The Fading Number of Single-Input Multiple-Output Fading Channels With Memory

Amos Lapidoth, *Fellow, IEEE,* and Stefan M. Moser, *Member, IEEE*

***Abstract*—We derive the fading number of stationary and ergodic (not necessarily Gaussian) single-input multiple-output (SIMO) fading channels with memory. This is the second term, after the double-logarithmic term, of the high signal-to-noise ratio (SNR) expansion of channel capacity. The transmitter and receiver are assumed to be cognizant of the probability law governing the fading but not of its realization.**

It is demonstrated that the fading number is achieved by independent and identically distributed (i.i.d.) circularly symmetric inputs of squared magnitude whose logarithm is uniformly distributed over an SNR-dependent interval. The upper limit of the interval is the logarithm of the allowed transmit power, and the lower limit tends to infinity sublogarithmically in the SNR. The converse relies *inter alia* on a new observation regarding input distributions that escape to infinity.

Lower and upper bounds on the fading number for Gaussian fading are also presented. These are related to the mean squared-errors of the one-step predictor and the one-gap interpolator of the fading process respectively. The bounds are computed explicitly for stationary mth-order autoregressive AR(m) Gaussian fading processes.

***Index Terms*—Autoregressive process, channel capacity, fading, fading number, high signal-to-noise ratio (SNR), memory, multiple antenna, single-input multiple-output (SIMO).**

I. Introduction

It has been recently shown in [1] that, whenever the matrix-valued fading process is of finite differential entropy rate, the capacity of multiple-input multiple-output (MIMO) fading channels typically grows only double logarithmically in the signal-to-noise ratio (SNR). To quantify the rates at which this poor power efficiency begins, [1] introduced the *fading number* as the second term in the high-SNR asymptotic expansion of channel capacity. Explicit expressions for the fading number were then given for a number of memoryless fading models. For channels with memory, only the fading number of single-input single-output (SISO) channels was derived.

Manuscript received November 5, 2004; revised October 6, 2005. The work of S. M. Moser was supported in part by the Swiss Federal Institute of Technology (ETH) under Grant TH-23 02-2. The material in this paper was presented at the 2004 International Symposium on Information Theory and Its Applications, Parma, Italy.

A. Lapidoth is with the Department of Information Technology and Electrical Engineering, Swiss Federal Institute of Technology (ETH), CH-8092 Zurich, Switzerland (e-mail: lapidoth@isi.ee.ethz.ch).

S. M. Moser was with the Department of Information Technology and Electrical Engineering, Swiss Federal Institute of Technology (ETH), CH-8092 Zurich, Switzerland. He is now with the Department of Communication Engineering, National Chiao Tung University (NCTU), Hsinchu 300, Taiwan (e-mail: stefan.moser@ieee.org).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2005.862104

In this paper, we extend the results of [1] and derive the fading number for single-input multiple-output (SIMO) fading channels with memory.

What makes SIMO channels difficult to analyze is the fact that even at asymptotically high SNR, the capacity achieving output distribution is not memoryless. This makes it critical in the direct part to utilize future outputs even if the channel inputs associated with them are unknown. In the converse things are even more complicated because a naive application of the chain rule yields an upper bound that is not tight. One must first argue that capacity can be achieved by "almost" stationary channel inputs, and one must then use a new result about input distributions that escape to infinity. This result is not specific to fading channels and finds application also in the asymptotic analysis of the capacity of the phase-noise channel [2], [3], the direct-detection Poisson channel [4], [5], the free-space optical channel [6], and the input-dependent noise optical channel [6].

The paper is structured as follows. After concluding this introductory section with some notes on notation, we proceed in Section II to introduce the channel model and to define the fading number. Section III discusses the validity of the model and the utility of the fading number in studying practical communication systems. Section IV summarizes some relevant known results, while Section V provides the main new result, i.e., the fading number of a general SIMO fading channel with memory. The special case of *Gaussian* fading is then discussed in Section VI, which also includes the example of stationary mth-order autoregressive AR(m) Gaussian fading processes. Section VII contains the proof of the main result. The new observation regarding "input distributions that escape to infinity" can be found in Section VII-C3, which is essentially self-contained. Section VIII concludes the paper with a summary, a discussion, and an open problem.

Throughout the paper, $\hat{U}$ denotes a complex random variable that is uniformly distributed over the unit circle

$$\hat{U} \sim \text{uniform on } \{z \in \mathbb{C} : |z| = 1\}. \tag{1}$$

When it appears in formulas with other random variables, $\hat{U}$ is always assumed to be independent of these other variables. Similarly, we use $\{\hat{U}_\ell\}$ to denote an independent and identically distributed (i.i.d.) sequence of complex random variables, each of which is uniformly distributed on the set $\{z \in \mathbb{C} : |z| = 1\}$. In any expression involving this sequence of random variables it is assumed that the sequence is independent of any other variables appearing with it.

We generally try to denote random variables and random vectors by upper case letters and to denote their realization as well as deterministic constants by lower case letters. An exception is the signal-to-noise ratio SNR, which we capitalize and the energy-per-symbol, which we denote by $\mathcal{E}_{\mathrm{s}}$. Both are deterministic. We use boldface fonts to denote vectors, e.g., $\boldsymbol{x}$ for a deterministic vector and $\boldsymbol{X}$ for a random vector. We use the shorthand H_a^b for $(H_a, H_{a+1}, \ldots, H_b)$. For more complicated expressions, such as $(H_a\hat{U}_a, H_{a+1}\hat{U}_{a+1}, \ldots, H_b\hat{U}_b)$, we use the dummy variable ℓ to clarify notation: $\{H_\ell \hat{U}_\ell\}_{\ell=a}^b$.

II. The Channel Model and the Fading Number

We consider a SIMO fading channel whose time-k output $\boldsymbol{Y}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ is given by

$$\boldsymbol{Y}_k = \boldsymbol{H}_k x_k + \boldsymbol{Z}_k \tag{2}$$

where $x_k \in \mathbb{C}$ denotes the time-k channel input; the random vector $\boldsymbol{H}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ denotes the time-k fading vector; and where $\boldsymbol{Z}_k$ denotes additive noise. Here $\mathbb{C}$ denotes the complex field, $\mathbb{C}^{n_{\mathrm{R}}}$ denotes the n_{R}-dimensional complex Euclidean space, and n_{R} denotes the number of receive antennas. We assume that the additive noise is a zero-mean temporally and spatially white Gaussian process of covariance matrix $\sigma^2 \mathsf{I}_{n_{\mathrm{R}}}$, where $\sigma^2 > 0$ and where $\mathsf{I}_{n_{\mathrm{R}}}$ denotes the $n_{\mathrm{R}} \times n_{\mathrm{R}}$ identity matrix. Thus, $\{\boldsymbol{Z}_k\}$ is a zero-mean, circularly symmetric, stationary, multivariate, Gaussian process such that $\mathsf{E}\left[\boldsymbol{Z}_k \boldsymbol{Z}_{k+m}^\dagger\right]$ is the zero matrix if $m \neq 0$, and is $\sigma^2 \mathsf{I}_{n_{\mathrm{R}}}$ for $m = 0$. Here $()^\dagger$ denotes Hermitian conjugation.

As for the multivariate fading process $\{\boldsymbol{H}_k\}$, we shall only assume that it is stationary, ergodic, of finite second moment

$$\mathsf{E}\left[\|\boldsymbol{H}_k\|^2\right] < \infty \tag{3}$$

and of finite differential entropy rate

$$h(\{\boldsymbol{H}_k\}) > -\infty. \tag{4}$$

Finally, we assume that the fading process $\{\boldsymbol{H}_k\}$ and the additive noise process $\{\boldsymbol{Z}_k\}$ are independent and of a joint law that does not depend on the channel input $\{x_k\}$.

As for the input, we consider two different constraints: a peak-power constraint and an average-power constraint. We use $\mathcal{E}_{\mathrm{s}}$ 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

$$\mathrm{SNR} \triangleq \frac{\mathcal{E}_{\mathrm{s}}}{\sigma^2}.$$

The capacity $C(\mathrm{SNR})$ of the channel (2) is given by

$$C(\mathrm{SNR}) = \lim_{n\to\infty} \frac{1}{n} \sup I\left(X_1^n; \boldsymbol{Y}_1^n\right)$$

where the supremum is over the set of all probability distributions on X_1^n satisfying the constraints, i.e.,

$$|X_k|^2 \leq \mathcal{E}_{\mathrm{s}}, \quad \text{almost surely}, \qquad k = 1, 2, \ldots, n \tag{5}$$

for a peak constraint, or

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[|X_k|^2\right] \leq \mathcal{E}_{\mathrm{s}} \tag{6}$$

for an average constraint.

Specializing [1, Theorem 4.2] to SIMO fading, we have

$$\varlimsup_{\mathrm{SNR}\uparrow\infty} \Big\{C(\mathrm{SNR}) - \log\log \mathrm{SNR}\Big\} < \infty. \tag{7}$$

The fading number χ is now defined as in [1, Definition 4.6] by

$$\chi(\{\boldsymbol{H}_k\}) = \varlimsup_{\mathrm{SNR}\uparrow\infty} \Big\{C(\mathrm{SNR}) - \log\log \mathrm{SNR}\Big\}. \tag{8}$$

Prima facie the fading number depends on whether a peak-power constraint (5) or an average-power constraint (6) is imposed on the input. However, as we shall see, for SIMO fading channels the two constraints lead to identical fading numbers.

III. Motivation

Before proceeding with a presentation of the known and new results on the fading number, we pause here to discuss the validity of the channel model and the applicability of the fading number in the analysis of practical communication schemes. We shall also present some further motivation for this study.

A. On the Channel Model

There are four core assumptions at the heart of our channel model: that time is discrete, that the fading is stationary, that it is ergodic, and that it is regular in the sense that it is of finite entropy rate.

The assumption that time is discrete is very difficult to justify. Unlike the additive Gaussian noise channel [7], it is not at all clear that, using some Karhunen–Loève type expansion, one can reduce the channel to discrete time without any loss of optimality. Indeed, even the sampling theorem cannot quite justify this model because the multiplicative nature of the fading causes the channel output to be of larger bandwidth than the input. Thus, one should either analyze a model where the output is sampled faster than the input, or one where both are sampled at the higher rate. But in the latter case one has to account for the bandwidth constraint on the input. Neither of the above approaches has, so far, led to channel models that are amenable to an information-theoretic analysis. In this sense, the present work (and other information-theoretic analyses of fading channels) is just a step in the long journey toward an analysis of the continuous-time model. There is, however, one additional redeeming justification for the discrete-time model. The systems being built today often (suboptimally) reduce the channel at some point to discrete time by employing some form of quadrature amplitude modulation (QAM) and a matched filter. The capacity results for the discrete-time model can thus be interpreted as results on the ultimate limit on reliable communication using such systems.

The second assumption in our model is stationarity. It is a reflection of our state of knowledge about the system. It is not that we believe that the fading law on the weekend (when we are more likely to be in the suburbs) is identical to the law during the week (when we are more likely to be in the office), it is just that we are ignorant of the dependence of the fading on time. (Perhaps we can also argue that the channel is "essentially" stationary for the duration of our codeword, though this is problematic when we let the block length tend to infinity.) Stationarity is at the heart of many models, e.g., Jakes's model [8]. Nevertheless, a number of models analyzed by information theorists are not stationary [9]–[11]. These block-fading models are only cyclostationary. While they can be stationarized by random time jitters, this can only be achieved at the cost of making the fading non-Gaussian. Such models can be justified for frequency hopping systems or by mathematical convenience.

It should be emphasized that the reason the block-constant fading model [9], [10] leads to optimistic high-SNR capacity estimates (logarithmic growth versus the double logarithmic of [1]) is not in its inherent assumption that the fading in the different blocks are independent [12]. This assumption is *pessimistic* and any stationary fading can be made to appear to behave in this manner using interleaving, which is a special type of coding. These models lead to optimistic high-SNR capacity estimates because they assume that the fading within each fading block is rank deficient [11]. This assumption allows at high SNR for near-perfect estimation of the fading in a given fading block using a finite number of pilot tones sent in a subset of the block.

The ergodicity assumption we make reflects the assumption that we are allowing for the use of block codes of very large block lengths so that during the transmission of a codeword the channel "averages out." For low delay systems, this assumption may not be justified and more complicated models may be called for.

Finally, we address our assumption that the fading is regular, i.e., of finite entropy rate. As shown in [12], [13], the asymptotic behavior of channel capacity depends heavily on this assumption. For SISO Gaussian fading this assumption is equivalent to the assumption that the present fading cannot be perfectly predicted from the past fading. Indeed, it is the residual error in this prediction that causes capacity to grow only double logarithmically in the SNR. For SISO Gaussian fading, if this assumption is dropped capacity can, for example, grow double logarithmically, logarithmically, or as a fractional power thereof [12], [13]. In fact, if the set of harmonics in which the fading power spectral density is zero is of positive Lebesgue measure (e.g., if the fading is bandlimited) capacity grows logarithmically with a pre-log that was found in [12], [13].

For Gaussian fading channels, the finite entropy rate assumption reflects the intuition that nature is not fully predictable and that the fading therefore cannot be perfectly predicted. On the other hand, Jakes's model, albeit in continuous time, suggests a predictable model. It is felt that both the predictable and the non-predictable models are of interest. Ultimately, it is the precise dependence of the *noisy* prediction error on the noise variance in the SNR of interest that will determine which assumption is more useful [12], [13].

We draw the reader's attention to the fact that the fading processes we analyze are general in the sense that we do not assume that the fading is Gaussian. This is not merely done for the sake of generality. It allows for a robustness analysis of the results with respect to the Gaussian assumption. Moreover, it turns out that the results for general (non-Gaussian) SIMO fading are much needed in the derivation of lower bounds on the fading number of MIMO fading channels [14, Theorem 4]. Nevertheless, the Gaussian case is certainly of great interest and Section VI specializes the general results to this case.

B. On the Fading Number

The fading number χ (8) was introduced in [1] as the second-order term in the high-SNR expansion of channel capacity in order to assess the rates above which capacity begins to grow double logarithmically. Here, we would like to discuss the extent to which χ may be of use in the analysis of practical systems operating at finite SNRs. We will argue that, while no single number can characterize the entire capacity versus SNR curve, the fading number can give an indication of the maximal rate at which power-efficient communication is feasible. Moreover, we shall argue that, though it is defined as $\text{SNR} \to \infty$, it can be of relevance at moderate SNRs.

If the limsup in (8) is actually a limit, then above some threshold SNR_0

$$C(\text{SNR}) \approx \log(1 + \log(1 + \text{SNR})) + \chi, \qquad \text{SNR} \geq \text{SNR}_0. \tag{9}$$

The precise value of SNR_0 depends on how good we insist that the approximation (9) be. Roughly speaking, we shall say that the system is operating at "high SNR" if the approximation (9) is valid. We refer to SNR_0 as the threshold delineating the high-SNR regime. At SNRs above SNR_0, each additional increase of channel capacity by 1 bit requires squaring the SNR, i.e., doubling the decibel value of the SNR. In this sense, the high-SNR regime corresponds to power-inefficient communication, and should thus best be avoided by most system designers. How can we identify this regime?

One approach is to estimate SNR_0. This approach requires a finer analysis of channel capacity and was performed in [12], [13] for Gaussian SISO channels. It was suggested that for such systems SNR_0 satisfies

$$\epsilon^2(1/\text{SNR}_0) \approx 2\epsilon^2(0) \tag{10}$$

where $\epsilon^2(\delta^2)$ is the prediction error in predicting the present fading from past values of the fading contaminated by i.i.d. Gaussian noise of variance δ^2.

A different approach to characterizing the high-SNR regime is via the *rates* rather than via the SNR. That is, instead of estimating SNR_0 we estimate $C(\text{SNR}_0)$. To that end, let us attempt to pull ourselves by our bootstraps and assume that

$$30 \text{ dB} < \text{SNR}_0 < 80 \text{ dB} \tag{11}$$

or equivalently

$$2.1 < \log(1 + \log(1 + \text{SNR}_0)) < 3.0 \tag{12}$$

in nats. In this case, we can substitute SNR_0 for SNR in (9) to obtain *roughly* that

$$\chi + 2.1 < C(\mathrm{SNR}_0) < \chi + 3.0 \tag{13}$$

thus obtaining the rule of thumb that a system operating at rates that appreciably exceed $\chi + 2$ is probably operating in the high-SNR regime and is thus extremely power inefficient. No system is likely to be designed to operate at rates exceeding $\chi + 2$ nats.

Of course, one can construct pathological fading laws where (11) is violated. For such fading laws, the rule of thumb may not be applicable. (Hence "rule of thumb" as opposed to "theorem.")

This brings us to the issue of the relevance of χ, which is defined as $\mathrm{SNR} \to \infty$, to finite SNRs. We contend that χ is relevant whenever

$$C(\mathrm{SNR}) \approx \log(1 + \log(1 + \mathrm{SNR})) + \chi. \tag{14}$$

Even for very slowly varying channels this has recently been shown to hold at moderate SNRs in the range of 30–50 dB [14].

A point that must be emphasized is that in (14) either of the two terms on the right-hand side (RHS) can be the dominating term. For example, for memoryless SISO Rayleigh fading, the fading number is negative and is thus dominated by the double-logarithmic term at all SNRs. On the other hand, for slowly varying channels, χ is often the dominant term in all but the extremely high SNRs. Indeed, the double-logarithmic term dominates χ only for SNRs larger than

$$e^{e^{\chi}-1} - 1$$

which can be formidable for a system with large fading numbers, e.g., memoryless Rician-fading channels of large specular components or for slowly varying channels.

We emphasize that even for systems, such as those on slowly varying channels, for which the double-logarithmic term only dominates the fading number at extremely high SNRs, the approximation (9) may still hold at relatively moderate SNRs [14]. Indeed, while Etkin and Tse in [15], [16] did show that for slowly varying channels it may require extremely high SNRs for the approximation $C(\mathrm{SNR}) \approx \log(1 + \log(1 + \mathrm{SNR}))$ to hold, they did not argue that such extremely high SNRs are required for (9) (where χ is added to the RHS) to hold. And, indeed, this is not the case [14].

We have glanced over the issue of whether the limsup in (8) is indeed a limit. For SISO systems this is, indeed, the case [1], and our present contribution demonstrates that this is also the case for SIMO systems; see Theorem 1. Whether the limit holds for general MIMO channels is still open. Note, however, that since we defined χ as a limsup and not as a liminf our rule of thumb would still be plausible even if it turned out that the limit does not exist for general MIMO fading.

IV. Previous Results

Among the fading numbers computed in [1] are the fading numbers of SISO fading channels with memory [1, Theorem 4.41]

$$\chi(\{H_k\}) = \log \pi + \mathsf{E}\left[\log |H_1|^2\right] - h\big(\{H_k\}\big) \tag{15}$$

and the fading number for memoryless SIMO fading [1, Proposition 4.30]

$$\chi_{\text{i.i.d.}}(\boldsymbol{H}) = I\big(\hat{U}; \boldsymbol{H}\hat{U}\big) + \mathsf{E}\left[\log \|\boldsymbol{H}\|\right] - h\big(\|\boldsymbol{H}\| \,\big|\, \hat{\boldsymbol{H}}\hat{U}\big) - \log 2 \tag{16}$$

where $\hat{U}$ is defined in (1), and where $\hat{\boldsymbol{H}} = \boldsymbol{H}/\|\boldsymbol{H}\|$. Alternatively, $\chi_{\text{i.i.d.}}(\boldsymbol{H})$ can be expressed as

$$\chi_{\text{i.i.d.}}(\boldsymbol{H}) = h_\lambda(\hat{\boldsymbol{H}}\hat{U}) - h(\boldsymbol{H}) + n_{\mathrm{R}}\mathsf{E}\left[\log \|\boldsymbol{H}\|^2\right] - \log 2 \tag{17}$$

where h_λ is the differential entropy on the sphere, so that if a random vector $\boldsymbol{G}$ takes value on the unit sphere and has the density $f^{\lambda}_{\boldsymbol{G}}(\boldsymbol{g})$ with respect to the surface-area measure λ, then

$$h_\lambda(\boldsymbol{G}) = -\mathsf{E}\left[\log f^{\lambda}_{\boldsymbol{G}}(\boldsymbol{G})\right].$$

The above is extended in [1, Note 4.31] to the case where the receiver has access to some side information $\boldsymbol{S}$ such that $(\boldsymbol{H}, \boldsymbol{S})$ are independent of $\boldsymbol{Z}$, the joint law of $(\boldsymbol{H}, \boldsymbol{S}, \boldsymbol{Z})$ does not depend on the input, and the mutual information $I(\boldsymbol{H}; \boldsymbol{S})$ is finite

$$I(\boldsymbol{H}; \boldsymbol{S}) < \infty. \tag{18}$$

In this case

$$\chi_{\text{i.i.d.}}(\boldsymbol{H}|\boldsymbol{S}) = h_\lambda(\hat{\boldsymbol{H}}\hat{U}|\boldsymbol{S}) - h(\boldsymbol{H}|\boldsymbol{S}) + n_{\mathrm{R}}\mathsf{E}\left[\log \|\boldsymbol{H}\|^2\right] - \log 2. \tag{19}$$

Here $h_\lambda(\hat{\boldsymbol{H}}\hat{U}|\boldsymbol{S})$ should be interpreted as the expectation over $\boldsymbol{S}$ of $h_\lambda(\hat{\boldsymbol{H}}\hat{U}|\boldsymbol{S} = \boldsymbol{s})$, where $h_\lambda(\hat{\boldsymbol{H}}\hat{U}|\boldsymbol{S} = \boldsymbol{s})$ is the differential entropy on the sphere of the conditional law of $\hat{\boldsymbol{H}}\hat{U}$ given $\boldsymbol{S} = \boldsymbol{s}$. That is, if $\boldsymbol{G}$ takes value on the unit sphere and if conditional on $\boldsymbol{S} = \boldsymbol{s}$ it has the density $f^{\lambda}_{\boldsymbol{G}|\boldsymbol{S}}(\boldsymbol{g}|\boldsymbol{s})$ with respect to the surface-area measure λ on the sphere, then

$$h_\lambda(\boldsymbol{G}|\boldsymbol{S} = \boldsymbol{s}) = -\int f_{\boldsymbol{G}|\boldsymbol{S}}(\boldsymbol{g}|\boldsymbol{s}) \log f_{\boldsymbol{G}|\boldsymbol{S}}(\boldsymbol{g}|\boldsymbol{s})\, \mathrm{d}\lambda(\boldsymbol{g}) \tag{20}$$

and

$$h_\lambda(\boldsymbol{G}|\boldsymbol{S}) = \int h_\lambda(\boldsymbol{G}|\boldsymbol{S} = \boldsymbol{s}) \mathrm{d}P_{\boldsymbol{S}}(\boldsymbol{s}). \tag{21}$$

It is further shown in [1, Sec. IV-D.8] that for the case of MIMO fading where the $n_{\mathrm{R}} \times n_{\mathrm{T}}$ random fading matrix $\mathbb{H}$ is of the form

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

where D is a deterministic $n_{\mathrm{R}} \times n_{\mathrm{T}}$ matrix and $\tilde{\mathbb{H}}$ is a random $n_{\mathrm{R}} \times n_{\mathrm{T}}$ matrix of i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ components, the fading number can be bounded as [1, eqs. (124) and (128)]

$$\chi_{\text{i.i.d.}}(\mathsf{D} + \tilde{\mathbb{H}}) \geq \log \|\mathsf{D}\|^2 - \mathrm{Ei}\big(-\|\mathsf{D}\|^2\big) - 1 \tag{22}$$

$$\chi_{\text{i.i.d.}}(\mathsf{D} + \tilde{\mathbb{H}}) \leq \min\{n_{\mathrm{R}}, n_{\mathrm{T}}\} \log\left(1 + \frac{\|\mathsf{D}\|^2}{\min\{n_{\mathrm{R}}, n_{\mathrm{T}}\}}\right) + n_{\mathrm{R}} \log n_{\mathrm{R}} - n_{\mathrm{R}} - \log \Gamma(n_{\mathrm{R}}) \tag{23}$$

where $\|\cdot\|$ denotes the matrix operator norm; $\mathrm{Ei}(\cdot)$ denotes the exponential integral function

$$\mathrm{Ei}(-\xi) = -\int_{\xi}^{\infty} \frac{e^{-t}}{t} \mathrm{d}t, \qquad \xi > 0; \tag{24}$$

$\Gamma(\cdot)$ denotes the Gamma function so that $\Gamma(n_{\mathrm{R}}) = (n_{\mathrm{R}} - 1)!$; and the term $\log(\xi) - \mathrm{Ei}(-\xi)$ is understood to take on the value $-\gamma$ at $\xi = 0$. (Here $\gamma \approx 0.577$ denotes Euler's constant.) This specializes for the SIMO case to

$$\chi_{\text{i.i.d.}}(\boldsymbol{d} + \tilde{\boldsymbol{H}}) \geq \log \|\boldsymbol{d}\|^2 - \mathrm{Ei}\big(-\|\boldsymbol{d}\|^2\big) - 1, \tag{25}$$

$$\chi_{\text{i.i.d.}}(\boldsymbol{d}+\tilde{\boldsymbol{H}}) \leq \log(1+\|\boldsymbol{d}\|^2) + n_{\text{R}}\log n_{\text{R}} - n_{\text{R}} - \log\Gamma(n_{\text{R}}) \tag{26}$$

where the n_{R} components of $\tilde{\boldsymbol{H}}$ are i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$. More generally, if $\boldsymbol{H}$ is a multivariate circularly symmetric complex Gaussian of mean $\boldsymbol{d}$ and covariance Σ

$$\boldsymbol{H} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{d}, \Sigma)$$

where Σ has eigenvalues $\lambda^{(1)},\ldots,\lambda^{(n_{\text{R}})}$, then

$$\chi_{\text{i.i.d.}}(\boldsymbol{H}) \geq \log\sum_{r=1}^{n_{\text{R}}}\left|\frac{(\mathsf{V}^{\mathsf{T}}\boldsymbol{d})^{(r)}}{\sqrt{\lambda^{(r)}}}\right|^2 - \operatorname{Ei}\left(-\sum_{r=1}^{n_{\text{R}}}\left|\frac{(\mathsf{V}^{\mathsf{T}}\boldsymbol{d})^{(r)}}{\sqrt{\lambda^{(r)}}}\right|^2\right) - 1 \tag{27}$$

$$\chi_{\text{i.i.d.}}(\boldsymbol{H}) \leq \log\left(1+\sum_{r=1}^{n_{\text{R}}}\left|\frac{(\mathsf{V}^{\mathsf{T}}\boldsymbol{d})^{(r)}}{\sqrt{\lambda^{(r)}}}\right|^2\right) + n_{\text{R}}\log n_{\text{R}} - n_{\text{R}} - \log\Gamma(n_{\text{R}}) \tag{28}$$

where V is unitary and diagonalizes Σ:

$$\Sigma\mathsf{V} = \mathsf{V}\operatorname{diag}\left(\lambda^{(1)},\ldots,\lambda^{(n_{\text{R}})}\right).$$

This follows because, by [1, Lemma 4.7], for any nonsingular deterministic $n_{\text{R}} \times n_{\text{R}}$ matrix G

$$\chi(\mathsf{G}\mathbb{H}) = \chi(\mathbb{H}).$$

The choice

$$\mathsf{G} = \operatorname{diag}\left(\frac{1}{\sqrt{\lambda^{(1)}}},\ldots,\frac{1}{\sqrt{\lambda^{(n_{\text{R}})}}}\right)\cdot\mathsf{V}^{\mathsf{T}}$$

leads to a fading vector $\mathsf{G}\boldsymbol{H}$ with components that are i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$, to which the above results can be applied.

In particular, if Σ is diagonal

$$\Sigma = \operatorname{diag}\left(\lambda^{(1)},\ldots,\lambda^{(n_{\text{R}})}\right)$$

then by (27) and (28)

$$\chi_{\text{i.i.d.}}(\boldsymbol{H}) \geq \log\sum_{r=1}^{n_{\text{R}}}\frac{|d^{(r)}|^2}{\lambda^{(r)}} - \operatorname{Ei}\left(-\sum_{r=1}^{n_{\text{R}}}\frac{|d^{(r)}|^2}{\lambda^{(r)}}\right) - 1, \quad \Sigma = \operatorname{diag}\left(\{\lambda^{(r)}\}_{r=1}^{n_{\text{R}}}\right) \tag{29}$$

$$\chi_{\text{i.i.d.}}(\boldsymbol{H}) \leq \log\left(1+\sum_{r=1}^{n_{\text{R}}}\frac{|d^{(r)}|^2}{\lambda^{(r)}}\right) + n_{\text{R}}\log n_{\text{R}} - n_{\text{R}} - \log\Gamma(n_{\text{R}}), \quad \Sigma = \operatorname{diag}\left(\{\lambda^{(r)}\}_{r=1}^{n_{\text{R}}}\right). \tag{30}$$

V. Main Result

Theorem 1: Consider a SIMO fading channel with memory (2) where the stationary and ergodic fading process $\{\boldsymbol{H}_k\}$ takes value in $\mathbb{C}^{n_{\text{R}}}$ and satisfies $h(\{\boldsymbol{H}_k\}) > -\infty$ and $\mathsf{E}\left[\|\boldsymbol{H}_k\|^2\right] < \infty$. Then, irrespective of whether a peak-power constraint (5) or an average-power constraint (6) is imposed on the input, the limsup in (8) is in fact a limit, and the fading number $\chi(\{\boldsymbol{H}_k\})$ is given by

$$\boxed{\chi(\{\boldsymbol{H}_k\}) = \chi_{\text{i.i.d.}}\left(\boldsymbol{H}_0 \,\middle|\, \boldsymbol{H}_{-\infty}^{-1}, \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\infty}\right).} \tag{31}$$

Here $\chi_{\text{i.i.d.}}(\boldsymbol{H}_0|\boldsymbol{S})$ is given in (19), the random process $\{\hat{U}_\ell\}$ is independent of $\{\boldsymbol{H}_k\}$ and constitutes of i.i.d. random variables that are uniformly distributed over the complex sphere, i.e.,

$$\hat{U}_\ell \sim \text{uniform on } \{z\in\mathbb{C} : |z| = 1\}$$

and $\hat{\boldsymbol{H}}_\ell$ is defined as

$$\hat{\boldsymbol{H}}_\ell \triangleq \frac{\boldsymbol{H}_\ell}{\|\boldsymbol{H}_\ell\|}, \qquad \forall \ell\in\mathbb{Z}.$$

Equivalently, the fading number is given by

$$\chi(\{\boldsymbol{H}_k\}) = \chi_{\text{i.i.d.}}(\boldsymbol{H}_0) + I(\boldsymbol{H}_0; \boldsymbol{H}_{-\infty}^{-1}) - I\left(\hat{\boldsymbol{H}}_0\hat{U}_0; \{\hat{\boldsymbol{H}}_\ell\hat{U}_\ell\}_{\ell=-\infty}^{-1}\right) \tag{32}$$

where $\chi_{\text{i.i.d.}}(\boldsymbol{H}_0)$ is defined in (17).

Moreover, this asymptotic behavior is achievable at high SNR by i.i.d. circularly symmetric inputs $\{X_k\}$ such that

$$\log|X_k|^2 \sim \text{uniform on } [\log\log\mathcal{E}_{\text{s}}, \log\mathcal{E}_{\text{s}}]. \tag{33}$$

Proof: See Section VII. □

Corollary 2: From Theorem 1 it follows that

$$\chi_{\text{i.i.d.}}\left(\boldsymbol{H}_0\middle|\boldsymbol{H}_{-\infty}^{-1}\right) \leq \chi(\{\boldsymbol{H}_k\}) \leq \chi_{\text{i.i.d.}}\left(\boldsymbol{H}_0\middle|\boldsymbol{H}_{-\infty}^{-1}, \boldsymbol{H}_1^{\infty}\right). \tag{34}$$

Remark 3: We can always lower-bound the capacity of a SIMO fading channel with memory (even with correlation between the antennas) by linearly combining the outputs of the n_{R} receive antennas and by then lower-bounding the capacity of the resulting SISO channel. In this way, we can use the expression for the fading number of SISO channels with memory (15) to obtain a lower bound on the fading number of a SIMO system

$$\chi(\{\boldsymbol{H}_k\}) \geq \sup\left\{\log\pi + \mathsf{E}\left[\log|\tilde{H}_1|^2\right] - h\left(\{\tilde{H}_k\}\right)\right\} \tag{35}$$

where $\tilde{H}_k = \sum_{r=1}^{n_{\text{R}}}\alpha^{(r)}H_k^{(r)}$ and where the supremum is over all linear combiners, i.e., over all $\alpha^{(1)},\ldots,\alpha^{(n_{\text{R}})}$ that fulfill $\sum_{r=1}^{n_{\text{R}}}|\alpha^{(r)}|^2 = 1$. This bound is generally not tight.

VI. Gaussian Fading With Memory

Since it is difficult to evaluate analytically the fading number (31) even for Gaussian fading, we shall next use the bounds (34) to approximate it. We shall only treat here the case of "spatially independent fading," i.e., the case where the fading processes experienced by the different links between the transmit antenna and the different receive antennas are statistically independent. That is, the n_{R} processes

$$\{H_k^{(1)}\}_{k=-\infty}^{\infty}, \{H_k^{(2)}\}_{k=-\infty}^{\infty}, \ldots, \{H_k^{(n_{\text{R}})}\}_{k=-\infty}^{\infty}$$

are independent.[1]

Let then $\boldsymbol{\mu}\in\mathbb{C}^{n_{\text{R}}}$ denote the mean vector of the stationary vector-valued fading process $\{\boldsymbol{H}_k\}$, and assume that $\{\boldsymbol{H}_k - \boldsymbol{\mu}\}$ is a stationary circularly symmetric vector-valued Gaussian process with a diagonal spectral distribution matrix

$$\mathsf{F} = \operatorname{diag}\left(\mathcal{F}^{(1)},\ldots,\mathcal{F}^{(n_{\text{R}})}\right).$$

Thus, the n_{R} components of the vector-valued process $\{\boldsymbol{H}_k\}$ are independent, and for each $1 \leq r \leq n_{\text{R}}$ the process

[1] In the more general case one may still resort to (35) which is, however, not tight.

$\{H_k^{(r)} - \mu^{(r)}\}_{k=-\infty}^{\infty}$ is a stationary circularly symmetric scalar Gaussian process of spectral distribution $\mathcal{F}^{(r)}$ so that

$$\mathsf{E}\left[\left(H_k^{(r)} - \mu^{(r)}\right)\left(H_{k+m}^{(r)} - \mu^{(r)}\right)^*\right] = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{i2\pi m\lambda} \mathrm{d}\mathcal{F}^{(r)}(\lambda).$$

Denote the derivative of $\mathcal{F}^{(r)}(\cdot)$ by $\mathcal{F}'^{(r)}(\cdot)$.

To evaluate the lower bound of (34) on the fading number we shall need the conditional law of the present fading given its past. To this end, we recall that the optimum *prediction error* in estimating $H_0^{(r)}$ from its infinite past $\{H_\ell^{(r)}\}_{\ell=-\infty}^{-1}$ is the optimum linear prediction error which is given by (see, e.g., [17], [18])

$$\epsilon_{\mathrm{pred},r}^2 = \exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \log \mathcal{F}'^{(r)}(\lambda)\mathrm{d}\lambda\right). \tag{36}$$

Moreover, conditional on $\{H_\ell^{(r)} = h_\ell^{(r)}\}_{\ell=-\infty}^{-1}$, the distribution of $H_0^{(r)}$ is Gaussian of mean

$$h_{\mathrm{pred}}^{(r)}\left(\{h_\ell^{(r)}\}_{\ell=-\infty}^{-1}\right) = \mathsf{E}\left[H_0^{(r)} \,\middle|\, \{H_\ell^{(r)} = h_\ell^{(r)}\}_{\ell=-\infty}^{-1}\right]$$

and of variance $\epsilon_{\mathrm{pred},r}^2$. Unconditionally, $H_{\mathrm{pred}}^{(r)}$ is Gaussian of mean $\mu^{(r)}$ and of variance

$$\begin{aligned}\mathsf{Var}\left(H_{\mathrm{pred}}^{(r)}\right) &= \mathsf{Var}\left(H_0^{(r)}\right) - \epsilon_{\mathrm{pred},r}^2 \\ &= \mathcal{F}^{(r)}(1/2) - \mathcal{F}^{(r)}(-1/2) - \epsilon_{\mathrm{pred},r}^2.\end{aligned}$$

Similarly, to evaluate the upper bound of (34) on the fading number we shall need the conditional law of the present fading given its past and future. To this end, we recall that the optimum *interpolation error* in estimating $H_0^{(r)}$ from its infinite past and future

$$\left(\{H_\ell^{(r)}\}_{\ell=-\infty}^{-1}, \{H_\ell^{(r)}\}_{\ell=1}^{\infty}\right)$$

is the optimum linear interpolation error given by (see [18, Sec. 37.2]–[20])

$$\epsilon_{\mathrm{int},r}^2 = \frac{1}{\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{\mathcal{F}'^{(r)}(\lambda)}\mathrm{d}\lambda}. \tag{37}$$

Moreover, conditional on

$$\left(\{H_\ell^{(r)} = h_\ell^{(r)}\}_{\ell=-\infty}^{-1}, \{H_\ell^{(r)} = h_\ell^{(r)}\}_{\ell=1}^{\infty}\right)$$

the distribution of $H_0^{(r)}$ is Gaussian of mean

$$\begin{aligned}&h_{\mathrm{int}}^{(r)}\left(\{h_\ell^{(r)}\}_{\ell=-\infty}^{-1}, \{h_\ell^{(r)}\}_{\ell=1}^{\infty}\right) \\ &\quad = \mathsf{E}\left[H_0^{(r)} \,\middle|\, \{H_\ell^{(r)} = h_\ell^{(r)}\}_{\ell=-\infty}^{-1}, \{H_\ell^{(r)} = h_\ell^{(r)}\}_{\ell=1}^{\infty}\right]\end{aligned}$$

and of variance $\epsilon_{\mathrm{int},r}^2$. Unconditionally, $H_{\mathrm{int}}^{(r)}$ is Gaussian of mean $\mu^{(r)}$ and of variance

$$\begin{aligned}\mathsf{Var}\left(H_{\mathrm{int}}^{(r)}\right) &= \mathsf{Var}\left(H_0^{(r)}\right) - \epsilon_{\mathrm{int},r}^2 \\ &= \mathcal{F}^{(r)}(1/2) - \mathcal{F}^{(r)}(-1/2) - \epsilon_{\mathrm{int},r}^2.\end{aligned}$$

Since we have assumed that the components of $\boldsymbol{H}_k$ are independent, we can use (29) and (30) to further bound the expressions in (34). We start with the upper bound

$$\begin{aligned}\chi(\{\boldsymbol{H}_k\}) &\le \chi_{\mathrm{i.i.d.}}\left(\boldsymbol{H}_0 \middle| \boldsymbol{H}_{-\infty}^{-1}, \boldsymbol{H}_1^{\infty}\right) \\ &\le \mathsf{E}\left[\log\left(1 + \sum_{r=1}^{n_\mathrm{R}} \frac{\left|H_{\mathrm{int}}^{(r)}\right|^2}{\epsilon_{\mathrm{int},r}^2}\right)\right] + n_\mathrm{R}\log n_\mathrm{R} \\ &\quad - n_\mathrm{R} - \log\Gamma(n_\mathrm{R}) \\ &\le \log\left(1 + \sum_{r=1}^{n_\mathrm{R}} \frac{\mathsf{E}\left[\left|H_{\mathrm{int}}^{(r)}\right|^2\right]}{\epsilon_{\mathrm{int},r}^2}\right) + n_\mathrm{R}\log n_\mathrm{R} \\ &\quad - n_\mathrm{R} - \log\Gamma(n_\mathrm{R}) \\ &= \log\left(1 + \sum_{r=1}^{n_\mathrm{R}} \frac{\mathsf{Var}\left(H_0^{(r)}\right) - \epsilon_{\mathrm{int},r}^2 + |\mu^{(r)}|^2}{\epsilon_{\mathrm{int},r}^2}\right) \\ &\quad + n_\mathrm{R}\log n_\mathrm{R} - n_\mathrm{R} - \log\Gamma(n_\mathrm{R}).\end{aligned} \tag{38}$$

Here, the first inequality is due to (34); the second inequality follows from (30); and the third inequality follows from Jensen's inequality.

For the lower bound we get

$$\begin{aligned}&\chi(\{\boldsymbol{H}_k\}) \\ &\ge \chi_{\mathrm{i.i.d.}}\left(\boldsymbol{H}_0 \middle| \boldsymbol{H}_{-\infty}^{-1}\right) \\ &\ge \mathsf{E}\left[\log \sum_{r=1}^{n_\mathrm{R}} \frac{\left|H_{\mathrm{pred}}^{(r)}\right|^2}{\epsilon_{\mathrm{pred},r}^2}\right] - \mathsf{E}\left[\mathrm{Ei}\left(-\sum_{r=1}^{n_\mathrm{R}} \frac{\left|H_{\mathrm{pred}}^{(r)}\right|^2}{\epsilon_{\mathrm{pred},r}^2}\right)\right] - 1 \\ &\ge \mathsf{E}\left[\log \sum_{r=1}^{n_\mathrm{R}} \frac{\left|H_{\mathrm{pred}}^{(r)}\right|^2}{\epsilon_{\mathrm{pred},r}^2}\right] - \mathrm{Ei}\left(-\sum_{r=1}^{n_\mathrm{R}} \frac{\mathsf{E}\left[\left|H_{\mathrm{pred}}^{(r)}\right|^2\right]}{\epsilon_{\mathrm{pred},r}^2}\right) - 1\end{aligned} \tag{39}$$

where the first inequality is due to (34); the second inequality follows from (29); and where the last inequality follows from Jensen's inequality. The analytic computation of the RHS of (39) is greatly simplified if each component process $\{H_k^{(r)}\}$ of the vector-valued fading process $\{\boldsymbol{H}_k\}$ is of an identical law, which in our case means that

$$\mathsf{F} = \mathrm{diag}\left(\tilde{\mathcal{F}}, \ldots, \tilde{\mathcal{F}}\right) \tag{40}$$

for a scalar spectral distribution function $\tilde{\mathcal{F}}$, and

$$\boldsymbol{\mu} = \left(\mu^{(1)}, \ldots, \mu^{(n_\mathrm{R})}\right)^\mathsf{T} = (\tilde{\mu}, \ldots, \tilde{\mu})^\mathsf{T} \tag{41}$$

for a mean $\tilde{\mu}$. In that case (using the expectation of the logarithm of a noncentral χ^2-distribution [1, Appendix X]) we obtain

$$\chi(\{\boldsymbol{H}_k\}) \ge \log \frac{\mathsf{Var}\left(H_0^{(1)}\right) - \epsilon_{\mathrm{pred}}^2}{\epsilon_{\mathrm{pred}}^2} - 1$$

$$+ g_{n_{\mathrm{R}}}\left(\frac{n_{\mathrm{R}}|\tilde{\mu}|^2}{\mathsf{Var}\left(H_0^{(1)}\right) - \epsilon_{\mathrm{pred}}^2}\right) - \mathrm{Ei}\left(-n_{\mathrm{R}}\frac{\mathsf{Var}\left(H_0^{(1)}\right) - \epsilon_{\mathrm{pred}}^2 + |\tilde{\mu}|^2}{\epsilon_{\mathrm{pred}}^2}\right) \quad (42)$$

where $g_m(\cdot)$ is defined as [1]

$$g_m(z) = \log z - \mathrm{Ei}(-z) + \sum_{j=1}^{m-1}(-1)^j\left(e^{-z}(j-1)! - \frac{(m-1)!}{j(m-1-j)!}\right)z^{-j}, \quad z \geq 0. \quad (43)$$

(Here, as before, the term $\log z - \mathrm{Ei}(-z)$ should be interpreted as minus Euler's constant at $z = 0$.)

Note that the simplifying assumptions (40) and (41) are not necessary if one resorts to the weaker lower bound described in Remark 3.

Example 4: Suppose that the fading process $\{\boldsymbol{H}_k\}$ is spatially i.i.d. so that the processes

$$\{H_k^{(1)}\}_{k=-\infty}^{\infty}, \{H_k^{(2)}\}_{k=-\infty}^{\infty}, \ldots, \{H_k^{(n_{\mathrm{R}})}\}_{k=-\infty}^{\infty}$$

are independent of each other and of identical (not necessarily temporally i.i.d.) law. Suppose that under this law, $\{H_k^{(r)}\}$ is a stationary, unit-variance, zero-mean, circularly symmetric, mth-order autoregressive AR(m) Gaussian process. That is, for all $1 \leq r \leq n_{\mathrm{R}}$

$$H_k^{(r)} = W_k^{(r)} - a_1 H_{k-1}^{(r)} - a_2 H_{k-2}^{(r)} - \cdots - a_m H_{k-m}^{(r)}. \quad (44)$$

Here $\{W_k^{(r)}\}$ is temporally i.i.d. $\mathcal{N}_{\mathbb{C}}(0, \varepsilon^2)$, where ε^2 denotes the innovation variance; the coefficients $a_1, \ldots, a_m$ satisfy the stability condition [21]

$$\sum_{j=1}^{m} a_j z^j \neq -1 \qquad \forall |z| \leq 1; \quad (45)$$

and ε^2 and $a_1, \ldots, a_m$ are such that

$$\mathsf{Var}\left(H_k^{(r)}\right) = 1. \quad (46)$$

Then [21]

$$\epsilon_{\mathrm{pred},r}^2 = \varepsilon^2 \quad (47)$$

$$H_{\mathrm{pred}}^{(r)} \sim \mathcal{N}_{\mathbb{C}}\left(0, 1 - \epsilon_{\mathrm{pred},r}^2\right) \quad (48)$$

$$\epsilon_{\mathrm{int},r}^2 = \frac{\varepsilon^2}{1 + \sum_{j=1}^{m}|a_j|^2} \quad (49)$$

$$H_{\mathrm{int}}^{(r)} \sim \mathcal{N}_{\mathbb{C}}\left(0, 1 - \epsilon_{\mathrm{int},r}^2\right) \quad (50)$$

which yields

$$\chi(\{\boldsymbol{H}_k\}) \geq \log\frac{1-\varepsilon^2}{\varepsilon^2} + \psi(n_{\mathrm{R}}) - \mathrm{Ei}\left(-n_{\mathrm{R}}\frac{1-\varepsilon^2}{\varepsilon^2}\right) - 1 \quad (51)$$

$$\chi(\{\boldsymbol{H}_k\}) \leq \log\left(1 + n_{\mathrm{R}}\frac{1 + \sum_{j=1}^{m}|a_j|^2 - \varepsilon^2}{\varepsilon^2}\right) + n_{\mathrm{R}}\log n_{\mathrm{R}} - n_{\mathrm{R}} - \log\Gamma(n_{\mathrm{R}}) \quad (52)$$

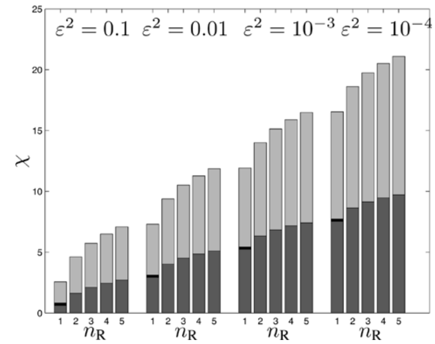

Fig. 1. The bounds (51) and (53) on the fading number for spatially white first-order Gauss–Markov fading laws of various innovation variances and for different numbers of receiver antennas. For the SISO case ($n_{\mathrm{R}} = 1$) the exact fading number is also depicted.

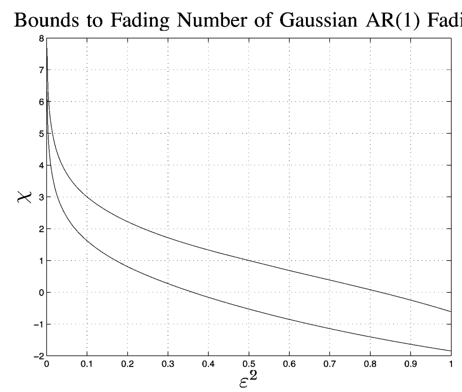

Fig. 2. Upper and lower bounds on the fading number χ for a zero-mean SIMO Gauss–Markov fading channel with memory 1 (AR(1)) and two receive antennas plotted as functions of the prediction error ε^2. The components of the fading vector $\boldsymbol{H}_k$ are assumed to be independent and identically distributed with variance 1.

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

$$\psi(n_{\mathrm{R}}) = -\gamma + \sum_{j=1}^{n_{\mathrm{R}}-1}\frac{1}{j}$$

and γ denotes Euler's constant.

For the case of Gauss–Markov fading ($m = 1$, $a_1 = -\sqrt{1-\varepsilon^2}$), the lower bound (51) is unchanged and the upper bound (52) becomes

$$\chi(\{\boldsymbol{H}_k\}) \leq \log\left(1 + 2n_{\mathrm{R}}\frac{1-\varepsilon^2}{\varepsilon^2}\right) + n_{\mathrm{R}}\log n_{\mathrm{R}} - n_{\mathrm{R}} - \log\Gamma(n_{\mathrm{R}}). \quad (53)$$

See Fig. 1 for a plot of the bounds (51) and (53) for various values of the innovation variance ϵ^2 of the Gauss–Markov process and for various numbers of receiver antennas n_{R}. Fig. 2 depicts these bounds as functions of the innovation variance ϵ^2 of the Gauss–Markov process for two receiver antennas $n_{\mathrm{R}} = 2$.

Note that for very slowly varying channels, i.e., for $\varepsilon^2 \ll 1$, one obtains from (51) and (53) the asymptotic bounds

$$\chi(\{\boldsymbol{H}_k\}) \geq \log\frac{1}{\varepsilon^2} + \psi(n_{\mathrm{R}}) - 1 + o(\varepsilon^2) \quad (54)$$

$$\chi(\{\boldsymbol{H}_k\}) \leq \log\frac{1}{\varepsilon^2} + \log 2 + (n_{\mathrm{R}}+1)\log n_{\mathrm{R}} - n_{\mathrm{R}} - \log\Gamma(n_{\mathrm{R}}) + o(\varepsilon^2) \quad (55)$$

where $o(\varepsilon^2)$ tends to zero as ε^2 tends to zero.

VII. Proof of Theorem 1

A. Proof Outline

The proof of Theorem 1 has three components. The first is an achievability result ("direct part") (Section VII-B) which provides a lower bound on channel capacity and hence a lower bound on the fading number. This lower bound is the RHS of (31). The second component is a "converse" (Section VII-C), which provides an upper bound on channel capacity, and hence an upper bound on the fading number. This upper bound is the RHS of (32). Finally, the last component (Appendix I) is a demonstration that the lower and upper bounds are in fact identical.

The inputs that are used to demonstrate the achievability of the RHS of (31) are peak-limited, whereas the converse is proved under an average-power constraint. Thus, the result for the fading number does not depend on the type of power constraint that is imposed.

B. The Direct Part

1) An Overview: The lower bound is based on choosing the input symbols to be i.i.d., circularly symmetric, with

$$\log |X_k|^2 \sim \text{uniform on } [\log x_{\min}^2, \log \mathcal{E}_{\rm s}]$$

where we choose $x_{\min}^2$ as[2]

$$x_{\min}^2 = \log \mathcal{E}_{\rm s}.$$

The motivation for using i.i.d. inputs is that it greatly simplifies the analysis and that our intuition (gained from the study of additive colored Gaussian noise channels [7] and from the study of SISO fading channels with memory [1]) is that at high SNR very little is to be gained from introducing memory into the input. In fact, we suspect that this is the case also for MIMO fading, but we have no proof of that.

The choice of the marginal distribution is motivated by two nice properties that it possesses. The first is that—irrespective of the partial side information at the receiver (assumed of finite mutual information with the fading)—this input distribution has been shown [1] to achieve the fading number of the memoryless SIMO fading channel. The second property has to do with "identification." Because with probability one $|X_k| \geq x_{\min}$ and because $x_{\min}$ tends to infinity (albeit slowly), it follows that at very high SNR we can identify the time-k fading vector with great accuracy by observing the time-k input X_k and the time-k output $\boldsymbol{Y}_k$. Indeed, in this regime, an excellent estimator for $\boldsymbol{H}_k$ is the estimator $\boldsymbol{Y}_k/X_k$. The other "identification" that this input distribution allows has to do with inference on $\boldsymbol{H}_k$ based on the channel output $\boldsymbol{Y}_k$ alone, i.e., when we know the channel output but not the corresponding input. In this scenario, our chosen input distribution allows us (at high SNR) to accurately estimate the "direction" of $\boldsymbol{H}_k$, namely, $\boldsymbol{H}_k/\|\boldsymbol{H}_k\|$, to within a multiple by a scalar complex random variable of unit magnitude and uniform phase. For this identification the estimator $\boldsymbol{Y}_k/\|\boldsymbol{Y}_k\|$ is most suitable. Indeed, while the circular symmetry of the input X_k renders the phase information in $\boldsymbol{Y}_k$ useless, the fact that X_k is, with probability one, very large guarantees that the additive noise has hardly any detrimental effect on the estimator, and the direction of $\boldsymbol{H}_k$ is—to within a random phase—almost identical to the direction of $\boldsymbol{Y}_k$.

[2] In fact, any choice of $x_{\min} = x_{\min}(\mathcal{E}_{\rm s})$ such that $x_{\min}(\mathcal{E}_{\rm s}) \to \infty$ as $\mathcal{E}_{\rm s} \to \infty$ and such that $\log x_{\min}^2(\mathcal{E}_{\rm s})/\log \mathcal{E}_{\rm s} \to 0$ as $\mathcal{E}_{\rm s} \to \infty$ would work.

The proof of the lower bound thus proceeds heuristically as follows: since the inputs are i.i.d., it follows from the chain rule that

$$\begin{aligned}\frac{1}{n} I\left(X_1^n; \boldsymbol{Y}_1^n\right) &= \frac{1}{n}\sum_{k=1}^{n} I\left(X_k; \boldsymbol{Y}_1^n | X_1^{k-1}\right)\\ &= \frac{1}{n}\sum_{k=1}^{n} I\left(X_k; \boldsymbol{Y}_1^n, X_1^{k-1}\right).\end{aligned}$$

We now analyze the individual terms in the sum

$$\begin{aligned}&I\left(X_k; \boldsymbol{Y}_1^n, X_1^{k-1}\right)\\ &\quad\approx I\left(X_k; \boldsymbol{H}_1^{k-1}, \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=k+1}^n, \boldsymbol{Y}_1^n, X_1^{k-1}\right)\\ &\quad= I\left(X_k; \boldsymbol{H}_1^{k-1}, \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=k+1}^n, \boldsymbol{Y}_k\right)\\ &\quad= I\left(X_k; \boldsymbol{Y}_k \middle| \boldsymbol{H}_1^{k-1}, \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=k+1}^n\right)\end{aligned}$$

which has the general form of a memoryless SIMO fading channel with side information consisting of the past fading vectors and the future fading "directions" corrupted by a random phase. The key here is the above approximation, which hinges on estimating the past fading $\boldsymbol{H}_1^{k-1}$ from the past inputs and outputs $(X_1^{k-1}, \boldsymbol{Y}_1^{k-1})$, and on estimating the future "directions" $\{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=k+1}^n$ based on the available future outputs $\boldsymbol{Y}_{k+1}^n$ (without their corresponding inputs). Note that if we were to ignore these future outputs we would not attain the fading number.

2) Proof of the Lower Bound: In this subsection, we derive a lower bound to capacity and use it to show that the RHS of (31) is a lower bound to the fading number. Let $\{X_k\}$ be i.i.d. circularly symmetric random variables with

$$\log |X_k|^2 \sim \text{uniform on } [\log x_{\min}^2, \log \mathcal{E}_{\rm s}] \tag{56}$$

where

$$x_{\min}^2 = \log \mathcal{E}_{\rm s}. \tag{57}$$

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

$$\frac{1}{n} I\left(X_1^n; \boldsymbol{Y}_1^n\right) = \frac{1}{n}\sum_{k=1}^{n} I\left(X_k; \boldsymbol{Y}_1^n | X_1^{k-1}\right) \tag{58}$$

$$\geq \frac{1}{n}\sum_{k=\kappa+1}^{n-\kappa} I\left(X_k; \boldsymbol{Y}_1^n | X_1^{k-1}\right). \tag{59}$$

Then for any $\kappa + 1 \leq k \leq n - \kappa$, we can use the fact that $\{X_k\}$ are i.i.d. and circularly symmetric to lower-bound $I\left(X_k; \boldsymbol{Y}_1^n | X_1^{k-1}\right)$ as follows:

$$\begin{aligned}&I\left(X_k; \boldsymbol{Y}_1^n \middle| X_1^{k-1}\right)\\ &\quad= I\left(X_k; X_1^{k-1}, \boldsymbol{Y}_1^n\right) &(60)\\ &\quad\geq I\left(X_k; X_{k-\kappa}^{k-1}, \boldsymbol{Y}_{k-\kappa}^{k-1}, \boldsymbol{Y}_k, \boldsymbol{Y}_{k+1}^{k+\kappa}\right) &(61)\end{aligned}$$

$$= I\big(X_k; X_{k-\kappa}^{k-1}, \mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{Z}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k+1}^{k+\kappa}\big) - \underbrace{I\big(X_k; \mathbf{Z}_{k-\kappa}^{k-1} \big| X_{k-\kappa}^{k-1}, \mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k+1}^{k+\kappa}\big)}_{\leq \epsilon_1(x_{\min},\kappa)} \quad (62)$$

$$\geq I\big(X_k; X_{k-\kappa}^{k-1}, \mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{Z}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k+1}^{k+\kappa}\big) - \epsilon_1(x_{\min}, \kappa) \quad (63)$$

$$= I\big(X_k; \mathbf{H}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k+1}^{k+\kappa}\big) - \epsilon_1(x_{\min}, \kappa) \quad (64)$$

$$= I\big(X_k; \mathbf{H}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k+1}^{k+\kappa}, \mathbf{Z}_{k+1}^{k+\kappa}\big) - \underbrace{I\big(X_k; \mathbf{Z}_{k+1}^{k+\kappa} \big| \mathbf{H}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k+1}^{k+\kappa}\big)}_{\leq \epsilon_2(x_{\min},\kappa)} - \epsilon_1(x_{\min}, \kappa) \quad (65)$$

$$\geq I\big(X_k; \mathbf{H}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k+1}^{k+\kappa}, \mathbf{Z}_{k+1}^{k+\kappa}\big) - \epsilon_1(x_{\min}, \kappa) - \epsilon_2(x_{\min}, \kappa) \quad (66)$$

$$= I\big(X_k; \mathbf{H}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \{\mathbf{H}_\ell X_\ell\}_{\ell=k+1}^{k+\kappa}\big) - \epsilon_1(x_{\min}, \kappa) - \epsilon_2(x_{\min}, \kappa) \quad (67)$$

$$= I\left(X_k; \mathbf{H}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \left\{\frac{\mathbf{H}_\ell X_\ell}{\|\mathbf{H}_\ell X_\ell\|}\right\}_{\ell=k+1}^{k+\kappa}, \{\|\mathbf{H}_\ell X_\ell\|\}_{\ell=k+1}^{k+\kappa}\right) - \epsilon_1(x_{\min}, \kappa) - \epsilon_2(x_{\min}, \kappa) \quad (68)$$

$$= I\big(X_k; \mathbf{H}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=k+1}^{k+\kappa}, \{\|\mathbf{H}_\ell X_\ell\|\}_{\ell=k+1}^{k+\kappa}\big) - \epsilon_1(x_{\min}, \kappa) - \epsilon_2(x_{\min}, \kappa) \quad (69)$$

$$\geq I\big(X_k; \mathbf{H}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=k+1}^{k+\kappa}\big) - \epsilon_1(x_{\min}, \kappa) - \epsilon_2(x_{\min}, \kappa) \quad (70)$$

$$= I\big(X_{\kappa+1}; \mathbf{H}_1^{\kappa}, \mathbf{Y}_{\kappa+1}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) - \epsilon_1(x_{\min}, \kappa) - \epsilon_2(x_{\min}, \kappa) \quad (71)$$

$$= I\big(X_{\kappa+1}; \mathbf{Y}_{\kappa+1} \big| \mathbf{H}_1^{\kappa}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) - \epsilon_1(x_{\min}, \kappa) - \epsilon_2(x_{\min}, \kappa), \qquad \kappa+1 \leq k \leq n-\kappa. \quad (72)$$

Here the first equality follows because $\{X_k\}$ is chosen to be i.i.d.; in the subsequent inequality we have dropped some arguments which reduces the mutual information; next we have used the chain rule; in (63) we lower-bound the second term by $-\epsilon_1(x_{\min}, \kappa)$ that—as shown in Appendix II—depends only on $x_{\min}$ and κ and tends to zero as $x_{\min} \uparrow \infty$; in the subsequent equality, we used $X_{k-\kappa}^{k-1}$ and $\mathbf{Z}_{k-\kappa}^{k-1}$ in order to extract $\mathbf{H}_{k-\kappa}^{k-1}$ from $\mathbf{Y}_{k-\kappa}^{k-1}$ and then we dropped $\{X_\ell, \mathbf{Y}_\ell, \mathbf{Z}_\ell\}_{\ell=k-\kappa}^{k-1}$ since given $\mathbf{H}_{k-\kappa}^{k-1}$ it is independent of the other random variables; the subsequent three steps are analogous to (62)–(64), where again it is shown in Appendix II that $\epsilon_2(x_{\min}, \kappa)$ depends only on $x_{\min}$ and κ and tends to zero as $x_{\min} \uparrow \infty$; in (69) $\hat{\mathbf{H}}$ denotes $\mathbf{H}/\|\mathbf{H}\|$; and the equality before last follows from stationarity.

From (72) and (59) we obtain

$$\frac{1}{n} I\big(X_1^n; \mathbf{Y}_1^n\big) \geq \left(1 - \frac{2\kappa}{n}\right) \Big(I\big(X_{\kappa+1}; \mathbf{Y}_{\kappa+1} \big| \mathbf{H}_1^{\kappa}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\kappa+2}^{2\kappa+1}\big) - \epsilon_1(x_{\min}, \kappa) - \epsilon_2(x_{\min}, \kappa)\Big). \quad (73)$$

Letting n tend to infinity we obtain

$$C(\text{SNR}) \geq I\big(X_{\kappa+1}; \mathbf{Y}_{\kappa+1} \big| \mathbf{H}_1^{\kappa}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\kappa+2}^{2\kappa+1}\big) - \epsilon_1(x_{\min}, \kappa) - \epsilon_2(x_{\min}, \kappa) \quad (74)$$

where the first term can be viewed as mutual information across a memoryless SIMO fading channel in the presence of the side information $\big(\mathbf{H}_1^{\kappa}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\kappa+2}^{2\kappa+1}\big)$.

We next let the power grow to infinity $\mathcal{E}_\text{s} \uparrow \infty$. Since the circularly symmetric law (56) achieves the fading number of i.i.d. SIMO fading with side information [1, Note 4.31] and since our choice (57) guarantees that $\epsilon_1(x_{\min}, \kappa)$ and $\epsilon_2(x_{\min}, \kappa)$ tend to zero as $\mathcal{E}_\text{s} \uparrow \infty$ (see Appendix II) we obtain the bound

$$\chi(\{\mathbf{H}_k\}) \geq \chi_{\text{i.i.d.}}\big(\mathbf{H}_{\kappa+1} \big| \mathbf{H}_1^{\kappa}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\kappa+2}^{2\kappa+1}\big). \quad (75)$$

Upon letting κ in the above tend to infinity we obtain the desired result, i.e., that the RHS of (31) is a lower bound to $\chi(\{\mathbf{H}_k\})$.

C. The Converse

Before presenting the derivation of the upper bound we begin with an overview of the proof.

1) An Overview: To upper-bound capacity we use the chain rule

$$\frac{1}{n} I(X_1^n; \mathbf{Y}_1^n) = \frac{1}{n} \sum_{k=1}^{n} I(X_1^n; \mathbf{Y}_k | \mathbf{Y}_1^{k-1}) \quad (76)$$

and upper-bound each term on the RHS of the above by

$$\begin{aligned} I(X_1^n; \mathbf{Y}_k | \mathbf{Y}_1^{k-1}) &= I(X_1^n, \mathbf{Y}_1^{k-1}; \mathbf{Y}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\ &= I(X_1^{k-1}, \mathbf{Y}_1^{k-1}, X_k; \mathbf{Y}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\ &\leq I(X_1^{k-1}, \mathbf{Y}_1^{k-1}, \mathbf{H}_1^{k-1}, X_k; \mathbf{Y}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\ &= I(\mathbf{H}_1^{k-1}, X_k; \mathbf{Y}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\ &= I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_1^{k-1}; \mathbf{Y}_k | X_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\ &= I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_1^{k-1}; X_k, \mathbf{Y}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\ &\leq I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_1^{k-1}; \mathbf{H}_k, X_k, \mathbf{Y}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\ &= I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_1^{k-1}; \mathbf{H}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\ &\leq I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}). \end{aligned} \quad (77)$$

Here, the first equality follows from the chain rule; the second because we prohibit feedback; the subsequent inequality from the inclusion of the additional random variables $\mathbf{H}_1^{k-1}$ in the mutual information term; the subsequent equality because, conditional on the past fading $\mathbf{H}_1^{k-1}$ and on the present input X_k, the past inputs and outputs $(X_1^{k-1}, \mathbf{Y}_1^{k-1})$ are independent of the present output $\mathbf{Y}_k$; the subsequent equality by the chain rule; the subsequent equality from the independence of the inputs and the fading; the subsequent inequality from the inclusion of the random vector $\mathbf{H}_k$ in the mutual information term; the subsequent equality because conditional on the present fading, the past fading $\mathbf{H}_1^{k-1}$ is independent of the present input and output $(X_k, \mathbf{Y}_k)$; and the final inequality from the stationarity of the fading.

A trivial upper bound can be now obtained from (77) by lower-bounding $I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1})$ by zero. This bound is, however, not tight. The main difficulty in the proof is that if we fix some k and maximize $I(X_k; \mathbf{Y}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1})$ over all joint distributions on $X_1, \ldots, X_n$ (satisfying the average-power constraint), then this nontight bound would be achievable. For example, we could choose $X_1, \ldots, X_{k-1}$ to be deterministically zero so that $I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) = I(\mathbf{Y}_k; \mathbf{Z}_1^{k-1}) = 0$ and choose X_k to maximize

$I(X_k;\boldsymbol{Y}_k)$, *i.e.*, to achieve the i.i.d. capacity. (The resulting average mutual information $n^{-1}I(X_1^n;\boldsymbol{Y}_1^n)$ would, of course, be very low but the single term $I(X_1^n;\boldsymbol{Y}_k|\boldsymbol{Y}_1^{k-1})$ would achieve this nontight bound.)

It would thus seem that to obtain an asymptotically tight upper bound on channel capacity we cannot upper-bound each of the individual terms in (76) in isolation. There is, however, a way to do just that. The trick is to consider only joint distributions on $(X_1,\ldots,X_n)$ that are stationary. In fact, it suffices to limit ourselves to joint distributions under which the random variables $X_1,\ldots,X_n$ all have the same law. The first step in the proof will thus be to show that one can approach capacity arbitrarily closely using such inputs. This is done in Lemma 5 ahead. (Actually, the inputs we use will not quite have equal marginals. Only $X_\eta,\ldots,X_{n-2\eta+2}$ will have equal marginals, where η is a fixed integer that depends on the SNR and on the required gap between capacity and mutual information but not on the block length n. The edge effects will wash out when we let $n\to\infty$ with η held fixed.)

Assume now that, except for some edge effects, we can get to within arbitrary $\epsilon>0$ of capacity using inputs $\{X_k\}$ of marginal Q (where the law Q depends on the SNR and on the gap to capacity ϵ, but not on n). Let $I(Q)$ denote $I(X_k;\boldsymbol{Y}_k)$ when X_k is distributed according to Q. Thus, for such inputs

$$C\le\varliminf_{n\to\infty}\frac{1}{n}I(X_1^n;\boldsymbol{Y}_1^n)+\epsilon. \tag{78}$$

By (76) and (77) we also have for such inputs

$$\begin{aligned}\frac{1}{n}&I(X_1^n;\boldsymbol{Y}_1^n)\\ &\le\frac{1}{n}\sum_{k=1}^{n}\Big(I(X_k;\boldsymbol{Y}_k)-I(\boldsymbol{Y}_k;\boldsymbol{Y}_1^{k-1})\Big)+I(\boldsymbol{H}_0;\boldsymbol{H}_{-\infty}^{-1})\\ &\approx I(Q)-\frac{1}{n}\sum_{k=1}^{n}I(\boldsymbol{Y}_k;\boldsymbol{Y}_1^{k-1})+I(\boldsymbol{H}_0;\boldsymbol{H}_{-\infty}^{-1})\end{aligned} \tag{79}$$

where the approximation results from ignoring the edge effects, i.e., from ignoring the fact that only $X_\eta,\ldots,X_{n-2\eta+2}$ are of marginal Q. In fact, as we let n tend to infinity the edge effects wash out and we obtain that for such marginal-Q inputs $\{X_k\}$

$$\varlimsup_{n\to\infty}\frac{1}{n}I(X_1^n;\boldsymbol{Y}_1^n)\le I(Q)-\varliminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}I(\boldsymbol{Y}_k;\boldsymbol{Y}_1^{k-1})+I(\boldsymbol{H}_0;\boldsymbol{H}_{-\infty}^{-1}). \tag{80}$$

The choice of Q (the distribution of X_k) affects the RHS of (79) and (80) in two different ways. It determines $I(Q)$ but it also influences the terms $I(\boldsymbol{Y}_k;\boldsymbol{Y}_1^{k-1})$. There is thus a tension between choosing Q to maximize $I(Q)$ (i.e., to make $I(Q)$ close to the i.i.d. channel capacity) and choosing Q to minimize the $I(\boldsymbol{Y}_k;\boldsymbol{Y}_1^{k-1})$ terms. It is important to note that at high SNR the relative importance of these conflicting objectives are vastly different. From $I(Q)$ stems the double-logarithmic growth of channel capacity, whereas the sum on the RHS of (79) and (80) merely influences the fading number. No matter how we choose Q, this sum cannot be smaller than zero.

We next study the input marginals Q. We note that for the marginal-Q input $\{X_k\}$ to satisfy (78) we must have

$$\lim_{\mathrm{SNR}\to\infty}\frac{I(Q)}{\log\log\mathrm{SNR}}=1. \tag{81}$$

This can be argued as follows. Because $C\ge I(Q)$, it follows by (7) that

$$\varlimsup_{\mathrm{SNR}\to\infty}\frac{I(Q)}{\log\log\mathrm{SNR}}\le 1.$$

On the other hand, from (80) and the nonnegativity of mutual information we obtain that for the marginal-Q input $\{X_k\}$ to satisfy (78) the marginal law Q must satisfy

$$I(Q)\ge C-\epsilon-I(\boldsymbol{H}_0;\boldsymbol{H}_{-\infty}^{-1}) \tag{82}$$

which combines with

$$C\ge\log\log\mathrm{SNR}+O(1) \tag{83}$$

(where the $O(1)$ term is bounded in the SNR) to imply

$$\varliminf_{\mathrm{SNR}\to\infty}\frac{I(Q)}{\log\log\mathrm{SNR}}\ge 1.$$

Here, the lower bound (83) follows, for example, from [1, Theorem 4.41], e.g., by considering the SISO channel that results when the signals in all but one of the receive antennas are ignored.

The next step in the proof of the converse is to show that (81) implies that Q "escapes to infinity," i.e., that

$$\lim_{\mathrm{SNR}\to\infty}Q(\{x:|x|\ge\xi_{\min}\})=1,\qquad\text{for any fixed }\xi_{\min}. \tag{84}$$

This is proved in greater generality for general cost-constrained channels in Section VII-C.3, where we also discuss how this result relates to the notion of "capacity achieving input distributions that escape to infinity" of [1].

It is thus seen that at high SNR, the marginal Q guarantees that with very high probability only very large inputs are used. In fact, using the union bound we can infer that the probability that a finite number of inputs will all exceed $\xi_{\min}$ also tends to one. The final step in the proof is then to show that if the inputs are large with high probability, then

$$I(\boldsymbol{Y}_k;\boldsymbol{Y}_1^{k-1})\approx I(\hat{\boldsymbol{H}}_k\hat{U}_k;\{\hat{\boldsymbol{H}}_\ell\hat{U}_\ell\}_{\ell=1}^{k-1}) \tag{85}$$

where $\{\hat{U}_\ell\}$ are i.i.d. uniformly distributed on the complex unit circle independently of the fading process (as described in (1)).

The intuition behind (85) is quite simple. If the inputs X_1^k are guaranteed to be very large with probability one, then we should be able from the past outputs $\boldsymbol{Y}_1^{k-1}$ to learn the past "direction" (corrupted by random rotations) $\{\hat{\boldsymbol{H}}_\ell\hat{U}_\ell\}_{\ell=1}^{k-1}$. Similarly, there would be an almost deterministic relationship between the present output $\boldsymbol{Y}_k$ and the present fading "direction" (again corrupted by a random phase) $\hat{\boldsymbol{H}}_k\hat{U}_k$.

Of course, the escape to infinity does not guarantee that the inputs exceed $\xi_{\min}$ with probability one but only with probability approaching one. To address this difficulty we introduce the binary random variable E_k in that part of the proof.

2) Stationarity Considerations:

Lemma 5: Fix some power $\mathcal{E}_s$ with corresponding SNR of $\mathcal{E}_s/\sigma^2$. Let $C(\text{SNR})$ denote the corresponding channel capacity under an average power $\mathcal{E}_s$ constraint. Then for any $\epsilon > 0$ there corresponds some positive integer $\eta = \eta(\text{SNR}, \epsilon)$ and some distribution $Q = Q(\text{SNR}, \epsilon)$ on $\mathbb{C}$ such that for any block length n sufficiently large there exists some input X_1^n satisfying the following:

1) the input X_1^n nearly achieves capacity in the sense that

$$\frac{1}{n} I\big(X_1^n; \boldsymbol{Y}_1^n\big) \geq C(\text{SNR}) - \epsilon; \tag{86}$$

2) except for the first $\eta - 1$ symbols $X_1^{\eta-1}$ and for at most the last $2(\eta - 1)$ symbols $X_{n-2\eta+3}^n$ the symbols

$$X_\eta, X_{\eta+1}, \ldots, X_{n-2\eta+2} \tag{87}$$

all have the same distribution Q;

3) this marginal distribution Q gives rise to a second moment $\mathcal{E}_s$

$$\mathsf{E}\left[|X_\ell|^2\right] = \mathcal{E}_s, \qquad \ell = \eta, \ldots, n - 2\eta + 2; \tag{88}$$

4) and the first $\eta - 1$ symbols and the last $2(\eta - 1)$ symbols satisfy the power constraint possibly strictly

$$\mathsf{E}\left[|X_\ell|^2\right] \leq \mathcal{E}_s, \quad \ell \in \{1, \ldots, \eta - 1\} \cup \{n - 2\eta + 3, \ldots, n\}. \tag{89}$$

Proof: The proof is by a simple shift-and-mix argument. Recalling that

$$C(\text{SNR}) = \lim_{n \to \infty} \frac{1}{n} \sup_{P_{X_1,\ldots,X_n}} I(X_1, \ldots, X_n; \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)$$

where the supremum is over all joint distributions on the random variables $X_1, \ldots, X_n$ under which $\sum_{\ell=1}^n \mathsf{E}\left[|X_\ell|^2\right] = n\mathcal{E}_s$, we conclude that there must exist some integer $\eta \geq 1$ and some joint distribution P^* on $\mathbb{C}^\eta$ such that if $(X_1, \ldots, X_\eta) \sim P^*$ then

$$\frac{1}{\eta} \sum_{\ell=1}^{\eta} \mathsf{E}\left[|X_\ell|^2\right] = \mathcal{E}_s \tag{90}$$

and

$$\frac{1}{\eta} I(X_1, \ldots, X_\eta; \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_\eta) > C(\text{SNR}) - \frac{\epsilon}{2}. \tag{91}$$

Let Q be the probability law on $\mathbb{C}$ that is the mixture of the η different marginals of P^*. That is, for any Borel set $\mathcal{B} \subset \mathbb{C}$

$$Q(\mathcal{B}) = \frac{1}{\eta} \sum_{\ell=1}^{\eta} P^*(X_\ell \in \mathcal{B}). \tag{92}$$

By (90) we have

$$\int_{\mathbb{C}} |x|^2 \mathrm{d}Q(x) = \mathcal{E}_s. \tag{93}$$

Let n now be given. We shall next describe the required input distribution as follows. Let

$$\nu = \left\lfloor \frac{n - \eta + 1}{\eta} \right\rfloor$$

and let the infinite random sequence $\tilde{\boldsymbol{X}}$ be defined by

$$\tilde{\boldsymbol{X}} = (\underbrace{0, \ldots, 0}_{\eta-1}, \underbrace{\Xi_1^{(1)}, \ldots, \Xi_\eta^{(1)}}_{\eta}, \ldots\ldots, \underbrace{\Xi_1^{(\nu)}, \ldots, \Xi_\eta^{(\nu)}}_{\eta}, 0, 0, \ldots$$

so that

$$\tilde{X}_\ell = \begin{cases} 0, & \text{if } 1 \leq \ell \leq \eta - 1 \\ \Xi_{(\ell \bmod \eta)+1}^{\lfloor \ell/\eta \rfloor}, & \text{if } \eta \leq \ell \leq (\nu + 1)\eta - 1 \\ 0, & \text{if } \ell \geq (\nu + 1)\,\eta. \end{cases}$$

Here

$$\left\{\left(\Xi_1^{(j)}, \ldots, \Xi_\eta^{(j)}\right)\right\}_{j=1}^{\nu} \sim \text{i.i.d. } P^*.$$

Notice that since the lead-in and trailing zeros have no effect on our channel, the unnormalized mutual information induced by $\tilde{\boldsymbol{X}}$ is lower-bounded by $\nu\eta(C(\text{SNR}) - \epsilon/2)$. Again, since the lead-in and trailing zeros are of no consequence, this same mutual information results if we shift $\tilde{\boldsymbol{X}}$ by t (provided that $0 \leq t \leq \eta - 1$). Consequently, if we define $X_1, \ldots, X_n$ by the mixture of the time shift of $\tilde{\boldsymbol{X}}$, i.e.,

$$X_\ell = \tilde{X}_{\ell+T}, \qquad 1 \leq \ell \leq n$$

where

$$T \sim \text{uniform on } \{0, \ldots, \eta - 1\}$$

is independent of $\tilde{\boldsymbol{X}}$, then by the concavity of mutual information in the input distribution we obtain that the unnormalized mutual information induced by X_1^n is lower-bounded by $\nu\eta(C(\text{SNR}) - \epsilon/2)$, so that the normalized mutual information satisfies

$$\begin{aligned} \frac{1}{n} I\left(X_1^n; \boldsymbol{Y}_1^n\right) &\geq \frac{\eta\nu}{n}\left(C(\text{SNR}) - \epsilon/2\right) \\ &= \frac{\eta\left\lfloor \frac{n-\eta+1}{\eta} \right\rfloor}{n}\left(C(\text{SNR}) - \epsilon/2\right) \end{aligned}$$

which exceeds $C(\text{SNR}) - \epsilon$ for sufficiently large n.

Except at the edges, the above mixture guarantees that all marginals are Q, and hence by (93) of average power $\mathcal{E}_s$. The power in the edges can be smaller than $\mathcal{E}_s$ because of the mixture with deterministic zeros. □

3) Input Distributions That Escape to Infinity Revisited: In this subsection, we revisit the notion of "capacity achieving input distributions that escape to infinity" that was introduced in [1]. Under slightly more restrictive conditions on the asymptotic behavior of channel capacity, we shall strengthen the results of [1] in the following sense. When specialized to the problem at hand, Theorem 4.13 of [1] demonstrates that the fading number *can* be achieved by input distributions that escape to infinity. That is, *there exist* input distributions satisfying the cost constraint and escaping to infinity that induce mutual informations whose *difference* from capacity tends to zero. Our present result, when specialized to the present setting, strengthens [1, Theorem 4.13] by showing that *any* sequence of input distributions satisfying the cost constraint and inducing a mutual information whose *ratio* to $\log\log \text{SNR}$ tends to 1 *must* escape to infinity.

To see how the new result implies the old one we need to demonstrate the existence of some sequence of input distributions satisfying the cost constraint; inducing mutual informations whose gap to capacity tends to zero; and escaping to infinity. But the new result demonstrates that any sequence of input distributions satisfying the first two conditions must satisfy the third, because if the *difference* between the mutual information and capacity tends to zero it follows that their ratios to $\log\log \mathrm{SNR}$ must tend to one.

Surprisingly, the proof of the present statement is easier. It should, however, be noted that while the new result—like [1, Theorem 4.13]—extends to general cost-constrained channels, the required assumptions on the functional form of the capacity-cost function are somewhat more stringent.

As in [1], for the sake of greater generality, we shall consider general memoryless channels over the input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ and general costs. As in [1], we shall assume that the input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ are separable metric spaces, and that for any set $\mathcal{B} \subset \mathcal{Y}$ the mapping $x \mapsto W(\mathcal{B}|x)$ from $\mathcal{X}$ to $[0,1]$ is Borel measurable. The cost function $g : \mathcal{X} \to [0,\infty)$ is assumed measurable.

Recall the following standard definition of the capacity-cost function.

Definition 6: Given a channel $W(\cdot|\cdot)$ over the input alphabet $\mathcal{X}$ and the output alphabet $\mathcal{Y}$ and given some nonnegative cost function $g : \mathcal{X} \to \mathbb{R}^+$, we define the capacity-cost function $C : \mathbb{R}^+ \to \mathbb{R}^+$ by

$$C(\Upsilon) = \sup_{Q:\mathsf{E}_Q[g(X)]\leq\Upsilon} I(Q,W), \qquad \Upsilon \geq 0. \tag{94}$$

Definition 7: Let $\{Q_\Upsilon, \Upsilon \geq 0\}$ be a family of input distributions on $\mathcal{X}$ parameterized by the cost Υ such that

$$\mathsf{E}_{Q_\Upsilon}[g(X)] \leq \Upsilon, \qquad \Upsilon \geq 0. \tag{95}$$

We say that the input distributions $\{Q_\Upsilon, \Upsilon \geq 0\}$ escape to infinity if for any $\Upsilon_0 > 0$

$$\lim_{\Upsilon\uparrow\infty} Q_\Upsilon\big(\{x \in \mathcal{X} : g(x) < \Upsilon_0\}\big) = 0. \tag{96}$$

Theorem 8: Let the cost function $g(\cdot)$ and the channel $W(\cdot|\cdot)$ over the alphabets $\mathcal{X}, \mathcal{Y}$ be as above. Let the capacity-cost function $C(\cdot)$ be finite but unbounded. Let $C_{\mathrm{asy}}(\Upsilon)$ be a function of the cost that captures the asymptotic behavior of the capacity-cost function $C(\Upsilon)$ in the sense that

$$\lim_{\Upsilon\uparrow\infty} \frac{C(\Upsilon)}{C_{\mathrm{asy}}(\Upsilon)} = 1. \tag{97}$$

Assume that $C_{\mathrm{asy}}(\cdot)$ satisfies the growth condition

$$\varliminf_{\Upsilon\uparrow\infty} \left\{ \sup_{\alpha\in(0,\alpha_0]} \frac{\alpha C_{\mathrm{asy}}\left(\frac{\Upsilon}{\alpha}\right)}{C_{\mathrm{asy}}(\Upsilon)} \right\} < 1, \qquad \forall\, 0<\alpha_0<1. \tag{98}$$

Let $\{Q_\Upsilon, \Upsilon \geq 0\}$ be a family of input distributions satisfying (95) and

$$\lim_{\Upsilon\uparrow\infty} \frac{I(Q_\Upsilon, W)}{C_{\mathrm{asy}}(\Upsilon)} = 1. \tag{99}$$

Then $\{Q_\Upsilon, \Upsilon \geq 0\}$ escape to infinity.

Remark 9: The growth condition (98) is related to the notion of "slowly varying in the Karamata sense," see [22, Sec.1.2]. Examples of functions $C_{\mathrm{asy}}(\Upsilon)$ that satisfy (98) include

$$\log(1+\log(1+\Upsilon)),\ \log(1+\Upsilon),\ (\log(1+\Upsilon))^\beta, \qquad \text{for } \beta > 0$$

and any positive multiple thereof. In this paper, we shall use this theorem with $C_{\mathrm{asy}}(\Upsilon) = \log(1+\log(1+\Upsilon))$.

Proof: In the following all expectations, probabilities, and mutual informations are computed with respect to the input law Q_Υ. Fix some $\Upsilon_0 > 0$ and let

$$E = \begin{cases} 1, & \text{if } g(X) \geq \Upsilon_0 \\ 0, & \text{otherwise} \end{cases} \tag{100}$$

$$\alpha = \Pr[E=1]. \tag{101}$$

Since $C(\cdot)$ is monotonically increasing and unbounded, it follows by (97) that

$$\lim_{\Upsilon\to\infty} C_{\mathrm{asy}}(\Upsilon) = \infty \tag{102}$$

which combines with (99) to imply that

$$\lim_{\Upsilon\to\infty} I(Q_\Upsilon, W) = \infty. \tag{103}$$

By (103), it follows that for all Υ sufficiently large we must have $\alpha > 0$, because $\alpha = 0$ implies $g(X) \leq \Upsilon_0$ Q_Υ-almost surely whence $I(Q_\Upsilon, W) \leq C(\Upsilon_0)$.

In the following, we shall thus assume that Υ is indeed sufficiently large so that $\alpha > 0$. Then

$$\begin{aligned} I(X;Y) &= I(X,E;Y) \\ &= I(E;Y) + I(X;Y|E) \\ &= I(E;Y) + I(X;Y|E{=}0)\Pr[E=0] \\ &\quad + I(X;Y|E{=}1)\Pr[E=1] \\ &\leq \log 2 + I(X;Y|E{=}0) + \alpha I(X;Y|E{=}1) \\ &\leq \log 2 + C(\Upsilon_0) + \alpha I(X;Y|E{=}1) \\ &\leq \log 2 + C(\Upsilon_0) + \alpha C\left(\frac{\Upsilon}{\alpha}\right). \end{aligned} \tag{104}$$

Here, the first inequality follows because E is a binary random variable and because $\Pr[E=0] \leq 1$; the following inequality because conditional on $E{=}0$ the input X satisfies $g(X) < \Upsilon_0$ with probability one, so that $\mathsf{E}[g(X)|E{=}0] \leq \Upsilon_0$; and the final inequality because

$$\mathsf{E}[g(X)\,|E=1] \leq \frac{\mathsf{E}[g(X)]}{\alpha} \leq \frac{\Upsilon}{\alpha}. \tag{105}$$

To show that $\alpha \to 1$ assume by contradiction that there is some sequence of costs $\Upsilon_n \uparrow \infty$ with corresponding

$$\alpha_n = Q_{\Upsilon_n}(g(X) \geq \Upsilon_0)$$

such that $\{\alpha_n\}$ converges to some $\alpha^* < 1$. It then follows that there exists some $\alpha_0 < 1$ such that

$$\alpha_n < \alpha_0, \qquad n \text{ sufficiently large.} \tag{106}$$

From (104) we now have

$$\underbrace{\frac{I(X;Y)}{C_{\text{asy}}(\Upsilon_n)}}_{\to 1} \le \underbrace{\frac{\log 2 + C(\Upsilon_0)}{C_{\text{asy}}(\Upsilon_n)}}_{\to 0} + \underbrace{\frac{C(\Upsilon_n/\alpha_n)}{C_{\text{asy}}(\Upsilon_n/\alpha_n)}}_{\to 1} \cdot \frac{\alpha_n C_{\text{asy}}(\Upsilon_n/\alpha_n)}{C_{\text{asy}}(\Upsilon_n)}.$$

Here, the limiting behavior of the left-hand side (LHS) follows from (99); the limiting value of $(\log 2 + C(\Upsilon_0))/C_{\text{asy}}(\Upsilon_n)$ follows by (102) because $C(\Upsilon_0) < \infty$; and the limiting behavior of the term $C(\Upsilon_n/\alpha_n)/C_{\text{asy}}(\Upsilon_n/\alpha_n) \to 1$ follows from (97) because $\Upsilon_n \uparrow \infty$ implies $\Upsilon_n/\alpha_n \uparrow \infty$. Upon letting n tend to infinity we obtain the contradiction

$$\begin{aligned} 1 &\le \varliminf_{n\to\infty} \frac{\alpha_n C_{\text{asy}}(\Upsilon_n/\alpha_n)}{C_{\text{asy}}(\Upsilon_n)} \\ &\le \varliminf_{\Upsilon\uparrow\infty} \left\{ \sup_{\alpha\in(0,\alpha_0]} \frac{\alpha C_{\text{asy}}\left(\frac{\Upsilon}{\alpha}\right)}{C_{\text{asy}}(\Upsilon)} \right\} \\ &< 1. \end{aligned}$$

Here the second inequality follows from (106) and the last inequality follows from (98). □

4) Proof of Converse: Fix $\mathcal{E}_{\text{s}} > 0$ and set $\text{SNR} = \mathcal{E}_{\text{s}}/\sigma^2$. Let the positive integer κ be arbitrary and let $\xi_{\min} > 0$ be also arbitrary. Fix $\epsilon > 0$ and let $\eta = \eta(\text{SNR}, \epsilon)$ and $Q = Q(\text{SNR}, \epsilon)$ be the integer and the input distribution on $\mathbb{C}$ whose existence is guaranteed in Lemma 5. Let X_1^n satisfy (86)–(89) of Lemma 5 so that, in particular

$$C \le \frac{1}{n} I\big(X_1^n; \boldsymbol{Y}_1^n\big) + \epsilon \tag{107}$$

$$= \frac{1}{n} \sum_{k=1}^{n} I\big(X_1^n; \boldsymbol{Y}_k \big| \boldsymbol{Y}_1^{k-1}\big) + \epsilon. \tag{108}$$

For $1 \le k \le \eta + \kappa - 1$ and for $n - 2\eta + 3 \le k \le n$ we use the crude bound

$$I\big(X_1^n; \boldsymbol{Y}_k \big| \boldsymbol{Y}_1^{k-1}\big) \le I\big(X_k; \boldsymbol{Y}_k\big) + I\big(\boldsymbol{H}_0; \boldsymbol{H}_{-\infty}^{-1}\big) \tag{109}$$

$$\le C_{\text{i.i.d.}}(\text{SNR}) + I\big(\boldsymbol{H}_0; \boldsymbol{H}_{-\infty}^{-1}\big) \tag{110}$$

which is uniformly bounded in n. Here, the first inequality follows from (77) and the second from (88) and (89). We conclude that

$$C \le \varliminf_{n\to\infty} \frac{1}{n} I\big(X_1^n; \boldsymbol{Y}_1^n\big) + \epsilon \tag{111}$$

$$= \varliminf_{n\to\infty} \frac{1}{n-\kappa-3(\eta-1)} \sum_{k=\eta+\kappa}^{n-2\eta+2} I\big(X_1^n; \boldsymbol{Y}_k \big| \boldsymbol{Y}_1^{k-1}\big) + \epsilon. \tag{112}$$

This allows us to focus on epochs k satisfying

$$\eta + \kappa \le k \le n - 2(\eta - 1)$$

and thus guaranteeing that X_k and its κ predecessors $X_{k-1}, \ldots, X_{k-\kappa}$ are all distributed according to Q. Any upper bound on $I\big(X_1^n; \boldsymbol{Y}_k \big| \boldsymbol{Y}_1^{k-1}\big)$ that does not depend on k will result in an upper bound on C via (112).

For k satisfying $\eta + \kappa \le k \le n - 2(\eta - 1)$, we upper-bound this term by

$$\begin{aligned} &I\big(X_1^n; \boldsymbol{Y}_k \big| \boldsymbol{Y}_1^{k-1}\big) \\ &\quad \le I(X_k; \boldsymbol{Y}_k) + I(\boldsymbol{H}_0; \boldsymbol{H}_{-\infty}^{-1}) - I(\boldsymbol{Y}_k; \boldsymbol{Y}_1^{k-1}) &(113)\\ &\quad \le I(X_k; \boldsymbol{Y}_k) + I(\boldsymbol{H}_0; \boldsymbol{H}_{-\infty}^{-1}) - I(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1}) &(114)\\ &\quad = I(Q) + I(\boldsymbol{H}_0; \boldsymbol{H}_{-\infty}^{-1}) - I(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1}) &(115)\end{aligned}$$

where we use $I(Q)$ to denote the mutual information $I(X_k; \boldsymbol{Y}_k)$ when X_k is distributed according to the law Q. From (115) and (112) we conclude that

$$C \le I(Q) + I(\boldsymbol{H}_0; \boldsymbol{H}_{-\infty}^{-1}) + \epsilon - \varliminf_{n\to\infty} \min_{\eta+\kappa\le k\le n-2(\eta-1)} I\big(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1}\big). \tag{116}$$

We thus proceed to lower-bound $I\big(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1}\big)$ for $\eta + \kappa \le k \le n - 2(\eta - 1)$. For such k, define

$$E_k = \begin{cases} 1, & \text{if } |X_j| \ge \xi_{\min}, \quad \forall\, k-\kappa \le j \le k \\ 0, & \text{otherwise.} \end{cases} \tag{117}$$

Let

$$\alpha_k = \Pr[E_k = 1]. \tag{118}$$

By the union of events bound

$$\alpha_k \ge 1 - \sum_{j=k-\kappa}^{k} \Pr[|X_j| < \xi_{\min}] \tag{119}$$

$$= 1 - (\kappa + 1) Q(|X| < \xi_{\min}) \tag{120}$$

where we have used the fact that for k in the range of interest $\eta + \kappa \le k \le n - 2(\eta - 1)$, the random variables $X_k, \ldots, X_{k-\kappa}$ are all distributed according to Q. Consequently

$$\alpha_k \ge \alpha \tag{121}$$

where $\alpha = \alpha(\xi_{\min}, Q, \kappa)$ is given by

$$\alpha = \max\big\{0, 1 - (\kappa + 1) Q(|X| < \xi_{\min})\big\}. \tag{122}$$

We now lower-bound $I\big(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1}\big)$ for $\eta + \kappa \le k \le n - 2(\eta - 1)$ by

$$\begin{aligned} &I\big(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1}\big) \\ &\quad = I\big(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1}, E_k\big) - I\big(\boldsymbol{Y}_k; E_k \big| \boldsymbol{Y}_{k-\kappa}^{k-1}\big) &(123)\\ &\quad = I\big(\boldsymbol{Y}_k; E_k\big) + I\big(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1} \big| E_k\big) \\ &\qquad - I\big(\boldsymbol{Y}_k; E_k \big| \boldsymbol{Y}_{k-\kappa}^{k-1}\big) &(124)\\ &\quad \ge I\big(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1} \big| E_k\big) - I\big(\boldsymbol{Y}_k; E_k \big| \boldsymbol{Y}_{k-\kappa}^{k-1}\big) &(125)\\ &\quad = I\big(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1} \big| E_k\big) - H\big(E_k \big| \boldsymbol{Y}_{k-\kappa}^{k-1}\big) \\ &\qquad + H\big(E_k \big| \boldsymbol{Y}_{k-\kappa}^{k}\big) &(126)\\ &\quad \ge I\big(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1} \big| E_k\big) - H\big(E_k \big| \boldsymbol{Y}_{k-\kappa}^{k-1}\big) &(127)\\ &\quad \ge I\big(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1} \big| E_k\big) - H(E_k) &(128)\\ &\quad = I\big(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1} \big| E_k\big) - H_{\text{b}}(\alpha_k) &(129)\\ &\quad \ge I\big(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1} \big| E_k\big) - H_{\text{b}}\left(\max\left\{\alpha_k, \frac{1}{2}\right\}\right) &(130)\\ &\quad \ge \alpha_k I\big(\boldsymbol{Y}_k; \boldsymbol{Y}_{k-\kappa}^{k-1} \big| E_k{=}1\big) - H_{\text{b}}\left(\max\left\{\alpha_k, \frac{1}{2}\right\}\right) &(131)\end{aligned}$$

$$\geq \alpha I\big(\boldsymbol{Y}_k;\boldsymbol{Y}_{k-\kappa}^{k-1}\big|E_k{=}1\big)-H_{\rm b}\left(\max\left\{\alpha,\frac{1}{2}\right\}\right) \quad (132)$$

where

$$H_{\rm b}(\xi) \triangleq -\xi\log\xi-(1-\xi)\log(1-\xi) \quad (133)$$

is the binary entropy function. Note that

$$H_{\rm b}(\xi) \leq H_{\rm b}\left(\max\left\{\xi,\frac{1}{2}\right\}\right)$$

and that $H_{\rm b}\left(\max\left\{\xi,\frac{1}{2}\right\}\right)$ is monotonically nonincreasing so that the last inequality follows from (121).

Inequalities (132) and (116) combine to yield

$$C \leq I(Q)+I(\boldsymbol{H}_0;\boldsymbol{H}_{-\infty}^{-1})+\epsilon+H_{\rm b}\left(\max\left\{\alpha,\frac{1}{2}\right\}\right) -\alpha \varliminf_{n\to\infty} \min_{\eta+\kappa\leq k\leq n-2(\eta-1)} I\big(\boldsymbol{Y}_k;\boldsymbol{Y}_{k-\kappa}^{k-1}\big|E_k=1\big) \quad (134)$$

and we now proceed to lower-bound $I\big(\boldsymbol{Y}_k;\boldsymbol{Y}_{k-\kappa}^{k-1}\big|E_k=1\big)$ (for $\eta+\kappa\leq k\leq n-2(\eta-1)$)

$$\begin{aligned} I\big(\boldsymbol{Y}_k;\boldsymbol{Y}_{k-\kappa}^{k-1}\big|E_k=1\big) &= I\big(\boldsymbol{Y}_k;\boldsymbol{Y}_{k-\kappa}^{k-1},\boldsymbol{Z}_{k-\kappa}^{k-1}\big|E_k=1\big) - I\big(\boldsymbol{Y}_k;\boldsymbol{Z}_{k-\kappa}^{k-1}\big|\boldsymbol{Y}_{k-\kappa}^{k-1},E_k=1\big) && (135)\\ &= I\big(\boldsymbol{Y}_k;\{\boldsymbol{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big) - I\big(\boldsymbol{Y}_k;\boldsymbol{Z}_{k-\kappa}^{k-1}\big|\boldsymbol{Y}_{k-\kappa}^{k-1},E_k=1\big) && (136)\\ &\geq I\big(\boldsymbol{Y}_k;\{\boldsymbol{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big)-\delta_1(\xi_{\min},\kappa) && (137)\end{aligned}$$

where $\delta_1(\xi_{\min},\kappa)$ is an upper bound

$$I\big(\boldsymbol{Y}_k;\boldsymbol{Z}_{k-\kappa}^{k-1}\big|\boldsymbol{Y}_{k-\kappa}^{k-1},E_k=1\big)\leq\delta_1(\xi_{\min},\kappa) \quad (138)$$

that is derived in Appendix III. Note that $\delta_1(\xi_{\min},\kappa)$ depends only on $\xi_{\min}$ and κ (and not on k or on the SNR) and that

$$\lim_{\xi_{\min}\uparrow\infty}\delta_1(\xi_{\min},\kappa)=0. \quad (139)$$

Continuing with the chain of inequalities we have

$$\begin{aligned} I\big(\boldsymbol{Y}_k;\boldsymbol{Y}_{k-\kappa}^{k-1}\big|E_k=1\big) &\geq I\big(\boldsymbol{Y}_k;\{\boldsymbol{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big)-\delta_1(\xi_{\min},\kappa) && (140)\\ &= I\big(\boldsymbol{Y}_k,\boldsymbol{Z}_k;\{\boldsymbol{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big) - I\big(\boldsymbol{Z}_k;\{\boldsymbol{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|\boldsymbol{Y}_k,E_k{=}1\big)-\delta_1(\xi_{\min},\kappa) && (141)\\ &= I\big(\boldsymbol{H}_k X_k;\{\boldsymbol{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big) - I\big(\boldsymbol{Z}_k;\{\boldsymbol{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|\boldsymbol{Y}_k,E_k{=}1\big)-\delta_1(\xi_{\min},\kappa) && (142)\\ &\geq I\big(\boldsymbol{H}_k X_k;\{\boldsymbol{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big)-\delta_1(\xi_{\min},\kappa) - \delta_2(\xi_{\min},\kappa) && (143)\end{aligned}$$

where $\delta_2(\xi_{\min},\kappa)$ is an upper bound

$$I\big(\boldsymbol{Z}_k;\{\boldsymbol{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|\boldsymbol{Y}_k,E_k=1\big)\leq\delta_2(\xi_{\min},\kappa) \quad (144)$$

that is also derived in Appendix III. It too depends only on $\xi_{\min}$ and κ and satisfies

$$\lim_{\xi_{\min}\uparrow\infty}\delta_2(\xi_{\min},\kappa)=0. \quad (145)$$

To continue the chain of inequalities, let us define $\{\hat{V}_\ell\}_{\ell=k-\kappa}^{k}$ to be i.i.d. complex random variables that are uniformly distributed on $\{z\in\mathbb{C}:|z|=1\}$ and independent of $\{X_k,\boldsymbol{H}_k\}$. Let $\{\hat{U}_\ell\}$ be similarly distributed. Then

$$\begin{aligned} I\big(\boldsymbol{Y}_k;\boldsymbol{Y}_{k-\kappa}^{k-1}\big|E_k=1\big) &\geq I\big(\boldsymbol{H}_k X_k;\{\boldsymbol{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big)-\delta_1(\xi_{\min},\kappa) - \delta_2(\xi_{\min},\kappa) && (146)\\ &\geq I\big(\boldsymbol{H}_k X_k\hat{V}_k;\{\boldsymbol{H}_\ell X_\ell\hat{V}_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big) - \delta_1(\xi_{\min},\kappa)-\delta_2(\xi_{\min},\kappa) && (147)\\ &\geq I\left(\frac{\boldsymbol{H}_k X_k\hat{V}_k}{\|\boldsymbol{H}_k X_k\|};\left\{\frac{\boldsymbol{H}_\ell X_\ell\hat{V}_\ell}{\|\boldsymbol{H}_\ell X_\ell\|}\right\}_{\ell=k-\kappa}^{k-1}\middle| E_k=1\right) - \delta_1(\xi_{\min},\kappa)-\delta_2(\xi_{\min},\kappa) && (148)\\ &= I\left(\frac{\boldsymbol{H}_k}{\|\boldsymbol{H}_k\|}\hat{U}_k;\left\{\frac{\boldsymbol{H}_\ell}{\|\boldsymbol{H}_\ell\|}\hat{U}_\ell\right\}_{\ell=k-\kappa}^{k-1}\right)-\delta_1(\xi_{\min},\kappa) - \delta_2(\xi_{\min},\kappa) && (149)\\ &= I\big(\hat{\boldsymbol{H}}_k\hat{U}_k;\{\hat{\boldsymbol{H}}_\ell\hat{U}_\ell\}_{\ell=k-\kappa}^{k-1}\big)-\delta_1(\xi_{\min},\kappa) - \delta_2(\xi_{\min},\kappa) && (150)\\ &= I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1};\{\hat{\boldsymbol{H}}_\ell\hat{U}_\ell\}_{\ell=1}^{\kappa}\big)-\delta_1(\xi_{\min},\kappa) - \delta_2(\xi_{\min},\kappa). && (151)\end{aligned}$$

Here, the first two inequalities follow from the data processing inequality; (149) follows because the law of $\frac{X_\ell}{|X_\ell|}\hat{V}_\ell$ is identical to the law of $\hat{U}_\ell$ (because the phase of $\hat{V}_\ell$ is uniformly distributed over $[-\pi,\pi)$ and independent of the phase of X_ℓ); and the final equality follows from stationarity.

From (151) and (134), we now have

$$C \leq I(Q)+I(\boldsymbol{H}_0;\boldsymbol{H}_{-\infty}^{-1})+\epsilon+H_{\rm b}\left(\max\left\{\alpha,\frac{1}{2}\right\}\right) - \alpha I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1};\{\hat{\boldsymbol{H}}_\ell\hat{U}_\ell\}_{\ell=1}^{\kappa}\big)-\delta_1(\xi_{\min},\kappa) - \delta_2(\xi_{\min},\kappa). \quad (152)$$

Note that at this point, all dependence on n has disappeared. The bound depends only on the SNR, on ϵ, on $\xi_{\min}$, on κ, and on Q.

We shall next study the limiting behavior of the RHS of (152) as the SNR tends to infinity. We shall begin by showing that

$$\lim_{\mathrm{SNR}\uparrow\infty}\alpha=1. \quad (153)$$

To this end it suffices, by (122), to show that

$$\lim_{\mathrm{SNR}\uparrow\infty}Q(|X|\geq\xi_{\min})=1. \quad (154)$$

But this follows from Theorem 8 because by (7) and (83)

$$\lim_{\mathrm{SNR}\uparrow\infty}\frac{C(\mathrm{SNR})}{\log\log\mathrm{SNR}}=1 \quad (155)$$

and because by (152) and the trivial bound

$$I(Q)\leq C_{\mathrm{i.i.d.}}(\mathrm{SNR})\leq C(\mathrm{SNR}) \quad (156)$$

it follows that

$$\lim_{\mathrm{SNR}\uparrow\infty}\frac{I(Q)}{C(\mathrm{SNR})}=1. \quad (157)$$

Applying the bound $I(Q) \leq C_{\text{i.i.d.}}(\text{SNR})$ to (152) and using (153) we obtain

$$\begin{aligned}\overline{\lim_{\text{SNR}\uparrow\infty}} &\{C(\text{SNR}) - C_{\text{i.i.d.}}(\text{SNR})\} \\ &\leq I(\boldsymbol{H}_0; \boldsymbol{H}_{-\infty}^{-1}) + \epsilon - I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\kappa}\big) \\ &\quad - \delta_1(\xi_{\min}, \kappa) - \delta_2(\xi_{\min}, \kappa). \end{aligned} \tag{158}$$

But since $\epsilon > 0$ can be chosen arbitrarily small and $\xi_{\min}$ can be taken arbitrarily large, it follows from from the above and from (139) and (145) that

$$\begin{aligned}\overline{\lim_{\text{SNR}\uparrow\infty}} &\{C(\text{SNR}) - C_{\text{i.i.d.}}(\text{SNR})\} \\ &\leq I(\boldsymbol{H}_0; \boldsymbol{H}_{-\infty}^{-1}) - I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\kappa}\big). \end{aligned} \tag{159}$$

Upon now letting κ tend to infinity, we obtain

$$\begin{aligned}\chi(\{\boldsymbol{H}_k\}) \leq \chi_{\text{i.i.d.}}(\boldsymbol{H}_0) &+ I(\boldsymbol{H}_0; \boldsymbol{H}_{-\infty}^{-1}) \\ &- I\big(\hat{\boldsymbol{H}}_0\hat{U}_0; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=-\infty}^{-1}\big) \end{aligned} \tag{160}$$

as we had set out to prove.

VIII. Conclusion

This paper provides the fading number for SIMO communication over regular stationary and ergodic fading channels. It demonstrates that, irrespective of the fading law, the fading number is achievable by circularly symmetric inputs of log squared magnitudes that are uniformly distributed over an SNR-dependent interval (33). The fact that this input distribution achieves the fading number for all fading laws has been recently used to derive a lower bound on the fading number of MIMO systems [14, Theorem 4]. As in the scalar case, the fading number does not depend on whether a peak-power constraint or an average-power constraint is imposed on the input. It remains an open problem to determine whether this is the case for MIMO systems as well.

We have also derived upper and lower bounds to the fading number (34) that can be significantly easier to compute than the precise fading number (31). It remains an open problem to determine whether the upper bound in (34) continues to hold also for MIMO systems.

A topic under current investigation is the relationship between the number of degrees of freedom in a MIMO system and its fading number. The number of degrees of freedom is defined as the minimum between the number of transmit antennas and the number of receive antennas. It is conjectured that for very slowly varying systems, where the entries in the fading matrix are i.i.d. stationary Gaussian processes ("spatially white Gaussian fading with memory"), the fading number is roughly proportional to the number of degrees of freedom. The direct part of this statement has been recently proved in [14], but the converse has only been established for the case where the number of transmit antennas is no smaller than the number of receive antennas. The present paper, however, shows that this is also the case for AR(m) SIMO systems. Indeed, comparing (51) and (52) demonstrates that for very slowly varying channels ($\epsilon^2 \ll 1$) the the dominant term in the fading number is $\log 1/\epsilon^2$ irrespective of the number of receive antennas. The coefficient of $\log 1/\epsilon^2$ is thus equal to the number of degrees of freedom (which for SIMO channels is one). This conjecture agrees with the results of Etkin and Tse [15], [16], who have shown that for systems operating at SNRs in which the term $\log(1+\log(1+\text{SNR}))$ is small over spatially white very slowly varying first-order Gauss–Markov fading, capacity is roughly proportional to the number of degrees of freedom.

Appendix I
Equivalence of (31) and (32)

The equivalence of (31) and (32) can be proved as follows. Using stationarity, (19), and(17) we get

$$\begin{aligned}\chi_{\text{i.i.d.}}&\big(\boldsymbol{H}_{\kappa+1}\big|\boldsymbol{H}_1^{\kappa}, \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) \\ &= h_\lambda\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1}\big|\boldsymbol{H}_1^{\kappa}, \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) \\ &\quad - h\big(\boldsymbol{H}_{\kappa+1}\big|\boldsymbol{H}_1^{\kappa}, \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) + n_{\text{R}}\mathsf{E}\left[\log\|\boldsymbol{H}_{\kappa+1}\|^2\right] \\ &\quad - \log 2 \\ &= \chi_{\text{i.i.d.}}(\boldsymbol{H}_{\kappa+1}) - I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \boldsymbol{H}_1^{\kappa}, \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) \\ &\quad + I\big(\boldsymbol{H}_{\kappa+1}; \boldsymbol{H}_1^{\kappa}, \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) \\ &= \chi_{\text{i.i.d.}}(\boldsymbol{H}_0) + I(\boldsymbol{H}_{\kappa+1}; \boldsymbol{H}_1^{\kappa}) \\ &\quad - I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\kappa}\big) + \delta_\kappa \end{aligned} \tag{161}$$

where the last equality should be read as the definition of δ_κ:

$$\begin{aligned}\delta_\kappa &= I\big(\boldsymbol{H}_{\kappa+1}; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big|\boldsymbol{H}_1^{\kappa}\big) \\ &\quad + I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\kappa}\big) \\ &\quad - I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \boldsymbol{H}_1^{\kappa}, \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) \\ &= I\big(\boldsymbol{H}_{\kappa+1}; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big|\boldsymbol{H}_1^{\kappa}\big) \\ &\quad + I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\kappa}\big) \\ &\quad - I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) \\ &\quad - I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \boldsymbol{H}_1^{\kappa}\big|\{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) \\ &= I\big(\boldsymbol{H}_{\kappa+1}; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big|\boldsymbol{H}_1^{\kappa}\big) \\ &\quad - I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \boldsymbol{H}_1^{\kappa}\big|\{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) \end{aligned} \tag{162}$$

$$\begin{aligned} &= h_\lambda\big(\{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big|\boldsymbol{H}_1^{\kappa}\big) - h_\lambda\big(\{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big|\boldsymbol{H}_1^{\kappa+1}\big) \\ &\quad - h\big(\boldsymbol{H}_1^{\kappa}\big|\{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) + h\big(\boldsymbol{H}_1^{\kappa}\big|\{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+1}^{2\kappa+1}\big) \\ &= h_\lambda\big(\{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) - h_\lambda\big(\{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big|\boldsymbol{H}_1^{\kappa+1}\big) \\ &\quad - h\big(\boldsymbol{H}_1^{\kappa}\big) + h\big(\boldsymbol{H}_1^{\kappa}\big|\{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+1}^{2\kappa+1}\big) \\ &= I\big(\boldsymbol{H}_1^{\kappa+1}; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) - I\big(\boldsymbol{H}_1^{\kappa}; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+1}^{2\kappa+1}\big) \\ &\to 0, \qquad \text{for } \kappa\uparrow\infty, \text{ by stationarity.} \end{aligned} \tag{163}$$

In (162) we have used

$$I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\kappa}\big) = I\big(\hat{\boldsymbol{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\boldsymbol{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big)$$

which follows by stationarity and in (163) we have used

$$h(A|B) - h(B|A) = h(A) - h(B).$$

Appendix II
Appendix for the Proof of the Lower Bound

In the derivation of the lower bound it remains to derive the upper bounds $\epsilon_1(x_{\min}, \kappa)$ and $\epsilon_2(x_{\min}, \kappa)$ to

$$I\big(X_k; \boldsymbol{Z}_{k-\kappa}^{k-1}\big|X_{k-\kappa}^{k-1}, \boldsymbol{Y}_{k-\kappa}^{k-1}, \boldsymbol{Y}_k^{k+\kappa}\big)$$

and

$$I\big(X_k; \boldsymbol{Z}_{k+1}^{k+\kappa}\big|\boldsymbol{H}_{k-\kappa}^{k-1}, \boldsymbol{Y}_k, \boldsymbol{Y}_{k+1}^{k+\kappa}\big)$$

respectively, and to show that both bounds do not depend on k and tend to zero as $x_{\min}$ tends to infinity.

We start with the former

$$\begin{aligned}
I\big(X_k; \boldsymbol{Z}_{k-\kappa}^{k-1}\big|X_{k-\kappa}^{k-1}, \boldsymbol{Y}_{k-\kappa}^{k-1}, \boldsymbol{Y}_k^{k+\kappa}\big)
&= h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big|X_{k-\kappa}^{k-1}, \boldsymbol{Y}_{k-\kappa}^{k-1}, \boldsymbol{Y}_k^{k+\kappa}\big) \\
&\quad - h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big|X_{k-\kappa}^{k-1}, \boldsymbol{Y}_{k-\kappa}^{k-1}, X_k, \boldsymbol{Y}_k^{k+\kappa}\big) \\
&\le h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big) \\
&\quad - h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big|X_{k-\kappa}^{k-1}, \boldsymbol{Y}_{k-\kappa}^{k-1}, X_k^{k+\kappa}, \boldsymbol{Z}_k^{k+\kappa}, \boldsymbol{Y}_k^{k+\kappa}\big) \qquad (164) \\
&= h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big) - h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big|X_{k-\kappa}^{k-1}, \boldsymbol{Y}_{k-\kappa}^{k-1}, \boldsymbol{H}_k^{k+\kappa}\big) \\
&\le h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big) - \min_{\substack{|x_{k-\kappa}|\ge x_{\min},\ldots,\\ |x_{k-1}|\ge x_{\min}}} h\Big(\boldsymbol{Z}_{k-\kappa}^{k-1}\Big|X_{k-\kappa}^{k-1} = x_{k-\kappa}^{k-1}, \\
&\qquad\qquad \{\boldsymbol{H}_\ell x_\ell + \boldsymbol{Z}_\ell\}_{\ell=k-\kappa}^{k-1}, \boldsymbol{H}_k^{k+\kappa}\Big) \\
&= h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big) - h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big|\{\boldsymbol{H}_\ell x_{\min} + \boldsymbol{Z}_\ell\}_{\ell=k-\kappa}^{k-1}, \boldsymbol{H}_k^{k+\kappa}\big) \\
&= I\left(\boldsymbol{Z}_{k-\kappa}^{k-1}; \left\{\boldsymbol{H}_\ell + \frac{\boldsymbol{Z}_\ell}{x_{\min}}\right\}_{\ell=k-\kappa}^{k-1}, \boldsymbol{H}_k^{k+\kappa}\right) \\
&= I\left(\boldsymbol{Z}_1^{\kappa}; \left\{\boldsymbol{H}_\ell + \frac{\boldsymbol{Z}_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa}, \boldsymbol{H}_{\kappa+1}^{2\kappa+1}\right) \\
&= I\left(\left\{\frac{\boldsymbol{Z}_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa}; \left\{\boldsymbol{H}_\ell + \frac{\boldsymbol{Z}_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa}\middle|\boldsymbol{H}_{\kappa+1}^{2\kappa+1}\right) \qquad (165) \\
&= h\left(\left\{\boldsymbol{H}_\ell + \frac{\boldsymbol{Z}_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa}\middle|\boldsymbol{H}_{\kappa+1}^{2\kappa+1}\right) - h\big(\boldsymbol{H}_1^{\kappa}\big|\boldsymbol{H}_{\kappa+1}^{2\kappa+1}\big) \\
&\triangleq \epsilon_1(x_{\min}, \kappa). \qquad (166)
\end{aligned}$$

Here (164) follows from conditioning that reduces entropy; and in the subsequent equality we used $X_{k-\kappa}^{k-1}$ and $\boldsymbol{Z}_k^{k+\kappa}$ in order to extract $\boldsymbol{H}_k^{k+\kappa}$ from $\boldsymbol{Y}_k^{k+\kappa}$, and then we dropped

$$\{X_\ell, \boldsymbol{Y}_\ell, \boldsymbol{Z}_\ell\}_{\ell=k}^{k+\kappa}$$

since given $\boldsymbol{H}_k^{k+\kappa}$ it is independent of the other random variables.

From [1, Lemma 6.11] we conclude that for any realization of $\boldsymbol{H}_{\kappa+1}^{2\kappa+1}$ the expression

$$h\left(\left\{\boldsymbol{H}_\ell + \frac{\boldsymbol{Z}_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa}\middle|\boldsymbol{H}_{\kappa+1}^{2\kappa+1} = \boldsymbol{h}_{\kappa+1}^{2\kappa+1}\right)$$

converges monotonically in $x_{\min}$ to $h\big(\boldsymbol{H}_1^{\kappa}\big|\boldsymbol{H}_{\kappa+1}^{2\kappa+1} = \boldsymbol{h}_{\kappa+1}^{2\kappa+1}\big)$. By the Monotone Convergence Theorem this is also true when we average over $\boldsymbol{H}_{\kappa+1}^{2\kappa+1}$.

Similarly, we get for $I\big(X_k; \boldsymbol{Z}_{k+1}^{k+\kappa}\big|\boldsymbol{H}_{k-\kappa}^{k-1}, \boldsymbol{Y}_k, \boldsymbol{Y}_{k+1}^{k+\kappa}\big)$

$$\begin{aligned}
I\big(X_k; \boldsymbol{Z}_{k+1}^{k+\kappa}\big|\boldsymbol{H}_{k-\kappa}^{k-1}, \boldsymbol{Y}_k, \boldsymbol{Y}_{k+1}^{k+\kappa}\big)
&= h\big(\boldsymbol{Z}_{k+1}^{k+\kappa}\big|\boldsymbol{H}_{k-\kappa}^{k-1}, \boldsymbol{Y}_k, \boldsymbol{Y}_{k+1}^{k+\kappa}\big) \\
&\quad - h\big(\boldsymbol{Z}_{k+1}^{k+\kappa}\big|\boldsymbol{H}_{k-\kappa}^{k-1}, X_k, \boldsymbol{Y}_k, \boldsymbol{Y}_{k+1}^{k+\kappa}\big) \\
&\le h\big(\boldsymbol{Z}_{k+1}^{k+\kappa}\big) - h\big(\boldsymbol{Z}_{k+1}^{k+\kappa}\big|\boldsymbol{H}_{k-\kappa}^{k-1}, X_k, \boldsymbol{Y}_k, \boldsymbol{Z}_k, \boldsymbol{Y}_{k+1}^{k+\kappa}, X_{k+1}^{k+\kappa}\big) \\
&= h\big(\boldsymbol{Z}_{k+1}^{k+\kappa}\big) - h\big(\boldsymbol{Z}_{k+1}^{k+\kappa}\big|\boldsymbol{H}_{k-\kappa}^{k}, \boldsymbol{Y}_{k+1}^{k+\kappa}, X_{k+1}^{k+\kappa}\big) \\
&\le h\big(\boldsymbol{Z}_{k+1}^{k+\kappa}\big) \\
&\quad - \min_{\substack{|x_{k+1}|\ge x_{\min},\ldots,\\ |x_{k+\kappa}|\ge x_{\min}}} h\Big(\boldsymbol{Z}_{k+1}^{k+\kappa}\Big|\boldsymbol{H}_{k-\kappa}^{k}, \{\boldsymbol{H}_\ell x_\ell + \boldsymbol{Z}_\ell\}_{\ell=k+1}^{k+\kappa}, \\
&\qquad\qquad X_{k+1}^{k+\kappa} = x_{k+1}^{k+\kappa}\Big) \\
&= h\big(\boldsymbol{Z}_{k+1}^{k+\kappa}\big) - h\big(\boldsymbol{Z}_{k+1}^{k+\kappa}\big|\boldsymbol{H}_{k-\kappa}^{k}, \{\boldsymbol{H}_\ell x_{\min} + \boldsymbol{Z}_\ell\}_{\ell=k+1}^{k+\kappa}\big) \\
&= I\left(\boldsymbol{Z}_{k+1}^{k+\kappa}; \left\{\boldsymbol{H}_\ell + \frac{\boldsymbol{Z}_\ell}{x_{\min}}\right\}_{\ell=k+1}^{k+\kappa}, \boldsymbol{H}_{k-\kappa}^{k}\right) \\
&= I\left(\boldsymbol{Z}_{\kappa+2}^{2\kappa+1}; \left\{\boldsymbol{H}_\ell + \frac{\boldsymbol{Z}_\ell}{x_{\min}}\right\}_{\ell=\kappa+2}^{2\kappa+1}, \boldsymbol{H}_1^{\kappa+1}\right) \\
&= I\left(\left\{\frac{\boldsymbol{Z}_\ell}{x_{\min}}\right\}_{\ell=\kappa+2}^{2\kappa+1}; \left\{\boldsymbol{H}_\ell + \frac{\boldsymbol{Z}_\ell}{x_{\min}}\right\}_{\ell=\kappa+2}^{2\kappa+1}\middle|\boldsymbol{H}_1^{\kappa+1}\right) \\
&\triangleq \epsilon_2(x_{\min}, \kappa) \qquad (167)
\end{aligned}$$

from which the results follows analogously to (165).

Appendix III
Appendix for the Proof of the Upper Bound

In the derivation of the upper bound, it remains to derive upper bounds $\delta_1(\xi_{\min}, \kappa)$ to

$$I\big(\boldsymbol{Y}_k; \boldsymbol{Z}_{k-\kappa}^{k-1}\big|\boldsymbol{Y}_{k-\kappa}^{k-1}, E_k{=}1\big)$$

and $\delta_2(\xi_{\min}, \kappa)$ to

$$I\big(\boldsymbol{Z}_k; \{\boldsymbol{H}_\ell X_\ell\}_{k-\kappa}^{k-1}\big|\boldsymbol{Y}_k, E_k{=}1\big)$$

that depend only on $\xi_{\min}$ and not on k or the SNR and that tend to zero as $\xi_{\min} \to \infty$.

We start with the bound on $I\big(\boldsymbol{Y}_k; \boldsymbol{Z}_{k-\kappa}^{k-1}\big|\boldsymbol{Y}_{k-\kappa}^{k-1}, E_k{=}1\big)$:

$$\begin{aligned}
I\big(\boldsymbol{Y}_k; \boldsymbol{Z}_{k-\kappa}^{k-1}\big|\boldsymbol{Y}_{k-\kappa}^{k-1}, E_k{=}1\big)
&= h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big|\boldsymbol{Y}_{k-\kappa}^{k-1}, E_k{=}1\big) - h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big|\boldsymbol{Y}_{k-\kappa}^{k-1}, \boldsymbol{Y}_k, E_k{=}1\big) \\
&\le h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big) - h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big|\boldsymbol{Y}_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, \boldsymbol{Y}_k, \boldsymbol{Z}_k, X_k, E_k{=}1\big) \\
&= h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big) - h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big|\boldsymbol{Y}_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, \boldsymbol{H}_k, E_k{=}1\big) \\
&\le h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big) - \min_{\substack{|x_{k-\kappa}|\ge \xi_{\min},\ldots,\\ |x_{k-1}|\ge \xi_{\min}}} h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big|\{\boldsymbol{H}_\ell x_\ell + \boldsymbol{Z}_\ell\}_{k-\kappa}^{k-1}, \\
&\qquad\qquad X_{k-\kappa}^{k-1} = x_{k-\kappa}^{k-1}, \boldsymbol{H}_k, E_k{=}1\big) \\
&= h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big) - h\big(\boldsymbol{Z}_{k-\kappa}^{k-1}\big|\{\boldsymbol{H}_\ell \xi_{\min} + \boldsymbol{Z}_\ell\}_{k-\kappa}^{k-1}, \boldsymbol{H}_k\big) \\
&= I\left(\boldsymbol{Z}_{k-\kappa}^{k-1}; \left\{\boldsymbol{H}_\ell + \frac{\boldsymbol{Z}_\ell}{\xi_{\min}}\right\}_{\ell=k-\kappa}^{k-1}, \boldsymbol{H}_k\right) \\
&= I\left(\boldsymbol{Z}_{k-\kappa}^{k-1}; \left\{\boldsymbol{H}_\ell + \frac{\boldsymbol{Z}_\ell}{\xi_{\min}}\right\}_{\ell=k-\kappa}^{k-1}\middle|\boldsymbol{H}_k\right) \\
&= h\left(\left\{\boldsymbol{H}_\ell + \frac{\boldsymbol{Z}_\ell}{\xi_{\min}}\right\}_{\ell=k-\kappa}^{k-1}\middle|\boldsymbol{H}_k\right) - h\big(\boldsymbol{H}_{k-\kappa}^{k-1}\big|\boldsymbol{H}_k\big) \\
&= h\left(\left\{\boldsymbol{H}_\ell + \frac{\boldsymbol{Z}_\ell}{\xi_{\min}}\right\}_{\ell=1}^{\kappa}\middle|\boldsymbol{H}_{\kappa+1}\right) - h\big(\boldsymbol{H}_1^{\kappa}\big|\boldsymbol{H}_{\kappa+1}\big) \\
&\triangleq \delta_1(\xi_{\min}, \kappa) \qquad (168)
\end{aligned}$$

where the last step follows from stationarity.

From [1, Lemma 6.11], we conclude that for any realization of $\boldsymbol{H}_{\kappa+1}$, the expression

$$h\left(\left\{\boldsymbol{H}_\ell + \frac{\boldsymbol{Z}_\ell}{\xi_{\min}}\right\}_{\ell=1}^{\kappa}\middle|\boldsymbol{H}_{\kappa+1} = \boldsymbol{h}_{\kappa+1}\right)$$

converges monotonically in $\xi_{\min}$ to $h\big(\boldsymbol{H}_1^{\kappa}\big|\boldsymbol{H}_{\kappa+1}=\boldsymbol{h}_{\kappa+1}\big)$. By the Monotone Convergence Theorem this is also true when we average over $\boldsymbol{H}_{\kappa+1}$.

Similarly, we bound $I\big(\boldsymbol{Z}_k;\{\boldsymbol{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|\boldsymbol{Y}_k,E_k{=}1\big)$

$$\begin{aligned}
I&\big(\boldsymbol{Z}_k;\{\boldsymbol{H}_\ell X_\ell\}_{k-\kappa}^{k-1}\big|\boldsymbol{Y}_k,E_k{=}1\big)\\
&=h\big(\boldsymbol{Z}_k\big|\boldsymbol{Y}_k,E_k{=}1\big)-h\big(\boldsymbol{Z}_k\big|\{\boldsymbol{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1},\boldsymbol{Y}_k,E_k{=}1\big)\\
&\le h\big(\boldsymbol{Z}_k\big)-h\big(\boldsymbol{Z}_k\big|\{\boldsymbol{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1},X_{k-\kappa}^{k-1},\boldsymbol{Y}_k,X_k,E_k{=}1\big)\\
&=h\big(\boldsymbol{Z}_k\big)-h\big(\boldsymbol{Z}_k\big|\boldsymbol{H}_{k-\kappa}^{k-1},\boldsymbol{Y}_k,X_k,E_k{=}1\big)\\
&\le h\big(\boldsymbol{Z}_k\big)-\min_{|x_k|\ge\xi_{\min}}h\big(\boldsymbol{Z}_k\big|\boldsymbol{H}_{k-\kappa}^{k-1},\boldsymbol{H}_k x_k+\boldsymbol{Z}_k,\\
&\qquad\qquad\qquad\qquad\qquad X_k=x_k,E_k{=}1\big)\\
&=h\big(\boldsymbol{Z}_k\big)-h\big(\boldsymbol{Z}_k\big|\boldsymbol{H}_{k-\kappa}^{k-1},\boldsymbol{H}_k\xi_{\min}+\boldsymbol{Z}_k\big)\\
&=I\left(\boldsymbol{Z}_k;\boldsymbol{H}_k+\frac{\boldsymbol{Z}_k}{\xi_{\min}},\boldsymbol{H}_{k-\kappa}^{k-1}\right)\\
&=I\left(\boldsymbol{Z}_k;\boldsymbol{H}_k+\frac{\boldsymbol{Z}_k}{\xi_{\min}}\middle|\boldsymbol{H}_{k-\kappa}^{k-1}\right)\\
&=h\left(\boldsymbol{H}_k+\frac{\boldsymbol{Z}_k}{\xi_{\min}}\middle|\boldsymbol{H}_{k-\kappa}^{k-1}\right)-h\big(\boldsymbol{H}_k\big|\boldsymbol{H}_{k-\kappa}^{k-1}\big)\\
&=h\left(\boldsymbol{H}_{\kappa+1}+\frac{\boldsymbol{Z}_{\kappa+1}}{\xi_{\min}}\middle|\boldsymbol{H}_1^{\kappa}\right)-h\big(\boldsymbol{H}_{\kappa+1}\big|\boldsymbol{H}_1^{\kappa}\big)\\
&\triangleq\delta_2(\xi_{\min},\kappa)
\end{aligned}\tag{169}$$

where the last step follows from stationarity.

Again, from [1, Lemma 6.11], we conclude that for any realization of $\boldsymbol{H}_1^{\kappa}$ the expression

$$h\left(\boldsymbol{H}_{\kappa+1}+\frac{1}{\xi_{\min}}\boldsymbol{Z}_{\kappa+1}\middle|\boldsymbol{H}_1^{\kappa}=\boldsymbol{h}_1^{\kappa}\right)$$

converges monotonically in $\xi_{\min}$ to $h\big(\boldsymbol{H}_{\kappa+1}\big|\boldsymbol{H}_1^{\kappa}=\boldsymbol{h}_1^{\kappa}\big)$. By the Monotone Convergence Theorem this is also true when we average over $\boldsymbol{H}_1^{\kappa}$.

ACKNOWLEDGMENT

The authors gratefully acknowledge İ. E. Telatar's contribution to the proofs in Appendices II and III. They thank the anonymous reviewers for their helpful comments and T. Koch for producing Fig. 1.

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] A. Lapidoth, "Capacity bounds via duality: A phase noise example," in *Proc. 2nd Asian-European Workshop on Information Theory*, Breisach, Germany, Jun. 2002, pp. 58–61.

[3] ——, "On phase noise channels at high SNR," in *Proc. IEEE Information Theory Workshop (ITW)*, Bangalore, India, Oct. 2002, pp. 1–4.

[4] A. Lapidoth and S. M. Moser, "The asymptotic capacity of the discrete-time Poisson channel," in *Proc. Winter School on Coding and Information Theory*, Monte Verità, Ascona, Switzerland, Feb. 2003, pp. 113–118.

[5] ——, "Bounds on the capacity of the discrete-time Poisson channel," in *Proc. 41st Allerton Conf. Communication, Control and Computing*, Monticello, IL, Oct. 2003, pp. 201–210.

[6] S. M. Moser, "Duality-Based Bounds on Channel Capacity," Ph.D. dissertation, Swiss Fed. Inst. Technol., Zurich, Switzerland, 2004.

[7] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[8] W. C. Jakes, Jr., *Microwave Mobile Communications*. New York: Wiley, 1974.

[9] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[10] L. Zheng and D. N. C. Tse, "Communicating on the Grassmann manifold: A geometric approach to the non coherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.

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

[12] A. Lapidoth, "On the high SNR capacity of stationary Gaussian fading channels," in *Proc. 41st Allerton Conf. Communication, Control and Computing*, Monticello, IL, Oct. 2003, pp. 410–419.

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

[14] T. Koch and A. Lapidoth, "Degrees of freedom in noncoherent stationary MIMO fading channels," in *Proc. Winter School on Coding and Information Theory*, Bratislava, Slovakia, Feb. 2005, pp. 91–97. See also: 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. Information Theory*, Adelaide, Australia, Sep. 2005, pp. 661–665.

[15] R. Etkin and D. N. C. Tse, "Degrees of freedom in underspread MIMO fading channels," *IEEE Trans. Inf. Theory*, submitted for publication.

[16] ——, "Degrees of freedom in underspread MIMO fading channels," in *Proc. IEEE Int. Symp. Information Theory*, Yokohama, Japan, Jun./Jul. 2003, p. 323.

[17] M. Pourahmadi, *Foundations of Time Series Analysis and Prediction Theory*. New York: Wiley, 2001.

[18] A. M. Yaglom, *An Introduction to the Theory of Stationary Random Functions*, R. A. Silverman, Ed. Englewood Cliffs, NJ: Prentice-Hall, 1962.

[19] A. N. Kolmogorov, "Stationary sequences in Hilbert space," *Bull. Moscow State Univ.*, vol. 2, no. 6, 1941.

[20] ——, "Interpolation and extrapolation of stationary random sequences," *Izv. Akad, Nauk SSSR*, ser. Mat., vol. 5, p. 3, 1941.

[21] J. Anděl, "A comparison of methods for interpolation in time series," in *Mathematical Statistics and Applications (Proc. 4th Pannonian Symp. Mathematical Statistics)*, vol. B, W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Eds., Bad Tatzmannsdorf, Austria, Sep. 1983, pp. 53–67.

[22] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular Variation*. ser. Encyclopedia of Mathematics and its Applications, G.-C. Rota, Ed. Cambridge, U.K.: Cambridge Univ. Press, 1987, vol. 27.