The Fading Number of IID MIMO Gaussian Fading Channels with a Scalar Line-of-Sight Component

Stefan M. Moser

***Abstract*— The capacity of regular noncoherent fading channels grows like $\log\log \mathsf{SNR} + \chi$ at high signal-to-noise ratios (SNR). Here, χ, denoted *fading number*, is a constant independent of the SNR, but dependent on the distribution of the fading process. Recently, an expression of the fading number has been derived for the situation of general memoryless multiple-input multiple-output (MIMO) fading channels. In this paper, this expression is evaluated in the special situation of an independent and identically distributed MIMO Gaussian fading channel with a scalar line-of-sight component d. It is shown that, for large $|d|$, the fading number grows like $\min\{n_{\mathrm{R}}, n_{\mathrm{T}}\} \log |d|^2$ where n_{R} and n_{T} denote the number of antennas at the receiver and transmitter, respectively.**

As a side-product along the way, closed-form expressions are derived for the expectation of the logarithm and for the expectation of the n-th power of the reciprocal value of a noncentral chi-square random variable. It is shown that these expectations can be expressed by a family of continuous functions $g_m(\cdot)$ and that these families have nice properties (monotonicity, concavity, *etc.*). Moreover, some tight upper and lower bounds are derived that are helpful in situations where the closed-form expression of $g_m(\cdot)$ is too complex for further analysis.

I. Introduction

In this paper we study noncoherent communication over a flat fading channel. The term *noncoherent* refers to the situation where neither transmitter nor receiver have *a priori* knowledge of the fading realization (or channel state). It is assumed that only the probability law of the fading process is known. We further assume that the process is *regular* which means that the differential entropy rate of the process is finite. In an engineering explanation this means that the process is "random enough" such that even with complete knowledge of the past fading realizations, the actual fading value cannot be predicted error-free, but there will be always (a possibly very small) nonzero prediction error.

It has been known for some time that the capacity of such noncoherent fading channels typically grows only double-logarithmically in the signal-to-noise ratio (SNR) at high SNR [1], [2]. Such a growth rate is extremely power-inefficient: increasing capacity by only one bit requires the SNR to be *squared* or, on a dB-scale, to be doubled! Hence, every communication system should be designed such that it does not operate in this regime.

To quantify the rates at which this poor power efficiency begins, [1], [2] introduce the *fading number* χ as the second term in the high-SNR asymptotic expansion of channel capacity. In detail this means that the capacity can be written as

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

where $o(1)$ tends to zero as the SNR tends to infinity. Note that the fading number χ is a constant, *i.e.*, it does not depend on the SNR, however, it does strongly depend on the particular probability distribution of the fading process. In particular, χ depends on channel parameters like line-of-sight component, number of antennas, etc.

Even though the fading number is defined for the (unrealistic) case of the SNR tending to infinity, it is relevant for practical purposes: it is an accurate indicator of the threshold SNR_0 between the power-efficient low-SNR regime where capacity grows logarithmically in the SNR and the highly inefficient high-SNR regime with double-logarithmic growth. To see this connection between χ and SNR_0 first note that a communication system is operating in the inefficient high-SNR region if and only if the approximation

$$C(\mathsf{SNR}) \approx \log(1 + \log(1 + \mathsf{SNR})) + \chi \tag{2}$$

is a good approximation (*i.e.*, the $o(1)$ terms in (1) have become negligible). In (2) χ is usually much larger than $\log\log \mathsf{SNR}$ unless the SNR is at extremely high values (larger than $e^{e^{\chi}}$!). Hence, for SNR values in the range of the threshold SNR_0, the fading number χ will dominate the $\log\log$-term so that the capacity around the threshold can be approximated by

$$C(\mathsf{SNR}_0) \approx \chi. \tag{3}$$

In other words, if $C(\mathsf{SNR})$ is significantly larger than χ, the SNR must be larger than SNR_0, *i.e.*, the system is in the power-inefficient $\log\log$-regime.

Recently, the fading number has been derived for general multiple-input multiple-output (MIMO) fading channels without temporal memory [3]. Here we use the term "general" to describe the fact that no specific distribution has been assumed for the fading process.[1] The given expression in [3] is therefore very general. On the other hand, the expression is rather difficult to evaluate and is therefore partially hiding some insight. In this paper we would like to evaluate this memoryless MIMO fading number in the special situation of an independent and identically distributed (IID) Gaussian

The author is with the Department of Communication Engineering, National Chiao Tung University (NCTU), Hsinchu, Taiwan.

This work was supported by the Industrial Technology Research Institute (ITRI), Zhudong, Taiwan, under contract G1-96001.

[1] The fading process is only assumed to be stationary, ergodic, regular, and of finite energy.

fading channel with a scalar line-of-sight component.[2] We will show that for a large line-of-sight component d the fading number basically grows like

$$\chi = n_{\mathrm{m}} \log |d|^2 \tag{4}$$

where

$$n_{\mathrm{m}} = \min\{n_{\mathrm{R}}, n_{\mathrm{T}}\} \tag{5}$$

is the degree of freedom of a MIMO fading channel as defined, *e.g.*, in [4]. Here n_{R} and n_{T} denote the number of antennas at the receiver and transmitter, respectively.

As an interesting side-product we will derive so far unknown closed-form expressions of some expectations of a noncentral chi-square random variable: we will give closed-form solutions to $\mathsf{E}[\log V]$ and $\mathsf{E}\left[\frac{1}{V^n}\right]$ for a noncentral chi-square random variable V with an even number of degrees of freedom. Note that in practice we often have an even number of degrees of freedom because we usually consider *complex* Gaussian random variables consisting of *two real* Gaussian components. We will see that these expectations are all related to a family of functions $g_m(\cdot)$ that is defined in Definition 3 in Section III-A. We will also derive some useful bounds on these functions.

The remainder of this paper is structured as follows. After some remarks about notation we will define the channel model and specify all assumptions in Section II. In Section III we give the preliminary results about some expectations of a noncentral chi-square random variable that are interesting by themselves. The main results concerning the MIMO Gaussian fading number are then summarized in Section IV. We conclude in Section V.

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 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. The $m \times m$ identity matrix is denoted by I_m, and $\mathsf{0}_{m\times n}$ stands for a $m \times n$ matrix with all components being zero.

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

$$\hat{\mathbf{v}} \triangleq \frac{\mathbf{v}}{\|\mathbf{v}\|}, \tag{6}$$

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

$$\|\hat{\mathbf{v}}\| = \|\hat{\mathbf{V}}\| = 1. \tag{7}$$

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 $p^{\lambda}_{\hat{\mathbf{V}}}(\hat{\mathbf{v}})$ with respect to λ, then we shall let

$$h_\lambda(\hat{\mathbf{V}}) \triangleq -\mathsf{E}\left[\log p^{\lambda}_{\hat{\mathbf{V}}}(\hat{\mathbf{V}})\right] \tag{8}$$

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

$$h_\lambda(\mathsf{U}\hat{\mathbf{V}}) = h_\lambda(\hat{\mathbf{V}}). \tag{9}$$

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

$$h_\lambda(\hat{\mathbf{V}}) = \log c_m \tag{10}$$

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

$$c_m = \frac{2\pi^m}{\Gamma(m)}. \tag{11}$$

The definition (8) 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 $p^{\lambda}_{\hat{\mathbf{V}}|\mathbf{W}}(\hat{\mathbf{v}}|\mathbf{w})$ then we can define

$$h_\lambda\big(\hat{\mathbf{V}} \bigm| \mathbf{W} = \mathbf{w}\big) \triangleq -\mathsf{E}\left[\log p^{\lambda}_{\hat{\mathbf{V}}|\mathbf{W}}(\hat{\mathbf{V}}|\mathbf{W}) \Bigm| \mathbf{W} = \mathbf{w}\right] \tag{12}$$

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

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 (6). Then

$$h(\mathbf{V}) = h(\|\mathbf{V}\|) + h_\lambda\big(\hat{\mathbf{V}} \bigm| \|\mathbf{V}\|\big) + (2m-1)\mathsf{E}[\log \|\mathbf{V}\|] \tag{13}$$

$$= h_\lambda\big(\hat{\mathbf{V}}\big) + h\big(\|\mathbf{V}\| \bigm| \hat{\mathbf{V}}\big) + (2m-1)\mathsf{E}[\log \|\mathbf{V}\|] \tag{14}$$

whenever all the quantities in (13) and (14), respectively, are defined. Here $h(\|\mathbf{V}\|)$ is the differential entropy of $\|\mathbf{V}\|$ when viewed as a real (scalar) random variable. Moreover, note that

$$h\big(\|\mathbf{V}\|^2\big) = h\big(\|\mathbf{V}\|\big) + \mathsf{E}[\log \|\mathbf{V}\|] + \log 2. \tag{15}$$

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, complex Gaussian random vector of covariance matrix $\mathsf{E}\big[(\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\dagger\big] = \mathsf{K}$. Similarly, $\mathcal{N}_{\mathbb{R}}(\boldsymbol{\mu}, \mathsf{K})$ denotes a *real* Gaussian random vector. By $X \sim$

[2]For precise definitions we refer to Section II.

$\mathcal{U}([a,b])$ 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.

We use "$\overset{\mathscr{L}}{=}$" to denote "equal in law," and "$\triangleq$" stands for "is defined as."

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

II. Channel Model and Definitions

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

$$\mathbf{Y} = \mathbb{H}\mathbf{x} + \mathbf{Z}. \tag{16}$$

Here $\mathbf{x} \in \mathbb{C}^{n_{\mathrm{T}}}$ denotes the input vector; the random matrix $\mathbb{H} \in \mathbb{C}^{n_{\mathrm{R}} \times n_{\mathrm{T}}}$ is the fading matrix; and the random vector $\mathbf{Z} \in \mathbb{C}^{n_{\mathrm{R}}}$ represents the additive noise vector.

We assume that the noise vector $\mathbf{Z}$ is a white, zero-mean, circularly symmetric, complex Gaussian random vector, *i.e.*, $\mathbf{Z} \sim \mathcal{N}_{\mathbb{C}}\big(\mathbf{0}, \sigma^2 \mathsf{I}_{n_{\mathrm{R}}}\big)$ for some $\sigma^2 > 0$.

We further assume that the fading matrix $\mathbb{H}$ can be written as

$$\mathbb{H} = \mathsf{D} + \tilde{\mathbb{H}} \tag{17}$$

where all components of the $n_{\mathrm{R}} \times n_{\mathrm{T}}$ random matrix $\tilde{\mathbb{H}}$ are IID $\sim \mathcal{N}_{\mathbb{C}}(0,1)$ and where the constant $n_{\mathrm{R}} \times n_{\mathrm{T}}$ line-of-sight matrix D is scalar in the sense that, for $n_{\mathrm{R}} \leq n_{\mathrm{T}}$,

$$\mathsf{D} = d \begin{pmatrix} \mathsf{I}_{n_{\mathrm{R}}} & \mathsf{0}_{n_{\mathrm{R}} \times (n_{\mathrm{T}} - n_{\mathrm{R}})} \end{pmatrix} \tag{18}$$

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

$$\mathsf{D} = d \begin{pmatrix} \mathsf{I}_{n_{\mathrm{T}}} \\ \mathsf{0}_{(n_{\mathrm{R}} - n_{\mathrm{T}}) \times n_{\mathrm{T}}} \end{pmatrix}, \tag{19}$$

where $d \in \mathbb{C}$ is a constant.

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

$$\mathrm{SNR} \triangleq \frac{\mathcal{E}}{\sigma^2}. \tag{20}$$

The capacity $\mathsf{C}(\mathrm{SNR})$ of the channel (16) is given by

$$\mathsf{C}(\mathrm{SNR}) = \sup_{Q_{\mathbf{X}}} I(\mathbf{X};\mathbf{Y}) \tag{21}$$

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

$$\|\mathbf{X}\|^2 \leq \mathcal{E}, \quad \text{almost surely} \tag{22}$$

for a peak-power constraint, or

$$\mathsf{E}\big[\|\mathbf{X}\|^2\big] \leq \mathcal{E} \tag{23}$$

for an average-power constraint.

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

$$\varlimsup_{\mathrm{SNR}\uparrow\infty} \Big\{ \mathsf{C}(\mathrm{SNR}) - \log\log \mathrm{SNR} \Big\} < \infty. \tag{24}$$

Note that [1, Theorem 4.2], [2, Theorem 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, (24) also holds in the situation of a peak-power constraint.

The fading number χ is now defined as [1, Definition 4.6], [2, Definition 6.13]

$$\chi(\mathbb{H}) \triangleq \varlimsup_{\mathrm{SNR}\uparrow\infty} \Big\{ \mathsf{C}(\mathrm{SNR}) - \log\log \mathrm{SNR} \Big\}. \tag{25}$$

It has been shown in [3] that the fading number is given as

$$\chi(\mathbb{H}) = \sup_{Q_{\hat{\mathbf{X}}}} \left\{ h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \right) + n_{\mathrm{R}} \mathsf{E}\Big[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2 \Big] - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) \right\}, \tag{26}$$

independently of the type of power constraint (22) or (23) that is imposed on the input. 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, it is shown in [3] that 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 supremum in (26) and where R is a nonnegative random variable independent of $\hat{\mathbf{X}}$ such that

$$\log R^2 \sim \mathcal{U}\left([\log\log \mathcal{E}, \log \mathcal{E}] \right). \tag{27}$$

In the following we will try to evaluate the expression (26) in the situation of IID MIMO Gaussian fading with a scalar line-of-sight component as defined in (17)–(19).

III. Preliminary Results: Some Expectations of a Noncentral Chi-Square Distribution

In the derivation of the main result we will need some expectations of a noncentral chi-square random variable. We state these results in a separate section because they are interesting by themselves. For space reason we omit the proofs. For them and more details we refer to [5], [6], and [7].

A. Definitions and Results

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

$$\sum_{j=1}^{n} (X_j + \mu_j)^2, \tag{28}$$

where $\{X_j\}_{j=1}^n$ are IID $\sim \mathcal{N}_{\mathbb{R}}(0,1)$ and the real constants $\{\mu_j\}_{j=1}^n$ satisfy

$$s^2 = \sum_{j=1}^{n} \mu_j^2. \tag{29}$$

(The distribution of (28) depends on the constants $\{\mu_j\}$ only via the sum of their squares.) The probability density function of such a distribution is given by [8, Chapter 29]

$$\frac{1}{2}\left(\frac{x}{s^2}\right)^{\frac{n-2}{4}} e^{-\frac{s^2+x}{2}} I_{n/2-1}\left(s\sqrt{x}\right), \qquad x \geq 0. \tag{30}$$

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

$$I_\nu(x) \triangleq \sum_{k=0}^{\infty} \frac{1}{k!\,\Gamma(\nu+k+1)}\left(\frac{x}{2}\right)^{\nu+2k}, \quad x \geq 0 \tag{31}$$

(see [9, Eq. 8.445]).

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

Definition 2: Let the random variable V have a noncentral chi-square distribution with an even number $2m$ of degrees of freedom, *i.e.*,

$$V \triangleq \sum_{j=1}^{m} \left|U_j + \mu_j\right|^2 \tag{32}$$

where $\{U_j\}_{j=1}^m$ are IID $\sim \mathcal{N}_{\mathbb{C}}(0,1)$, and $\{\mu_j\}_{j=1}^m$ are complex constants. Let further the noncentrality parameter s^2 be defined as

$$s^2 \triangleq \sum_{j=1}^{m} |\mu_j|^2. \tag{33}$$

Next we define the following continuous functions.

Definition 3: The functions $g_m(\cdot)$ are defined as follows:

$$g_m(\xi) \triangleq \begin{cases} \log(\xi) - \mathrm{Ei}(-\xi) + \sum_{j=1}^{m-1}(-1)^j \Big[e^{-\xi}(j-1)! \\ \qquad - \frac{(m-1)!}{j(m-1-j)!}\Big]\left(\frac{1}{\xi}\right)^j, & \xi > 0 \\ \psi(m), & \xi = 0 \end{cases} \tag{34}$$

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

$$\mathrm{Ei}(-\xi) \triangleq -\int_{\xi}^{\infty} \frac{e^{-t}}{t}\,\mathrm{d}t, \qquad \xi > 0 \tag{35}$$

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

$$\psi(m) \triangleq -\gamma + \sum_{j=1}^{m-1} \frac{1}{j}, \qquad m \in \mathbb{N} \tag{36}$$

with $\gamma \approx 0.577$ denoting Euler's constant.

Note that $g_m(\xi)$ is continuous for all $\xi \geq 0$, *i.e.*, in particular

$$\lim_{\xi \downarrow 0} \left\{ \log(\xi) - \mathrm{Ei}(-\xi) + \sum_{j=1}^{m-1}(-1)^j \Big[e^{-\xi}(j-1)! - \frac{(m-1)!}{j(m-1-j)!}\Big]\left(\frac{1}{\xi}\right)^j \right\} = \psi(m) \tag{37}$$

for all $m \in \mathbb{N}$. Therefore its first derivative is defined for all $\xi \geq 0$ and can be evaluated to

$$g_m'(\xi) \triangleq \frac{\partial g_m(\xi)}{\partial \xi} = \frac{(-1)^m \Gamma(m)}{\xi^m}\left(e^{-\xi} - \sum_{j=0}^{m-1} \frac{(-1)^j}{j!}\xi^j\right) \tag{38}$$

(see [1, Eq. (417)], [2, Eq. (A.39)]). Note that $g_m'(\cdot)$ is also continuous, *i.e.*, in particular

$$\lim_{\xi \downarrow 0} \left\{ \frac{(-1)^m \Gamma(m)}{\xi^m}\left(e^{-\xi} - \sum_{j=0}^{m-1} \frac{(-1)^j}{j!}\xi^j\right) \right\} = \frac{1}{m} = g_m'(0). \tag{39}$$

Now we will give closed-form expressions for some expectations of a noncentral chi-square random variable. We start with the logarithm.

Theorem 4: The expected value of the logarithm of a noncentral chi-square random variable with an even number $2m$ of degrees of freedom is given as

$$\boxed{\mathsf{E}[\log V] = g_m(s^2)} \tag{40}$$

where V and s^2 are defined in (32) and (33). Hence, we have found the solution to the following integral:

$$\int_0^\infty \log v \cdot \left(\frac{v}{s^2}\right)^{\frac{m-1}{2}} e^{-v-s^2} I_{m-1}(2s\sqrt{v})\,\mathrm{d}v = g_m(s^2) \tag{41}$$

for any $m \in \mathbb{N}$ and $s^2 \geq 0$.

Proof: A proof can be found in [1, Lemma 10.1], [2, Lemma A.6] ■

Next we look at the reciprocal value.

Theorem 5: Let $n \in \mathbb{N}$ with $n < m$. The expected value of the n-th power reciprocal value of a noncentral chi-square random variable with an even number $2m$ of degrees of freedom is given as

$$\boxed{\mathsf{E}\left[\frac{1}{V^n}\right] = \frac{(-1)^{n-1}}{(n-1)!} \cdot g_{m-n}^{(n)}(s^2), \qquad m > n} \tag{42}$$

where

$$g_m^{(\ell)}(\xi) = \frac{\partial^\ell g_m(\xi)}{\partial \xi^\ell} \tag{43}$$

denotes the ℓ-th derivative of $g_m(\cdot)$ and where V and s^2 are defined in (32) and (33). In particular, for $m > 1$

$$\boxed{\mathsf{E}\left[\frac{1}{V}\right] = g_{m-1}'(s^2).} \tag{44}$$

Hence, we have found the solution to the following integral:

$$\int_0^\infty \frac{1}{v^n} \cdot \left(\frac{v}{s^2}\right)^{\frac{m-1}{2}} e^{-v-s^2} I_{m-1}(2s\sqrt{v})\, \mathrm{d}v = \frac{(-1)^{n-1}}{(n-1)!} \cdot g_{m-n}^{(n)}(s^2) \quad (45)$$

for any $m, n \in \mathbb{N}$, $m > n$, and any real $s^2 \geq 0$.
Note that in the cases where $m \leq n$ the expectation is unbounded.

Proof: The proof is based on a series expansion of the modified Bessel function similarly to the proof of [1, Lemma 10.1], [2, Lemma A.6]. For more details see [5]. ∎

B. Properties and Bounds of $g_m(\cdot)$ and $g'_m(\cdot)$

We will next summarize some properties of the family of functions $g_m(\cdot)$ and $g'_m(\cdot)$ and state some useful bounds. For proofs we refer to [5]. We start with a lemma that shows that these functions are well-behaved.

Lemma 6: The functions $g_m(\xi)$ are monotonically strictly increasing and strictly concave in ξ for all $m \in \mathbb{N}$, and monotonically strictly increasing in m for all $\xi \geq 0$. The functions $g'_m(\xi)$ are positive, monotonically strictly decreasing, and strictly convex functions in ξ for all $m \in \mathbb{N}$, and monotonically strictly decreasing in m for all $\xi \geq 0$.

Next we state some interesting relations between $g_m(\cdot)$ and $g'_m(\cdot)$.

Lemma 7: We have the following relations:

$$g_{m+1}(\xi) = g_m(\xi) + g'_m(\xi) \quad (46)$$

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

$$g'_{m+1}(\xi) = \frac{1}{\xi} - \frac{m}{\xi} g'_m(\xi), \qquad \xi > 0, \quad (47)$$

and
$$g'_m(\xi) = \frac{1}{m} - \frac{\xi}{m} g'_{m+1}(\xi), \qquad \xi \geq 0, \quad (48)$$

for all $m \in \mathbb{N}$.

Finally, we give some tight bounds.

Lemma 8: For the functions $g_m(\cdot)$ we state two sets of bounds. The first set is tighter:

$$g_m(\xi) \geq \log \xi - \mathrm{Ei}(-\xi) + \sum_{j=1}^{m-1} \frac{1}{\xi + j}, \quad (49)$$

$$g_m(\xi) \leq \log \xi - \mathrm{Ei}(-\xi) + \sum_{j=1}^{m-1} \min\left\{\frac{j+1}{j(\xi+j+1)}, \frac{1}{\xi+j-1}\right\}. \quad (50)$$

Secondly, we give a set of bounds that is slightly less tight, but that is very appealing because the expressions are simple and easy to use for further analysis:

$$\log(\xi + m - 1) \leq g_m(\xi) \leq \log(\xi + m). \quad (51)$$

The bounds (49) and (50) are depicted in Figure 1 and the bounds (51) in Figure 2, both times for the cases of $m = 1$, $m = 3$, and $m = 8$.

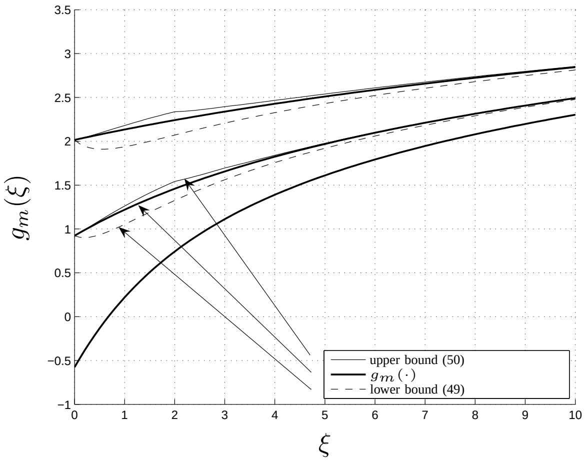

Fig. 1. Upper and lower bounds on $g_m(\cdot)$ according to (49) and (50) in Lemma 8. The lowest curve corresponds to $m = 1$ (in this case all bounds coincides with $g_1(\cdot)$), the next three curves correspond to $m = 3$, and the top three to $m = 8$.

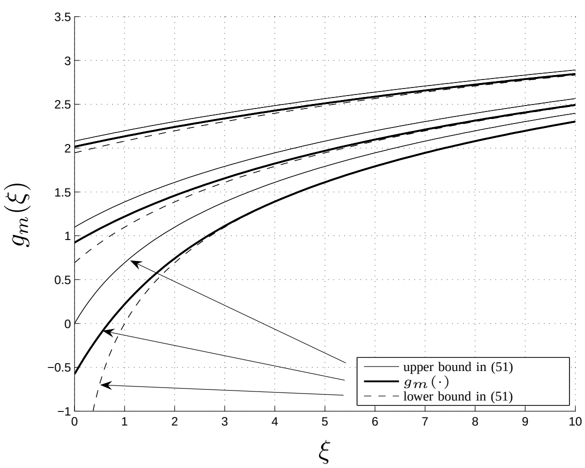

Fig. 2. Upper and lower bounds on $g_m(\cdot)$ according to (51) in Lemma 8. The lowest three curves correspond to $m = 1$, the next three to $m = 3$, and the top three to $m = 8$.

Lemma 9: The function $g'_m(\cdot)$ can be bounded as follows:

$$\frac{1}{\xi + m} \leq g'_m(\xi) \leq \min\left\{\frac{m+1}{m(\xi+m+1)}, \frac{1}{\xi+m-1}\right\}. \quad (52)$$

Note that for $\xi < m + 1$ the first of the two upper bounds is tighter than second (*i.e.*, the first argument of the min-operator achieves the minimum), while for $\xi > m + 1$ the second is tighter (the second argument is smaller). Moreover, the first upper bound coincides with $g'_m(\xi)$ for $\xi = 0$, and the second upper bound is asymptotically tight when ξ tends to infinity.

The bounds (52) are depicted in Figure 3 for the cases of $m = 1$, $m = 3$, and $m = 8$.

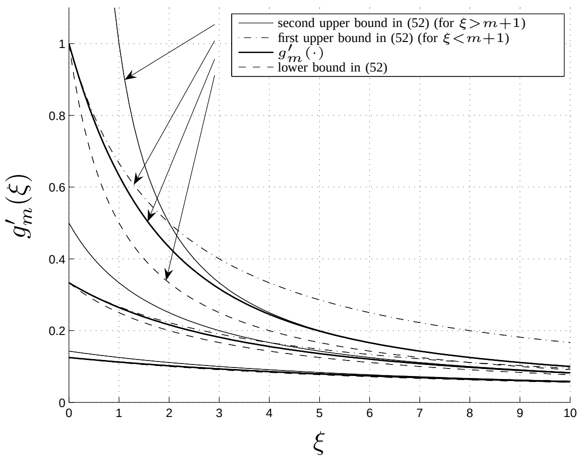

Fig. 3. Upper and lower bounds on $g'_m(\cdot)$ according to (52) in Lemma 9. The top four curves correspond to $m = 1$, the middle four to $m = 3$, and the lowest group of four curves to $m = 8$.

IV. Main Result: Fading Number of an IID Gaussian Fading Channel with a Scalar Line-of-Sight Component

To state our main result we need to make a case distinction. We start with the situation $n_R \leq n_T$ which turns out to be easier to solve.

Theorem 10: Assume $n_R \leq n_T$ and a Gaussian fading matrix as given in (17) and (18). Then

$$\boxed{\chi(\mathbb{H}) = n_R g_{n_R}\left(|d|^2\right) - n_R - \log\Gamma(n_R)} \tag{53}$$

where $g_m(\cdot)$ is defined in (34). The fading number is achievable by an input $\mathbf{X} = R \cdot \hat{\mathbf{X}}^*$ with $R \perp\!\!\!\perp \hat{\mathbf{X}}^*$, where the distribution of $R \in \mathbb{R}^+$ is specified in (27) and where

$$\hat{\mathbf{X}}^* = \begin{pmatrix} \mathbf{\Xi}^* \\ \mathbf{0} \end{pmatrix} \tag{54}$$

with $\mathbf{\Xi}^* \in \mathbb{C}^{n_R}$ being an *isotropically distributed* unit vector.

Proof: We write for the unit vector $\hat{\mathbf{X}}$ of the maximization in (26)

$$\hat{\mathbf{X}} = \begin{pmatrix} \mathbf{\Xi} \\ \mathbf{\Xi}' \end{pmatrix} \tag{55}$$

where $\mathbf{\Xi} \in \mathbb{C}^{n_R}$ and $\mathbf{\Xi}' \in \mathbb{C}^{n_T - n_R}$. Note that from $\|\hat{\mathbf{X}}\|^2 = 1$ it follows that $\|\mathbf{\Xi}\|^2 \leq 1$. Then

$$\mathbb{H}\hat{\mathbf{X}} = \mathsf{D}\hat{\mathbf{X}} + \tilde{\mathbb{H}}\hat{\mathbf{X}} \stackrel{\mathscr{L}}{=} d\mathbf{\Xi} + \tilde{\mathbf{H}} \tag{56}$$

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

$$h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) = h\big(\tilde{\mathbf{H}}\big) = n_R \log \pi e; \tag{57}$$

$$n_R \mathsf{E}\left[\log \|\mathbb{H}\hat{\mathbf{X}}\|^2\right] = n_R g_{n_R}\left(|d|^2 \|\mathbf{\Xi}\|^2\right) \tag{58}$$

$$\leq n_R g_{n_R}\left(|d|^2\right); \tag{59}$$

$$h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) \leq \log \frac{2\pi^{n_R}}{\Gamma(n_R)}. \tag{60}$$

Here, (58) follows from the fact that $\|d\mathbf{\Xi} + \tilde{\mathbf{H}}\|^2$ is noncentral chi-square distributed and from (40). The inequality (59) follows from the monotonicity of $g_m(\cdot)$ and the fact that $\|\mathbf{\Xi}\|^2 \leq 1$. It is tight if $\|\mathbf{\Xi}\|^2 = 1$, *i.e.*, $\mathbf{\Xi}' = \mathbf{0}$. The inequality (60) follows from (10) and (11) and is tight if $\mathbf{\Xi}$ is uniformly distributed on the unit sphere in $\mathbb{C}^{n_R}$ so that $\mathbb{H}\hat{\mathbf{X}}$ is isotropically distributed. The result now follows from (26). ■

The case $n_R > n_T$ is more difficult since then (60) is in general not tight. We firstly need to introduce some notation. Note that

$$\mathbb{H}\hat{\mathbf{X}} = \mathsf{D}\hat{\mathbf{X}} + \tilde{\mathbb{H}}\hat{\mathbf{X}} \stackrel{\mathscr{L}}{=} \begin{pmatrix} d\hat{\mathbf{X}} \\ \mathbf{0} \end{pmatrix} + \tilde{\mathbf{H}} \tag{61}$$

where $\tilde{\mathbf{H}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathsf{I}_{n_R})$. Let us split the vector $\tilde{\mathbf{H}}$ into two parts:

$$\tilde{\mathbf{H}} = \begin{pmatrix} \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \end{pmatrix} \tag{62}$$

where $\tilde{\mathbf{H}}_1 \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathsf{I}_{n_T})$ and $\tilde{\mathbf{H}}_2 \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathsf{I}_{n_R - n_T})$ are two independent white Gaussian random vectors in $\mathbb{C}^{n_T}$ and $\mathbb{C}^{n_R - n_T}$, respectively. Then we can write

$$\mathbb{H}\hat{\mathbf{X}} \stackrel{\mathscr{L}}{=} \begin{pmatrix} d\hat{\mathbf{X}} + \tilde{\mathbf{H}}_1 \\ \tilde{\mathbf{H}}_2 \end{pmatrix}. \tag{63}$$

Next we define

$$S_1 \triangleq \left\|d\hat{\mathbf{X}} + \tilde{\mathbf{H}}_1\right\|^2, \tag{64}$$

$$S_2 \triangleq \left\|\tilde{\mathbf{H}}_2\right\|^2. \tag{65}$$

Note that S_1 is noncentral chi-square distributed with $2n_T$ degrees of freedom and noncentrality parameter $\|d\hat{\mathbf{X}}\|^2 = |d|^2$ independently of the distribution of $\hat{\mathbf{X}}$, and that S_2 is central chi-square distributed with $2(n_R - n_T)$ degrees of freedom. Moreover, S_1 and S_2 are independent of each other.

Theorem 11: Assume $n_R > n_T$ and a Gaussian fading matrix as given in (17) and (19). Then

$$\boxed{\chi(\mathbb{H}) = n_T g_{n_T}\left(|d|^2\right) - n_T - \log\Gamma(n_T) + I\left(S_1; \frac{S_2}{S_1}\right)} \tag{66}$$

where $g_m(\cdot)$ is defined in (34). The fading number is achievable by an input $\mathbf{X} = R \cdot \hat{\mathbf{X}}^*$ with $R \perp\!\!\!\perp \hat{\mathbf{X}}^*$, where the distribution of $R \in \mathbb{R}^+$ is specified in (27) and where $\hat{\mathbf{X}}^*$ is an *isotropically distributed* unit vector.

Proof: A proof is given in the Appendix. ■

Unfortunately, we have not succeeded in deriving the term $I(S_1; S_2/S_1)$ precisely. Instead, we will state an upper and a lower bound that both do not depend on d. This shows that $I(S_1; S_2/S_1)$ is bounded in d.

Claim 12: For $n_R > n_T$ let S_1 and S_2 be independent random variables defined in (64) and (65), respectively. Then for $n_T > 1$

$$0 \leq I\left(S_1; \frac{S_2}{S_1}\right) \leq (n_R - n_T - 1)\psi(n_R - n_T) - (n_R - n_T - 1) - \log\frac{\Gamma(n_R - n_T)}{n_R - n_T} + \log\frac{n_T}{n_T - 1}, \tag{67}$$

and for $n_\mathrm{T}=1$

$$0 \le I\left(S_1; \frac{S_2}{S_1}\right) \le (n_\mathrm{R}-2)\psi(n_\mathrm{R}-1) - (n_\mathrm{R}-2) - \log\frac{\Gamma(n_\mathrm{R}-1)}{n_\mathrm{R}-1} + \log\frac{\pi}{2} + 1. \tag{68}$$

Proof: The proof relies strongly on the results from Section III about the noncentral chi-square distribution. For space reasons we omit the details. ∎

Hence, combining Theorem 10, Theorem 11, and Claim 12 we get the following corollary.

Corollary 13: The fading number of the IID MIMO Gaussian fading channel as defined in (16)–(19) is given by

$$\boxed{\chi(\mathbb{H}) = n_\mathrm{m} g_{n_\mathrm{m}}\big(|d|^2\big) - n_\mathrm{m} - \log\Gamma\big(n_\mathrm{m}\big) + f(n_\mathrm{R}, n_\mathrm{T}; d)} \tag{69}$$

where

$$n_\mathrm{m} = \min\{n_\mathrm{R}, n_\mathrm{T}\} \tag{70}$$

denotes the degree of freedom of a MIMO fading channel and where $f(n_\mathrm{R}, n_\mathrm{T}; d)$ depends primarily on n_T and n_R and is bounded in d:

$$0 \le f(n_\mathrm{R}, n_\mathrm{T}; d) \le (n_\mathrm{R}-n_\mathrm{T}-1)\psi(n_\mathrm{R}-n_\mathrm{T}) - (n_\mathrm{R}-n_\mathrm{T}-1) - \log\frac{\Gamma(n_\mathrm{R}-n_\mathrm{T})}{n_\mathrm{R}-n_\mathrm{T}} + \log\left(\min\left\{\frac{n_\mathrm{T}}{n_\mathrm{T}-1}, \frac{\pi e}{2}\right\}\right). \tag{71}$$

Using (51) we therefore have for $|d| \gg 1$

$$\chi \sim n_\mathrm{m}\log\big(|d|^2 + n_\mathrm{m}\big). \tag{72}$$

V. Conclusions

We have evaluated the expression of the fading number of a memoryless IID Gaussian fading channel with scalar line-of-sight component d. We have shown that for large d, the fading number grows like $n_\mathrm{m} g_{n_\mathrm{m}}\big(|d|^2\big)$ where $g_{n_\mathrm{m}}\big(|d|^2\big)$ grows like $\log|d|^2$ and where n_m, denoted *degree of freedom of a MIMO fading channel*, is the smaller of the number of antennas at transmitter and receiver, respectively.

Moreover, we have shown that the optimal input is basically isotropically distributed, *i.e.,* for $n_\mathrm{R} \le n_\mathrm{T}$

$$\hat{\mathbf{X}}^* = \begin{pmatrix} \boldsymbol{\Xi}^* \\ \mathbf{0} \end{pmatrix} \tag{73}$$

with $\boldsymbol{\Xi}^* \in \mathbb{C}^{n_\mathrm{R}}$ being an *isotropically distributed* unit vector, and for $n_\mathrm{R} > n_\mathrm{T}$

$$\hat{\mathbf{X}}^* = \boldsymbol{\Xi}^* \tag{74}$$

with $\boldsymbol{\Xi}^* \in \mathbb{C}^{n_\mathrm{T}}$ being an *isotropically distributed* unit vector.

Furthermore, we have derived closed-form expressions for some important expectations in the fields of information theory and communications. In particular, we have computed $\mathsf{E}[\log V]$ and $\mathsf{E}\left[\frac{1}{V^n}\right]$ for a noncentral chi-square random variable V with an even number of degrees of freedom. Moreover, we have shown that the resulting functions behave nicely and we have found tight upper and lower bounds to them.

Appendix

We will give here a derivation of Theorem 11. We start with the derivation of an optimal input distribution. From (26) we know that the optimal input is given as $\mathbf{X} = R\cdot\hat{\mathbf{X}}$ and (27) specifies an optimal choice of R. It remains to prove that an isotropic unit vector $\hat{\mathbf{X}}$ is optimal. To that goal let V be an arbitrary (deterministic) unitary $n_\mathrm{T}\times n_\mathrm{T}$ matrix and define the $n_\mathrm{R}\times n_\mathrm{R}$ matrix U as follows:

$$\mathsf{U} \triangleq \begin{pmatrix} \mathsf{V} & \mathsf{0}_{n_\mathrm{T}\times(n_\mathrm{R}-n_\mathrm{T})} \\ \mathsf{0}_{(n_\mathrm{R}-n_\mathrm{T})\times n_\mathrm{T}} & \mathsf{I}_{n_\mathrm{R}-n_\mathrm{T}} \end{pmatrix}. \tag{75}$$

It is not hard to show that U is unitary and that for any choice of V we have $\mathsf{U}\mathbb{H}\hat{\mathbf{X}} \overset{\mathscr{L}}{=} \mathbb{H}\mathsf{V}\hat{\mathbf{X}}$.

Let's now make V random, *i.e.,* we define a $n_\mathrm{T}\times n_\mathrm{T}$ unitary matrix $\mathbb{V}$ that is Haar distributed[3] and independent of $(\mathbb{H}, \mathbf{X}, \mathbf{Z})$. Further we define a random unitary $n_\mathrm{R}\times n_\mathrm{R}$ matrix $\mathbb{U}$ analogously to (75):

$$\mathbb{U} \triangleq \begin{pmatrix} \mathbb{V} & \mathsf{0}_{n_\mathrm{T}\times(n_\mathrm{R}-n_\mathrm{T})} \\ \mathsf{0}_{(n_\mathrm{R}-n_\mathrm{T})\times n_\mathrm{T}} & \mathsf{I}_{n_\mathrm{R}-n_\mathrm{T}} \end{pmatrix}. \tag{76}$$

Then we get:

$$\begin{aligned}
&I\big(\mathbf{X}; \mathbb{H}\mathbf{X}+\mathbf{Z}\big) \\
&\quad= I\big(\mathbf{X}; \mathbb{H}\mathbf{X}+\mathbf{Z}\,\big|\,\mathbb{U}\big) &(77)\\
&\quad= I\big(\mathbb{V}\mathbf{X}; \mathbb{U}\mathbb{H}\mathbf{X}+\mathbb{U}\mathbf{Z}\,\big|\,\mathbb{U}\big) &(78)\\
&\quad= I\big(\mathbb{V}\mathbf{X}; \mathbb{H}\mathbb{V}\mathbf{X}+\mathbf{Z}\,\big|\,\mathbb{U}\big) &(79)\\
&\quad= I\big(\check{\mathbf{X}}; \mathbb{H}\check{\mathbf{X}}+\mathbf{Z}\,\big|\,\mathbb{U}\big) &(80)\\
&\quad= h\big(\mathbb{H}\check{\mathbf{X}}+\mathbf{Z}\,\big|\,\mathbb{U}\big) - h\big(\mathbb{H}\check{\mathbf{X}}+\mathbf{Z}\,\big|\,\check{\mathbf{X}},\mathbb{U}\big) &(81)\\
&\quad= h\big(\mathbb{H}\check{\mathbf{X}}+\mathbf{Z}\,\big|\,\mathbb{U}\big) - h\big(\mathbb{H}\check{\mathbf{X}}+\mathbf{Z}\,\big|\,\check{\mathbf{X}}\big) &(82)\\
&\quad\le h\big(\mathbb{H}\check{\mathbf{X}}+\mathbf{Z}\big) - h\big(\mathbb{H}\check{\mathbf{X}}+\mathbf{Z}\,\big|\,\check{\mathbf{X}}\big) &(83)\\
&\quad= I\big(\check{\mathbf{X}}; \mathbb{H}\check{\mathbf{X}}+\mathbf{Z}\big). &(84)
\end{aligned}$$

Here, the first equality follows since $\mathbb{U}$ is independent of the other random quantities; the subsequent equality follows because given $\mathbb{U}$ also $\mathbb{V}$ is known and because mutual information is not changed when the arguments are multiplied by known invertible quantities; in the subsequent equality we use that $\mathbb{U}\mathbb{H}\hat{\mathbf{X}} \overset{\mathscr{L}}{=} \mathbb{H}\mathbb{V}\hat{\mathbf{X}}$ as shown above and that $\mathbb{U}\mathbf{Z} \overset{\mathscr{L}}{=} \mathbf{Z}$; in (80) we introduce $\check{\mathbf{X}} \triangleq \mathbb{V}\mathbf{X}$; the subsequent equality follows from the definition of mutual information; then we use the fact that $\mathbb{H}\check{\mathbf{X}}+\mathbf{Z}\big|\check{\mathbf{X}}$ is independent of $\mathbb{U}$ or $\mathbb{V}$; and the inequality follows from conditioning that cannot increase entropy.

Note that $\check{\mathbf{X}}$ is isotropically distributed independently of the distribution of $\mathbf{X}$. Hence, an isotropic input will achieve at least the same mutual information as any other input.

Next, we will derive the expression (66). To that goal we start with Lemma 1 and plug (15) into (13):

$$h_\lambda\big(\hat{\mathbf{V}}\,\big|\,\|\mathbf{V}\|\big) = h(\mathbf{V}) - h(\|\mathbf{V}\|^2) + \log 2 - (m-1)\mathsf{E}\left[\log\|\mathbf{V}\|^2\right]. \tag{85}$$

We then choose $\mathbf{V} \triangleq \mathbb{H}\hat{\mathbf{X}}$ and plug this expression into (26). Note that we can drop the supremum since we have proven

[3]A random matrix $\mathbb{T}$ is Haar distributed if for any deterministic unitary matrix M we have that $\mathsf{M}\mathbb{T} \overset{\mathscr{L}}{=} \mathbb{T}$.

above that the supremum is achieved by an isotropically distributed $\hat{\mathbf{X}}$. We get the following:

$$\chi = I\left(\|\mathbb{H}\hat{\mathbf{X}}\|; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + h_\lambda\left(\frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|} \,\middle|\, \|\mathbb{H}\hat{\mathbf{X}}\|\right)$$
$$+ n_\mathrm{R}\mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\right] - \log 2 - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) \tag{86}$$
$$= I\left(\|\mathbb{H}\hat{\mathbf{X}}\|; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + h\big(\mathbb{H}\hat{\mathbf{X}}\big) - h\big(\|\mathbb{H}\hat{\mathbf{X}}\|^2\big)$$
$$+ \mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\right] - h\big(\mathbb{H}\hat{\mathbf{X}} \,\big|\, \hat{\mathbf{X}}\big) \tag{87}$$
$$= I\left(\|\mathbb{H}\hat{\mathbf{X}}\|^2; \frac{\mathbb{H}\hat{\mathbf{X}}}{\|\mathbb{H}\hat{\mathbf{X}}\|}\right) + I\big(\hat{\mathbf{X}}; \mathbb{H}\hat{\mathbf{X}}\big) - h\big(\|\mathbb{H}\hat{\mathbf{X}}\|^2\big)$$
$$+ \mathsf{E}\left[\log\|\mathbb{H}\hat{\mathbf{X}}\|^2\right] \tag{88}$$
$$= I\left(S_1 + S_2; \frac{\mathbb{H}\hat{\mathbf{X}}}{\sqrt{S_1+S_2}}\right) + I\big(\hat{\mathbf{X}}; \mathbb{H}\hat{\mathbf{X}}\big) - h\big(S_1+S_2\big)$$
$$+ \mathsf{E}[\log(S_1+S_2)]\,. \tag{89}$$

Here, the first equality follows from (26); in the subsequent equality we use (85); the next step follows from the definition of mutual information and the fact that squaring a nonnegative argument of mutual information does not change its value; and in the final equality we apply our definitions (64) and (65).

The first term in (89) can be simplified as follows:

$$I\left(S_1+S_2; \frac{\mathbb{H}\hat{\mathbf{X}}}{\sqrt{S_1+S_2}}\right)$$
$$= I\left(S_1+S_2; \frac{d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1}{\sqrt{S_1+S_2}}, \frac{\tilde{\mathbf{H}}_2}{\sqrt{S_1+S_2}}\right) \tag{90}$$
$$= I\left(S_1+S_2; \frac{d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1}{\sqrt{S_1}}, \frac{\tilde{\mathbf{H}}_2}{\sqrt{S_2}}, \frac{S_2}{S_1}\right) \tag{91}$$
$$= h(S_1+S_2) - h\left(S_1+S_2 \,\middle|\, \frac{d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1}{\sqrt{S_1}}, \frac{\tilde{\mathbf{H}}_2}{\sqrt{S_2}}, \frac{S_2}{S_1}\right) \tag{92}$$
$$= h(S_1+S_2) - h\left(S_1+S_2 \,\middle|\, \frac{S_2}{S_1}\right) \tag{93}$$
$$= h(S_1+S_2) - h\left(S_1\left(1+\frac{S_2}{S_1}\right) \,\middle|\, \frac{S_2}{S_1}\right) \tag{94}$$
$$= h(S_1+S_2) - h\left(S_1 \,\middle|\, \frac{S_2}{S_1}\right) - \mathsf{E}\left[\log\left(1+\frac{S_2}{S_1}\right)\right]. \tag{95}$$

Here the first equality follows from splitting the vector into two subvectors, one with n_T and one with $n_\mathrm{R}-n_\mathrm{T}$ components (see (61)); in the subsequent equality we perform some one-to-one transformations that do not change the value of the mutual information. In (93) we note that $\hat{\mathbf{X}}$ is isotropically distributed and that therefore both $d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1$ and $\tilde{\mathbf{H}}_2$ are isotropic. Hence, the magnitudes S_1 and S_2 are independent of the directions $\frac{d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1}{\sqrt{S_1}}$ and $\frac{\tilde{\mathbf{H}}_2}{\sqrt{S_2}}$. And finally, the last equality follows from the scaling property of differential entropy of a *real* argument.

For the second term in (89) we note that it is independent of the last $n_\mathrm{R}-n_\mathrm{T}$ rows of $\mathbb{H}$:

$$I\big(\hat{\mathbf{X}}; \mathbb{H}\hat{\mathbf{X}}\big) = I\left(\hat{\mathbf{X}}; \begin{pmatrix} d\hat{\mathbf{X}} \\ \mathbf{0} \end{pmatrix} + \tilde{\mathbf{H}}\right) \tag{96}$$
$$= I\big(\hat{\mathbf{X}}; d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1\big); \tag{97}$$

and for the last term we get

$$\mathsf{E}[\log(S_1+S_2)] = \mathsf{E}[\log S_1] + \mathsf{E}\left[\log\left(1+\frac{S_2}{S_1}\right)\right]. \tag{98}$$

Putting this together we yield

$$\chi = I\big(\hat{\mathbf{X}}; d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1\big) + \mathsf{E}[\log S_1] - h\left(S_1 \,\middle|\, \frac{S_2}{S_1}\right) \tag{99}$$
$$= I\big(\hat{\mathbf{X}}; d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1\big) + \mathsf{E}[\log S_1] - h(S_1)$$
$$+ I\left(S_1; \frac{S_2}{S_1}\right). \tag{100}$$

Now note that the same derivation can be done for the situation $n_\mathrm{R} = n_\mathrm{T}$ in which case $S_2 = 0$. We get exactly the same result (100) apart from the last term $I\left(S_1; \frac{S_2}{S_1}\right)$ which is in this situation equal to zero. However, we know from Theorem 10 the exact value of the fading number for $n_\mathrm{R} = n_\mathrm{T}$, and hence we see that

$$I\big(\hat{\mathbf{X}}; d\hat{\mathbf{X}}+\tilde{\mathbf{H}}_1\big) + \mathsf{E}[\log S_1] - h(S_1)$$
$$= n_\mathrm{T} g_{n_\mathrm{T}}\big(|d|^2\big) - n_\mathrm{T} - \log\Gamma(n_\mathrm{T}). \tag{101}$$

Plugging this into (100) then yields

$$\chi = n_\mathrm{T} g_{n_\mathrm{T}}\big(|d|^2\big) - n_\mathrm{T} - \log\Gamma(n_\mathrm{T}) + I\left(S_1; \frac{S_2}{S_1}\right) \tag{102}$$

which proves the claim.

ACKNOWLEDGMENT

This work was supported by the Industrial Technology Research Institute (ITRI), Zhudong, Taiwan, under contract G1-96001.

REFERENCES

[1] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Trans. Inf. 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 Technol. (ETH), Zurich, Oct. 2004, Diss. ETH No. 15769. [Online]. Available: http://moser.cm.nctu.edu.tw/

[3] S. M. Moser, "The fading number of memoryless multiple-input multiple-output fading channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2652–2666, Jul. 2007.

[4] R. Etkin and D. Tse, "Degrees of freedom in underspread MIMO fading channels," in *IEEE Int. Symp. Inf. Theory*, Yokohama, Japan, June 29 – July 4, 2003, p. 323.

[5] S. M. Moser, "Some expectations of a non-central chi-square distribution with an even number of degrees of freedom," Apr. 2007, subm. [Online]. Available: http://moser.cm.nctu.edu.tw/publications.html

[6] A. Lapidoth and S. M. Moser, "The expected logarithm of a noncentral chi-square random variable," website. [Online]. Available: http://moser.cm.nctu.edu.tw/explog.html

[7] S. M. Moser, "Some expectations of a non-central chi-square distribution with an even number of degrees of freedom," To be pres. at *IEEE TENCON*, Taipei, Taiwan, Oct. 31 – Nov. 2, 2007.

[8] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd ed. John Wiley & Sons, 1995, vol. 2.

[9] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed., A. Jeffrey, Ed. Academic Press, San Diego, 1996.