The Fading Number of Memoryless Multiple-Input Multiple-Output Fading Channels

Stefan M. Moser, *Member, IEEE*

***Abstract*—We derive the fading number of multiple-input multiple-output (MIMO) flat fading channels of general (not necessarily Gaussian) regular law without temporal memory. The channel is assumed to be non-coherent, *i.e.*, neither receiver nor transmitter have knowledge about the channel state, but they only know the probability law of the fading process. The fading number is the second term, after the double-logarithmic term, of the high signal-to-noise ratio (SNR) expansion of channel capacity. Hence the asymptotic channel capacity of memoryless MIMO fading channels is derived exactly.**

The result is then specialized to the known cases of single-input multiple-output (SIMO), multiple-input single-output (MISO), and single-input single-output (SISO) fading channels, as well as to the situation of Gaussian fading.

***Index Terms*—Channel capacity, fading number, Gaussian fading, general flat fading, high signal-to-noise ratio (SNR), multiple-input multiple-output (MIMO), multiple antenna, non-coherent.**

I. Introduction

It has been recently shown in [1], [2] that, whenever the matrix-valued fading process is of finite differential entropy rate (a so-called *regular* process), the capacity of non-coherent multiple-input multiple-output (MIMO) fading channels typically grows only double-logarithmically in the signal-to-noise ratio (SNR).

This is in stark contrast to both, the *coherent* fading channel where the receiver has *perfect* knowledge about the channel state, and to the non-coherent fading channel with *non-regular* channel law, *i.e.*, the differential entropy rate of the fading process is not finite. In the former case the capacity grows logarithmically in the SNR with a factor in front of the logarithm that is related to the number of receive and transmit antennas [3].

In the latter case the asymptotic growth rate of the capacity depends highly on the specific details of the fading process. In the case of Gaussian fading, non-regularity means that the present fading realization can be predicted *precisely* from the past realizations. However, in every non-coherent system the past realizations are not known *a priori*, but need to be estimated either by known past channel inputs and outputs or by means of special training signals. Depending on the spectral distribution of the fading process, the dependence of such estimations on the available power can vary largely which gives rise to a huge variety of possible high-SNR capacity behaviors: it is shown in [4], [5], and [6] that depending on the spectrum of the non-regular Gaussian fading process, the asymptotic behavior of the channel capacity can be varied in a large range: it is possible to have very slow double-logarithmic growth, fast logarithmic growth, or even exotic situations where the capacity grows proportionally to a fractional power of $\log \textnormal{SNR}$.

Similarly, Liang and Veeravalli show in [7] that the capacity of a Gaussian block-fading channel depends critically on the assumptions one makes about the time-correlation of the fading process: if the correlation matrix is rank deficient, the capacity grows logarithmically in the SNR, otherwise double-logarithmically.

In this paper we will only consider non-coherent channels with regular fading processes, *i.e.*, the capacity at high SNR will be growing double-logarithmically. To quantify the rates at which this poor power efficiency begins, [1], [2] introduce the *fading number* as the second term in the high-SNR asymptotic expansion of channel capacity. Hence, the capacity can be written as

$$\mathsf{C}(\textnormal{SNR}) = \log(1 + \log(1 + \textnormal{SNR})) + \chi + o(1) \tag{1}$$

where $o(1)$ tends to zero as the SNR tends to infinity, and where χ is a constant, denoted *fading number*, that does not depend on the SNR.

Explicit expressions of the fading number are known for a number of fading models. For channels with memory, the fading number of single-input single-output (SISO) fading channels is derived in [1], [2] and the single-input multiple-output (SIMO) case is derived in [8], [2].

For memoryless fading channels, the fading number is known in the situation of only one antenna at transmitter and receiver (SISO)

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

in the situation of a SIMO fading channel[1]

$$\chi(\mathbf{H}) = h_\lambda\bigl(\hat{\mathbf{H}} e^{\mathrm{i}\Theta}\bigr) + n_\mathrm{R}\mathsf{E}\left[\log \|\mathbf{H}\|^2\right] - \log 2 - h(\mathbf{H}) \tag{3}$$

(both are special cases from the corresponding situation with memory); and also in the case of a multiple-input single-output (MISO) fading channel [1], [2]

$$\chi\bigl(\mathbf{H}^\mathsf{T}\bigr) = \sup_{\hat{\mathbf{x}}} \left\{ \log \pi + \mathsf{E}\left[\log \left|\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}\right|^2\right] - h\bigl(\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}\bigr) \right\}. \tag{4}$$

The most general situation of multiple antennas at both transmitter and receiver, however, has been solved so far only in the special case of a particular rotational symmetry of the fading process: if every rotation of the input vector of the channel can be "undone" by a corresponding rotation of the output vector, and vice-versa, then the fading number has been shown in [1], [2] to be

$$\chi(\mathbb{H}) = \log \frac{\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})} + n_\mathrm{R}\mathsf{E}\left[\log \|\mathbb{H}\hat{\mathbf{e}}\|^2\right] - h(\mathbb{H}\hat{\mathbf{e}}) \tag{5}$$

where $\hat{\mathbf{e}} \in \mathbb{C}^{n_\mathrm{T}}$ is an arbitrary constant vector of unit length, and where n_R denotes the number of receive antennas. Such fading channels are called *rotation-commutative in the generalized sense* (for a detailed definition see Section V).

In this paper we will extend these results and derive the fading number of general memoryless MIMO fading channels.

The remaining of this paper is structured as follows: before we proceed in Section III to introduce the channel model in detail, the following section will clarify our notation. We will then present the main result, *i.e.*, the fading number of the general memoryless MIMO fading channel in Section IV. The corresponding proof is found in Section VII.

In Section V the known fading numbers of SISO, SIMO, MISO, and rotation-commutative MIMO fading channels are derived once more as special cases of the new general result from Section IV. In Section VI we investigate the situation of Gaussian fading processes. We will conclude in Section VIII.

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 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

The author is with the Department of Communication Engineering, at National Chiao Tung University (NCTU) in Hsinchu, Taiwan. This work was supported by the Industrial Technology Research Institute (ITRI), Zhudong, Taiwan, under contract G1-95003.

[1]For a precise definition of the notation used in this paper we refer to Section II.

written in lower-case, *e.g.*, x. For random vectors we use bold face capitals, *e.g.*, $\mathbf{X}$ and bold lower-case for their realizations, *e.g.*, $\mathbf{x}$. Deterministic matrices are denoted by upper-case letters but of a special font, *e.g.*, H; and random matrices are denoted using another special upper-case font, *e.g.*, $\mathbb{H}$. The capacity is denoted by C, the energy per symbol by $\mathcal{E}$, and the signal-to-noise ratio SNR is denoted by SNR.

We use the shorthand H_a^b for $(H_a, H_{a+1}, \ldots, H_b)$. For more complicated expressions, such as $(\mathbf{H}_a^{\mathsf{T}}\hat{\mathbf{x}}_a, \mathbf{H}_{a+1}^{\mathsf{T}}\hat{\mathbf{x}}_{a+1}, \ldots, \mathbf{H}_b^{\mathsf{T}}\hat{\mathbf{x}}_b)$, we use the dummy variable ℓ to clarify notation: $\{\mathbf{H}_\ell^{\mathsf{T}}\hat{\mathbf{x}}_\ell\}_{\ell=a}^{b}$.

Hermitian conjugation is denoted by $(\cdot)^\dagger$, and $(\cdot)^{\mathsf{T}}$ stands for the transpose (without conjugation) of a matrix or vector. The trace of a matrix is denoted by $\mathrm{tr}\,(\cdot)$.

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

$$\|\mathbf{x}\| \triangleq \sqrt{\sum_{t=1}^{m} |x^{(t)}|^2}, \quad \mathbf{x} \in \mathbb{C}^m \tag{6}$$

$$\|\mathsf{A}\| \triangleq \max_{\|\hat{\mathbf{w}}\|=1} \|\mathsf{A}\hat{\mathbf{w}}\|. \tag{7}$$

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.*,

$$\|\mathsf{A}\|_{\mathrm{F}} \triangleq \sqrt{\mathrm{tr}\,(\mathsf{A}^\dagger\mathsf{A})}. \tag{8}$$

Note that for every matrix A

$$\|\mathsf{A}\| \leq \|\mathsf{A}\|_{\mathrm{F}} \tag{9}$$

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

We will often split a complex vector $\mathbf{v} \in \mathbb{C}^m$ up into its magnitude $\|\mathbf{v}\|$ and its *direction*

$$\hat{\mathbf{v}} \triangleq \frac{\mathbf{v}}{\|\mathbf{v}\|} \tag{10}$$

where we reserve this notation exclusively for unit vectors, *i.e.*, throughout the paper every vector carrying a hat, $\hat{\mathbf{v}}$ or $\hat{\mathbf{V}}$, denotes a (deterministic or random, respectively) vector of unit length

$$\|\hat{\mathbf{v}}\| = \|\hat{\mathbf{V}}\| = 1. \tag{11}$$

To be able to work with such *direction vectors* 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 $\mathbb{C}^m$. If a random vector $\hat{\mathbf{V}}$ takes value in the unit sphere and has the density $p_{\hat{\mathbf{V}}}^{\lambda}(\hat{\mathbf{v}})$ with respect to λ, then we shall let

$$h_\lambda(\hat{\mathbf{V}}) \triangleq -\mathsf{E}\left[\log p_{\hat{\mathbf{V}}}^{\lambda}(\hat{\mathbf{V}})\right] \tag{12}$$

if the expectation is defined.

We note that just as ordinary differential entropy is invariant under translation, so is $h_\lambda(\hat{\mathbf{V}})$ invariant under rotation. That is, if U is a deterministic unitary matrix, then

$$h_\lambda(\mathsf{U}\hat{\mathbf{V}}) = h_\lambda(\hat{\mathbf{V}}). \tag{13}$$

Also note that $h_\lambda(\hat{\mathbf{V}})$ is maximized if $\hat{\mathbf{V}}$ is uniformly distributed on the unit sphere, in which case

$$h_\lambda(\hat{\mathbf{V}}) = \log c_m \tag{14}$$

where c_m denotes the surface area of the unit sphere in $\mathbb{C}^m$

$$c_m = \frac{2\pi^m}{\Gamma(m)}. \tag{15}$$

The definition (12) can be easily extended to conditional entropies: if $\mathbf{W}$ is some random vector, and if conditional on $\mathbf{W} = \mathbf{w}$ the random vector $\hat{\mathbf{V}}$ has density $p_{\hat{\mathbf{V}}|\mathbf{W}}^{\lambda}(\hat{\mathbf{v}}|\mathbf{w})$ then we can define

$$h_\lambda\big(\hat{\mathbf{V}} \,\big|\, \mathbf{W} = \mathbf{w}\big) \triangleq -\mathsf{E}\left[\log p_{\hat{\mathbf{V}}|\mathbf{W}}^{\lambda}(\hat{\mathbf{V}}|\mathbf{W}) \,\Big|\, \mathbf{W} = \mathbf{w}\right] \tag{16}$$

and we can define $h_\lambda\big(\hat{\mathbf{V}} \,\big|\, \mathbf{W}\big)$ as the expectation (with respect to $\mathbf{W}$) of $h_\lambda\big(\hat{\mathbf{V}} \,\big|\, \mathbf{W} = \mathbf{w}\big)$.

Based on these definitions we have the following lemma:

Lemma 1: Let $\mathbf{V}$ be a complex random vector taking value in $\mathbb{C}^m$ and having differential entropy $h(\mathbf{V})$. Let $\|\mathbf{V}\|$ denote its norm and $\hat{\mathbf{V}}$ denote its direction as in (10). Then

$$h(\mathbf{V}) = h(\|\mathbf{V}\|) + h_\lambda\big(\hat{\mathbf{V}} \,\big|\, \|\mathbf{V}\|\big) + (2m-1)\mathsf{E}[\log\|\mathbf{V}\|] \tag{17}$$

$$= h_\lambda(\hat{\mathbf{V}}) + h\big(\|\mathbf{V}\| \,\big|\, \hat{\mathbf{V}}\big) + (2m-1)\mathsf{E}[\log\|\mathbf{V}\|] \tag{18}$$

whenever all the quantities in (17) and (18), respectively, are defined. Here $h(\|\mathbf{V}\|)$ is the differential entropy of $\|\mathbf{V}\|$ when viewed as a real (scalar) random variable.

Proof: Omitted. □

We shall write $\mathbf{X} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \mathsf{K})$ if $\mathbf{X} - \boldsymbol{\mu}$ is a circularly symmetric, zero-mean, Gaussian random vector of covariance matrix $\mathsf{E}\left[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\dagger\right] = \mathsf{K}$. By $X \sim \mathcal{U}\left([a,b]\right)$ we denote a random variable that is uniformly distributed on the interval $[a,b]$. The probability distribution of a random variable X or random vector $\mathbf{X}$ is denoted by Q_X or $Q_{\mathbf{X}}$, respectively.

Throughout the paper $e^{\mathrm{i}\Theta}$ denotes a complex random variable that is uniformly distributed over the unit circle

$$e^{\mathrm{i}\Theta} \sim \text{Uniform on } \{z \in \mathbb{C} : |z| = 1\}. \tag{19}$$

When it appears in formulas with other random variables, $e^{\mathrm{i}\Theta}$ is always assumed to be independent of these other variables.

All rates specified in this paper are in nats per channel use, *i.e.*, $\log(\cdot)$ denotes the natural logarithmic function.

III. The Channel Model

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

$$\mathbf{Y}_k = \mathbb{H}_k\mathbf{x}_k + \mathbf{Z}_k. \tag{20}$$

Here $\mathbf{x}_k \in \mathbb{C}^{n_{\mathrm{T}}}$ 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; and the random vector $\mathbf{Z}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ denotes the time-k additive noise vector.

We assume that the random vectors $\{\mathbf{Z}_k\}$ are spatially and temporally white, zero-mean, circularly symmetric, complex Gaussian random vectors, *i.e.*, $\{\mathbf{Z}_k\} \sim$ IID $\mathcal{N}_{\mathbb{C}}\big(\mathbf{0}, \sigma^2\mathsf{I}\big)$ for some $\sigma^2 > 0$. Here I denotes the identity matrix.

As for the matrix-valued fading process $\{\mathbb{H}_k\}$ we will not specify a particular distribution, but shall only assume that it is stationary, ergodic, of a finite-energy fading gain, *i.e.*,

$$\mathsf{E}\left[\|\mathbb{H}_k\|_{\mathrm{F}}^2\right] < \infty \tag{21}$$

and *regular*, *i.e.*, its differential entropy rate is finite

$$h(\{\mathbb{H}_k\}) \triangleq \lim_{n\uparrow\infty} \frac{1}{n} h(\mathbb{H}_1, \ldots, \mathbb{H}_n) > -\infty. \tag{22}$$

Furthermore, we will restrict ourselves to the memoryless case, *i.e.*, we assume that $\{\mathbb{H}_k\}$ is IID with respect to time k. Since there is no memory in the channel, an IID input process $\{\mathbf{X}_k\}$ will be sufficient to achieve capacity and we will therefore drop the time index k hereafter, *i.e.*, (20) simplifies to

$$\mathbf{Y} = \mathbb{H}\mathbf{x} + \mathbf{Z}. \tag{23}$$

Note that while we assume that there is no temporal memory in the channel, we do not restrict the spatial memory, *i.e.*, the different fading components $H^{(i,j)}$ of the fading matrix $\mathbb{H}$ may be dependent.

We assume that the fading $\mathbb{H}$ and the additive noise $\mathbf{Z}$ are independent and of a joint law that does not depend on the channel input $\mathbf{x}$.

As for the input, we consider two different constraints: a peak-power constraint and an average-power constraint. We use $\mathcal{E}$ to denote the maximal allowed instantaneous power in the former case, and to denote the allowed average power in the latter case. For both cases we set

$$\text{SNR} \triangleq \frac{\mathcal{E}}{\sigma^2}. \tag{24}$$

The capacity $\mathsf{C}(\text{SNR})$ of the channel (23) is given by

$$\mathsf{C}(\text{SNR}) = \sup_{Q_{\mathbf{X}}} I(\mathbf{X};\mathbf{Y}) \tag{25}$$

where the supremum is over the set of all probability distributions on $\mathbf{X}$ satisfying the constraints, *i.e.*,

$$\|\mathbf{X}\|^2 \leq \mathcal{E}, \quad \text{almost surely} \tag{26}$$

for a peak-power constraint, or

$$\mathsf{E}\left[\|\mathbf{X}\|^2\right] \leq \mathcal{E} \tag{27}$$

for an average-power constraint.

Specializing [1, Th. 4.2], [2, Th. 6.10] to memoryless MIMO fading, we have

$$\varlimsup_{\text{SNR}\uparrow\infty} \Big\{\mathsf{C}(\text{SNR}) - \log\log \text{SNR}\Big\} < \infty. \tag{28}$$

Note that [1, Th. 4.2], [2, Th. 6.10] is stated under the assumption of an average-power constraint only. However, since a peak-power constraint is more stringent than an average-power constraint, (28) also holds in the situation of a peak-power constraint.

The fading number χ is now defined as in [1, Def. 4.6], [2, Def. 6.13] by

$$\chi(\mathbb{H}) \triangleq \varlimsup_{\text{SNR}\uparrow\infty} \Big\{\mathsf{C}(\text{SNR}) - \log\log \text{SNR}\Big\}. \tag{29}$$

Prima facie the fading number depends on whether a peak-power constraint (26) or an average-power constraint (27) is imposed on the input. However, it will turn out that the memoryless MIMO fading number is identical for both cases.

IV. Main Result

A. Preliminaries

Before we can state our main result we need to introduce three concepts: The first concerns probability distributions that escape to infinity, the second a technique of upper-bounding mutual information, and the third concept concerns circular symmetry.

1) Escaping to Infinity: We start with a discussion about the concept of capacity-achieving input distributions that escape to infinity.

A sequence of input distributions parameterized by the allowed cost (in our case of fading channels the cost is the available power or SNR) is said to *escape to infinity* if it assigns to every fixed compact set a probability that tends to zero as the allowed cost tends to infinity. In other words this means that in the limit—when the allowed cost tends to infinity—such a distribution does not use finite-cost symbols.

This notion is important because the asymptotic capacity of many channels of interest can only be achieved by input distributions that escape to infinity. As a matter of fact one can show that every input distribution that only achieves a mutual information of identical asymptotic *growth rate* as the capacity *must* escape to infinity. Loosely speaking, for many channels it is not favorable to use finite-cost input symbols whenever the cost constraint is loosened completely.

In the following we will only state this result specialized to the situation at hand. For a more general description and for all proofs we refer to [8, Sec. VII.C.3], [2, Sec. 2.6].

Definition 2: Let $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ be a family of input distributions for the memoryless fading channel (23), where this family is parameterized by the available average power $\mathcal{E}$ such that

$$\mathsf{E}_{Q_{\mathbf{X},\mathcal{E}}}\left[\|\mathbf{X}\|^2\right] \leq \mathcal{E}, \qquad \mathcal{E} \geq 0. \tag{30}$$

We say that the input distributions $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ *escape to infinity* if for every $\mathcal{E}_0 > 0$

$$\lim_{\mathcal{E}\uparrow\infty} Q_{\mathbf{X},\mathcal{E}}\left(\|\mathbf{X}\|^2 \leq \mathcal{E}_0\right) = 0. \tag{31}$$

We now have the following:

Lemma 3: Let the memoryless MIMO fading channel be given as in (23) and let $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ be a family of distributions on the channel input that satisfy the power constraint (30). Let $I(Q_{\mathbf{X},\mathcal{E}})$ denote the mutual information between input and output of channel (23) when the input is distributed according to the law $Q_{\mathbf{X},\mathcal{E}}$. Assume that the family of input distributions $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ is such that the following condition is satisfied:

$$\lim_{\mathcal{E}\uparrow\infty} \frac{I(Q_{\mathbf{X},\mathcal{E}})}{\log\log \mathcal{E}} = 1. \tag{32}$$

Then $\{Q_{\mathbf{X},\mathcal{E}}\}_{\mathcal{E}\geq 0}$ must escape to infinity.

Proof: A proof can be found in [8, Th. 8, Rem. 9], [2, Cor. 2.8]. □

2) An Upper Bound on Channel Capacity: In [1], [2] a new approach of deriving upper bounds to channel capacity has been introduced. Since capacity is by definition a maximization of mutual information, it is implicitly difficult to find *upper* bounds to it. The proposed technique bases on a dual expression of mutual information that leads to an expression of capacity as a minimization instead of a maximization. This way it becomes much easier to find upper bounds.

Again, here we only state the upper bound in a form needed in the derivation of Theorem 7. For a more general form, for more mathematical details, and for all proofs we refer to [1, Sec. IV], [2, Sec. 2.4].

Lemma 4: Consider a memoryless channel with input $\mathbf{s} \in \mathbb{C}^{n_\text{R}}$ and output $T \in \mathbb{C}$. Then for an arbitrary distribution on the input $\mathbf{S}$ the mutual information between input and output of the channel is upper-bounded as follows:

$$I(\mathbf{S};T) \leq -h(T|\mathbf{S}) + \log\pi + \alpha\log\beta + \log\Gamma\left(\alpha, \frac{\nu}{\beta}\right) + (1-\alpha)\mathsf{E}\left[\log\left(|T|^2+\nu\right)\right] + \frac{1}{\beta}\mathsf{E}\left[|T|^2\right] + \frac{\nu}{\beta} \tag{33}$$

where $\alpha, \beta > 0$ and $\nu \geq 0$ are parameters that can be chosen freely, but must not depend on the distribution of $\mathbf{S}$.

Proof: A proof can be found in [1, Sec. IV], [2, Sec. 2.4]. □

3) Capacity-Achieving Input Distributions and Circular Symmetry: The final preliminary remark concerns circular symmetry. We say that a random vector $\mathbf{W}$ is *circularly symmetric* if

$$\mathbf{W} \stackrel{\mathscr{L}}{=} \mathbf{W} \cdot e^{\mathrm{i}\Theta} \tag{34}$$

where $\Theta \sim \mathcal{U}([0, 2\pi])$ is independent of $\mathbf{W}$ and where $\stackrel{\mathscr{L}}{=}$ stands for "equal in law". Note that this is not to be confused with *isotropically distributed*, which means that a vector has equal probability to point in every direction. Circular symmetry only concerns the phase of the components of a vector, not the vector's direction.

The following lemma says that for our channel model an optimal input can be assumed to be circularly symmetric:

Lemma 5: Assume a channel as given in (23). Then the capacity-achieving input distribution can be assumed to be circularly symmetric, *i.e.*, the input vector $\mathbf{X}$ can be replaced by $\mathbf{X}e^{\mathrm{i}\Theta}$, where $\Theta \sim \mathcal{U}([0, 2\pi])$ is independent of every other random quantity.

Proof: A proof is given in Appendix A. □

Remark 6: Note that the proof of Lemma 5 relies only on the fact that the additive noise is assumed to be circularly symmetric.

B. Fading Number of General Memoryless MIMO Fading Channels

We are now ready for the main result, *i.e.*, the fading number of a memoryless MIMO fading channel:

Theorem 7: Consider a memoryless MIMO fading channel (23) where the random fading matrix $\mathbb{H}$ takes value in $\mathbb{C}^{n_\mathrm{R}\times n_\mathrm{T}}$ and satisfies

$$h(\mathbb{H}) > -\infty \tag{35}$$

and

$$\mathsf{E}\left[\|\mathbb{H}\|_\mathrm{F}^2\right] < \infty. \tag{36}$$

Then, irrespective of whether a peak-power constraint (26) or an average-power constraint (27) is imposed on the input, the fading number $\chi(\mathbb{H})$ is given by

$$\boxed{\begin{aligned}\chi(\mathbb{H}) = \sup_{Q_{\hat{\mathbf{X}}}} \Bigg\{ & h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + n_\mathrm{R}\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\right] \\ & - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big)\Bigg\}.\end{aligned}} \tag{37}$$

Here $\hat{\mathbf{X}}$ denotes a random vector of unit length and $Q_{\hat{\mathbf{X}}}$ denotes its probability law, *i.e.*, the supremum is taken over all distributions of the random unit-vector $\hat{\mathbf{X}}$. Note that the expectation in the second term is understood jointly over $\mathbb{H}$ and $\hat{\mathbf{X}}$.

Moreover, this fading number is achievable by a random vector $\mathbf{X} = \hat{\mathbf{X}} \cdot R$ where $\hat{\mathbf{X}}$ is distributed according to the distribution that achieves the fading number in (37) and where R is a non-negative random variable independent of $\hat{\mathbf{X}}$ such that

$$\log R^2 \sim \mathcal{U}\left([\log\log\mathcal{E}, \log\mathcal{E}]\right). \tag{38}$$

Proof: A proof is given in Section VII. □

Note that—even if it might not be obvious at first sight—it is not hard to show that the distribution $Q_{\hat{\mathbf{X}}}$ that achieves the supremum in (37) is circularly symmetric. This is in agreement with Lemma 5.

The evaluation of (37) can be pretty awkward mainly due to the first term, *i.e.*, the differential entropy with respect to the surface area measure λ. We therefore will derive next an upper bound to the fading number that is easier to evaluate.

To that goal firstly note that for an arbitrary constant non-singular $n_\mathrm{R}\times n_\mathrm{R}$ matrix A and an arbitrary constant non-singular $n_\mathrm{T}\times n_\mathrm{T}$ matrix B

$$\chi(\mathsf{A}\mathbb{H}\mathsf{B}) = \chi(\mathbb{H}), \tag{39}$$

see [1, Lem. 4.7], [2, Lem. 6.14]. Secondly, note that for an arbitrary random unit vector $\hat{\mathbf{Y}} \in \mathbb{C}^{n_\mathrm{R}}$

$$h_\lambda\big(\hat{\mathbf{Y}}\big) \le \log c_{n_\mathrm{R}} = \log\frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})} \tag{40}$$

where c_{n_R} denotes the surface area of the unit sphere in $\mathbb{C}^{n_\mathrm{R}}$ as defined in (15) and where the upper bound is achieved with equality only if $\hat{\mathbf{Y}}$ is uniformly distributed on the sphere, *i.e.*, $\hat{\mathbf{Y}}$ is isotropically distributed.

Using these two observations we get the following upper bound on the fading number:

Corollary 8: The fading number of a memoryless MIMO fading channel as given in Theorem 7 can be upper-bounded as follows:

$$\begin{aligned}\chi(\mathbb{H}) \le{}& n_\mathrm{R}\log\pi - \log\Gamma(n_\mathrm{R}) \\ &+ \inf_{\mathsf{A},\mathsf{B}}\sup_{\hat{\mathbf{x}}}\left\{n_\mathrm{R}\mathsf{E}\left[\log\|\mathsf{A}\mathbb{H}\mathsf{B}\hat{\mathbf{x}}\|^2\right] - h\big(\mathsf{A}\mathbb{H}\mathsf{B}\hat{\mathbf{x}}\big)\right\}\end{aligned} \tag{41}$$

where the infimum is over all non-singular $n_\mathrm{R}\times n_\mathrm{R}$ complex matrices A and non-singular $n_\mathrm{T}\times n_\mathrm{T}$ complex matrices B.

Proof: Using the two observations (39) and (40) we immediately get from Theorem 7:

$$\begin{aligned}\chi(\mathbb{H}) \le \inf_{\mathsf{A},\mathsf{B}}\sup_{Q_{\hat{\mathbf{X}}}} \mathsf{E}_{\hat{\mathbf{X}}}\Bigg[& n_\mathrm{R}\log\pi - \log\Gamma(n_\mathrm{R}) \\ &+ n_\mathrm{R}\mathsf{E}\left[\log\|\mathsf{A}\mathbb{H}\mathsf{B}\hat{\mathbf{X}}\|^2 \,\Big|\, \hat{\mathbf{X}} = \hat{\mathbf{x}}\right] \\ &- h\big(\mathsf{A}\mathbb{H}\mathsf{B}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}} = \hat{\mathbf{x}}\big)\Bigg].\end{aligned} \tag{42}$$

The result now follows by noting that (41) can always be achieved by choosing $Q_{\hat{\mathbf{X}}}$ in (42) to be the distribution which with probability 1 takes on the value $\hat{\mathbf{x}}$ that achieves the maximum in (41). □

This upper bound is possibly tighter than the upper bound given in [1, Lem. 4.14], [2, Lem. 6.16] because of the additional infimum over B.

V. Some Known Special Cases

In this section we will briefly show how some already known results of various fading numbers can be derived as special cases from this new more general result.

We start with the situation of a fading matrix that is *rotation-commutative in the generalized sense*, *i.e.*, the fading matrix $\mathbb{H}$ is such that for every constant unitary $n_\mathrm{T}\times n_\mathrm{T}$ matrix V_t there exists an $n_\mathrm{R}\times n_\mathrm{R}$ constant unitary matrix V_r such that

$$\mathsf{V}_r\mathbb{H} \stackrel{\mathscr{L}}{=} \mathbb{H}\mathsf{V}_t \tag{43}$$

where $\stackrel{\mathscr{L}}{=}$ stands for "has the same law"; and for every constant unitary $n_\mathrm{R}\times n_\mathrm{R}$ matrix V_r there exists a constant unitary $n_\mathrm{T}\times n_\mathrm{T}$ matrix V_t such that (43) holds [1, Def. 4.37], [2, Def. 6.37].

The property of rotation-commutativity for random matrices is a generalization of the isotropic distribution of random vectors, *i.e.*, we have the following:

Lemma 9: Let $\mathbb{H}$ be rotation-commutative in the generalized sense. Then the following two statements hold:

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

$$\|\mathbb{H}\hat{\mathbf{e}}\| \stackrel{\mathscr{L}}{=} \|\mathbb{H}\hat{\mathbf{e}}'\|, \quad \|\hat{\mathbf{e}}\| = \|\hat{\mathbf{e}}'\| = 1 \tag{44}$$

$$h\big(\mathbb{H}\hat{\mathbf{e}}\big) = h\big(\mathbb{H}\hat{\mathbf{e}}'\big), \quad \|\hat{\mathbf{e}}\| = \|\hat{\mathbf{e}}'\| = 1. \tag{45}$$

Proof: For a proof see, *e.g.*, [1, Lem. 4.38], [2, Lem. 6.38]. □

From Lemma 9 it immediately follows that in the situation of rotation-commutative fading the only term in the expression of the fading number (37) that depends on $Q_{\hat{\mathbf{X}}}$ is

$$h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right).$$

This entropy is maximized if $\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}$ is uniformly distributed on the surface of the n_R-dimensional complex unit sphere, which can be achieved according to Lemma 9 by the choice of an isotropic distribution for $Q_{\hat{\mathbf{X}}}$. Then according to (14) and (15)

$$h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) = \log\frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})}. \tag{46}$$

The expression of the fading number (37) then reduces to (5):

$$\chi(\mathbb{H}) = \log \frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})} - \log 2 + n_\mathrm{R} \mathsf{E}\left[\log \|\mathbb{H}\hat{\mathbf{e}}\|^2\right] - h(\mathbb{H}\hat{\mathbf{e}}) \tag{47}$$

where $\hat{\mathbf{e}}$ is an arbitrary constant unit vector in $\mathbb{C}^{n_\mathrm{T}}$.

In case of a SIMO fading channel, the direction vector $\hat{\mathbf{X}}$ reduces to a phase term $e^{\mathrm{i}\Phi}$. From Lemma 5 we know that an optimal choice of $e^{\mathrm{i}\Phi}$ is circularly symmetric, such that (37) becomes

$$\chi(\mathbf{H}) = h_\lambda\big(\hat{\mathbf{H}} e^{\mathrm{i}\Theta}\big) + n_\mathrm{R} \mathsf{E}\left[\log \|\mathbf{H}\|^2\right] - \log 2 - h(\mathbf{H}). \tag{48}$$

Before we continue with the MISO case, we would like to remark that the only term in (37) that depends on the distribution of the phase of each component of $\mathbf{X}$ is

$$h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right).$$

Since we know from Lemma 5 that $\hat{\mathbf{X}}$ is circularly symmetric, we can therefore equivalently write

$$h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) = h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} e^{\mathrm{i}\Theta}\right). \tag{49}$$

Turning to the MISO case now note that the distribution of

$$\frac{\mathbf{H}^\mathsf{T}\hat{\mathbf{X}}}{|\mathbf{H}^\mathsf{T}\hat{\mathbf{X}}|} e^{\mathrm{i}\Theta}$$

is identical to the distribution of $e^{\mathrm{i}\Theta}$, independently of the distribution of $\mathbf{H}$ and $\hat{\mathbf{X}}$. Hence,

$$h_\lambda\left(\frac{\mathbf{H}^\mathsf{T}\hat{\mathbf{X}}}{|\mathbf{H}^\mathsf{T}\hat{\mathbf{X}}|} e^{\mathrm{i}\Theta}\right) = h_\lambda\big(e^{\mathrm{i}\Theta}\big) = \log 2\pi. \tag{50}$$

The fading number (37) then becomes

$$\begin{aligned}
\chi(\mathbf{H}^\mathsf{T}) &= \sup_{Q_{\hat{\mathbf{X}}}} \bigg\{ \log 2\pi + \mathsf{E}\Big[\log |\mathbf{H}^\mathsf{T}\hat{\mathbf{X}}|^2\Big] - \log 2 \\
&\qquad - h\big(\mathbf{H}^\mathsf{T}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) \bigg\} \qquad (51)\\
&= \sup_{Q_{\hat{\mathbf{X}}}} \mathsf{E}_{\hat{\mathbf{X}}} \Big[\log \pi + \mathsf{E}\Big[\log |\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}|^2 \,\Big|\, \hat{\mathbf{X}} = \hat{\mathbf{x}} \Big] \\
&\qquad - h\big(\mathbf{H}^\mathsf{T}\hat{\mathbf{x}} \,\big|\, \hat{\mathbf{X}} = \hat{\mathbf{x}}\big) \Big] \qquad (52)\\
&\le \sup_{\hat{\mathbf{x}}} \left\{ \log \pi + \mathsf{E}\left[\log |\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}|^2\right] - h\big(\mathbf{H}^\mathsf{T}\hat{\mathbf{x}}\big) \right\} \qquad (53)
\end{aligned}$$

which can be achieved for a distribution of $Q_{\hat{\mathbf{X}}}$ that with probability 1 takes on the value $\hat{\mathbf{x}}$ that achieves the fading number in (53).

Finally, the SISO case is a combination of the arguments of the SIMO and MISO case, *i.e.*, using

$$h_\lambda\big(e^{\mathrm{i}\Theta}\big) = \log 2\pi \tag{54}$$

we get

$$\begin{aligned}
\chi(H) &= \log 2\pi + \mathsf{E}\left[\log |H|^2\right] - \log 2 - h(H) \qquad (55)\\
&= \log \pi + \mathsf{E}\left[\log |H|^2\right] - h(H). \qquad (56)
\end{aligned}$$

VI. Gaussian Fading

The evaluation of the fading number is rather difficult even for the usually simpler situation of Gaussian fading processes. However, we are able to give the exact value for some important special cases, and we will give bounds on some others.

Throughout this section we assume that the fading matrix $\mathbb{H}$ can be written as

$$\mathbb{H} = \mathsf{D} + \tilde{\mathbb{H}} \tag{57}$$

where all components of $\tilde{\mathbb{H}}$ are independent of each other and zero-mean, unit-variance Gaussian distributed, and where D denotes a constant line-of-sight matrix.

Note that for some constant unitary $n_\mathrm{R} \times n_\mathrm{R}$ matrix U and some constant unitary $n_\mathrm{T} \times n_\mathrm{T}$ matrix V the law of $\mathsf{U}\tilde{\mathbb{H}}\mathsf{V}$ is identical to the law of $\tilde{\mathbb{H}}$. Therefore, without loss of generality, we may restrict ourselves to matrices D that are "diagonal", *i.e.*, for $n_\mathrm{R} \le n_\mathrm{T}$,

$$\mathsf{D} = \begin{pmatrix} \tilde{\mathsf{D}} & \mathsf{0}_{n_\mathrm{R}\times(n_\mathrm{T}-n_\mathrm{R})} \end{pmatrix} \tag{58}$$

or, for $n_\mathrm{R} > n_\mathrm{T}$,

$$\mathsf{D} = \begin{pmatrix} \tilde{\mathsf{D}} \\ \mathsf{0}_{(n_\mathrm{R}-n_\mathrm{T})\times n_\mathrm{T}} \end{pmatrix} \tag{59}$$

where $\tilde{\mathsf{D}}$ is a $\min\{n_\mathrm{R}, n_\mathrm{T}\} \times \min\{n_\mathrm{R}, n_\mathrm{T}\}$ diagonal matrix with the singular values of D on the diagonal.

A. Scalar Line-of-Sight Matrix

We start with a scalar line-of-sight matrix, *i.e.*, we assume $\tilde{\mathsf{D}} = d\mathsf{I}$ where I denotes the identity matrix.

Under these assumptions the fading number has been known already for $n_\mathrm{R} = n_\mathrm{T} = m$, in which case the fading matrix $\mathbb{H}$ is rotation commutative [1], [2]:

$$\chi(\mathbb{H}) = m g_m\big(|d|^2\big) - m - \log\Gamma(m). \tag{60}$$

Here $g_m(\cdot)$ is a continuous, monotonically increasing, concave function defined as

$$g_m(\xi) \triangleq \begin{cases} \log(\xi) - \mathrm{Ei}(-\xi) + \displaystyle\sum_{j=1}^{m-1} (-1)^j \bigg[e^{-\xi}(j-1)! \\ \qquad - \dfrac{(m-1)!}{j(m-1-j)!}\bigg] \left(\dfrac{1}{\xi}\right)^j, & \xi > 0 \\ \psi(m), & \xi = 0 \end{cases} \tag{61}$$

for $m \in \mathbb{N}$, where $\mathrm{Ei}(\cdot)$ denotes the exponential integral function defined as

$$\mathrm{Ei}(-x) \triangleq -\int_x^\infty \frac{e^{-t}}{t}\,\mathrm{d}t, \qquad x > 0 \tag{62}$$

and $\psi(\cdot)$ is Euler's psi function given by

$$\psi(m) \triangleq -\gamma + \sum_{j=1}^{m-1} \frac{1}{j} \tag{63}$$

with $\gamma \approx 0.577$ denoting Euler's constant. This function $g_m(\cdot)$ is the expected value of the logarithm of a non-central chi-square random variable, *i.e.*, for some Gaussian random variables $\{U_j\}_{j=1}^m$ IID $\sim \mathcal{N}_\mathbb{C}(0,1)$ and for some complex constants $\{\mu_j\}_{j=1}^m$ we have

$$\mathsf{E}\left[\log\left(\sum_{j=1}^m |U_j + \mu_j|^2\right)\right] = g_m(s^2) \tag{64}$$

where

$$s^2 \triangleq \sum_{j=1}^m |\mu_j|^2 \tag{65}$$

(see [9], [1, Lem. 10.1], [2, Lem. A.6] for more details and a proof).

We would like to emphasize that $g_m(\xi)$ is continuous for all $\xi \ge 0$, *i.e.*, in particular

$$\lim_{\xi\downarrow 0} \left\{ \log(\xi) - \mathrm{Ei}(-\xi) + \sum_{j=1}^{m-1} (-1)^j \bigg[e^{-\xi}(j-1)! - \frac{(m-1)!}{j(m-1-j)!}\bigg] \left(\frac{1}{\xi}\right)^j \right\} = \psi(m) \tag{66}$$

for all $m \in \mathbb{N}$. Moreover, for all $m \in \mathbb{N}$ and $\xi \geq 0$

$$\log \xi - \mathrm{Ei}(-\xi) \leq g_m(\xi) \leq \log(m+\xi). \tag{67}$$

A derivation of (67) can be found in Appendix B.

We now consider the case where $n_\mathrm{R} \leq n_\mathrm{T}$:

Corollary 10: Assume $n_\mathrm{R} \leq n_\mathrm{T}$ and a Gaussian fading matrix as given in (57). Let the line-of-sight matrix D be given as

$$\mathsf{D} = d \left(\mathsf{I}_{n_\mathrm{R}} \quad \mathsf{0}_{n_\mathrm{R}\times(n_\mathrm{T}-n_\mathrm{R})} \right). \tag{68}$$

Then

$$\chi(\mathbb{H}) = n_\mathrm{R} g_{n_\mathrm{R}}\left(|d|^2\right) - n_\mathrm{R} - \log\Gamma(n_\mathrm{R}) \tag{69}$$

where $g_m(\cdot)$ is defined in (61).

Proof: We write for the unit vector $\hat{\mathbf{X}}$

$$\hat{\mathbf{X}} = \begin{pmatrix} \mathbf{\Xi} \\ \mathbf{\Xi}' \end{pmatrix} \tag{70}$$

where $\mathbf{\Xi} \in \mathbb{C}^{n_\mathrm{R}}$ and $\mathbf{\Xi}' \in \mathbb{C}^{n_\mathrm{T}-n_\mathrm{R}}$. Then

$$\mathbb{H}\hat{\mathbf{X}} = \mathsf{D}\hat{\mathbf{X}} + \tilde{\mathbb{H}}\hat{\mathbf{X}} = d\mathbf{\Xi} + \tilde{\mathbf{H}} \tag{71}$$

where $\tilde{\mathbf{H}} \sim \mathcal{N}_\mathbb{C}(\mathbf{0}, \mathsf{I}_{n_\mathrm{R}})$. Hence,

$$h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) = h\big(\tilde{\mathbf{H}}\big) = n_\mathrm{R}\log \pi e; \tag{72}$$

$$n_\mathrm{R}\mathsf{E}\Big[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] = n_\mathrm{R} g_{n_\mathrm{R}}\left(|d|^2\|\mathbf{\Xi}\|^2\right) \leq n_\mathrm{R} g_{n_\mathrm{R}}\left(|d|^2\right); \tag{73}$$

$$h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) \leq \log \frac{2\pi^{n_\mathrm{R}}}{\Gamma(n_\mathrm{R})}. \tag{74}$$

Here, the equality in (73) follows from the fact that $\|d\mathbf{\Xi} + \tilde{\mathbf{H}}\|^2$ is non-central chi-square distributed and from (64); the inequality in (73) follows from the monotonicity of $g_m(\cdot)$ and is tight if $\|\mathbf{\Xi}\| = 1$, *i.e.*, $\mathbf{\Xi}' = \mathbf{0}$; and the inequality in (74) follows from (14) and (15) and is tight if $\mathbf{\Xi}$ is uniformly distributed on the unit sphere in $\mathbb{C}^{n_\mathrm{R}}$ so that $\mathbb{H}\hat{\mathbf{X}}$ is isotropically distributed. The result now follows from Theorem 7. □

The case $n_\mathrm{R} > n_\mathrm{T}$ is more difficult since then (74) is in general not tight. We will only state an upper bound:

Proposition 11: Assume $n_\mathrm{R} > n_\mathrm{T}$ and a Gaussian fading matrix as given in (57). Let the line-of-sight matrix D be given as

$$\mathsf{D} = d \begin{pmatrix} \mathsf{I}_{n_\mathrm{T}} \\ \mathsf{0}_{(n_\mathrm{R}-n_\mathrm{T})\times n_\mathrm{T}} \end{pmatrix}. \tag{75}$$

Then

$$\chi(\mathbb{H}) \leq n_\mathrm{T} \log\left(1 + \frac{|d|^2}{n_\mathrm{T}}\right) + n_\mathrm{R}\log n_\mathrm{R} - n_\mathrm{R} - \log\Gamma(n_\mathrm{R}). \tag{76}$$

Proof: This result is a special case of Proposition 13 and has been published before in [1, (128)], [2, (6.224)]. □

B. General Line-of-Sight Matrix

Next we assume Gaussian fading as defined in (57) with a general line-of-sight matrix D having singular values $d_1, \ldots, d_{\min\{n_\mathrm{R},n_\mathrm{T}\}}$. Hence, $\tilde{\mathsf{D}}$, defined in (58) and (59), is given as

$$\tilde{\mathsf{D}} = \mathrm{diag}\left(d_1, \ldots, d_{\min\{n_\mathrm{R},n_\mathrm{T}\}}\right) \tag{77}$$

where $|d_1| \geq |d_2| \geq \ldots \geq |d_{\min\{n_\mathrm{R},n_\mathrm{T}\}}| > 0$.

We again start with the case $n_\mathrm{R} \leq n_\mathrm{T}$.

Corollary 12: Assume $n_\mathrm{R} \leq n_\mathrm{T}$ and a Gaussian fading matrix as given in (57). Let the line-of-sight matrix D have singular values $d_1, \ldots, d_{n_\mathrm{R}}$, where $|d_1| \geq |d_2| \geq \ldots \geq |d_{n_\mathrm{R}}| > 0$. Then

$$\chi(\mathbb{H}) \leq n_\mathrm{R} g_{n_\mathrm{R}}\left(\|\mathsf{D}\|^2\right) - n_\mathrm{R} - \log\Gamma(n_\mathrm{R}) \tag{78}$$

where $g_m(\cdot)$ is given in (61) and where $\|\mathsf{D}\|^2 = |d_1|^2$.

Proof: A proof is given in Appendix C. □

The situation $n_\mathrm{R} > n_\mathrm{T}$ is again more complicated. We include this case in a new upper bound based on (41) which holds independently of the particular relation between n_R and n_T:

Proposition 13: Assume a Gaussian fading matrix as given in (57) and let the line-of-sight matrix D be general with singular values $d_1, \ldots, d_{\min\{n_\mathrm{R},n_\mathrm{T}\}}$, where $|d_1| \geq |d_2| \geq \ldots \geq |d_{\min\{n_\mathrm{R},n_\mathrm{T}\}}| > 0$. Then the fading number is upper-bounded as follows:

$$\chi(\mathbb{H}) \leq \min\{n_\mathrm{R},n_\mathrm{T}\}\log\frac{\delta^2}{\min\{n_\mathrm{R},n_\mathrm{T}\}} + n_\mathrm{R}\log n_\mathrm{R} - n_\mathrm{R} - \log\Gamma(n_\mathrm{R}) \tag{79}$$

where

$$\delta^2 \triangleq \left(|d_1|^2 \cdot \cdots \cdot |d_{\min\{n_\mathrm{R},n_\mathrm{T}\}}|^2\right)^{1/\min\{n_\mathrm{R},n_\mathrm{T}\}} \cdot \left(1 + \frac{1}{|d_1|^2} + \cdots + \frac{1}{|d_{\min\{n_\mathrm{R},n_\mathrm{T}\}}|^2}\right). \tag{80}$$

Proof: A proof is given in Appendix D. □

VII. Proof of the Main Result

The proof of Theorem 7 consists of two parts: firstly we derive an upper bound to the fading number assuming an average-power constraint (27) on the input. The key ingredients here are the preliminary results from Section IV-A.

In a second part we then show that this upper bound can actually be achieved by an input that satisfies the peak-power constraint (26). Since a peak-power constraint is more restrictive than the corresponding average-power constraint, the theorem follows.

Because the proof is rather technical, we will give a short overview to clarify the main ideas.

The upper bound relies strongly on Lemma 3 which says that the input can be assumed to take on large values only, *i.e.*, at high SNR the additive noise will become negligible so that we can bound

$$I(\mathbf{X};\mathbf{Y}) \lessapprox I(\mathbf{X};\mathbb{H}\mathbf{X}). \tag{81}$$

This term is then split into a term that only considers the magnitude of $\mathbb{H}\mathbf{X}$ and a term that takes into account the direction:

$$I(\mathbf{X};\mathbb{H}\mathbf{X}) = I(\mathbf{X};\|\mathbb{H}\mathbf{X}\|) + I\left(\mathbf{X};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\mathbf{X}\|\right). \tag{82}$$

For the first term—which is related to MISO fading—we then use the bounding technique of Lemma 4.

Because Lemma 3 only holds in the limit when $\mathcal{E}$ tends to infinity, we introduce an event $\|\mathbf{X}\|^2 > \mathcal{E}_0$ for some fixed $\mathcal{E}_0 \geq 0$ and condition everything on this event.

To derive a lower bound on capacity we choose a specific input distribution of the form

$$\mathbf{X} = R \cdot \hat{\mathbf{X}} \tag{83}$$

where the distribution of R is such that it achieves the fading number of a SIMO fading channel and where the distribution of $\hat{\mathbf{X}}$ is independent of R and will be only specified at the very end of the derivation (it will be chosen to maximize the fading number). We then split the mutual information into two terms:

$$I(\mathbf{X};\mathbf{Y}) = I\big(R;\mathbf{Y}\,\big|\,\hat{\mathbf{X}}\big) + I\big(\hat{\mathbf{X}};\mathbf{Y}\big). \tag{84}$$

The first term (almost) corresponds to a SIMO fading channel with side-information for which the fading number is known. The second term is treated separately.

A. Derivation of an Upper Bound

In the following we will use the notation $R \triangleq \|\mathbf{X}\|$ to denote the magnitude of the input vector $\mathbf{X}$, *i.e.*, we have $\mathbf{X} = R \cdot \hat{\mathbf{X}}$. Note that in this section we are not allowed to assume that R is independent of $\hat{\mathbf{X}}$.

From Lemma 3 we know that the capacity-achieving input distribution must escape to infinity. Hence, we fix an arbitrary finite $\mathcal{E}_0 \geq 0$ and define an indicator random variable as follows:

$$E \triangleq \begin{cases} 1 & \text{if } \|\mathbf{X}\|^2 \geq \mathcal{E}_0, \\ 0 & \text{otherwise.} \end{cases} \tag{85}$$

Let

$$p \triangleq \Pr[E = 1] = \Pr\big[\|\mathbf{X}\|^2 \geq \mathcal{E}_0\big] \tag{86}$$

where we know from Lemma 3 that

$$\lim_{\mathcal{E}\uparrow\infty} p = 1. \tag{87}$$

We now bound as follows:

$$\begin{aligned}
I(\mathbf{X};\mathbf{Y}) &\leq I(\mathbf{X},E;\mathbf{Y}) && (88)\\
&= I(E;\mathbf{Y}) + I\big(\mathbf{X};\mathbf{Y}\,\big|\,E\big) && (89)\\
&= H(E) - H\big(E\,\big|\,\mathbf{Y}\big) + I\big(\mathbf{X};\mathbf{Y}\,\big|\,E\big) && (90)\\
&\leq H(E) + I\big(\mathbf{X};\mathbf{Y}\,\big|\,E\big) && (91)\\
&= H_{\mathrm{b}}(p) + pI\big(\mathbf{X};\mathbf{Y}\,\big|\,E=1\big) \\
&\quad + (1-p)I\big(\mathbf{X};\mathbf{Y}\,\big|\,E=0\big) && (92)\\
&\leq H_{\mathrm{b}}(p) + I\big(\mathbf{X};\mathbf{Y}\,\big|\,E=1\big) + (1-p)\mathsf{C}(\mathcal{E}_0) && (93)
\end{aligned}$$

where

$$H_{\mathrm{b}}(\xi) \triangleq -\xi\log\xi - (1-\xi)\log(1-\xi) \tag{94}$$

is the binary entropy function. Here, (88) follows from adding an additional random variable to mutual information; the subsequent two equalities follow from the chain rule and from the definition of mutual information (notice that we use entropy and not differential entropy because E is a binary random variable); in the subsequent inequality we rely on the non-negativity of entropy; and the last inequality follows from bounding $p \leq 1$ and from upper-bounding the mutual information term by the capacity C for the available power which—conditional on $E = 0$—is $\mathcal{E}_0$.

We remark that even though $\mathsf{C}(\mathcal{E}_0)$ is unknown, we know that it is finite and independent of $\mathcal{E}$ so that from (87) we have

$$\lim_{\mathcal{E}\uparrow\infty} \big\{H_{\mathrm{b}}(p) + (1-p)\mathsf{C}(\mathcal{E}_0)\big\} = 0. \tag{95}$$

We continue with the second term of (93) as follows:

$$\begin{aligned}
&I\big(\mathbf{X};\mathbf{Y}\,\big|\,E=1\big)\\
&= I\big(\mathbf{X};\mathbb{H}\mathbf{X}+\mathbf{Z}\,\big|\,E=1\big) && (96)\\
&\leq I\big(\mathbf{X};\mathbb{H}\mathbf{X}+\mathbf{Z},\mathbf{Z}\,\big|\,E=1\big) && (97)\\
&= I\big(\mathbf{X};\mathbb{H}\mathbf{X},\mathbf{Z}\,\big|\,E=1\big) && (98)\\
&= I\big(\mathbf{X};\mathbb{H}\mathbf{X}\,\big|\,E=1\big) + I\big(\mathbf{X};\mathbf{Z}\,\big|\,\mathbb{H}\mathbf{X},E=1\big) && (99)\\
&= I\big(\mathbf{X};\mathbb{H}\mathbf{X}\,\big|\,E=1\big) && (100)\\
&= I\left(\mathbf{X};\|\mathbb{H}\mathbf{X}\|,\frac{\mathbb{H}\mathbf{X}}{\|\mathbb{H}\mathbf{X}\|}\,\middle|\,E=1\right) && (101)\\
&= I\left(\mathbf{X};\|\mathbb{H}\mathbf{X}\|,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,E=1\right) && (102)\\
&= I\big(\mathbf{X};\|\mathbb{H}\mathbf{X}\|\,\big|\,E=1\big)\\
&\quad + I\left(\mathbf{X};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\mathbf{X}\|,E=1\right) && (103)\\
&\leq I\big(\mathbf{X};\|\mathbb{H}\mathbf{X}\|,e^{\mathrm{i}\Theta}\,\big|\,E=1\big)\\
&\quad + I\left(\mathbf{X};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\mathbf{X}\|,E=1\right) && (104)\\
&= I\big(\mathbf{X};\|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta}\,\big|\,E=1\big)\\
&\quad + I\left(\mathbf{X};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\mathbf{X}\|,E=1\right). && (105)
\end{aligned}$$

Here, (97) follows from adding an additional random vector $\mathbf{Z}$ to the argument of the mutual information; the subsequent equality from subtracting the known vector $\mathbf{Z}$ from $\mathbf{Y}$; the subsequent two equalities follow from the chain rule and the independence between the noise and all other random quantities; then we split $\mathbb{H}\mathbf{X}$ into magnitude and direction vector and use the chain rule again; (104) follows from adding a random variable to mutual information: we introduce $e^{\mathrm{i}\Theta}$ that is independent of all the other random quantities and that is uniformly distributed on the complex unit circle; and the last equality holds because from $\|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta}$ we can easily get back $\|\mathbb{H}\mathbf{X}\|$ and $e^{\mathrm{i}\Theta}$.

We next apply Lemma 4 to the first term in (105), *i.e.*, we choose $\mathbf{S} = \mathbf{X}$ and $T = \|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta}$. Note that we need to condition everything on the event $E = 1$:

$$\begin{aligned}
&I\big(\mathbf{X};\|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta}\,\big|\,E=1\big)\\
&\leq -h\big(\|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta}\,\big|\,\mathbf{X},E=1\big) + \log\pi + \alpha\log\beta\\
&\quad + \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right) + (1-\alpha)\mathsf{E}\left[\log\left(\|\mathbb{H}\mathbf{X}\|^2+\nu\right)\,\middle|\,E=1\right]\\
&\quad + \frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}\mathbf{X}\|^2\,\middle|\,E=1\right] + \frac{\nu}{\beta} && (106)
\end{aligned}$$

where $\alpha, \beta > 0$, and $\nu \geq 0$ can be chosen freely, but must not depend on $\mathbf{X}$.

Notice that from a conditional version of Lemma 1 with $m = 1$ follows that

$$\begin{aligned}
&h\big(\|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta}\,\big|\,\mathbf{X}=\mathbf{x},E=1\big)\\
&= h_\lambda\big(e^{\mathrm{i}\Theta}\,\big|\,\mathbf{X}=\mathbf{x},E=1\big) + h\big(\|\mathbb{H}\mathbf{X}\|\,\big|\,e^{\mathrm{i}\Theta},\mathbf{X}=\mathbf{x},E=1\big)\\
&\quad + \mathsf{E}\big[\log\|\mathbb{H}\mathbf{X}\|\,\big|\,\mathbf{X}=\mathbf{x},E=1\big] && (107)\\
&= \log 2\pi + h\big(\|\mathbb{H}\mathbf{X}\|\,\big|\,\mathbf{X}=\mathbf{x},E=1\big)\\
&\quad + \mathsf{E}\big[\log\|\mathbb{H}\mathbf{X}\|\,\big|\,\mathbf{X}=\mathbf{x},E=1\big] && (108)
\end{aligned}$$

where we have used that $e^{\mathrm{i}\Theta}$ is independent of all other random quantities and uniformly distributed on the unit circle. Taking the expectation over $\mathbf{X}$ conditional on $E = 1$ we then yield

$$\begin{aligned}
&h\big(\|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta}\,\big|\,\mathbf{X},E=1\big)\\
&= \log 2\pi + h\big(\|\mathbb{H}\mathbf{X}\|\,\big|\,\mathbf{X},E=1\big) + \mathsf{E}\big[\log\|\mathbb{H}\mathbf{X}\|\,\big|\,E=1\big] && (109)\\
&= \log 2\pi + h\big(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\big|\,\hat{\mathbf{X}},R,E=1\big)\\
&\quad + \mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,E=1\right] && (110)\\
&= \log 2\pi + h\big(\|\mathbb{H}\hat{\mathbf{X}}\|\,\big|\,\hat{\mathbf{X}},R,E=1\big) + \mathsf{E}\big[\log R\,\big|\,E=1\big]\\
&\quad + \mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|\,\middle|\,E=1\right] + \mathsf{E}\big[\log R\,\big|\,E=1\big] && (111)\\
&= \log 2\pi + h\big(\|\mathbb{H}\hat{\mathbf{X}}\|\,\big|\,\hat{\mathbf{X}},E=1\big) + 2\mathsf{E}\big[\log R\,\big|\,E=1\big]\\
&\quad + \mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|\,\middle|\,E=1\right] && (112)
\end{aligned}$$

where the second equality follows from the definition of $R \triangleq \|\mathbf{X}\|$; where the third equality follows from the scaling property of entropy with a *real* argument; and where the last equality follows because given $\hat{\mathbf{X}}$, $\|\mathbb{H}\hat{\mathbf{X}}\|$ is independent of R.

Next we assume $0 < \alpha < 1$ such that $1 - \alpha > 0$. Then we define

$$\epsilon_\nu \triangleq \sup_{\|\mathbf{x}\|^2\geq\mathcal{E}_0}\left\{\mathsf{E}\left[\log\left(\|\mathbb{H}\mathbf{x}\|^2+\nu\right)\right] - \mathsf{E}\left[\log\|\mathbb{H}\mathbf{x}\|^2\right]\right\} \tag{113}$$

such that

$$
\begin{aligned}
&(1-\alpha)\mathsf{E}\left[\log\left(\|\mathbb{H}\mathbf{X}\|^2+\nu\right)\,\middle|\, E=1\right]\\
&\quad=(1-\alpha)\mathsf{E}\left[\log\left(\|\mathbb{H}\mathbf{X}\|^2+\nu\right)-\log\|\mathbb{H}\mathbf{X}\|^2\,\middle|\, E=1\right]\\
&\qquad+(1-\alpha)\mathsf{E}\left[\log\|\mathbb{H}\mathbf{X}\|^2\,\middle|\, E=1\right] && (114)\\
&\quad\le(1-\alpha)\sup_{\|\mathbf{x}\|^2\ge\mathcal{E}_0}\left\{\mathsf{E}\left[\log\left(\|\mathbb{H}\mathbf{x}\|^2+\nu\right)\right]-\mathsf{E}\left[\log\|\mathbb{H}\mathbf{x}\|^2\right]\right\}\\
&\qquad+(1-\alpha)\mathsf{E}\left[\log\|\mathbb{H}\mathbf{X}\|^2\,\middle|\, E=1\right] && (115)\\
&\quad=(1-\alpha)\epsilon_\nu+(1-\alpha)\mathsf{E}\left[\log\|\mathbb{H}\mathbf{X}\|^2\,\middle|\, E=1\right] && (116)\\
&\quad\le\epsilon_\nu+(1-\alpha)\mathsf{E}\left[\log\|\mathbb{H}\mathbf{X}\|^2\,\middle|\, E=1\right]. && (117)
\end{aligned}
$$

Note that in the first inequality we have made use of the fact that $E=1$, *i.e.*, that $\|\mathbf{X}\|^2\ge\mathcal{E}_0$.

Finally, we bound

$$
\begin{aligned}
\frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}\mathbf{X}\|^2\,\middle|\, E=1\right]&=\frac{1}{\beta}\mathsf{E}\left[\|\mathbb{H}\hat{\mathbf{X}}\|^2\cdot R^2\,\middle|\, E=1\right] && (118)\\
&\le\frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\left[\|\mathbb{H}\hat{\mathbf{x}}\|^2\cdot R^2\,\middle|\, E=1\right] && (119)\\
&=\frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\left[\|\mathbb{H}\hat{\mathbf{x}}\|^2\right]\cdot\mathsf{E}\left[R^2\,\middle|\, E=1\right] && (120)\\
&\le\frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\left[\|\mathbb{H}\hat{\mathbf{x}}\|^2\right]\cdot\frac{\mathcal{E}}{p} && (121)
\end{aligned}
$$

where we have used the fact that R needs to satisfy the average-power constraint (27) to get the following bound:

$$
\begin{aligned}
\mathcal{E}&\ge\mathsf{E}\left[R^2\right] && (122)\\
&=p\mathsf{E}\left[R^2\,\middle|\, E=1\right]+(1-p)\mathsf{E}\left[R^2\,\middle|\, E=0\right] && (123)\\
&\ge p\mathsf{E}\left[R^2\,\middle|\, E=1\right]. && (124)
\end{aligned}
$$

Plugging (112), (117), and (121) into (106) we yield

$$
\begin{aligned}
&I\left(\mathbf{X};\|\mathbb{H}\mathbf{X}\|e^{\mathrm{i}\Theta}\,\middle|\, E=1\right)\\
&\quad\le-\log 2-h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\,\middle|\,\hat{\mathbf{X}},E=1\right)-2\mathsf{E}[\log R\mid E=1]\\
&\qquad-\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|\,\middle|\, E=1\right]+\alpha\log\beta+\log\Gamma\left(\alpha,\frac{\nu}{\beta}\right)\\
&\qquad+(1-\alpha)\mathsf{E}\left[\log\|\mathbb{H}\mathbf{X}\|^2\,\middle|\, E=1\right]+\epsilon_\nu\\
&\qquad+\frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\left[\|\mathbb{H}\hat{\mathbf{x}}\|^2\right]\frac{\mathcal{E}}{p}+\frac{\nu}{\beta}. && (125)
\end{aligned}
$$

Next we continue with the second term in (105):

$$
\begin{aligned}
&I\left(\mathbf{X};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\mathbf{X}\|,E=1\right)\\
&\quad=h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\mathbf{X}\|,E=1\right)\\
&\qquad-h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R,\hat{\mathbf{X}},R,E=1\right) && (126)\\
&\quad=h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\mathbf{X}\|,E=1\right)\\
&\qquad-h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\hat{\mathbf{X}}\|,\hat{\mathbf{X}},R,E=1\right) && (127)\\
&\quad\le h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\, E=1\right)\\
&\qquad-h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\hat{\mathbf{X}}\|,\hat{\mathbf{X}},E=1\right). && (128)
\end{aligned}
$$

Here, the last inequality follows because conditioning cannot increase entropy and because given $\hat{\mathbf{X}}$ and $\|\mathbb{H}\hat{\mathbf{X}}\|$, the term $\mathbb{H}\hat{\mathbf{X}}/\|\mathbb{H}\hat{\mathbf{X}}\|$ does not depend on R.

Hence, using (128), (125), and (105) in (93) we get

$$
\begin{aligned}
&I(\mathbf{X};\mathbf{Y})\\
&\quad\le H_{\mathrm{b}}(p)+(1-p)\mathsf{C}(\mathcal{E}_0)-\log 2-h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\,\middle|\,\hat{\mathbf{X}},E=1\right)\\
&\qquad-2\mathsf{E}[\log R\mid E=1]-\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|\,\middle|\, E=1\right]+\alpha\log\beta\\
&\qquad+\log\Gamma\left(\alpha,\frac{\nu}{\beta}\right)+(1-\alpha)\mathsf{E}\left[\log\|\mathbb{H}\mathbf{X}\|^2\,\middle|\, E=1\right]\\
&\qquad+\epsilon_\nu+\frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\left[\|\mathbb{H}\hat{\mathbf{x}}\|^2\right]\frac{\mathcal{E}}{p}+\frac{\nu}{\beta}+h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\, E=1\right)\\
&\qquad-h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\|\mathbb{H}\hat{\mathbf{X}}\|,\hat{\mathbf{X}},E=1\right) && (129)\\
&\quad=H_{\mathrm{b}}(p)+(1-p)\mathsf{C}(\mathcal{E}_0)-\log 2-h\left(\mathbb{H}\hat{\mathbf{X}}\,\middle|\,\hat{\mathbf{X}},E=1\right)\\
&\qquad+(2n_{\mathrm{R}}-1)\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|\,\middle|\, E=1\right]-2\mathsf{E}[\log R\mid E=1]\\
&\qquad-\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|\,\middle|\, E=1\right]+\alpha\log\beta+\log\Gamma\left(\alpha,\frac{\nu}{\beta}\right)\\
&\qquad+2\mathsf{E}[\log\|\mathbb{H}\mathbf{X}\|\mid E=1]-\alpha\mathsf{E}\left[\log\|\mathbb{H}\mathbf{X}\|^2\,\middle|\, E=1\right]+\epsilon_\nu\\
&\qquad+\frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\left[\|\mathbb{H}\hat{\mathbf{x}}\|^2\right]\frac{\mathcal{E}}{p}+\frac{\nu}{\beta}+h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\, E=1\right) && (130)\\
&\quad=h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\, E=1\right)-h\left(\mathbb{H}\hat{\mathbf{X}}\,\middle|\,\hat{\mathbf{X}},E=1\right)\\
&\qquad+n_{\mathrm{R}}\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\,\middle|\, E=1\right]-\log 2+\log\Gamma\left(\alpha,\frac{\nu}{\beta}\right)\\
&\qquad+\frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\left[\|\mathbb{H}\hat{\mathbf{x}}\|^2\right]\frac{\mathcal{E}}{p}+\frac{\nu}{\beta}+\epsilon_\nu+H_{\mathrm{b}}(p)+(1-p)\mathsf{C}(\mathcal{E}_0)\\
&\qquad+\alpha\left(\log\beta-\mathsf{E}\left[\log\|\mathbb{H}\mathbf{X}\|^2\,\middle|\, E=1\right]\right) && (131)\\
&\quad\le h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\, E=1\right)-h\left(\mathbb{H}\hat{\mathbf{X}}\,\middle|\,\hat{\mathbf{X}},E=1\right)\\
&\qquad+n_{\mathrm{R}}\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\,\middle|\, E=1\right]-\log 2+\log\Gamma\left(\alpha,\frac{\nu}{\beta}\right)\\
&\qquad+\frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\left[\|\mathbb{H}\hat{\mathbf{x}}\|^2\right]\frac{\mathcal{E}}{p}+\frac{\nu}{\beta}+\epsilon_\nu+H_{\mathrm{b}}(p)+(1-p)\mathsf{C}(\mathcal{E}_0)\\
&\qquad+\alpha\left(\log\beta-\log\mathcal{E}_0-\xi\right). && (132)
\end{aligned}
$$

Here, (130) follows again from a conditional version of Lemma 1 similar to (107)–(112) which allows us to combine the fourth and the last term in (129); in the subsequent equality we arithmetically rearrange the terms; and the final inequality follows from the following bound:

$$
\begin{aligned}
\mathsf{E}\left[\log\|\mathbb{H}\mathbf{X}\|^2\,\middle|\, E=1\right]&\ge\inf_{\|\mathbf{x}\|^2\ge\mathcal{E}_0}\mathsf{E}\left[\log\|\mathbb{H}\mathbf{x}\|^2\right] && (133)\\
&=\log\mathcal{E}_0+\inf_{\hat{\mathbf{x}}}\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{x}}\|^2\right] && (134)\\
&\triangleq\log\mathcal{E}_0+\xi && (135)
\end{aligned}
$$

where the last line should be taken as a definition for ξ. Notice that

$$
-\infty<\xi<\infty \qquad (136)
$$

as can be argued as follows: the lower bound on ξ follows from [1, Lem. 6.7f)], [2, Lem. A.15f)] because $h(\mathbb{H})>-\infty$ and $\mathsf{E}\left[\|\mathbb{H}\|_{\mathrm{F}}^2\right]<\infty$. The upper bound on ξ can be verified using the concavity of the logarithm function and Jensen's inequality.

Note that (132) does not depend on the distribution of R anymore, but only on $\hat{\mathbf{X}}$! Hence, we can get an upper bound on capacity by taking the supremum over all possible distributions $Q_{\hat{\mathbf{X}}}$. This then gives us the following upper bound on the fading number:

$$\chi(\mathbb{H}) = \varlimsup_{\mathcal{E}\uparrow\infty} \left\{ \mathsf{C}(\mathcal{E}) - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} \tag{137}$$

$$= \varlimsup_{\mathcal{E}\uparrow\infty} \left\{ \sup_{Q_{\mathbf{X},\mathcal{E}}} I(\mathbf{X};\mathbf{Y}) - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} \tag{138}$$

$$\begin{aligned}\le \varlimsup_{\mathcal{E}\uparrow\infty} \Bigg\{ \sup_{Q_{\hat{\mathbf{X}}}} \Bigg\{ & h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) - \log 2 \\ & + n_\mathrm{R}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] + \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right) \\ & + \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\,\frac{\mathcal{E}}{p} + \frac{\nu}{\beta} + \epsilon_\nu + H_\mathrm{b}(p) \\ & + (1-p)\mathsf{C}(\mathcal{E}_0) + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi\Big)\Bigg\} \\ & - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\Bigg\} \end{aligned} \tag{139}$$

$$\begin{aligned}= \varlimsup_{\mathcal{E}\uparrow\infty} \Bigg\{ \sup_{Q_{\hat{\mathbf{X}}}} \Bigg\{ & h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) - \log 2 \\ & + n_\mathrm{R}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big]\Bigg\} + \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right) \\ & + \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\,\frac{\mathcal{E}}{p} + \frac{\nu}{\beta} + \epsilon_\nu + H_\mathrm{b}(p) \\ & + (1-p)\mathsf{C}(\mathcal{E}_0) + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi\Big) \\ & - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\Bigg\} \end{aligned} \tag{140}$$

$$\begin{aligned}= \sup_{Q_{\hat{\mathbf{X}}}} \Bigg\{ & h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) - \log 2 \\ & + n_\mathrm{R}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big]\Bigg\} \\ + \varlimsup_{\mathcal{E}\uparrow\infty} \Bigg\{ & \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right) - \log\frac{1}{\alpha} \\ & + \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\,\frac{\mathcal{E}}{p} + \frac{\nu}{\beta} + \epsilon_\nu + H_\mathrm{b}(p) \\ & + (1-p)\mathsf{C}(\mathcal{E}_0) + \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi\Big) \\ & + \log\frac{1}{\alpha} - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\Bigg\} \end{aligned} \tag{141}$$

$$\begin{aligned}= \sup_{Q_{\hat{\mathbf{X}}}} \Bigg\{ & h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) + n_\mathrm{R}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] \\ & - \log 2\Bigg\} + \log\left(1-e^{-\nu}\right) + \nu + \epsilon_\nu - \log\nu. \end{aligned} \tag{142}$$

Here the first two equalities follows from the definition of the fading number (29); the subsequent inequality from (132); (140) follows because the parameters α, β, and ν must not depend on the input distribution $Q_{\hat{\mathbf{X}}}$ (however, note that we are allowed to let them depend on $\mathcal{E}$); the subsequent equality follows since the first four terms do not depend on $\mathcal{E}$; and in the last equality we have used (95) and made the following choices on the free parameters α and β:

$$\alpha \triangleq \alpha(\mathcal{E}) = \frac{\nu}{\log\mathcal{E} + \log\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]} \tag{143}$$

$$\beta \triangleq \beta(\mathcal{E}) = \frac{1}{\alpha(\mathcal{E})}e^{\nu/\alpha} \tag{144}$$

for some constant $\nu \ge 0$. For this choice note that

$$\varlimsup_{\mathcal{E}\uparrow\infty} \left\{ \log\Gamma\left(\alpha,\frac{\nu}{\beta}\right) - \log\frac{1}{\alpha}\right\} = \log\left(1-e^{-\nu}\right); \tag{145}$$

$$\varlimsup_{\mathcal{E}\uparrow\infty} \alpha\Big(\log\beta - \log\mathcal{E}_0 - \xi\Big) = \nu; \tag{146}$$

$$\varlimsup_{\mathcal{E}\uparrow\infty} \left\{ \frac{1}{\beta}\sup_{\hat{\mathbf{x}}}\mathsf{E}\big[\|\mathbb{H}\hat{\mathbf{x}}\|^2\big]\,\frac{\mathcal{E}}{p} + \frac{\nu}{\beta}\right\} = 0; \tag{147}$$

$$\varlimsup_{\mathcal{E}\uparrow\infty} \left\{ \log\frac{1}{\alpha} - \log\left(1+\log\left(1+\frac{\mathcal{E}}{\sigma^2}\right)\right)\right\} = -\log\nu. \tag{148}$$

(Compare with [1, App. VII], [2, Sec. B.5.9].)

To finish the derivation of the upper bound, we let ν go to zero. Note that $\epsilon_\nu \to 0$ as $\nu \downarrow 0$ as can be seen from (113). Note further that

$$\lim_{\nu\downarrow 0} \left\{ \log\left(1-e^{-\nu}\right) - \log\nu\right\} = 0. \tag{149}$$

Therefore, we get

$$\begin{aligned}\chi(\mathbb{H}) \le \sup_{Q_{\hat{\mathbf{X}}}} \Bigg\{ & h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) + n_\mathrm{R}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] \\ & - \log 2\Bigg\}. \end{aligned} \tag{150}$$

B. Derivation of a Lower Bound

To derive a lower bound on capacity (or the fading number, respectively) we choose a specific input distribution. Let $\mathbf{X}$ be of the form

$$\mathbf{X} = R\cdot\hat{\mathbf{X}}. \tag{151}$$

Here $\hat{\mathbf{X}} \in \mathbb{C}^{n_\mathrm{T}}$ is assumed to be a random unit-vector that is circularly symmetric, but whose exact distribution will be specified later. The random variable $R \in \mathbb{R}_0^+$ is chosen to be independent of $\hat{\mathbf{X}}$ and such that

$$\log R^2 \sim \mathcal{U}\left(\left[\log x_\text{min}^2, \log\mathcal{E}\right]\right) \tag{152}$$

where we choose x_min^2 as

$$x_\text{min}^2 \triangleq \log\mathcal{E}. \tag{153}$$

Note that this choice of R satisfies the peak-power constraint (26) and therefore also the average-power constraint (27).

Using such an input to our MIMO fading channel we get the following lower bound to channel capacity:

$$\begin{aligned}\mathsf{C}(\mathcal{E}) &\ge I(\mathbf{X};\mathbf{Y}) && (154)\\ &= I\big(R,\hat{\mathbf{X}};\mathbf{Y}\big) && (155)\\ &= I\big(\hat{\mathbf{X}};\mathbf{Y}\big) + I\big(R;\mathbf{Y}\,\big|\,\hat{\mathbf{X}}\big) && (156)\\ &= I\big(\hat{\mathbf{X}};\mathbf{Y}\big) + I\big(R;\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) - I\big(R;\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) \\ &\quad + I\big(R;\mathbf{Y}\,\big|\,\hat{\mathbf{X}}\big) && (157)\\ &= I\big(\hat{\mathbf{X}};\mathbf{Y}\big) + I\big(R,e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) - I\big(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}},R\big) \\ &\quad - I\big(R;\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + I\big(R;\mathbf{Y}\,\big|\,\hat{\mathbf{X}}\big). && (158)\end{aligned}$$

Here we have introduced a new random variable $\Theta \sim \mathcal{U}([0,2\pi])$ which is assumed to be independent of every other random quantity.

The last two terms can be rearranged as follows:

$$\begin{aligned}& -I\big(R;\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + I\big(R;\mathbf{Y}\,\big|\,\hat{\mathbf{X}}\big) \\ &\quad = -h\big(\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + h\big(\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}},R\big) + h\big(\mathbf{Y}\,\big|\,\hat{\mathbf{X}}\big) \\ &\qquad - h\big(\mathbf{Y}\,\big|\,\hat{\mathbf{X}},R\big) && (159)\\ &\quad = -h\big(\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + h\big(\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}},R\big) + h\big(\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}},e^{\mathrm{i}\Theta}\big) \\ &\qquad - h\big(\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}},R,e^{\mathrm{i}\Theta}\big) && (160)\\ &\quad = -I\big(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + I\big(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}},R\big). && (161)\end{aligned}$$

Here the second equality follows because $e^{\mathrm{i}\Theta}$ is independent of everything else so that we can add it to the conditioning part of the entropy without changing its values, and because differential entropy remains unchanged if its argument is multiplied by a constant complex number of magnitude 1.

Combining this with (158) we yield

$$\begin{aligned}\mathsf{C}(\mathcal{E}) &\geq I\big(\hat{\mathbf{X}};\mathbf{Y}\big) + I\big(R, e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta} \,\big|\, \hat{\mathbf{X}}\big) - I\big(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta} \,\big|\, \hat{\mathbf{X}}\big) && (162)\\ &= I\big(\hat{\mathbf{X}};\mathbf{Y}\big) + I\big(Re^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta} \,\big|\, \hat{\mathbf{X}}\big) - I\big(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta} \,\big|\, \hat{\mathbf{X}}\big) && (163)\end{aligned}$$

where the last equality follows because from $Re^{\mathrm{i}\Theta}$ the random variables R and $e^{\mathrm{i}\Theta}$ can be gained back.

We continue with bounding the first term in (163):

$$\begin{aligned}I\big(\hat{\mathbf{X}};\mathbf{Y}\big) &= I\big(\hat{\mathbf{X}};\mathbf{Y},\mathbf{Z}\big) - \underbrace{I\big(\hat{\mathbf{X}};\mathbf{Z}\,\big|\,\mathbf{Y}\big)}_{\leq \epsilon(x_{\text{min}})} && (164)\\ &\geq I\big(\hat{\mathbf{X}};\mathbf{Y},\mathbf{Z}\big) - \epsilon(x_{\text{min}}) && (165)\\ &= I\big(\hat{\mathbf{X}};\mathbb{H}\hat{\mathbf{X}}R\big) - \epsilon(x_{\text{min}}) && (166)\\ &= I\left(\hat{\mathbf{X}};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}, \|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\right) - \epsilon(x_{\text{min}}) && (167)\\ &= I\left(\hat{\mathbf{X}};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + I\left(\hat{\mathbf{X}};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R \,\middle|\, \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right)\\ &\quad - \epsilon(x_{\text{min}}). && (168)\end{aligned}$$

Here the first equality follows from the chain rule; in the subsequent inequality we lower-bound the second term by $-\epsilon(x_{\text{min}})$ which is defined in Appendix E and is shown there to be independent of the input distribution $Q_{\mathbf{X}}$ and to tend to zero as $x_{\text{min}} \uparrow \infty$; in the subsequent equality we use $\mathbf{Z}$ in order to extract $\mathbb{H}\hat{\mathbf{X}}R$ from $\mathbf{Y}$ and then drop $(\mathbf{Y},\mathbf{Z})$ since given $\mathbb{H}\hat{\mathbf{X}}R$ it is independent of the other random variables; and the last equality follows again from the chain rule.

Similarly, we bound the third term in (163):

$$\begin{aligned}&I\big(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big)\\ &\quad\leq I\big(e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta},\mathbf{Z}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) && (169)\\ &\quad= I\big(e^{\mathrm{i}\Theta};\mathbb{H}\mathbf{X}e^{\mathrm{i}\Theta},\mathbf{Z}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) && (170)\\ &\quad= I\big(e^{\mathrm{i}\Theta};\mathbb{H}\mathbf{X}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + I\big(e^{\mathrm{i}\Theta};\mathbf{Z}e^{\mathrm{i}\Theta}\,\big|\,\mathbb{H}\mathbf{X}e^{\mathrm{i}\Theta},\hat{\mathbf{X}}\big) && (171)\\ &\quad= I\big(e^{\mathrm{i}\Theta};\mathbb{H}\mathbf{X}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) && (172)\\ &\quad= I\left(e^{\mathrm{i}\Theta};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R, \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right) && (173)\\ &\quad= I\left(e^{\mathrm{i}\Theta};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right)\\ &\qquad + I\left(e^{\mathrm{i}\Theta};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta},\hat{\mathbf{X}}\right). && (174)\end{aligned}$$

Hence, plugging these results into (163) we get:

$$\begin{aligned}\mathsf{C}(\mathcal{E}) &\geq I\big(Re^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + I\left(\hat{\mathbf{X}};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right)\\ &\quad + I\left(\hat{\mathbf{X}};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right)\\ &\quad - I\left(e^{\mathrm{i}\Theta};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right)\\ &\quad - I\left(e^{\mathrm{i}\Theta};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta},\hat{\mathbf{X}}\right) - \epsilon(x_{\text{min}}). && (175)\end{aligned}$$

We next bound the third and fifth mutual information term in (175):

$$\begin{aligned}&I\left(\hat{\mathbf{X}};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - I\left(e^{\mathrm{i}\Theta};\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}e^{\mathrm{i}\Theta}}{\|\mathbb{H}\hat{\mathbf{X}}\|},\hat{\mathbf{X}}\right)\\ &\quad= h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|},\hat{\mathbf{X}}\right)\\ &\qquad - h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta},\hat{\mathbf{X}}\right)\\ &\qquad + h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta},\hat{\mathbf{X}},e^{\mathrm{i}\Theta}\right) && (176)\\ &\quad= h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|},\hat{\mathbf{X}}\right)\\ &\qquad - h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta},\hat{\mathbf{X}}\right)\\ &\qquad + h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|},\hat{\mathbf{X}}\right) && (177)\\ &\quad= h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta},\hat{\mathbf{X}}\right) && (178)\\ &\quad\geq h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}e^{\mathrm{i}\Theta}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) && (179)\\ &\quad= h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h\left(\|\mathbb{H}\hat{\mathbf{X}}\|\cdot R\,\middle|\,\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) && (180)\\ &\quad= 0. && (181)\end{aligned}$$

Here, the inequality follows from conditioning that reduces entropy; and the second last equality holds because we have assumed $\hat{\mathbf{X}}$ to be circularly symmetric, *i.e.*, $\hat{\mathbf{X}}$ "destroys" the random phase shift of $e^{\mathrm{i}\Theta}$.

Therefore, we are left with the following bound:

$$\begin{aligned}\mathsf{C}(\mathcal{E}) &\geq I\big(Re^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta}\,\big|\,\hat{\mathbf{X}}\big) + I\left(\hat{\mathbf{X}};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right)\\ &\quad - I\left(e^{\mathrm{i}\Theta};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right) - \epsilon(x_{\text{min}}). && (182)\end{aligned}$$

Now, we rewrite the second and third term as follows:

$$\begin{aligned}&I\left(\hat{\mathbf{X}};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - I\left(e^{\mathrm{i}\Theta};\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right)\\ &\quad= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right)\\ &\qquad + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}},e^{\mathrm{i}\Theta}\right) && (183)\\ &\quad= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right)\\ &\qquad + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\,\middle|\,\hat{\mathbf{X}}\right) && (184)\\ &\quad= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta}\,\middle|\,\hat{\mathbf{X}}\right) && (185)\end{aligned}$$

where the second equality follows from (13) with a choice $\mathsf{U} = e^{-\mathrm{i}\theta}\mathsf{I}_{n_\mathrm{R}}$ and from the fact that $e^{\mathrm{i}\Theta}$ is independent of all other random quantities.

This leaves us with

$$C(\mathcal{E}) \geq I\big(Re^{\mathrm{i}\Theta}; \mathbf{Y}e^{\mathrm{i}\Theta} \,\big|\, \hat{\mathbf{X}}\big) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta} \,\middle|\, \hat{\mathbf{X}}\right) - \epsilon(x_{\min}). \tag{186}$$

We next let the power grow to infinity $\mathcal{E} \to \infty$ and use the definition of the fading number (29). Since $Re^{\mathrm{i}\Theta}$ is circularly symmetric with a magnitude distributed according to (152), we know from [1, (108) and Th. 4.8], [2, (6.194) and Th. 6.15], that $Re^{\mathrm{i}\Theta}$ achieves the fading number of a memoryless SIMO fading channel with partial side-information. In our situation we have

$$I\big(Re^{\mathrm{i}\Theta}; \mathbf{Y}e^{\mathrm{i}\Theta} \,\big|\, \hat{\mathbf{X}}\big) = I\big(Re^{\mathrm{i}\Theta}; \mathbb{H}\hat{\mathbf{X}}Re^{\mathrm{i}\Theta} + \mathbf{Z} \,\big|\, \hat{\mathbf{X}}\big) \tag{187}$$

$$= I\big(Re^{\mathrm{i}\Theta}; \mathbb{H}\hat{\mathbf{X}}Re^{\mathrm{i}\Theta} + \mathbf{Z}, \hat{\mathbf{X}}\big) \tag{188}$$

where $\hat{\mathbf{X}}$ serves as partial receiver side-information (that is independent of the SIMO input $Re^{\mathrm{i}\Theta}$). Note that a random vector $\mathbf{A}$ is said to contain only *partial* side-information about $\mathbf{B}$ if $h(\mathbf{B}|\mathbf{A}) > -\infty$, *i.e.*, in our case we need

$$h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) > -\infty \tag{189}$$

which is satisfied since we assume that $h(\mathbb{H}) > -\infty$ and $\mathsf{E}\big[\|\mathbb{H}\|_{\mathrm{F}}^2\big] < \infty$ (see [1, Lem. 6.6], [2, Lem. A.14]).

Hence,

$$\chi(\mathbb{H}) \geq \varlimsup_{\mathcal{E}\uparrow\infty} \Bigg\{ I\big(Re^{\mathrm{i}\Theta}; \mathbb{H}\hat{\mathbf{X}}Re^{\mathrm{i}\Theta} + \mathbf{Z} \,\big|\, \hat{\mathbf{X}}\big) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta} \,\middle|\, \hat{\mathbf{X}}\right) - \epsilon(x_{\min}) - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right) \Bigg\} \tag{190}$$

$$= \varlimsup_{\mathcal{E}\uparrow\infty} \Bigg\{ I\big(Re^{\mathrm{i}\Theta}; \mathbb{H}\hat{\mathbf{X}}Re^{\mathrm{i}\Theta} + \mathbf{Z} \,\big|\, \hat{\mathbf{X}}\big) - \log\left(1 + \log\left(1 + \frac{\mathcal{E}}{\sigma^2}\right)\right) - \epsilon(x_{\min}) \Bigg\} + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta} \,\middle|\, \hat{\mathbf{X}}\right) \tag{191}$$

$$= \chi\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta} \,\middle|\, \hat{\mathbf{X}}\right) \tag{192}$$

$$= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta} \,\middle|\, \hat{\mathbf{X}}\right) + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) - h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}e^{\mathrm{i}\Theta} \,\middle|\, \hat{\mathbf{X}}\right) \tag{193}$$

$$= h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + n_{\mathrm{R}}\mathsf{E}\Big[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\Big] - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big). \tag{194}$$

Here in (192) we have used the fact that our choice (153) guarantees that $\epsilon(x_{\min})$ tends to zero as $\mathcal{E} \to \infty$ (see Appendix E) and that we achieve the SIMO fading number for a channel with input $Re^{\mathrm{i}\Theta}$ and output $\mathbb{H}\hat{\mathbf{x}}Re^{\mathrm{i}\Theta} + \mathbf{Z}$; the subsequent equality follows from the fading number of a memoryless SIMO fading channel where the receiver has access to some partial side-information [1, (108)], [2, (6.194)]:

$$\chi(\mathbf{H}|\mathbf{S}) = h_\lambda\big(\hat{\mathbf{H}}e^{\mathrm{i}\Theta} \,\big|\, \mathbf{S}\big) + n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{H}\|^2\big] - \log 2 - h(\mathbf{H}|\mathbf{S}). \tag{195}$$

The result now follows by choosing the distribution $Q_{\hat{\mathbf{X}}}$ such as to maximize the lower bound (194) to the fading number.

VIII. Conclusions

We have derived the fading number of a MIMO fading channel of general fading law including spatial, but without temporal memory. Since the fading number is the second term after the double-logarithmic term of the high-SNR expansion of channel capacity, this means that we have precisely specified the behavior of the channel capacity asymptotically when the power grows to infinity. The result shows that the asymptotic capacity can be achieved by an input that consists of the product of two independent random quantities: a circularly symmetric random unit vector (the *direction*) and a non-negative (*i.e.*, real) random variable (the *magnitude*). The distribution of the random direction is chosen such as to maximize the fading number and therefore depends on the particular law of the fading process. The non-negative random variable is such that (38) is satisfied. This is the well-known choice that also achieves the fading number in the SISO and SIMO case and is also used in the MISO case where it is multiplied by a constant beam-direction $\hat{\mathbf{x}}$. All these special cases follow nicely from this new result.

We have then derived some new results for the important special situation of Gaussian fading. For the case of a scalar line-of-sight matrix (68) assuming at least as many transmit as receive antennas $n_{\mathrm{R}} \leq n_{\mathrm{T}}$ we have been able to state the fading number precisely:

$$\chi = n_{\mathrm{R}}g_{n_{\mathrm{R}}}\big(|d|^2\big) - n_{\mathrm{R}} - \log\Gamma(n_{\mathrm{R}}) \tag{196}$$

where $g_m(\cdot)$ denotes the expected value of a non-central chi-square random variable (see (61)). We see that the asymptotic capacity only depends on the number of receive antennas and is growing proportionally to $n_{\mathrm{R}}\log|d|^2$.

For a general line-of-sight matrix we have shown an upper bound that grows like $\min\{n_{\mathrm{R}}, n_{\mathrm{T}}\}\log\delta^2$ where δ^2 is a certain kind of average of all singular values of the line-of-sight matrix (see (79) and (80)).

We would like to emphasize that even though all results on the fading number are asymptotic results for the theoretical situation of infinite power, they are still of relevance for finite SNR values: it has been shown that the approximation

$$C(\textsc{snr}) \approx \log(1 + \log(1 + \textsc{snr})) + \chi \tag{197}$$

holds already for moderate values of the SNR. Actually, pulling ourselves by our bootstraps, let us consider for the moment that (197) starts to be valid for an SNR somewhere in the range of 30 to 80 dB. In this case $\log(1 + \log(1 + \textsc{snr}))$ will have a value between 2 and 3 nats. Hence, once the capacity is appreciably above $\chi + 2$ nats, the approximation (197) is likely to be valid [10], [11].

Therefore, the fading number can be seen as an indicator of the maximal rate at which power efficient communication is possible on the channel. For a further discussion about the practical relevance of the fading number we refer to [10] and [12].

Appendix A
Proof of Lemma 5

Assume that $\Theta \sim \mathcal{U}([0, 2\pi])$, independent of every other random quantity. Then

$$I(\mathbf{X};\mathbf{Y}) = I\big(\mathbf{X};\mathbf{Y} \,\big|\, e^{\mathrm{i}\Theta}\big) \tag{198}$$

$$= I\big(\mathbf{X}e^{\mathrm{i}\Theta};\mathbf{Y}e^{\mathrm{i}\Theta} \,\big|\, e^{\mathrm{i}\Theta}\big) \tag{199}$$

$$= I\big(\mathbf{X}e^{\mathrm{i}\Theta}; \mathbb{H}\mathbf{X}e^{\mathrm{i}\Theta} + \mathbf{Z} \,\big|\, e^{\mathrm{i}\Theta}\big) \tag{200}$$
$$= I\big(\tilde{\mathbf{X}}; \mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z} \,\big|\, e^{\mathrm{i}\Theta}\big) \tag{201}$$
$$= h\big(\mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z} \,\big|\, e^{\mathrm{i}\Theta}\big) - h\big(\mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z} \,\big|\, \tilde{\mathbf{X}}, e^{\mathrm{i}\Theta}\big) \tag{202}$$
$$= h\big(\mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z} \,\big|\, e^{\mathrm{i}\Theta}\big) - h\big(\mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z} \,\big|\, \tilde{\mathbf{X}}\big) \tag{203}$$
$$\leq h\big(\mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z}\big) - h\big(\mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z} \,\big|\, \tilde{\mathbf{X}}\big) \tag{204}$$
$$= I\big(\tilde{\mathbf{X}}; \mathbb{H}\tilde{\mathbf{X}} + \mathbf{Z}\big). \tag{205}$$

Here the first equality follows because Θ is independent of every other random quantity; the third equality follows because $\mathbf{Z}$ is circularly symmetric; in the subsequent equality we substitute $\tilde{\mathbf{X}} = \mathbf{X}e^{\mathrm{i}\Theta}$; and the inequality follows since conditioning reduces entropy.

Hence, a circularly symmetric input achieves a mutual information that is at least as big as the original mutual information.

Appendix B
Derivation of Bounds (67)

In this appendix we will derive the bounds (67) on $g_m(\cdot)$. We start with the upper bound which follows directly from (64) and (65) and from Jensen's inequality:

$$g_m(s^2) = \mathsf{E}\left[\log\left(\sum_{j=1}^{m} \left|U_j + \mu_j\right|^2\right)\right] \tag{206}$$
$$\leq \log\left(\sum_{j=1}^{m} \mathsf{E}\left[\left|U_j + \mu_j\right|^2\right]\right) \tag{207}$$
$$= \log\left(\sum_{j=1}^{m} \left(1 + |\mu_j|^2\right)\right) \tag{208}$$
$$= \log(m + s^2). \tag{209}$$

For the lower bound we also start with (64) and choose $\mu_1 = s$ and $\mu_2 = \cdots = \mu_m = 0$. Then we get

$$g_m(s^2) = \mathsf{E}\left[\log\left(\sum_{j=1}^{m} \left|U_j + \mu_j\right|^2\right)\right] \tag{210}$$
$$\geq \mathsf{E}\left[\log\left(\left|U_1 + \mu_1\right|^2\right)\right] \tag{211}$$
$$= g_1(s^2) \tag{212}$$
$$= \log s^2 - \mathrm{Ei}\big(-s^2\big). \tag{213}$$

Here, (211) follows from dropping some non-negative terms in the sum; and in the subsequent two equalities we use the definition of $g_1(\cdot)$.

Appendix C
Proof of Corollary 12

We choose a constant $n_{\mathrm{T}} \times n_{\mathrm{T}}$ matrix B as follows

$$\mathsf{B} \triangleq \mathrm{diag}\left(\frac{1}{d_1}, \ldots, \frac{1}{d_{n_\mathrm{R}}}, \frac{1}{d_1}, \ldots, \frac{1}{d_1}\right) \tag{214}$$

and then we note that for a unit vector $\hat{\mathbf{x}} = \left(\hat{x}^{(1)}, \ldots, \hat{x}^{(n_\mathrm{T})}\right)^{\mathsf{T}}$

$$\mathbb{H}\mathsf{B}\hat{\mathbf{x}} = \mathsf{D}\mathsf{B}\hat{\mathbf{x}} + \tilde{\mathbb{H}}\mathsf{B}\hat{\mathbf{x}} = \begin{pmatrix} \hat{x}^{(1)} \\ \vdots \\ \hat{x}^{(n_\mathrm{R})} \end{pmatrix} + \tilde{\mathbb{H}}\mathsf{B}\hat{\mathbf{x}} \triangleq \boldsymbol{\xi} + \tilde{\mathbf{H}} \tag{215}$$

where $\tilde{\mathbf{H}} \sim \mathcal{N}_{\mathbb{C}}\big(\mathbf{0}, \zeta^2(\hat{\mathbf{x}})\mathsf{I}_{n_\mathrm{R}}\big)$ with

$$\zeta^2(\hat{\mathbf{x}}) \triangleq \frac{|\hat{x}^{(1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{x}^{(n_\mathrm{R})}|^2}{|d_{n_\mathrm{R}}|^2} + \frac{|\hat{x}^{(n_\mathrm{R}+1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{x}^{(n_\mathrm{T})}|^2}{|d_1|^2} \tag{216}$$

and where $\boldsymbol{\xi} \in \mathbb{C}^{n_\mathrm{R}}$ with $\|\boldsymbol{\xi}\| \leq 1$. Therefore,

$$h\big(\mathbb{H}\mathsf{B}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}} = \hat{\mathbf{x}}\big) = n_\mathrm{R} \log \pi e \zeta^2(\hat{\mathbf{x}}); \tag{217}$$
$$\mathsf{E}\left[\log\|\mathbb{H}\mathsf{B}\hat{\mathbf{x}}\|^2\right] = \log\zeta^2(\hat{\mathbf{x}}) + g_{n_\mathrm{R}}\left(\frac{\|\boldsymbol{\xi}\|^2}{\zeta^2(\hat{\mathbf{x}})}\right) \tag{218}$$

(where the last equality follows from (64)); and hence

$$n_\mathrm{R}\mathsf{E}\left[\log\|\mathbb{H}\mathsf{B}\hat{\mathbf{X}}\|^2\right] - h\big(\mathbb{H}\mathsf{B}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big)$$
$$= n_\mathrm{R}\mathsf{E}\left[g_{n_\mathrm{R}}\left(\frac{|\hat{X}^{(1)}|^2 + \cdots + |\hat{X}^{(n_\mathrm{R})}|^2}{\zeta^2(\hat{\mathbf{X}})}\right)\right] - n_\mathrm{R}\log\pi e. \tag{219}$$

The upper bound on the fading number now follows from (39); from Theorem 7 by upper-bounding the h_λ-term by $\log c_{n_\mathrm{R}}$; and from the additional observations that $g_m(\cdot)$ is a monotonically increasing function, that

$$|\hat{X}^{(1)}|^2 + \cdots + |\hat{X}^{(n_\mathrm{R})}|^2 \leq 1 \tag{220}$$

and that

$$\zeta^2(\hat{\mathbf{X}}) = \frac{|\hat{X}^{(1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{X}^{(n_\mathrm{R})}|^2}{|d_{n_\mathrm{R}}|^2} + \frac{|\hat{X}^{(n_\mathrm{R}+1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{X}^{(n_\mathrm{T})}|^2}{|d_1|^2} \tag{221}$$
$$\geq \frac{|\hat{X}^{(1)}|^2}{|d_1|^2} + \cdots + \frac{|\hat{X}^{(n_\mathrm{T})}|^2}{|d_1|^2} \tag{222}$$
$$= \frac{1}{|d_1|^2}\left(|\hat{X}^{(1)}|^2 + \cdots + |\hat{X}^{(n_\mathrm{T})}|^2\right) \tag{223}$$
$$= \frac{1}{|d_1|^2} = \frac{1}{\|\mathsf{D}\|^2} \tag{224}$$

where the inequality follows since $|d_1| \geq |d_2| \geq \ldots \geq |d_{n_\mathrm{R}}|$.

Appendix D
Proof of Proposition 13

This upper bound is based on the upper bound given in Corollary 8 for a choice of $\mathsf{B} = \mathsf{I}_{n_\mathrm{T}}$. If $n_\mathrm{R} > n_\mathrm{T}$ we choose for A

$$\mathsf{A} \triangleq \mathrm{diag}\left(\frac{a}{d_1}, \ldots, \frac{a}{d_{n_\mathrm{T}}}, b, \ldots, b\right) \tag{225}$$

with

$$b \triangleq \left(\frac{\delta^2}{n_\mathrm{T}}\right)^{\frac{n_\mathrm{T}}{2n_\mathrm{R}}} \tag{226}$$

for δ as given in (80), and with a such that $\det\mathsf{A} = 1$, *i.e.*,

$$a \triangleq \left(d_1 \cdot \cdots \cdot d_{n_\mathrm{T}}\right)^{\frac{1}{n_\mathrm{T}}} \cdot b^{\frac{n_\mathrm{T} - n_\mathrm{R}}{n_\mathrm{T}}}. \tag{227}$$

For such a choice we note that

$$\mathsf{A}\mathbb{H}\hat{\mathbf{x}} = a\begin{pmatrix}\hat{\mathbf{x}} \\ \mathbf{0}\end{pmatrix} + \left(\mathcal{N}_{\mathbb{C}}\left(0, \frac{|a|^2}{|d_1|^2}\right), \ldots, \mathcal{N}_{\mathbb{C}}\left(0, \frac{|a|^2}{|d_{n_\mathrm{T}}|^2}\right), \mathcal{N}_{\mathbb{C}}\big(0, b^2\big), \ldots, \mathcal{N}_{\mathbb{C}}\big(0, b^2\big)\right)^{\mathsf{T}} \tag{228}$$

so that

$$\mathsf{E}\left[\|\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\|^2\right] = \delta^2\left(b^2\right)^{\frac{n_\mathrm{T} - n_\mathrm{R}}{n_\mathrm{T}}} + (n_\mathrm{R} - n_\mathrm{T})b^2 \tag{229}$$
$$= n_\mathrm{R}\left(\frac{\delta^2}{n_\mathrm{T}}\right)^{\frac{n_\mathrm{T}}{n_\mathrm{R}}}. \tag{230}$$

Hence, using Jensen's inequality and the fact that $\det\mathsf{A} = 1$ we get

$$n_\mathrm{R}\mathsf{E}\left[\log\|\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\|^2\right] - h\big(\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\big)$$
$$\leq n_\mathrm{R}\log\mathsf{E}\left[\|\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\|^2\right] - \log\det\mathsf{A} - h\big(\mathbb{H}\hat{\mathbf{x}}\big) \tag{231}$$
$$= n_\mathrm{R}\log\left(n_\mathrm{R}\left(\frac{\delta^2}{n_\mathrm{T}}\right)^{n_\mathrm{T}/n_\mathrm{R}}\right) - n_\mathrm{R}\log\pi e. \tag{232}$$

Plugging this into the upper bound (41) of Corollary 8 we yield

$$\chi \leq n_{\mathrm{R}} \log \pi - \log \Gamma(n_{\mathrm{R}}) + n_{\mathrm{R}} \log n_{\mathrm{R}} + n_{\mathrm{T}} \log \left(\frac{\delta^2}{n_{\mathrm{T}}} \right) - n_{\mathrm{R}} \log \pi e \tag{233}$$

$$= n_{\mathrm{T}} \log \left(\frac{\delta^2}{n_{\mathrm{T}}} \right) + n_{\mathrm{R}} \log n_{\mathrm{R}} - \log \Gamma(n_{\mathrm{R}}) - n_{\mathrm{R}}. \tag{234}$$

If $n_{\mathrm{R}} \leq n_{\mathrm{T}}$ we choose for A

$$\mathsf{A} = \operatorname{diag} \left(\frac{a}{d_1}, \ldots, \frac{a}{d_{n_{\mathrm{R}}}} \right) \tag{235}$$

with a such that $\det \mathsf{A} = 1$, *i.e.*,

$$a \triangleq \left(d_1 \cdot \cdots \cdot d_{n_{\mathrm{R}}} \right)^{\frac{1}{n_{\mathrm{R}}}}. \tag{236}$$

For such a choice we note that

$$\mathsf{A}\mathbb{H}\hat{\mathbf{x}} = a \left(\hat{x}^{(1)}, \ldots, \hat{x}^{(n_{\mathrm{R}})} \right)^{\mathsf{T}} + \left(\mathcal{N}_{\mathbb{C}}\left(0, \frac{|a|^2}{|d_1|^2} \right), \ldots, \mathcal{N}_{\mathbb{C}}\left(0, \frac{|a|^2}{|d_{n_{\mathrm{R}}}|^2} \right) \right)^{\mathsf{T}} \tag{237}$$

so that

$$\mathsf{E}\left[\|\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\|^2 \right] = |a|^2 \left(\left| \hat{x}^{(1)} \right|^2 + \cdots + \left| \hat{x}^{(n_{\mathrm{R}})} \right|^2 \right) + \frac{|a|^2}{|d_1|^2} + \cdots + \frac{|a|^2}{|d_{n_{\mathrm{R}}}|^2} \tag{238}$$

$$\leq \delta^2 \tag{239}$$

where we have bounded $|\hat{x}^{(1)}|^2 + \cdots + |\hat{x}^{(n_{\mathrm{R}})}|^2 \leq 1$. Hence, using Jensen's inequality and the fact that $\det \mathsf{A} = 1$ we get

$$n_{\mathrm{R}} \mathsf{E}\left[\log \|\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\|^2 \right] - h\big(\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\big) \leq n_{\mathrm{R}} \log \mathsf{E}\left[\|\mathsf{A}\mathbb{H}\hat{\mathbf{x}}\|^2 \right] - \log \det \mathsf{A} - h\big(\mathbb{H}\hat{\mathbf{x}}\big) \tag{240}$$

$$\leq n_{\mathrm{R}} \log \delta^2 - n_{\mathrm{R}} \log \pi e. \tag{241}$$

Plugging this into the upper bound (41) of Corollary 8 we yield

$$\chi \leq n_{\mathrm{R}} \log \pi - \log \Gamma(n_{\mathrm{R}}) + n_{\mathrm{R}} \log \delta^2 - n_{\mathrm{R}} \log \pi e \tag{242}$$

$$= n_{\mathrm{R}} \log \frac{\delta^2}{n_{\mathrm{R}}} + n_{\mathrm{R}} \log n_{\mathrm{R}} - \log \Gamma(n_{\mathrm{R}}) - n_{\mathrm{R}}. \tag{243}$$

The result now follows by combining (234) and (243).

APPENDIX E
ADDITIONAL DERIVATION FOR THE PROOF OF THE LOWER BOUND

In the derivation of the lower bound to the fading number we need to find the following upper bound

$$I\big(\hat{\mathbf{X}}; \mathbf{Z} \,\big|\, \mathbf{Y}\big) \leq \epsilon(x_{\min}) \tag{244}$$

and to show that $\epsilon(x_{\min})$ does not depend on the input distribution $Q_{\mathbf{X}}$ and tends to zero as $x_{\min}$ tends to infinity.

Such a bound can be found as follows:

$$I\big(\hat{\mathbf{X}}; \mathbf{Z} \,\big|\, \mathbf{Y}\big) = h(\mathbf{Z}|\mathbf{Y}) - h\big(\mathbf{Z} \,\big|\, \mathbf{Y}, \hat{\mathbf{X}}\big) \tag{245}$$

$$\leq h(\mathbf{Z}) - h\big(\mathbf{Z} \,\big|\, \mathbf{Y}, \hat{\mathbf{X}}, R\big) \tag{246}$$

$$= h(\mathbf{Z}) - h\big(\mathbf{Z} \,\big|\, \mathbb{H}\hat{\mathbf{X}}R + \mathbf{Z}, \hat{\mathbf{X}}, R\big) \tag{247}$$

$$\leq h(\mathbf{Z}) - \inf_{\hat{\mathbf{x}}} \inf_{r \geq x_{\min}} h\big(\mathbf{Z} \,\big|\, \mathbb{H}\hat{\mathbf{x}}r + \mathbf{Z}\big) \tag{248}$$

$$= h(\mathbf{Z}) - \inf_{\hat{\mathbf{x}}} h\big(\mathbf{Z} \,\big|\, \mathbb{H}\hat{\mathbf{x}}x_{\min} + \mathbf{Z}\big) \tag{249}$$

$$= \sup_{\hat{\mathbf{x}}} I\big(\mathbf{Z}; \mathbb{H}\hat{\mathbf{x}}x_{\min} + \mathbf{Z}\big) \tag{250}$$

$$= \sup_{\hat{\mathbf{x}}} I\left(\frac{\mathbf{Z}}{x_{\min}}; \mathbb{H}\hat{\mathbf{x}} + \frac{\mathbf{Z}}{x_{\min}} \right) \tag{251}$$

$$= \sup_{\hat{\mathbf{x}}} \left\{ h\left(\mathbb{H}\hat{\mathbf{x}} + \frac{\mathbf{Z}}{x_{\min}} \right) - h\big(\mathbb{H}\hat{\mathbf{x}}\big) \right\} \tag{252}$$

$$\triangleq \epsilon(x_{\min}) \tag{253}$$

where we have used the fact that we have chosen R such that $R \geq x_{\min}$. Note that (252) does not depend on the input $\mathbf{X}$ anymore. The convergence

$$\lim_{x_{\min} \uparrow \infty} \epsilon(x_{\min}) = 0 \tag{254}$$

follows from [1, Lem. 6.11], [2, Lem. A.19].

ACKNOWLEDGMENT

This work was supported by the Industrial Technology Research Institute (ITRI), Zhudong, Taiwan, under contract G1-95003.

The author would like to thank Amos Lapidoth for long, fruitful discussions and Tobias Koch for giving the right hint to fix a serious bug in an earlier version of the proof. Moreover, he thanks the anonymous reviewer for his detailed and very helpful comments.

REFERENCES

[1] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.

[2] S. M. Moser, "Duality-based bounds on channel capacity," Ph.D. dissertation, Swiss Fed. Inst. of Techn., Zurich, Oct. 2004, Diss. ETH No. 15769. [Online]. Available: http://moser.cm.nctu.edu.tw/

[3] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Europ. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov.–Dec. 1999.

[4] A. Lapidoth, "On the high SNR capacity of stationary Gaussian fading channels," in *Proc. 41st Allerton Conf. Comm., Contr. and Comp.*, Allerton H., Monticello, Il, Oct. 1–3, 2003, pp. 410–419.

[5] T. Koch, "On the asymptotic capacity of multiple-input single-output fading channels with memory," Master's thesis, Signal and Inform. Proc. Lab., ETH Zurich, Switzerland, Apr. 2004, supervised by Prof. Dr. Amos Lapidoth.

[6] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.

[7] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective Rayleigh-fading channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 12, pp. 3095–3110, Dec. 2004.

[8] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 437–453, Feb. 2006.

[9] ——, "The expected logarithm of a non-central chi-square random variable," website. [Online]. Available: http://moser.cm.nctu.edu.tw/explog.html

[10] T. Koch and A. Lapidoth, "The fading number and degrees of freedom in non-coherent MIMO fading channels: a peace pipe," in *Proc. IEEE Int. Symp. Inf. Theory*, Adelaide, Australia, Sept. 4–9, 2005, pp. 661–665.

[11] A. Lapidoth, "The fading number and degrees of freedom: A peace pipe," talk, Shushan Purim, Israel, Mar. 27, 2005.

[12] T. Koch and A. Lapidoth, "Degrees of freedom in non-coherent stationary MIMO fading channels," in *Proc. Winter School Cod. and Inform. Theory*, Bratislava, Slovakia, Feb. 20–25, 2005, pp. 91–97.

Stefan M. Moser (S'01, M'05) was born in Switzerland. He received the M.Sc. degree in electrical engineering (with distinction) in 1999, the M.Sc. degree in industrial management (MBA) in 2003, and the Ph.D. degree in the field of information theory in 2004, all from the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland.

In the years 1999 to 2003, Dr. Moser was a Research and Teaching Assistant and from 2004 to 2005 he was a Senior Research Assistant at the Signal and Information Processing Laboratory at ETH in Zurich. Since August 2005, he has been an Assistant Professor at the Department of Communication Engineering at National Chiao Tung University (NCTU) in Hsinchu, Taiwan. His research interests are in digital communications and information theory.

Dr. Moser received the National Chiao Tung University Outstanding Researchers Award in 2007, the Willi Studer Award of ETH in 1999, and the ETH Silver Medal for an excellent master thesis in 1999.