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

Amos Lapidoth Stefan M. Moser*

October 29, 2004

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 IID 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 sub-logarithmically 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 m-th order auto-regressive AR(m) Gaussian fading processes.

KEYWORDS: Auto-regressive process, channel capacity, fading, fading number, high SNR, memory, multiple-antenna, SIMO.

1 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 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.

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

*The authors are with the Department of Information Technology and Electrical Engineering, at the Swiss Federal Institute of Technology (ETH) in Zurich, Switzerland. The work of S. M. Moser was supported in part by the ETH under TH-23 02-2.

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 new result holds for general channels and not only for fading channels. It is hoped that it may be of some independent interest and that it may find applications in the study of other channels as well.

The paper is structured as follows. After concluding this introductory section with some notes on notation, we proceed in Section 2 to introduce the channel model and to define the fading number. Section 3 summarizes some relevant known results, while Section 4 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 5, which also includes the example of stationary m-th order auto-regressive AR(m) Gaussian fading processes. Finally, Section 6 contains the proof of the main result. The new observation regarding "input distributions that escape to infinity" can be found in Section 6.3.3, which is essentially self contained.

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 IID 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 that we capitalize and the energy-per-symbol, which we denote by $\mathcal{E}_\text{s}$. Both are deterministic. We use boldface fonts to denote vectors, *e.g.*, $\mathbf{x}$ for a deterministic vector and $\mathbf{X}$ for a random vector. We use the shorthand H_a^b for $(H_a, H_{a+1}, \ldots, H_b)$. In case the expression is more complicated like, *e.g.*, $(H_a\hat{U}_a, H_{a+1}\hat{U}_{a+1}, \ldots, H_b\hat{U}_b)$, we use the dummy variable ℓ to clarify notation: $\{H_\ell\hat{U}_\ell\}_{\ell=a}^b$.

2 The Channel Model and the Fading Number

We consider a single-input multiple-output (SIMO) fading channel whose time-k output $\mathbf{Y}_k \in \mathbb{C}^{n_\text{R}}$ is given by

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

where $x_k \in \mathbb{C}$ denotes the time-k channel input; the random vector $\mathbf{H}_k \in \mathbb{C}^{n_\text{R}}$ denotes the time-k fading vector; and where $\mathbf{Z}_k$ denotes additive noise. Here $\mathbb{C}$ denotes the complex field, $\mathbb{C}^{n_\text{R}}$ 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 $\sigma^2 \mathsf{I}_{n_\text{R}}$, where $\sigma^2 > 0$ and where I_{n_R} denotes the $n_\text{R} \times n_\text{R}$ identity matrix. Thus, $\{\mathbf{Z}_k\}$ is a zero-mean, circularly symmetric, stationary, multi-variate, Gaussian process such that $\mathsf{E}\left[\mathbf{Z}_k\mathbf{Z}_{k+m}^\dagger\right]$ is the zero matrix if $m \neq 0$, and is $\sigma^2\mathsf{I}_{n_\text{R}}$ for $m = 0$. Here $()^\dagger$ denotes Hermitian conjugation.

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

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

and of finite differential entropy rate

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

Finally, we assume that the fading process $\{\mathbf{H}_k\}$ and the additive noise process $\{\mathbf{Z}_k\}$ 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

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

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

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

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

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

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.

3 Previous Results

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

$$\chi(\{H_k\}) = \log\pi + \mathsf{E}\left[\log|H_1|^2\right] - h\big(\{H_k\}\big), \tag{9}$$

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

$$\chi_{\mathrm{IID}}(\mathbf{H}) = I\big(\hat{U}; \mathbf{H}\hat{U}\big) + \mathsf{E}[\log\|\mathbf{H}\|] - h\big(\|\mathbf{H}\| \,\big|\, \hat{\mathbf{H}}\hat{U}\big) - \log 2, \tag{10}$$

where $\hat{U}$ is defined in (1), and where $\hat{\mathbf{H}} = \mathbf{H}/\|\mathbf{H}\|$. Alternatively, $\chi_{\mathrm{IID}}(\mathbf{H})$ can be expressed as

$$\chi_{\mathrm{IID}}(\mathbf{H}) = h_\lambda(\hat{\mathbf{H}}\hat{U}) - h(\mathbf{H}) + n_\mathrm{R}\mathsf{E}\left[\log\|\mathbf{H}\|^2\right] - \log 2, \tag{11}$$

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

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

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

$$I(\mathbf{H}; \mathbf{S}) < \infty. \tag{12}$$

In this case

$$\chi_{\text{IID}}(\mathbf{H}|\mathbf{S}) = h_\lambda(\hat{\mathbf{H}}\hat{U}|\mathbf{S}) - h(\mathbf{H}|\mathbf{S}) + n_\text{R}\mathsf{E}\left[\log \|\mathbf{H}\|^2\right] - \log 2. \tag{13}$$

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

$$h_\lambda(\mathbf{G}|\mathbf{S} = \mathbf{s}) = -\int f_{\mathbf{G}|\mathbf{S}}(\mathbf{g}|\mathbf{s}) \log f_{\mathbf{G}|\mathbf{S}}(\mathbf{g}|\mathbf{s}) \,\mathrm{d}\lambda(\mathbf{g}) \tag{14}$$

and

$$h_\lambda(\mathbf{G}|\mathbf{S}) = \int h_\lambda(\mathbf{G}|\mathbf{S} = \mathbf{s}) \,\mathrm{d}P(\mathbf{s}). \tag{15}$$

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

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

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

$$\chi_{\text{IID}}(\mathsf{D} + \tilde{\mathbb{H}}) \geq \log \|\mathsf{D}\|^2 - \text{Ei}\big(-\|\mathsf{D}\|^2\big) - 1, \tag{16}$$

$$\chi_{\text{IID}}(\mathsf{D} + \tilde{\mathbb{H}}) \leq \min\{n_\text{R}, n_\text{T}\} \log\left(1 + \frac{\|\mathsf{D}\|^2}{\min\{n_\text{R}, n_\text{T}\}}\right) + n_\text{R} \log n_\text{R} - n_\text{R} - \log \Gamma(n_\text{R}), \tag{17}$$

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

$$\text{Ei}(-\xi) = -\int_\xi^\infty \frac{e^{-t}}{t} \,\mathrm{d}t, \qquad \xi > 0; \tag{18}$$

$\Gamma(\cdot)$ denotes the Gamma function so that $\Gamma(n_\text{R}) = (n_\text{R} - 1)!$; and the term $\log(\xi) - \text{Ei}(-\xi)$ 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

$$\chi_{\text{IID}}(\mathbf{d} + \tilde{\mathbf{H}}) \geq \log \|\mathbf{d}\|^2 - \text{Ei}\big(-\|\mathbf{d}\|^2\big) - 1, \tag{19}$$

$$\chi_{\text{IID}}(\mathbf{d} + \tilde{\mathbf{H}}) \leq \log(1 + \|\mathbf{d}\|^2) + n_\text{R} \log n_\text{R} - n_\text{R} - \log \Gamma(n_\text{R}), \tag{20}$$

where the n_R components of $\tilde{\mathbf{H}}$ are IID $\mathcal{N}_\mathbb{C}(0, 1)$. More generally, if $\mathbf{H}$ is a multi-variate circularly symmetric complex Gaussian of mean $\mathbf{d}$ and covariance $\boldsymbol{\Sigma}$

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

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

$$\chi_{\text{IID}}(\mathbf{H}) \geq \log \sum_{r=1}^{n_\text{R}} \left| \frac{(\mathsf{V}^\mathsf{T}\mathbf{d})^{(r)}}{\sqrt{\lambda^{(r)}}} \right|^2 - \text{Ei}\left(-\sum_{r=1}^{n_\text{R}} \left| \frac{(\mathsf{V}^\mathsf{T}\mathbf{d})^{(r)}}{\sqrt{\lambda^{(r)}}} \right|^2\right) - 1, \tag{21}$$

$$\chi_{\text{IID}}(\mathbf{H}) \leq \log\left(1 + \sum_{r=1}^{n_\text{R}} \left| \frac{(\mathsf{V}^\mathsf{T}\mathbf{d})^{(r)}}{\sqrt{\lambda^{(r)}}} \right|^2\right) + n_\text{R} \log n_\text{R} - n_\text{R} - \log \Gamma(n_\text{R}), \tag{22}$$

where V is unitary and diagonalizes $\mathsf{\Sigma}$:

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

This follows because, by [1, Lemma 4.7], for any non-singular deterministic $n_\mathrm{R} \times n_\mathrm{R}$ matrix G

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

The choice

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

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

In particular, if $\mathsf{\Sigma}$ is diagonal,

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

then by (21) and (22)

$$\chi_\mathrm{IID}(\mathbf{H}) \geq \log \sum_{r=1}^{n_\mathrm{R}} \frac{\left|d^{(r)}\right|^2}{\lambda^{(r)}} - \mathrm{Ei}\left(-\sum_{r=1}^{n_\mathrm{R}} \frac{\left|d^{(r)}\right|^2}{\lambda^{(r)}}\right) - 1, \quad \mathsf{\Sigma} = \mathrm{diag}\left(\{\lambda^{(r)}\}_{r=1}^{n_\mathrm{R}}\right), \tag{23}$$

$$\chi_\mathrm{IID}(\mathbf{H}) \leq \log\left(1 + \sum_{r=1}^{n_\mathrm{R}} \frac{\left|d^{(r)}\right|^2}{\lambda^{(r)}}\right) + n_\mathrm{R}\log n_\mathrm{R} - n_\mathrm{R} - \log\Gamma(n_\mathrm{R}),$$
$$\mathsf{\Sigma} = \mathrm{diag}\left(\{\lambda^{(r)}\}_{r=1}^{n_\mathrm{R}}\right). \tag{24}$$

4 Main Result

Theorem 1. *Consider a SIMO fading channel with memory (2) where the fading process $\{\mathbf{H}_k\}$ takes value in $\mathbb{C}^{n_\mathrm{R}}$ and satisfies $h(\{\mathbf{H}_k\}) > -\infty$ and $\mathsf{E}\left[\|\mathbf{H}_k\|^2\right] < \infty$. 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 $\chi(\{\mathbf{H}_k\})$ is given by*

$$\boxed{\chi(\{\mathbf{H}_k\}) = \chi_\mathrm{IID}\left(\mathbf{H}_0 \,\middle|\, \mathbf{H}_{-\infty}^{-1}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\infty}\right).} \tag{25}$$

Here $\chi_\mathrm{IID}(\mathbf{H}_0 | \mathbf{S})$ is given in (13)*, the random process $\{\hat{U}_\ell\}$ is independent of $\{\mathbf{H}_k\}$ and constitutes of IID 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 $\hat{\mathbf{H}}_\ell$ is defined as

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

Equivalently, the fading number is given by

$$\chi(\{\mathbf{H}_k\}) = \chi_\mathrm{IID}(\mathbf{H}_0) + I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) - I\left(\hat{\mathbf{H}}_0 \hat{U}_0; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=-\infty}^{-1}\right), \tag{26}$$

where $\chi_\mathrm{IID}(\mathbf{H}_0)$ is defined in (11)*.*

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

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

Proof. See Section 6. □

Corollary 2. *From Theorem 1 it follows that*

$$\chi_{\text{IID}}\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}\big) \leq \chi(\{\mathbf{H}_k\}) \leq \chi_{\text{IID}}\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, \mathbf{H}_1^{\infty}\big). \tag{27}$$

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* (9) *to obtain a lower bound on the fading number of a SIMO system:*

$$\chi(\{\mathbf{H}_k\}) \geq \sup\left\{\log \pi + \mathsf{E}\left[\log |\tilde{H}_1|^2\right] - h\big(\{\tilde{H}_k\}\big)\right\}, \tag{28}$$

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

5 Gaussian Fading with Memory

Since it is difficult to evaluate analytically the fading number (25) even for Gaussian fading, we shall next use the bounds (27) to approximate it. We shall only treat here 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

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

are independent.[1]

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

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

Thus, the n_{R} components of the vector-valued process $\{\mathbf{H}_k\}$ are independent and for each $1 \leq r \leq n_{\mathrm{R}}$ the process $\{H_k^{(r)} - \mu^{(r)}\}_{k=-\infty}^{\infty}$ is a stationary circularly symmetric scalar Gaussian process of spectral distribution $\mathsf{F}^{(r)}$ so that

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

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

To evaluate the lower bound of (27) 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.*, [2], [3])

$$\epsilon_{\text{pred},r}^2 = \exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \log \mathsf{F}'^{(r)}(\lambda)\, \mathrm{d}\lambda\right). \tag{29}$$

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

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

[1] In the more general case one may still resort to (28) which is, however, not tight.

and of variance $\epsilon^2_{\text{pred},r}$. Unconditionally, $H^{(r)}_{\text{pred}}$ is Gaussian of mean $\mu^{(r)}$ and of variance

$$\begin{aligned}\mathsf{Var}\left(H^{(r)}_{\text{pred}}\right) &= \mathsf{Var}\left(H^{(r)}_0\right) - \epsilon^2_{\text{pred},r}\\ &= \mathsf{F}^{(r)}(1/2) - \mathsf{F}^{(r)}(-1/2) - \epsilon^2_{\text{pred},r}.\end{aligned}$$

Similarly, to evaluate the upper bound of (27) on the fading number we shall need the conditional law of the present fading given its past & future. To this end we recall that the optimum *interpolation error* in estimating $H^{(r)}_0$ from its infinite past and future

$$(\{H^{(r)}_\ell\}^{-1}_{\ell=-\infty}, \{H^{(r)}_\ell\}^{\infty}_{\ell=1})$$

is the optimum *linear interpolation error* given by (see [3, Sec. 37.2], [4], [5])

$$\epsilon^2_{\text{int},r} = \frac{4\pi^2}{\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{\mathsf{F}'^{(r)}(\lambda)}\,\mathrm{d}\lambda}. \tag{30}$$

Moreover, conditional on

$$(\{H^{(r)}_\ell = h^{(r)}_\ell\}^{-1}_{\ell=-\infty}, \{H^{(r)}_\ell = h^{(r)}_\ell\}^{\infty}_{\ell=1})$$

the distribution of $H^{(r)}_0$ is Gaussian of mean

$$h^{(r)}_{\text{int}}(\{h^{(r)}_\ell\}^{-1}_{\ell=-\infty}, \{h^{(r)}_\ell\}^{\infty}_{\ell=1}) = \mathsf{E}\left[H^{(r)}_0 \,\middle|\, \{H^{(r)}_\ell = h^{(r)}_\ell\}^{-1}_{\ell=-\infty}, \{H^{(r)}_\ell = h^{(r)}_\ell\}^{\infty}_{\ell=1}\right]$$

and of variance $\epsilon^2_{\text{int},r}$. Unconditionally, $H^{(r)}_{\text{int}}$ is Gaussian of mean $\mu^{(r)}$ and of variance

$$\begin{aligned}\mathsf{Var}\left(H^{(r)}_{\text{int}}\right) &= \mathsf{Var}\left(H^{(r)}_0\right) - \epsilon^2_{\text{int},r}\\ &= \mathsf{F}^{(r)}(1/2) - \mathsf{F}^{(r)}(-1/2) - \epsilon^2_{\text{int},r}.\end{aligned}$$

Since we have assumed that the components of $\mathbf{H}_k$ are independent, we can use (23) and (24) to further bound the expressions in (27). We start with the upper bound:

$$\begin{aligned}\chi(\{\mathbf{H}_k\}) &\le \chi_{\text{IID}}\left(\mathbf{H}_0 \middle| \mathbf{H}^{-1}_{-\infty}, \mathbf{H}^{\infty}_1\right)\\ &\le \mathsf{E}\left[\log\left(1 + \sum_{r=1}^{n_{\text{R}}} \frac{\left|H^{(r)}_{\text{int}}\right|^2}{\epsilon^2_{\text{int},r}}\right)\right] + n_{\text{R}}\log n_{\text{R}} - n_{\text{R}} - \log\Gamma(n_{\text{R}})\\ &\le \log\left(1 + \sum_{r=1}^{n_{\text{R}}} \frac{\mathsf{E}\left[\left|H^{(r)}_{\text{int}}\right|^2\right]}{\epsilon^2_{\text{int},r}}\right) + n_{\text{R}}\log n_{\text{R}} - n_{\text{R}} - \log\Gamma(n_{\text{R}})\\ &= \log\left(1 + \sum_{r=1}^{n_{\text{R}}} \frac{\mathsf{Var}\left(H^{(r)}_0\right) - \epsilon^2_{\text{int},(r)} + |\mu^{(r)}|^2}{\epsilon^2_{\text{int},r}}\right) + n_{\text{R}}\log n_{\text{R}}\\ &\quad - n_{\text{R}} - \log\Gamma(n_{\text{R}}).\end{aligned} \tag{31}$$

Here the first inequality is due to (27); the second inequality follows from (24); and the third inequality follows from Jensen's inequality.

For the lower bound we get

$$
\begin{aligned}
\chi(\{\mathbf{H}_k\}) &\geq \chi_{\mathrm{IID}}\big(\mathbf{H}_0 \big| \mathbf{H}_{-\infty}^{-1}\big) \\
&\geq \mathsf{E}\left[\log \sum_{r=1}^{n_{\mathrm{R}}} \frac{\left|H_{\mathrm{pred}}^{(r)}\right|^2}{\epsilon_{\mathrm{pred},r}^2}\right] - \mathsf{E}\left[\mathrm{Ei}\left(-\sum_{r=1}^{n_{\mathrm{R}}} \frac{\left|H_{\mathrm{pred}}^{(r)}\right|^2}{\epsilon_{\mathrm{pred},r}^2}\right)\right] - 1 \\
&\geq \mathsf{E}\left[\log \sum_{r=1}^{n_{\mathrm{R}}} \frac{\left|H_{\mathrm{pred}}^{(r)}\right|^2}{\epsilon_{\mathrm{pred},r}^2}\right] - \mathrm{Ei}\left(-\sum_{r=1}^{n_{\mathrm{R}}} \frac{\mathsf{E}\left[\left|H_{\mathrm{pred}}^{(r)}\right|^2\right]}{\epsilon_{\mathrm{pred},r}^2}\right) - 1,
\end{aligned} \tag{32}
$$

where the first inequality is due to (27); the second inequality follows from (23); and where the last inequality follows from Jensen's inequality. The analytic computation of the RHS of (32) is greatly simplified if each component process $\{H_k^{(r)}\}$ of the vector-valued fading process $\{\mathbf{H}_k\}$ is of an identical law, which in our case means that

$$
\mathsf{F} = \mathrm{diag}\left(\tilde{\mathsf{F}}, \ldots, \tilde{\mathsf{F}}\right) \tag{33}
$$

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

$$
\boldsymbol{\mu} = \left(\mu^{(1)}, \ldots, \mu^{(n_{\mathrm{R}})}\right)^{\mathsf{T}} = (\tilde{\mu}, \ldots, \tilde{\mu})^{\mathsf{T}} \tag{34}
$$

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

$$
\begin{aligned}
\chi(\{\mathbf{H}_k\}) \geq{}& \log \frac{\mathsf{Var}\left(H_0^{(1)}\right) - \epsilon_{\mathrm{pred}}^2}{\epsilon_{\mathrm{pred}}^2} - 1 + g_{n_{\mathrm{R}}}\left(\frac{n_{\mathrm{R}}|\tilde{\mu}|^2}{\mathsf{Var}\left(H_0^{(1)}\right) - \epsilon_{\mathrm{pred}}^2}\right) \\
& - \mathrm{Ei}\left(-n_{\mathrm{R}} \frac{\mathsf{Var}\left(H_0^{(1)}\right) - \epsilon_{\mathrm{pred}}^2 + |\tilde{\mu}|^2}{\epsilon_{\mathrm{pred}}^2}\right),
\end{aligned} \tag{35}
$$

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

$$
g_m(z) = \log z - \mathrm{Ei}(-z) + \sum_{j=1}^{m-1} (-1)^j \left(e^{-z}(j-1)! - \frac{(m-1)!}{j(m-1-j)!}\right) z^{-j}, \qquad z \geq 0. \tag{36}
$$

(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 (33) and (34) are not necessary if one resorts to the weaker lower bound described in Remark 3.

Example 4. *Suppose that the fading process $\{\mathbf{H}_k\}$ is spatially IID 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 IID) law. Suppose that under this law $\{H_k^{(r)}\}$ is a stationary, unit-variance, zero-mean, circularly symmetric, m-th order auto-regressive AR(m) Gaussian process. That is, for all $1 \leq r \leq n_{\mathrm{R}}$,

$$
H_k^{(r)} = W_k^{(r)} - a_1 H_{k-1}^{(r)} - a_2 H_{k-2}^{(r)} - \cdots - a_m H_{k-m}^{(r)}. \tag{37}
$$

Here $\{W_k^{(r)}\}$ is temporally IID $\mathcal{N}_{\mathbb{C}}(0, \varepsilon^2)$, where ε^2 denotes the innovation variance; the coefficients $a_1, \ldots, a_m$ satisfy the stability condition [6]

$$\sum_{j=1}^{m} a_j z^j \neq -1 \qquad \forall\, |z| \leq 1; \tag{38}$$

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

$$\mathsf{Var}\left(H_k^{(r)}\right) = 1. \tag{39}$$

Then [6]

$$\epsilon^2_{\text{pred},r} = \varepsilon^2, \tag{40}$$

$$H^{(r)}_{\text{pred}} \sim \mathcal{N}_{\mathbb{C}}\left(0, 1 - \epsilon^2_{\text{pred},r}\right), \tag{41}$$

$$\epsilon^2_{\text{int},r} = \frac{\varepsilon^2}{1 + \sum_{j=1}^{m} |a_j|^2}, \tag{42}$$

$$H^{(r)}_{\text{int}} \sim \mathcal{N}_{\mathbb{C}}\left(0, 1 - \epsilon^2_{\text{int},r}\right), \tag{43}$$

which yields

$$\chi(\{\mathbf{H}_k\}) \geq \log \frac{1 - \varepsilon^2}{\varepsilon^2} + \psi(n_{\mathrm{R}}) - \mathrm{Ei}\left(-n_{\mathrm{R}} \frac{1 - \varepsilon^2}{\varepsilon^2}\right) - 1, \tag{44}$$

$$\chi(\{\mathbf{H}_k\}) \leq \log\left(1 + n_{\mathrm{R}} \frac{1 + \sum_{j=1}^{m} |a_j|^2 - \varepsilon^2}{\varepsilon^2}\right) + n_{\mathrm{R}} \log n_{\mathrm{R}} - n_{\mathrm{R}} - \log \Gamma(n_{\mathrm{R}}), \tag{45}$$

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 (44) is unchanged and the upper bound becomes

$$\chi(\{\mathbf{H}_k\}) \leq \log\left(1 + 2n_{\mathrm{R}} \frac{1 - \varepsilon^2}{\varepsilon^2}\right) + n_{\mathrm{R}} \log n_{\mathrm{R}} - n_{\mathrm{R}} - \log \Gamma(n_{\mathrm{R}}). \tag{46}$$

For $\varepsilon^2 \ll 1$ one obtains the asymptotic bounds

$$\chi(\{\mathbf{H}_k\}) \geq \log \frac{1}{\varepsilon^2} + \psi(n_{\mathrm{R}}) - 1 + o(\varepsilon^2), \tag{47}$$

$$\chi(\{\mathbf{H}_k\}) \leq \log \frac{1}{\varepsilon^2} + \log 2 + (n_{\mathrm{R}} + 1) \log n_{\mathrm{R}} - n_{\mathrm{R}} - \log \Gamma(n_{\mathrm{R}}) + o(\varepsilon^2), \tag{48}$$

where $o(\varepsilon^2)$ tends to zero as ε^2 tends to zero. Further specializing to the case of two receive antennas ($n_{\mathrm{R}} = 2$) we obtain the bounds that are depicted in Figure 1.

6 Proof of Theorem 1

6.1 Proof Outline

The proof of Theorem 1 has three components. The first is an achievability result ("direct part") (Section 6.2) which provides a lower bound on channel capacity and hence a lower

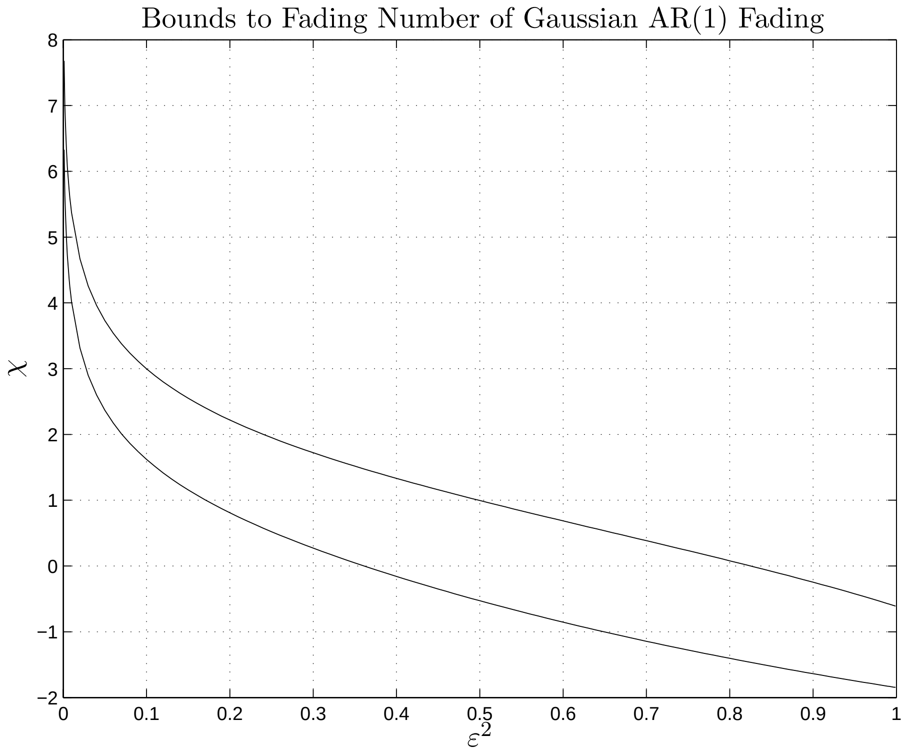

Figure 1: Upper and lower bound of a zero-mean SIMO Gaussian-Markov fading channel with memory 1 (AR(1)) and two receive antennas plotted in function of the prediction error ε^2. Both components of the fading vector $\mathbf{H}_k$ are assumed to be independent and identically distributed with variance 1.

bound on the fading number. This lower bound is the RHS of (25). The second component is a "converse" (Section 6.3), which provides for an upper bound on channel capacity, and hence for an upper bound on the fading number. This upper bound is the RHS of (26). Finally, the last component (Appendix A) is a demonstration that the lower and upper bounds are in fact identical.

The inputs that are used to demonstrate the achievability of RHS of (25) 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.

6.2 The Direct Part

6.2.1 An Overview

The lower bound is based on choosing the input symbols to be IID, circularly symmetric, with

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

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

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

The motivation for using IID 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

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

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 $\mathbf{Y}_k$. Indeed, in this regime, an excellent estimator for $\mathbf{H}_k$ is the estimator $\mathbf{Y}_k/X_k$. The other "identification" that this input distribution allows has to do with inference on $\mathbf{H}_k$ based on the channel output $\mathbf{Y}_k$ 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 $\mathbf{H}_k$, namely $\mathbf{H}_k/\|\mathbf{H}_k\|$, to within a multiple by a scalar complex random variable of unit magnitude and uniform phase. For this identification the estimator $\mathbf{Y}_k/\|\mathbf{Y}_k\|$ is most suitable. Indeed, while the circular symmetry of the input X_k renders the phase information in $\mathbf{Y}_k$ 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 $\mathbf{H}_k$ is—to within a random phase—almost identical to the direction of $\mathbf{Y}_k$.

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

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

We now analyze the individual terms in the sum:

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

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 $\mathbf{H}_1^{k-1}$ from the past inputs & outputs $(X_1^{k-1}, \mathbf{Y}_1^{k-1})$, and on estimating the future "directions" $\{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=k+1}^n$ based on the available future outputs $\mathbf{Y}_{k+1}^n$ (without their corresponding inputs). Note that if we were to ignore these future outputs we would not attain the fading number.

6.2.2 Proof of the Lower Bound

In this section we derive a lower bound to capacity and use it to show that the RHS of (25) is a lower bound to the fading number. Let $\{X_k\}$ be IID circularly symmetric random variables with

$$\log |X_k|^2 \sim \text{Uniform on } [\log x_{\min}^2, \log \mathcal{E}_{\mathrm{s}}], \tag{49}$$

where

$$x_{\min}^2 = \log \mathcal{E}_{\mathrm{s}}. \tag{50}$$

Fix some (large) positive integer κ and use the chain rule and the non-negativity of mutual information to obtain:

$$\frac{1}{n} I\big(X_1^n; \mathbf{Y}_1^n\big) = \frac{1}{n} \sum_{k=1}^{n} I\big(X_k; \mathbf{Y}_1^n \big| X_1^{k-1}\big) \tag{51}$$

$$\geq \frac{1}{n} \sum_{k=\kappa+1}^{n-\kappa} I\big(X_k; \mathbf{Y}_1^n \big| X_1^{k-1}\big). \tag{52}$$

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

$$\begin{aligned}
& I\big(X_k; \mathbf{Y}_1^n \big| X_1^{k-1}\big) \\
&\quad = I\big(X_k; X_1^{k-1}, \mathbf{Y}_1^n\big) && (53)\\
&\quad \geq I\big(X_k; X_{k-\kappa}^{k-1}, \mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k+1}^{k+\kappa}\big) && (54)\\
&\quad = 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)} && (55)\\
&\quad \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) && (56)\\
&\quad = 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) && (57)\\
&\quad = 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) && (58)\\
&\quad \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) && (59)\\
&\quad = 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) && (60)\\
&\quad = 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) \\
&\qquad - \epsilon_2(x_{\min}, \kappa) && (61)\\
&\quad = 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) && (62)\\
&\quad \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) && (63)\\
&\quad = 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) && (64)\\
&\quad = 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\qquad \kappa + 1 \leq k \leq n - \kappa. && (65)
\end{aligned}$$

Here the first equality follows because $\{X_k\}$ is chosen to be IID; in the subsequent inequality we have dropped some arguments which reduces the mutual information; next we have used the chain rule; in (56) we lower bound the second term by $-\epsilon_1(x_{\min}, \kappa)$ that—as shown in Appendix B—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 (55)–(57), where again it is shown in Appendix B that $\epsilon_2(x_{\min}, \kappa)$ depends only on $x_{\min}$ and κ and tends to zero as $x_{\min} \uparrow \infty$; in (62) $\hat{\mathbf{H}}$ denotes $\mathbf{H}/\|\mathbf{H}\|$; and the equality before last follows from stationarity.

From (65) and (52) we obtain

$$\begin{aligned}
& \frac{1}{n} I\big(X_1^n; \mathbf{Y}_1^n\big) \\
&\quad \geq \left(1 - \frac{2\kappa}{n}\right) \Big(I\big(X_{\kappa+1}; \mathbf{Y}_{\kappa+1} \big| \mathbf{H}_1^\kappa, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\kappa+2}^{2\kappa+1}\big) - \epsilon_1(x_{\min}, \kappa) - \epsilon_2(x_{\min}, \kappa)\Big). && (66)
\end{aligned}$$

Letting n tend to infinity we obtain

$$C(\mathrm{SNR}) \geq I\big(X_{\kappa+1}; \mathbf{Y}_{\kappa+1} \big| \mathbf{H}_1^\kappa, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\kappa+2}^{2\kappa+1}\big) - \epsilon_1(x_{\min}, \kappa) - \epsilon_2(x_{\min}, \kappa), \tag{67}$$

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

We next let the power grow to infinity $\mathcal{E}_\mathrm{s} \uparrow \infty$. Since the circular-symmetric law (49) achieves the fading number of IID SIMO fading with side-information [1, Note 4.31] and since our choice (50) guarantees that $\epsilon_1(x_{\min}, \kappa)$ and $\epsilon_2(x_{\min}, \kappa)$ tend to zero as $\mathcal{E}_\mathrm{s} \uparrow \infty$ (see Appendix B) we obtain the bound

$$\chi(\{\mathbf{H}_k\}) \geq \chi_{\mathrm{IID}}\bigl(\mathbf{H}_{\kappa+1} \big| \mathbf{H}_1^{\kappa}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\kappa+2}^{2\kappa+1}\bigr). \tag{68}$$

Upon letting κ in the above tend to infinity we obtain the desired result, *i.e.*, that the RHS of (25) is a lower bound to $\chi(\{\mathbf{H}_k\})$.

6.3 The Converse

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

6.3.1 An Overview

To upper bound capacity we use the chain rule

$$\frac{1}{n} I(X_1^n; \mathbf{Y}_1^n) = \frac{1}{n} \sum_{k=1}^{n} I(X_1^n; \mathbf{Y}_k | \mathbf{Y}_1^{k-1}), \tag{69}$$

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

$$\begin{aligned}
I(X_1^n; \mathbf{Y}_k | \mathbf{Y}_1^{k-1}) &= I(X_1^n, \mathbf{Y}_1^{k-1}; \mathbf{Y}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\
&= I(X_1^{k-1}, \mathbf{Y}_1^{k-1}, X_k; \mathbf{Y}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\
&\leq I(X_1^{k-1}, \mathbf{Y}_1^{k-1}, \mathbf{H}_1^{k-1}, X_k; \mathbf{Y}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\
&= I(\mathbf{H}_1^{k-1}, X_k; \mathbf{Y}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\
&= I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_1^{k-1}; \mathbf{Y}_k | X_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\
&= I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_1^{k-1}; X_k, \mathbf{Y}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\
&\leq I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_1^{k-1}; \mathbf{H}_k, X_k, \mathbf{Y}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\
&= I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_1^{k-1}; \mathbf{H}_k) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \\
&\leq I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}).
\end{aligned} \tag{70}$$

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 & 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 & 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 (70) 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 non-tight 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 $I(X_k; \mathbf{Y}_k)$, *i.e.*, to achieve the IID capacity. (The resulting average mutual information $n^{-1}I(X_1^n;\mathbf{Y}_1^n)$ would, of course, be very low but the single term $I(X_1^n;\mathbf{Y}_k|\mathbf{Y}_1^{k-1})$ would achieve this non-tight 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 (69) 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 in Section 6.3.2. (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 blocklength 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 $I(X_k;\mathbf{Y}_k)$ when X_k is distributed according to Q. Thus, for such inputs

$$C \le \varliminf_{n\to\infty}\frac{1}{n}I(X_1^n;\mathbf{Y}_1^n)+\epsilon. \tag{71}$$

By (69) and (70) we also have for such inputs

$$\begin{aligned}\frac{1}{n}I(X_1^n;\mathbf{Y}_1^n) &\le \frac{1}{n}\sum_{k=1}^{n}\Big(I(X_k;\mathbf{Y}_k)-I(\mathbf{Y}_k;\mathbf{Y}_1^{k-1})\Big)+I(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1})\\ &\approx I(Q)-\frac{1}{n}\sum_{k=1}^{n}I(\mathbf{Y}_k;\mathbf{Y}_1^{k-1})+I(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1})\end{aligned} \tag{72}$$

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 $\{X_k\}$

$$\varlimsup_{n\to\infty}\frac{1}{n}I(X_1^n;\mathbf{Y}_1^n)\le I(Q)-\varliminf_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}I(\mathbf{Y}_k;\mathbf{Y}_1^{k-1})+I(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}). \tag{73}$$

The choice of Q (the distribution of X_k) affects the RHS of (72) and (73) in two different ways. It determines $I(Q)$ but it also influences the terms $I(\mathbf{Y}_k;\mathbf{Y}_1^{k-1})$. There is thus a tension between choosing Q to maximize $I(Q)$ (*i.e.*, to make $I(Q)$ close to the IID channel capacity) and choosing Q to minimize the $I(\mathbf{Y}_k;\mathbf{Y}_1^{k-1})$ 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 (72) and (73) 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 (71) we must have

$$\lim_{\mathrm{SNR}\to\infty}\frac{I(Q)}{\log\log\mathrm{SNR}}=1. \tag{74}$$

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 (73) and the non-negativity of mutual information we obtain that for the marginal-Q input $\{X_k\}$ to satisfy (71) the marginal law Q must satisfy

$$I(Q) \geq C - \epsilon - I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}), \tag{75}$$

which combines with

$$C \geq \log\log \mathrm{SNR} + O(1) \tag{76}$$

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

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

Here the lower bound (76) follows, for example, from [1, Theorem 4.41], *e.g.*, by considering the use of only one transmit antenna and only one receive antenna.

The next step in the proof of the converse is to show that (74) implies that Q "escapes to infinity", *i.e.*, that

$$\lim_{\mathrm{SNR}\to\infty} Q(\{x : |x| \geq \xi_{\min}\}) = 1, \quad \text{for any fixed } \xi_{\min}. \tag{77}$$

This is proven in greater generality for general cost-constrained channels in Section 6.3.3, 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

$$I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \approx I(\hat{\mathbf{H}}_k \hat{U}_k; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{k-1}), \tag{78}$$

where $\{\hat{U}_j\}$ are IID uniformly distributed on the complex unit-circle independently of the fading process (as described in (1)).

The intuition behind (78) 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 $\mathbf{Y}_1^{k-1}$ to learn the past "direction" (corrupted by random rotations) $\{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{k-1}$. Similarly, there would be an almost deterministic relationship between the present output $\mathbf{Y}_k$ and the present fading "direction" (again corrupted by a random phase) $\hat{\mathbf{H}}_k \hat{U}_k$.

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.

6.3.2 Stationarity Considerations

Lemma 5. *Fix some power $\mathcal{E}_\mathrm{s}$ with corresponding* SNR *of $\mathcal{E}_\mathrm{s}/\sigma^2$. Let $C(\mathrm{SNR})$ denote the corresponding channel capacity under an average power $\mathcal{E}_\mathrm{s}$ constraint. Then for any $\epsilon > 0$ there corresponds some positive integer $\eta = \eta(\mathrm{SNR}, \epsilon)$ and some distribution $Q = Q(\mathrm{SNR}, \epsilon)$ on $\mathbb{C}$ such that for any blocklength 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*

$$\frac{1}{n} I\big(X_1^n; \mathbf{Y}_1^n\big) \geq C(\mathrm{SNR}) - \epsilon; \tag{79}$$

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*

$$X_\eta, X_{\eta+1}, \ldots, X_{n-2\eta+2} \tag{80}$$

all have the same distribution Q*;*

3. *this marginal distribution* Q *gives rise to a second moment* $\mathcal{E}_\mathrm{s}$

$$\mathsf{E}\left[|X_\ell|^2\right] = \mathcal{E}_\mathrm{s}, \qquad \ell = \eta, \ldots, n - 2\eta + 2; \tag{81}$$

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

$$\mathsf{E}\left[|X_\ell|^2\right] \leq \mathcal{E}_\mathrm{s}, \qquad \ell \in \{1, \ldots, \eta - 1\} \cup \{n - 2\eta + 3, \ldots, n\}. \tag{82}$$

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

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

where the supremum is over all joint distributions on the random variables $X_1, \ldots, X_n$ under which $\sum_{\ell=1}^n \mathsf{E}\left[|X_\ell|^2\right] = n\mathcal{E}_\mathrm{s}$, 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

$$\frac{1}{\eta}\sum_{\ell=1}^{\eta} \mathsf{E}\left[|X_\ell|^2\right] = \mathcal{E}_\mathrm{s} \quad \text{and} \quad \frac{1}{\eta} I(X_1, \ldots, X_\eta; \mathbf{Y}_1, \ldots, \mathbf{Y}_\eta) > C(\mathrm{SNR}) - \frac{\epsilon}{2}. \tag{83}$$

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 $\mathcal{B} \subset \mathbb{C}$

$$Q(\mathcal{B}) = \frac{1}{\eta}\sum_{\ell=1}^{\eta} P^*(X_\ell \in \mathcal{B}). \tag{84}$$

By (83) we have

$$\int_{\mathbb{C}} |x|^2 \,\mathrm{d}Q(x) = \mathcal{E}_\mathrm{s}. \tag{85}$$

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 $\tilde{\mathbf{X}}$ be defined by

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

so that

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

Here

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

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

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

where

$$T \sim \text{Uniform on } \{0, \dots, \eta - 1\}$$

is independent of $\tilde{\mathbf{X}}$, 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 $\nu\eta(C(\mathrm{SNR}) - \epsilon/2)$, so that the normalized mutual information satisfies

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

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

Except at the edges, the above mixture guarantees that all marginals are Q, and hence by (85) of average power $\mathcal{E}_\mathrm{s}$. The power in the edges can be smaller than $\mathcal{E}_\mathrm{s}$ because of the mixture with deterministic zeros. □

6.3.3 Input Distributions that Escape to Infinity Revisited

In this section 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 $\log\log\mathrm{SNR}$ 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 non-negative cost function $g : \mathcal{X} \to \mathbb{R}^+$, we define the capacity-cost function $C : \mathbb{R}^+ \to \mathbb{R}^+$ by*

$$C(\Upsilon) = \sup_{Q:\mathsf{E}_Q[g(X)]\leq\Upsilon} I(Q, W), \qquad \Upsilon \geq 0. \tag{86}$$

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

$$\mathsf{E}_{Q_\Upsilon}[g(X)] \leq \Upsilon, \qquad \Upsilon \geq 0. \tag{87}$$

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

$$\lim_{\Upsilon\uparrow\infty} Q_\Upsilon\big(\{x \in \mathcal{X} : g(x) < \Upsilon_0\}\big) = 0. \tag{88}$$

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 $C_{\text{asy}}(\Upsilon)$ be a function of the cost that captures the asymptotic behavior of the capacity-cost function $C(\Upsilon)$ in the sense that*

$$\lim_{\Upsilon\uparrow\infty} \frac{C(\Upsilon)}{C_{\text{asy}}(\Upsilon)} = 1. \tag{89}$$

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

$$\varliminf_{\Upsilon\uparrow\infty} \left\{ \sup_{\alpha\in(0,\alpha_0]} \frac{\alpha C_{\text{asy}}\left(\frac{\Upsilon}{\alpha}\right)}{C_{\text{asy}}(\Upsilon)} \right\} < 1, \qquad \forall\, 0 < \alpha_0 < 1. \tag{90}$$

Let $\{Q_\Upsilon, \Upsilon \geq 0\}$ be a family of input distributions satisfying (87) and

$$\lim_{\Upsilon\uparrow\infty} \frac{I(Q_\Upsilon, W)}{C_{\text{asy}}(\Upsilon)} = 1. \tag{91}$$

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

Remark 9. *The growth condition (90) is related to the notion of "slowly varying in the Karamata sense", see [8, Sec. 1.2]. Examples of functions $C_{\text{asy}}(\Upsilon)$ that satisfy (90) include:*

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

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

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

$$E = \begin{cases} 1 & \text{if } g(X) \geq \Upsilon_0 \\ 0 & \text{otherwise} \end{cases}, \tag{92}$$

$$\alpha = \Pr[E = 1]\,. \tag{93}$$

Since $C(\cdot)$ is monotonically increasing and unbounded, it follows by (89) that

$$\lim_{\Upsilon\to\infty} C_{\text{asy}}(\Upsilon) = \infty \tag{94}$$

which combines with (91) to imply that

$$\lim_{\Upsilon\to\infty} I(Q_\Upsilon, W) = \infty. \tag{95}$$

By (95) 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

$$\begin{aligned} I(X;Y) &= I(X,E;Y) \\ &= I(E;Y) + I(X;Y|E) \\ &= I(E;Y) + I(X;Y|E{=}0)\Pr[E=0] + I(X;Y|E{=}1)\Pr[E=1] \\ &\leq \log 2 + I(X;Y|E{=}0) + \alpha I(X;Y|E{=}1) \\ &\leq \log 2 + C(\Upsilon_0) + \alpha I(X;Y|E{=}1) \\ &\leq \log 2 + C(\Upsilon_0) + \alpha C\left(\frac{\Upsilon}{\alpha}\right). \end{aligned} \tag{96}$$

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

$$\mathsf{E}[g(X) \mid E=1] \leq \frac{\mathsf{E}[g(X)]}{\alpha} \leq \frac{\Upsilon}{\alpha}. \tag{97}$$

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

$$\alpha_n < \alpha_0, \qquad n \text{ sufficiently large.} \tag{98}$$

From (96) we now have

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

Here the limiting behavior of the LHS follows from (91); the limiting value of $(\log 2 + C(\Upsilon_0))/C_{\text{asy}}(\Upsilon_n)$ follows by (94) 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 (89) because $\Upsilon_n \uparrow \infty$ implies $\Upsilon_n/\alpha_n \uparrow \infty$. Upon letting n tend to infinity we obtain the contradiction

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

Here the second inequality follows from (98) and the last inequality follows from (90). □

6.3.4 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 (79)–(82) of Lemma 5 so that, in particular,

$$C \leq \frac{1}{n} I\big(X_1^n; \mathbf{Y}_1^n\big) + \epsilon \tag{99}$$

$$= \frac{1}{n}\sum_{k=1}^{n} I\big(X_1^n; \mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}\big) + \epsilon. \tag{100}$$

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

$$I\big(X_1^n; \mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}\big) \le I\big(X_k; \mathbf{Y}_k\big) + I\big(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}\big) \tag{101}$$

$$\le C_{\mathrm{IID}}\,(\mathrm{SNR}) + I\big(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}\big), \tag{102}$$

which is uniformly bounded in n. Here the first inequality follows from (70) and the second from (81) and (82). We conclude that

$$C \le \varliminf_{n\to\infty} \frac{1}{n} I\big(X_1^n; \mathbf{Y}_1^n\big) + \epsilon \tag{103}$$

$$= \varliminf_{n\to\infty} \frac{1}{n - \kappa - 3(\eta-1)} \sum_{k=\eta+\kappa}^{n-2\eta+2} I\big(X_1^n; \mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}\big) + \epsilon. \tag{104}$$

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 $I\big(X_1^n; \mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}\big)$ that does not depend on k will result in an upper bound on C via (104).

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

$$I\big(X_1^n; \mathbf{Y}_k \big| \mathbf{Y}_1^{k-1}\big) \le I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) - I(\mathbf{Y}_k; \mathbf{Y}_1^{k-1}) \tag{105}$$

$$\le I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) - I(\mathbf{Y}_k; \mathbf{Y}_{k-\kappa}^{k-1}) \tag{106}$$

$$= I(Q) + I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) - I(\mathbf{Y}_k; \mathbf{Y}_{k-\kappa}^{k-1}), \tag{107}$$

where we use $I(Q)$ to denote the mutual information $I(X_k; \mathbf{Y}_k)$ when X_k is distributed according to the law Q. From (107) and (104) we conclude that

$$C \le I(Q) + I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) + \epsilon - \varliminf_{n\to\infty} \min_{\eta+\kappa \le k \le n-2(\eta-1)} I\big(\mathbf{Y}_k; \mathbf{Y}_{k-\kappa}^{k-1}\big). \tag{108}$$

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

$$E_k = \begin{cases} 1 & \text{if } |X_j| \ge \xi_{\min} \quad \forall\, k - \kappa \le j \le k, \\ 0 & \text{otherwise.} \end{cases} \tag{109}$$

Let

$$\alpha_k = \Pr[E_k = 1]\,. \tag{110}$$

By the union of events bound

$$\alpha_k \ge 1 - \sum_{j=k-\kappa}^{k} \Pr[|X_j| < \xi_{\min}] \tag{111}$$

$$= 1 - (\kappa+1) Q(|X| < \xi_{\min}), \tag{112}$$

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,

$$\alpha_k \ge \alpha, \tag{113}$$

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

$$\alpha = \max\big\{0, 1 - (\kappa+1) Q(|X| < \xi_{\min})\big\}. \tag{114}$$

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

$$\begin{aligned}
I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big) &= I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1},E_k\big)-I\big(\mathbf{Y}_k;E_k\big|\mathbf{Y}_{k-\kappa}^{k-1}\big) && (115)\\
&= I\big(\mathbf{Y}_k;E_k\big)+I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k\big)-I\big(\mathbf{Y}_k;E_k\big|\mathbf{Y}_{k-\kappa}^{k-1}\big) && (116)\\
&\ge I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k\big)-I\big(\mathbf{Y}_k;E_k\big|\mathbf{Y}_{k-\kappa}^{k-1}\big) && (117)\\
&= I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k\big)-H\big(E_k\big|\mathbf{Y}_{k-\kappa}^{k-1}\big)+H\big(E_k\big|\mathbf{Y}_{k-\kappa}^{k}\big) && (118)\\
&\ge I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k\big)-H\big(E_k\big|\mathbf{Y}_{k-\kappa}^{k-1}\big) && (119)\\
&\ge I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k\big)-H(E_k) && (120)\\
&= I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k\big)-H_{\mathrm{b}}(\alpha_k) && (121)\\
&\ge I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k\big)-H_{\mathrm{b}}\left(\max\left\{\alpha_k,\frac{1}{2}\right\}\right) && (122)\\
&\ge \alpha_k I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k{=}1\big)-H_{\mathrm{b}}\left(\max\left\{\alpha_k,\frac{1}{2}\right\}\right) && (123)\\
&\ge \alpha I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k{=}1\big)-H_{\mathrm{b}}\left(\max\left\{\alpha,\frac{1}{2}\right\}\right), && (124)
\end{aligned}$$

where

$$H_{\mathrm{b}}(\xi)\triangleq-\xi\log\xi-(1-\xi)\log(1-\xi) \tag{125}$$

is the binary entropy function. Note that $H_{\mathrm{b}}(\xi)\le H_{\mathrm{b}}\left(\max\left\{\xi,\frac{1}{2}\right\}\right)$ and that $H_{\mathrm{b}}\left(\max\left\{\xi,\frac{1}{2}\right\}\right)$ is monotonically non-increasing so that the last inequality follows from (113).

Inequalities (124) and (108) combine to yield

$$\begin{aligned}
C\le I(Q)+I(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1})+\epsilon+H_{\mathrm{b}}\left(\max\left\{\alpha,\frac{1}{2}\right\}\right)&\\
-\alpha\varliminf_{n\to\infty}\min_{\eta+\kappa\le k\le n-2(\eta-1)}I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k=1\big)& && (126)
\end{aligned}$$

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

$$\begin{aligned}
I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k=1\big) &= I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1},\mathbf{Z}_{k-\kappa}^{k-1}\big|E_k=1\big)-I\big(\mathbf{Y}_k;\mathbf{Z}_{k-\kappa}^{k-1}\big|\mathbf{Y}_{k-\kappa}^{k-1},E_k=1\big) && (127)\\
&= I\big(\mathbf{Y}_k;\{\mathbf{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big)-I\big(\mathbf{Y}_k;\mathbf{Z}_{k-\kappa}^{k-1}\big|\mathbf{Y}_{k-\kappa}^{k-1},E_k=1\big) && (128)\\
&\ge I\big(\mathbf{Y}_k;\{\mathbf{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big)-\delta_1(\xi_{\min},\kappa), && (129)
\end{aligned}$$

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

$$I\big(\mathbf{Y}_k;\mathbf{Z}_{k-\kappa}^{k-1}\big|\mathbf{Y}_{k-\kappa}^{k-1},E_k=1\big)\le\delta_1(\xi_{\min},\kappa) \tag{130}$$

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

$$\lim_{\xi_{\min}\uparrow\infty}\delta_1(\xi_{\min},\kappa)=0. \tag{131}$$

Continuing with the chain of inequalities we have

$$\begin{aligned}
&I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k=1\big)\\
&\quad\ge I\big(\mathbf{Y}_k;\{\mathbf{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big)-\delta_1(\xi_{\min},\kappa) && (132)\\
&\quad= I\big(\mathbf{Y}_k,\mathbf{Z}_k;\{\mathbf{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big)-I\big(\mathbf{Z}_k;\{\mathbf{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|\mathbf{Y}_k,E_k=1\big)\\
&\qquad-\delta_1(\xi_{\min},\kappa) && (133)\\
&\quad= I\big(\mathbf{H}_k X_k;\{\mathbf{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big)-I\big(\mathbf{Z}_k;\{\mathbf{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|\mathbf{Y}_k,E_k=1\big)\\
&\qquad-\delta_1(\xi_{\min},\kappa) && (134)\\
&\quad\ge I\big(\mathbf{H}_k X_k;\{\mathbf{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}\big|E_k=1\big)-\delta_1(\xi_{\min},\kappa)-\delta_2(\xi_{\min},\kappa), && (135)
\end{aligned}$$

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

$$I\big(\mathbf{Z}_k; \{\mathbf{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1} \big| \mathbf{Y}_k, E_k = 1\big) \le \delta_2(\xi_{\min}, \kappa) \tag{136}$$

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

$$\lim_{\xi_{\min} \uparrow \infty} \delta_2(\xi_{\min}, \kappa) = 0. \tag{137}$$

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

$$\begin{aligned}
& I\big(\mathbf{Y}_k; \mathbf{Y}_{k-\kappa}^{k-1} \big| E_k = 1\big) \\
& \quad \ge I\big(\mathbf{H}_k X_k; \{\mathbf{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1} \big| E_k = 1\big) - \delta_1(\xi_{\min}, \kappa) - \delta_2(\xi_{\min}, \kappa) && (138) \\
& \quad \ge I\big(\mathbf{H}_k X_k \hat{V}_k; \{\mathbf{H}_\ell X_\ell \hat{V}_\ell\}_{\ell=k-\kappa}^{k-1} \big| E_k = 1\big) - \delta_1(\xi_{\min}, \kappa) - \delta_2(\xi_{\min}, \kappa) && (139) \\
& \quad \ge I\left(\frac{\mathbf{H}_k X_k \hat{V}_k}{\|\mathbf{H}_k X_k\|}; \left\{ \frac{\mathbf{H}_\ell X_\ell \hat{V}_\ell}{\|\mathbf{H}_\ell X_\ell\|} \right\}_{\ell=k-\kappa}^{k-1} \middle| E_k = 1 \right) - \delta_1(\xi_{\min}, \kappa) - \delta_2(\xi_{\min}, \kappa) && (140) \\
& \quad = I\left(\frac{\mathbf{H}_k}{\|\mathbf{H}_k\|} \hat{U}_k; \left\{ \frac{\mathbf{H}_\ell}{\|\mathbf{H}_\ell\|} \hat{U}_\ell \right\}_{\ell=k-\kappa}^{k-1} \right) - \delta_1(\xi_{\min}, \kappa) - \delta_2(\xi_{\min}, \kappa) && (141) \\
& \quad = I\big(\hat{\mathbf{H}}_k \hat{U}_k; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=k-\kappa}^{k-1}\big) - \delta_1(\xi_{\min}, \kappa) - \delta_2(\xi_{\min}, \kappa) && (142) \\
& \quad = I\big(\hat{\mathbf{H}}_{\kappa+1} \hat{U}_{\kappa+1}; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\kappa}\big) - \delta_1(\xi_{\min}, \kappa) - \delta_2(\xi_{\min}, \kappa). && (143)
\end{aligned}$$

Here the first two inequalities follow from the data processing inequality; (141) 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 (143) and (126) we now have

$$\begin{aligned}
C \le\ & I(Q) + I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) + \epsilon + H_{\mathrm{b}}\left(\max\left\{ \alpha, \frac{1}{2} \right\} \right) \\
& - \alpha I\big(\hat{\mathbf{H}}_{\kappa+1} \hat{U}_{\kappa+1}; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\kappa}\big) - \delta_1(\xi_{\min}, \kappa) - \delta_2(\xi_{\min}, \kappa).
\end{aligned} \tag{144}$$

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 (144) as the SNR tends to infinity. We shall begin by showing that

$$\lim_{\mathrm{SNR} \uparrow \infty} \alpha = 1. \tag{145}$$

To this end it suffices, by (114), to show that

$$\lim_{\mathrm{SNR} \uparrow \infty} Q(|X| \ge \xi_{\min}) = 1. \tag{146}$$

But this follows from Theorem 8 because by (7) and (76)

$$\lim_{\mathrm{SNR} \uparrow \infty} \frac{C(\mathrm{SNR})}{\log \log \mathrm{SNR}} = 1 \tag{147}$$

and because by (144) and the trivial bound

$$I(Q) \le C_{\mathrm{IID}}(\mathrm{SNR}) \le C(\mathrm{SNR}) \tag{148}$$

it follows that

$$\lim_{\mathrm{SNR}\uparrow\infty} \frac{I(Q)}{C(\mathrm{SNR})} = 1. \tag{149}$$

Applying the bound $I(Q) \le C_{\mathrm{IID}}(\mathrm{SNR})$ to (144) and using (145) we obtain

$$\begin{aligned}&\varlimsup_{\mathrm{SNR}\uparrow\infty} \{C(\mathrm{SNR}) - C_{\mathrm{IID}}(\mathrm{SNR})\} \\ &\qquad \le I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) + \epsilon - I\big(\hat{\mathbf{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\kappa}\big) - \delta_1(\xi_{\min}, \kappa) - \delta_2(\xi_{\min}, \kappa).\end{aligned} \tag{150}$$

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 (131) & (137) that

$$\varlimsup_{\mathrm{SNR}\uparrow\infty} \{C(\mathrm{SNR}) - C_{\mathrm{IID}}(\mathrm{SNR})\} \le I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) - I\big(\hat{\mathbf{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\kappa}\big). \tag{151}$$

Upon now letting κ tend to infinity we obtain

$$\chi(\{\mathbf{H}_k\}) \le \chi_{\mathrm{IID}}(\mathbf{H}_0) + I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) - I\big(\hat{\mathbf{H}}_0 \hat{U}_0; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=-\infty}^{-1}\big) \tag{152}$$

as we had set out to prove.

A Equivalence of (25) and (26)

The equivalence of (25) and (26) can be proved as follows. Using stationarity, (13), and (11) we get

$$\begin{aligned}&\chi_{\mathrm{IID}}\big(\mathbf{H}_{\kappa+1} \big| \mathbf{H}_1^{\kappa}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) \\ &\quad = h_\lambda\big(\hat{\mathbf{H}}_{\kappa+1}\hat{U}_{\kappa+1} \big| \mathbf{H}_1^{\kappa}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) - h\big(\mathbf{H}_{\kappa+1} \big| \mathbf{H}_1^{\kappa}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) \\ &\qquad + n_{\mathrm{R}} \mathsf{E}\big[\log \|\mathbf{H}_{\kappa+1}\|^2\big] - \log 2 &(153)\\ &\quad = \chi_{\mathrm{IID}}(\mathbf{H}_{\kappa+1}) - I\big(\hat{\mathbf{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \mathbf{H}_1^{\kappa}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) \\ &\qquad + I\big(\mathbf{H}_{\kappa+1}; \mathbf{H}_1^{\kappa}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) &(154)\\ &\quad = \chi_{\mathrm{IID}}(\mathbf{H}_0) + I(\mathbf{H}_{\kappa+1}; \mathbf{H}_1^{\kappa}) - I\big(\hat{\mathbf{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\kappa}\big) + \delta_\kappa, &(155)\end{aligned}$$

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

$$\begin{aligned}\delta_\kappa &= I\big(\mathbf{H}_{\kappa+1}; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1} \big| \mathbf{H}_1^{\kappa}\big) + I\big(\hat{\mathbf{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\kappa}\big) \\ &\quad - I\big(\hat{\mathbf{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \mathbf{H}_1^{\kappa}, \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) &(156)\\ &= I\big(\mathbf{H}_{\kappa+1}; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1} \big| \mathbf{H}_1^{\kappa}\big) \\ &\quad + \underbrace{I\big(\hat{\mathbf{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=1}^{\kappa}\big) - I\big(\hat{\mathbf{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big)}_{=0 \text{ by stationarity}} \\ &\quad - I\big(\hat{\mathbf{H}}_{\kappa+1}\hat{U}_{\kappa+1}; \mathbf{H}_1^{\kappa} \big| \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) &(157)\\ &= h_\lambda\big(\{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1} \big| \mathbf{H}_1^{\kappa}\big) - h_\lambda\big(\{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1} \big| \mathbf{H}_1^{\kappa+1}\big) - h\big(\mathbf{H}_1^{\kappa} \big| \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) \\ &\quad + h\big(\mathbf{H}_1^{\kappa} \big| \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+1}^{2\kappa+1}\big) &(158)\\ &= h_\lambda\big(\{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) - h_\lambda\big(\{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1} \big| \mathbf{H}_1^{\kappa+1}\big) - h\big(\mathbf{H}_1^{\kappa}\big) + h\big(\mathbf{H}_1^{\kappa} \big| \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+1}^{2\kappa+1}\big) &(159)\\ &= I\big(\mathbf{H}_1^{\kappa+1}; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+2}^{2\kappa+1}\big) - I\big(\mathbf{H}_1^{\kappa}; \{\hat{\mathbf{H}}_\ell \hat{U}_\ell\}_{\ell=\kappa+1}^{2\kappa+1}\big) &(160)\\ &\to 0 \qquad \text{for } \kappa \uparrow \infty, \text{ by stationarity.} &(161)\end{aligned}$$

In (159) we have used

$$h(A|B) - h(B|A) = h(A) - h(B). \tag{162}$$

B 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 $I\big(X_k; \mathbf{Z}_{k-\kappa}^{k-1} \big| X_{k-\kappa}^{k-1}, \mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{Y}_k^{k+\kappa}\big)$ and $I\big(X_k; \mathbf{Z}_{k+1}^{k+\kappa} \big| \mathbf{H}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k+1}^{k+\kappa}\big)$, respectively, that do not depend on k and that tend to zero as $x_{\min}$ tends to infinity.

We start with $I\big(X_k; \mathbf{Z}_{k-\kappa}^{k-1} \big| X_{k-\kappa}^{k-1}, \mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{Y}_k^{k+\kappa}\big)$:

$$
\begin{aligned}
&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^{k+\kappa}\big) \\
&\quad = h\big(\mathbf{Z}_{k-\kappa}^{k-1} \big| X_{k-\kappa}^{k-1}, \mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{Y}_k^{k+\kappa}\big) - h\big(\mathbf{Z}_{k-\kappa}^{k-1} \big| X_{k-\kappa}^{k-1}, \mathbf{Y}_{k-\kappa}^{k-1}, X_k, \mathbf{Y}_k^{k+\kappa}\big) && (163) \\
&\quad \leq h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big) - h\big(\mathbf{Z}_{k-\kappa}^{k-1} \big| X_{k-\kappa}^{k-1}, \mathbf{Y}_{k-\kappa}^{k-1}, X_k^{k+\kappa}, \mathbf{Z}_k^{k+\kappa}, \mathbf{Y}_k^{k+\kappa}\big) && (164) \\
&\quad = h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big) - h\big(\mathbf{Z}_{k-\kappa}^{k-1} \big| X_{k-\kappa}^{k-1}, \mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{H}_k^{k+\kappa}\big) && (165) \\
&\quad \leq h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big) - \min_{\substack{|x_{k-\kappa}| \geq x_{\min}, \ldots, \\ |x_{k-1}| \geq x_{\min}}} h\Big(\mathbf{Z}_{k-\kappa}^{k-1} \Big| X_{k-\kappa}^{k-1} = x_{k-\kappa}^{k-1}, \{\mathbf{H}_\ell x_\ell + \mathbf{Z}_\ell\}_{\ell=k-\kappa}^{k-1}, \mathbf{H}_k^{k+\kappa}\Big) && (166) \\
&\quad = h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big) - h\big(\mathbf{Z}_{k-\kappa}^{k-1} \big| \{\mathbf{H}_\ell x_{\min} + \mathbf{Z}_\ell\}_{\ell=k-\kappa}^{k-1}, \mathbf{H}_k^{k+\kappa}\big) && (167) \\
&\quad = I\left(\mathbf{Z}_{k-\kappa}^{k-1}; \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=k-\kappa}^{k-1}, \mathbf{H}_k^{k+\kappa}\right) && (168) \\
&\quad = I\left(\mathbf{Z}_1^{\kappa}; \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa}, \mathbf{H}_{\kappa+1}^{2\kappa+1}\right) && (169) \\
&\quad = I\left(\left\{\frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa}; \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa} \middle| \mathbf{H}_{\kappa+1}^{2\kappa+1}\right) && (170) \\
&\quad = h\left(\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa} \middle| \mathbf{H}_{\kappa+1}^{2\kappa+1}\right) - h\big(\mathbf{H}_1^{\kappa} \big| \mathbf{H}_{\kappa+1}^{2\kappa+1}\big) && (171) \\
&\quad \triangleq \epsilon_1(x_{\min}, \kappa). && (172)
\end{aligned}
$$

Here (164) follows from conditioning that reduces entropy; and in the subsequent equality we used $X_{k-\kappa}^{k-1}$ and $\mathbf{Z}_k^{k+\kappa}$ in order to extract $\mathbf{H}_k^{k+\kappa}$ from $\mathbf{Y}_k^{k+\kappa}$, and then we dropped $\{X_\ell, \mathbf{Y}_\ell, \mathbf{Z}_\ell\}_{\ell=k}^{k+\kappa}$ since given $\mathbf{H}_k^{k+\kappa}$ it is independent of the other random variables.

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

$$
h\left(\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=1}^{\kappa} \middle| \mathbf{H}_{\kappa+1}^{2\kappa+1} = \mathbf{h}_{\kappa+1}^{2\kappa+1}\right) \tag{173}
$$

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

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

$$
\begin{aligned}
&I\big(X_k; \mathbf{Z}_{k+1}^{k+\kappa} \big| \mathbf{H}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k+1}^{k+\kappa}\big) \\
&\quad = h\big(\mathbf{Z}_{k+1}^{k+\kappa} \big| \mathbf{H}_{k-\kappa}^{k-1}, \mathbf{Y}_k, \mathbf{Y}_{k+1}^{k+\kappa}\big) - h\big(\mathbf{Z}_{k+1}^{k+\kappa} \big| \mathbf{H}_{k-\kappa}^{k-1}, X_k, \mathbf{Y}_k, \mathbf{Y}_{k+1}^{k+\kappa}\big) && (174) \\
&\quad \leq h\big(\mathbf{Z}_{k+1}^{k+\kappa}\big) - h\big(\mathbf{Z}_{k+1}^{k+\kappa} \big| \mathbf{H}_{k-\kappa}^{k-1}, X_k, \mathbf{Y}_k, \mathbf{Z}_k, \mathbf{Y}_{k+1}^{k+\kappa}, X_{k+1}^{k+\kappa}\big) && (175) \\
&\quad = h\big(\mathbf{Z}_{k+1}^{k+\kappa}\big) - h\big(\mathbf{Z}_{k+1}^{k+\kappa} \big| \mathbf{H}_{k-\kappa}^{k}, \mathbf{Y}_{k+1}^{k+\kappa}, X_{k+1}^{k+\kappa}\big) && (176) \\
&\quad \leq h\big(\mathbf{Z}_{k+1}^{k+\kappa}\big) - \min_{\substack{|x_{k+1}| \geq x_{\min}, \ldots, \\ |x_{k+\kappa}| \geq x_{\min}}} h\Big(\mathbf{Z}_{k+1}^{k+\kappa} \Big| \mathbf{H}_{k-\kappa}^{k}, \{\mathbf{H}_\ell x_\ell + \mathbf{Z}_\ell\}_{\ell=k+1}^{k+\kappa}, X_{k+1}^{k+\kappa} = x_{k+1}^{k+\kappa}\Big) && (177) \\
&\quad = h\big(\mathbf{Z}_{k+1}^{k+\kappa}\big) - h\big(\mathbf{Z}_{k+1}^{k+\kappa} \big| \mathbf{H}_{k-\kappa}^{k}, \{\mathbf{H}_\ell x_{\min} + \mathbf{Z}_\ell\}_{\ell=k+1}^{k+\kappa}\big) && (178) \\
&\quad = I\left(\mathbf{Z}_{k+1}^{k+\kappa}; \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=k+1}^{k+\kappa}, \mathbf{H}_{k-\kappa}^{k}\right) && (179) \\
&\quad = I\left(\mathbf{Z}_{\kappa+2}^{2\kappa+1}; \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=\kappa+2}^{2\kappa+1}, \mathbf{H}_1^{\kappa+1}\right) && (180)
\end{aligned}
$$

$$= I\left(\left\{\frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=\kappa+2}^{2\kappa+1}; \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=\kappa+2}^{2\kappa+1}\middle| \mathbf{H}_1^{\kappa+1}\right) \tag{181}$$

$$\triangleq \epsilon_2(x_{\min}, \kappa), \tag{182}$$

from which the results follows analogously to (170).

C Appendix for the Proof of the Upper Bound

In the derivation of the upper bound it remains to derive upper bounds to

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

and

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

that depend only on $\xi_{\min}$ and not on k or the SNR.

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

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

$$= h\big(\mathbf{Z}_{k-\kappa}^{k-1} \big| \mathbf{Y}_{k-\kappa}^{k-1}, E_k{=}1\big) - h\big(\mathbf{Z}_{k-\kappa}^{k-1} \big| \mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{Y}_k, E_k{=}1\big) \tag{183}$$

$$\leq h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big) - h\big(\mathbf{Z}_{k-\kappa}^{k-1} \big| \mathbf{Y}_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, \mathbf{Y}_k, \mathbf{Z}_k, X_k, E_k{=}1\big) \tag{184}$$

$$= h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big) - h\big(\mathbf{Z}_{k-\kappa}^{k-1} \big| \mathbf{Y}_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, \mathbf{H}_k, E_k{=}1\big) \tag{185}$$

$$\leq h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big) - \min_{\substack{|x_{k-\kappa}|\geq\xi_{\min},\ldots,\\ |x_{k-1}|\geq\xi_{\min}}} h\big(\mathbf{Z}_{k-\kappa}^{k-1} \big| \{\mathbf{H}_\ell x_\ell + \mathbf{Z}_\ell\}_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1} = x_{k-\kappa}^{k-1}, \mathbf{H}_k, E_k{=}1\big) \tag{186}$$

$$= h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big) - h\big(\mathbf{Z}_{k-\kappa}^{k-1} \big| \{\mathbf{H}_\ell \xi_{\min} + \mathbf{Z}_\ell\}_{k-\kappa}^{k-1}, \mathbf{H}_k\big) \tag{187}$$

$$= I\left(\mathbf{Z}_{k-\kappa}^{k-1}; \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{k-\kappa}^{k-1}, \mathbf{H}_k\right) \tag{188}$$

$$= I\left(\mathbf{Z}_{k-\kappa}^{k-1}; \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{k-\kappa}^{k-1} \middle| \mathbf{H}_k\right) \tag{189}$$

$$= h\left(\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{k-\kappa}^{k-1} \middle| \mathbf{H}_k\right) - h\big(\mathbf{H}_{k-\kappa}^{k-1} \big| \mathbf{H}_k\big) \tag{190}$$

$$= h\left(\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{1}^{\kappa} \middle| \mathbf{H}_{\kappa+1}\right) - h\big(\mathbf{H}_1^{\kappa} \big| \mathbf{H}_{\kappa+1}\big) \tag{191}$$

$$\triangleq \delta_1(\xi_{\min}, \kappa), \tag{192}$$

where the last equality follows from stationarity.

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

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

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

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

$$
\begin{aligned}
&I\big(\mathbf{Z}_k;\{\mathbf{H}_\ell X_\ell\}_{k-\kappa}^{k-1}\big|\mathbf{Y}_k,E_k{=}1\big)\\
&\quad= h\big(\mathbf{Z}_k\big|\mathbf{Y}_k,E_k{=}1\big)-h\big(\mathbf{Z}_k\big|\{\mathbf{H}_\ell X_\ell\}_{k-\kappa}^{k-1},\mathbf{Y}_k,E_k{=}1\big) && (193)\\
&\quad\le h\big(\mathbf{Z}_k\big)-h\big(\mathbf{Z}_k\big|\{\mathbf{H}_\ell X_\ell\}_{k-\kappa}^{k-1},X_{k-\kappa}^{k-1},\mathbf{Y}_k,X_k,E_k{=}1\big) && (194)\\
&\quad= h\big(\mathbf{Z}_k\big)-h\big(\mathbf{Z}_k\big|\mathbf{H}_{k-\kappa}^{k-1},\mathbf{Y}_k,X_k,E_k{=}1\big) && (195)\\
&\quad\le h\big(\mathbf{Z}_k\big)-\min_{|x_k|\ge\xi_{\min}} h\big(\mathbf{Z}_k\big|\mathbf{H}_{k-\kappa}^{k-1},\mathbf{H}_k x_k+\mathbf{Z}_k,X_k=x_k,E_k{=}1\big) && (196)\\
&\quad= h\big(\mathbf{Z}_k\big)-h\big(\mathbf{Z}_k\big|\mathbf{H}_{k-\kappa}^{k-1},\mathbf{H}_k\xi_{\min}+\mathbf{Z}_k\big) && (197)\\
&\quad= I\left(\mathbf{Z}_k;\mathbf{H}_k+\frac{\mathbf{Z}_k}{\xi_{\min}},\mathbf{H}_{k-\kappa}^{k-1}\right) && (198)\\
&\quad= I\left(\mathbf{Z}_k;\mathbf{H}_k+\frac{\mathbf{Z}_k}{\xi_{\min}}\middle|\,\mathbf{H}_{k-\kappa}^{k-1}\right) && (199)\\
&\quad= h\left(\mathbf{H}_k+\frac{\mathbf{Z}_k}{\xi_{\min}}\middle|\,\mathbf{H}_{k-\kappa}^{k-1}\right)-h\big(\mathbf{H}_k\big|\mathbf{H}_{k-\kappa}^{k-1}\big) && (200)\\
&\quad= h\left(\mathbf{H}_{\kappa+1}+\frac{\mathbf{Z}_{\kappa+1}}{\xi_{\min}}\middle|\,\mathbf{H}_1^{\kappa}\right)-h\big(\mathbf{H}_{\kappa+1}\big|\mathbf{H}_1^{\kappa}\big) && (201)\\
&\quad\triangleq \delta_2(\xi_{\min},\kappa), && (202)
\end{aligned}
$$

where the last equality follows from stationarity.

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

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

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

Acknowledgment

The authors gratefully acknowledge İ. E. Telatar's contribution to the proofs in Appendices B and C.

References

[1] Amos Lapidoth and Stefan M. Moser. Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels. *IEEE Transactions on Information Theory*, 49(10):2426–2467, October 2003.

[2] Mohsen Pourahmadi. *Foundations of Time Series Analysis and Prediction Theory.* John Wiley & Sons, 2001.

[3] A. M. Yaglom. *An Introduction to the Theory of Stationary Random Functions.* Prentice Hall, 1962.

[4] A. N. Kolmogorov. Stationary sequences in Hilbert space. *Bulletin Moscow State University*, 2(6), 1941.

[5] A. N. Kolmogorov. Interpolation and extrapolation of stationary random sequences. *Izvestija Akademii Nauk SSSR*, 5 (Ser. Mat.):3, 1941.

[6] Jiří Anděl. A comparison of methods for interpolation in time series. In W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, editors, *Mathematical Statistics and Applications*, volume B (Bad Tatzmannsdorf), pages 53–67. 1983.

[7] Robert G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.

[8] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular Variation*, volume 27 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 1987.