsymbol{X} + \boldsymbol{c}\|^{-1} \cdot I\{\|\boldsymbol{X} + \boldsymbol{c}\| \leq \delta\}] = \int_{\|\boldsymbol{x}+\boldsymbol{c}\| \leq \delta} f_{\boldsymbol{X}}(\boldsymbol{x}) \log \|\boldsymbol{x} + \boldsymbol{c}\|^{-1} \, \mathrm{d}\boldsymbol{x} \quad (261)$$

as the sum of two integrals

$$\begin{aligned} I_1 &= \int_{\substack{\|\boldsymbol{x}+\boldsymbol{c}\| \leq \delta \\ f_{\boldsymbol{X}}(\boldsymbol{x}) \leq \|\boldsymbol{x}+\boldsymbol{c}\|^{-\alpha}}} f_{\boldsymbol{X}}(\boldsymbol{x}) \log \|\boldsymbol{x} + \boldsymbol{c}\|^{-1} \, \mathrm{d}\boldsymbol{x} \\ &\leq \int_{\substack{\|\boldsymbol{x}+\boldsymbol{c}\| \leq \delta \\ f_{\boldsymbol{X}}(\boldsymbol{x}) \leq \|\boldsymbol{x}+\boldsymbol{c}\|^{-\alpha}}} f_{\boldsymbol{X}}(\boldsymbol{x}) \log^+ \|\boldsymbol{x} + \boldsymbol{c}\|^{-1} \, \mathrm{d}\boldsymbol{x} \\ &\leq \int_{\|\boldsymbol{x}+\boldsymbol{c}\| \leq \delta} \|\boldsymbol{x} + \boldsymbol{c}\|^{-\alpha} \log^+ \|\boldsymbol{x} + \boldsymbol{c}\|^{-1} \, \mathrm{d}\boldsymbol{x} \\ &= \int_{\|\boldsymbol{x}\| \leq \delta} \|\boldsymbol{x}\|^{-\alpha} \log^+ \|\boldsymbol{x}\|^{-1} \, \mathrm{d}\boldsymbol{x} \\ &= \epsilon(n, \delta, \alpha) \end{aligned}$$

and

$$\begin{aligned} I_2 &= \int_{\substack{\|\boldsymbol{x}+\boldsymbol{c}\| \leq \delta \\ f_{\boldsymbol{X}}(\boldsymbol{x}) > \|\boldsymbol{x}+\boldsymbol{c}\|^{-\alpha}}} f_{\boldsymbol{X}}(\boldsymbol{x}) \log \|\boldsymbol{x} + \boldsymbol{c}\|^{-1} \, \mathrm{d}\boldsymbol{x} \\ &\leq \frac{1}{\alpha} \int_{\substack{\|\boldsymbol{x}+\boldsymbol{c}\| \leq \delta \\ f_{\boldsymbol{X}}(\boldsymbol{x}) > \|\boldsymbol{x}+\boldsymbol{c}\|^{-\alpha}}} f_{\boldsymbol{X}}(\boldsymbol{x}) \log f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \\ &\leq \frac{1}{\alpha} \int_{\substack{\|\boldsymbol{x}+\boldsymbol{c}\| \leq \delta \\ f_{\boldsymbol{X}}(\boldsymbol{x}) > 1}} f_{\boldsymbol{X}}(\boldsymbol{x}) \log f_{\boldsymbol{X}}(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}. \end{aligned}$$

Parts b) and c) follow directly from Part a). Part d) follows from Part a) by choosing $\delta = 1$ and $\boldsymbol{c} = \boldsymbol{0}$.

To prove Part e) we note that since the logarithm function is monotonic, it follows that if the expected-log of the squared-norm of a subset of the components of $\boldsymbol{Y}$ is greater than $-\infty$, then so is the expected-log of the norm-squared of all its components. Consequently, it suffices to prove the lemma for the case that $h(\boldsymbol{Y}) > -\infty$, i.e., in the case where $h^-(\boldsymbol{Y}) < +\infty$. But in this case the claim follows directly from Part d).

Part f) follows from Part d) and Lemma 6.6, i.e., from (257) and (247). Part g) follows by conditioning on $\boldsymbol{X}$ and applying Part c) with $\delta = 1$. □

The following lemma extends Lemma 6.7 to provide uniform bounds for the case where a random matrix operates on deterministic vectors. It will allow us to limit attention to integrations over sets that avoid the singularity of the logarithmic function at the origin.

Lemma 6.8: Let the $n_{\mathrm{T}} \times n_{\mathrm{R}}$ complex random matrix $\mathbb{H}$ satisfy $\mathsf{E}[\|\mathbb{H}\|_{\mathrm{F}}^2] < \infty$ and $h(\mathbb{H}) > -\infty$. Then

$$\lim_{\delta \downarrow 0} \sup_{\|\hat{\boldsymbol{x}}\|=1} \mathsf{E}[\log \|\mathbb{H}\hat{\boldsymbol{x}}\|^{-1} \cdot I\{\|\mathbb{H}\hat{\boldsymbol{x}}\| \leq \delta\}] = 0. \quad (262)$$

Proof: Since the lemma only addresses limiting behaviors as $\delta \downarrow 0$, we shall restrict ourselves throughout the proof to

$$0 < \delta \leq 1/\sqrt{n_{\mathrm{T}}}. \quad (263)$$

We shall prove the result by showing that for any integer $1 \leq \tau \leq n_{\mathrm{T}}$

$$\lim_{\delta \downarrow 0} \sup_{\hat{\boldsymbol{x}} \in \mathcal{D}^{(\tau)}} \mathsf{E}[\log \|\mathbb{H}\hat{\boldsymbol{x}}\|^{-1} \cdot I\{\|\mathbb{H}\hat{\boldsymbol{x}}\| \leq \delta\}] = 0 \quad (264)$$

where

$$\mathcal{D}^{(\tau)} = \left\{\hat{\boldsymbol{x}} \in \mathbb{C}^{n_{\mathrm{T}}} : \|\hat{\boldsymbol{x}}\| = 1, \left|\hat{x}^{(\tau)}\right| \geq 1/\sqrt{n_{\mathrm{T}}}\right\}. \quad (265)$$

To this end, we shall show that[9]

$$\lim_{\delta \downarrow 0} \sup_{\hat{\boldsymbol{x}} \in \mathcal{D}^{(\tau)}} \mathsf{E}\left[\log \|\mathbb{H}\hat{\boldsymbol{x}}\|^{-1} \cdot I\{\|\mathbb{H}\hat{\boldsymbol{x}}\| \leq \delta\} \,\middle|\, \{\boldsymbol{H}^{(t)}\}_{t \neq \tau}\right] = 0, \quad \{\boldsymbol{H}^{(t)}\}_{t \neq \tau}\text{-a.s.} \quad (266)$$

and exhibit a dominating function that will allow us to use the Dominated Convergence Theorem to infer that

$$\lim_{\delta \downarrow 0} \mathsf{E}_{\{\boldsymbol{H}^{(t)}\}_{t \neq \tau}}\left[\sup_{\hat{\boldsymbol{x}} \in \mathcal{D}^{(\tau)}} \mathsf{E}\left[\log \|\mathbb{H}\hat{\boldsymbol{x}}\|^{-1} \cdot I\{\|\mathbb{H}\hat{\boldsymbol{x}}\| \leq \delta\} \,\middle|\, \{\boldsymbol{H}^{(t)}\}_{t \neq \tau}\right]\right] = 0. \quad (267)$$

Since (267) implies (264), the lemma will follow.

To prove (266) fix then some integer $1 \leq \tau \leq n_{\mathrm{T}}$, and for any $\{\boldsymbol{h}^{(t)}\}_{t \neq \tau}$ consider

$$\sup_{\hat{\boldsymbol{x}} \in \mathcal{D}^{(\tau)}} \mathsf{E}\left[\log \|\mathbb{H}\hat{\boldsymbol{x}}\|^{-1} \cdot I\{\|\mathbb{H}\hat{\boldsymbol{x}}\| \leq \delta\} \,\middle|\, \{\boldsymbol{H}^{(t)} = \boldsymbol{h}^{(t)}\}_{t \neq \tau}\right]. \quad (268)$$

The argument in the conditional expectation in (268) can be written as

$$\log \left\|\hat{x}^{(\tau)}\left(\boldsymbol{H}^{(\tau)} + \tilde{\boldsymbol{c}}\right)\right\|^{-1} \cdot I\left\{\left\|\hat{x}^{(\tau)}\left(\boldsymbol{H}^{(\tau)} + \tilde{\boldsymbol{c}}\right)\right\| \leq \delta\right\} \quad (269)$$

[9] Recall that $\boldsymbol{H}^{(t)}$ is the tth column of $\mathbb{H}$.

where

$$\tilde{\boldsymbol{c}} = \frac{1}{\hat{x}^{(\tau)}} \sum_{\substack{1 \le t \le n_{\mathrm{T}} \\ t \ne \tau}} \hat{x}^{(t)} \boldsymbol{h}^{(t)}. \tag{270}$$

Since any element $\hat{\boldsymbol{x}}$ in $\mathcal{D}^{(\tau)}$ satisfies

$$1/\sqrt{n_{\mathrm{T}}} \le \left|\hat{x}^{(\tau)}\right| \le 1 \tag{271}$$

it follows from (263), from the lower bound in (271), and from the upper bound in (271) that

$$\delta \le \delta\sqrt{n_{\mathrm{T}}}\left|\hat{x}^{(\tau)}\right| \le \left|\hat{x}^{(\tau)}\right| \le 1 \tag{272}$$

so that for $\hat{\boldsymbol{x}} \in \mathcal{D}^{(\tau)}$ and for δ satisfying (263) the integrand (269) can be upper-bounded by

$$\log\left\|\hat{x}^{(\tau)}\left(\boldsymbol{H}^{(\tau)} + \tilde{\boldsymbol{c}}\right)\right\|^{-1} \cdot I\left\{\left\|\hat{x}^{(\tau)}\left(\boldsymbol{H}^{(\tau)} + \tilde{\boldsymbol{c}}\right)\right\| \le \delta\sqrt{n_{\mathrm{T}}}\left|\hat{x}^{(\tau)}\right|\right\}. \tag{273}$$

Using (271) again we can further upper-bound (273) (for $\hat{\boldsymbol{x}} \in \mathcal{D}^{(\tau)}$ and for δ satisfying (263)) by

$$\log\left(\sqrt{n_{\mathrm{T}}}\right) \cdot I\left\{\left\|\boldsymbol{H}^{(\tau)} + \tilde{\boldsymbol{c}}\right\| \le \delta\sqrt{n_{\mathrm{T}}}\right\} + \log\left\|\boldsymbol{H}^{(\tau)} + \tilde{\boldsymbol{c}}\right\|^{-1} \cdot I\left\{\left\|\boldsymbol{H}^{(\tau)} + \tilde{\boldsymbol{c}}\right\| \le \delta\sqrt{n_{\mathrm{T}}}\right\}. \tag{274}$$

To prove (266), we shall now consider the supremum (over $\tilde{\mathbf{c}}$) of the conditional (on $\{\boldsymbol{H}^{(t)}\}$, $t \ne \tau$) expectation (over $\boldsymbol{H}^{(\tau)}$) of the two terms in (274) separately. With regard to the first, we have that

$$\sup_{\tilde{\boldsymbol{c}} \in \mathbb{C}^{n_{\mathrm{R}}}} \Pr\left(\left\|\boldsymbol{H}^{(\tau)} + \tilde{\boldsymbol{c}}\right\| \le \delta\sqrt{n_{\mathrm{T}}} \,\middle|\, \{\boldsymbol{H}^{(t)} = \boldsymbol{h}^{(t)}\}_{t \ne \tau}\right) \tag{275}$$

can be upper-bounded by

$$\sup_{\tilde{\boldsymbol{c}}} \int_{\substack{\|\boldsymbol{h}^{(\tau)} + \tilde{\boldsymbol{c}}\| \le \delta\sqrt{n_{\mathrm{T}}} \\ f_{\boldsymbol{H}^{(\tau)}|\{\boldsymbol{H}^{(t)}\}_{t\ne\tau}}(\boldsymbol{h}^{(\tau)}|\{\boldsymbol{h}^{(t)}\}_{t\ne\tau}) \ge e}} f_{\boldsymbol{H}^{(\tau)}|\{\boldsymbol{H}^{(t)}\}_{t\ne\tau}}\left(\boldsymbol{h}^{(\tau)}\middle|\{\boldsymbol{h}^{(t)}\}_{t\ne\tau}\right) \cdot \log f_{\boldsymbol{H}^{(\tau)}|\{\boldsymbol{H}^{(t)}\}_{t\ne\tau}}\left(\boldsymbol{h}^{(\tau)}\middle|\{\boldsymbol{h}^{(t)}\}_{t\ne\tau}\right) \mathrm{d}\boldsymbol{h}^{(\tau)} + \int_{\|\boldsymbol{h}^{(\tau)}\| \le \delta\sqrt{n_{\mathrm{T}}}} e \,\mathrm{d}\boldsymbol{h}^{(\tau)} \tag{276}$$

which for almost every $\{\boldsymbol{h}^{(t)}\}_{t\ne\tau}$ converges to zero as $\delta \downarrow 0$ because by Lemma 6.5

$$h^{-}\left(\boldsymbol{H}^{(\tau)}\middle|\{\boldsymbol{H}^{(t)} = \boldsymbol{h}^{(t)}\}_{t\ne\tau}\right) < \infty, \qquad \{\boldsymbol{H}^{(t)}\}_{t\ne\tau}\text{-a.s.} \tag{277}$$

Having established the pointwise convergence of this term to zero, we can now justify the swapping of the limit $\delta \downarrow 0$ with the expectation (with respect to $\{\boldsymbol{H}^{(t)}\}_{t\ne\tau}$) by exhibiting an integrable (with respect to $\{\boldsymbol{H}^{(t)}\}_{t\ne\tau}$) dominating function (of $\{\boldsymbol{h}^{(t)}\}_{t\ne\tau}$ but not depending on $0 < \delta < 1/\sqrt{n_{\mathrm{T}}}$ or $\tilde{\mathbf{c}}$). An appropriate function is, for example,

$$h^{-}\left(\boldsymbol{H}^{(\tau)}\middle|\{\boldsymbol{H}^{(t)} = \boldsymbol{h}^{(t)}\}_{t\ne\tau}\right) + e \cdot \mathrm{Vol}(1, n_{\mathrm{R}}) \tag{278}$$

which is integrable by Lemma 6.5. Here we used $\mathrm{Vol}(1, n_{\mathrm{R}})$ to denote the volume of the unit ball in $\mathbb{C}^{n_{\mathrm{R}}}$.

The supremum (over $\tilde{\boldsymbol{c}}$) of the conditional (on $\{\boldsymbol{H}^{(t)}\}$, $t \ne \tau$) expectation (over $\boldsymbol{H}^{(\tau)}$) of the second term in (274) can be similarly analyzed using Lemma 6.7 to show that for almost every realization of $\{\boldsymbol{H}^{(t)}\}_{t\ne\tau}$

$$\sup_{\tilde{\boldsymbol{c}}} \mathsf{E}\left[\log\left\|\boldsymbol{H}^{(\tau)} + \tilde{\boldsymbol{c}}\right\|^{-1} \cdot I\left\{\left\|\boldsymbol{H}^{(\tau)} + \tilde{\boldsymbol{c}}\right\| \le \delta\sqrt{n_{\mathrm{T}}}\right\} \,\middle|\, \{\boldsymbol{H}^{(t)} = \boldsymbol{h}^{(t)}\}_{t\ne\tau}\right] \tag{279}$$

converges to zero as $\delta \downarrow 0$. This can be argued because for almost every realization of $\{\boldsymbol{H}^{(t)}\}_{t\ne\tau}$ it follows from Lemma 6.5 that $h^{-}(\boldsymbol{H}^{(\tau)}|\{\boldsymbol{h}^{(t)}\}_{t\ne\tau}) < \infty$ so that the limiting behavior in (279) follows from a conditional version of Lemma 6.7, Part b). To demonstrate a dominating function, we can rely on the conditional version of Lemma 6.7, Part c) and use Lemma 6.5 to demonstrate the integrability of the dominating function. □

5) Continuity Issues: We turn now to address some continuity claims about differential entropy and related functions.

Lemma 6.9: Let X and Y be two independent real random variables satisfying $\mathsf{E}[|X|] < \infty$ and $\mathsf{E}[|Y|] < \infty$. Then

$$\lim_{\epsilon \to 0} h(X + \epsilon Y) = h(X). \tag{280}$$

Proof: The idea of the proof is to relate differential entropy to relative entropy, and to then use the lower semicontinuity of the latter. This approach is surveyed in [29]. We begin by first treating the case where X is not deterministically zero, i.e., $\mathsf{E}[|X|] > 0$. Let

$$\alpha_\epsilon = \mathsf{E}[|X + \epsilon Y|].$$

Note that the assumption $\mathsf{E}[|X|] > 0$ guarantees that $\alpha_\epsilon > 0$ for all sufficiently small values of ϵ.

Let P_ϵ denote the probability measure corresponding to $X + \epsilon Y$, and let $P_{\epsilon,\mathcal{L}}$ be the probability measure corresponding to a Laplacian random variable L_ϵ of equal expected magnitude, i.e., of expected magnitude α_ϵ. Thus, $P_{\epsilon,\mathcal{L}}$ has the density

$$\frac{1}{2\alpha_\epsilon} e^{-\frac{|\xi|}{\alpha_\epsilon}}, \qquad \xi \in \mathbb{R}. \tag{281}$$

Since P_ϵ converges weakly to P_0 and since $P_{\epsilon,\mathcal{L}}$ converges weakly to $P_{0,\mathcal{L}}$ [30, 17.1.d], it follows by the lower semicontinuity of relative entropy (see, e.g., [21, Proof of Lemma 4] and references therein) that

$$\varliminf_{\epsilon \to 0} D\left(P_\epsilon \| P_{\epsilon,\mathcal{L}}\right) \ge D\left(P_0 \| P_{0,\mathcal{L}}\right). \tag{282}$$

But using the explicit form of the density of $P_{\epsilon,\mathcal{L}}$ (281) we have

$$D\left(P_\epsilon \| P_{\epsilon,\mathcal{L}}\right) = 1 + \log\left(2\alpha_\epsilon\right) - h(X + \epsilon Y). \tag{283}$$

It thus follows from (282), (283) (and from our assumption $\alpha_0 = \mathsf{E}[|X|] > 0$, which implies the continuity of $\log(2\alpha_\epsilon)$) that

$$\varlimsup_{\epsilon \to 0} h(X + \epsilon Y) \le h(X). \tag{284}$$

This completes the proof of the lemma (for the case $\mathsf{E}[|X|] > 0$) because the independence of X and Y guarantees that $h(X + \epsilon Y) \ge h(X)$, and hence $\varliminf h(X + \epsilon Y) \ge h(X)$.

The case $\mathsf{E}[|X|] = 0$ can be treated by other methods. For this case, $h(X) = -\infty$ and $h(X + \epsilon Y) = h(\epsilon Y) = \log \epsilon + h(Y)$. The result now follows by noting that because $\mathsf{E}[|Y|] < \infty$ it follows that $h(Y) < \infty$. (The max-entropy distribution under an expected magnitude constraint is the Laplacian distribution, which has a finite differential entropy.) □

Lemma 6.10: Let H be a complex random variable satisfying $\mathsf{E}[|H|^2] < \infty$ and $h(H) > -\infty$. Let $U = \log|H|^2$ and let V be independent of U and uniformly distributed over an interval of length β. Then

$$\lim_{\beta\to\infty} h(V+U) - h(V) = 0. \tag{285}$$

Proof: Since differential entropy is invariant under translation, it follows that there is no loss in generality in assuming that V is uniformly distributed over the interval $[0, \beta]$. By the scaling property of differential entropy [12, Theorem 9.6.4] it follows that

$$h(V+U) - h(V) = h(V' + \beta^{-1}U) - h(V') \tag{286}$$

where $V' = V/\beta$ is uniformly distributed over the interval $[0, 1]$ and independent of U. The claim now follows from Lemma 6.9 because the conditions $\mathsf{E}[|H|^2] < \infty$ and $h(H) > -\infty$ guarantee that $\mathsf{E}[|\log|H||] < \infty$; see Lemma 6.7. □

Lemma 6.11: The following continuity results of differential entropy with respect to Gaussian perturbations hold.

a) Let $\boldsymbol{W} \in \mathbb{C}^\nu$ be a random vector satisfying $\mathsf{E}[\|\boldsymbol{W}\|^2] < \infty$ and $h(\boldsymbol{W}) > -\infty$. Let $\boldsymbol{Z} \in \mathbb{C}^\nu$ be a Gaussian random vector that is independent of $\boldsymbol{W}$. Then

$$\lim_{\sigma\to 0}\{h(\boldsymbol{W}+\sigma\boldsymbol{Z}) - h(\boldsymbol{W})\} = 0. \tag{287}$$

b) Let $\mathbb{H}$ be a random $n_{\mathrm{R}} \times n_{\mathrm{T}}$ matrix such that $\mathsf{E}[\|\mathbb{H}\|_{\mathrm{F}}^2] < \infty$ and $h(\mathbb{H}) > -\infty$. Let $\boldsymbol{Z} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathsf{I}_{n_{\mathrm{R}}})$ be independent of $\mathbb{H}$. Then

$$\lim_{\sigma\to 0}\sup_{\|\hat{\boldsymbol{x}}\|=1}\{h(\mathbb{H}\hat{\boldsymbol{x}}+\sigma\boldsymbol{Z}) - h(\mathbb{H}\hat{\boldsymbol{x}})\} = 0. \tag{288}$$

Proof: The proof of Part a) is very similar to the proof of Lemma 6.9 except that rather than considering relative entropies with respect to Laplacian random variables we need to consider relative entropies with respect to multivariate Gaussians of equal covariance matrices. We begin by noting that since differential entropy is invariant under deterministic translation [12], we may assume without loss of generality that $\mathsf{E}[\boldsymbol{W}] = 0$. Consider now a sequence $\{\sigma_n^2\}$ converging to zero. Let $\boldsymbol{W}_n = \boldsymbol{W} + \sigma_n\boldsymbol{Z}$. Since conditioning on $\boldsymbol{Z}$ cannot increase differential entropy, we have

$$h(\boldsymbol{W}_n) \geq h(\boldsymbol{W}) \tag{289}$$

so that

$$\varliminf_{n\to\infty} h(\boldsymbol{W}_n) \geq h(\boldsymbol{W}). \tag{290}$$

To study the $\limsup$ of $h(\boldsymbol{W}_n)$ we first note that

$$\boldsymbol{W}_n \Longrightarrow \boldsymbol{W}$$

where we use the symbol "$\Longrightarrow$" to denote weak convergence.[10] Denoting the law of $\boldsymbol{W}$ by $P_{\boldsymbol{W}}$, the law of $\boldsymbol{W}_n$ by $P_{\boldsymbol{W}_n}$, the law of a zero-mean covariance $\mathsf{E}[\boldsymbol{W}\boldsymbol{W}^\dagger]$ Gaussian $\mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathsf{E}[\boldsymbol{W}\boldsymbol{W}^\dagger])$ by $P_{\boldsymbol{W},\mathrm{G}}$, and the law of a zero-mean covariance $\mathsf{E}[\boldsymbol{W}_n\boldsymbol{W}_n^\dagger]$ Gaussian $\mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathsf{E}[\boldsymbol{W}_n\boldsymbol{W}_n^\dagger])$ by $P_{\boldsymbol{W}_n,\mathrm{G}}$, we obtain from the lower semicontinuity of relative entropy that

$$\varliminf_{n\to\infty} D(P_{\boldsymbol{W}_n}\|P_{\boldsymbol{W}_n,\mathrm{G}}) \geq D(P_{\boldsymbol{W}}\|P_{\boldsymbol{W},\mathrm{G}}). \tag{291}$$

But

$$D(P_{\boldsymbol{W}_n}\|P_{\boldsymbol{W}_n,\mathrm{G}}) = \log(\pi e)^\nu + \log\det\mathsf{E}[\boldsymbol{W}_n\boldsymbol{W}_n^\dagger] - h(\boldsymbol{W}_n) \tag{292}$$

and

$$D(P_{\boldsymbol{W}}\|P_{\boldsymbol{W},\mathrm{G}}) = \log(\pi e)^\nu + \log\det\mathsf{E}[\boldsymbol{W}\boldsymbol{W}^\dagger] - h(\boldsymbol{W}) \tag{293}$$

so that since $\log\det\mathsf{E}[\boldsymbol{W}_n\boldsymbol{W}_n^\dagger] \to \log\det\mathsf{E}[\boldsymbol{W}\boldsymbol{W}^\dagger]$ we obtain

$$\varlimsup_{n\to\infty} h(\boldsymbol{W}_n) \leq h(\boldsymbol{W}) \tag{294}$$

which combines with (290) to prove Part a).

To prove Part b), define the $n_{\mathrm{R}} \times n_{\mathrm{T}}$ random matrix $\mathbb{Z}$ to have independent $\mathcal{N}_{\mathbb{C}}(0, 1)$ components. We will show that

$$h(\mathbb{H}\hat{\boldsymbol{x}}+\sigma\boldsymbol{Z}) - h(\mathbb{H}\hat{\boldsymbol{x}}) \leq h(\mathbb{H}+\sigma\mathbb{Z}) - h(\mathbb{H}), \qquad \|\hat{\boldsymbol{x}}\| = 1. \tag{295}$$

Once this relation is established, the result will follow by noting that the RHS of (295) converges to zero as $\sigma^2 \downarrow 0$. (This convergence of the RHS follows from Part a) by stacking the $n_{\mathrm{R}} \times n_{\mathrm{T}}$ components of $\mathbb{H}$ and $\mathbb{Z}$ into vectors.)

To prove (295), express the difference in the differential entropies as a mutual information as follows:

$$\begin{aligned} h(\mathbb{H}+\sigma\mathbb{Z}) - h(\mathbb{H}) &= I(\mathbb{Z}; \mathbb{H}+\sigma\mathbb{Z}) \\ &\geq I(\mathbb{Z}; (\mathbb{H}+\sigma\mathbb{Z})\hat{\boldsymbol{x}}) \\ &= I(\mathbb{Z}\hat{\boldsymbol{x}}; \mathbb{H}\hat{\boldsymbol{x}}+\sigma\mathbb{Z}\hat{\boldsymbol{x}}) \\ &= I(\boldsymbol{Z}; \mathbb{H}\hat{\boldsymbol{x}}+\sigma\boldsymbol{Z}) \\ &= h(\mathbb{H}\hat{\boldsymbol{x}}+\sigma\boldsymbol{Z}) - h(\mathbb{H}\hat{\boldsymbol{x}}). \end{aligned}$$

Here the inequality follows from the data processing theorem; the subsequent equality because $\mathbb{Z}\hat{\boldsymbol{x}}$ is sufficient statistics for $\mathbb{H}\hat{\boldsymbol{x}} + \sigma\mathbb{Z}\hat{\boldsymbol{x}}$; and the subsequent equality because $\mathbb{Z}\hat{\boldsymbol{x}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathsf{I}_{n_{\mathrm{R}}})$. □

Lemma 6.12: Let the $n_{\mathrm{R}} \times n_{\mathrm{T}}$ random matrix $\mathbb{H}$ satisfy $\mathsf{E}[\|\mathbb{H}\|_{\mathrm{F}}^2] < \infty$, and let $\boldsymbol{Z} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathsf{I}_{n_{\mathrm{R}}})$ be independent of $\mathbb{H}$. Then for any $\sigma > 0$, the mapping from $\mathbb{C}^{n_{\mathrm{T}}}$ to the real line

$$\boldsymbol{x} \mapsto h(\mathbb{H}\boldsymbol{x}+\sigma\boldsymbol{Z})$$

is continuous in $\boldsymbol{x}$.

Proof: Let the sequence $\{\boldsymbol{x}_n\}$ converge to $\boldsymbol{x}$. It then follows that the sequence $\{\mathbb{H}\boldsymbol{x}_n + \sigma\boldsymbol{Z}\}$ converges weakly to $\mathbb{H}\boldsymbol{x} + \sigma\boldsymbol{Z}$. By relating the differential entropy $h(\mathbb{H}\boldsymbol{x}_n + \sigma\boldsymbol{Z})$ to the relative entropy between the distribution of $\mathbb{H}\boldsymbol{x}_n + \sigma\boldsymbol{Z}$ and a Gaussian distribution of equal second-order moments, we can infer from the lower semicontinuity of relative entropy that

$$\varlimsup_{n\to\infty} h(\mathbb{H}\boldsymbol{x}_n+\sigma\boldsymbol{Z}) \leq h(\mathbb{H}\boldsymbol{x}+\sigma\boldsymbol{Z}). \tag{296}$$

It, therefore, remains to prove the reverse inequality

$$\varliminf_{n\to\infty} h(\mathbb{H}\boldsymbol{x}_n+\sigma\boldsymbol{Z}) \geq h(\mathbb{H}\boldsymbol{x}+\sigma\boldsymbol{Z}). \tag{297}$$

By the behavior of differential entropy under scaling, it suffices to prove the lemma for $\boldsymbol{x} = 0$ and for all unit vectors $\|\boldsymbol{x}\| = 1$. The case $\boldsymbol{x} = 0$ is straightforward because the inequality $h(\mathbb{H}\boldsymbol{x}_n+\sigma\boldsymbol{Z}) > h(\sigma\boldsymbol{Z})$ suffices, in this case, to demon-

[10] By weak convergence we refer to the standard definition of weak convergence of probability measures [30, Ch. 17]. Functional analysts would have perhaps preferred to refer to it as weak* convergence.

strate (297) and, therefore, to prove continuity with the aid of (296). We, therefore, now focus on the case where $\boldsymbol{x}$ is a unit vector. As a reminder that this is the case at hand, we replace $\boldsymbol{x}$ with the symbol $\hat{\boldsymbol{x}}$, where the hat is an indication that $\|\hat{\boldsymbol{x}}\| = 1$. Thus, $\{\boldsymbol{x}_n\}$ now converges to $\hat{\boldsymbol{x}}$. Let $\tilde{\boldsymbol{Z}} \sim \mathcal{N}(\mathbf{0}, \mathsf{I}_{n_\mathrm{R}})$ be independent of $\mathbb{H}$ and $\boldsymbol{Z}$. For the purposes of obtaining uniform bounds that do not depend on the transmitted vector it will be helpful to consider matrix extensions of $\boldsymbol{Z}$ and $\tilde{\boldsymbol{Z}}$. Let $\mathbb{Z}$ and $\tilde{\mathbb{Z}}$ be two independent $n_\mathrm{R} \times n_\mathrm{T}$ random matrices, both independent of $\mathbb{H}$, each of which has $\mathcal{N}_\mathbb{C}(0, 1)$ i.i.d. components. By stacking the components of the matrices $\mathbb{H}$ and $\mathbb{Z}$ into $n_\mathrm{R} \cdot n_\mathrm{T}$ arrays, we can invoke [31]–[33] to infer that the entropy power of $\mathbb{H} + \sigma\mathbb{Z}$ is concave in σ, so that, in particular, $h(\mathbb{H} + \sigma\mathbb{Z})$ is continuous in σ, for $\sigma > 0$. Since any continuous function is also uniformly continuous on compact intervals, it follows that there exists some $\overline{\sigma} > 0$ such that

$$h\Big(\mathbb{H} + \beta\cdot\sigma\mathbb{Z} + \overline{\sigma}\tilde{\mathbb{Z}}\Big) - h(\mathbb{H} + \beta\cdot\sigma\mathbb{Z}) < \epsilon, \qquad |\beta - 1| \le \frac{1}{2}. \tag{298}$$

By the monotonicity of the LHS of (298) in $\overline{\sigma}$ it thus follows that if we define $\tilde{\sigma} = \overline{\sigma}/1.5$, then

$$I\Big(\mathbb{H} + \beta\cdot\sigma\mathbb{Z} + \beta\cdot\tilde{\sigma}\tilde{\mathbb{Z}};\, \tilde{\mathbb{Z}}\Big) < \epsilon, \qquad |\beta - 1| \le \frac{1}{2}. \tag{299}$$

It now follows from (299) that for any $\boldsymbol{x}'$ satisfying $\big|\|\boldsymbol{x}'\|^{-1} - 1\big| \le 1/2$

$$\begin{aligned}
&I\Big(\mathbb{H}\boldsymbol{x}' + \sigma\boldsymbol{Z} + \tilde{\sigma}\tilde{\boldsymbol{Z}};\, \tilde{\boldsymbol{Z}}\Big)\\
&\quad= h\Big(\mathbb{H}\boldsymbol{x}' + \sigma\boldsymbol{Z} + \tilde{\sigma}\tilde{\boldsymbol{Z}}\Big) - h(\mathbb{H}\boldsymbol{x}' + \sigma\boldsymbol{Z})\\
&\quad= I\bigg(\mathbb{H}\,\frac{\boldsymbol{x}'}{\|\boldsymbol{x}'\|} + \frac{\sigma}{\|\boldsymbol{x}'\|}\boldsymbol{Z} + \frac{\tilde{\sigma}}{\|\boldsymbol{x}'\|}\tilde{\boldsymbol{Z}};\, \tilde{\boldsymbol{Z}}\bigg)\\
&\quad= I\bigg(\Big(\mathbb{H} + \frac{\sigma}{\|\boldsymbol{x}'\|}\mathbb{Z} + \frac{\tilde{\sigma}}{\|\boldsymbol{x}'\|}\tilde{\mathbb{Z}}\Big)\frac{\boldsymbol{x}'}{\|\boldsymbol{x}'\|};\, \tilde{\mathbb{Z}}\,\frac{\boldsymbol{x}'}{\|\boldsymbol{x}'\|}\bigg)\\
&\quad\le I\bigg(\mathbb{H} + \frac{\sigma}{\|\boldsymbol{x}'\|}\mathbb{Z} + \frac{\tilde{\sigma}}{\|\boldsymbol{x}'\|}\tilde{\mathbb{Z}};\, \tilde{\mathbb{Z}}\bigg)\\
&\quad< \epsilon, \qquad \big|\|\boldsymbol{x}'\|^{-1} - 1\big| \le \frac{1}{2}
\end{aligned} \tag{300}$$

where the first inequality follows by the data processing theorem, because

$$\tilde{\mathbb{Z}}\,\frac{\boldsymbol{x}'}{\|\boldsymbol{x}'\|} \multimap \tilde{\mathbb{Z}} \multimap \mathbb{H} + \frac{\sigma}{\|\boldsymbol{x}'\|}\mathbb{Z} + \frac{\tilde{\sigma}}{\|\boldsymbol{x}'\|}\tilde{\mathbb{Z}} \multimap \Big(\mathbb{H} + \frac{\sigma}{\|\boldsymbol{x}'\|}\mathbb{Z} + \frac{\tilde{\sigma}}{\|\boldsymbol{x}'\|}\tilde{\mathbb{Z}}\Big)\frac{\boldsymbol{x}'}{\|\boldsymbol{x}'\|}$$

form a Markov chain, and where the last inequality follows from (299) with $\beta = 1/\|\boldsymbol{x}'\|$.

It now follows from (300) that for $\big|\|\boldsymbol{x}'\|^{-1} - 1\big| \le 1/2$

$$\begin{aligned}
&h(\mathbb{H}\boldsymbol{x}' + \sigma\boldsymbol{Z})\\
&\quad\ge h\Big(\mathbb{H}\boldsymbol{x}' + \sigma\boldsymbol{Z} + \tilde{\sigma}\tilde{\boldsymbol{Z}}\Big) - \epsilon\\
&\quad\ge h\Big(\mathbb{H}\hat{\boldsymbol{x}} + \mathbb{H}(\boldsymbol{x}' - \hat{\boldsymbol{x}}) + \sigma\boldsymbol{Z} + \tilde{\sigma}\tilde{\boldsymbol{Z}}\Big|\mathbb{H}(\boldsymbol{x}' - \hat{\boldsymbol{x}}) + \tilde{\sigma}\tilde{\boldsymbol{Z}}\Big) - \epsilon\\
&\quad= h\Big(\mathbb{H}\hat{\boldsymbol{x}} + \sigma\boldsymbol{Z}\Big|\mathbb{H}(\boldsymbol{x}' - \hat{\boldsymbol{x}}) + \tilde{\sigma}\tilde{\boldsymbol{Z}}\Big) - \epsilon\\
&\quad= h(\mathbb{H}\hat{\boldsymbol{x}} + \sigma\boldsymbol{Z}) - I\Big(\mathbb{H}\hat{\boldsymbol{x}} + \sigma\boldsymbol{Z};\, \mathbb{H}(\boldsymbol{x}' - \hat{\boldsymbol{x}}) + \tilde{\sigma}\tilde{\boldsymbol{Z}}\Big) - \epsilon
\end{aligned} \tag{301}$$

where the second inequality follows because conditioning cannot increase differential entropy. Expanding the mutual information term we obtain

$$\begin{aligned}
&I\Big(\mathbb{H}\hat{\boldsymbol{x}} + \sigma\boldsymbol{Z};\, \mathbb{H}(\boldsymbol{x}' - \hat{\boldsymbol{x}}) + \tilde{\sigma}\tilde{\boldsymbol{Z}}\Big)\\
&\quad= h\Big(\mathbb{H}(\boldsymbol{x}' - \hat{\boldsymbol{x}}) + \tilde{\sigma}\tilde{\boldsymbol{Z}}\Big) - h\Big(\mathbb{H}(\boldsymbol{x}' - \hat{\boldsymbol{x}}) + \tilde{\sigma}\tilde{\boldsymbol{Z}}\Big|\mathbb{H}\hat{\boldsymbol{x}} + \sigma\boldsymbol{Z}\Big)
\end{aligned} \tag{302}$$

$$\le n_\mathrm{R}\log\bigg(\frac{\mathsf{E}[\|\mathbb{H}\|_\mathrm{F}^2]}{n_\mathrm{R}}\,\|\boldsymbol{x}' - \hat{\boldsymbol{x}}\|^2 + \tilde{\sigma}^2\bigg) - n_\mathrm{R}\log\tilde{\sigma}^2. \tag{303}$$

Here the inequality follows by noting that because $\tilde{\boldsymbol{Z}}$ is Gaussian and independent of $\mathbb{H}(\boldsymbol{x}' - \hat{\boldsymbol{x}})$, even conditional on $\mathbb{H}\hat{\boldsymbol{x}} + \sigma\boldsymbol{Z}$

$$\begin{aligned}
h\Big(\mathbb{H}(\boldsymbol{x}' - \hat{\boldsymbol{x}}) + \tilde{\sigma}\tilde{\boldsymbol{Z}}\Big|\mathbb{H}\hat{\boldsymbol{x}} + \sigma\boldsymbol{Z}\Big) &\ge h\Big(\tilde{\sigma}\tilde{\boldsymbol{Z}}\Big)\\
&= n_\mathrm{R}\log\left(\pi e\tilde{\sigma}^2\right)
\end{aligned}$$

and because among all random vectors of a given expected squared-norm, differential entropy is maximized by the one whose components are i.i.d. Gaussian, so that

$$\begin{aligned}
&h\Big(\mathbb{H}(\boldsymbol{x}' - \hat{\boldsymbol{x}}) + \tilde{\sigma}\tilde{\boldsymbol{Z}}\Big)\\
&\quad\le n_\mathrm{R}\log\bigg(\pi e\mathsf{E}\Big[\Big\|\mathbb{H}(\boldsymbol{x}' - \hat{\boldsymbol{x}}) + \tilde{\sigma}\tilde{\boldsymbol{Z}}\Big\|^2\Big]\Big/ n_\mathrm{R}\bigg)\\
&\quad= n_\mathrm{R}\log\big(\pi e\big(\mathsf{E}\big[\|\mathbb{H}(\boldsymbol{x}' - \hat{\boldsymbol{x}})\|^2\big] + n_\mathrm{R}\tilde{\sigma}^2\big)/ n_\mathrm{R}\big)\\
&\quad\le n_\mathrm{R}\log\big(\pi e\big(\mathsf{E}\big[\|\mathbb{H}\|_\mathrm{F}^2\big]\,\|\boldsymbol{x}' - \hat{\boldsymbol{x}})\|^2/ n_\mathrm{R} + \tilde{\sigma}^2\big)\big).
\end{aligned}$$

Inequalities (301) and (303) combine to prove that

$$\varliminf_{n\to\infty} h(\mathbb{H}\boldsymbol{x}_n + \sigma\boldsymbol{Z}) \ge h(\mathbb{H}\hat{\boldsymbol{x}} + \sigma\boldsymbol{Z}) - \epsilon \tag{304}$$

and since $\epsilon > 0$ was arbitrary, (297) is proved, which combines with (296) to prove the lemma. □

6) Differences Between Expected-Logarithms and Entropies:

Lemma 6.13: Let $\mathbb{H}$ be a complex random $n_\mathrm{R} \times n_\mathrm{T}$ matrix satisfying $h(\mathbb{H}) > -\infty$ and $\mathsf{E}[\|\mathbb{H}\|_\mathrm{F}^2] < \infty$. Let $\boldsymbol{Z}'$ be a zero-mean complex random vector in $\mathbb{C}^{n_\mathrm{R}}$ satisfying $\mathsf{E}[\|\boldsymbol{Z}'\|^2] < \infty$ and $h(\boldsymbol{Z}') > -\infty$. Then the function

$$\boldsymbol{x} \mapsto n_\mathrm{R}\mathsf{E}\left[\log\|\mathbb{H}\boldsymbol{x} + \boldsymbol{Z}'\|^2\right] - h(\mathbb{H}\boldsymbol{x} + \boldsymbol{Z}') \tag{305}$$

is a bounded function of $\boldsymbol{x} \in \mathbb{C}^{n_\mathrm{R}}$.

Proof: We begin by using Jensen's inequality to upper-bound $n_\mathrm{R}\mathsf{E}[\log\|\mathbb{H}\boldsymbol{x} + \boldsymbol{Z}'\|^2]$ to obtain for every $\boldsymbol{x} \neq 0$

$$\begin{aligned}
&n_\mathrm{R}\mathsf{E}\left[\log\|\mathbb{H}\boldsymbol{x} + \boldsymbol{Z}'\|^2\right]\\
&\quad\le n_\mathrm{R}\log\mathsf{E}\left[\|\mathbb{H}\boldsymbol{x} + \boldsymbol{Z}'\|^2\right]\\
&\quad= n_\mathrm{R}\log\left(\mathsf{E}\left[\|\boldsymbol{Z}'\|^2\right] + \mathsf{E}\left[\|\mathbb{H}\boldsymbol{x}\|^2\right]\right)\\
&\quad\le n_\mathrm{R}\log\left(\mathsf{E}\left[\|\boldsymbol{Z}'\|^2\right] + \mathsf{E}\left[\|\mathbb{H}\|^2\right]\|\boldsymbol{x}\|^2\right) \qquad (306)\\
&\quad= n_\mathrm{R}\log\bigg(\frac{\mathsf{E}\left[\|\boldsymbol{Z}'\|^2\right] + \mathsf{E}\left[\|\mathbb{H}\|^2\right]\|\boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^2}\bigg) + n_\mathrm{R}\log\|\boldsymbol{x}\|^2\\
&\quad\le n_\mathrm{R}\log\bigg(\frac{\mathsf{E}\left[\|\boldsymbol{Z}'\|^2\right] + \mathsf{E}\left[\|\mathbb{H}\|_\mathrm{F}^2\right]\|\boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^2}\bigg) + n_\mathrm{R}\log\|\boldsymbol{x}\|^2.
\end{aligned} \tag{307}$$

We next lower-bound $h(\mathbb{H}\boldsymbol{x}+\boldsymbol{Z}')$ by conditioning on $\boldsymbol{Z}'$ to obtain

$$\begin{aligned} h(\mathbb{H}\boldsymbol{x}+\boldsymbol{Z}') &\geq h(\mathbb{H}\boldsymbol{x}) \\ &= h\left(\mathbb{H}\,\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right) + 2\log\|\boldsymbol{x}\|^{n_{\mathrm{R}}} \\ &\geq \inf_{\|\hat{\boldsymbol{x}}\|=1} h(\mathbb{H}\hat{\boldsymbol{x}}) + 2\log\|\boldsymbol{x}\|^{n_{\mathrm{R}}} \end{aligned} \tag{308}$$

where the equality follows from the behavior of differential entropy under the scaling of *complex* random vectors, and where the RHS of the last inequality is finite by Lemma 6.6.

Combining (307) and (308) we obtain the upper bound

$$\begin{aligned} &n_{\mathrm{R}}\mathsf{E}\left[\log\|\mathbb{H}\boldsymbol{x}+\boldsymbol{Z}'\|^2\right] - h(\mathbb{H}\boldsymbol{x}+\boldsymbol{Z}') \\ &\quad\leq n_{\mathrm{R}}\log\left(\frac{\mathsf{E}[\|\boldsymbol{Z}'\|^2] + \mathsf{E}[\|\mathbb{H}\|_{\mathrm{F}}^2]\|\boldsymbol{x}\|^2}{\|\boldsymbol{x}\|^2}\right) - \inf_{\|\hat{\boldsymbol{x}}\|=1} h(\mathbb{H}\hat{\boldsymbol{x}}) \end{aligned} \tag{309}$$

which demonstrates that the mapping (305) is bounded outside the unit ball. To demonstrate that the mapping is also bounded inside the unit ball, we note that

$$\begin{aligned} &n_{\mathrm{R}}\mathsf{E}\left[\log\|\mathbb{H}\boldsymbol{x}+\boldsymbol{Z}'\|^2\right] - h(\mathbb{H}\boldsymbol{x}+\boldsymbol{Z}') \\ &\quad\leq n_{\mathrm{R}}\log\left(\mathsf{E}\left[\|\boldsymbol{Z}'\|^2\right] + \mathsf{E}\left[\|\mathbb{H}\|_{\mathrm{F}}^2\right]\|\boldsymbol{x}\|^2\right) - h(\boldsymbol{Z}') \end{aligned} \tag{310}$$

which follows from (306) and the inequality

$$h(\mathbb{H}\boldsymbol{x}+\boldsymbol{Z}') \geq h(\boldsymbol{Z}'). \qquad \square$$

The following lemma is provided for completeness. It will not be used in subsequent sections. It demonstrates that if $\boldsymbol{Z}$ is Gaussian, then the mapping (305) is continuous.

Lemma 6.14: Let $\mathbb{H}$ be a complex random $n_{\mathrm{R}} \times n_{\mathrm{T}}$ matrix satisfying $h(\mathbb{H}) > -\infty$ and $\mathsf{E}[\|\mathbb{H}\|_{\mathrm{F}}^2] < \infty$. Fix some $\sigma^2 > 0$ and let $\boldsymbol{Z} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \sigma^2 \mathsf{I}_{n_{\mathrm{R}}})$ be independent of $\mathbb{H}$. Then the function

$$\boldsymbol{x} \mapsto n_{\mathrm{R}}\mathsf{E}\left[\log\|\mathbb{H}\boldsymbol{x}+\boldsymbol{Z}\|^2\right] - h(\mathbb{H}\boldsymbol{x}+\boldsymbol{Z}) \tag{311}$$

is a continuous function of $\boldsymbol{x} \in \mathbb{C}^{n_{\mathrm{R}}}$.

Proof: By Lemma 6.12, continuity will be established once we demonstrate that the mapping

$$\boldsymbol{x} \mapsto \mathsf{E}\left[\log\|\mathbb{H}\boldsymbol{x}+\boldsymbol{Z}\|^2\right] \tag{312}$$

is continuous. Let $\{\boldsymbol{x}_n\}$ converge to $\boldsymbol{x}$. We now have

$$\mathsf{E}\left[\log\|\mathbb{H}\boldsymbol{x}_n+\boldsymbol{Z}\|^2\right] = \int_{\mathsf{H}}\int_{\boldsymbol{z}} \log\|\mathsf{H}\boldsymbol{x}_n+\boldsymbol{z}\|^2\,\mathrm{d}F_{\boldsymbol{Z}}(\boldsymbol{z})\,\mathrm{d}F_{\mathbb{H}}(\mathsf{H}).$$

By the explicit formula for the expectation of the logarithm of a noncentral chi-squared random variable (209) it follows that the inner integral converges, i.e., that

$$\begin{aligned} &\lim_{n\to\infty}\int \log\|\mathsf{H}\boldsymbol{x}_n+\boldsymbol{z}\|^2\,\mathrm{d}F_{\boldsymbol{Z}}(\boldsymbol{z}) \\ &\qquad = \int \log\|\mathsf{H}\boldsymbol{x}+\boldsymbol{z}\|^2\,\mathrm{d}F_{\boldsymbol{Z}}(\boldsymbol{z}), \qquad \forall\,\mathsf{H}. \end{aligned} \tag{313}$$

The required continuity now follows from the Dominated Convergence Theorem as follows. We first note that $\|\mathsf{H}\boldsymbol{x}_n+\boldsymbol{Z}\|^2$ is by Lemma 6.2 stochastically larger than $\|\boldsymbol{Z}\|^2$ so that by the monotonicity of the logarithmic function

$$\begin{aligned} \mathsf{E}\left[\log\|\mathsf{H}\boldsymbol{x}_n+\boldsymbol{Z}\|^2\right] &\geq \mathsf{E}\left[\log\|\boldsymbol{Z}\|^2\right] \\ &= \log\sigma^2 + \psi(n_{\mathrm{R}}), \qquad \forall\,\mathsf{H}. \end{aligned} \tag{314}$$

Next, we note that by Jensen's inequality and the definition of the norm

$$\begin{aligned} \mathsf{E}\left[\log\|\mathsf{H}\boldsymbol{x}_n+\boldsymbol{Z}\|^2\right] &\leq \log\mathsf{E}\left[\|\mathsf{H}\boldsymbol{x}_n+\boldsymbol{Z}\|^2\right] \\ &= \log\left(\|\mathsf{H}\boldsymbol{x}_n\|^2 + n_{\mathrm{R}}\sigma^2\right) \\ &\leq \log\left(\|\mathsf{H}\|^2\|\boldsymbol{x}_n\|^2 + n_{\mathrm{R}}\sigma^2\right), \qquad \forall\,\mathsf{H}. \end{aligned} \tag{315}$$

The RHS of (314) is integrable with respect to H, and the condition $\mathsf{E}[\|\mathbb{H}\|_{\mathrm{F}}^2] < \infty$ implies that the RHS of (315) is also integrable with respect to H. Consequently, the conditions of the Dominated Convergence Theorem hold, and the required continuity is established. □

7) Change of Coordinates: The behavior of differential entropy under coordinate transformations is governed by the corresponding behavior of joint densities. Here we mention some of the results that will be used repeatedly.

We begin by relating the differential entropy of a positive random variable to the differential entropy of its logarithm and of its square.

Lemma 6.15: Let $S \geq 0$ be a nonnegative real random variable. Then

$$h(\log S) = h(S) - \mathsf{E}[\log S] \tag{316}$$

and

$$h(S^2) = h(S) + \mathsf{E}[\log S] + \log 2. \tag{317}$$

Proof: To prove (316) let $T = \log S$, and let $f_S(s)$ be the density function of S. The density $f_T(t)$ of T is then

$$f_T(t) = e^t \cdot f_S(e^t), \qquad t \in \mathbb{R}$$

so that

$$-\log f_T(T) = -\log f_S(S) - \log S$$

from which (316) follows upon taking expectations.

To prove (317) let $T = S^2$, and let $f_S(s)$ be the density function of S. Then the density $f_T(t)$ of T is

$$f_T(t) = \frac{1}{2\sqrt{t}} \cdot f_S\left(\sqrt{t}\right), \qquad t \geq 0$$

so that

$$-\log f_T(T) = -\log f_S(S) + \log S + \log 2$$

from which (317) follows again upon taking expectations. □

We next relate the differential entropy of a complex random variable to the joint differential entropy of it magnitude and phase.

Lemma 6.16: Let W be a complex random variable of differential entropy $h(W)$. Let $|W| \geq 0$ and $-\pi \leq \Theta_w < \pi$ be two real random variables designating the magnitude and phase of W, so that $W = |W| \cdot e^{j\Theta_w}$. Let $h(|W|, \Theta_w)$ denote the joint differential entropy of the pair $(|W|, \Theta_w)$ when the pair is viewed as a pair of real random variables. Then

$$h(W) = h(|W|, \Theta_w) + \mathsf{E}[\log |W|]. \tag{318}$$

In particular, if W is circularly symmetric then

$$\begin{aligned} h(W) &= \log 2\pi + h(|W|) + \mathsf{E}[\log |W|] & (319)\\ &= \log \pi + h(|W|^2), \qquad W \text{ circularly symmetric.} & (320) \end{aligned}$$

Proof: The differential entropy $h(W)$ is given by $-\mathsf{E}[\log f_W(W)]$, where f_W is the joint density function of the real and imaginary parts of W with respect to the Lebesgue measure on $\mathbb{R}^2$. The result now follows by relating the density f_W and the density $f_{|W|,\Theta_w}(|w|, \theta_w)$ with respect to polar coordinates

$$f_W(w) \cdot |w| = f_{|W|,\Theta_w}(|w|, \theta_w).$$

Equation (320) is a consequence of (317). □

The extension of this result to complex vectors is slightly more intricate. In the following, we shall relate the differential entropy $h(\boldsymbol{W})$ of a complex random vector in $\mathbb{C}^m$ to some entropy-like quantities related to its magnitude $\|\boldsymbol{W}\|$ and its direction

$$\hat{\boldsymbol{W}} = \frac{\boldsymbol{W}}{\|\boldsymbol{W}\|}. \tag{321}$$

To express the desired result, we shall need a differential entropy-like quantity for random vectors that take value on the unit-sphere in $\mathbb{C}^m$.

Let λ denote the area measure on the unit-sphere in $\mathbb{C}^m$. Let c_m denote the area of the entire unit-sphere, so that

$$c_m = \frac{2\pi^m}{\Gamma(m)}. \tag{322}$$

If a random vector $\boldsymbol{G}$ takes value in the unit-sphere and has the density $f^{\lambda}_{\boldsymbol{G}}(\boldsymbol{g})$ with respect to λ, then we shall let

$$h_\lambda(\boldsymbol{G}) = -\mathsf{E}[\log f^{\lambda}_{\boldsymbol{G}}(\boldsymbol{G})] \tag{323}$$

if the expectation is defined.

We note that just as ordinary differential entropy is invariant under translation, so is $h_\lambda(\boldsymbol{G})$ invariant under rotation. That is, if V is a deterministic unitary matrix, then

$$h_\lambda(\mathsf{V}\boldsymbol{G}) = h_\lambda(\boldsymbol{G}). \tag{324}$$

Also note that if $\boldsymbol{G}$ is uniformly distributed on the unit sphere, then $h_\lambda(\boldsymbol{G}) = \log c_m$. (Recall that we use c_m to denote the surface of the unit-sphere—see (322).)

If $\boldsymbol{W}$ is any random vector, and if conditional on $\boldsymbol{W} = \boldsymbol{w}$ the random vector $\boldsymbol{G}$ has density $f^{\lambda}_{\boldsymbol{G}|\boldsymbol{W}}(\boldsymbol{g}|\boldsymbol{W} = \boldsymbol{w})$ then we can define

$$h_\lambda(\boldsymbol{G}|\boldsymbol{W} = \boldsymbol{w}) = -\mathsf{E}\Big[\log f^{\lambda}_{\boldsymbol{G}|\boldsymbol{W}}(\boldsymbol{G}|\boldsymbol{W} = \boldsymbol{w})\Big] \tag{325}$$

and we can define $h_\lambda(\boldsymbol{G}|\boldsymbol{W})$ as the expectation (with respect to $\boldsymbol{W}$) of $h_\lambda(\boldsymbol{G}|\boldsymbol{W} = \boldsymbol{w})$.

Lemma 6.17: Let $\boldsymbol{W}$ be a complex random vector taking value in $\mathbb{C}^m$ and of differential entropy $h(\boldsymbol{W})$. Let $\|\boldsymbol{W}\|$ denote its norm and $\hat{\boldsymbol{W}}$ denotes its direction as in (321). Then

$$h(\boldsymbol{W}) = h(\|\boldsymbol{W}\|) + h_\lambda\Big(\hat{\boldsymbol{W}} \;\Big|\; \|\boldsymbol{W}\|\Big) + (2m-1)\mathsf{E}[\log \|\boldsymbol{W}\|] \tag{326}$$

whenever all the quantities in (326) are defined. Here the first term on the right is the differential entropy of $\|\boldsymbol{W}\|$ when viewed as a real (scalar) random variable.

Proof: Omitted. □

E. Isotropic Distribution

In this subsection, we recall the definition and some properties of isotropically distributed vectors and matrices.

Definition 6.18: A random vector $\boldsymbol{X}$ taking value in $\mathbb{C}^n$ is said to be uniformly distributed over the unit sphere if $\|\boldsymbol{X}\| = 1$ with probability one, and if for any deterministic unitary matrix V the distribution of $\mathsf{V}\boldsymbol{X}$ is identical to the distribution of $\boldsymbol{X}$.

Definition 6.19: A random vector $\boldsymbol{X} \in \mathbb{C}^n$ is said to be *isotropically distributed* if one of the following equivalent conditions holds.

a) For any deterministic $n \times n$ unitary matrix V, the distribution of $\mathsf{V}\boldsymbol{X}$ is identical to the distribution of $\boldsymbol{X}$.
b) The random vector $\boldsymbol{X}$ can be written in the form $\boldsymbol{X} = R\hat{\boldsymbol{X}}$, where $R \geq 0$ is a nonnegative random variable; $\hat{\boldsymbol{X}}$ is uniformly distributed over the unit sphere in $\mathbb{C}^n$; and the pair $(R, \hat{\boldsymbol{X}})$ are independent.
c) For any random $n \times n$ unitary matrix $\mathbb{V}$ that is independent of $\boldsymbol{X}$, the law of $\boldsymbol{X}$ is identical to the law of $\mathbb{V}\boldsymbol{X}$.

The most important example of an isotropically distributed random vector $\boldsymbol{X}$ is the one whose components are i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$. Any multiplication of such a vector by an independent nonnegative random variable also results in an isotropically distributed random vector.

Definition 6.20: We shall say that a random $n \times n$ unitary matrix $\mathbb{U}$ is Haar distributed if it is distributed according to the Haar measure on the set of all unitary $n \times n$ matrices. That is, if $\mathbb{U}$ is with probability one unitary, and if for any $n \times n$ deterministic unitary matrix V the law of $\mathsf{V}\mathbb{U}$ is identical to the law of $\mathbb{U}$.

Definition 6.21: We shall say that an $n \times n$ random matrix $\mathbb{A}$ is isotropically distributed if for any deterministic $n \times n$ unitary matrix V the law of $\mathsf{V}\mathbb{A}$ is identical to the law of $\mathbb{A}$.

Lemma 6.22: Let $\mathbb{A}$ by an $n \times n$ isotropically distributed random matrix. Then we have the following.

1) For any deterministic vector $\boldsymbol{x} \in \mathbb{C}^n$, the vector $\mathbb{A}\boldsymbol{x}$ is isotropically distributed.
2) For any random vector $\boldsymbol{X} \in \mathbb{C}^n$ that is independent of $\mathbb{A}$, the vector $\mathbb{A}\boldsymbol{X}$ is isotropically distributed.
3) Each of the columns of $\mathbb{A}$ is an isotropically distributed random vector.
4) If V is any deterministic unitary $n \times n$ complex matrix, then the law of $\mathbb{A}\mathsf{V}$ is identical to the law of $\mathbb{A}$.

5) If $\mathbb{U}$ is any random unitary $n \times n$ matrix that is independent of $\mathbb{A}$, then the law of $\mathbb{A}\mathbb{U}$ and the law of $\mathbb{U}\mathbb{A}$ are both identical to the law of $\mathbb{A}$.

Proof: Omitted. □

We now turn now to rotation-commutative matrices. See Definition 4.34.

Lemma 6.23: Let $\mathbb{A}$ be an $n \times n$ random rotation commutative matrix.

a) If $\hat{\boldsymbol{e}}_1$ and $\hat{\boldsymbol{e}}_2$ are two deterministic unit-vectors, then the distributions of $\|\mathbb{A}\hat{\boldsymbol{e}}_1\|$ and of $\|\mathbb{A}\hat{\boldsymbol{e}}_2\|$ are identical. Equivalently, if $\boldsymbol{X}$ is independent of $\mathbb{A}$, then $\|\mathbb{A}\boldsymbol{X}\|/\|\boldsymbol{X}\|$ is independent of $\boldsymbol{X}$ and has a law that is identical to the law of $\|\mathbb{A}\hat{\boldsymbol{e}}\|$, for any deterministic unit vector $\hat{\boldsymbol{e}}$.

b) If $\boldsymbol{X}$ is an isotropically distributed random n-vector that is independent of $\mathbb{A}$, then the random vector $\mathbb{A}\boldsymbol{X}$ is isotropically distributed.

c) Let $\hat{\boldsymbol{e}} \in \mathbb{C}^n$ be some arbitrary deterministic unit-vector. For every deterministic unit vector $\hat{\boldsymbol{x}} \in \mathbb{C}^n$ let $\mathsf{V}_{\hat{\boldsymbol{x}}}$ be some deterministic unitary matrix such that $\mathsf{V}_{\hat{\boldsymbol{x}}}\hat{\boldsymbol{e}} = \hat{\boldsymbol{x}}$. Let $\boldsymbol{X}$ be an arbitrary random vector in $\mathbb{C}^n$ that is independent of $\mathbb{A}$. Then conditional on $\boldsymbol{X}/\|\boldsymbol{X}\| = \hat{\boldsymbol{x}}$, on $\|\boldsymbol{X}\| = \|\boldsymbol{x}\|$, and on $\|\mathbb{A}\boldsymbol{X}\|/\|\boldsymbol{X}\|$, the distribution of $\mathsf{V}_{\hat{\boldsymbol{x}}}^\dagger\mathbb{A}\boldsymbol{X}/\|\mathbb{A}\boldsymbol{X}\|$ does not depend on $(\hat{\boldsymbol{x}}, \|\boldsymbol{x}\|)$.

Proof: To prove Part a), let V be some deterministic unitary matrix such that $V\hat{\boldsymbol{e}}_1 = \hat{\boldsymbol{e}}_2$. Then because V is unitary we have

$$\begin{aligned}\|\mathbb{A}\hat{\boldsymbol{e}}_2\| &= \|\mathsf{V}\mathsf{V}^\dagger\mathbb{A}\mathsf{V}\hat{\boldsymbol{e}}_1\| \\ &= \|\mathsf{V}^\dagger\mathbb{A}\mathsf{V}\hat{\boldsymbol{e}}_1\|\end{aligned}$$

and the result now follows because A is rotation commutative so that the law of $\mathsf{V}^\dagger\mathbb{A}\mathsf{V}$ is the same as the law of $\mathbb{A}$.

To prove Part b), we shall show that for any deterministic unitary $n \times n$ matrix V, the law of $\mathsf{V}\mathbb{A}\boldsymbol{X}$ is identical to the law of $\mathbb{A}\boldsymbol{X}$. Let V be such a matrix. Then $\mathsf{V}\mathbb{A}\boldsymbol{X} = \mathsf{V}\mathbb{A}\mathsf{V}^\dagger\mathsf{V}\boldsymbol{X}$, which can also be written as $(\mathsf{V}\mathbb{A}\mathsf{V}^\dagger)(\mathsf{V}\boldsymbol{X})$ from which the result is apparent because the fact that $\mathbb{A}$ is isotropically distributed implies that the law of $(\mathsf{V}\mathbb{A}\mathsf{V}^\dagger)$ is the same as the law of $\mathbb{A}$, and the fact that $\boldsymbol{X}$ is isotropically distributed implies that the law of $(\mathsf{V}\boldsymbol{X})$ is the same as the law of $\boldsymbol{X}$. The independence of $\mathbb{A}$ and $\boldsymbol{X}$ guarantees that the law of the product $(\mathsf{V}\mathbb{A}\mathsf{V}^\dagger)(\mathsf{V}\boldsymbol{X})$ is determined by the individual laws of each of the terms.

The proof of Part c) relies on expressing $\mathsf{V}_{\hat{\boldsymbol{x}}}^\dagger\mathbb{A}\boldsymbol{X}/\|\mathbb{A}\boldsymbol{X}\|$ as

$$\mathsf{V}_{\hat{\boldsymbol{x}}}^\dagger \frac{\mathbb{A}\boldsymbol{X}}{\|\mathbb{A}\boldsymbol{X}\|} = \frac{\left(\mathsf{V}_{\hat{\boldsymbol{x}}}^\dagger\mathbb{A}\mathsf{V}_{\hat{\boldsymbol{x}}}\right)\hat{\boldsymbol{e}}}{\left\|\left(\mathsf{V}_{\hat{\boldsymbol{x}}}^\dagger\mathbb{A}\mathsf{V}_{\hat{\boldsymbol{x}}}\right)\hat{\boldsymbol{e}}\right\|} \tag{327}$$

and noting that, because $\mathbb{A}$ is isotropically distributed, the law of $\mathsf{V}_{\hat{\boldsymbol{x}}}^\dagger\mathbb{A}\mathsf{V}_{\hat{\boldsymbol{x}}}$ is identical to the law of $\mathbb{A}$. □

VII. Summary and Conclusion

In this paper, we proposed a technique for deriving upper bounds on channel capacity and demonstrated its use by studying the capacity of multiple-antenna systems operating over flat-fading channels with neither receiver nor transmitter side information. Extensions to receivers with partial side information were also considered. This technique has been subsequently successfully employed in the study of the capacity of other channels such as finite-state channels with only intersymbol interference memory [18, Sec. 4.6], [3], channels with both additive noise and phase noise [4], and Poisson channels [5]. Extensions to the study of error exponents are discussed in [6].

The technique is based on the inequality (11), which we extended to continuous alphabets in (186) of Theorem 5.1. To derive an upper bound on mutual information (and ultimately on channel capacity) one would typically start by judiciously picking some family of distributions on the output alphabet. Applying (186) to any output distribution $R(\cdot)$ in the family leads to an upper bound on mutual information, and if the family is sufficiently rich, the tightest such bound may be quite good.

In the study of multiple-antenna flat-fading channels we have had some success with the family of output distributions of densities (204). By applying (186) to output distributions $R(\cdot)$ in this family we obtained the upper bound (25). It should, however, be noted that (25) is not specific to fading channels. It can be applied to any channel taking output in Euclidean space.

For channels taking value in the nonnegative reals, we considered the family of regularized Gamma distributions (199), which leads (via Theorem 5.1) to the inequality (30). This inequality can be useful in the study of noncoherent channels where the channel output has a nonnegative sufficient statistic. See, for example, [4] for an application of this inequality to noncoherent communication and [5] for applications to the Poisson channel.

Using inequality (25), we derived upper bounds on the capacity of Rayleigh- and Ricean-fading channels in the absence of any receiver side information. These bounds were complemented by some new lower bounds. The bounds are tight in the sense that their difference from capacity tends to zero at high SNR and in the sense that their ratio to capacity approaches one at low SNR. For the Ricean model, where the fading is $\mathcal{N}_{\mathbb{C}}(d, 1)$ distributed, our bounds indicate the following: Up to some threshold, which is nearly a rate of $\log(1+|d|^2)$, Gaussian inputs with nearest neighbor decoding that ignores the fading are nearly optimal. Above this threshold, one needs more sophisticated coding and decoding techniques to achieve capacity, but capacity soon grows so slowly with the SNR that communication becomes extremely power inefficient, and thus of only limited engineering interest.

The poor power efficiency of communication over flat-fading channels at high SNR is not specific to the Ricean model. We have demonstrated that under very general conditions, even allowing for memory and partial receiver side information, channel capacity typically grows only double-logarithmically with the SNR; see Theorem 4.2. In an attempt to better understand the high-SNR behavior of channel capacity and to assess the rate above which capacity only increases double-logarithmically in the SNR, we have introduced the "fading number" χ as

$$\chi = \overline{\lim_{\text{SNR}\to\infty}} \left\{C(\text{SNR}) - \log\log \text{SNR}\right\}. \tag{328}$$

For channels for which the limsup in (328) is actually a limit, capacity at high SNR can be expressed as

$$C(\mathrm{SNR}) = \log\log \mathrm{SNR} + \chi + o(1) \tag{329}$$

where the $o(1)$ term decays to zero as the SNR tends to infinity. In this sense, χ is the second-order term in the high-SNR capacity expansion. It is also in this sense that it would seem that at rates that are significantly higher than χ, communication becomes power inefficient and capacity increases only double-logarithmically in the SNR.

Motivated by these interpretations of the fading number, we set out to compute its value. This computation was significantly simplified by Theorem 4.8, which allowed us to replace the additive noise with a minimal input power constraint without altering the high-SNR asymptotics. The proof of this theorem hinges on the notion of "capacity-achieving input distributions that escape to infinity," which also has applications in the high-SNR analysis of other channels with costs; see [4] and [5].

Loosely speaking, we say that the capacity of a channel can be achieved by input distributions that escape to infinity if the high-SNR channel capacity asymptotics can be achieved even if the channel input are subjected to an additional input constraint that only allows inputs of a given minimal cost. It turns out that, in addition to some technical conditions, for a channel to have this property it suffices that its capacity grows sublinearly to infinity in the cost; see Theorem 4.13. We hope that this observation may be of some use in the study of the high-SNR capacity of other channels.

By replacing the additive noise with a minimal energy constraint we were able to compute the fading number of some memoryless fading channels including SISO channels (with Rayleigh and Ricean channels as special cases), the fading number for SIMO channels, and the fading number for MISO channels. The latter was shown to be achievable using beamforming. Note, however, that the beamforming "direction" (103) is not typically the one that maximizes the SNR in the resulting SISO channel. In the Gaussian fading case, it is the direction that maximizes the specular-to-granular fading ratio (105).

For fading channels with memory the computation of the fading number is more intricate. Theorem 4.41 solves the problem for the SISO case. It hinges on the fact that the fading number of a memoryless SISO system can be achieved by input distributions that do not depend on the fading law and that are bounded away from zero so that from past inputs and outputs one can arbitrarily well estimate past fading levels. These properties also hold for SIMO channels, but not for MISO systems.

If the fading process is a stationary and ergodic regular Gaussian process then the fading number takes on a particularly simple form; see Corollary 4.42. It is interesting that for such fading and in the absence of any receiver side information, the fading number is determined by the (normalized) specular component and the mean squared error in predicting the fading value from its past. It is not directly related to such concepts as the Doppler-spread or coherence-time.

For an analysis of the case where the fading process is *not regular* (i.e., $\epsilon_{\mathrm{est}}^2 = 0$) see [34]. There a necessary and sufficient condition for a double-logarithmic capacity growth is derived; the "pre-log" for channels with logarithmic capacity growths is computed; and examples of channels with growths of the form $(\log \mathrm{SNR})^\alpha$ for $0 < \alpha < 1$ are presented.

It is instructive to compare the high-SNR behavior of channel capacity $C(\mathrm{SNR})$ in the absence of side information (329) with the behavior in its presence $C_{\mathrm{PSI}}(\mathrm{SNR})$. In its presence—if perfect—capacity (for i.i.d. zero-mean Gaussian fading) typically grows logarithmically in the SNR [17], [35], with

$$\lim_{\mathrm{SNR}\to\infty} \left\{ C_{\mathrm{PSI}}(\mathrm{SNR}) - \min\{n_{\mathrm{T}}, n_{\mathrm{R}}\} \cdot \log \mathrm{SNR} \right\} > -\infty \tag{330}$$

which is in stark contrast to (329).

(The formula we derived for the expectation of the logarithm of a noncentral chi-square random variable (209) also allows us to evaluate the high-SNR channel capacity of some fading channels with a nonzero mean in the presence of perfect receiver side information, e.g., the Ricean channel (181) and SIMO Gaussian channels with mean.)

Other models also lead to results that are dramatically different from (329). For example, the block constant fading model of [36] was analyzed at high SNR in [37]. It was shown there that at high-SNR capacity is given asymptotically as

$$M^*(1 - M^*/T)\log \mathrm{SNR} + O(1) \tag{331}$$

where $M^* = \min\{n_{\mathrm{T}}, n_{\mathrm{R}}, \lfloor T/2 \rfloor\}$ and $T \geq 2$ is the number of symbols over which the channel remains constant.

It is thus seen that at high SNR, the behavior of the capacity of fading channels depends critically on the assumed fading model.

APPENDIX I
LOW-SNR ANALYSIS

In this appendix, we discuss the bound (27) at low SNR. In particular, we show that the choice $\alpha = n_{\mathrm{R}}$ and A satisfying $\mathsf{A}^\dagger\mathsf{A} = (\mathsf{E}[\boldsymbol{Y}\boldsymbol{Y}^\dagger])^{-1}$ yields the max-entropy bound (28) and demonstrate that this bound is tight enough to give the right capacity–energy slope at zero SNR. We begin with the former task.

Substituting $\alpha = n_{\mathrm{R}}$ in (27) yields the bound

$$I(Q; W) \leq \log(\pi \cdot e)^{n_{\mathrm{R}}} + n_{\mathrm{R}} \log \mathsf{E}_Q\left[\|\mathsf{A}\boldsymbol{Y}\|^2\right] - \log\det(\mathsf{A}\mathsf{A}^\dagger) - n_{\mathrm{R}}\log n_{\mathrm{R}} - h_Q(\boldsymbol{Y}|\boldsymbol{X}).$$

Computing $\mathsf{E}_Q[\|\mathsf{A}\boldsymbol{Y}\|^2]$ and substituting

$$\mathsf{A}^\dagger\mathsf{A} = (\mathsf{E}[\boldsymbol{Y}\boldsymbol{Y}^\dagger])^{-1}$$

yields

$$\begin{aligned}
\mathsf{E}_Q\left[\|\mathsf{A}\boldsymbol{Y}\|^2\right] &= \mathrm{tr}\left(\mathsf{E}\left[\mathsf{A}\boldsymbol{Y}(\mathsf{A}\boldsymbol{Y})^\dagger\right]\right)\\
&= \mathrm{tr}\left(\mathsf{A}\mathsf{E}\left[\boldsymbol{Y}\boldsymbol{Y}^\dagger\right]\mathsf{A}^\dagger\right)\\
&= \mathrm{tr}\left(\mathsf{A}\mathsf{E}\left[\boldsymbol{Y}\boldsymbol{Y}^\dagger\right]\mathsf{A}^\dagger\mathsf{A}\mathsf{A}^{-1}\right)\\
&= \mathrm{tr}\left(\mathsf{A}\mathsf{A}^{-1}\right)\\
&= n_{\mathrm{R}}
\end{aligned}$$

and

$$-\log\det\left(\mathsf{A}\mathsf{A}^\dagger\right) = \log\det \mathsf{E}\left[\boldsymbol{Y}\boldsymbol{Y}^\dagger\right]$$

thus establishing (28).

We now turn to demonstrating tightness at low SNR. Using the bound $h(\boldsymbol{Y}|\boldsymbol{X}) \geq h(\boldsymbol{Y}|\boldsymbol{X}, \mathbb{H}) = h(\boldsymbol{Z})$ we obtain from (28)

$$\begin{aligned} I(\boldsymbol{X}; \boldsymbol{Y}) &\leq \log\det\left(\mathsf{I} + \frac{\mathsf{E}[\mathbb{H}\boldsymbol{X}\boldsymbol{X}^\dagger\mathbb{H}^\dagger]}{\sigma^2}\right) \\ &\leq \frac{1}{\sigma^2}\operatorname{tr}\left(\mathsf{E}\left[\mathbb{H}\boldsymbol{X}\boldsymbol{X}^\dagger\mathbb{H}^\dagger\right]\right) \\ &= \frac{1}{\sigma^2}\mathsf{E}\left[\|\mathbb{H}\boldsymbol{X}\|^2\right] \\ &\leq \frac{1}{\sigma^2}\mathsf{E}\left[\|\boldsymbol{X}\|^2\right]\max_{\|\hat{\boldsymbol{x}}\|=1}\mathsf{E}\left[\|\mathbb{H}\hat{\boldsymbol{x}}\|^2\right] \\ &\leq \frac{\mathcal{E}_{\mathrm{s}}}{\sigma^2}\max_{\|\hat{\boldsymbol{x}}\|=1}\hat{\boldsymbol{x}}^\dagger\mathsf{E}\left[\mathbb{H}^\dagger\mathbb{H}\right]\hat{\boldsymbol{x}} \\ &= \frac{\mathcal{E}_{\mathrm{s}}}{\sigma^2}\lambda_{\max}\left(\mathsf{E}\left[\mathbb{H}^\dagger\mathbb{H}\right]\right) \end{aligned}$$

where $\lambda_{\max}(\mathsf{E}[\mathbb{H}^\dagger\mathbb{H}])$ denotes the maximal eigenvalue of the matrix $\mathsf{E}[\mathbb{H}^\dagger\mathbb{H}]$. At low SNR this agrees with the asymptotic expression [38]

$$\lim_{\mathcal{E}_{\mathrm{s}}\downarrow 0}\frac{\sigma^2}{\mathcal{E}_{\mathrm{s}}}C(\mathcal{E}_{\mathrm{s}}/\sigma^2) = \lambda_{\max}\left(\mathsf{E}\left[\mathbb{H}^\dagger\mathbb{H}\right]\right). \tag{332}$$

Appendix II
A Proof of Theorem 4.2

Proof: In view of Lemma 4.5 it suffices to prove this theorem in the case where the fading is memoryless, the additive noise is memoryless, and the side information is null. Consequently, to simplify notation, we remove all time indexes. Also, since the mean of the noise can be subtracted off at the receiver, we shall assume throughout $\mathsf{E}[\boldsymbol{Z}'] = 0$.

The proof of the theorem is based on a study of the bound (27) when it is applied to input distributions Q satisfying the average power constraint $\mathsf{E}_Q[\|\boldsymbol{X}\|^2] \leq \mathcal{E}_{\mathrm{s}}$. In fact, it will suffice to consider (27) with the possibly suboptimal choice of the matrix A as the following identity matrix:

$$\begin{aligned} I(Q; W) \leq{}& \log\pi^{n_{\mathrm{R}}} - \log\Gamma(n_{\mathrm{R}}) \\ &+ n_{\mathrm{R}}\mathsf{E}_Q\left[\log\|\boldsymbol{Y}\|^2\right] - h_Q(\boldsymbol{Y}|\boldsymbol{X}) \\ &+ \alpha\left(1 + \log\mathsf{E}_Q\left[\|\boldsymbol{Y}\|^2\right] - \mathsf{E}_Q\left[\log\|\boldsymbol{Y}\|^2\right]\right) \\ &+ \log\Gamma(\alpha) - \alpha\log\alpha, \qquad \alpha > 0. \end{aligned} \tag{333}$$

We begin noting that by Lemma 6.13

$$\sup_{\boldsymbol{x}\in\mathbb{C}^{n_{\mathrm{R}}}}\left\{n_{\mathrm{R}}\mathsf{E}\left[\log\|\boldsymbol{Y}\|^2\middle|\boldsymbol{X} = \boldsymbol{x}\right] - h(\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x})\right\} < \infty \tag{334}$$

and consequently

$$\sup_{Q}\left\{n_{\mathrm{R}}\mathsf{E}\left[\log\|\boldsymbol{Y}\|^2\right] - h_Q(\boldsymbol{Y}|\boldsymbol{X})\right\} < \infty \tag{335}$$

where the supremum is over *all* input distributions, irrespective of their power. We now continue the proof of the theorem with a study of the remaining terms in (333). We begin by noting that if $\log\mathsf{E}[\|\boldsymbol{Y}\|^2] - \mathsf{E}[\log\|\boldsymbol{Y}\|^2]$ does not approach infinity with the SNR, then the theorem follows directly from (333) and (335) by choosing any fixed $\alpha > 0$. In fact, in this case the capacity is bounded in the SNR.

The more interesting case is, of course, when $\log\mathsf{E}[\|\boldsymbol{Y}\|^2] - \mathsf{E}[\log\|\boldsymbol{Y}\|^2]$ does tend to infinity with the SNR. In this case, we shall derive the theorem by focusing on (333) with the choice

$$\alpha^* = \left(1 + \log\mathsf{E}\left[\|\boldsymbol{Y}\|^2\right] - \mathsf{E}\left[\log\|\boldsymbol{Y}\|^2\right]\right)^{-1} \tag{336}$$

where $\alpha^* \downarrow 0$ with the SNR.

For small values of α^*, we note that since $\Gamma(z)$ is analytic at $z = 1$ with $\Gamma(1) = 1$, we obtain from the relationship

$$\Gamma(\alpha) = \frac{1}{\alpha}\Gamma(\alpha + 1)$$

that

$$\log\Gamma(\alpha^*) = \log\frac{1}{\alpha^*} + o(1) \tag{337}$$

where the correction term $o(1)$ tends to zero as α^* tends to zero. Consequently, since $\alpha^*\log\alpha^* = o(1)$ we obtain from (333) and (336) the bound

$$\begin{aligned} C \leq{}& n_{\mathrm{R}}\log\pi - \log\Gamma(n_{\mathrm{R}}) \\ &+ \sup_{\boldsymbol{x}\in\mathbb{C}^{n_{\mathrm{R}}}}\left\{n_{\mathrm{R}}\mathsf{E}\left[\log\|\boldsymbol{Y}\|^2\middle|\boldsymbol{X} = \boldsymbol{x}\right] - h(\boldsymbol{Y}|\boldsymbol{X} = \boldsymbol{x})\right\} \\ &+ 1 + \log\frac{1}{\alpha^*} + o(1). \end{aligned} \tag{338}$$

The theorem will now follow from (334) and (338) once we obtain the logarithmic bound on $1/\alpha^*$

$$\begin{aligned} \frac{1}{\alpha^*} &= 1 + \log\mathsf{E}\left[\|\boldsymbol{Y}\|^2\right] - \mathsf{E}\left[\log\|\boldsymbol{Y}\|^2\right] \\ &\leq 1 + \log\left(\mathsf{E}\left[\|\mathbb{H}\|_{\mathrm{F}}^2\right]\mathcal{E}_{\mathrm{s}} + \mathsf{E}\left[\|\boldsymbol{Z}'\|^2\right]\right) \\ &\quad - \inf_{\boldsymbol{c}\in\mathbb{C}^{n_{\mathrm{R}}}}\mathsf{E}\left[\log\|\boldsymbol{Z}' + \boldsymbol{c}\|^2\right] \end{aligned} \tag{339}$$

where the last term in the above is finite by (256) (evaluated at $\delta = 1$).

This bound can be derived by upper-bounding $\log\mathsf{E}[\|\boldsymbol{Y}\|^2]$ by

$$\begin{aligned} \log\mathsf{E}\left[\|\boldsymbol{Y}\|^2\right] &= \log\left(\mathsf{E}\left[\|\mathbb{H}\boldsymbol{X}\|^2\right] + \mathsf{E}\left[\|\boldsymbol{Z}'\|^2\right]\right) \\ &\leq \log\left(\mathsf{E}\left[\|\mathbb{H}\|^2\right]\mathsf{E}\left[\|\boldsymbol{X}\|^2\right] + \mathsf{E}\left[\|\boldsymbol{Z}'\|^2\right]\right) \\ &\leq \log\left(\mathsf{E}\left[\|\mathbb{H}\|_{\mathrm{F}}^2\right]\mathcal{E}_{\mathrm{s}} + \mathsf{E}\left[\|\boldsymbol{Z}'\|^2\right]\right) \end{aligned} \tag{340}$$

and by lower-bounding $\mathsf{E}[\log\|\boldsymbol{Y}\|^2]$ by

$$\mathsf{E}\left[\log\|\boldsymbol{Y}\|^2\right] = \mathsf{E}\left[\log\|\mathbb{H}\boldsymbol{X} + \boldsymbol{Z}'\|^2\right] \tag{341}$$

$$\geq \inf_{\boldsymbol{c}\in\mathbb{C}^{n_{\mathrm{R}}}}\mathsf{E}\left[\log\|\boldsymbol{Z}' + \boldsymbol{c}\|^2\right] \tag{342}$$

as can be justified by the independence of $\mathbb{H}\boldsymbol{X}$ and $\boldsymbol{Z}'$. □

Appendix III
A Proof of Theorem 4.3

Proof: In view of Lemma 4.5 it suffices to prove this in the memoryless case. We shall, therefore, proceed to treat this case, and dispense with all time indexes. Expanding the mutual information term we obtain using the data processing inequality

$$\begin{aligned} & I\left(\sqrt{\mathcal{E}_{\mathrm{s}}}\boldsymbol{X};\, \sqrt{\mathcal{E}_{\mathrm{s}}}\mathbb{H}\boldsymbol{X} + \boldsymbol{Z}'\right) \\ &\quad \leq I\left(\sqrt{\mathcal{E}_{\mathrm{s}}}\boldsymbol{X};\, \sqrt{\mathcal{E}_{\mathrm{s}}}\mathbb{H}\boldsymbol{X}\right) \\ &\quad = I\left(\boldsymbol{X};\, \mathbb{H}\boldsymbol{X}\right) \\ &\quad = h(\mathbb{H}\boldsymbol{X}) - h(\mathbb{H}\boldsymbol{X}|\boldsymbol{X}) \\ &\quad = h(\mathbb{H}\boldsymbol{X}) - 2n_{\mathrm{R}}\mathsf{E}[\log\|\boldsymbol{X}\|] - h\left(\mathbb{H}\frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}\middle|\boldsymbol{X}\right). \end{aligned}$$

The proof is now concluded by noting that, since i.i.d. Gaussians maximize differential entropy subject to an expected squared-norm constraint

$$\begin{aligned} h(\mathbb{H}\boldsymbol{X}) &\leq n_{\mathrm{R}} \log \frac{\pi e \mathsf{E}[\|\mathbb{H}\boldsymbol{X}\|^2]}{n_{\mathrm{R}}} \\ &\leq n_{\mathrm{R}} \log \frac{\pi e \mathsf{E}[\|\mathbb{H}\|^2 \cdot \|\boldsymbol{X}\|^2]}{n_{\mathrm{R}}} \\ &= n_{\mathrm{R}} \log \frac{\pi e \mathsf{E}[\|\mathbb{H}\|^2]}{n_{\mathrm{R}}} \\ &< \infty \end{aligned}$$

and by noting that by Lemma 6.6

$$\begin{aligned} h\left(\mathbb{H}\, \frac{\tilde{\boldsymbol{X}}}{\|\tilde{\boldsymbol{X}}\|} \middle| \tilde{\boldsymbol{X}}\right) &\geq \inf_{\|\hat{\boldsymbol{x}}\|=1} h(\mathbb{H}\hat{\boldsymbol{x}}) \\ &> -\infty. \end{aligned}$$

□

APPENDIX IV
A PROOF OF LEMMA 4.5

Proof: The first inequality is a simple consequence of the chain rule and the basic properties of mutual information

$$\begin{aligned} I(\boldsymbol{X}_1^n;\boldsymbol{Y}_1^n,\boldsymbol{S}_1^n) &= I(\boldsymbol{X}_1^n;\boldsymbol{Y}_1^n) + I(\boldsymbol{X}_1^n;\boldsymbol{S}_1^n|\boldsymbol{Y}_1^n) \\ &= I(\boldsymbol{X}_1^n;\boldsymbol{Y}_1^n) + I(\boldsymbol{X}_1^n,\boldsymbol{Y}_1^n;\boldsymbol{S}_1^n) - I(\boldsymbol{Y}_1^n;\boldsymbol{S}_1^n) \\ &\leq I(\boldsymbol{X}_1^n;\boldsymbol{Y}_1^n) + I(\boldsymbol{X}_1^n,\boldsymbol{Y}_1^n;\boldsymbol{S}_1^n) \\ &\leq I(\boldsymbol{X}_1^n;\boldsymbol{Y}_1^n) + I(\boldsymbol{X}_1^n,\boldsymbol{Y}_1^n,\mathbb{H}_1^n;\boldsymbol{S}_1^n) \\ &= I(\boldsymbol{X}_1^n;\boldsymbol{Y}_1^n) + I(\mathbb{H}_1^n;\boldsymbol{S}_1^n). \end{aligned}$$

The second inequality can be argued by using the chain rule to write

$$I(\boldsymbol{X}_1^n;\boldsymbol{Y}_1^n) = \sum_{k=1}^{n} I(\boldsymbol{X}_1^n;\boldsymbol{Y}_k|\boldsymbol{Y}_1^{k-1}) \tag{343}$$

and by studying the term $I(\boldsymbol{X}_1^n;\boldsymbol{Y}_k|\boldsymbol{Y}_1^{k-1})$ as follows:

$$\begin{aligned} &I(\boldsymbol{X}_1^n;\boldsymbol{Y}_k|\boldsymbol{Y}_1^{k-1}) \\ &\quad = I(\boldsymbol{X}_1^n,\boldsymbol{Y}_1^{k-1};\boldsymbol{Y}_k) - I(\boldsymbol{Y}_k;\boldsymbol{Y}_1^{k-1}) \\ &\quad \leq I(\boldsymbol{X}_1^n,\boldsymbol{Y}_1^{k-1};\boldsymbol{Y}_k) \\ &\quad = I(\boldsymbol{X}_1^{k-1},\boldsymbol{Y}_1^{k-1},\boldsymbol{X}_k;\boldsymbol{Y}_k) \\ &\quad \leq I(\boldsymbol{X}_1^{k-1},\boldsymbol{Y}_1^{k-1},\mathbb{H}_1^{k-1},\boldsymbol{Z}_1'^{k-1},\boldsymbol{X}_k;\boldsymbol{Y}_k) \\ &\quad = I(\mathbb{H}_1^{k-1},\boldsymbol{Z}_1'^{k-1},\boldsymbol{X}_k;\boldsymbol{Y}_k) \\ &\quad = I(\boldsymbol{X}_k;\boldsymbol{Y}_k) + I(\mathbb{H}_1^{k-1},\boldsymbol{Z}_1'^{k-1};\boldsymbol{Y}_k|\boldsymbol{X}_k) \\ &\quad = I(\boldsymbol{X}_k;\boldsymbol{Y}_k) + I(\mathbb{H}_1^{k-1},\boldsymbol{Z}_1'^{k-1};\boldsymbol{X}_k,\boldsymbol{Y}_k) \\ &\quad \leq I(\boldsymbol{X}_k;\boldsymbol{Y}_k) + I(\mathbb{H}_1^{k-1},\boldsymbol{Z}_1'^{k-1};\boldsymbol{Z}_k',\mathbb{H}_k,\boldsymbol{X}_k,\boldsymbol{Y}_k) \\ &\quad = I(\boldsymbol{X}_k;\boldsymbol{Y}_k) + I(\mathbb{H}_1^{k-1},\boldsymbol{Z}_1'^{k-1};\boldsymbol{Z}_k',\mathbb{H}_k) \\ &\quad = I(\boldsymbol{X}_k;\boldsymbol{Y}_k) + I(\mathbb{H}_k;\mathbb{H}_1^{k-1}) + I(\boldsymbol{Z}_k';\boldsymbol{Z}_1'^{k-1}) \\ &\quad \leq I(\boldsymbol{X}_k;\boldsymbol{Y}_k) + I(\mathbb{H}_n;\mathbb{H}_1^{n-1}) + I(\boldsymbol{Z}_n';\boldsymbol{Z}_1'^{n-1}). \end{aligned}$$

□

APPENDIX V
A PROOF OF LEMMA 4.7

Proof: To simplify notation, we shall prove this lemma for the memoryless case. The general case follows along similar lines but with more cumbersome notation. We prove this lemma in two steps. In the first, we shall show that $\chi(\mathbb{H})$ is invariant with respect to multiplication by a deterministic matrix on the right, namely,

$$\chi(\mathbb{H}\mathsf{F}) = \chi(\mathbb{H}), \qquad \det \mathsf{F} \neq 0. \tag{344}$$

In the second step, we shall demonstrate invariance on the left, i.e.,

$$\chi(\mathsf{G}\mathbb{H}) = \chi(\mathbb{H}), \qquad \det \mathsf{G} \neq 0. \tag{345}$$

To prove (344) it suffices to show that

$$\chi(\mathbb{H}) \leq \chi(\mathbb{H}\mathsf{F}), \qquad \det \mathsf{F} \neq 0, \tag{346}$$

because an application of the inequality to the random matrix $\mathbb{H}\mathsf{F}$ and the deterministic matrix F^{-1} will prove the reverse inequality.

To proceed with the proof of (346) we write

$$\begin{aligned} I(\boldsymbol{X};\, \mathbb{H}\boldsymbol{X} + \boldsymbol{Z}) &= I(\mathsf{F}^{-1}\boldsymbol{X};\, \mathbb{H}\boldsymbol{X} + \boldsymbol{Z}) \\ &= I(\mathsf{F}^{-1}\boldsymbol{X};\, \mathbb{H}\mathsf{F}\mathsf{F}^{-1}\boldsymbol{X} + \boldsymbol{Z}) \\ &= I\left(\tilde{\boldsymbol{X}};\, \mathbb{H}\mathsf{F}\tilde{\boldsymbol{X}} + \boldsymbol{Z}\right) \end{aligned}$$

where $\tilde{\boldsymbol{X}} = \mathsf{F}^{-1}\boldsymbol{X}$. We now note that

$$\mathsf{E}\left[\left\|\tilde{\boldsymbol{X}}\right\|^2\right] \leq \|\mathsf{F}^{-1}\|^2 \mathsf{E}\left[\|\boldsymbol{X}\|^2\right] \tag{347}$$

so that the capacity of the channel $\boldsymbol{X} \mapsto \mathbb{H}\boldsymbol{X} + \boldsymbol{Z}$ with average power $\mathcal{E}_{\mathrm{s}}$ is no larger than the capacity of the channel $\tilde{\boldsymbol{X}} \mapsto \mathbb{H}\mathsf{F}\tilde{\boldsymbol{X}} + \boldsymbol{Z}$ with average power $\|\mathsf{F}^{-1}\|^2\mathcal{E}_{\mathrm{s}}$. This proves (346) in view of (46).

We next prove (345) for which it suffices to show

$$\chi(\mathbb{H}) \leq \chi(\mathsf{G}\mathbb{H}), \qquad \det \mathsf{G} \neq 0. \tag{348}$$

To prove (348) write

$$\begin{aligned} I(\boldsymbol{X};\, \mathbb{H}\boldsymbol{X} + \boldsymbol{Z}) &= I(\boldsymbol{X};\, \mathsf{G}\mathbb{H}\boldsymbol{X} + \mathsf{G}\boldsymbol{Z}) \\ &= I\left(\boldsymbol{X};\, \mathsf{G}\mathbb{H}\boldsymbol{X} + \tilde{\boldsymbol{Z}}\right) \end{aligned}$$

where $\tilde{\boldsymbol{Z}} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0},\, \sigma^2\mathsf{G}\mathsf{G}^\dagger)$ can be written as $\tilde{\boldsymbol{Z}} = \tilde{\boldsymbol{Z}}_1 + \tilde{\boldsymbol{Z}}_2$ where $\tilde{\boldsymbol{Z}}_1$ and $\tilde{\boldsymbol{Z}}_2$ are independent zero-mean circularly symmetric multivariate Gaussians of covariances $\sigma^2\lambda_{\min}\mathsf{I}_{n_{\mathrm{R}}}$ and $\sigma^2\mathsf{G}\mathsf{G}^\dagger - \sigma^2\lambda_{\min}\mathsf{I}_{n_{\mathrm{R}}}$, respectively. Here $\lambda_{\min}$ denotes the minimal eigenvalue of $\mathsf{G}\mathsf{G}^\dagger$. Thus, we have

$$\begin{aligned} I(\boldsymbol{X};\, \mathbb{H}\boldsymbol{X} + \boldsymbol{Z}) &= I\left(\boldsymbol{X};\, \mathsf{G}\mathbb{H}\boldsymbol{X} + \tilde{\boldsymbol{Z}}\right) \\ &= I\left(\boldsymbol{X};\, \mathsf{G}\mathbb{H}\boldsymbol{X} + \tilde{\boldsymbol{Z}}_1 + \tilde{\boldsymbol{Z}}_2\right) \\ &\leq I\left(\boldsymbol{X};\, \mathsf{G}\mathbb{H}\boldsymbol{X} + \tilde{\boldsymbol{Z}}_1\right). \end{aligned}$$

We thus conclude that the capacity of the channel $\boldsymbol{X} \mapsto \mathbb{H}\boldsymbol{X}+\boldsymbol{Z}$ with noise variance σ^2 (and SNR given by $\mathcal{E}_{\mathrm{s}}/\sigma^2$) is no larger than the capacity of the channel with fading matrix $\mathsf{G}\mathbb{H}$ and noise variance $\sigma^2\lambda_{\min}$ (and SNR given by $\mathcal{E}_{\mathrm{s}}/(\sigma^2\lambda_{\min})$). This demonstrates (348) in view of (46). □

APPENDIX VI
A PROOF OF THEOREM 4.13

Proof: Fix some $\Upsilon_0 > 0$ and let $\mathcal{K} = \{x \in \mathcal{X}: g(x) \leq \Upsilon_0\}$. For any $\upsilon > 0$ and any probability law Q on $\mathcal{X}$ define

$$\begin{aligned}\tilde{I}_\upsilon(Q) &= I(Q; W) - \upsilon \mathsf{E}_Q[g(X)] \\ &= \int_{\mathcal{X}} (D(W(\cdot|x)\|(QW)(\cdot)) - \upsilon g(x))\, \mathrm{d}Q(x). \quad (349)\end{aligned}$$

Define also the function $F(\upsilon)$ for any $\upsilon > 0$ by

$$\begin{aligned}F(\upsilon) &= \sup_Q \tilde{I}_\upsilon(Q) \quad (350)\\ &= \max_{\Upsilon \geq 0}\{C(\Upsilon) - \upsilon\Upsilon\}. \quad (351)\end{aligned}$$

Note that the concavity of $C(\cdot)$ and its sublinear growth imply

$$\lim_{\Upsilon\uparrow\infty} C'_-(\Upsilon) = \lim_{\Upsilon\uparrow\infty} C'_+(\Upsilon) = 0 \quad (352)$$

where $C'_-(\Upsilon)$ and $C'_+(\Upsilon)$ denote the left and right derivatives of $C(\cdot)$ at Υ. This observation combined with the duality $F(\cdot) \leftrightarrow C(\cdot)$ [1, Lemma 3.1, Sec. 2.3] and the observation that if two probabilitiy measures assign a small probability to $\mathcal{K}$ then so does any convex combination of the two, show that to prove the theorem it suffices to show that for any $\epsilon > 0$ and any $\upsilon > 0$ there exists some law Q_υ such that

$$\tilde{I}_\upsilon(Q_\upsilon) > F(\upsilon) - \epsilon \quad (353)$$

and such that

$$\lim_{\upsilon\downarrow 0} Q_\upsilon(\mathcal{K}) = 0. \quad (354)$$

To prove this we begin by noting that

$$\lim_{\upsilon\downarrow 0} F(\upsilon) = \infty. \quad (355)$$

This follows because

$$\begin{aligned}F(\upsilon) &= \max_{\Upsilon\geq 0}\{C(\Upsilon) - \upsilon\Upsilon\}\\ &\geq C(\Upsilon) - \upsilon\Upsilon|_{\Upsilon=1/\upsilon}\\ &= C(1/\upsilon) - 1\\ &\to \infty\end{aligned}$$

where the last step follows by the unboundedness of $C(\cdot)$.

Fix now some $\epsilon > 0$. For any $\upsilon > 0$ it follows from the definition of (350) that there exists some probability measure $Q_\upsilon^{(1)}$ such that

$$\tilde{I}_\upsilon\left(Q_\upsilon^{(1)}\right) > F(\upsilon) - \frac{\epsilon}{2}. \quad (356)$$

Define now $I_{\max}(\mathcal{K})$ as the supremum of the mutual information corresponding to the channel $W(\cdot|\cdot)$ over all input distributions whose support is contained in $\mathcal{K}$

$$I_{\max}(\mathcal{K}) = \sup_{Q:\, Q(\mathcal{K})=1} I(Q; W). \quad (357)$$

Notice that $I_{\max}(\mathcal{K}) \leq C(\Upsilon_0)$, so that in particular

$$I_{\max}(\mathcal{K}) < \infty. \quad (358)$$

By (357) and (355), it now follows that for all sufficiently small $\upsilon > 0$ we have $Q_\upsilon^{(1)}(\mathcal{K}^c) > 0$, so that for such small υ's we can define a probability measure $Q_\upsilon^{(0)}$ as the conditional distribution of $Q_\upsilon^{(1)}$ conditioned on $x \notin \mathcal{K}$

$$Q_\upsilon^{(0)}(\mathcal{B}) = \frac{Q_\upsilon^{(1)}(\mathcal{B}\cap\mathcal{K}^c)}{Q_\upsilon^{(1)}(\mathcal{K}^c)}, \qquad \mathcal{B} \subset \mathcal{X} \text{ Borel.} \quad (359)$$

Note that in particular

$$Q_\upsilon^{(0)}(\mathcal{K}) = 0. \quad (360)$$

While $Q_\upsilon^{(1)}$ does not necessarily assign a small probability to the set $\mathcal{K}$, we shall now proceed to find a possibly different measure Q_υ that satisfies

$$\tilde{I}_\upsilon(Q_\upsilon) > F(\upsilon) - \epsilon \quad (361)$$

and does. The probability Q_υ will be found in the convex hull of $\{Q_\upsilon^{(1)}, Q_\upsilon^{(0)}\}$. Let

$$Q_{\upsilon,\theta} = \theta Q_\upsilon^{(1)} + (1-\theta)Q_\upsilon^{(0)}, \qquad 0 \leq \theta \leq 1 \quad (362)$$

and consider the mapping from $[0, 1]$ to the reals $\mathbb{R}$

$$\theta \mapsto \tilde{I}_\upsilon(Q_{\upsilon,\theta}). \quad (363)$$

By the concavity of mutual information and the linearity of expectation it follows that this mapping is concave and thus continuous on $(0, 1)$ and lower semicontinuous at the endpoints. Consequently, exactly one of the following must hold:

a) the mapping's supremum over $[0, 1]$ is equal to the mapping's limit as $\theta \downarrow 0$; or
b) the supremum is equal to the mapping's limit as $\theta \uparrow 1$ and a) does not hold; or
c) the mapping's supremum over $[0, 1]$ is achieved in the open interval $(0, 1)$ and it attained neither in the limit $\theta \downarrow 0$ nor in the limit $\theta \uparrow 1$.

In the former case, we can guarantee (361) and that at the same time $Q_\upsilon(\mathcal{K})$ will be arbitrarily close to zero, by choosing Q_υ to be of the form $Q_{\upsilon,\theta}$, for θ very close to 0. This follows from (360), which implies $Q_{\upsilon,\theta}(\mathcal{K}) \leq \theta$.

In the latter two cases, we can find some θ^* such that for $Q_\upsilon = Q_{\upsilon,\theta^*}$ (361) holds and such that at the same time the left derivative of the mapping is negative so that

$$\frac{\partial}{\partial\lambda}\tilde{I}_\upsilon\left((1-\lambda)Q_\upsilon + \lambda Q_\upsilon^{(0)}\right)\Big|_{\lambda=0} \leq 0. \quad (364)$$

(The discussion of the three cases is superfluous if the supremum is achieved by some θ^* because then the choice of $Q_\upsilon = Q_{\upsilon,\theta^*}$ would result in $\tilde{I}_\upsilon(Q_\upsilon)$ being greater or equal to $\tilde{I}_\upsilon(Q_\upsilon^{(1)})$ and of (364) being satisfied unless θ^* were zero, with the latter case resulting in $Q_\upsilon(\mathcal{K}) = 0$.)

Note that because Q_υ is in the convex hull of $\{Q_\upsilon^{(1)}, Q_\upsilon^{(0)}\}$ and because $Q_\upsilon^{(0)}$ is the conditional of $Q_\upsilon^{(1)}$ given $x \notin \mathcal{K}$ (359), it follows that $Q_\upsilon^{(0)}$ is also the conditional version of Q_υ given $x \notin \mathcal{K}$

$$Q_\upsilon^{(0)}(\mathcal{B}) = \frac{Q_\upsilon(\mathcal{B}\cap\mathcal{K}^c)}{Q_\upsilon(\mathcal{K}^c)}, \qquad \mathcal{B} \subset \mathcal{X} \text{ Borel.} \quad (365)$$

We now proceed to demonstrate that (364) implies that $Q_v(\mathcal{K})$ must be small. From (364) we have

$$\begin{aligned}0 &\geq \frac{\partial}{\partial\lambda}\tilde{I}_v\left((1-\lambda)Q_v+\lambda Q_v^{(0)}\right)\Big|_{\lambda=0}\\ &= \int (D(W(\cdot|x)\|(Q_vW)(\cdot)) - vg(x))\,\mathrm{d}Q_v^{(0)} - \tilde{I}_v(Q_v)\\ &= \frac{1}{Q_v(\mathcal{K}^c)}\int_{\mathcal{K}^c}(D(W(\cdot|x)\|(Q_vW)(\cdot)) - vg(x))\,\mathrm{d}Q_v\\ &\quad - \tilde{I}_v(Q_v)\\ &= \frac{1}{Q_v(\mathcal{K}^c)}\left(\tilde{I}_v(Q_v)\right.\\ &\quad \left. - \int_{\mathcal{K}}(D(W(\cdot|x)\|(Q_vW)(\cdot)) - vg(x))\,\mathrm{d}Q_v\right) - \tilde{I}_v(Q_v)\\ &= \left(\frac{1}{Q_v(\mathcal{K}^c)}-1\right)\tilde{I}_v(Q_v)\\ &\quad - \frac{1}{Q_v(\mathcal{K}^c)}\int_{\mathcal{K}}(D(W(\cdot|x)\|(Q_vW)(\cdot)) - vg(x))\,\mathrm{d}Q_v\\ &\geq \left(\frac{1}{Q_v(\mathcal{K}^c)}-1\right)\tilde{I}_v(Q_v)\\ &\quad - \frac{1}{Q_v(\mathcal{K}^c)}\int_{\mathcal{K}}D(W(\cdot|x)\|(Q_vW)(\cdot))\,\mathrm{d}Q_v. \end{aligned} \tag{366}$$

Here, the first equality follows from the expression for the directional derivative of mutual information (76); the subsequent equality by (360) and because $Q_v^{(0)}$ is the conditional probability Q_v conditional on $x \notin \mathcal{K}$ (365); the subsequent equality by expressing mutual information as an integral (over $\mathcal{X}$) of relative entropies as in (349), and by writing (349) as the sum of two integrals, over $\mathcal{K}$ and $\mathcal{K}^c$ ($\mathcal{X} = \mathcal{K}\cup\mathcal{K}^c$); the subsequent equality by a simple algebraic manipulation; and the following inequality by the nonnegativity of the cost function $g(\cdot)$.

To continue with the above chain of inequalities we next define the probability measure $\overline{Q}_v$ as the conditional law of Q_v conditioned on $x \in \mathcal{K}$

$$\overline{Q}_v(\mathcal{B}) = \frac{Q_v(\mathcal{B}\cap\mathcal{K})}{Q_v(\mathcal{K})}, \qquad \mathcal{B}\subset\mathcal{X} \text{ Borel} \tag{367}$$

so that in particular

$$\overline{Q}_v(\mathcal{K}) = 1. \tag{368}$$

(If $Q_v(\mathcal{K}) = 0$ then our discussion is over—Q_v has already escaped.) Alternatively

$$Q_v(\mathcal{B}) = Q_v(\mathcal{K})\overline{Q}_v(\mathcal{B}) + Q_v(\mathcal{K}^c)Q_v^{(0)}(\mathcal{B}), \qquad \mathcal{B}\subset\mathcal{X} \text{ Borel} \tag{369}$$

and, hence,

$$\begin{aligned}(Q_vW)(\mathcal{B}) &= Q_v(\mathcal{K})\left(\overline{Q}_vW\right)(\mathcal{B})\\ &\quad + Q_v(\mathcal{K}^c)\left(\overline{Q}_v^{(0)}W\right)(\mathcal{B}), \qquad \mathcal{B}\subset\mathcal{Y} \text{ Borel}.\end{aligned} \tag{370}$$

We shall now continue to lower-bound the RHS of (366) by first noting that by (370)

$$\begin{aligned}&D(W(\cdot|x)\|(Q_vW)(\cdot))\\ &\qquad \leq \log\frac{1}{Q_v(\mathcal{K})} + D\left(W(\cdot|x)\middle\|\left(\overline{Q}_vW\right)(\cdot)\right).\end{aligned} \tag{371}$$

We next note that by (367)

$$\begin{aligned}&\int_{\mathcal{K}} D\left(W(\cdot|x)\middle\|\left(\overline{Q}_vW\right)(\cdot)\right)\mathrm{d}Q_v\\ &\qquad = Q_v(\mathcal{K})\int_{\mathcal{K}} D\left(W(\cdot|x)\middle\|\left(\overline{Q}_vW\right)(\cdot)\right)\mathrm{d}\overline{Q}_v\end{aligned} \tag{372}$$

so that by (371) and (372)

$$\begin{aligned}&\int_{\mathcal{K}} D(W(\cdot|x)\|(Q_vW)(\cdot))\,\mathrm{d}Q_v\\ &\quad \leq Q_v(\mathcal{K})\log\frac{1}{Q_v(\mathcal{K})}\\ &\qquad + Q_v(\mathcal{K})\int_{\mathcal{K}} D\left(W(\cdot|x)\middle\|\left(\overline{Q}_vW\right)(\cdot)\right)\mathrm{d}\overline{Q}_v\\ &\quad = Q_v(\mathcal{K})\log\frac{1}{Q_v(\mathcal{K})} + Q_v(\mathcal{K})\cdot I\left(\overline{Q}_v; W\right)\\ &\quad \leq \frac{1}{e} + Q_v(\mathcal{K})\cdot I_{\max}(\mathcal{K}).\end{aligned} \tag{373}$$

It now follows from (366) and (373) that

$$\begin{aligned}0 &\geq \left(\frac{1}{Q_v(\mathcal{K}^c)}-1\right)\tilde{I}_v(Q_v)\\ &\quad - \frac{1}{Q_v(\mathcal{K}^c)}\left(\frac{1}{e} + Q_v(\mathcal{K})\cdot I_{\max}(\mathcal{K})\right)\\ &\geq \left(\frac{1}{Q_v(\mathcal{K}^c)}-1\right)\tilde{I}_v(Q_v) - \frac{1}{Q_v(\mathcal{K}^c)}\left(\frac{1}{e} + I_{\max}(\mathcal{K})\right)\\ &= \frac{1}{Q_v(\mathcal{K}^c)}\left(\tilde{I}_v(Q_v) - I_{\max}(\mathcal{K}) - \frac{1}{e}\right) - \tilde{I}_v(Q_v)\end{aligned} \tag{374}$$

so that

$$Q_v(\mathcal{K}) \leq \frac{I_{\max}(\mathcal{K}) + \frac{1}{e}}{\tilde{I}_v(Q_v)}. \tag{375}$$

By (361), (355), and (358) we thus obtain

$$\lim_{v\downarrow 0} Q_v(\mathcal{K}) = 0 \tag{376}$$

thus demonstrating that $\{Q_v\}$ escapes to infinity. □

APPENDIX VII
A PROOF OF THEOREM 4.14

Proof: In view of Theorem 4.8, any upper bound on the capacity of the channel

$$\boldsymbol{x} \mapsto \mathbb{H}\boldsymbol{x} \tag{377}$$

subject to the constraints

$$\mathsf{E}\left[\|\boldsymbol{X}\|^2\right] \leq \mathcal{E}_{\mathrm{s}}, \qquad \Pr\left(\|\boldsymbol{X}\|^2 \geq \mathcal{E}_0\right) = 1 \tag{378}$$

will give rise to an upper bound on the fading number of the channel $\boldsymbol{x} \mapsto \mathbb{H}\boldsymbol{x} + \boldsymbol{Z}$. The upper bound of our choice is the one that is based on (25). Our proof will thus be based on an asymptotic analysis of the bound (25) applied to the channel $\boldsymbol{x} \mapsto \mathbb{H}\boldsymbol{x}$.

Let $\{\mathcal{E}_{\mathrm{s}}^{(n)}\}_{n=1}^{\infty}$ be monotonically increasing to infinity, and let $\{Q^{(n)}\}$ be a corresponding sequence of input distributions satisfying

$$\mathsf{E}_{Q^{(n)}}\left[\|\boldsymbol{X}\|^2\right] \leq \mathcal{E}_{\mathrm{s}}^{(n)}, \qquad Q^{(n)}\left(\|\boldsymbol{X}\|^2 \geq \mathcal{E}_0\right) = 1 \tag{379}$$

and

$$\lim_{n\to\infty}\left\{I\left(Q^{(n)};\,\tilde{W}\right)-\log\log\frac{\mathcal{E}_{\rm s}^{(n)}}{\mathcal{E}_0}\right\}=\chi$$

where $\tilde{W}$ denotes the channel $\boldsymbol{x}\mapsto\mathbb{H}\boldsymbol{x}$.

Fix some $\beta,\,\delta>0$, and $0<\alpha<1$ and define

$$\epsilon_{\delta,\,\mathsf{A}}=\sup_{\|\boldsymbol{x}\|^2\geq\mathcal{E}_0}\left\{\mathsf{E}[\log(\|\mathsf{A}\mathbb{H}\boldsymbol{x}\|^2+\delta)]-\mathsf{E}[\log\|\mathsf{A}\mathbb{H}\boldsymbol{x}\|^2]\right\} \tag{380}$$

so that the restriction $0<\alpha<1$ implies

$$(1-\alpha)\mathsf{E}[\log(\|\mathsf{A}\mathbb{H}\boldsymbol{x}\|^2+\delta)] \leq(1-\alpha)\mathsf{E}[\log\|\mathsf{A}\mathbb{H}\boldsymbol{x}\|^2]+\epsilon_{\delta,\,\mathsf{A}},\qquad \|\boldsymbol{x}\|^2\geq\mathcal{E}_0.$$

Consequently, by (24)

$$\begin{aligned}D\left(\tilde{W}(\cdot|\boldsymbol{x})\middle\|R(\cdot)\right)\leq&\,\epsilon_{\delta,\,\mathsf{A}}+\delta/\beta+\log\pi^{n_{\rm R}}-\log\Gamma(n_{\rm R})\\&-\log|\det\mathsf{A}|^2+\log\beta^\alpha+\log\Gamma(\alpha,\,\delta/\beta)\\&+(n_{\rm R}-\alpha)\mathsf{E}[\log\|\mathsf{A}\mathbb{H}\boldsymbol{x}\|^2]+\frac{1}{\beta}\mathsf{E}[\|\mathsf{A}\mathbb{H}\boldsymbol{x}\|^2]\\&-h(\mathbb{H}\boldsymbol{x}),\qquad \|\boldsymbol{x}\|^2\geq\mathcal{E}_0.\end{aligned}$$

It thus follows from Theorem 5.1 that for any input distribution Q satisfying

$$\|X\|^2\geq\mathcal{E}_0,\qquad Q\text{-a.s.} \tag{381}$$

the mutual information $I(Q;\,\tilde{W})$ can be bounded as

$$\begin{aligned}I\left(Q;\,\tilde{W}\right)\leq&\,\epsilon_{\delta,\,\mathsf{A}}+\delta/\beta\\&+\log\pi^{n_{\rm R}}-\log\Gamma(n_{\rm R})-\log|\det\mathsf{A}|^2\\&+\sup_{\|\boldsymbol{x}\|^2\geq\mathcal{E}_0}\left\{n_{\rm R}\mathsf{E}[\log\|\mathsf{A}\mathbb{H}\boldsymbol{x}\|^2]-h(\mathbb{H}\boldsymbol{x})\right\}\\&+\log\Gamma(\alpha,\delta/\beta)+\frac{1}{\beta}\mathsf{E}_Q\left[\|\mathsf{A}\mathbb{H}\boldsymbol{X}\|^2\right]\\&+\alpha\left(\log\beta-\mathsf{E}_Q\left[\log\|\mathsf{A}\mathbb{H}\boldsymbol{X}\|^2\right]\right).\end{aligned} \tag{382}$$

The rest of the proof is dedicated to the study of the asymptotic behavior of the various terms in the above inequality. We begin by arguing using Lemma 6.8 that for any nonsingular matrix A

$$\lim_{\delta\downarrow0}\epsilon_{\delta,\,\mathsf{A}}=0. \tag{383}$$

To that end, we first note that by the behavior of logarithms under scaling it suffices to show

$$\lim_{\delta\downarrow0}\sup_{\|\hat{\boldsymbol{x}}\|^2=1}\{\mathsf{E}[\log(\|\mathsf{A}\mathbb{H}\hat{\boldsymbol{x}}\|^2+\delta)]-\mathsf{E}[\log\|\mathsf{A}\mathbb{H}\hat{\boldsymbol{x}}\|^2]\}=0. \tag{384}$$

Let $\epsilon>0$ be arbitrary. By Lemma 6.8 (applied to the matrix $\mathsf{A}\mathbb{H}$), there exists some $0<\delta'<1/2$ such that

$$\mathsf{E}[\log\|\mathsf{A}\mathbb{H}\hat{\boldsymbol{x}}\|] \geq\int_{\|\mathsf{AH}\hat{\boldsymbol{x}}\|^2>\delta'}\log\|\mathsf{AH}\hat{\boldsymbol{x}}\|\,\mathrm{d}P_{\mathbb{H}}(\mathsf{H})-\epsilon,\quad \|\hat{\boldsymbol{x}}\|=1. \tag{385}$$

Also,

$$\mathsf{E}[\log(\|\mathsf{A}\mathbb{H}\hat{\boldsymbol{x}}\|^2+\delta)]\leq\int_{\|\mathsf{AH}\hat{\boldsymbol{x}}\|^2>\delta'}\log(\|\mathsf{AH}\hat{\boldsymbol{x}}\|^2+\delta)\,\mathrm{d}P_{\mathbb{H}}(\mathsf{H}),\qquad \delta+\delta'<1,\,\hat{\boldsymbol{x}}\in\mathbb{C}^{n_{\rm T}}. \tag{386}$$

Consequently, for all $\delta<1/2$

$$\begin{aligned}\sup_{\|\hat{\boldsymbol{x}}\|=1}\mathsf{E}\left[\log\frac{\|\mathsf{A}\mathbb{H}\hat{\boldsymbol{x}}\|^2+\delta}{\|\mathsf{A}\mathbb{H}\hat{\boldsymbol{x}}\|^2}\right]&\leq\sup_{\|\hat{\boldsymbol{x}}\|=1}\int_{\|\mathsf{AH}\hat{\boldsymbol{x}}\|^2>\delta'}\log\frac{\|\mathsf{AH}\hat{\boldsymbol{x}}\|^2+\delta}{\|\mathsf{AH}\hat{\boldsymbol{x}}\|^2}\,\mathrm{d}P_{\mathbb{H}}(\mathsf{H})+\epsilon\\&\leq\sup_{\|\hat{\boldsymbol{x}}\|=1}\int_{\|\mathsf{AH}\hat{\boldsymbol{x}}\|^2>\delta'}\log\frac{\delta'+\delta}{\delta'}\,\mathrm{d}P_{\mathbb{H}}(\mathsf{H})+\epsilon\\&\leq\log\left(1+\frac{\delta}{\delta'}\right)+\epsilon\\&\xrightarrow{\delta\downarrow0}\epsilon\end{aligned}$$

from which (384) follows because $\epsilon>0$ was arbitrary.

The next term in (382) to be studied is the term $\mathsf{E}_Q[\log\|\mathsf{A}\mathbb{H}\boldsymbol{X}\|^2]$. For distributions Q satisfying (381) this term can be lower-bounded as

$$\begin{aligned}\mathsf{E}_Q\left[\log\|\mathsf{A}\mathbb{H}\boldsymbol{X}\|^2\right]&\geq\inf_{\|\boldsymbol{x}\|^2\geq\mathcal{E}_0}\mathsf{E}\left[\log\|\mathsf{A}\mathbb{H}\boldsymbol{x}\|^2\right]\\&=\log\mathcal{E}_0+\inf_{\|\hat{\boldsymbol{x}}\|=1}\mathsf{E}\left[\log\|\mathsf{A}\mathbb{H}\hat{\boldsymbol{x}}\|^2\right]\\&=\log\mathcal{E}_0+\eta(\mathsf{A})\end{aligned} \tag{387}$$

where the last equality should be taken as a definition for $\eta(\mathsf{A})$.

Notice that

$$-\infty<\eta(\mathsf{A})<\infty \tag{388}$$

as can be argued as follows. The lower bound on $\eta(\mathsf{A})$ in (388) can be argued using Lemma 6.7, Part f) applied to the matrix $\mathsf{A}\mathbb{H}$. The upper bound on $\eta(\mathsf{A})$ in (388) can be verified using the concavity of the logarithm function and Jensen's inequality.

Having established (388), we continue with the proof of the theorem. Combining (382) and (387) we obtain

$$\begin{aligned}I\left(Q;\,\tilde{W}\right)\leq&\,\epsilon_{\delta,\,\mathsf{A}}+\delta/\beta+\log\pi^{n_{\rm R}}-\log\Gamma(n_{\rm R})-\log|\det\mathsf{A}|^2\\&+\sup_{\|\hat{\boldsymbol{x}}\|=1}\{n_{\rm R}\mathsf{E}[\log\|\mathsf{A}\mathbb{H}\boldsymbol{x}\|^2]-h(\mathbb{H}\boldsymbol{x})\}\\&+\log\Gamma(\alpha,\,\delta/\beta)+\frac{1}{\beta}\mathsf{E}_Q[\|\mathsf{A}\mathbb{H}\boldsymbol{X}\|^2]\\&+\alpha\left(\log\beta-\log\mathcal{E}_0-\eta(\mathsf{A})\right).\end{aligned} \tag{389}$$

It now follows from (389) and (383) that in order to conclude the proof of the theorem it will suffice to establish that there exist parameters $\alpha^{(n)}$ and $\beta^{(n)}$ such that

$$\begin{aligned}\lim_{n\uparrow\infty}\Bigg\{&\frac{\delta}{\beta^{(n)}}+\left(\log\Gamma(\alpha^{(n)},\delta/\beta^{(n)})-\log\frac{1}{\alpha^{(n)}}\right)\\&+\alpha^{(n)}\left(\log\beta^{(n)}-\log\mathcal{E}_0-\eta(\mathsf{A})\right)\\&+\frac{1}{\beta^{(n)}}\mathsf{E}_{Q^{(n)}}[\|\mathsf{A}\mathbb{H}\boldsymbol{X}\|^2]\\&+\log\frac{1}{\alpha^{(n)}}-\log\log\frac{\mathcal{E}_{\rm s}^{(n)}}{\mathcal{E}_0}\Bigg\}<o(\delta).\end{aligned} \tag{390}$$

This inequality is trivial if $\mathsf{E}_{Q^{(n)}}[\|\mathsf{A}\mathbb{H}\boldsymbol{X}\|^2]$ is bounded, because in that case the LHS tends to $-\infty$ for any fixed values of $\alpha^{(n)}$ and $\beta^{(n)}$. Therefore, by possibly passing to a subsequence, it suffices to treat the case where

$$\lim_{n\to\infty} \mathsf{E}_{Q^{(n)}}[\|\mathsf{A}\mathbb{H}\boldsymbol{X}\|^2] = \infty. \tag{391}$$

A choice of the parameters $\alpha^{(n)}$ and $\beta^{(n)}$ that will demonstrate (390) is

$$\alpha^{(n)} = \frac{\delta}{\log \mathsf{E}_{Q^{(n)}}[\|\mathsf{A}\mathbb{H}\boldsymbol{X}\|^2]} \tag{392}$$

$$\beta^{(n)} = \frac{1}{\alpha^{(n)}} \cdot e^{\delta/\alpha^{(n)}}. \tag{393}$$

With this choice of parameters, we have for the various terms in (390)

$$\lim_{n\to\infty} \frac{\delta}{\beta^{(n)}} = 0 \tag{394}$$

$$\lim_{n\to\infty} \left\{ \log \Gamma(\alpha^{(n)}, \delta/\beta^{(n)}) - \log \frac{1}{\alpha^{(n)}} \right\} = \log(1 - e^{-\delta}) \tag{395}$$

(see Appendix XI)

$$\lim_{n\to\infty} \alpha^{(n)} \left(\log \beta^{(n)} - \log \mathcal{E}_0 - \eta(\mathsf{A}) \right) = \delta \tag{396}$$

$$\frac{1}{\beta^{(n)}} \mathsf{E}_{Q^{(n)}}[\|\mathsf{A}\mathbb{H}\boldsymbol{X}\|^2] = \alpha^{(n)} \to 0 \tag{397}$$

$$\begin{aligned} &\log \frac{1}{\alpha^{(n)}} - \log\log \frac{\mathcal{E}_{\mathrm{s}}^{(n)}}{\mathcal{E}_0} \\ &\quad = \log\log \mathsf{E}_{Q^{(n)}}[\|\mathsf{A}\mathbb{H}\boldsymbol{X}\|^2] - \log\delta - \log\log \frac{\mathcal{E}_{\mathrm{s}}^{(n)}}{\mathcal{E}_0} \\ &\quad \le \log\log \left(\|\mathsf{A}\|^2 \mathsf{E}[\|\mathbb{H}\|^2] \mathcal{E}_{\mathrm{s}}^{(n)} \right) - \log\delta - \log\log \frac{\mathcal{E}_{\mathrm{s}}^{(n)}}{\mathcal{E}_0} \\ &\quad \to -\log\delta. \end{aligned} \tag{398}$$

Expressions (394)–(398) prove (390) and thus conclude the proof of the theorem. □

Appendix VIII
A Proof of Proposition 4.23

Proof: Let $Y = HX + Z$. To derive an upper bound on the fading number in the presence of receiver side information we invoke Lemma 4.5 (in the memoryless case) to obtain

$$I(X;\, Y,\, S) \le I(X;\, Y) + I(H;\, S) \tag{399}$$

and then upper bound $I(X;\, Y)$ by invoking Theorem 4.16.

We now proceed with a lower bound. To that end, we choose X to be circularly symmetric with $\log|X|^2$ being uniformly distributed over the interval $(\log x_{\min}^2,\, \log \mathcal{E}_{\mathrm{s}})$. Using Lemma 4.9 we have

$$\begin{aligned} &I(X;\, HX+Z|S=s) \\ &\quad \ge I(X;\, HX|S=s) \\ &\qquad - \left(h\left(H + \frac{Z}{x_{\min}} \middle| S=s \right) - h(H|S=s) \right) \\ &\quad \ge \log\log \frac{\mathcal{E}_{\mathrm{s}}}{x_{\min}^2} + \log\pi + \mathsf{E}[\log|H|^2|S=s] - h(H|S=s) \\ &\qquad - \left(h\left(H + \frac{Z}{x_{\min}} \middle| S=s \right) - h(H|S=s) \right) \\ &\quad \triangleq C_l(\mathcal{E}_{\mathrm{s}},\, x_{\min},\, s) \end{aligned}$$

where the second inequality follows from (83). We next note that for a fixed s, the term

$$h\left(H + \frac{Z}{x_{\min}} \middle| S=s \right) - h(H|S=s) \tag{400}$$

is monotonically decreasing in $x_{\min}$. Choose now

$$x_{\min}^2 = \log \mathcal{E}_{\mathrm{s}}$$

so that $x_{\min}^2$ is monotonically increasing in $\mathcal{E}_{\mathrm{s}}$; it tends to infinity as $\mathcal{E}_{\mathrm{s}} \uparrow \infty$; and $\log\log\mathcal{E}_{\mathrm{s}} - \log\log(\mathcal{E}_{\mathrm{s}}/x_{\min}^2)$ approaches zero as $\mathcal{E}_{\mathrm{s}} \uparrow \infty$. With this choice, the function

$$\begin{aligned} &C_l(\mathcal{E}_{\mathrm{s}},\, x_{\min},\, s) - \log\log \frac{\mathcal{E}_{\mathrm{s}}}{x_{\min}^2} - \log\pi \\ &\qquad - \mathsf{E}[\log|H|^2|S=s] + h(H|S=s) \\ &\quad = -\left(h\left(H + \frac{Z}{x_{\min}} \middle| S=s \right) - h(H|S=s) \right) \end{aligned} \tag{401}$$

is for every s monotonically increasing (in $\mathcal{E}_{\mathrm{s}}$) to zero. In addition, the RHS of (401) is integrable as a function of s. The desired lower bound now follows by the Monotone Convergence Theorem. □

Appendix IX
A Proof of Theorem 4.41

The proof of this theorem was outlined in the text. There remains to show that ϵ, defined in (136), tends to zero as the SNR tends to infinity. This is what we prove in this appendix.

Rewriting ϵ we have

$$\begin{aligned} \epsilon &= I\left(X_k; H_{k-\kappa}^{k-1} \middle| Y_k, Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa} \right) \\ &\le h\left(H_{k-\kappa}^{k-1} \middle| Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa} \right) \\ &\quad - h\left(H_{k-\kappa}^{k-1} \middle| Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa}, X_k, Y_k \right) \\ &\le h\left(H_{k-\kappa}^{k-1} \middle| Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa} \right) \\ &\quad - h\left(H_{k-\kappa}^{k-1} \middle| Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa}, H_k \right) \\ &= I\left(H_k; H_{k-\kappa}^{k-1} \middle| Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa} \right) \\ &\le h\left(H_k \middle| \frac{Y_{k-1}}{X_{k-1}}, \ldots, \frac{Y_{k-\kappa}}{X_{k-\kappa}}, \boldsymbol{S}_{k-\kappa}^{k+\kappa} \right) \\ &\quad - h\left(H_k \middle| H_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa} \right) \\ &\le \max_{\substack{\|x_{k-1}\| \ge x_{\min}, \ldots, \\ \|x_{k-\kappa}\| \ge x_{\min}}} h\left(H_k \middle| \frac{Y_{k-1}}{x_{k-1}}, \ldots, \frac{Y_{k-\kappa}}{x_{k-\kappa}}, \boldsymbol{S}_{k-\kappa}^{k+\kappa} \right) \\ &\quad - h\left(H_k \middle| H_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa} \right) \\ &= h\left(H_k \middle| H_{k-1} + \frac{\sigma}{x_{\min}} W_1, \ldots, H_{k-\kappa} + \frac{\sigma}{x_{\min}} W_\kappa, \boldsymbol{S}_{k-\kappa}^{k+\kappa} \right) \\ &\quad - h\left(H_k \middle| H_{k-\kappa}^{k-1}, \boldsymbol{S}_{k-\kappa}^{k+\kappa} \right) \\ &\le h\left(H_k, H_{k-1} + \frac{\sigma}{x_{\min}} W_1, \ldots, H_{k-\kappa} + \frac{\sigma}{x_{\min}} W_\kappa \middle| \boldsymbol{S}_{k-\kappa}^{k+\kappa} \right) \\ &\quad - h\left(H_k, H_{k-\kappa}^{k-1} \middle| \boldsymbol{S}_{k-\kappa}^{k+\kappa} \right) \end{aligned}$$

where $W_1,\, \ldots,\, W_\kappa$ are i.i.d. $\mathcal{N}_{\mathbb{C}}(0,\, 1)$.

We thus conclude from Lemma 6.11 that for any realization of $\boldsymbol{S}_{k-\kappa}^{k+\kappa}$ the conditional law of ϵ converges monotonically in

$x_{\min}$ to zero. Choosing $\log x_{\min}^2 = \log\log \mathcal{E}_s$ gives the desired convergence in the SNR. The result now follows by averaging over $\boldsymbol{S}_{k-\kappa}^{k+\kappa}$ using the Monotone Convergence Theorem.

Appendix X
Expected-Log of a Noncentral χ^2

Lemma 10.1: Let the random variable W have a noncentral χ^2 distribution with an even number of degrees of freedom, i.e.,

$$W = \sum_{j=1}^{m} \left| X_j + \mu_j \right|^2 \tag{402}$$

where $\{X_j\}_{j=1}^m$ are i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$, and $\{\mu_j\}_{j=1}^m$ are deterministic complex constants. Then

$$\mathsf{E}[\log W] = g_m(s^2) \tag{403}$$

$$= \log(s^2) - \mathrm{Ei}(-s^2) + \sum_{j=1}^{m-1} (-1)^j \cdot \left[e^{-s^2}(j-1)! - \frac{(m-1)!}{j(m-1-j)!} \right] \left(\frac{1}{s^2} \right)^j \tag{404}$$

where

$$s^2 = \sum_{j=1}^{m} |\mu_j|^2. \tag{405}$$

Moreover, the functions $g_m(\cdot)$ are monotonically increasing and concave in the interval $[0, \infty)$.

Proof: The density $f_W(w)$ of W is given by [24, Ch. 29]

$$f_W(w) = \left(\frac{w}{s^2} \right)^{\frac{m-1}{2}} e^{-w-s^2} I_{m-1}\left(2s\sqrt{w}\right), \qquad w \geq 0 \tag{406}$$

where $I_{m-1}(\cdot)$ denotes the modified Bessel function of the first kind of order $m-1$. Thus, the required expectation can be written as

$$\mathsf{E}[\log W] = \int_0^\infty \log w \cdot p_W(w)\,\mathrm{d}w = \int_0^\infty \log w \cdot \left(\frac{w}{s^2} \right)^{\frac{m-1}{2}} e^{-w-s^2} I_{m-1}\left(2s\sqrt{w}\right) \mathrm{d}w. \tag{407}$$

Expressing $I_{m-1}(\cdot)$ as a power series

$$I_{m-1}(z) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(m+k)} \left(\frac{z}{2} \right)^{m-1+2k} \tag{408}$$

we obtain from [15, 4.352 (1)]

$$\mathsf{E}[\log W] = \frac{1}{s^{m-1}} e^{-s^2} \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(m+k)} s^{2k+m-1} \cdot \int_0^\infty w^{k+m-1} e^{-w} \log w \,\mathrm{d}w \tag{409}$$

$$= \frac{1}{s^{m-1}} e^{-s^2} \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(m+k)} s^{2k+m-1} \cdot \Gamma(m+k)\psi(k+m) \tag{410}$$

$$= e^{-s^2} \sum_{k=0}^{\infty} \frac{1}{k!} (s^2)^k \left[-\gamma + \sum_{i=1}^{k+m-1} \frac{1}{i} \right] \tag{411}$$

$$= -\gamma + e^{-s^2} \sum_{k=0}^{\infty} \frac{1}{k!} (s^2)^k \sum_{j=1}^{k+m-1} \frac{1}{j} \tag{412}$$

$$= \tilde{g}_m(s^2) \tag{413}$$

where $\psi(\cdot)$ denotes Euler's psi-function (213) and where we define the function $\tilde{g}_m(z)$ as

$$\tilde{g}_m(z) = -\gamma + e^{-z} \sum_{k=0}^{\infty} \frac{1}{k!} z^k \sum_{j=1}^{k+m-1} \frac{1}{j} \tag{414}$$

with derivatives

$$\tilde{g}'_m(z) = e^{-z} \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \frac{1}{k+m} \cdot z^k \tag{415}$$

$$\tilde{g}''_m(z) = -e^{-z} \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \frac{1}{(k+m)(k+m+1)} \cdot z^k. \tag{416}$$

To conclude the proof of (403), it is now required to show that $\tilde{g}_m(\cdot)$ is identical to $g_m(\cdot)$. We shall begin by studying the derivative $\tilde{g}'_m$ and show that it can be expressed as

$$\tilde{g}'_m(z) = \frac{(-1)^m \Gamma(m)}{z^m} \left(e^{-z} - \sum_{j=0}^{m-1} \frac{(-1)^j}{j!} z^j \right). \tag{417}$$

Indeed

$$\begin{aligned}
\tilde{g}'_m(z) &= \frac{(-1)^m \Gamma(m)}{z^m} e^{-z} \left(1 - e^z \sum_{j=0}^{m-1} \frac{(-1)^j}{j!} z^j \right) \\
&= \frac{(-1)^m \Gamma(m)}{z^m} e^{-z} \left(1 - \sum_{k=0}^{\infty} \sum_{j=0}^{m-1} \frac{(-1)^j}{j!k!} z^{k+j} \right) \\
&= \frac{(-1)^m \Gamma(m)}{z^m} e^{-z} \Bigg(1 - 1 - z \cdot 0 - z^2 \cdot 0 - \cdots \\
&\qquad - z^{m-1} \cdot 0 - \sum_{i=m}^{\infty} \sum_{j=0}^{m-1} \frac{(-1)^j}{j!(i-j)!} z^i \Bigg) \\
&= -\frac{(-1)^m \Gamma(m)}{z^m} e^{-z} \sum_{i=m}^{\infty} z^i \sum_{j=0}^{m-1} \frac{(-1)^j}{j!(i-j)!} \\
&= -\frac{(-1)^m \Gamma(m)}{z^m} e^{-z} \sum_{i=m}^{\infty} z^i \frac{(-1)^{m-1}(i-1)!}{i!(m-1)!(i-m)!} \\
&= e^{-z} \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \frac{1}{k+m} \cdot z^k.
\end{aligned}$$

Integrating this series term-by-term we obtain

$$\tilde{g}_m(z) = c_m + \log z - \mathrm{Ei}(-z) + \sum_{j=1}^{m-1} (-1)^j \left(e^{-z}(j-1)! - \frac{(m-1)!}{j(m-1-j)!} \right) z^{-j} \tag{418}$$

for some constant c_m. By evaluating the RHS of (418) at $z=0$ and comparing with the evaluation at $z=0$ of the RHS of (414) we conclude that $c_m = 0$. With this value of c_m it is readily seen that the RHS of (418) is identical to $g_m(z)$, and we thus conclude that $\tilde{g}_m(z) = g_m(z)$. Since $\tilde{g}_m$ is identical to g_m, the monotonicity and concavity of $g_m(z)$ follow from those of $\tilde{g}_m$, which can be verified from (415) and (416). □

APPENDIX XI
A LIMIT OF THE INCOMPLETE GAMMA FUNCTION

Lemma 11.1: Let the sequences $\{\xi^{(n)}\}$ and $\{\alpha^{(n)}\}$ take value in the open interval $(0, 1)$. Assume

$$\lim_{n\to\infty} \xi^{(n)} = 0 \tag{419}$$

$$\lim_{n\to\infty} \alpha^{(n)} = 0 \tag{420}$$

$$\lim_{n\to\infty} (\xi^{(n)})^{\alpha^{(n)}} = \zeta \tag{421}$$

for some $0 \le \zeta < 1$. Then

$$\lim_{n\to\infty} \left\{ \log \Gamma\Big(\alpha^{(n)}, \xi^{(n)}\Big) - \log \frac{1}{\alpha^{(n)}} \right\} = \log(1-\zeta). \tag{422}$$

Proof: Integrating (200) by parts we obtain

$$\Gamma(\alpha, \xi) = \int_{\xi}^{\infty} t^{\alpha-1} e^{-t} \, \mathrm{d}t = -e^{-\xi} \cdot \frac{\xi^{\alpha}}{\alpha} + \frac{1}{\alpha} \Gamma(\alpha+1, \xi)$$

thus establishing

$$\log \Gamma(\alpha, \xi) = \log \frac{1}{\alpha} + \log \left(\Gamma(\alpha+1, \xi) - e^{-\xi} \xi^{\alpha} \right).$$

The claim now follows from the continuity of $\Gamma(\alpha+1, \xi)$ around $(\alpha, \xi) = (0, 0)$. □

ACKNOWLEDGMENT

Discussions with İ. Emre Telatar and Ram Zamir are gratefully acknowledged.

REFERENCES

[1] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic, 1981.
[2] F. Topsœ, "An information theoretical identity and a problem involving capacity," *Studia Scientiarum Math. Hung.*, vol. 2, pp. 291–292, 1967.
[3] P. Vontobel and D. Arnold, "An upper bound on the capacity of channels with memory and constraint input," in *Proc. 2001 IEEE Information Theory Workshop*, Cairns, Australia, Sept. 2–7, 2001, pp. 147–149.
[4] A. Lapidoth, "Capacity bounds via duality: A phase noise example," in *Proc. 2nd Asian-European Workshop on Information Theory*, Breisach, Germany, June 26–29, 2002, pp. 58–61. See also A. Lapidoth, "On phase noise channels at high SNR," in *Proc. 2002 IEEE Information Theory Workshop*, Bangalore, India, Oct. 20–25, 2002, pp. 1–4.
[5] A. Lapidoth and S. Moser, "The asymptotic capacity of the discrete-time Poisson channel," in *Proc. 2003 Winter School on Coding and Information Theory*. Ascona, Switzerland, Feb. 24–27, 2003.
[6] A. Lapidoth and N. Miliou, "Duality based bounds on the cut-off rate of a discrete-time memoryless Rayleigh fading channel," in *Proc. 2003 Winter School on Coding and Information Theory*. Ascona, Switzerland, Feb. 24–27, 2003.
[7] S. Karlin and W. Studen, *Tchebycheff Systems: With Applications in Analysis and Statistics*. New York: Wiley, 1966.
[8] M. Shaked and J. Shanthikumar, *Stochastic Orders and Their Applications*. San Diego, CA: Academic, 1994.
[9] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need "perfect side information" be?," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1118–1134, May 2002.
[10] N. Merhav and M. Feder, "A strong version of the redundancy-capacity theorem of universal coding," *IEEE Trans. Inform. Theory*, vol. 41, pp. 714–722, May 1995.
[11] I. Abou-Faycal, M. Trott, and S. Shamai (Shitz), "The capacity of discrete time Rayleigh fading channels," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1290–1301, May 2001.
[12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
[13] A. Sengupta and P. Mitra, "Capacity of multivariate channels with multiplicative noise: I. Random matrix techniques and large- n expansion for full transfer matrices," Bell Labs., Lucent Technologies, Murray Hill, NJ, Tech. Rep., Oct. 2000.
[14] G. Taricco and M. Elia, "Capacity of fading channels with no side information," *Electron. Lett.*, vol. 33, pp. 1368–1370, July 31, 1997.
[15] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. San Diego, CA: Academic, 1996.
[16] T. Ericson, "A Gaussian channel with slow fading," *IEEE Trans. Inform. Theory*, vol. IT-16, pp. 353–355, May 1970.
[17] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, 1999.
[18] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
[19] D. Haussler, "A general minimax result for relative entropy," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1276–1280, July 1997.
[20] I. Vajda, *Theory of Statistical Inference and Information*. Norwell, MA: Kluwer, 1989.
[21] I. Csiszár, "Arbitrarily varying channel with general alphabets and states," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1725–1742, Nov. 1992.
[22] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inform. Theory*, vol. 39, pp. 1293–1302, July 1993.
[23] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd ed. New York: Wiley, 1995, vol. 1.
[24] ——, *Continuous Univariate Distributions*, 2nd ed. New York: Wiley, 1995, vol. 2.
[25] E. L. Lehmann, "Ordered families of distributions," *Ann. Math. Statist.*, vol. 26, pp. 399–419, 1955.
[26] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.
[27] Y. L. Tong, *The Multivariate Normal Distribution*. New York: Springer-Verlag, 1990.
[28] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[29] T. Linder and R. Zamir, "On the asymptotic tightness of the Shannon lower bound," *IEEE Trans. Inform. Theory*, vol. 40, pp. 2026–2031, Nov. 1994.
[30] D. Williams, *Probability with Martingales*. Cambridge, U.K.: Cambridge Mathematical Textbooks, Cambridge Univ. Press, 1991.
[31] M. Costa, "A new entropy power inequality," *IEEE Trans. Inform. Theory*, vol. IT-31, pp. 751–760, Nov. 1985.
[32] A. Dembo, "Simple proof of the concavity of the entropy power with respect to added Gaussian noise," *IEEE Trans. Inform. Theory*, vol. 35, pp. 887–888, July 1989.
[33] C. Villani, "A short proof of the concavity of entropy power," *IEEE Trans. Inform. Theory*, vol. 46, pp. 1695–1696, July 2000.
[34] A. Lapidoth, "On the asymptotic capacity of fading channels," preprint, May 2003, submitted for publication.
[35] Ö Oyman, R. Nabar, H. Bölcskei, and A. Paulraj, "Tight lower bounds on the ergodic capacity of Rayleigh fading MIMO channels," in *Proc. IEEE GLOBECOM 2002*, Taipei, Taiwan, R.O.C., Nov. 2002.
[36] T. Marzetta and B. Hochwald, "Capacity of a multiple-antenna communication link in Rayleigh flat fading environment," *IEEE Trans. Inform. Theory*, vol. 45, pp. 139–157, Jan. 1999.
[37] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inform. Theory*, vol. 48, pp. 359–383, Feb. 2002.
[38] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1319–1343, June 2002.