Impact of Feedback and Side-Information on the Asymptotic Capacity of Single-Input Multiple-Output Fading Channels With Memory

Stefan M. Moser, *Senior Member, IEEE*

***Abstract*—An analytic expression for the asymptotic capacity of noncoherent single-input multiple-output (SIMO) regular fading channels with memory and with partial receiver side-information is derived and is shown to remain unchanged by causal or acausal side-information at the transmitter and by a noiseless feedback link. In particular, the corresponding fading numbers are identical. Furthermore, the asymptotic capacity of a single-input single-output (SISO) nonregular Gaussian fading channel with memory is investigated, and it is shown that the prelog is unaffected by noiseless feedback.**

***Index Terms*—Causality, channel capacity, fading channel, fading number, feedback, high signal-to-noise ratio (SNR), multiple antennas, nonregular Gaussian fading, prelog, regular fading, side-information.**

I. Introduction

In LITERATURE, there exists a large variety of different channel models that try to describe the behavior of mobile wireless communication systems. The historically oldest models are the *coherent fading channels* that assume that the receiver has free and *noiseless* access to the current fading realization [1]. This simplification can sometimes be justified, e.g., at low power when the errors are dominated by the additive noise. At high power, the *noncoherent* fading models describe real systems more accurately because there it is assumed that transmitter and receiver only have statistical knowledge of the fading process, but no direct access to the current fading realization.

There are again many families of such noncoherent models, reaching from *block-fading models* (fading remains perfectly unchanged during a certain time, before it takes on a new, possibly dependent value [2]–[4]), to *underspread fading channels* (the fading process is wide-sense stationary and uncorrelated in the delay, and the product of the delay and Doppler spread is small [5] (and references therein)), and to *stationary fading models* [6]–[8].

In this paper, we will focus on the last family. We will assume that the fading process is some stationary and ergodic stochastic process of finite energy. It has been shown that depending on the exact assumptions about this process, the high-SNR capacity of such a noncoherent fading channel can vary largely. For example, so-called *regular* fading channels[1] exhibit an extremely slow, double-logarithmic growth of the capacity in the available power [6]. To describe the exact asymptotic behavior, [6] introduces the *fading number* χ as the second term in the high-SNR asymptotic expansion of the channel capacity C:

$$\chi \triangleq \varlimsup_{\mathsf{E_s}\uparrow\infty} \{\mathsf{C}(\mathsf{E_s}) - \log\log \mathsf{E_s}\}. \tag{1}$$

(Here $\mathsf{E_s}$ denotes either the available average power or the available peak power at the transmitter.) An analytic expression for its value for general multiple-input multiple-output fading channels with memory has been derived in [6], [9], [10]. In particular, the single-input multiple-output (SIMO) fading number with memory is given as[2]

$$\begin{aligned}\chi(\{\mathbf{H}_k\}) = h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\infty}^{-1}\Big) - \log 2 \\ + n_\mathsf{R}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big] - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}\big)\end{aligned} \tag{2}$$

where $\{\Theta_k\}$ is independent and identically distributed (IID) $\sim \mathcal{U}((-\pi,\pi])$ and independent of $\{\mathbf{H}_k\}$.

We remark that beside describing the asymptotic capacity, the fading number is also of importance as an indicator of the borderline between the logarithmically growing low-SNR regime and the double-logarithmically growing high-SNR regime (for more details see [10, Section I]).

For *nonregular* fading processes, the growth rate can range from double-logarithmic up to logarithmic, depending on the specific assumptions about the process [7]. In the situation of a logarithmic growth, we are particularly interested in the factor in front of the logarithm, the so-called *prelog*,[3]

$$\Pi \triangleq \varlimsup_{\mathsf{E_s}\uparrow\infty} \frac{\mathsf{C}(\mathsf{E_s})}{\log \mathsf{E_s}}. \tag{3}$$

Manuscript received December 3, 2013; revised March 13, 2014; accepted March 13, 2014. Date of publication March 19, 2014; date of current version May 15, 2014. This work was supported in part by the National Science Council of Taiwan under Grant NSC 101-2628-E-009-002 and in part by ETH under Grant TH-23 02-2. This paper was presented in parts at the 2004 IEEE Convention of Electrical and Electronics Engineers, the 2005 Winter School on Coding and Information Theory, and the 2013 IEEE International Symposium on Information Theory.

The author is with ETH Zurich, Zurich 8092, Switzerland, and also with the National Chiao Tung University, Hsinchu 300, Taiwan (e-mail: stefan.moser@ieee.org).

Communicated by A. Lozano, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2014.2312394

[1]Loosely speaking, a fading process is regular if its current value cannot be predicted precisely even if the infinite past is known exactly. We define regularity based on the value of the differential entropy rate of the process, see Section II.

[2]For a definition of $\hat{\mathbf{H}}$ and $h_\lambda(\cdot)$, see Section III-A.

[3]In literature, this term sometimes is also called 'multiplexing gain'.

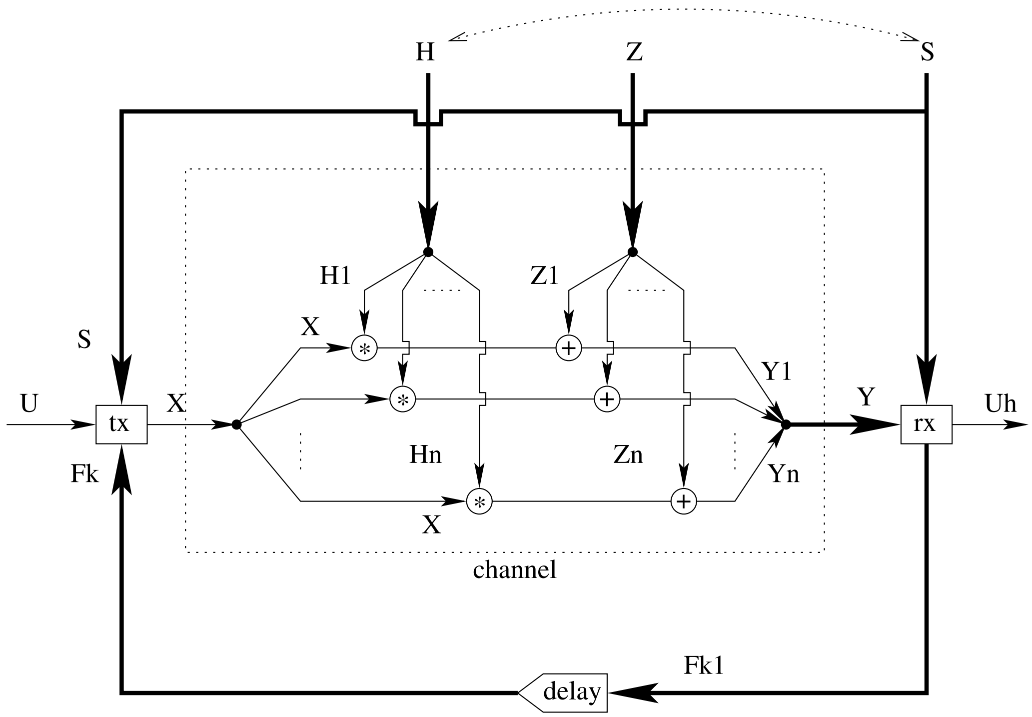

Fig. 1. SIMO fading channel with causal side-information at both transmitter and receiver and a noiseless causal feedback link from the receiver back to the transmitter.

For the situation of a single-antenna stationary complex Gaussian fading channel with a line-of-sight component and a given spectral distribution function $\mathsf{F}(\lambda)$, $-\frac{1}{2} < \lambda \leq \frac{1}{2}$, the prelog (under a peak-power constraint) has been derived in [7]:

$$\Pi^{\text{pp}}(\{H_k\}) = \mu_{\mathcal{L}}\bigl(\{\lambda\colon \mathsf{F}'(\lambda) = 0\}\bigr) \tag{4}$$

where $\mu_{\mathcal{L}}(\cdot)$ denotes the Lebesgue measure on the interval $[-1/2, 1/2]$.

In this work, we return to these results and investigate what the impact of a feedback link and of side-information is on the asymptotic capacity of these fading channels. Our contributions are as follows:

1) We derive the fading number of a general SIMO regular fading channel with memory in the situation when the receiver has access to some partial side-information about the fading process.
2) We derive the fading number of a general SIMO regular fading channel with memory in the situation when there exists an arbitrary causal feedback link from the receiver back to the transmitter and when both receiver and transmitter have access to some partial side-information about the fading process.
3) We generalize the result of Contribution 2 to the situation when the transmitter has *acausal* access to the side-information.
4) We derive the prelog of a single-input single-output (SISO) Gaussian fading channel with temporal memory in the situation when there exists an arbitrary causal feedback link from the receiver back to the transmitter.

The remainder of this paper is structured as follows. Section II starts with a detailed definition of the channel model and then gives some comments about our notation. Section III summarizes some mathematical preliminaries before, in Section IV, we present our main results together with a discussion.

The rest of the paper then focuses on the derivations: in Section V, we find a lower bound to the fading number of a SIMO fading channel with receiver side-information (no feedback); in Sections VI and VII, we present the derivations of an upper bound on the fading number of a SIMO fading channel with feedback and with causal and acausal side-information, respectively; and in Section VIII, we investigate the prelog of a SISO Gaussian fading channel with memory and with feedback.

Some parts of the derivations have been moved to the appendix. In particular, in Appendix A we focus on detailed investigations of dependencies between different random quantities, where we rely on a graphical tool presented in [11], [12].

II. Channel Model

We consider a communication system as depicted in Fig. 1. A message M is transmitted over a single-input multiple-output (SIMO) fading channel with memory. The channel output $\mathbf{Y}_k \in \mathbb{C}^{n_\text{R}}$ at time k (with n_R components corresponding to the n_R antennas at the receiver) is given by

$$\mathbf{Y}_k = \mathbf{H}_k x_k + \mathbf{Z}_k \tag{5}$$

where $x_k \in \mathbb{C}$ denotes the input of the channel at time k; the random vector $\mathbf{H}_k \in \mathbb{C}^{n_\text{R}}$ denotes the time-k fading vector; and $\mathbf{Z}_k \in \mathbb{C}^{n_\text{R}}$ denotes the time-k additive noise vector. It is assumed that $\{\mathbf{Z}_k\}$ and $\{\mathbf{H}_k\}$ are independent and that their joint law does not depend on the channel input. The fading is noncoherent, i.e., neither transmitter nor receiver know the realization of the fading process $\{\mathbf{H}_k\}$; they only know its law.

The additive noise process $\{\mathbf{Z}_k\}$ is assumed to be a spatially and temporally white, zero-mean, circularly symmetric, complex Gaussian vector process,

$$\{\mathbf{Z}_k\} \text{ IID} \sim \mathcal{N}_{\mathbb{C}}\big(\mathbf{0}, \sigma^2 \mathsf{I}\big) \tag{6}$$

for some $\sigma^2 > 0$.

The fading $\{\mathbf{H}_k\}$ is foremost assumed to be a stationary and ergodic stochastic process of finite energy

$$\mathsf{E}\left[\|\mathbf{H}_k\|^2\right] < \infty. \tag{7}$$

Then we consider two different scenarios. In the first scenario, we do not specify a particular distribution, but only make the additional assumption of the fading being a *regular* stochastic process, i.e., $\{\mathbf{H}_k\}$ is of finite differential entropy rate

$$h(\{\mathbf{H}_k\}) > -\infty. \tag{8}$$

In the second scenario we address the more general case where $h(\{H_k\})$ need not be finite, but we restrict ourselves to Gaussian fading and to only one antenna at the receiver $n_{\text{R}} = 1$. Specifically, we assume that $\{\tilde{H}_k\} \triangleq \{H_k - d\}$ is a zero-mean, finite-variance, stationary, circularly-symmetric, Gaussian process of some arbitrary spectral distribution function

$$\mathsf{F}(\lambda), \quad -\frac{1}{2} \le \lambda \le \frac{1}{2}. \tag{9}$$

The constant $d \in \mathbb{C}$ denotes the spectral component of the fading process $\{H_k\}$.

Furthermore, in the first scenario of a regular fading process, we also consider a side-information process $\{\mathbf{S}_k\}$, $\mathbf{S}_k \in \mathbb{C}^{n_{\text{S}}}$, that carries partial information about the fading process. It is assumed that the fading process $\{\mathbf{H}_k\}$ and the side-information process $\{\mathbf{S}_k\}$ are jointly stationary, ergodic, of finite energy, and of finite joint differential entropy rate

$$h(\{\mathbf{H}_k, \mathbf{S}_k\}) \triangleq \lim_{n\to\infty} \frac{1}{n} h\big(\mathbf{H}_1^n, \mathbf{S}_1^n\big) > -\infty. \tag{10}$$

It is still assumed that $(\{\mathbf{H}_k\}, \{\mathbf{S}_k\})$ are independent of the additive noise $\{\mathbf{Z}_k\}$, and the joint law of $(\{\mathbf{H}_k\}, \{\mathbf{S}_k\}, \{\mathbf{Z}_k\})$ does not depend on the channel input.

Finally, we allow a feedback link from the receiver back to the transmitter. The feedback is assumed to be noiseless (i.e., of infinite capacity), but delayed by one time-step, so that the feedback random vector $\mathbf{F}_k$ that is available to the transmitter at time instant k consists of the complete knowledge of the receiver at time $k-1$, i.e.,

$$\mathbf{F}_k = \mathbf{Y}_1^{k-1} \tag{11}$$

or, if the receiver also has side-information available,

$$\mathbf{F}_k = \big(\mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}\big). \tag{12}$$

Such a feedback link is, of course, overly optimistic for any realistic feedback system, which will always be limited by a finite rate. However, it will serve as an upper bound on what is possible with any type of realistic feedback. In fact, we will show that at high power, feedback will not increase capacity in spite of it being noiseless and in spite of the memory in the channel.

We consider two types of power constraints. Either we will constrain the input power on average,

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[\left|X_k\big(M, \mathbf{S}_1^k, \mathbf{F}_1^k\big)\right|^2\right] \le \mathsf{E}_{\mathsf{s}} \tag{13}$$

or we impose a peak-power constraint,

$$\left|X_k\big(M, \mathbf{S}_1^k, \mathbf{F}_1^k\big)\right|^2 \le \mathsf{E}_{\mathsf{s}}, \quad \text{a.s.}, \ \forall\, k = 1, \ldots, n. \tag{14}$$

We end this section with a few remarks on our notation. Since this paper compares various different channel models, we try to be very careful in specifying the current assumptions. So, we use superscripts "avg" and "pp" to denote an average-power constraint and a peak-power constraint, respectively. If neither superscript is given, then the result holds for both types of power constraints.

Whenever feedback is available, this is highlighted by a subscript "FB". For receiver side-information, we use a "conditioning" notation, e.g., $\mathsf{R}(\mathsf{E}_{\mathsf{s}}|\{\mathbf{S}_k\})$ represents the rate for power E_{s} and with side-information $\{\mathbf{S}_k\}$ available at the receiver. If the side-information is also available at the transmitter, we use the subscript "c" or "ac" for causal and acausal side-information at the transmitter, respectively.

Clearly,

$$\mathsf{C}^{\text{pp}}(\mathsf{E}_{\mathsf{s}}) \le \mathsf{C}^{\text{avg}}(\mathsf{E}_{\mathsf{s}}) \le \mathsf{C}^{\text{avg}}_{\text{FB}}(\mathsf{E}_{\mathsf{s}}) \tag{15}$$

$$\mathsf{C}^{\text{pp}}(\mathsf{E}_{\mathsf{s}}) \le \mathsf{C}^{\text{pp}}_{\text{FB}}(\mathsf{E}_{\mathsf{s}}) \tag{16}$$

$$\mathsf{C}^{i}(\mathsf{E}_{\mathsf{s}}|\{\mathbf{S}_k\}) \le \mathsf{C}^{i}_{\text{c}}(\mathsf{E}_{\mathsf{s}}|\{\mathbf{S}_k\}) \le \mathsf{C}^{i}_{\text{ac}}(\mathsf{E}_{\mathsf{s}}|\{\mathbf{S}_k\}), \quad i \in \{\text{avg}, \text{pp}\}. \tag{17}$$

We meticulously distinguish between random and nonrandom quantities. A random variable is denoted by a capital Roman letter, e.g., M, while its realization is denoted by the corresponding small Roman letter, e.g., m. Vectors are boldfaced, e.g., $\mathbf{H}$ denotes a random vector and $\mathbf{h}$ its realization. Constants are typeset either in small Romans, in Greek letters or in a capital sans-serif font, e.g., $\mathsf{F}(\lambda)$ or E_{s}. The only exception to these rules is the mutual information functional that is typeset as $I(\cdot;\cdot)$ as is very common in literature. Sets are typeset in calligraphic font, e.g., $\mathcal{V}$, and $\mathbb{C}$ and $\mathbb{R}$ denote the fields of the complex and the real numbers, respectively.

Entropy is typeset as $\mathsf{H}(\cdot)$, differential entropy by $h(\cdot)$, and by $\mathsf{H}_{\mathsf{b}}(\cdot)$ we denote the binary entropy function

$$\mathsf{H}_{\mathsf{b}}(p) \triangleq -p\log p - (1-p)\log(1-p), \quad p \in [0,1]. \tag{18}$$

For a unit vector $\hat{\mathbf{H}}$, we write $h_\lambda(\hat{\mathbf{H}})$ for the differential entropy with respect to the surface area of a unit sphere in $\mathbb{C}^m$, see Section III-A.

By $M \sim \mathcal{U}(\mathcal{M})$ we mean that the random variable M is uniformly distributed over the set $\mathcal{M}$, and $\mathbf{H}_i^j$ stands for $(\mathbf{H}_i, \ldots, \mathbf{H}_j)$.

We exclusively use the natural logarithm, and all rates are therefore specified in nats.

III. Mathematical Preliminaries

A. Differential Entropy and Expected Logarithms

In [6, Section VI.D] (and in much abbreviated form also in [10, Section II]), one finds an extensive discussion of fundamental properties of differential entropy and expected logarithms. The main point made there is that for a random vector $\mathbf{H}$ that is of finite second moment $\mathsf{E}\big[\|\mathbf{H}\|^2\big] < \infty$ and of finite differential entropy $h(\mathbf{H}) > -\infty$, expected logarithms and differential entropy expressions are well-defined and finite.

This can be understood when realizing that differential entropy can be written as the difference of two nonnegative parts:

$$h(\mathbf{H}) = h^+(\mathbf{H}) - h^-(\mathbf{H}) \tag{19}$$

where

$$h^+(\mathbf{H}) \triangleq \int_{\{\mathbf{h}\colon 0<f_{\mathbf{H}}(\mathbf{h})<1\}} f_{\mathbf{H}}(\mathbf{h}) \log \frac{1}{f_{\mathbf{H}}(\mathbf{h})} \,\mathrm{d}\mathbf{h} \geq 0 \tag{20}$$

$$h^-(\mathbf{H}) \triangleq \int_{\{\mathbf{h}\colon f_{\mathbf{H}}(\mathbf{h})>1\}} f_{\mathbf{H}}(\mathbf{h}) \log f_{\mathbf{H}}(\mathbf{h}) \,\mathrm{d}\mathbf{h} \geq 0. \tag{21}$$

Note that for the differential entropy to be defined, at least one of the two integrals (20) or (21) must be finite.

The assumption of a finite second moment now guarantees a finite upper bound on $h^+(\mathbf{H})$ and therefore on $h(\mathbf{H})$, and the assumption of a finite differential entropy rate makes sure that $h^-(\mathbf{H})$ is finite. Moreover, this analysis can be extended to the expected logarithm: The assumption of finite energy trivially guarantees that $\mathsf{E}\big[\log\|\mathbf{H}\|^2\big]$ is bounded from above (by Jensen's inequality), while the boundedness of $h^-(\mathbf{H})$ provides a finite lower bound on $\mathsf{E}\big[\log\|\mathbf{H}\|^2\big]$.

These arguments can also be generalized to a situation with memory. There a finite second moment and a finite lower bound on the *entropy rate* will make sure that all differential entropy expressions and expected logarithms are well-defined and finite. Moreover, we can also include situations when we condition a differential entropy expression or an expected logarithm to one out of a finite number of disjoint events that form a partition. Again, the assumption of a finite second moment and a finite differential entropy rate will make sure that such conditional entropy or logarithm expressions remain finite. Compare also with the derivations shown in Appendix E.

Moreover, all these results also hold for the differential entropy $h_\lambda(\cdot)$ with respect to the surface of the unit sphere in $\mathbb{C}^m$ as defined in [6, Section VI.D]. Recall that if we split a complex random vector $\mathbf{H} \in \mathbb{C}^m$ up into magnitude $\|\mathbf{H}\| \in \mathbb{R}_0^+$ and direction

$$\hat{\mathbf{H}} \triangleq \frac{\mathbf{H}}{\|\mathbf{H}\|} \tag{22}$$

then the latter is a unit vector and therefore has zero measure with respect to a probability distribution over $\mathbb{C}^m$. We therefore define a new probability density function (PDF) $f_{\hat{\mathbf{H}}}^{\lambda}(\cdot)$ with respect to a measure λ that lives on the m-dimensional complex unit sphere. The corresponding differential entropy then simply is defined as

$$h_\lambda\big(\hat{\mathbf{H}}\big) \triangleq \mathsf{E}\big[-\log f_{\hat{\mathbf{H}}}^{\lambda}(\hat{\mathbf{H}})\big]. \tag{23}$$

The connection between $h(\mathbf{H})$ and $h_\lambda(\hat{\mathbf{H}})$ is as follows.

Lemma 1: Let $\mathbf{H} \in \mathbb{C}^m$ be a complex random vector with a finite differential entropy $h(\mathbf{H})$. Let $\|\mathbf{H}\|$ denote its magnitude and $\hat{\mathbf{H}}$ its direction as in (22). Then

$$h(\mathbf{H}) = h(\|\mathbf{H}\|) + h_\lambda\big(\hat{\mathbf{H}}\big|\|\mathbf{H}\|\big) + (2m-1)\,\mathsf{E}[\log\|\mathbf{H}\|] \tag{24}$$

$$= h_\lambda\big(\hat{\mathbf{H}}\big) + h\big(\|\mathbf{H}\|\big|\hat{\mathbf{H}}\big) + (2m-1)\,\mathsf{E}[\log\|\mathbf{H}\|]. \tag{25}$$

Proof: See, e.g., [10, Lemma 2]. ■

B. Stationarity and Joint Differential Entropy Rate

The joint differential entropy rate of two jointly stochastic processes $\{\mathbf{H}_k, \mathbf{S}_k\}$ is defined as

$$h(\{\mathbf{H}_k, \mathbf{S}_k\}) \triangleq \lim_{n\uparrow\infty} \frac{1}{n}\, h\big(\mathbf{H}_1^n, \mathbf{S}_1^n\big) \tag{26}$$

if the limit exists. By our assumptions of $\{\mathbf{H}_k, \mathbf{S}_k\}$ being jointly stationary, ergodic, of finite energy and of finite differential entropy rate we make sure that the limit indeed exists and is finite. Moreover, it is not difficult to show that — under above mentioned assumptions — also

$$h(\{\mathbf{H}_k, \mathbf{S}_k\}) = \lim_{n\uparrow\infty} h\big(\mathbf{H}_n, \mathbf{S}_n\big|\mathbf{H}_1^{n-1}, \mathbf{S}_1^{n-1}\big). \tag{27}$$

For the latter we usually write

$$\begin{aligned}
&\lim_{n\uparrow\infty} h\big(\mathbf{H}_n, \mathbf{S}_n\big|\mathbf{H}_1^{n-1}, \mathbf{S}_1^{n-1}\big) \\
&\quad= \lim_{n\uparrow\infty} h\big(\mathbf{H}_0, \mathbf{S}_0\big|\mathbf{H}_{-n+1}^{-1}, \mathbf{S}_{-n+1}^{-1}\big) && (28)\\
&\quad= h\big(\mathbf{H}_0, \mathbf{S}_0\big|\mathbf{H}_{-\infty}^{-1}, \mathbf{S}_{-\infty}^{-1}\big) && (29)
\end{aligned}$$

where the first equality follows from stationarity and the second is a convenient shorthand. One needs to be aware, however, that this shorthand hides a limit (which exists and is finite).

Note that there exist many variations of such entropy definitions. Under our assumptions, all these different expressions are well-defined and finite. For example, the conditional entropy rate can be given in various different equivalent forms:

$$\begin{aligned}
&h\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big) \\
&\quad\triangleq \lim_{n\uparrow\infty} \frac{1}{n}\, h(\mathbf{H}_1^n|\mathbf{S}_1^n) && (30)\\
&\quad= \lim_{n\uparrow\infty} \frac{1}{n}\big(h(\mathbf{H}_1^n, \mathbf{S}_1^n) - h(\mathbf{S}_1^n)\big) && (31)\\
&\quad= h(\{\mathbf{H}_k, \mathbf{S}_k\}) - h(\{\mathbf{S}_k\}) && (32)\\
&\quad= \lim_{n\uparrow\infty} \Big\{ h\big(\mathbf{H}_n, \mathbf{S}_n\big|\mathbf{H}_1^{n-1}, \mathbf{S}_1^{n-1}\big) - h\big(\mathbf{S}_n\big|\mathbf{S}_1^{n-1}\big) \Big\} && (33)\\
&\quad= \lim_{n\uparrow\infty} \Big\{ h\big(\mathbf{H}_n\big|\mathbf{H}_1^{n-1}, \mathbf{S}_1^{n}\big) + h\big(\mathbf{S}_n\big|\mathbf{H}_1^{n-1}, \mathbf{S}_1^{n-1}\big) \\
&\qquad\qquad - h\big(\mathbf{S}_n\big|\mathbf{S}_1^{n-1}\big) \Big\} && (34)\\
&\quad= \lim_{n\uparrow\infty} \Big\{ h\big(\mathbf{H}_n\big|\mathbf{H}_1^{n-1}, \mathbf{S}_1^{n}\big) - I\big(\mathbf{S}_n; \mathbf{H}_1^{n-1}\big|\mathbf{S}_1^{n-1}\big) \Big\} && (35)\\
&\quad= \lim_{n\uparrow\infty} h\big(\mathbf{H}_0\big|\mathbf{H}_{-n+1}^{-1}, \mathbf{S}_{-n+1}^{0}\big) && (36)\\
&\quad= h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, \mathbf{S}_{-\infty}^{0}\big). && (37)
\end{aligned}$$

Here, (36) can be argued using a tool explained in Appendix A. Again, the reader is warned to keep in mind that the shorthand (37) actually involves a limit, but that this limit does exist and is finite.

C. Duality-Based Bounds on Mutual Information

Based on an identity given in [13] and [14, Section 2.3, Equation (3.7)], a duality-based upper bound on the mutual information between input and output of a memoryless channel is presented in [6], [15]. For convenience, we review this bound here quickly, however, we state it without proof and only in the form needed for our derivations. For a more general version and for the proofs, we refer to [6], [15].

Lemma 2: Consider a memoryless channel with input $X \in \mathbb{C}$ and output $T \in \mathbb{C}$. Then

$$I(X;T) \leq -h(T|X) + \log\pi + \mu\log\eta + \log\Gamma\left(\mu, \frac{\nu}{\eta}\right) + \frac{\nu}{\eta} + (1-\mu)\,\mathsf{E}\left[\log\left(|T|^2+\nu\right)\right] + \frac{1}{\eta}\,\mathsf{E}\left[|T|^2\right] \tag{38}$$

where $\mu, \eta > 0$ and $\nu \geq 0$ are free parameters.

IV. Main Results

A. Regular Fading with Memory and Receiver Side-Information

Theorem 3: Consider a SIMO regular fading channel with memory as given in (5) and (8), where the receiver has access to some partial side-information $\{\mathbf{S}_k\}$ according to the description around (10), and where the input is subject to either an average-power constraint (13) or a peak-power constraint (14). Then the capacity grows double-logarithmically in the power and the fading number is given as

$$\chi\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big) \triangleq \varlimsup_{\mathsf{E}_s\uparrow\infty}\Big\{\mathsf{C}\big(\mathsf{E}_s\big|\{\mathbf{S}_k\}\big) - \log\log\mathsf{E}_s\Big\} \tag{39}$$

$$= h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\infty}^{-1}, \mathbf{S}_{-\infty}^0\Big) - \log 2 + n_\mathrm{R}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big] - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, \mathbf{S}_{-\infty}^0\big) \tag{40}$$

where $\{\Theta_k\}$ is an independent random process that is IID $\sim \mathcal{U}((-\pi,\pi])$.

Proof: A lower bound under the assumption of a peak-power constraint is derived in Section V. The result then follows because

$$\chi\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big) \leq \chi_{\mathrm{FB,c}}^{\mathrm{avg}}\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big) \tag{41}$$

and from Theorem 4 below. ■

Note that $\chi\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big)$ can be expressed with the help of the SIMO fading number $\chi(\{\mathbf{H}_k\})$ with memory, but without side-information as given in (2) as follows:

$$\chi\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big) = \chi(\{\mathbf{H}_k\}) + I\big(\mathbf{H}_0; \mathbf{S}_{-\infty}^0\big|\mathbf{H}_{-\infty}^{-1}\big) - I\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}; \mathbf{S}_{-\infty}^0\Big|\big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\infty}^{-1}\Big). \tag{42}$$

B. Regular Fading with Memory, Feedback, and Causal Side-Information

Theorem 4: Consider a SIMO regular fading channel as given in Theorem 3, but additionally assume a noiseless causal feedback link from the receiver back to the transmitter and assume that the side-information also is revealed causally to the transmitter (see Fig. 1). Then the capacity remains as given in Theorem 3, i.e., it grows double-logarithmically in the power and has the same fading number

$$\chi_{\mathrm{FB,c}}\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big) = \chi\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big). \tag{43}$$

Proof: Using

$$\chi_{\mathrm{FB,c}}\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big) \geq \chi^{\mathrm{pp}}\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big) \tag{44}$$

and Theorem 3, we get a lower bound under the assumption of a peak-power constraint. An upper bound, under the assumption of an average-power constraint, is derived in Section VI. ■

We see that feedback and transmitter side-information do not change the asymptotic capacity in spite of the memory in the channel.

Corollary 5: Any type of feedback or causal side-information at the transmitter does not increase the fading number of a general SIMO fading channel with memory. As a matter of fact, it is not difficult to adapt the proof to show that even the revelation of the past fading realizations at the transmitter does not increase the asymptotic capacity.

C. Regular Fading with Memory, Feedback, and Acausal Side-Information

Theorem 6: Consider a SIMO regular fading channel with feedback and side-information as given in Theorem 4, but assume that the side-information is revealed **acausally** to the transmitter in advance. Then the capacity still grows only double-logarithmically in the power and the fading number remains unchanged:

$$\chi_{\mathrm{FB,ac}}\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big) = \chi_{\mathrm{FB,c}}\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big) \tag{45}$$

$$= \chi\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big). \tag{46}$$

Proof: A lower bound follows from

$$\chi_{\mathrm{FB,ac}}\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big) \geq \chi^{\mathrm{pp}}\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big) \tag{47}$$

and Theorem 3. An upper bound, under the assumption of an average-power constraint, is derived in Section VII. ■

Note again that the result continues to hold even if the past fading realizations are revealed to the transmitter.

D. Nonregular Gaussian Fading with Memory and Feedback

Theorem 7: Consider a SISO nonregular Gaussian fading channel with spectral distribution function $\mathsf{F}(\cdot)$ as described around (9), and consider a peak-power constraint (14). Then the prelog of the asymptotic capacity with a causal noiseless feedback link is identical to the prelog without feedback and is given as

$$\Pi_{\mathrm{FB}}^{\mathrm{pp}}(\{H_k\}) = \Pi^{\mathrm{pp}}(\{H_k\}) = \mu_{\mathcal{L}}\big(\{\lambda\colon \mathsf{F}'(\lambda) = 0\}\big). \tag{48}$$

Proof: A lower bound follows from

$$\Pi_{\mathrm{FB}}^{\mathrm{pp}}(\{H_k\}) \geq \Pi^{\mathrm{pp}}(\{H_k\}) \tag{49}$$

and from the results given in [7]. An upper bound is derived in Section VIII. ■

E. Discussion

We see that the asymptotic capacity of the large class of SIMO regular fading channels remains unchanged even if one allows causal noiseless feedback and transmitter side-information. This once more exemplifies the extremely unattractive behavior of regular fading channels at high SNR: besides the double-logarithmic growth [6] and the very poor performance in a multiple-user setup (where the maximum sum-rate only can be achieved if the channel is used exclusively by one user only and the other users can never communicate at all [16]), we now also have shown that any type of feedback does not increase capacity in spite of the memory in the channel.

Similarly, also the capacity of nonregular Gaussian fading channels is not strongly increased by feedback as the factor in front of the logarithm is not improved by feedback. Note that for proof-technical reasons we have only proven the case of a peak-power constraint. We believe that the channel will exhibit the same behavior also under an average-power constraint.

We would like to point out the main challenges for the derivations given in Sections V–VIII:

1) Due to the feedback, the channel input, the fading, and the additive noise become dependent.
2) We cannot rely on the important auxiliary result given in [10, Theorem 3] that shows that the optimal input is stationary. Indeed, since the transmitter continually learns more about the fading process through the feedback, the optimal input changes, i.e., the system is inherently nonstationary in spite of the stationary fading and noise processes.
3) We cannot rely on the important auxiliary result given in [9, Theorem 8] that shows that the capacity-achieving input distribution *escapes to infinity*, i.e., $X_k \to \infty$ as $\mathsf{E}_\mathrm{s} \to \infty$ almost surely.

Particularly the implicit dependence between the input and the channel noise introduces very subtle challenges. Indeed, we stumbled over this in [15], [17], [18]: While the asymptotic results given there are correct, their derivations contain a flaw that we only managed to fix very recently. Moreover, the results in [15] and [18] with respect to the feedback capacity for finite power (i.e., [15, Section 8.2.2], [18, Theorem 1]) are wrong or at least remain unproven.

Even though the derivations given in Sections V–VII turn out to be quite elaborate, by the following hand-waving argument, one can nevertheless intuitively understand why regular fading channels behave so poorly. To that goal note that since the fading process is assumed to be regular with a finite differential entropy rate, it is not possible to perfectly predict the future realizations of the process even if one is presented with the exact realizations of the infinite past. Nevertheless the feedback allows the transmitter to make an estimate of future realizations. Based on these estimates, the transmitter can then perform elaborate schemes of optimal power allocation over time: if the channel state is likely to be poor, it saves power and uses it once the channel state is likely to be good again. Unfortunately, due to the double-logarithmic behavior of capacity, such power allocation has no effect at all: for any constant $t > 0$ (t can be chosen arbitrarily large!),

$$\begin{aligned}
&\varlimsup_{\mathsf{E}_\mathrm{s}\uparrow\infty} \{\log\log(t\mathsf{E}_\mathrm{s}) - \log\log \mathsf{E}_\mathrm{s}\} \\
&\quad = \varlimsup_{\mathsf{E}_\mathrm{s}\uparrow\infty} \{\log(\log t + \log \mathsf{E}_\mathrm{s}) - \log\log \mathsf{E}_\mathrm{s}\} && (50)\\
&\quad = \varlimsup_{\mathsf{E}_\mathrm{s}\uparrow\infty} \{\log(\log \mathsf{E}_\mathrm{s}) - \log\log \mathsf{E}_\mathrm{s}\} && (51)\\
&\quad = 0. && (52)
\end{aligned}$$

So not only the double-logarithmic growth is left untouched, but also the second term, i.e., the fading number, remains unchanged.

V. A Lower Bound on the Fading Number with Receiver Side-Information

To derive a lower bound on capacity of the channel model described in Theorem 3 (i.e., without feedback or transmitter side-information), we choose a specific input distribution. This naturally yields a lower bound. Let $\{X_k\}$ be of the form

$$X_k \triangleq R_k\, e^{\mathrm{i}\Theta_k}. \tag{53}$$

Here $\{\Theta_k\}$ is a sequence of IID random variables that are uniformly distributed on $(-\pi, \pi]$. The stochastic process $\{R_k\}$ is chosen to be independent of $\{e^{\mathrm{i}\Theta_k}\}$ and to consist of random variables $R_k \in \mathbb{R}_0^+$ that are IID with

$$\log R_k^2 \sim \mathcal{U}\big([\log x_{\min}^2, \log \mathsf{E}_\mathrm{s}]\big) \tag{54}$$

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

$$x_{\min}^2 \triangleq \log \mathsf{E}_\mathrm{s}. \tag{55}$$

Note that this choice of $\{X_k\}$ satisfies the peak-power constraint (14) and therefore also the average-power constraint (13).

We now fix some (large) positive integer κ and use the chain rule and the nonnegativity of mutual information to bound:

$$\begin{aligned}
\frac{1}{n} I(X_1^n; \mathbf{Y}_1^n, \mathbf{S}_1^n) &= \frac{1}{n}\sum_{k=1}^{n} I\big(X_k; \mathbf{Y}_1^n, \mathbf{S}_1^n \big| X_1^{k-1}\big) && (56)\\
&\ge \frac{1}{n}\sum_{k=\kappa+1}^{n-\kappa} I\big(X_k; \mathbf{Y}_1^n, \mathbf{S}_1^n \big| X_1^{k-1}\big). && (57)
\end{aligned}$$

Then for every $\kappa + 1 \le k \le n - \kappa$, we can use the fact that $\{X_k\}$ is IID to lower-bound $I\big(X_k; \mathbf{Y}_1^n, \mathbf{S}_1^n \big| X_1^{k-1}\big)$ as follows:

$$\begin{aligned}
&I\big(X_k; \mathbf{Y}_1^n, \mathbf{S}_1^n \big| X_1^{k-1}\big) \\
&\quad \ge I\big(X_k; \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa} \big| X_1^{k-1}\big) && (58)\\
&\quad = I\big(X_k; \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa} \big| X_{k-\kappa}^{k-1}\big) && (59)\\
&\quad = I\big(X_k; \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}, \mathbf{Z}_{k-\kappa}^{k-1}, \mathbf{Z}_{k+1}^{k+\kappa} \big| X_{k-\kappa}^{k-1}\big) \\
&\qquad - \underbrace{I\big(X_k; \mathbf{Z}_{k-\kappa}^{k-1}, \mathbf{Z}_{k+1}^{k+\kappa} \big| \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}, X_{k-\kappa}^{k-1}\big)}_{\le \delta_1(x_{\min},\kappa)} && (60)\\
&\quad \ge I\big(X_k; \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{Z}_{k-\kappa}^{k-1}, \mathbf{Z}_{k+1}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa} \big| X_{k-\kappa}^{k-1}\big) \\
&\qquad - \delta_1(x_{\min}, \kappa) && (61)\\
&\quad = I\big(X_k; \mathbf{Y}_k, \{\mathbf{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}, \{\mathbf{H}_\ell X_\ell\}_{\ell=k+1}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa} \big| X_{k-\kappa}^{k-1}\big) \\
&\qquad - \delta_1(x_{\min}, \kappa) && (62)
\end{aligned}$$

$$= I\Big(X_k; \mathbf{Y}_k, \mathbf{H}_{k-\kappa}^{k-1}, \{\|\mathbf{H}_\ell\| R_\ell\}_{\ell=k+1}^{k+\kappa}, \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=k+1}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}\Big) - \delta_1(x_{\min},\kappa) \tag{63}$$

$$\geq I\Big(X_k; \mathbf{Y}_k, \mathbf{H}_{k-\kappa}^{k-1}, \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=k+1}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}\Big) - \delta_1(x_{\min},\kappa) \tag{64}$$

$$= I\Big(X_k; \mathbf{Y}_k \Big| \mathbf{H}_{k-\kappa}^{k-1}, \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=k+1}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}\Big) - \delta_1(x_{\min},\kappa). \tag{65}$$

Here, (58) follows from dropping some random quantities in the argument of the mutual information term; (59) holds because $\{X_k\}$ is IID and independent of fading and side-information; (60) results from the chain rule; and (61) follows from the following lemma.

Lemma 8: Let $\{X_k\}$ be as specified in (53)–(55). Then

$$I\big(X_k; \mathbf{Z}_{k-\kappa}^{k-1}, \mathbf{Z}_{k+1}^{k+\kappa} \big| \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}, X_{k-\kappa}^{k-1}\big) \leq \delta_1(x_{\min},\kappa) \tag{66}$$

where $\delta_1(x_{\min},\kappa)$ is defined in Appendix B, is independent of k and the distribution of $\{X_k\}$, and tends to zero as $x_{\min}\uparrow\infty$.

Proof: See Appendix B. ■

In (62) we firstly extract $\{\mathbf{H}_\ell X_\ell\}$ from $\{\mathbf{Y}_\ell\}$ using the noise $\{\mathbf{Z}_\ell\}$ and then drop $\{\mathbf{Y}_\ell, \mathbf{Z}_\ell\}$ because it is independent of $(X_k, \mathbf{Y}_k)$ when conditioned on $\{\mathbf{H}_\ell X_\ell\}$; in (63) we split $\mathbf{H}_\ell X_\ell$ into magnitude and direction for $\ell = k+1,\ldots,k+\kappa$, and extract $\mathbf{H}_\ell$ for $\ell = k-\kappa,\ldots,k-1$ and then drop the conditioning since it is independent of the remaining terms; in (64) we drop some arguments; and (65) follows from our choice of $\{X_k\}$ being IID.

Hence, using (65) in (57) we get

$$\frac{1}{n} I\big(X_1^n; \mathbf{Y}_1^n, \mathbf{S}_1^n\big) \geq \frac{1}{n}\sum_{k=\kappa+1}^{n-\kappa}\Big(I\Big(X_k; \mathbf{Y}_k \Big| \mathbf{H}_{k-\kappa}^{k-1}, \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=k+1}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}\Big) - \delta_1(x_{\min},\kappa)\Big) \tag{67}$$

$$= \frac{n-2\kappa}{n}\Big(I\Big(X_0; \mathbf{Y}_0 \Big| \mathbf{H}_{-\kappa}^{-1}, \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_{-\kappa}^{\kappa}\Big) - \delta_1(x_{\min},\kappa)\Big) \tag{68}$$

where (68) follows from stationarity. Letting n tend to infinity we obtain

$$\mathsf{C}(\mathsf{E}_s|\{\mathbf{S}_k\}) \geq I\Big(X_0; \mathbf{Y}_0 \Big| \mathbf{H}_{-\kappa}^{-1}, \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_{-\kappa}^{\kappa}\Big) - \delta_1(x_{\min},\kappa). \tag{69}$$

We next let the power grow to infinity $\mathsf{E}_s\uparrow\infty$ and use the definition of the fading number. Note that the distribution of X_0 (the product of (54) with the circularly symmetric law from $e^{\mathrm{i}\Theta_0}$) achieves the fading number of a memoryless SIMO fading channel with side-information [6, Proposition 4.23], [15, Proposition 6.23]. Moreover, our choice (55) guarantees that $\delta_1(x_{\min},\kappa)$ tends to zero as $\mathsf{E}_s\uparrow\infty$. Therefore, we obtain the following bound:

$$\chi\big(\{\mathbf{H}_k\}\big|\{\mathbf{S}_k\}\big) = \varlimsup_{\mathsf{E}_s\uparrow\infty}\big\{\mathsf{C}(\mathsf{E}_s|\{\mathbf{S}_k\}) - \log\log\mathsf{E}_s\big\} \tag{70}$$

$$\geq \varlimsup_{\mathsf{E}_s\uparrow\infty}\Big\{I\Big(X_0; \mathbf{Y}_0 \Big| \mathbf{H}_{-\kappa}^{-1}, \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_{-\kappa}^{\kappa}\Big) - \log\log\mathsf{E}_s - \delta_1(x_{\min},\kappa)\Big\} \tag{71}$$

$$= \chi_{\mathrm{IID}}\Big(\mathbf{H}_0 \Big| \mathbf{H}_{-\kappa}^{-1}, \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_{-\kappa}^{\kappa}\Big) \tag{72}$$

$$= h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0} \Big| \mathbf{H}_{-\kappa}^{-1}, \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_{-\kappa}^{\kappa}\Big) - \log 2 + n_{\mathsf{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big] - h\Big(\mathbf{H}_0 \Big| \mathbf{H}_{-\kappa}^{-1}, \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_{-\kappa}^{\kappa}\Big). \tag{73}$$

Here, in (73) we use the expression for the memoryless SIMO fading number with side-information [6, Equation (108)]:

$$\chi_{\mathrm{IID}}(\mathbf{H}|\mathbf{S}) = h_\lambda\big(\hat{\mathbf{H}}\, e^{\mathrm{i}\Theta}\big|\mathbf{S}\big) - \log 2 + n_{\mathsf{R}}\,\mathsf{E}\big[\log\|\mathbf{H}\|^2\big] - h(\mathbf{H}|\mathbf{S}). \tag{74}$$

Next, by using the definition of mutual information and by repeatedly applying the chain rule, we note that

$$h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0} \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{H}_{-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{\kappa}\Big)$$

$$= h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0} \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_0^{\kappa}\Big) - I\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}; \mathbf{H}_{-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{-1} \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_0^{\kappa}\Big) \tag{75}$$

$$= h_\lambda\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=0}^{\kappa} \Big| \mathbf{S}_0^{\kappa}\Big) - h_\lambda\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa} \Big| \mathbf{S}_0^{\kappa}\Big) - I\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}; \mathbf{H}_{-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{-1} \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_0^{\kappa}\Big) \pm I\Big(\mathbf{S}_0; \mathbf{H}_{-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{-1} \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_1^{\kappa}\Big) \tag{76}$$

$$= h_\lambda\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=0}^{\kappa}\Big) - I\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=0}^{\kappa}; \mathbf{S}_0^{\kappa}\Big) - h_\lambda\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}\Big) + I\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}; \mathbf{S}_0^{\kappa}\Big) - I\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}, \mathbf{S}_0; \mathbf{H}_{-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{-1} \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_1^{\kappa}\Big) + h\Big(\mathbf{S}_0 \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_1^{\kappa}\Big) - h\Big(\mathbf{S}_0 \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_1^{\kappa}, \mathbf{H}_{-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{-1}\Big) \tag{77}$$

$$= h_\lambda\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=0}^{\kappa}\Big) - h\big(\mathbf{S}_0^{\kappa}\big) + h\Big(\mathbf{S}_0^{\kappa} \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=0}^{\kappa}\Big) - h_\lambda\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}\Big) + h\big(\mathbf{S}_0^{\kappa}\big) - h\Big(\mathbf{S}_0^{\kappa} \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}\Big) - I\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=0}^{\kappa}, \mathbf{S}_0^{\kappa}; \mathbf{H}_{-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{-1}\Big) + I\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_1^{\kappa}; \mathbf{H}_{-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{-1}\Big) + h\Big(\mathbf{S}_0^{\kappa} \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}\Big) - h\Big(\mathbf{S}_1^{\kappa} \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}\Big) - h\Big(\mathbf{S}_0 \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_1^{\kappa}, \mathbf{H}_{-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{-1}\Big) \tag{78}$$

$$= h_\lambda\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\kappa}^{0}\Big) + h\Big(\mathbf{S}_{-\kappa}^{0} \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\kappa}^{0}\Big) - h_\lambda\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\kappa}^{-1}\Big) - I\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=0}^{\kappa}, \mathbf{S}_0^{\kappa}; \mathbf{H}_{-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{-1}\Big) + I\Big(\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_1^{\kappa}; \mathbf{H}_{-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{-1}\Big) - h\Big(\mathbf{S}_{-\kappa}^{-1} \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\kappa}^{-1}\Big) - h\Big(\mathbf{S}_0 \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=1}^{\kappa}, \mathbf{S}_1^{\kappa}, \mathbf{H}_{-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{-1}\Big)$$

$$\begin{aligned}
&\pm I\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}; \mathbf{S}^0_{-\kappa} \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}\Big) &(79)\\
&= h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0} \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}\Big)\\
&\quad + h\Big(\mathbf{S}^0_{-\kappa} \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{0}\Big)\\
&\quad - I\Big(\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=0}^{\kappa}, \mathbf{S}_0^\kappa; \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\quad + I\Big(\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa; \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\quad - h\Big(\mathbf{S}^{-1}_{-\kappa} \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}\Big)\\
&\quad - h\Big(\mathbf{S}_0 \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa, \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\quad + h\Big(\mathbf{S}^0_{-\kappa} \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}\Big) - h\Big(\mathbf{S}^0_{-\kappa} \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{0}\Big)\\
&\quad - h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0} \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}\Big)\\
&\quad + h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0} \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}^0_{-\kappa}\Big) &(80)\\
&= -I\Big(\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=0}^{\kappa}, \mathbf{S}_0^\kappa; \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\quad + I\Big(\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa; \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\quad - h\Big(\mathbf{S}_0 \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa, \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\quad + h\Big(\mathbf{S}_0 \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\quad + h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0} \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}^0_{-\kappa}\Big). &(81)
\end{aligned}$$

Here, in (76) we add and subtract the same term; and (79) follows by stationarity and again by adding and subtracting the same term.

Furthermore,

$$\begin{aligned}
&-h\Big(\mathbf{H}_0 \Big| \mathbf{H}^{-1}_{-\kappa}, \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}^\kappa_{-\kappa}\Big)\\
&\quad = -h\Big(\mathbf{H}_0, \mathbf{S}_0 \Big| \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}, \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa\Big)\\
&\qquad + h\Big(\mathbf{S}_0 \Big| \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}, \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa\Big) &(82)\\
&\quad = -h\big(\mathbf{H}_0, \mathbf{S}_0 \big| \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\big)\\
&\qquad + I\Big(\mathbf{H}_0, \mathbf{S}_0; \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa \Big| \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\qquad + h\Big(\mathbf{S}_0 \Big| \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}, \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa\Big) &(83)\\
&\quad = -h\big(\mathbf{H}_0, \mathbf{S}_0 \big| \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\big)\\
&\qquad + I\Big(\mathbf{H}^0_{-\kappa}, \mathbf{S}^0_{-\kappa}; \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa\Big)\\
&\qquad - I\Big(\mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}; \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa\Big)\\
&\qquad + h\Big(\mathbf{S}_0 \Big| \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}, \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa\Big). &(84)
\end{aligned}$$

Plugging (84) and (81) into (73) we hence obtain

$$\begin{aligned}
&\chi(\{\mathbf{H}_k\}|\{\mathbf{S}_k\})\\
&\quad \geq -I\Big(\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=0}^{\kappa}, \mathbf{S}_0^\kappa; \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\qquad + I\Big(\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa; \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\qquad - h\Big(\mathbf{S}_0 \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa, \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\qquad + h\Big(\mathbf{S}_0 \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\qquad + h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0} \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}^0_{-\kappa}\Big)\\
&\qquad - \log 2 + n_{\mathsf{R}}\, \mathsf{E}\big[\log \|\mathbf{H}_0\|^2\big] - h\big(\mathbf{H}_0, \mathbf{S}_0 \big| \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\big)\\
&\qquad + I\Big(\mathbf{H}^0_{-\kappa}, \mathbf{S}^0_{-\kappa}; \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa\Big)\\
&\qquad - I\Big(\mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}; \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa\Big)\\
&\qquad + h\Big(\mathbf{S}_0 \Big| \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}, \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa\Big) &(85)\\
&\quad = h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0} \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}^0_{-\kappa}\Big)\\
&\qquad - \log 2 + n_{\mathsf{R}}\, \mathsf{E}\big[\log \|\mathbf{H}_0\|^2\big] - h\big(\mathbf{H}_0 \big| \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{0}_{-\kappa}\big)\\
&\qquad + h\Big(\mathbf{S}_0 \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}^{-1}_{-\kappa}\Big) - h\big(\mathbf{S}_0 \big| \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\big)\\
&\qquad + I\Big(\mathbf{H}^0_{-\kappa}, \mathbf{S}^0_{-\kappa}; \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=1}^{\kappa}, \mathbf{S}_1^\kappa\Big)\\
&\qquad - I\Big(\mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}; \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=0}^{\kappa}, \mathbf{S}_0^\kappa\Big). &(86)
\end{aligned}$$

We next point out that

$$\begin{aligned}
&h\Big(\mathbf{S}_0 \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}^{-1}_{-\kappa}\Big) - h\big(\mathbf{S}_0 \big| \mathbf{H}^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\big)\\
&\quad = h\Big(\mathbf{S}_0 \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\qquad - h\big(\mathbf{S}_0 \big| \{\mathbf{H}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\kappa}^{-1}, \Theta^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\big) &(87)\\
&\quad = h\Big(\mathbf{S}_0 \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}^{-1}_{-\kappa}\Big)\\
&\qquad - h\Big(\mathbf{S}_0 \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\kappa}^{-1}, \{\|\mathbf{H}_\ell\|\}_{\ell=-\kappa}^{-1}, \Theta^{-1}_{-\kappa}, \mathbf{S}^{-1}_{-\kappa}\Big) &(88)\\
&\quad = I\Big(\mathbf{S}_0; \{\|\mathbf{H}_\ell\|\}_{\ell=-\kappa}^{-1}, \Theta^{-1}_{-\kappa} \Big| \{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\kappa}^{-1}, \mathbf{S}^{-1}_{-\kappa}\Big) &(89)\\
&\quad \geq 0. &(90)
\end{aligned}$$

Note that, using the tool from Appendix A, one can show that the inequality actually holds with equality.[4] Finally, we let κ go to infinity. Stationarity of the input and the fading process now makes sure that the two mutual information terms in (86) cancel, and we obtain

$$\begin{aligned}
&\chi(\{\mathbf{H}_k\}|\{\mathbf{S}_k\})\\
&\quad \geq h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0} \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\infty}^{-1}, \mathbf{S}^0_{-\infty}\Big) - \log 2\\
&\qquad + n_{\mathsf{R}}\, \mathsf{E}\big[\log \|\mathbf{H}_0\|^2\big] - h\big(\mathbf{H}_0 \big| \mathbf{H}^{-1}_{-\infty}, \mathbf{S}^0_{-\infty}\big). &(91)
\end{aligned}$$

VI. An Upper Bound on the Fading Number With Feedback and Causal Side-Information

A. Overview

While the basic structure of the following derivation is relatively straightforward, there are many subtle details that need to be taken care of and that complicate the proof considerably. We therefore try to give a rough outline of the proof first.

We start with the standard approach for deriving a converse using Fano's inequality:

$$\mathsf{R}_{\mathrm{FB,c}}(\mathsf{E}_{\mathrm{s}}|\{\mathbf{S}_k\}) \leq \frac{1}{n}\sum_{k=1}^{n} I\big(M; \mathbf{Y}_k, \mathbf{S}_k \big| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}\big) + \tilde{\epsilon}_n. \tag{92}$$

[4] This also can be argued indirectly: Since in Section VI we derive an *upper* bound on the fading number without this term, it follows that the term must be zero.

We then would like to split the mutual information term into three parts:

$$I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\big) \leq I\big(X_k;\mathbf{Y}_k\big|\mathbf{S}_1^k\big)+I\big(\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|X_k,\mathbf{S}_1^k\big) - I\big(\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k\big) \tag{93}$$

where the first term basically corresponds to the memoryless SIMO fading channel without feedback, and where the other two terms are correction terms taking care of the memory. While indeed these three terms are bounded separately in Sections VI-D to VI-F and then combined in Section VI-G to the final result, there are many obstacles on the way that need to be circumvented.

First of all, note that for small k, the term $I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\big)$ goes through a transitional phase because for small k the transmitter has only a very limited knowledge of past fading realizations. Only for large k the transmitter can properly rely on the statistical knowledge from the feedback. In order to handle this transition phase, we split the sum in (92) up into two parts:

$$\frac{1}{n}\sum_{k=1}^{n} I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\big) = \frac{1}{n}\sum_{k=1}^{\kappa} I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\big) + \frac{1}{n}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\big). \tag{94}$$

The first part will then be bounded very roughly with the only aim to make sure that it will disappear once we let n tend to infinity. So we can focus on the second sum.

Then, as we are interested in the asymptotic capacity, we would like to think that an optimal input satisfies $X_k\to\infty$ as $\mathsf{E}_\mathrm{s}\to\infty$. Unfortunately, while it is possible to use stationarity of the channel model to prove such a result for the capacity-achieving input of a channel without feedback, in the case with feedback, the system is inherently nonstationary because the knowledge at the transmitter grows at every time step and therefore changes the optimal input. To solve this dilemma, we introduce a case distinction on whether $|X_k|$ is larger or smaller than some given threshold $\xi_{\min}$, with β_k denoting the probability of the former case. In the derivations below, these two cases are expressed by the indicator random variable B_k as $B_k=1$ or $B_k=0$, respectively. We then need to prove that as E_s becomes large, it is optimal to have $\beta_k\to 1$. Note that we have to take particular care here to make sure that we only start to twiddle with E_s once we have loosened $n\to\infty$. For that reason we make an effort in deriving rough bounds (that are independent of k) for terms that have a factor $1-\beta_k$ in front (i.e., terms that will disappear anyway once we prove that $\beta_k\to 1$).

Finally, one needs to be aware that via the feedback and the side-information, the current channel input X_k depends on the past fading and therefore also on the current fading. So, for example, the expression $\mathsf{E}\big[\|\mathbf{H}_k\|^2|X_k|^2\big]$ that in the case without feedback can be evaluated trivially as

$$\mathsf{E}\big[\|\mathbf{H}_k\|^2|X_k|^2\big]=\mathsf{E}\big[\|\mathbf{H}_k\|^2\big]\,\mathsf{E}\big[|X_k|^2\big] \tag{95}$$

becomes an unsolvable problem as the exact dependence of the unknown optimal input distribution on $\{\mathbf{H}_k\}$ is intractable. Purely because of this expression, we need to introduce a second case distinction on whether $\|\mathbf{H}_k\|^2$ is larger or smaller than some chosen threshold t. In the derivations below, these two cases are expressed by the indicator random variable A_k as $A_k=1$ or $A_k=0$, respectively. The case of $\|\mathbf{H}_k\|^2\geq t$ is then bounded very roughly with the only aim to make sure that all terms belonging to this case will disappear once we let t tend to infinity towards the end of the derivation. For the case of $\|\mathbf{H}_k\|^2<t$, we can then bound

$$\mathsf{E}\big[\|\mathbf{H}_k\|^2|X_k|^2\big|A_k=0\big]<t\,\mathsf{E}\big[|X_k|^2\big|A_k=0\big]. \tag{96}$$

There are many other places where one has to be very careful with dependencies. For example, a conditioning on $B_k=1$ cannot simply be dropped even if the expression only involves the fading process because the fading process depends via the feedback on the input and therefore also on B_k.

In many situations, we rely on a graphical tool described by Massey [11], [12] that allows to figure out whether two sets of random variables are independent of each other when conditioned on some more random variables. Note that the dependencies are often so subtle that it is essential to have a graphical proof, rather than using hand-waving explanations and engineering intuition. All these independence investigations are presented in Appendix A.

B. A Useful Inequality

In the following we will often make use of the following inequality.

Lemma 9: Let $T\geq 0$ be a nonnegative RV and let $J\in\{0,1\}$ be a related binary indicator RV with $\Pr[J=0]=p$. Then

$$\mathsf{E}[T]\geq p\,\mathsf{E}[T|J=0]. \tag{97}$$

Proof: By the rule of total expectation we have

$$\mathsf{E}[T]=p\,\mathsf{E}[T|J=0]+(1-p)\,\mathsf{E}[T|J=1] \tag{98}$$

$$\geq p\,\mathsf{E}[T|J=0] \tag{99}$$

where the inequality holds because T is nonnegative. ■

C. Setup

We start by assuming that there exists a sequence of coding schemes with $\big\lfloor e^{n\mathsf{R}_{\mathrm{FB,c}}(\mathsf{E}_\mathrm{s}|\{\mathbf{S}_k\})}\big\rfloor$ codewords of blocklength n — i.e., for each n the rate of the code is not larger than $\mathsf{R}_{\mathrm{FB,c}}(\mathsf{E}_\mathrm{s}|\{\mathbf{S}_k\})$ — that all satisfy the average-power constraint (13) such that the error probability $\Pr[M\neq\hat{M}]$ tends to zero as n tends to infinity. Then

$$\mathsf{H}(M)=\log\big\lfloor e^{n\mathsf{R}_{\mathrm{FB,c}}(\mathsf{E}_\mathrm{s}|\{\mathbf{S}_k\})}\big\rfloor \tag{100}$$

$$\geq\log\big(e^{n\mathsf{R}_{\mathrm{FB,c}}(\mathsf{E}_\mathrm{s}|\{\mathbf{S}_k\})}-1\big) \tag{101}$$

$$=n\mathsf{R}_{\mathrm{FB,c}}(\mathsf{E}_\mathrm{s}|\{\mathbf{S}_k\})-\epsilon_n \tag{102}$$

with $\epsilon_n \downarrow 0$ as $n \uparrow \infty$. Next recall Fano's inequality [19, Section 9.6]: Let M take on $|\mathcal{M}|$ values. Then

$$\mathsf{H}\bigl(M\big|\hat{M}\bigr) \leq \log 2 + \Pr\bigl[M \neq \hat{M}\bigr] \log |\mathcal{M}|. \tag{103}$$

We therefore have

$$\begin{aligned}
&\mathsf{R}_{\text{FB,c}}(\mathsf{E}_{\text{s}}|\{\mathbf{S}_k\}) \\
&\leq \frac{1}{n}\mathsf{H}(M) + \frac{\epsilon_n}{n} && (104)\\
&= \frac{1}{n}I\bigl(M;\hat{M}\bigr) + \frac{1}{n}\mathsf{H}\bigl(M\big|\hat{M}\bigr) + \frac{\epsilon_n}{n} && (105)\\
&\leq \frac{1}{n}I\bigl(M;\hat{M}\bigr) \\
&\quad + \frac{\log 2 + \Pr\bigl[M \neq \hat{M}\bigr]\log\bigl\lfloor e^{n\mathsf{R}_{\text{FB,c}}(\mathsf{E}_{\text{s}}|\{\mathbf{S}_k\})}\bigr\rfloor}{n} + \frac{\epsilon_n}{n} && (106)\\
&\leq \frac{1}{n}I\bigl(M;\hat{M}\bigr) + \frac{\log 2}{n} + \Pr\bigl[M \neq \hat{M}\bigr]\,\mathsf{R}_{\text{FB,c}}(\mathsf{E}_{\text{s}}|\{\mathbf{S}_k\}) \\
&\quad + \frac{\epsilon_n}{n} && (107)\\
&\leq \frac{1}{n}I\bigl(M;\mathbf{Y}_1^n,\mathbf{S}_1^n\bigr) + \frac{\log 2}{n} \\
&\quad + \Pr\bigl[M \neq \hat{M}\bigr]\,\mathsf{R}_{\text{FB,c}}(\mathsf{E}_{\text{s}}|\{\mathbf{S}_k\}) + \frac{\epsilon_n}{n} && (108)\\
&= \frac{1}{n}\sum_{k=1}^{n} I\bigl(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\bigr) + \frac{\log 2}{n} \\
&\quad + \Pr\bigl[M \neq \hat{M}\bigr]\,\mathsf{R}_{\text{FB,c}}(\mathsf{E}_{\text{s}}|\{\mathbf{S}_k\}) + \frac{\epsilon_n}{n} && (109)\\
&= \frac{\kappa}{n}\,\frac{1}{\kappa}\sum_{k=1}^{\kappa} I\bigl(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\bigr) \\
&\quad + \frac{1}{n}\sum_{k=\kappa+1}^{n} I\bigl(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\bigr) + \frac{\log 2}{n} \\
&\quad + \Pr\bigl[M \neq \hat{M}\bigr]\,\mathsf{R}_{\text{FB,c}}(\mathsf{E}_{\text{s}}|\{\mathbf{S}_k\}) + \frac{\epsilon_n}{n}. && (110)
\end{aligned}$$

Here, (104) follows from (102); (106) follows from (103); and in (108) we apply the data processing inequality [19, Section 9.6].

We next need the following lemma.

Lemma 10: For the channel model as given in Theorem 4,

$$\begin{aligned}
&I\bigl(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\bigr) \\
&\quad\leq n_{\text{R}}\log\left(1 + \frac{1}{n_{\text{R}}}\,\mathsf{E}\bigl[\|\mathbf{H}_0\|^2\bigr]\,\frac{\mathsf{E}_k}{\sigma^2}\right) \\
&\qquad + I\bigl(\mathbf{H}_1^{k-1},\mathbf{S}_1^{k-1};\mathbf{H}_k,\mathbf{S}_k\bigr) && (111)
\end{aligned}$$

where E_k denotes the average power of X_k at time k, averaged over all realizations of message, feedback, and side-information, i.e.,

$$\mathsf{E}_k \triangleq \mathsf{E}\Bigl[\bigl|X_k\bigl(M,\mathbf{S}_1^k,\mathbf{F}_1^k\bigr)\bigr|^2\Bigr]. \tag{112}$$

Proof: See Appendix C. ∎

We apply (111) to the first sum in (110), and use Jensen's inequality and stationarity to obtain:

$$\begin{aligned}
&\frac{1}{\kappa}\sum_{k=1}^{\kappa} I\bigl(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\bigr) \\
&\quad\leq \frac{1}{\kappa}\sum_{k=1}^{\kappa} n_{\text{R}}\log\left(1 + \frac{1}{n_{\text{R}}}\,\mathsf{E}\bigl[\|\mathbf{H}_k\|^2\bigr]\,\frac{\mathsf{E}_k}{\sigma^2}\right) \\
&\qquad + \frac{1}{\kappa}\sum_{k=1}^{\kappa} h(\mathbf{H}_k,\mathbf{S}_k) \\
&\qquad - \frac{1}{\kappa}\sum_{k=1}^{\kappa} h\bigl(\mathbf{H}_k,\mathbf{S}_k\big|\mathbf{H}_1^{k-1},\mathbf{S}_1^{k-1}\bigr) && (113)\\
&\quad\leq n_{\text{R}}\log\left(1 + \frac{1}{n_{\text{R}}}\,\mathsf{E}\bigl[\|\mathbf{H}_k\|^2\bigr]\frac{1}{\kappa}\sum_{k=1}^{\kappa}\frac{\mathsf{E}_k}{\sigma^2}\right) + h(\mathbf{H}_0,\mathbf{S}_0) \\
&\qquad - \frac{1}{\kappa}h\bigl(\mathbf{H}_1^{\kappa},\mathbf{S}_1^{\kappa}\bigr). && (114)
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathsf{R}_{\text{FB,c}}(\mathsf{E}_{\text{s}}|\{\mathbf{S}_k\}) \\
&\leq \frac{\kappa}{n}\,n_{\text{R}}\log\left(1 + \frac{1}{n_{\text{R}}}\,\mathsf{E}\bigl[\|\mathbf{H}_0\|^2\bigr]\,\frac{1}{\kappa}\sum_{k=1}^{\kappa}\frac{\mathsf{E}_k}{\sigma^2}\right) + \frac{\kappa}{n}\,h(\mathbf{H}_0,\mathbf{S}_0) \\
&\quad - \frac{1}{n}\,h(\mathbf{H}_1^{\kappa},\mathbf{S}_1^{\kappa}) + \frac{\log 2}{n} + \Pr\bigl[M \neq \hat{M}\bigr]\,\mathsf{R}_{\text{FB,c}}(\mathsf{E}_{\text{s}}|\{\mathbf{S}_k\}) \\
&\quad + \frac{\epsilon_n}{n} + \frac{1}{n}\sum_{k=\kappa+1}^{n} I\bigl(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\bigr). && (115)
\end{aligned}$$

Note that apart from the last term, all terms will tend to zero as n tends to infinity. We therefore henceforth concentrate on the terms inside of the sum in (115), i.e., we look at $I\bigl(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\bigr)$ for $\kappa+1 \leq k \leq n$.

We introduce the indicator random variables A_k defined as

$$A_k \triangleq \begin{cases} 1 & \text{if } \|\mathbf{H}_k\|^2 \geq t \\ 0 & \text{otherwise} \end{cases} \tag{116}$$

for some given $t > 0$ that will be specified later. Moreover, we define

$$\alpha_k \triangleq \Pr[A_k = 1] = \Pr\bigl[\|\mathbf{H}_k\|^2 \geq t\bigr]. \tag{117}$$

It follows from Markov's inequality [20, Section 5] that

$$\alpha_k \leq \frac{\mathsf{E}\bigl[\|\mathbf{H}_k\|^2\bigr]}{t} = \frac{\mathsf{E}\bigl[\|\mathbf{H}_0\|^2\bigr]}{t} \tag{118}$$

and therefore (by conditioning that reduces entropy)

$$\begin{aligned}
\mathsf{H}\bigl(A_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\bigr) &\leq \mathsf{H}(A_k) && (119)\\
&= \mathsf{H}_{\text{b}}(\alpha_k) && (120)\\
&\leq \mathsf{H}_{\text{b}}\left(\frac{\mathsf{E}\bigl[\|\mathbf{H}_0\|^2\bigr]}{t}\right) && (121)
\end{aligned}$$

where $\mathsf{H}_{\text{b}}(\cdot)$ denotes the binary entropy function (18) and where we choose t large enough such that

$$\frac{\mathsf{E}\bigl[\|\mathbf{H}_0\|^2\bigr]}{t} \leq \frac{1}{2}. \tag{122}$$

We now bound as follows:

$$\begin{aligned}
&I\bigl(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\bigr) \\
&\leq I\bigl(M;\mathbf{Y}_k,\mathbf{S}_k,A_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\bigr) && (123)\\
&= I\bigl(M;A_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\bigr) \\
&\quad + I\bigl(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},A_k\bigr) && (124)\\
&= \mathsf{H}\bigl(A_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\bigr) - \underbrace{\mathsf{H}\bigl(A_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},M\bigr)}_{\geq 0}
\end{aligned}$$

$$
\begin{aligned}
&+ \alpha_k\, I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},A_k=1\big)\\
&+ \underbrace{(1-\alpha_k)}_{\leq 1} I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},A_k=0\big) \quad (125)\\
\leq{}& \mathsf{H}\big(A_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\big)\\
&+ \alpha_k n_\mathsf{R}\log\left(1+\frac{1}{n_\mathsf{R}}\,\mathsf{E}\left[\|\mathbf{H}_0\|^2\middle|A_0=1\right]\,\frac{\mathsf{E}_k}{\alpha_k\,\sigma^2}\right)\\
&+ \alpha_k I\big(\mathbf{H}_1^{k-1},\mathbf{S}_1^{k-1};\mathbf{H}_k,\mathbf{S}_k\big|A_k=1\big)\\
&+ I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},A_k=0\big) \quad (126)\\
\leq{}& \mathsf{H}_\mathsf{b}\left(\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\right)\\
&+ \frac{n_\mathsf{R}\,\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\log\left(1+\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\middle|A_0=1\right]}{n_\mathsf{R}\,\mathsf{E}[\|\mathbf{H}_0\|^2]}\,\frac{t\mathsf{E}_k}{\sigma^2}\right)\\
&+ \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\, I\big(\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^{-1};\mathbf{H}_0,\mathbf{S}_0\big|A_0=1\big)\\
&+ I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},A_k=0\big). \quad (127)
\end{aligned}
$$

Here (126) follows from a conditional version of (111) of Lemma 10 with E_k replaced by E_k/α_k because

$$
\begin{aligned}
&\mathsf{E}\left[\left|X_k\big(M,\mathbf{S}_1^k,\mathbf{F}_1^k\big)\right|^2\middle|A_k=1\right]\\
&\quad\leq \frac{1}{\alpha_k}\,\mathsf{E}\left[\left|X_k\big(M,\mathbf{S}_1^k,\mathbf{F}_1^k\big)\right|^2\right] \quad (128)\\
&\quad= \frac{\mathsf{E}_k}{\alpha_k} \quad (129)
\end{aligned}
$$

(see Lemma 9); and (127) follows from (121), from (118) and the monotonicity of $\alpha_k \mapsto \alpha_k\log(1+\text{const}/\alpha_k)$, and from stationarity combined with additional random variables in the argument of mutual information.

Hence,

$$
\begin{aligned}
&\frac{1}{n}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1}\big)\\
&\leq \frac{1}{n}\sum_{k=1}^{n}\mathsf{H}_\mathsf{b}\left(\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\right)\\
&\quad+ \frac{n_\mathsf{R}\,\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t\cdot n}\sum_{k=1}^{n}\log\left(1+\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\middle|A_0=1\right]}{n_\mathsf{R}\,\mathsf{E}[\|\mathbf{H}_0\|^2]}\frac{t\mathsf{E}_k}{\sigma^2}\right)\\
&\quad+ \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\,\frac{1}{n}\sum_{k=1}^{n} I\big(\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^{-1};\mathbf{H}_0,\mathbf{S}_0\big|A_0=1\big)\\
&\quad+ \frac{1}{n}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},A_k=0\big) \quad (130)\\
&\leq \mathsf{H}_\mathsf{b}\left(\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\right)\\
&\quad+ \frac{n_\mathsf{R}\,\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\log\left(1+\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\middle|A_0=1\right]}{n_\mathsf{R}\,\mathsf{E}[\|\mathbf{H}_0\|^2]}\,\frac{t\mathsf{E}_\mathsf{s}}{\sigma^2}\right)\\
&\quad+ \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\, I\big(\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^{-1};\mathbf{H}_0,\mathbf{S}_0\big|A_0=1\big)\\
&\quad+ \frac{1}{n}\,\frac{n-\kappa}{n-\kappa}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},A_k=0\big)
\end{aligned}
\quad (131)
$$

where we have added some nonnegative terms to some of the sums, used Jensen's inequality, and relied on the average-power constraint (13) that guarantees that

$$
\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}_k \leq \mathsf{E}_\mathsf{s}. \quad (132)
$$

Hence, combining (131) with (115), we have

$$
\begin{aligned}
&\mathsf{R}_{\mathsf{FB,c}}(\mathsf{E}_\mathsf{s}|\{\mathbf{S}_k\})\\
&\leq \frac{\kappa}{n}\,n_\mathsf{R}\log\left(1+\frac{1}{n_\mathsf{R}}\,\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]\,\frac{1}{\kappa}\sum_{k=1}^{\kappa}\frac{\mathsf{E}_k}{\sigma^2}\right)+\frac{\kappa}{n}\,h(\mathbf{H}_0,\mathbf{S}_0)\\
&\quad- \frac{1}{n}\,h\big(\mathbf{H}_1^\kappa,\mathbf{S}_1^\kappa\big)+\frac{\log 2}{n}+\Pr\big[M\neq\hat{M}\big]\,\mathsf{R}_{\mathsf{FB,c}}(\mathsf{E}_\mathsf{s}|\{\mathbf{S}_k\})\\
&\quad+ \frac{\epsilon_n}{n}+\mathsf{H}_\mathsf{b}\left(\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\right)\\
&\quad+ \frac{n_\mathsf{R}\,\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\log\left(1+\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\middle|A_0=1\right]}{n_\mathsf{R}\,\mathsf{E}[\|\mathbf{H}_0\|^2]}\,\frac{t\mathsf{E}_\mathsf{s}}{\sigma^2}\right)\\
&\quad+ \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\, I\big(\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^{-1};\mathbf{H}_0,\mathbf{S}_0\big|A_0=1\big)\\
&\quad+ \frac{1}{n}\,\frac{n-\kappa}{n-\kappa}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},A_k=0\big).
\end{aligned}
\quad (133)
$$

Next we introduce a second family of indicator random variables. For some $\xi_{\min} > 0$, we define

$$
B_k \triangleq \begin{cases} 1 & \text{if } |X_\ell| \geq \xi_{\min}, \quad \forall\,\ell=1,\ldots,k\\ 0 & \text{otherwise}\end{cases} \quad (134)
$$

and

$$
\beta_k \triangleq \Pr[B_k=1|A_k=0]. \quad (135)
$$

Then

$$
\begin{aligned}
&I\big(B_k;\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0\big)\\
&\quad= \mathsf{H}\big(B_k\big|\mathbf{S}_1^k,A_k=0\big)-\mathsf{H}\big(B_k\big|\mathbf{Y}_k,\mathbf{S}_1^k,A_k=0\big) \quad (136)\\
&\quad\leq \mathsf{H}(B_k|A_k=0) \quad (137)\\
&\quad= \mathsf{H}_\mathsf{b}(\beta_k). \quad (138)
\end{aligned}
$$

Note that in the situation without feedback, it has been shown in [9] that asymptotically for $\mathsf{E}_\mathsf{s}\uparrow\infty$ the probability β_k tends to 1. We cannot use this result here due to the feedback. It will turn out, however, that the result still holds.

We now bound each term in the sum in (133) as follows:

$$
\begin{aligned}
&I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},A_k=0\big)\\
&= \underbrace{I\big(M;\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},A_k=0\big)}_{=0,\text{ see Fig. 3 in Appendix A}}\\
&\quad+ I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k},A_k=0\big) \quad (139)\\
&= I\big(M,\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0\big)\\
&\quad- I\big(\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0\big) \quad (140)\\
&\leq I\big(M,\mathbf{Y}_1^{k-1},X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0\big)\\
&\quad- I\big(\mathbf{Y}_{k-\kappa}^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0\big) \quad (141)\\
&= I\big(X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0\big)
\end{aligned}
$$

$$\begin{aligned}&+\underbrace{I\big(M,\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big|X_k,\mathbf{H}_1^{k-1},\mathbf{S}_1^k,A_k=0\big)}_{=0,\text{ see Fig. 4 in Appendix A}}\\&-I\big(\mathbf{Y}_{k-\kappa}^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0\big) &(142)\\ \le\,& I\big(X_k,B_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0\big)\\&-I\big(\mathbf{Y}_{k-\kappa}^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0\big) &(143)\\ =\,& I\big(B_k;\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0\big)\\&+\beta_k\,I\big(X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0,B_k=1\big)\\&+(1-\beta_k)\,I\big(X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0,B_k=0\big)\\&-I\big(\mathbf{Y}_{k-\kappa}^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0\big) &(144)\\ \le\,&\beta_k\,I\big(X_k;\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0,B_k=1\big)\\&+\beta_k\,I\big(\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|X_k,\mathbf{S}_1^k,A_k=0,B_k=1\big)\\&-I\big(\mathbf{Y}_{k-\kappa}^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0\big)+\mathsf{H}_\mathsf{b}(\beta_k)\\&+(1-\beta_k)\,I\big(X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0,B_k=0\big) &(145)\end{aligned}$$

where in the last inequality we used (138).

We next investigate each of the first three terms in (145) separately. As a shorthand, we introduce the event

$$\mathcal{V}\triangleq\{A_k=0,B_k=1\}. \tag{146}$$

D. Bound on $I\big(X_k;\mathbf{Y}_k\big|\mathbf{S}_1^k,\mathcal{V}\big)$

Using the notation $X_k\triangleq|X_k|\,e^{\mathrm{i}\Phi_k}$ and introducing IID random variables $\{\Theta_k\}$ that are uniformly distributed and independent of any other random variables, we bound the first term as follows:

$$\begin{aligned}&I\big(X_k;\mathbf{Y}_k\big|\mathbf{S}_1^k,\mathcal{V}\big)\\&\le I\big(X_k;\mathbf{Y}_k,\mathbf{H}_kX_k\big|\mathbf{S}_1^k,\mathcal{V}\big) &(147)\\&=I\big(X_k;\mathbf{H}_kX_k\big|\mathbf{S}_1^k,\mathcal{V}\big)+I\big(X_k;\mathbf{Y}_k\big|\mathbf{H}_kX_k,\mathbf{S}_1^k,\mathcal{V}\big) &(148)\\&=I\big(X_k;\|\mathbf{H}_k\||X_k|,\hat{\mathbf{H}}_k\,e^{\mathrm{i}\Phi_k}\big|\mathbf{S}_1^k,\mathcal{V}\big)\\&\quad+\underbrace{I\big(X_k;\mathbf{Z}_k\big|\mathbf{H}_kX_k,\mathbf{S}_1^k,\mathcal{V}\big)}_{=0} &(149)\\&=I\big(X_k;\|\mathbf{H}_k\||X_k|,\hat{\mathbf{H}}_k\,e^{\mathrm{i}\Phi_k},\Theta_k\big|\mathbf{S}_1^k,\mathcal{V}\big) &(150)\\&=I\big(X_k;\|\mathbf{H}_k\||X_k|,\hat{\mathbf{H}}_k\,e^{\mathrm{i}(\Phi_k+\Theta_k)},\Theta_k\big|\mathbf{S}_1^k,\mathcal{V}\big) &(151)\\&=I\big(X_k;\|\mathbf{H}_k\||X_k|,\Theta_k\big|\mathbf{S}_1^k,\mathcal{V}\big)\\&\quad+I\big(X_k;\hat{\mathbf{H}}_k\,e^{\mathrm{i}(\Phi_k+\Theta_k)}\big|\|\mathbf{H}_k\||X_k|,\Theta_k,\mathbf{S}_1^k,\mathcal{V}\big) &(152)\\&=I\big(X_k;\|\mathbf{H}_k\||X_k|\,e^{\mathrm{i}\Theta_k}\big|\mathbf{S}_1^k,\mathcal{V}\big)\\&\quad+h_\lambda\big(\hat{\mathbf{H}}_k\,e^{\mathrm{i}(\Phi_k+\Theta_k)}\big|\|\mathbf{H}_k\||X_k|,\Theta_k,\mathbf{S}_1^k,\mathcal{V}\big)\\&\quad-h_\lambda\big(\hat{\mathbf{H}}_k\,e^{\mathrm{i}(\Phi_k+\Theta_k)}\big|\|\mathbf{H}_k\||X_k|,\Theta_k,X_k,\mathbf{S}_1^k,\mathcal{V}\big) &(153)\\&\le I\big(X_k;\|\mathbf{H}_k\||X_k|\,e^{\mathrm{i}\Theta_k}\big|\mathbf{S}_1^k,\mathcal{V}\big)\\&\quad+h_\lambda\big(\hat{\mathbf{H}}_k\,e^{\mathrm{i}(\Phi_k+\Theta_k)}\big|\mathbf{S}_1^k,\mathcal{V}\big)\\&\quad-h_\lambda\big(\hat{\mathbf{H}}_k\big|\|\mathbf{H}_k\|,X_k,\mathbf{S}_1^k,\mathcal{V}\big) &(154)\\&=I\big(X_k;\|\mathbf{H}_k\||X_k|\,e^{\mathrm{i}\Theta_k}\big|\mathbf{S}_1^k,\mathcal{V}\big)+h_\lambda\big(\hat{\mathbf{H}}_k\,e^{\mathrm{i}\Theta_k}\big|\mathbf{S}_1^k,\mathcal{V}\big)\\&\quad-h_\lambda\big(\hat{\mathbf{H}}_k\big|\|\mathbf{H}_k\|,X_k,\mathbf{S}_1^k,\mathcal{V}\big). &(155)\end{aligned}$$

Here, in (153) we use that both $\|\mathbf{H}_k\||X_k|$ and Θ_k can be recovered from $\|\mathbf{H}_k\||X_k|\,e^{\mathrm{i}\Theta_k}$; and (154) follows from conditioning that reduces entropy. We next apply a conditional version of Lemma 2 to the first term in (155) where we substitute $X=X_k$ and $T=\|\mathbf{H}_k\||X_k|\,e^{\mathrm{i}\Theta_k}$:

$$\begin{aligned}&I\big(X_k;\|\mathbf{H}_k\||X_k|\,e^{\mathrm{i}\Theta_k}\big|\mathbf{S}_1^k,\mathcal{V}\big)\\&\quad\le -h\big(\|\mathbf{H}_k\||X_k|\,e^{\mathrm{i}\Theta_k}\big|X_k,\mathbf{S}_1^k,\mathcal{V}\big)+\log\pi+\mu\log\eta\\&\qquad+\log\Gamma\left(\mu,\frac{\nu}{\eta}\right)+(1-\mu)\,\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2\,|X_k|^2+\nu\big)\big|\mathcal{V}\big]\\&\qquad+\frac{1}{\eta}\,\mathsf{E}\big[\|\mathbf{H}_k\|^2\,|X_k|^2\big|\mathcal{V}\big]+\frac{\nu}{\eta} &(156)\end{aligned}$$

with free parameters $\mu,\eta>0$ and $\nu\ge 0$. We restrict the choice of μ further to $0<\mu\le 1$. Note that (see Lemma 1)

$$\begin{aligned}&h\big(\|\mathbf{H}_k\||X_k|\,e^{\mathrm{i}\Theta_k}\big|X_k,\mathbf{S}_1^k,\mathcal{V}\big)\\&\quad=\mathsf{E}\big[\log|X_k|^2\big|\mathcal{V}\big]+h\big(\|\mathbf{H}_k\|\,e^{\mathrm{i}\Theta_k}\big|X_k,\mathbf{S}_1^k,\mathcal{V}\big) &(157)\\&\quad=\mathsf{E}\big[\log|X_k|^2\big|\mathcal{V}\big]+\log 2\pi+h\big(\|\mathbf{H}_k\|\big|X_k,\mathbf{S}_1^k,\mathcal{V}\big)\\&\qquad+\mathsf{E}\big[\log\|\mathbf{H}_k\|\big|\mathcal{V}\big]. &(158)\end{aligned}$$

Moreover, we define

$$\begin{aligned}\epsilon_{\nu,k}&\triangleq\sup_{r\ge\xi_{\min}}\Big\{\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2r^2+\nu\big)\big|\mathcal{V}\big]\\&\qquad-\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2r^2\big)\big|\mathcal{V}\big]\Big\} &(159)\\ \epsilon_\nu&\triangleq\sup_{r\ge\xi_{\min}}\Big\{\mathsf{E}\big[\log\big(\|\mathbf{H}_0\|^2r^2+\nu\big)\big|A_0=0\big]\\&\qquad-\mathsf{E}\big[\log\big(\|\mathbf{H}_0\|^2r^2\big)\big|A_0=0\big]\Big\} &(160)\end{aligned}$$

such that

$$\begin{aligned}&\beta_k\epsilon_{\nu,k}\\&=\sup_{r\ge\xi_{\min}}\Big\{\beta_k\,\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2r^2+\nu\big)\big|A_k=0,B_k=1\big]\\&\qquad-\beta_k\,\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2r^2\big)\big|A_k=0,B_k=1\big]\Big\} &(161)\\&\le\sup_{r\ge\xi_{\min}}\Big\{\beta_k\,\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2r^2+\nu\big)\big|A_k=0,B_k=1\big]\\&\qquad-\beta_k\,\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2r^2\big)\big|A_k=0,B_k=1\big]\\&\qquad+(1-\beta_k)\,\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2r^2+\nu\big)\big|A_k=0,B_k=0\big]\\&\qquad-(1-\beta_k)\,\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2r^2\big)\big|A_k=0,B_k=0\big]\Big\} &(162)\\&=\sup_{r\ge\xi_{\min}}\Big\{\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2r^2+\nu\big)\big|A_k=0\big]\\&\qquad-\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2r^2\big)\big|A_k=0\big]\Big\} &(163)\\&=\epsilon_\nu &(164)\end{aligned}$$

(where the last equality follows from stationarity), and such that

$$\begin{aligned}&(1-\mu)\,\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2\,|X_k|^2+\nu\big)\big|\mathcal{V}\big]\\&\quad=(1-\mu)\,\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2\,|X_k|^2\big)\big|\mathcal{V}\big]\\&\qquad+(1-\mu)\,\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2\,|X_k|^2+\nu\big)\big|\mathcal{V}\big]\\&\qquad-(1-\mu)\,\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2\,|X_k|^2\big)\big|\mathcal{V}\big] &(165)\\&\quad\le(1-\mu)\,\mathsf{E}\big[\log\big(\|\mathbf{H}_k\|^2\,|X_k|^2\big)\big|\mathcal{V}\big]+(1-\mu)\epsilon_{\nu,k} &(166)\\&\quad\le(1-\mu)\,\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|\mathcal{V}\big]\\&\qquad+(1-\mu)\,\mathsf{E}\big[\log|X_k|^2\big|\mathcal{V}\big]+\epsilon_{\nu,k} &(167)\end{aligned}$$

where we use that $\mu \leq 1$ and that conditional on $\mathcal{V}$ we have $|X_k| \geq \xi_{\min}$.

Plugging this all back into (155) now yields

$$\begin{aligned}
&I\big(X_k; \mathbf{Y}_k \big| \mathbf{S}_1^k, \mathcal{V}\big) \\
&\leq -\log 2 - h\big(\|\mathbf{H}_k\| \big| X_k, \mathbf{S}_1^k, \mathcal{V}\big) - \mathsf{E}\big[\log \|\mathbf{H}_k\| \big| \mathcal{V}\big] \\
&\quad + \mu \log \eta + \log \Gamma\left(\mu, \frac{\nu}{\eta}\right) + (1-\mu)\,\mathsf{E}\big[\log \|\mathbf{H}_k\|^2 \big| \mathcal{V}\big] \\
&\quad - \mu\,\mathsf{E}\big[\log |X_k|^2 \big| \mathcal{V}\big] + \epsilon_{\nu,k} + \frac{1}{\eta}\,\mathsf{E}\big[\|\mathbf{H}_k\|^2\,|X_k|^2 \big| \mathcal{V}\big] + \frac{\nu}{\eta} \\
&\quad + h_\lambda\big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k} \big| \mathbf{S}_1^k, \mathcal{V}\big) - h_\lambda\big(\hat{\mathbf{H}}_k \big| \|\mathbf{H}_k\|, X_k, \mathbf{S}_1^k, \mathcal{V}\big) && (168) \\
&= -\log 2 - \mathsf{E}\big[\log \|\mathbf{H}_k\| \big| \mathcal{V}\big] + \mu \log \eta + \log \Gamma\left(\mu, \frac{\nu}{\eta}\right) \\
&\quad + (1-\mu)\,\mathsf{E}\big[\log \|\mathbf{H}_k\|^2 \big| \mathcal{V}\big] - \mu\,\mathsf{E}\big[\log |X_k|^2 \big| \mathcal{V}\big] + \epsilon_{\nu,k} \\
&\quad + \frac{1}{\eta}\,\mathsf{E}\big[\|\mathbf{H}_k\|^2\,|X_k|^2 \big| \mathcal{V}\big] + \frac{\nu}{\eta} + h_\lambda\big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k} \big| \mathbf{S}_1^k, \mathcal{V}\big) \\
&\quad - h\big(\mathbf{H}_k \big| X_k, \mathbf{S}_1^k, \mathcal{V}\big) + (2n_{\mathrm{R}} - 1)\,\mathsf{E}\big[\log \|\mathbf{H}_k\| \big| \mathcal{V}\big] && (169) \\
&= -\log 2 + \mu \log \eta + \log \Gamma\left(\mu, \frac{\nu}{\eta}\right) \\
&\quad + (n_{\mathrm{R}} - \mu)\,\mathsf{E}\big[\log \|\mathbf{H}_k\|^2 \big| \mathcal{V}\big] - \mu\,\mathsf{E}\big[\log |X_k|^2 \big| \mathcal{V}\big] \\
&\quad + \epsilon_{\nu,k} + \frac{1}{\eta}\,\mathsf{E}\big[\|\mathbf{H}_k\|^2\,|X_k|^2 \big| \mathcal{V}\big] + \frac{\nu}{\eta} \\
&\quad + h_\lambda\big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k} \big| \mathbf{S}_1^k, \mathcal{V}\big) - h\big(\mathbf{H}_k \big| X_k, \mathbf{S}_1^k, \mathcal{V}\big) && (170)
\end{aligned}$$

where in (169) we have made use of Lemma 1 once more.

E. Bound on $I\big(\mathbf{H}_1^{k-1}; \mathbf{Y}_k \big| X_k, \mathbf{S}_1^k, \mathcal{V}\big)$

We bound the second term as follows:

$$\begin{aligned}
&I\big(\mathbf{H}_1^{k-1}; \mathbf{Y}_k \big| X_k, \mathbf{S}_1^k, \mathcal{V}\big) \\
&\leq I\big(\mathbf{H}_1^{k-1}; \mathbf{Y}_k, \mathbf{H}_k \big| X_k, \mathbf{S}_1^k, \mathcal{V}\big) && (171) \\
&= I\big(\mathbf{H}_1^{k-1}; \mathbf{H}_k \big| X_k, \mathbf{S}_1^k, \mathcal{V}\big) \\
&\quad + \underbrace{I\big(\mathbf{H}_1^{k-1}; \mathbf{Y}_k \big| \mathbf{H}_k, X_k, \mathbf{S}_1^k, \mathcal{V}\big)}_{=0,\ \text{see Fig. 5 in Appendix A}} && (172) \\
&= h\big(\mathbf{H}_k \big| X_k, \mathbf{S}_1^k, \mathcal{V}\big) - h\big(\mathbf{H}_k \big| \mathbf{H}_1^{k-1}, X_k, \mathbf{S}_1^k, \mathcal{V}\big) && (173) \\
&= h\big(\mathbf{H}_k \big| X_k, \mathbf{S}_1^k, \mathcal{V}\big) - h\big(\mathbf{H}_k \big| \mathbf{H}_1^{k-1}, \mathbf{S}_1^k, A_k = 0\big) && (174)
\end{aligned}$$

where the last equality holds because conditional on $\mathbf{H}_1^{k-1}$ and $\mathbf{S}_1^k$, the input X_k is independent of $\mathbf{H}_k$ (see Fig. 6 in Appendix A).

F. Bound on $I\big(\mathbf{Y}_{k-\kappa}^{k-1}; \mathbf{Y}_k \big| \mathbf{S}_1^k, A_k = 0\big)$

The third term is bounded as follows:

$$\begin{aligned}
&I\big(\mathbf{Y}_{k-\kappa}^{k-1}; \mathbf{Y}_k \big| \mathbf{S}_1^k, A_k = 0\big) \\
&= I\big(\mathbf{Y}_{k-\kappa}^{k-1}, B_k; \mathbf{Y}_k \big| \mathbf{S}_1^k, A_k = 0\big) \\
&\quad - I\big(B_k; \mathbf{Y}_k \big| \mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{S}_1^k, A_k = 0\big) && (175) \\
&= \underbrace{I\big(B_k; \mathbf{Y}_k \big| \mathbf{S}_1^k, A_k = 0\big)}_{\geq 0} \\
&\quad + \beta_k\, I\big(\mathbf{Y}_{k-\kappa}^{k-1}; \mathbf{Y}_k \big| \mathbf{S}_1^k, A_k = 0, B_k = 1\big) \\
&\quad + (1-\beta_k)\, \underbrace{I\big(\mathbf{Y}_{k-\kappa}^{k-1}; \mathbf{Y}_k \big| \mathbf{S}_1^k, A_k = 0, B_k = 0\big)}_{\geq 0} \\
&\quad - \mathsf{H}\big(B_k \big| \mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{S}_1^k, A_k = 0\big) \\
&\quad + \underbrace{\mathsf{H}\big(B_k \big| \mathbf{Y}_{k-\kappa}^{k}, \mathbf{S}_1^k, A_k = 0\big)}_{\geq 0} && (176) \\
&\geq \beta_k\, I\big(\mathbf{Y}_{k-\kappa}^{k-1}; \mathbf{Y}_k \big| \mathbf{S}_1^k, A_k = 0, B_k = 1\big) \\
&\quad - \mathsf{H}(B_k | A_k = 0) && (177) \\
&= \beta_k\, I\big(\mathbf{Y}_{k-\kappa}^{k-1}; \mathbf{Y}_k \big| \mathbf{S}_1^k, \mathcal{V}\big) - \mathsf{H}_{\mathrm{b}}(\beta_k) && (178) \\
&= \beta_k\, I\big(\mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{Z}_{k-\kappa}^{k-1}; \mathbf{Y}_k \big| \mathbf{S}_1^k, \mathcal{V}\big) \\
&\quad - \underbrace{\beta_k\, I\big(\mathbf{Z}_{k-\kappa}^{k-1}; \mathbf{Y}_k \big| \mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{S}_1^k, \mathcal{V}\big)}_{\leq \delta_2(\xi_{\min},\kappa) + \mathsf{H}_{\mathrm{b}}(\beta_k)} - \mathsf{H}_{\mathrm{b}}(\beta_k) && (179) \\
&\geq \beta_k\, I\big(\mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{Z}_{k-\kappa}^{k-1}; \mathbf{Y}_k \big| \mathbf{S}_1^k, \mathcal{V}\big) - \delta_2(\xi_{\min}, \kappa) \\
&\quad - 2\,\mathsf{H}_{\mathrm{b}}(\beta_k) && (180) \\
&= \beta_k\, I\big(\mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{Z}_{k-\kappa}^{k-1}; \mathbf{Y}_k, \mathbf{Z}_k \big| \mathbf{S}_1^k, \mathcal{V}\big) \\
&\quad - \underbrace{\beta_k\, I\big(\mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{Z}_{k-\kappa}^{k-1}; \mathbf{Z}_k \big| \mathbf{Y}_k, \mathbf{S}_1^k, \mathcal{V}\big)}_{\leq \delta_3(\xi_{\min},\kappa) + \mathsf{H}_{\mathrm{b}}(\beta_k)} - \delta_2(\xi_{\min}, \kappa) \\
&\quad - 2\,\mathsf{H}_{\mathrm{b}}(\beta_k) && (181) \\
&\geq \beta_k\, I\big(\mathbf{Y}_{k-\kappa}^{k-1}, \mathbf{Z}_{k-\kappa}^{k-1}; \mathbf{Y}_k, \mathbf{Z}_k \big| \mathbf{S}_1^k, \mathcal{V}\big) - \delta_2(\xi_{\min}, \kappa) \\
&\quad - \delta_3(\xi_{\min}, \kappa) - 3\,\mathsf{H}_{\mathrm{b}}(\beta_k) && (182) \\
&= \beta_k\, I\big(\{\mathbf{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}, \mathbf{Z}_{k-\kappa}^{k-1}; \mathbf{H}_k X_k, \mathbf{Z}_k \big| \mathbf{S}_1^k, \mathcal{V}\big) \\
&\quad - \delta_2(\xi_{\min}, \kappa) - \delta_3(\xi_{\min}, \kappa) - 3\,\mathsf{H}_{\mathrm{b}}(\beta_k) && (183) \\
&= \beta_k\, I\big(\{\mathbf{H}_\ell X_\ell\}_{\ell=k-\kappa}^{k-1}, \mathbf{Z}_{k-\kappa}^{k-1}, \Theta_{k-\kappa}^{k-1}; \\
&\qquad\qquad \mathbf{H}_k X_k, \mathbf{Z}_k, \Theta_k \big| \mathbf{S}_1^k, \mathcal{V}\big) \\
&\quad - \delta_2(\xi_{\min}, \kappa) - \delta_3(\xi_{\min}, \kappa) - 3\,\mathsf{H}_{\mathrm{b}}(\beta_k) && (184) \\
&= \beta_k\, I\big(\{\mathbf{H}_\ell X_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=k-\kappa}^{k-1}, \mathbf{Z}_{k-\kappa}^{k-1}, \Theta_{k-\kappa}^{k-1}; \\
&\qquad\qquad \mathbf{H}_k X_k\, e^{\mathrm{i}\Theta_k}, \mathbf{Z}_k, \Theta_k \big| \mathbf{S}_1^k, \mathcal{V}\big) \\
&\quad - \delta_2(\xi_{\min}, \kappa) - \delta_3(\xi_{\min}, \kappa) - 3\,\mathsf{H}_{\mathrm{b}}(\beta_k) && (185) \\
&\geq \beta_k\, I\big(\{\mathbf{H}_\ell X_\ell\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=k-\kappa}^{k-1}; \mathbf{H}_k X_k\, e^{\mathrm{i}\Theta_k} \big| \mathbf{S}_1^k, \mathcal{V}\big) \\
&\quad - \delta_2(\xi_{\min}, \kappa) - \delta_3(\xi_{\min}, \kappa) - 3\,\mathsf{H}_{\mathrm{b}}(\beta_k) && (186) \\
&= \beta_k\, I\big(\{\mathbf{H}_\ell |X_\ell|\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=k-\kappa}^{k-1}; \mathbf{H}_k |X_k|\, e^{\mathrm{i}\Theta_k} \big| \mathbf{S}_1^k, \mathcal{V}\big) \\
&\quad - \delta_2(\xi_{\min}, \kappa) - \delta_3(\xi_{\min}, \kappa) - 3\,\mathsf{H}_{\mathrm{b}}(\beta_k). && (187)
\end{aligned}$$

Here, (177) follows by dropping some nonnegative terms and by conditioning that reduces entropy; the bounds in (180) and (182) are derived in Appendix D with $\delta_2(\xi_{\min}, \kappa)$ and $\delta_3(\xi_{\min}, \kappa)$ defined there and shown to tend to 0 as $\xi_{\min}$ tends to infinity; in (184) we add IID and uniformly distributed random variables Θ_ℓ that are independent of all other random quantities; and in (186) we drop some arguments of the mutual information functional.

For the first term in (187), we continue as follows:

$$\begin{aligned}
&\beta_k I\big(\{\mathbf{H}_\ell |X_\ell|\, e^{\mathrm{i}\Theta_\ell}\}_{\ell=k-\kappa}^{k-1}; \mathbf{H}_k |X_k|\, e^{\mathrm{i}\Theta_k} \big| \mathbf{S}_1^k, \mathcal{V}\big) \\
&= \beta_k\, I\Big(\big\{\|\mathbf{H}_\ell\|\,|X_\ell|\big\}_{\ell=k-\kappa}^{k-1}, \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}; \\
&\qquad\qquad \|\mathbf{H}_k\|\,|X_k|, \hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k} \Big| \mathbf{S}_1^k, \mathcal{V}\Big) && (188) \\
&\geq \beta_k\, I\Big(\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}; \hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k} \Big| \mathbf{S}_1^k, \mathcal{V}\Big) && (189) \\
&= \beta_k\, h_\lambda\big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k} \big| \mathbf{S}_1^k, \mathcal{V}\big) \\
&\quad - \beta_k\, h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k} \Big| \big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}, \mathbf{S}_1^k, \mathcal{V}\Big)
\end{aligned}$$

$$\pm (1-\beta_k)\, h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}, \mathbf{S}_1^k, A_k=0, B_k=0\Big) \tag{190}$$

$$= \beta_k\, h_\lambda\big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\big|\mathbf{S}_1^k, \mathcal{V}\big) - h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}, \mathbf{S}_1^k, A_k=0, B_k\Big) + (1-\beta_k)\, h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}, \mathbf{S}_1^k, A_k=0, B_k=0\Big) \tag{191}$$

$$\geq \beta_k\, h_\lambda\big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\big|\mathbf{S}_1^k, \mathcal{V}\big) - h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}, \mathbf{S}_{k-\kappa}^k, A_k=0\Big) + (1-\beta_k)\, h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}, \mathbf{S}_1^k, A_k=0, B_k=0\Big). \tag{192}$$

Here, in (189), we drop some arguments; and in (192), we remove some arguments from the conditioning of the second differential entropy term.

G. Combination of Three Bounds

We combine the three bounds (170), (174), and (192) in (145) as follows:

$$\begin{aligned}
&I\big(M; \mathbf{Y}_k, \mathbf{S}_k\big|\mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}, A_k=0\big)\\
&\leq -\beta_k \log 2 + \beta_k \mu \log\eta + \beta_k \log\Gamma\Big(\mu, \frac{\nu}{\eta}\Big)\\
&\quad + \beta_k(n_{\mathsf{R}}-\mu)\, \mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|\mathcal{V}\big] - \beta_k\mu\, \mathsf{E}\big[\log|X_k|^2\big|\mathcal{V}\big]\\
&\quad + \beta_k\epsilon_{\nu,k} + \frac{\beta_k}{\eta}\, \mathsf{E}\big[\|\mathbf{H}_k\|^2\, |X_k|^2\big|\mathcal{V}\big] + \beta_k\frac{\nu}{\eta}\\
&\quad + \underline{\beta_k\, h_\lambda\big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\big|\mathbf{S}_1^k, \mathcal{V}\big)} - \underline{\beta_k\, h\big(\mathbf{H}_k\big|X_k, \mathbf{S}_1^k, \mathcal{V}\big)}\\
&\quad + \underline{\beta_k\, h\big(\mathbf{H}_k\big|X_k, \mathbf{S}_1^k, \mathcal{V}\big)} - \beta_k\, h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1}, \mathbf{S}_1^k, A_k=0\big)\\
&\quad - \underline{\beta_k\, h_\lambda\big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\big|\mathbf{S}_1^k, \mathcal{V}\big)}\\
&\quad + h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}, \mathbf{S}_{k-\kappa}^k, A_k=0\Big)\\
&\quad - (1-\beta_k)\, h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}, \mathbf{S}_1^k, A_k=0, B_k=0\Big)\\
&\quad + \delta_2(\xi_{\min},\kappa) + \delta_3(\xi_{\min},\kappa) + 3\,\mathsf{H}_\mathsf{b}(\beta_k) + \mathsf{H}_\mathsf{b}(\beta_k)\\
&\quad + (1-\beta_k)\, I\big(X_k, \mathbf{H}_1^{k-1}; \mathbf{Y}_k\big|\mathbf{S}_1^k, A_k=0, B_k=0\big).
\end{aligned} \tag{193}$$

Note that the four underlined terms cancel, that

$$\begin{aligned}
\mathsf{E}\big[\log|X_k|^2\big|\mathcal{V}\big] &= \mathsf{E}\big[\log|X_k|^2\big|A_k=0, B_k=1\big] &(194)\\
&\geq \log\xi_{\min}^2 &(195)
\end{aligned}$$

and that

$$\begin{aligned}
&\beta_k\, \mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|\mathcal{V}\big]\\
&\quad = \beta_k\, \mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=0, B_k=1\big] &(196)\\
&\quad = \mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=0\big]\\
&\qquad - (1-\beta_k)\, \mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=0, B_k=0\big]. &(197)
\end{aligned}$$

Hence, with (164), we get

$$\begin{aligned}
&I\big(M; \mathbf{Y}_k, \mathbf{S}_k\big|\mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}, A_k=0\big)\\
&\leq h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}, \mathbf{S}_{k-\kappa}^k, A_k=0\Big)\\
&\quad - \beta_k\log 2 + n_\mathsf{R}\, \mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=0\big]\\
&\quad - \beta_k\, h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1}, \mathbf{S}_1^k, A_k=0\big) + \beta_k\log\Gamma\Big(\mu, \frac{\nu}{\eta}\Big)\\
&\quad + \frac{\beta_k}{\eta}\, \mathsf{E}\big[\|\mathbf{H}_k\|^2\, |X_k|^2\big|A_k=0, B_k=1\big] + \beta_k\frac{\nu}{\eta}\\
&\quad + \mu\Big(\beta_k\log\eta - \mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=0\big]\\
&\qquad + (1-\beta_k)\, \mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=0, B_k=0\big]\\
&\qquad - \beta_k\log\xi_{\min}^2\Big)\\
&\quad + (1-\beta_k)\Big(I\big(X_k, \mathbf{H}_1^{k-1}; \mathbf{Y}_k\big|\mathbf{S}_1^k, A_k=0, B_k=0\big)\\
&\qquad - h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}, \mathbf{S}_1^k, A_k=0, B_k=0\Big)\\
&\qquad - n_\mathsf{R}\, \mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=0, B_k=0\big]\Big)\\
&\quad + 4\,\mathsf{H}_\mathsf{b}(\beta_k) + \delta_2(\xi_{\min},\kappa) + \delta_3(\xi_{\min},\kappa) + \epsilon_\nu.
\end{aligned} \tag{198}$$

Next, we bound

$$\begin{aligned}
&(1-\beta_k)\, \mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=0, B_k=0\big]\\
&\quad \leq (1-\beta_k)\log\mathsf{E}\big[\|\mathbf{H}_k\|^2\big|A_k=0, B_k=0\big] &(199)\\
&\quad \leq (1-\beta_k)\log\frac{\mathsf{E}\big[\|\mathbf{H}_k\|^2\big|A_k=0\big]}{1-\beta_k} &(200)\\
&\quad = (1-\beta_k)\log\mathsf{E}\big[\|\mathbf{H}_k\|^2\big|A_k=0\big]\\
&\qquad - (1-\beta_k)\log(1-\beta_k) &(201)\\
&\quad \leq (1-\beta_k)\log\mathsf{E}\big[\|\mathbf{H}_k\|^2\big|A_k=0\big] + e^{-1} &(202)\\
&\quad \leq (1-\beta_k)\log\big(2\,\mathsf{E}\big[\|\mathbf{H}_k\|^2\big]\big) + e^{-1} &(203)
\end{aligned}$$

where (199) follows from Jensen's inequality, where (200) follows from Lemma 9, and where (203) can be derived from Lemma 9 and (122) as follows:

$$\begin{aligned}
\mathsf{E}\big[\|\mathbf{H}_k\|^2\big|A_k=0\big] &\leq \frac{1}{1-\alpha_k}\, \mathsf{E}\big[\|\mathbf{H}_k\|^2\big] &(204)\\
&\leq \frac{1}{1-\frac{1}{2}}\, \mathsf{E}\big[\|\mathbf{H}_k\|^2\big] &(205)\\
&= 2\,\mathsf{E}\big[\|\mathbf{H}_k\|^2\big]. &(206)
\end{aligned}$$

Furthermore, using that conditional on $A_k = 0$ we have $\|\mathbf{H}_k\|^2 \leq t$ and relying once more on Lemma 9 and (122), we obtain

$$\begin{aligned}
&\frac{\beta_k}{\eta}\, \mathsf{E}\big[\|\mathbf{H}_k\|^2\, |X_k|^2\big|A_k=0, B_k=1\big]\\
&\quad \leq \frac{1}{\eta}\, \mathsf{E}\big[\|\mathbf{H}_k\|^2\, |X_k|^2\big|A_k=0\big] &(207)\\
&\quad \leq \frac{t}{\eta}\, \mathsf{E}\big[|X_k|^2\big|A_k=0\big] &(208)\\
&\quad \leq \frac{t}{\eta}\, \frac{1}{1-\alpha_k}\, \mathsf{E}\big[|X_k|^2\big] &(209)\\
&\quad \leq \frac{2t}{\eta}\, \mathsf{E}\big[|X_k|^2\big]. &(210)
\end{aligned}$$

Moreover, we bound

$$\begin{aligned}
-\beta_k\, h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},\mathbf{S}_1^k,A_k=0\big)
&= -h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},\mathbf{S}_1^k,A_k=0\big)\\
&\quad + (1-\beta_k)\, h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},\mathbf{S}_1^k,A_k=0\big) \qquad (211)\\
&\le -h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},\mathbf{S}_1^k,A_k=0\big)\\
&\quad + (1-\beta_k)\, h(\mathbf{H}_k|A_k=0). \qquad (212)
\end{aligned}$$

Finally, the following bound is proven in Appendix E:

$$\begin{aligned}
&I\big(X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|\mathbf{S}_1^k,A_k=0,B_k=0\big)\\
&- h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1},\mathbf{S}_1^k,A_k=0,B_k=0\Big)\\
&- n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=0,B_k=0\big]\\
&\quad\le -4n_{\mathrm{R}}h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0,A_0=0\big)+\frac{4n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e}\\
&\qquad - h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\Big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0,A_0=0\Big)\\
&\qquad + \frac{16n_{\mathrm{R}}\pi^{n_{\mathrm{R}}}}{\Gamma(n_{\mathrm{R}})(4n_{\mathrm{R}}-1)^2}+4n_{\mathrm{R}}^2\log\left(1+\frac{2\pi e}{n_{\mathrm{R}}}\,\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{1-\beta_k}\right)\\
&\qquad + n_{\mathrm{R}}\log\left(1+\frac{2\xi_{\min}^2}{n_{\mathrm{R}}\,\sigma^2}\,\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{1-\beta_k}\right). \qquad (213)
\end{aligned}$$

Hence, using all these bounds in (198), we obtain

$$\begin{aligned}
&I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},A_k=0\big)\\
&\le h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1},\mathbf{S}_{k-\kappa}^k,A_k=0\Big)\\
&\quad -\log 2+n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=0\big]\\
&\quad - h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},\mathbf{S}_1^k,A_k=0\big)+(1-\beta_k)\log 2\\
&\quad + (1-\beta_k)\,h(\mathbf{H}_k|A_k=0)+\beta_k\log\Gamma\left(\mu,\frac{\nu}{\eta}\right)\\
&\quad + \frac{2t}{\eta}\,\mathsf{E}\big[|X_k|^2\big]+\beta_k\frac{\nu}{\eta}\\
&\quad + \mu\Big(\beta_k\log\eta-\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=0\big]\\
&\qquad\quad + (1-\beta_k)\log\big(2\,\mathsf{E}\big[\|\mathbf{H}_k\|^2\big]\big)+e^{-1}-\beta_k\log\xi_{\min}^2\Big)\\
&\quad + (1-\beta_k)\bigg(-4n_{\mathrm{R}}h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0,A_0=0\big)\\
&\qquad\quad + \frac{4n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e}+\frac{16n_{\mathrm{R}}\pi^{n_{\mathrm{R}}}}{\Gamma(n_{\mathrm{R}})(4n_{\mathrm{R}}-1)^2}\\
&\qquad\quad - h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\Big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0,A_0=0\Big)\bigg)\\
&\quad + (1-\beta_k)4n_{\mathrm{R}}^2\log\left(1+\frac{2\pi e}{n_{\mathrm{R}}}\,\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{1-\beta_k}\right)\\
&\quad + (1-\beta_k)n_{\mathrm{R}}\log\left(1+\frac{2\xi_{\min}^2}{n_{\mathrm{R}}\,\sigma^2}\,\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{1-\beta_k}\right)\\
&\quad + 4\,\mathsf{H}_{\mathrm{b}}(\beta_k)+\delta_2(\xi_{\min},\kappa)+\delta_3(\xi_{\min},\kappa)+\epsilon_\nu. \qquad (214)
\end{aligned}$$

We next bound the first four terms as follows:

$$\begin{aligned}
&h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1},\mathbf{S}_{k-\kappa}^k,A_k=0\Big)-\log 2\\
&\quad + n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=0\big]-h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},\mathbf{S}_1^k,A_k=0\big)\\
&= \frac{1}{1-\alpha_k}\bigg(h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1},\mathbf{S}_{k-\kappa}^k,A_k\Big)\\
&\qquad -\log 2+n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big]\\
&\qquad - h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},\mathbf{S}_1^k,A_k\big)\bigg)\\
&\quad - \frac{\alpha_k}{1-\alpha_k}\bigg(h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1},\mathbf{S}_{k-\kappa}^k,\\
&\qquad\qquad A_k=1\Big)\\
&\qquad -\log 2+n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=1\big]\\
&\qquad - h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},\mathbf{S}_1^k,A_k=1\big)\bigg) \qquad (215)\\
&\le \frac{1}{1-\alpha_k}\bigg(h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1},\mathbf{S}_{k-\kappa}^k\Big)-\log 2\\
&\qquad + n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big]-h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},\mathbf{S}_1^k\big)\bigg)\\
&\quad + \frac{1}{1-\alpha_k}\,I\big(A_k;\mathbf{H}_k\big|\mathbf{H}_1^{k-1},\mathbf{S}_1^k\big)\\
&\quad - \frac{\alpha_k}{1-\alpha_k}\bigg(h_\lambda\Big(\hat{\mathbf{H}}_k\, e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1},\mathbf{S}_{k-\kappa}^k,\\
&\qquad\qquad A_k=1\Big)\\
&\qquad -\log 2+n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|A_k=1\big]\\
&\qquad - h\big(\mathbf{H}_k\big|A_k=1\big)\bigg) \qquad (216)\\
&\le \frac{1}{1-\alpha_k}\bigg(h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^0\Big)-\log 2\\
&\qquad + n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big]-h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0\big)\bigg)\\
&\quad + \frac{1}{1-\alpha_k}\,\mathsf{H}_{\mathrm{b}}(\alpha_k)\\
&\quad + \frac{\alpha_k}{1-\alpha_k}\,\bigg|h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^0,A_0=1\Big)\\
&\qquad -\log 2+n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|A_0=1\big]\\
&\qquad - h\big(\mathbf{H}_0\big|A_0=1\big)\bigg| \qquad (217)\\
&\le \frac{1}{1-\delta_4(t,\kappa)}\bigg(h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^0\Big)\\
&\qquad -\log 2+n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big]\\
&\qquad - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0\big)\bigg)\\
&\quad + 2\,\mathsf{H}_{\mathrm{b}}\left(\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\right)+\frac{2\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\cdot\\
&\qquad\cdot\bigg|h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^0,A_0=1\Big)\\
&\qquad -\log 2+n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|A_0=1\big]\\
&\qquad - h\big(\mathbf{H}_0\big|A_0=1\big)\bigg|. \qquad (218)
\end{aligned}$$

Here, (215) rewrites the expression with respect to A_k; in (216) we drop terms in the conditioning of the first and the last differential entropy term; (217) follows from stationarity plus adding some conditioning, from taking the absolute value, and from dropping a nonnegative entropy term; and the last inequality (218) we rely on (118), (121), $\alpha_k \leq \frac{1}{2}$, and define

$$\begin{aligned}
&\delta_4(t,\kappa)\\
&\triangleq \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\,\mathscr{I}\Bigg\{h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^0\Big)\\
&\qquad -\log 2+n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big]\\
&\qquad -h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0\big)\geq 0\Bigg\}
\end{aligned}\tag{219}$$

where $\mathscr{I}\{\text{statement}\}$ denotes the indicator function that takes on the value one if the statement holds true and zero otherwise. Note that $\lim_{t\uparrow\infty}\delta_4(t,\kappa)=0$.

Hence, using this in (214) and applying stationarity where possible, we have

$$\begin{aligned}
&I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},A_k=0\big)\\
&\leq \frac{1}{1-\delta_4(t,\kappa)}\Bigg(h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^0\Big)\\
&\qquad -\log 2+n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big]\\
&\qquad -h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0\big)\Bigg)\\
&\quad +2\,\mathsf{H}_\mathsf{b}\left(\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\right)+\frac{2\,\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\cdot\\
&\qquad\cdot\Bigg|h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^0,A_0=1\Big)\\
&\qquad -\log 2+n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|A_0=1\big]\\
&\qquad -h\big(\mathbf{H}_0\big|A_0=1\big)\Bigg|\\
&\quad +(1-\beta_k)\log 2+(1-\beta_k)\,h(\mathbf{H}_0|A_0=0)\\
&\quad +\beta_k\log\Gamma\left(\mu,\frac{\nu}{\eta}\right)+\frac{2t}{\eta}\,\mathsf{E}\big[|X_k|^2\big]+\beta_k\frac{\nu}{\eta}\\
&\quad +\mu\Big(\beta_k\log\eta-\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|A_0=0\big]\\
&\qquad +(1-\beta_k)\log\big(2\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]\big)+e^{-1}-\beta_k\log\xi_{\min}^2\Big)\\
&\quad +(1-\beta_k)\Bigg(-4n_{\mathrm{R}}h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0,A_0=0\big)\\
&\qquad +\frac{4n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e}+\frac{16n_{\mathrm{R}}\pi^{n_{\mathrm{R}}}}{\Gamma(n_{\mathrm{R}})(4n_{\mathrm{R}}-1)^2}\\
&\qquad -h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0,A_0=0\Big)\Bigg)\\
&\quad +(1-\beta_k)\,4n_{\mathrm{R}}^2\log\left(1+\frac{2\pi e}{n_{\mathrm{R}}}\,\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{1-\beta_k}\right)\\
&\quad +(1-\beta_k)\,n_{\mathrm{R}}\log\left(1+\frac{2\xi_{\min}^2}{n_{\mathrm{R}}\,\sigma^2}\,\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{1-\beta_k}\right)\\
&\quad +4\,\mathsf{H}_\mathsf{b}(\beta_k)+\delta_2(\xi_{\min},\kappa)+\delta_3(\xi_{\min},\kappa)+\epsilon_\nu.
\end{aligned}\tag{220}$$

We now define

$$\beta_{n,\kappa}\triangleq\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n}\beta_k\tag{221}$$

and note that $\beta_k\mapsto(1-\beta_k)\log\big(1+\text{const}/(1-\beta_k)\big)$ and $\mathsf{H}_\mathsf{b}(\cdot)$ are concave such that by Jensen's inequality

$$\begin{aligned}
&\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n}(1-\beta_k)4n_{\mathrm{R}}^2\log\left(1+\frac{2\pi e}{n_{\mathrm{R}}}\,\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{1-\beta_k}\right)\\
&\quad\leq(1-\beta_{n,\kappa})\,4n_{\mathrm{R}}^2\log\left(1+\frac{2\pi e}{n_{\mathrm{R}}}\,\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{1-\beta_{n,\kappa}}\right)
\end{aligned}\tag{222}$$

$$\begin{aligned}
&\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n}(1-\beta_k)n_{\mathrm{R}}\log\left(1+\frac{2\xi_{\min}^2}{n_{\mathrm{R}}\,\sigma^2}\,\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{1-\beta_k}\right)\\
&\quad\leq(1-\beta_{n,\kappa})\,n_{\mathrm{R}}\log\left(1+\frac{2\xi_{\min}^2}{n_{\mathrm{R}}\,\sigma^2}\,\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{1-\beta_{n,\kappa}}\right)
\end{aligned}\tag{223}$$

and

$$\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n}\mathsf{H}_\mathsf{b}(\beta_k)\leq\mathsf{H}_\mathsf{b}(\beta_{n,\kappa}).\tag{224}$$

Moreover, by adding some nonnegative terms and using the average-power constraint (13),

$$\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n}\mathsf{E}\big[|X_k|^2\big]\leq\frac{n}{n-\kappa}\,\frac{1}{n}\sum_{k=1}^{n}\mathsf{E}\big[|X_k|^2\big]\tag{225}$$

$$\leq\frac{n}{n-\kappa}\,\mathsf{E}_\mathsf{s}.\tag{226}$$

Plugging (220) back into its corresponding summation from (133) hence yields:

$$\begin{aligned}
&\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n}I\big(M;\mathbf{Y}_k,\mathbf{S}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^{k-1},A_k=0\big)\\
&\leq\frac{1}{1-\delta_4(t,\kappa)}\Bigg(h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^0\Big)\\
&\qquad -\log 2+n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big]\\
&\qquad -h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0\big)\Bigg)\\
&\quad +2\,\mathsf{H}_\mathsf{b}\left(\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\right)+\frac{2\,\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{t}\cdot\\
&\qquad\cdot\Bigg|h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^0,A_0=1\Big)\\
&\qquad -\log 2+n_{\mathrm{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|A_0=1\big]\\
&\qquad -h\big(\mathbf{H}_0\big|A_0=1\big)\Bigg|\\
&\quad +(1-\beta_{n,\kappa})\log 2+(1-\beta_{n,\kappa})\,h(\mathbf{H}_0|A_0=0)\\
&\quad +\beta_{n,\kappa}\log\Gamma\left(\mu,\frac{\nu}{\eta}\right)+\frac{2t}{\eta}\,\frac{n}{n-\kappa}\,\mathsf{E}_\mathsf{s}+\beta_{n,\kappa}\,\frac{\nu}{\eta}\\
&\quad +\mu\Big(\beta_{n,\kappa}\log\eta-\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|A_0=0\big]\\
&\qquad +(1-\beta_{n,\kappa})\log\big(2\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]\big)+e^{-1}\\
&\qquad -\beta_{n,\kappa}\log\xi_{\min}^2\Big)\\
&\quad +(1-\beta_{n,\kappa})\Bigg(-4n_{\mathrm{R}}h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0,A_0=0\big)\\
&\qquad +\frac{4n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e}+\frac{16n_{\mathrm{R}}\pi^{n_{\mathrm{R}}}}{\Gamma(n_{\mathrm{R}})(4n_{\mathrm{R}}-1)^2}
\end{aligned}$$

$$
\begin{aligned}
&\qquad - h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\Big|\mathbf{H}_{-\infty}^{-1}, \mathbf{S}_{-\infty}^{0}, A_0 = 0\Big)\Big) \\
&\quad + (1-\beta_{n,\kappa})\, 4n_{\mathrm{R}}^2 \log\left(1 + \frac{2\pi e}{n_{\mathrm{R}}}\, \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{1-\beta_{n,\kappa}}\right) \\
&\quad + (1-\beta_{n,\kappa})\, n_{\mathrm{R}} \log\left(1 + \frac{2\xi_{\min}^2}{n_{\mathrm{R}}\,\sigma^2}\, \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{1-\beta_{n,\kappa}}\right) \\
&\quad + 4\,\mathsf{H}_{\mathrm{b}}(\beta_{n,\kappa}) + \delta_2(\xi_{\min},\kappa) + \delta_3(\xi_{\min},\kappa) + \epsilon_\nu \qquad (227) \\
&\le \frac{1}{1-\delta_4(t,\kappa)} \Big(h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{0}\Big) \\
&\qquad - \log 2 + n_{\mathrm{R}}\, \mathsf{E}\big[\log \|\mathbf{H}_0\|^2\big] \\
&\qquad - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, \mathbf{S}_{-\infty}^{0}\big)\Big) \\
&\quad + 2\,\mathsf{H}_{\mathrm{b}}\left(\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\right) + \frac{2\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t} \cdot \\
&\qquad \cdot \Big|h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{0}, A_0 = 1\Big) \\
&\qquad - \log 2 + n_{\mathrm{R}}\, \mathsf{E}\big[\log \|\mathbf{H}_0\|^2 \big| A_0 = 1\big] \\
&\qquad - h\big(\mathbf{H}_0\big|A_0 = 1\big)\Big| \\
&\quad + \frac{2t}{\eta}\,\frac{n}{n-\kappa}\,\mathsf{E}_{\mathrm{s}} + \delta_2(\xi_{\min},\kappa) + \delta_3(\xi_{\min},\kappa) + \epsilon_\nu \\
&\quad + \sup_{0\le\beta\le 1}\Big\{(1-\beta)\log 2 + (1-\beta)\, h(\mathbf{H}_0|A_0 = 0) \\
&\qquad + \beta \log \Gamma\left(\mu, \frac{\nu}{\eta}\right) + \beta\, \frac{\nu}{\eta} \\
&\qquad + \mu\Big(\beta\log\eta - \mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|A_0 = 0\big] \\
&\qquad\quad + (1-\beta)\log\big(2\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]\big) + e^{-1} \\
&\qquad\quad - \beta\log\xi_{\min}^2\Big) \\
&\qquad - (1-\beta)\, 4n_{\mathrm{R}}\, h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, \mathbf{S}_{-\infty}^{0}, A_0 = 0\big) \\
&\qquad + (1-\beta)\, \frac{4n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e} \\
&\qquad + (1-\beta)\, \frac{16 n_{\mathrm{R}} \pi^{n_{\mathrm{R}}}}{\Gamma(n_{\mathrm{R}})(4n_{\mathrm{R}}-1)^2} \\
&\qquad - (1-\beta)\, h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\Big|\mathbf{H}_{-\infty}^{-1}, \mathbf{S}_{-\infty}^{0}, A_0 = 0\Big) \\
&\qquad + (1-\beta)\, 4n_{\mathrm{R}}^2 \log\left(1 + \frac{2\pi e}{n_{\mathrm{R}}}\, \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{1-\beta}\right) \\
&\qquad + (1-\beta)\, n_{\mathrm{R}} \log\left(1 + \frac{2\xi_{\min}^2}{n_{\mathrm{R}}\,\sigma^2}\, \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{1-\beta}\right) \\
&\qquad + 4\,\mathsf{H}_{\mathrm{b}}(\beta)\Big\} \qquad (228)
\end{aligned}
$$

where in the last step we take the supremum over $\beta_{n,\kappa} \in [0,1]$.

Next, we plug (228) back into (133) and let n tend to infinity:

$$
\begin{aligned}
&\mathsf{R}_{\mathrm{FB,c}}(\mathsf{E}_{\mathrm{s}}|\{\mathbf{S}_k\}) \\
&\le \mathsf{H}_{\mathrm{b}}\left(\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\right) \\
&\quad + \frac{n_{\mathrm{R}}\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t} \log\left(1 + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big|A_0 = 1\big]}{n_{\mathrm{R}}\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}\, \frac{t\mathsf{E}_{\mathrm{s}}}{\sigma^2}\right) \\
&\quad + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\, I\big(\mathbf{H}_{-\infty}^{-1}, \mathbf{S}_{-\infty}^{-1}; \mathbf{H}_0, \mathbf{S}_0\big|A_0 = 1\big) \\
&\quad + \frac{1}{1-\delta_4(t,\kappa)} \Big(h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{0}\Big) \\
&\qquad - \log 2 + n_{\mathrm{R}}\, \mathsf{E}\big[\log \|\mathbf{H}_0\|^2\big] \\
&\qquad - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, \mathbf{S}_{-\infty}^{0}\big)\Big) \\
&\quad + 2\,\mathsf{H}_{\mathrm{b}}\left(\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\right) + \frac{2\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t} \cdot \\
&\qquad \cdot \Big|h_\lambda\Big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\, e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1}, \mathbf{S}_{-\kappa}^{0}, A_0 = 1\Big) \\
&\qquad - \log 2 + n_{\mathrm{R}}\, \mathsf{E}\big[\log \|\mathbf{H}_0\|^2 \big| A_0 = 1\big] \\
&\qquad - h\big(\mathbf{H}_0\big|A_0 = 1\big)\Big| \\
&\quad + \frac{2t\mathsf{E}_{\mathrm{s}}}{\eta} + \delta_2(\xi_{\min},\kappa) + \delta_3(\xi_{\min},\kappa) + \epsilon_\nu \\
&\quad + \sup_{0\le\beta\le 1}\Big\{(1-\beta)\log 2 + (1-\beta)\, h(\mathbf{H}_0|A_0 = 0) \\
&\qquad + \beta \log \Gamma\left(\mu, \frac{\nu}{\eta}\right) + \beta\, \frac{\nu}{\eta} \\
&\qquad + \mu\Big(\beta\log\eta - \mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|A_0 = 0\big] \\
&\qquad\quad + (1-\beta)\log\big(2\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]\big) + e^{-1} \\
&\qquad\quad - \beta\log\xi_{\min}^2\Big) \\
&\qquad - (1-\beta)\, 4n_{\mathrm{R}}\, h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, \mathbf{S}_{-\infty}^{0}, A_0 = 0\big) \\
&\qquad + (1-\beta)\, \frac{4n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e} \\
&\qquad + (1-\beta)\, \frac{16 n_{\mathrm{R}} \pi^{n_{\mathrm{R}}}}{\Gamma(n_{\mathrm{R}})(4n_{\mathrm{R}}-1)^2} \\
&\qquad - (1-\beta)\, h_\lambda\big(\hat{\mathbf{H}}_0\, e^{\mathrm{i}\Theta_0}\big|\mathbf{H}_{-\infty}^{-1}, \mathbf{S}_{-\infty}^{0}, A_0 = 0\big) \\
&\qquad + (1-\beta)\, 4n_{\mathrm{R}}^2 \log\left(1 + \frac{2\pi e}{n_{\mathrm{R}}}\, \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{1-\beta}\right) \\
&\qquad + (1-\beta)\, n_{\mathrm{R}} \log\left(1 + \frac{2\xi_{\min}^2}{n_{\mathrm{R}}\,\sigma^2}\, \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{1-\beta}\right) \\
&\qquad + 4\,\mathsf{H}_{\mathrm{b}}(\beta)\Big\}. \qquad (229)
\end{aligned}
$$

We make the following choice for the free parameters μ, η, t, and $\xi_{\min}$:

$$\mu \triangleq \frac{\nu}{\log \mathsf{E}_{\mathrm{s}}} \qquad (230)$$

$$\eta \triangleq \frac{\mathsf{E}_{\mathrm{s}} \log^2 \mathsf{E}_{\mathrm{s}}}{\nu} \qquad (231)$$

$$t \triangleq \log^2 \mathsf{E}_{\mathrm{s}} \qquad (232)$$

$$\xi_{\min} \triangleq \sqrt{\log\log \mathsf{E}_{\mathrm{s}}} \qquad (233)$$

for some fixed value $\nu > 0$, and for E_{s} large enough such that $\mu \le 1$. Note that then inside the supremum, the term $\beta \log \Gamma(\mu, \nu/\eta)$ grows like $\beta \log\log \mathsf{E}_{\mathrm{s}}$, while one more term grows like $(1-\beta)\log\log\log \mathsf{E}_s$ and the rest is bounded from above in E_s. Hence, for any E_s larger than some threshold, $\beta\log\Gamma(\mu,\nu/\eta)$ becomes the dominant term and the supremum is therefore achieved for $\beta=1$.

Note that the bound (229) holds for any system, hence also for an optimal system. So we can use (229) to upper-bound the asymptotic feedback capacity with causal side-information at transmitter and receiver:

$$
\begin{aligned}
&\chi_{\text{FB,c}}(\{\mathbf{H}_k\}|\{\mathbf{S}_k\}) \\
&\triangleq \varlimsup_{\mathsf{E}_s\uparrow\infty}\left\{\mathsf{C}_{\text{FB,c}}(\mathsf{E}_s|\{\mathbf{S}_k\})-\log\log\mathsf{E}_s\right\} \qquad (234)\\
&\le \varlimsup_{\mathsf{E}_s\uparrow\infty}\Bigg\{\mathsf{H}_b\left(\frac{\mathsf{E}[\|\mathbf{H}_0\|^2]}{\log^2\mathsf{E}_s}\right) \\
&\quad+\frac{n_R\,\mathsf{E}[\|\mathbf{H}_0\|^2]}{\log^2\mathsf{E}_s}\log\left(1+\frac{\mathsf{E}[\|\mathbf{H}_0\|^2|A_0=1]}{n_R\,\mathsf{E}[\|\mathbf{H}_0\|^2]}\frac{\mathsf{E}_s\log^2\mathsf{E}_s}{\sigma^2}\right)\\
&\quad+\frac{\mathsf{E}[\|\mathbf{H}_0\|^2]}{\log^2\mathsf{E}_s}I\big(\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^{-1};\mathbf{H}_0,\mathbf{S}_0\big|A_0=1\big)\\
&\quad+\frac{1}{1-\delta_4(\log^2\mathsf{E}_s,\kappa)}\cdot\\
&\qquad\cdot\Bigg(h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^0\Big)-\log 2\\
&\qquad\qquad+n_R\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big]-h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0\big)\Bigg)\\
&\quad+\frac{2\,\mathsf{E}[\|\mathbf{H}_0\|^2]}{\log^2\mathsf{E}_s}\cdot\\
&\qquad\cdot\Bigg|h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^0,A_0=1\Big)\\
&\qquad\quad-\log 2+n_R\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|A_0=1\big]\\
&\qquad\quad-h\big(\mathbf{H}_0\big|A_0=1\big)\Bigg|\\
&\quad+2\,\mathsf{H}_b\left(\frac{\mathsf{E}[\|\mathbf{H}_0\|^2]}{\log^2\mathsf{E}_s}\right)+2\nu+\delta_2\Big(\sqrt{\log\log\mathsf{E}_s},\kappa\Big)\\
&\quad+\delta_3\Big(\sqrt{\log\log\mathsf{E}_s},\kappa\Big)+\epsilon_\nu+\log\Gamma\left(\frac{\nu}{\log\mathsf{E}_s},\frac{\nu^2}{\mathsf{E}_s\log^2\mathsf{E}_s}\right)\\
&\quad+\frac{\nu}{\log\mathsf{E}_s}(\log\mathsf{E}_s+2\log\log\mathsf{E}_s-\log\nu)\\
&\quad-\frac{\nu}{\log\mathsf{E}_s}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|A_0=0\big]-\frac{\nu\log\log\log\mathsf{E}_s}{\log\mathsf{E}_s}\\
&\quad+\frac{\nu^2}{\mathsf{E}_s\log^2\mathsf{E}_s}+\frac{\nu}{e\log\mathsf{E}_s}-\log\log\mathsf{E}_s\Bigg\} \qquad (235)\\
&=h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^0\Big)-\log 2\\
&\quad+n_R\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big]-h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0\big)\\
&\quad+3\nu+\log(1-e^{-\nu})-\log\nu+\epsilon_\nu \qquad (236)
\end{aligned}
$$

where we have used that [6, Appendix XI]

$$
\begin{aligned}
&\varlimsup_{\mathsf{E}_s\uparrow\infty}\left\{\log\Gamma\left(\frac{\nu}{\log\mathsf{E}_s},\frac{\nu^2}{\mathsf{E}_s\log^2\mathsf{E}_s}\right)-\log\log\mathsf{E}_s\right\}\\
&\quad=\log(1-e^{-\nu})-\log\nu. \qquad (237)
\end{aligned}
$$

Note further that

$$
\lim_{\nu\downarrow 0}\left\{\log(1-e^{-\nu})-\log\nu+3\nu+\epsilon_\nu\right\}=0. \qquad (238)
$$

Hence, by choosing κ very large and ν very small, we see that

$$
\begin{aligned}
&\chi_{\text{FB,c}}(\{\mathbf{H}_k\}|\{\mathbf{S}_k\})\\
&\quad\le h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\infty}^{-1},\mathbf{S}_{-\infty}^0\Big)-\log 2\\
&\qquad+n_R\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big]-h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0\big). \qquad (239)
\end{aligned}
$$

VII. An Upper Bound on the Fading Number with Feedback and Acausal Side-Information

This derivation is very similar to the one given in Section VI. We will therefore only point out the main differences and omit many details.

We start with (108):

$$
\begin{aligned}
&\mathsf{R}_{\text{FB,ac}}(\mathsf{E}_s|\{\mathbf{S}_k\})\\
&\quad\le\frac{1}{n}I\big(M;\mathbf{Y}_1^n,\mathbf{S}_1^n\big)+\frac{\log 2}{n}\\
&\qquad+\Pr\big[M\ne\hat{M}\big]\,\mathsf{R}_{\text{FB,ac}}(\mathsf{E}_s|\{\mathbf{S}_k\})+\frac{\epsilon_n}{n} \qquad (240)\\
&\quad=\frac{1}{n}I\big(M;\mathbf{Y}_1^n\big|\mathbf{S}_1^n\big)+\frac{\log 2}{n}\\
&\qquad+\Pr\big[M\ne\hat{M}\big]\,\mathsf{R}_{\text{FB,ac}}(\mathsf{E}_s|\{\mathbf{S}_k\})+\frac{\epsilon_n}{n} \qquad (241)\\
&\quad=\frac{\kappa}{n}\,\frac{1}{\kappa}\sum_{k=1}^{\kappa}I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^n\big)\\
&\qquad+\frac{1}{n}\sum_{k=\kappa+1}^{n}I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^n\big)\\
&\qquad+\frac{\log 2}{n}+\Pr\big[M\ne\hat{M}\big]\,\mathsf{R}_{\text{FB,ac}}(\mathsf{E}_s|\{\mathbf{S}_k\})+\frac{\epsilon_n}{n}. \qquad (242)
\end{aligned}
$$

Lemma 10 is adapted as follows.

Lemma 11: In the case of acausal side-information, Lemma 10 reads

$$
\begin{aligned}
&I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^n\big)\\
&\quad\le n_R\log\left(1+\frac{1}{n_R}\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]\frac{\mathsf{E}_k}{\sigma^2}\right)+I\big(\mathbf{H}_1^{k-1};\mathbf{H}_k\big|\mathbf{S}_1^n\big). \qquad (243)
\end{aligned}
$$

Proof: Omitted. ∎

Applied to the first sum in (242) this yields

$$
\begin{aligned}
&\frac{1}{\kappa}\sum_{k=1}^{\kappa}I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^n\big)\\
&\quad\le\frac{1}{\kappa}\sum_{k=1}^{\kappa}n_R\log\left(1+\frac{1}{n_R}\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]\frac{\mathsf{E}_k}{\sigma^2}\right)\\
&\qquad+\frac{1}{\kappa}\sum_{k=1}^{\kappa}h\big(\mathbf{H}_k\big|\mathbf{S}_1^n\big)-\frac{1}{\kappa}\sum_{k=1}^{\kappa}h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},\mathbf{S}_1^n\big) \qquad (244)\\
&\quad\le n_R\log\left(1+\frac{1}{n_R}\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]\frac{1}{\kappa}\sum_{k=1}^{\kappa}\frac{\mathsf{E}_k}{\sigma^2}\right)+h(\mathbf{H}_0)\\
&\qquad-\frac{1}{\kappa}h\big(\mathbf{H}_1^k\big|\mathbf{S}_1^n\big) \qquad (245)
\end{aligned}
$$

such that

$$\begin{aligned}
&\mathsf{R}_{\text{FB,ac}}(\mathsf{E}_\text{s}|\{\mathbf{S}_k\}) \\
&\le \frac{\kappa}{n}\, n_\text{R} \log\left(1 + \frac{1}{n_\text{R}}\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]\,\frac{1}{\kappa}\sum_{k=1}^{\kappa}\frac{\mathsf{E}_k}{\sigma^2}\right) + \frac{\kappa}{n}\,h(\mathbf{H}_0) \\
&\quad - \frac{1}{n}\,h\big(\mathbf{H}_1^\kappa\big|\mathbf{S}_1^n\big) + \frac{\log 2}{n} + \Pr\big[M \neq \hat{M}\big]\,\mathsf{R}_{\text{FB,ac}}(\mathsf{E}_\text{s}|\{\mathbf{S}_k\}) \\
&\quad + \frac{\epsilon_n}{n} + \frac{1}{n}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^n\big).
\end{aligned} \tag{246}$$

The expression (133) is then adapted as follows:

$$\begin{aligned}
&\mathsf{R}_{\text{FB,ac}}(\mathsf{E}_\text{s}|\{\mathbf{S}_k\}) \\
&\le \frac{\kappa}{n}\, n_\text{R} \log\left(1 + \frac{1}{n_\text{R}}\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]\,\frac{1}{\kappa}\sum_{k=1}^{\kappa}\frac{\mathsf{E}_k}{\sigma^2}\right) + \frac{\kappa}{n}\,h(\mathbf{H}_0) \\
&\quad - \frac{1}{n}\,h\big(\mathbf{H}_1^\kappa\big|\mathbf{S}_1^n\big) + \frac{\log 2}{n} + \Pr\big[M \neq \hat{M}\big]\,\mathsf{R}_{\text{FB,ac}}(\mathsf{E}_\text{s}|\{\mathbf{S}_k\}) \\
&\quad + \frac{\epsilon_n}{n} + \mathsf{H}_\text{b}\left(\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\right) \\
&\quad + \frac{n_\text{R}\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\log\left(1 + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big|A_0=1\big]}{n_\text{R}\,\mathsf{E}[\|\mathbf{H}_0\|^2]}\,\frac{t\mathsf{E}_\text{s}}{\sigma^2}\right) \\
&\quad + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\, I\big(\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^{\infty};\mathbf{H}_0\big|A_0=1\big) \\
&\quad + \frac{n-\kappa}{n}\,\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^n,A_k=0\big).
\end{aligned} \tag{247}$$

The derivation of the bound on $I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1},\mathbf{S}_1^n,A_k=0\big)$ is completely analogous to (134)–(228). After having n tending to infinity, we end up with

$$\begin{aligned}
&\mathsf{R}_{\text{FB,ac}}(\mathsf{E}_\text{s}|\{\mathbf{S}_k\}) \\
&\le \mathsf{H}_\text{b}\left(\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\right) \\
&\quad + \frac{n_\text{R}\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\log\left(1 + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big|A_0=1\big]}{n_\text{R}\,\mathsf{E}[\|\mathbf{H}_0\|^2]}\,\frac{t\mathsf{E}_\text{s}}{\sigma^2}\right) \\
&\quad + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\, I\big(\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^{\infty};\mathbf{H}_0\big|A_0=1\big) \\
&\quad + \frac{1}{1-\tilde{\delta}_4(t,\kappa)}\bigg(h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^{\kappa}\Big) \\
&\qquad\qquad - \log 2 + n_\text{R}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big] \\
&\qquad\qquad - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^{\infty}\big)\bigg) \\
&\quad + 2\,\mathsf{H}_\text{b}\left(\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\right) + \frac{2\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\cdot \\
&\qquad \cdot\bigg|h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^{\kappa},A_0=1\Big) \\
&\qquad\quad - \log 2 + n_\text{R}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|A_0=1\big] \\
&\qquad\quad - h\big(\mathbf{H}_0\big|A_0=1\big)\bigg| \\
&\quad + \frac{2t\mathsf{E}_\text{s}}{\eta} + \tilde{\delta}_2(\xi_\text{min},\kappa) + \tilde{\delta}_3(\xi_\text{min},\kappa) + \epsilon_\nu \\
&\quad + \sup_{0\le\beta\le1}\bigg\{(1-\beta)\log 2 + (1-\beta)\,h(\mathbf{H}_0|A_0=0) \\
&\qquad + \beta\log\Gamma\left(\mu,\frac{\nu}{\eta}\right) + \beta\,\frac{\nu}{\eta} \\
&\qquad + \mu\Big(\beta\log\eta - \mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|A_0=0\big] \\
&\qquad\qquad + (1-\beta)\log\big(2\,\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]\big) + e^{-1} \\
&\qquad\qquad - \beta\log\xi_\text{min}^2\Big) \\
&\qquad - (1-\beta)\,4n_\text{R}\,h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^{\infty},A_0=0\big) \\
&\qquad + (1-\beta)\,\frac{4n_\text{R}(n_\text{R}+1)}{e} \\
&\qquad + (1-\beta)\,\frac{16n_\text{R}\pi^{n_\text{R}}}{\Gamma(n_\text{R})(4n_\text{R}-1)^2} \\
&\qquad - (1-\beta)\,h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^{0},A_0=0\Big) \\
&\qquad + (1-\beta)\,4n_\text{R}^2\log\left(1 + \frac{2\pi e}{n_\text{R}}\,\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{1-\beta}\right) \\
&\qquad + (1-\beta)\,n_\text{R}\log\left(1 + \frac{2\xi_\text{min}^2}{n_\text{R}\,\sigma^2}\,\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{1-\beta}\right) \\
&\qquad + 4\,\mathsf{H}_\text{b}(\beta)\bigg\}
\end{aligned} \tag{248}$$

where $\tilde{\delta}_2(\xi_\text{min},\kappa)$, $\tilde{\delta}_3(\xi_\text{min},\kappa)$, and $\tilde{\delta}_4(t,\kappa)$ are correspondingly adapted versions of $\delta_2(\xi_\text{min},\kappa)$, $\delta_3(\xi_\text{min},\kappa)$, and $\delta_4(t,\kappa)$, respectively.

Before we conclude the proof in the same manner as in (234)–(239) using the same choice of the free parameters as given in (230)–(233), we point out that by an argument based on the tool of Appendix A, one can show that

$$\begin{aligned}
&h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^{\kappa}\Big) \\
&\quad = h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\kappa}^{-1},\mathbf{S}_{-\kappa}^{0}\Big)
\end{aligned} \tag{249}$$

and

$$h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^{\infty}\big) = h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^{0}\big). \tag{250}$$

Hence, we end up with the following upper bound on the fading number with feedback and acausal side-information at the transmitter:

$$\begin{aligned}
&\chi_{\text{FB,ac}}(\{\mathbf{H}_k\}|\{\mathbf{S}_k\}) \\
&\quad \le h_\lambda\Big(\hat{\mathbf{H}}_0\,e^{\mathrm{i}\Theta_0}\Big|\big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=-\infty}^{-1},\mathbf{S}_{-\infty}^{0}\Big) - \log 2 \\
&\qquad + n_\text{R}\,\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big] - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^{0}\big).
\end{aligned} \tag{251}$$

VIII. Upper Bounds on the Prelog with Feedback

In this section, we focus on a SISO nonregular Gaussian fading process $\{H_k\}$ with $\{H_k - d\}$ being a zero-mean, unit-variance, stationary, circularly symmetric, Gaussian process of arbitrary spectral distribution function $\mathsf{F}(\lambda)$, $-\frac{1}{2} < \lambda \le \frac{1}{2}$, and with $d \in \mathbb{C}$ denoting the spectral component. Again we allow a noiseless, but delayed feedback link from the receiver back to the transmitter. As input constraint we only consider the peak-power constraint (14).

We start as in Section VI-C with Fano's inequality:

$$\begin{aligned}\mathsf{R}^{\text{pp}}_{\text{FB}}(\mathsf{E}_\text{s}) &\leq \frac{1}{n} I\big(M; Y_1^n\big) + \frac{\log 2}{n} \\ &\quad + \Pr\big[M \neq \hat{M}\big]\, \mathsf{R}^{\text{pp}}_{\text{FB}}(\mathsf{E}_\text{s}) + \frac{\epsilon_n}{n} \quad (252)\\ &= \frac{1}{n}\sum_{k=1}^{n} I\big(M; Y_k \big| Y_1^{k-1}\big) + \frac{\log 2}{n} \\ &\quad + \Pr\big[M \neq \hat{M}\big]\, \mathsf{R}^{\text{pp}}_{\text{FB}}(\mathsf{E}_\text{s}) + \frac{\epsilon_n}{n}. \quad (253)\end{aligned}$$

Next we bound each term in the sum separately:

$$\begin{aligned}&I\big(M; Y_k \big| Y_1^{k-1}\big) \\ &\quad \leq I\big(M, X_1^k; Y_k \big| Y_1^{k-1}\big) \quad (254)\\ &\quad = I\big(X_1^k; Y_k \big| Y_1^{k-1}\big) + \underbrace{I\big(M; Y_k \big| Y_1^{k-1}, X_1^k\big)}_{=0} \quad (255)\\ &\quad = I\big(X_1^k; Y_k \big| Y_1^{k-1}\big) \quad (256)\\ &\quad \leq I\big(X_1^k, Y_1^{k-1}; Y_k\big) \quad (257)\\ &\quad = I(X_k; Y_k) + I\big(X_1^{k-1}, Y_1^{k-1}; Y_k \big| X_k\big) \quad (258)\end{aligned}$$

where (256) can be shown using the tool from Appendix A (see Fig. 9). The second term in (258) is bounded as follows:

$$\begin{aligned}&I\big(X_1^{k-1}, Y_1^{k-1}; Y_k \big| X_k\big) \\ &\quad \leq I\big(X_1^{k-1}, Y_1^{k-1}; Y_k, H_k \big| X_k\big) \quad (259)\\ &\quad = I\big(X_1^{k-1}, Y_1^{k-1}; H_k \big| X_k\big) \\ &\qquad + \underbrace{I\big(X_1^{k-1}, Y_1^{k-1}; Y_k \big| H_k, X_k\big)}_{=0} \quad (260)\\ &\quad = I\big(X_1^{k-1}, Y_1^{k-1}; H_k \big| X_k\big) \quad (261)\\ &\quad = h(H_k|X_k) - h\big(H_k \big| Y_1^{k-1}, X_1^k\big) \quad (262)\\ &\quad = h(H_k|X_k) - \mathsf{E}_{X_1^k}\big[h\big(H_k \big| Y_1^{k-1}, \{X_\ell = x_\ell\}_{\ell=1}^k\big)\big] \quad (263)\\ &\quad = h(H_k|X_k) - \mathsf{E}_{X_1^k}\big[h\big(H_k \big| \{H_\ell x_\ell + Z_\ell\}_{\ell=1}^{k-1}\big)\big] \quad (264)\\ &\quad \leq h(H_k) - \inf_{\substack{x_\ell:\ |x_\ell| \leq \sqrt{\mathsf{E}_\text{s}} \\ \ell=1,\ldots,k-1}} h\big(H_k \big| \{H_\ell x_\ell + Z_\ell\}_{\ell=1}^{k-1}\big) \quad (265)\\ &\quad = h(H_k) - \inf_{\substack{x_\ell:\ |x_\ell| \leq \sqrt{\mathsf{E}_\text{s}} \\ \ell=1,\ldots,k-1}} h\left(H_k \middle| \left\{H_\ell + \frac{Z_\ell}{x_\ell}\right\}_{\ell=1}^{k-1}\right) \quad (266)\\ &\quad = h(H_k) - h\left(H_k \middle| \left\{H_\ell + \frac{Z_\ell}{\sqrt{\mathsf{E}_\text{s}}}\right\}_{\ell=1}^{k-1}\right) \quad (267)\\ &\quad = I\left(\left\{H_\ell + \frac{Z_\ell}{\sqrt{\mathsf{E}_\text{s}}}\right\}_{\ell=1}^{k-1}; H_k\right) \quad (268)\\ &\quad \leq I\left(\left\{H_\ell + \frac{Z_\ell}{\sqrt{\mathsf{E}_\text{s}}}\right\}_{\ell=-\infty}^{k-1}; H_k\right) \quad (269)\\ &\quad = \log \frac{1}{\epsilon^2_{\text{pred}}\big(\frac{\sigma^2}{\mathsf{E}_\text{s}}\big)}. \quad (270)\end{aligned}$$

Note that in (264) we use the fact that conditional on the past outputs Y_1^{k-1}, the current fading H_k is independent of the inputs X_1^k. In (265), we bound by dropping some conditioning argument and by replacing an expectation by an infimum; and (267) holds because the infimum is achieved for $x_\ell = \sqrt{\mathsf{E}_\text{s}}$ (and an arbitrary phase since Z_ℓ is circularly symmetric).

In (270), $\epsilon^2_{\text{pred}}(\delta^2)$ denotes the noisy prediction error as described in [7, Equation (11)]: When predicting H_0 based on the noisy observations

$$H_{-1} + W_{-1}, H_{-2} + W_{-2}, H_{-3} + W_{-3}, \ldots \quad (271)$$

where $\{W_k\}$ is a sequence of random variables that are IID $\sim \mathcal{N}_{\mathbb{C}}\big(0, \delta^2\big)$, an optimal predictor will achieve a mean-squared error

$$\epsilon^2_{\text{pred}}(\delta^2) = \exp\left(\int_{-\frac{1}{2}}^{\frac{1}{2}} \log\big(\mathsf{F}'(\lambda) + \delta^2\big)\,\mathrm{d}\lambda\right) - \delta^2. \quad (272)$$

Hence, we have

$$\begin{aligned}\mathsf{R}^{\text{pp}}_{\text{FB}}(\mathsf{E}_\text{s}) &\leq \frac{1}{n}\sum_{k=1}^{n} I(X_k; Y_k) + \log \frac{1}{\epsilon^2_{\text{pred}}\big(\frac{\sigma^2}{\mathsf{E}_\text{s}}\big)} + \frac{\log 2}{n} \\ &\quad + \Pr\big[M \neq \hat{M}\big]\, \mathsf{R}^{\text{pp}}_{\text{FB}}(\mathsf{E}_\text{s}) + \frac{\epsilon_n}{n}. \quad (273)\end{aligned}$$

It only remains to show that the first term grows like $\log\log \mathsf{E}_\text{s}$. This basically follows from the upper bound given in Section VI. Unfortunately, we need to be careful with the order of limits. So, to avoid any hand-waving arguments, we derive a rough bound showing a double-logarithmic growth (where we not need to worry about the correct second-order terms).

Similarly to Section VI-C, we introduce an indicator random variable B_k,

$$B_k \triangleq \begin{cases} 1 & \text{if } |X_k| \geq \xi_{\min} \\ 0 & \text{otherwise} \end{cases} \quad (274)$$

with $\beta_k \triangleq \Pr[B_k = 1]$, and obtain

$$\begin{aligned}I(X_k; Y_k) &\leq I(X_k, B_k; Y_k) \quad (275)\\ &= I(B_k; Y_k) + I(X_k; Y_k|B_k) \quad (276)\\ &\leq \mathsf{H}_\text{b}(\beta_k) + \beta_k\, I(X_k; Y_k|B_k = 1) \\ &\quad + (1-\beta_k)\, I(X_k; Y_k|B_k = 0). \quad (277)\end{aligned}$$

The last term we bound as follows:

$$\begin{aligned}&I(X_k; Y_k|B_k = 0) \\ &\quad = h(Y_k|B_k = 0) - h(Y_k|X_k, B_k = 0) \quad (278)\\ &\quad \leq \log\big(\pi e\big(\mathsf{E}\big[|H_k|^2 \big| B_k = 0\big]\xi^2_{\min} + \sigma^2\big)\big) \\ &\qquad - \underbrace{h(Y_k|X_k, H_k X_k, B_k = 0)}_{=\log(\pi e \sigma^2)} \quad (279)\\ &\quad = \log\left(1 + \mathsf{E}\big[|H_k|^2 \big| B_k = 0\big]\frac{\xi^2_{\min}}{\sigma^2}\right) \quad (280)\\ &\quad \leq \log\left(1 + \frac{\mathsf{E}\big[|H_k|^2\big]}{1-\beta_k}\,\frac{\xi^2_{\min}}{\sigma^2}\right) \quad (281)\end{aligned}$$

where the last inequality follows from Lemma 9.

To bound the second term in (277), we now rely on a conditional version of Lemma 2 with $X = X_k$ and $T = Y_k$ and with the choice $\nu = 0$ and $\mu \leq 1$:

$$\begin{aligned}&I(X_k; Y_k|B_k = 1) \\ &\leq -h(Y_k|X_k, B_k = 1) + \log\pi + \mu\log\eta + \log\Gamma(\mu) \\ &\quad + (1-\mu)\,\mathsf{E}\big[\log|Y_k|^2 \big| B_k = 1\big] + \frac{1}{\eta}\,\mathsf{E}\big[|Y_k|^2 \big| B_k = 1\big] \\ &\hspace{20em} (282)\end{aligned}$$

$$
\begin{aligned}
&= -\mathsf{E}_{X_k}\Big[\log\Big(\pi e\big(\mathsf{E}\big[|H_k|^2\big|X_k = x_k, B_k = 1\big]|X_k|^2 + \sigma^2\big)\Big) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Big| B_k = 1\Big] \\
&\quad + \log\pi + \mu\log\eta + \log\Gamma(\mu) \\
&\quad + (1-\mu)\,\mathsf{E}\big[\log|H_k X_k + Z_k|^2\big|B_k = 1\big] \\
&\quad + \frac{1}{\eta}\,\mathsf{E}\big[|H_k X_k + Z_k|^2\big|B_k = 1\big] \qquad (283) \\
&= -\mathsf{E}_{X_k}\Bigg[\log\Bigg(\underbrace{\mathsf{E}\big[|H_k|^2\big|X_k = x_k, B_k = 1\big]}_{\geq 0} + \frac{\sigma^2}{|X_k|^2}\Bigg) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Bigg| B_k = 1\Bigg] \\
&\quad - 1 - \mathsf{E}\big[\log|X_k|^2\big|B_k = 1\big] + \mu\log\eta + \log\Gamma(\mu) \\
&\quad + (1-\mu)\,\mathsf{E}\big[\log|X_k|^2\big|B_k = 1\big] \\
&\quad + (1-\mu)\,\mathsf{E}\left[\log\left|H_k + \frac{Z_k}{X_k}\right|^2\middle|B_k = 1\right] \\
&\quad + \frac{1}{\eta}\,\mathsf{E}\big[|H_k|^2|X_k|^2\big|B_k = 1\big] + \frac{\sigma^2}{\eta} \qquad (284) \\
&\leq -1 - \log\left(\frac{\sigma^2}{\xi_{\min}^2}\right) + \mu\log\eta + \log\Gamma(\mu) \\
&\quad - \mu\,\mathsf{E}\big[\log|X_k|^2\big|B_k = 1\big] \\
&\quad + (1-\mu)\log\mathsf{E}\left[|H_k|^2 + \frac{|Z_k|^2}{\xi_{\min}^2}\middle|B_k = 1\right] \\
&\quad + \frac{1}{\eta}\,\mathsf{E}\big[|H_k|^2\big|B_k = 1\big]\,\mathsf{E}_s + \frac{\sigma^2}{\eta} \qquad (285) \\
&\leq -1 - \log\sigma^2 + \mu\log\eta + \log\Gamma(\mu) + (1-\mu)\log\xi_{\min}^2 \\
&\quad + (1-\mu)\log\left(\frac{\mathsf{E}\big[|H_k|^2\big]}{\beta_k} + \frac{\sigma^2}{\xi_{\min}^2}\right) + \frac{\mathsf{E}_s\,\mathsf{E}\big[|H_k|^2\big]}{\eta\beta_k} \\
&\quad + \frac{\sigma^2}{\eta}. \qquad (286)
\end{aligned}
$$

Here, (282) follows from Lemma 2; in (285), we use Jensen's inequality and the fact that conditional on $B_k = 1$, we have $\xi_{\min}^2 \leq |X_k|^2 \leq \mathsf{E}_s$; and in (286), we again use $|X_k|^2 \geq \xi_{\min}^2$ (conditional on $B_k = 1$) and we apply Lemma 9.

Hence, from (277), (281), and (286), and taking the supremum over β_k, we finally obtain

$$
\begin{aligned}
I(X_k;Y_k) \leq \sup_{0\leq\beta\leq 1}\Bigg\{&(1-\beta)\log\left(1 + \frac{\mathsf{E}\big[|H_0|^2\big]}{1-\beta}\,\frac{\xi_{\min}^2}{\sigma^2}\right) \\
&+ \mathsf{H}_b(\beta) - \beta - \beta\log\sigma^2 + \beta\mu\log\eta \\
&+ \beta\log\Gamma(\mu) + \beta(1-\mu)\log\xi_{\min}^2 \\
&+ \beta(1-\mu)\log\left(\frac{\mathsf{E}\big[|H_0|^2\big]}{\beta} + \frac{\sigma^2}{\xi_{\min}^2}\right) \\
&+ \frac{\mathsf{E}_s\,\mathsf{E}\big[|H_0|^2\big]}{\eta} + \frac{\beta\sigma^2}{\eta}\Bigg\}. \qquad (287)
\end{aligned}
$$

Note that (due to stationarity) this expression does not depend on k or n anymore. So, plugging this back into (273) and letting n tend to infinity, we obtain

$$
\begin{aligned}
\mathsf{R}_{\mathrm{FB}}^{\mathrm{pp}}(\mathsf{E}_s) \leq \sup_{0\leq\beta\leq 1}\Bigg\{&(1-\beta)\log\left(1 + \frac{\mathsf{E}\big[|H_0|^2\big]}{1-\beta}\,\frac{\xi_{\min}^2}{\sigma^2}\right) \\
&+ \mathsf{H}_b(\beta) - \beta - \beta\log\sigma^2 + \beta\mu\log\eta \\
&+ \beta\log\Gamma(\mu) + \beta(1-\mu)\log\xi_{\min}^2 \\
&+ \beta(1-\mu)\log\left(\frac{\mathsf{E}\big[|H_0|^2\big]}{\beta} + \frac{\sigma^2}{\xi_{\min}^2}\right) \\
&+ \frac{\mathsf{E}_s\,\mathsf{E}\big[|H_0|^2\big]}{\eta} + \frac{\beta\sigma^2}{\eta}\Bigg\} \\
&+ \log\frac{1}{\epsilon_{\mathrm{pred}}^2\left(\frac{\sigma^2}{\mathsf{E}_s}\right)}. \qquad (288)
\end{aligned}
$$

Choosing

$$\mu \triangleq \frac{1}{\log\mathsf{E}_s} \qquad (289)$$

$$\eta \triangleq \mathsf{E}_s\log\mathsf{E}_s \qquad (290)$$

$$\xi_{\min} \triangleq \sqrt{\log\log\mathsf{E}_s} \qquad (291)$$

and using the same argument as in Section VI-G, we see that for E_s large enough, the supremum is achieved for $\beta = 1$. Hence,

$$
\begin{aligned}
&\mathsf{R}_{\mathrm{FB}}^{\mathrm{pp}}(\mathsf{E}_s) \\
&\leq -1 - \log\sigma^2 + 1 + \frac{\log\log\mathsf{E}_s}{\log\mathsf{E}_s} + \log\Gamma\left(\frac{1}{\log\mathsf{E}_s}\right) \\
&\quad + \left(1 - \frac{1}{\log\mathsf{E}_s}\right)\Bigg(\log\log\log\mathsf{E}_s \\
&\qquad\qquad + \log\left(\mathsf{E}\big[|H_0|^2\big] + \frac{\sigma^2}{\log\log\mathsf{E}_s}\right)\Bigg) \\
&\quad + \frac{\mathsf{E}\big[|H_0|^2\big]}{\log\mathsf{E}_s} + \frac{\sigma^2}{\mathsf{E}_s\log\mathsf{E}_s} + \log\frac{1}{\epsilon_{\mathrm{pred}}^2\left(\frac{\sigma^2}{\mathsf{E}_s}\right)} \qquad (292) \\
&= \log\log\mathsf{E}_s + \log\log\log\mathsf{E}_s + O(1) + \log\frac{1}{\epsilon_{\mathrm{pred}}^2\left(\frac{\sigma^2}{\mathsf{E}_s}\right)}. \qquad (293)
\end{aligned}
$$

Here, $O(1)$ expresses some unspecified terms that are bounded in E_s, and we used that

$$\log\Gamma\left(\frac{1}{\log\mathsf{E}_s}\right) = \log\log\mathsf{E}_s + o(1) \qquad (294)$$

where $o(1)$ represent terms that tend to zero as E_s tends to infinity. Note that this upper bound holds for any system, i.e., it also holds for a capacity-achieving system. Hence, we have succeeded in deriving a bound similar to [7, Equation (48)] (it contains an additional term $\log\log\log\mathsf{E}_s$). Note that under the assumption that

$$\lim_{\delta^2\downarrow 0}\frac{\epsilon_{\mathrm{pred}}^2(\delta^2)}{\delta^2} = \infty \qquad (295)$$

we have

$$\log\frac{1}{\epsilon_{\mathrm{pred}}^2\left(\frac{\sigma^2}{\mathsf{E}_s}\right)} + O(1) = \log\frac{1}{\epsilon_{\mathrm{pred}}^2\left(\frac{\sigma^2}{\mathsf{E}_s}\right) + \frac{\sigma^2}{\mathsf{E}_s}} + O(1) \qquad (296)$$

(compare with [7, Equation (54)]).

Next, we will derive an upper bound that is tight for very small errors when predicting the actual fading level from noisy observations of its past. To that end note that $\mathsf{C}_{\mathrm{FB}}^{\mathrm{pp}}(\mathsf{E}_s)$ is always upper-bounded by the achievable rate in the case when both receiver and transmitter have *perfect* knowledge about

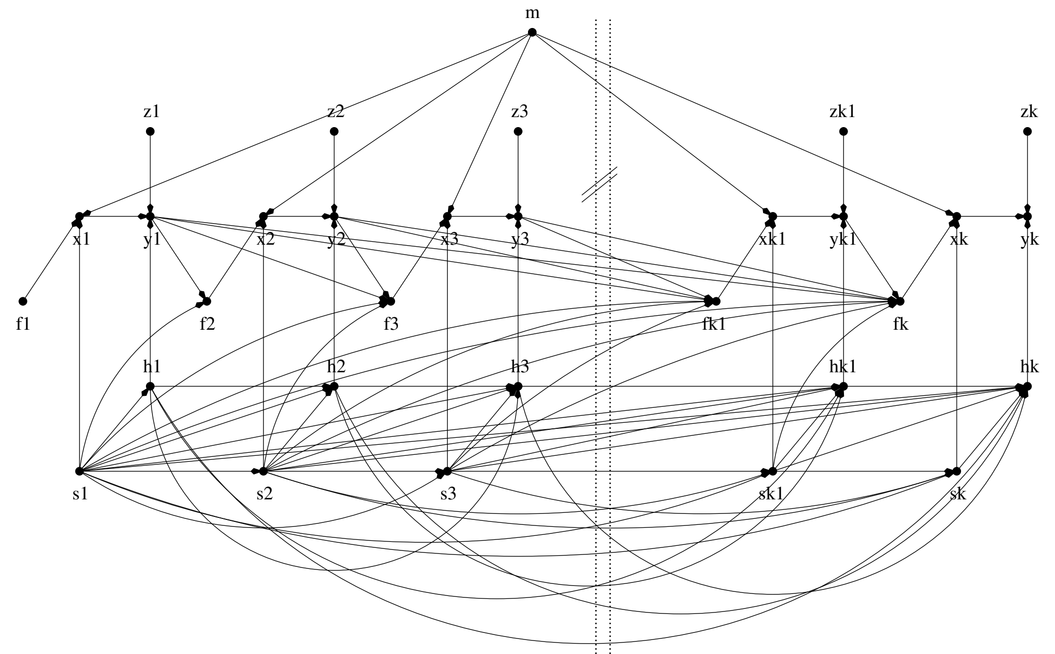

Fig. 2. The causality graph of the causal-order expansion given in (303).

the actual fading realization. In such a situation the channel appears memoryless and therefore feedback does not increase capacity. Note that since a peak-power constraint is in effect, there is no water-filling possible, but in an optimal scheme, the transmitter will transmit always with highest allowed power E_s. Thus

$$\begin{aligned} \mathsf{C}_{\mathsf{FB}}^{\mathrm{pp}}(\mathsf{E}_\mathrm{s}) &\leq \mathsf{R}_{\mathrm{PSI@Tx}}(\mathsf{E}_\mathrm{s}) && (297)\\ &\leq \mathsf{E}\left[\log\left(1+\frac{\mathsf{E}_\mathrm{s}\cdot|H_k|^2}{\sigma^2}\right)\right] && (298)\\ &\leq \log\left(1+\frac{\mathsf{E}_\mathrm{s}\cdot \mathsf{E}\left[|H_k|^2\right]}{\sigma^2}\right) && (299)\\ &= \log\left(1+\left(|d|^2+1\right)\frac{\mathsf{E}_\mathrm{s}}{\sigma^2}\right) && (300)\\ &= \log\frac{\mathsf{E}_\mathrm{s}}{\sigma^2}+\log\left(|d|^2+1\right)+o(1) && (301) \end{aligned}$$

where in (298) we upper-bound further by allowing Gaussian inputs (that are not peak-limited). The bound (301) is identical to [7, Equation (49)].

Furthermore note that any lower bound on the capacity without feedback trivially also is a lower bound on the capacity with feedback, i.e., the lower bounds given in [7, Section VI] continue to hold in our context. Hence, we can now easily adapt the derivations of [7, Section VIII] to the situation with noiseless feedback, such that the results given in [7] also hold in the case with noiseless feedback.

APPENDIX A INDEPENDENCE AND CAUSALITY

In [11], [12], James L. Massey introduces a way of graphically determining independence of random variables based on *causal interpretations*. A causal interpretation is an ordered list of random variables, where the choice of a specific order is based on the causality of the system. Loosely speaking, we like to think of some random variables being generated "first" and some "later based on the first." Note that *a priori* every ordered list is a valid causal interpretation, but some choices will be more useful than others keeping in mind the engineering understanding of causality.

In our case of the random variables depicted in Fig. 1

$$\left(M, X_1^k, \mathbf{Y}_1^k, \mathbf{H}_1^k, \mathbf{Z}_1^k, \mathbf{S}_1^k, \mathbf{F}_1^k\right) \tag{302}$$

we choose the following causal interpretation:

$$\begin{aligned}(M, \mathbf{Z}_1, \ldots, \mathbf{Z}_k, \mathbf{S}_1, \mathbf{H}_1, \mathbf{S}_2, \mathbf{H}_2, \ldots, \mathbf{S}_k, \mathbf{H}_k,\\ \mathbf{F}_1, X_1, \mathbf{Y}_1, \mathbf{F}_2, X_2, \mathbf{Y}_2, \ldots, \mathbf{F}_k, X_k, \mathbf{Y}_k).\end{aligned} \tag{303}$$

For some given sets $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ of random variables, we now would like to know whether $\mathcal{A}$ is independent of $\mathcal{B}$ when conditioned on $\mathcal{C}$. In order to answer this question, we make use of the "Markov structure" of the random variables given by the causal interpretation of the system. For example, for (303), we think of

- M being generated independently;
- $\mathbf{Z}_1$ being generated independently;

$\vdots$

- $\mathbf{Z}_k$ being generated independently;
- $\mathbf{S}_1$ being generated independently;
- $\mathbf{H}_1$ being generated based on $\mathbf{S}_1$;
- $\mathbf{S}_2$ being generated based on $\mathbf{S}_1$;
- $\mathbf{H}_2$ being generated based on $\mathbf{H}_1, \mathbf{S}_1^2$;

$\vdots$

- $\mathbf{S}_k$ being generated based on $\mathbf{S}_1^{k-1}$;
- $\mathbf{H}_k$ being generated based on $\mathbf{H}_1^{k-1}, \mathbf{S}_1^k$;
- $\mathbf{F}_1$ being generated independently;

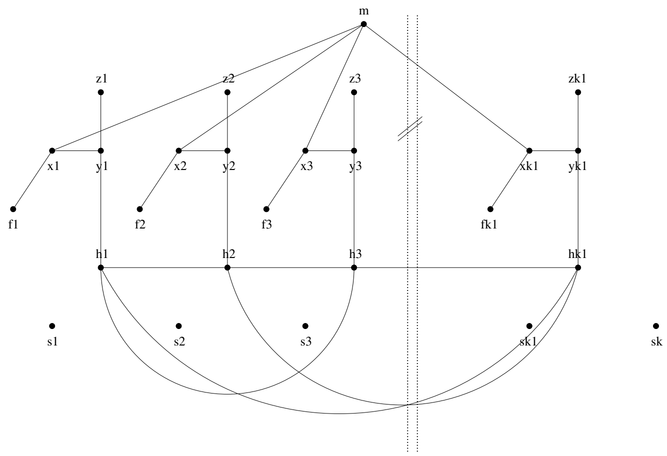

Fig. 3. The causally relevant subgraph of (303) showing the independence of M and $\mathbf{S}_k$ when conditioned on $\left(\mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}\right)$.

- X_1 being generated based on $M, \mathbf{F}_1, \mathbf{S}_1$;
- $\mathbf{Y}_1$ being generated based on $X_1, \mathbf{H}_1, \mathbf{Z}_1$;
- $\mathbf{F}_2$ being generated based on $\mathbf{Y}_1, \mathbf{S}_1$;
- X_2 being generated based on $M, \mathbf{F}_2, \mathbf{S}_2$;
- $\mathbf{Y}_2$ being generated based on $X_2, \mathbf{H}_2, \mathbf{Z}_2$;

  $\vdots$
- $\mathbf{F}_k$ being generated based on $\mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}$;
- X_k being generated based on $M, \mathbf{F}_k, \mathbf{S}_k$; and
- $\mathbf{Y}_k$ being generated based on $X_k, \mathbf{H}_k, \mathbf{Z}_k$.

Massey calls this a *causal-order expansion* of (302). It can easily be depicted graphically in a *causality graph*, which is a directed graph with an edge from vertex V_i to V_j if, and only if, the generation of V_j is directly based on V_i. For the example (303) the corresponding graph is shown in Fig. 2.

In order to prove the independence of $\mathcal{A}$ and $\mathcal{B}$ when conditioned on $\mathcal{C}$, we then consider the subgraph of Fig. 2 that is *causally relevant* to $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$, i.e., we consider only nodes (and the corresponding edges stemming from them) that are either member of $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$ or causally prior to $\mathcal{A} \cup \mathcal{B} \cup \mathcal{C}$. Then we delete all edges leaving any component of $\mathcal{C}$. If now all components of $\mathcal{A}$ are unconnected (when the edges are considered without direction) to the components of $\mathcal{B}$, the conditional independence is proven. Note that this graphical proof only constitutes a sufficient condition for independence, i.e., if the proof fails, then this does not imply that $\mathcal{A}$ and $\mathcal{B}$ must be conditionally dependent conditional on $\mathcal{C}$.

In Figs. 3–13, several independence claims are proven that are used in this paper and that are based on the causal interpretation (303).

APPENDIX B
PROOF OF LEMMA 8

We derive the following bound:

$$
\begin{aligned}
& I\big(X_k; \mathbf{Z}_{k-\kappa}^{k-1}, \mathbf{Z}_{k+1}^{k+\kappa} \big| \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}, X_{k-\kappa}^{k-1}\big) \\
&\quad \le I\big(X_k; \mathbf{Z}_{k-\kappa}^{k+\kappa} \big| \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}, X_{k-\kappa}^{k-1}\big) && (304) \\
&\quad = h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa} \big| \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}, X_{k-\kappa}^{k-1}\big) \\
&\qquad - h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa} \big| \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}, X_{k-\kappa}^{k}\big) && (305)
\end{aligned}
$$

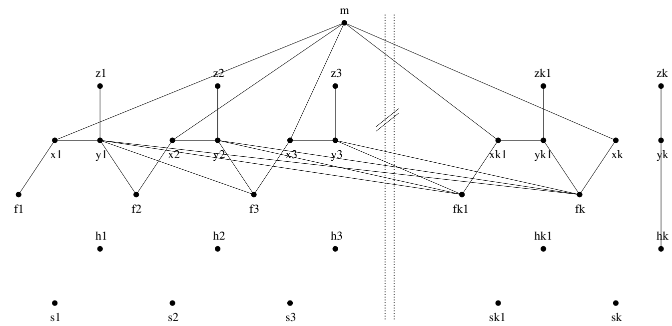

Fig. 4. The causally relevant subgraph of (303) showing the independence of $\left(M, \mathbf{Y}_1^{k-1}\right)$ and $\mathbf{Y}_k$ when conditioned on $\left(X_k, \mathbf{H}_1^{k-1}, \mathbf{S}_1^{k}\right)$.

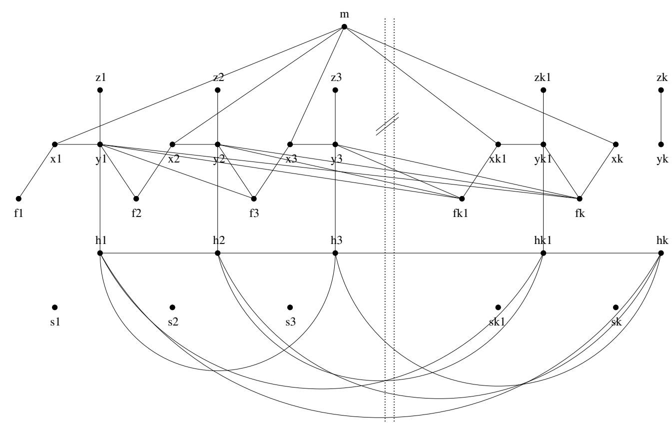

Fig. 5. The causally relevant subgraph of (303) showing the independence of $\mathbf{Y}_k$ and $\mathbf{H}_1^{k-1}$ when conditioned on $\left(X_k, \mathbf{H}_k, \mathbf{S}_1^{k}\right)$.

$$
\begin{aligned}
& \le h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\big) - h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa} \big| \mathbf{Y}_{k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}, X_{k-\kappa}^{k+\kappa}\big) && (306) \\
& = h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\big) \\
& \quad - h\left(\mathbf{Z}_{k-\kappa}^{k+\kappa} \middle| \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{X_\ell}\right\}_{\ell=k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}, X_{k-\kappa}^{k+\kappa}\right) && (307) \\
& \le h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\big) \\
& \quad - \inf_{\substack{x_\ell \colon |x_\ell| \ge x_{\min} \\ \ell = k-\kappa, \ldots, k+\kappa}} h\left(\mathbf{Z}_{k-\kappa}^{k+\kappa} \middle| \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_\ell}\right\}_{\ell=k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}\right) \\
& && (308) \\
& = h\big(\mathbf{Z}_{k-\kappa}^{k+\kappa}\big) - h\left(\mathbf{Z}_{k-\kappa}^{k+\kappa} \middle| \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}\right) \\
& && (309) \\
& = I\left(\mathbf{Z}_{k-\kappa}^{k+\kappa}; \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=k-\kappa}^{k+\kappa}, \mathbf{S}_{k-\kappa}^{k+\kappa}\right) && (310) \\
& = I\left(\left\{\frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=k-\kappa}^{k+\kappa}; \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=k-\kappa}^{k+\kappa} \middle| \mathbf{S}_{k-\kappa}^{k+\kappa}\right) && (311) \\
& = I\left(\left\{\frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=-\kappa}^{\kappa}; \left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=-\kappa}^{\kappa} \middle| \mathbf{S}_{-\kappa}^{\kappa}\right) && (312) \\
& = h\left(\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=-\kappa}^{\kappa} \middle| \mathbf{S}_{-\kappa}^{\kappa}\right) - h\big(\mathbf{H}_{-\kappa}^{\kappa} \big| \mathbf{S}_{-\kappa}^{\kappa}\big) && (313) \\
& \triangleq \delta_1(x_{\min}, \kappa). && (314)
\end{aligned}
$$

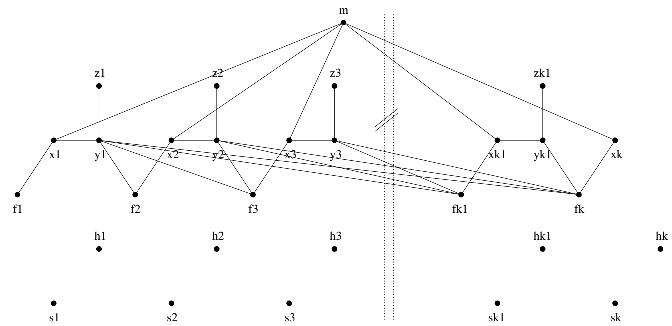

Fig. 6. The causally relevant subgraph of (303) showing the independence of X_k and $\mathbf{H}_k$ when conditioned on $\left(\mathbf{H}_1^{k-1}, \mathbf{S}_1^k\right)$.

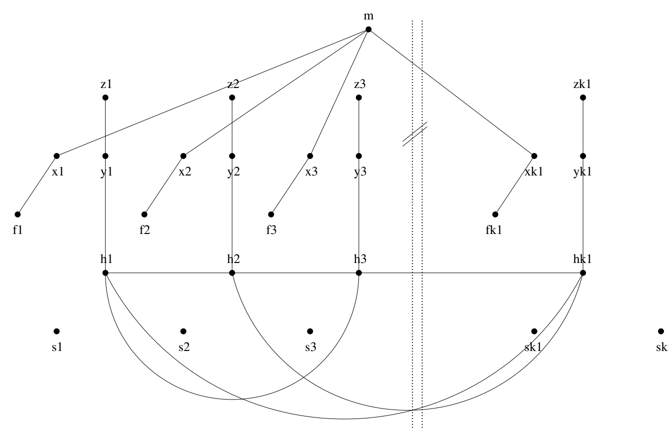

Fig. 7. The causally relevant subgraph of (303) showing the independence of $\mathbf{S}_k$ and M when conditioned on $\left(\mathbf{Y}_1^{k-1}, X_1^{k-1}, \mathbf{S}_1^{k-1}\right)$.

Here (304) follows from adding $\mathbf{Z}_k$ to the arguments of the mutual information; (306) follows from conditioning that reduces entropy; in (308) we replace the expectation over $X_{k-\kappa}^{k+\kappa}$ by a corresponding minimization; for (309) we note that the minimum is achieved for $x_\ell = x_{\min}$; and (312) follows from stationarity.

From [6, Lemma 6.11], [15, Lemma A.19] we conclude that the expression

$$h\left(\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_{\min}}\right\}_{\ell=-\kappa}^{\kappa} \middle| \mathbf{S}_{-\kappa}^{\kappa}\right) \tag{315}$$

converges monotonically in $x_{\min}$ to $h\left(\mathbf{H}_{-\kappa}^{\kappa} \middle| \mathbf{S}_{-\kappa}^{\kappa}\right)$.

Appendix C
Proof of Lemma 10

By the chain rule, we have

$$\begin{aligned} I\left(M; \mathbf{Y}_k, \mathbf{S}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}\right) = {} & I\left(M; \mathbf{S}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}\right) \\ & + I\left(M; \mathbf{Y}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k}\right) \end{aligned} \tag{316}$$

where for the first term we obtain

$$\begin{aligned} & I\left(M; \mathbf{S}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}\right) \\ & \quad \le I\left(M, X_1^{k-1}; \mathbf{S}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}\right) && (317) \\ & \quad = I\left(X_1^{k-1}; \mathbf{S}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}\right) && (318) \\ & \quad = I\left(X_1^{k-1}, \mathbf{Y}_1^{k-1}; \mathbf{S}_k \middle| \mathbf{S}_1^{k-1}\right) - \underbrace{I\left(\mathbf{Y}_1^{k-1}; \mathbf{S}_k \middle| \mathbf{S}_1^{k-1}\right)}_{\ge 0} && (319) \\ & \quad \le I\left(X_1^{k-1}, \mathbf{Y}_1^{k-1}, \mathbf{H}_1^{k-1}; \mathbf{S}_k \middle| \mathbf{S}_1^{k-1}\right) && (320) \\ & \quad = I\left(\mathbf{H}_1^{k-1}; \mathbf{S}_k \middle| \mathbf{S}_1^{k-1}\right) \\ & \qquad + I\left(X_1^{k-1}, \mathbf{Y}_1^{k-1}; \mathbf{S}_k \middle| \mathbf{H}_1^{k-1}, \mathbf{S}_1^{k-1}\right) && (321) \\ & \quad = I\left(\mathbf{H}_1^{k-1}; \mathbf{S}_k \middle| \mathbf{S}_1^{k-1}\right). && (322) \end{aligned}$$

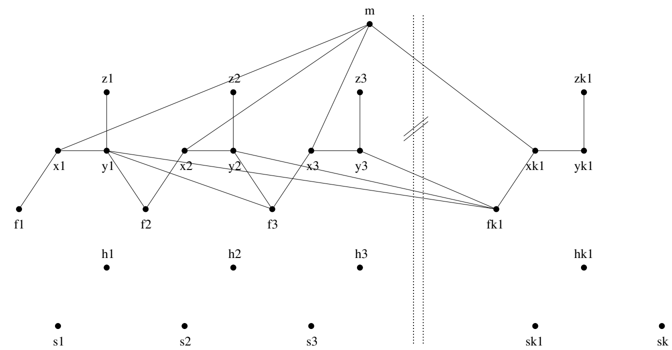

Fig. 8. The causally relevant subgraph of (303) showing the independence of $\mathbf{S}_k$ and $\left(X_1^{k-1}, \mathbf{Y}_1^{k-1}\right)$ when conditioned on $\left(\mathbf{H}_1^{k-1}, \mathbf{S}_1^{k-1}\right)$.

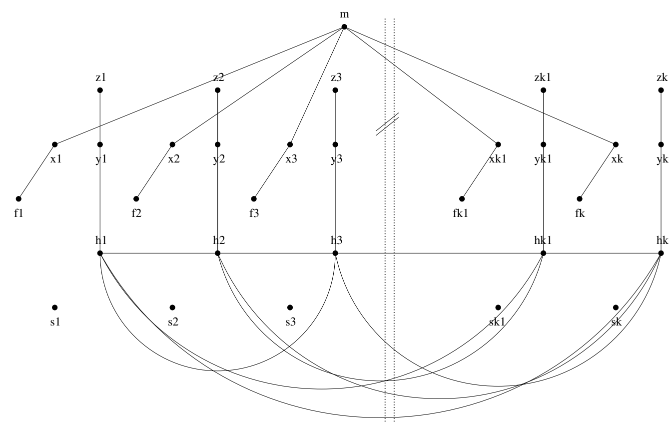

Fig. 9. The causally relevant subgraph of (303) showing the independence of $\mathbf{Y}_k$ and M when conditioned on $\left(\mathbf{Y}_1^{k-1}, X_1^{k}, \mathbf{S}_1^{k}\right)$.

Here (317) and (320) follow from adding an additional argument to the mutual information functional, and (318) and (322) are proven in Appendix A (see Figs. 7 and 8, respectively).

We bound the second term in (316) as follows:

$$\begin{aligned} & I\left(M; \mathbf{Y}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k}\right) \\ & \quad = h\left(\mathbf{Y}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k}\right) - h\left(\mathbf{Y}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k}, M\right) && (323) \\ & \quad = h\left(\mathbf{Y}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k}\right) - h\left(\mathbf{Y}_k \middle| \mathbf{Y}_1^{k-1}, X_1^k, \mathbf{S}_1^{k}, M\right) && (324) \\ & \quad = h\left(\mathbf{Y}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k}\right) - h\left(\mathbf{Y}_k \middle| \mathbf{Y}_1^{k-1}, X_1^k, \mathbf{S}_1^{k}\right) && (325) \\ & \quad = I\left(X_1^k; \mathbf{Y}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k}\right) && (326) \\ & \quad = I\left(X_1^k; \mathbf{Y}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k}, \mathbf{F}_1^k\right) && (327) \\ & \quad = I\left(X_1^k, \mathbf{Y}_1^{k-1}; \mathbf{Y}_k \middle| \mathbf{S}_1^{k}, \mathbf{F}_1^k\right) - I\left(\mathbf{Y}_1^{k-1}; \mathbf{Y}_k \middle| \mathbf{S}_1^{k}, \mathbf{F}_1^k\right) && (328) \\ & \quad \le I\left(X_1^k, \mathbf{Y}_1^{k-1}, \mathbf{H}_1^{k-1}; \mathbf{Y}_k \middle| \mathbf{S}_1^{k}, \mathbf{F}_1^k\right) && (329) \\ & \quad = I\left(X_k, \mathbf{H}_1^{k-1}; \mathbf{Y}_k \middle| \mathbf{S}_1^{k}, \mathbf{F}_1^k\right) \\ & \qquad + I\left(X_1^{k-1}, \mathbf{Y}_1^{k-1}; \mathbf{Y}_k \middle| X_k, \mathbf{H}_1^{k-1}, \mathbf{S}_1^{k}, \mathbf{F}_1^k\right) && (330) \end{aligned}$$

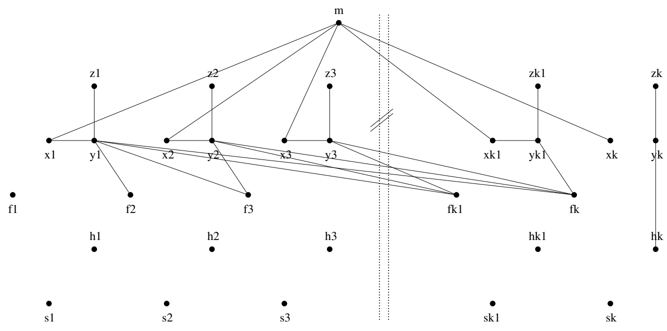

Fig. 10. The causally relevant subgraph of (303) showing the independence of $\mathbf{Y}_k$ and $\left(X_1^{k-1}, \mathbf{Y}_1^{k-1}\right)$ when conditioned on $\left(X_k, \mathbf{H}_1^{k-1}, \mathbf{S}_1^k, \mathbf{F}_1^k\right)$.

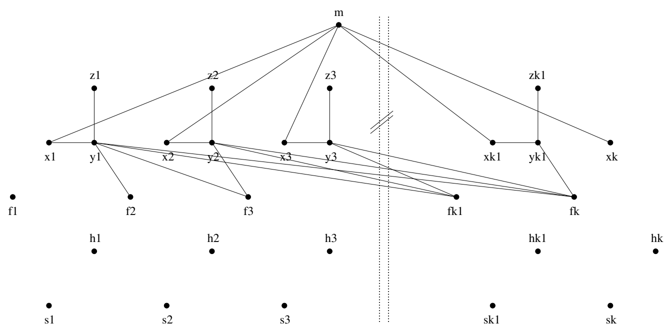

Fig. 11. The causally relevant subgraph of (303) showing the independence of X_k and $\mathbf{H}_k$ when conditioned on $\left(\mathbf{H}_1^{k-1}, \mathbf{S}_1^k, \mathbf{F}_1^k\right)$.

$$
\begin{aligned}
&= I\left(X_k, \mathbf{H}_1^{k-1}; \mathbf{Y}_k \middle| \mathbf{S}_1^k, \mathbf{F}_1^k\right) && (331)\\
&\leq I\left(X_k, \mathbf{H}_1^{k-1}; \mathbf{Y}_k, \mathbf{H}_k \middle| \mathbf{S}_1^k, \mathbf{F}_1^k\right) && (332)\\
&= I\left(\mathbf{H}_1^{k-1}; \mathbf{H}_k \middle| \mathbf{S}_1^k, \mathbf{F}_1^k\right) + I\left(X_k; \mathbf{H}_k \middle| \mathbf{H}_1^{k-1}, \mathbf{S}_1^k, \mathbf{F}_1^k\right)\\
&\quad + I\left(X_k; \mathbf{Y}_k \middle| \mathbf{H}_k, \mathbf{S}_1^k, \mathbf{F}_1^k\right)\\
&\quad + I\left(\mathbf{H}_1^{k-1}; \mathbf{Y}_k \middle| \mathbf{H}_k, X_k, \mathbf{S}_1^k, \mathbf{F}_1^k\right) && (333)\\
&= I\left(\mathbf{H}_1^{k-1}; \mathbf{H}_k \middle| \mathbf{S}_1^k, \mathbf{F}_1^k\right) + I\left(X_k; \mathbf{Y}_k \middle| \mathbf{H}_k, \mathbf{S}_1^k, \mathbf{F}_1^k\right) && (334)
\end{aligned}
$$

where in (324) equality holds because X_1^k is a function of the message M, the feedback $\left(\mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}\right)$ and the side-information $\mathbf{S}_k$; where in (327) we introduce the feedback $\mathbf{F}_1^k \triangleq \left(\mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}\right)$; and where (325), (331), and (334) are proven in Appendix A (see Figs. 9, 10, 11, and 12, respectively).

Combining (334), and (322), and (316) now yields

$$
\begin{aligned}
&I\left(M; \mathbf{Y}_k, \mathbf{S}_k \middle| \mathbf{Y}_1^{k-1}, \mathbf{S}_1^{k-1}\right)\\
&\quad \leq I\left(\mathbf{H}_1^{k-1}; \mathbf{S}_k \middle| \mathbf{S}_1^{k-1}\right) + I\left(\mathbf{H}_1^{k-1}; \mathbf{H}_k \middle| \mathbf{S}_1^k, \mathbf{F}_1^k\right)\\
&\quad\quad + I\left(X_k; \mathbf{Y}_k \middle| \mathbf{H}_k, \mathbf{S}_1^k, \mathbf{F}_1^k\right) && (335)
\end{aligned}
$$

where we continue to bound the first two terms as follows:

$$
\begin{aligned}
&I\left(\mathbf{H}_1^{k-1}; \mathbf{S}_k \middle| \mathbf{S}_1^{k-1}\right) + I\left(\mathbf{H}_1^{k-1}; \mathbf{H}_k \middle| \mathbf{S}_1^k, \mathbf{F}_1^k\right)\\
&\quad = I\left(\mathbf{H}_1^{k-1}; \mathbf{S}_k \middle| \mathbf{S}_1^{k-1}\right) + h\left(\mathbf{H}_k \middle| \mathbf{S}_1^k, \mathbf{F}_1^k\right)\\
&\quad\quad - h\left(\mathbf{H}_k \middle| \mathbf{H}_1^{k-1}, \mathbf{S}_1^k, \mathbf{F}_1^k\right) && (336)\\
&\quad \leq I\left(\mathbf{H}_1^{k-1}; \mathbf{S}_k \middle| \mathbf{S}_1^{k-1}\right) + h\left(\mathbf{H}_k \middle| \mathbf{S}_1^k\right)\\
&\quad\quad - h\left(\mathbf{H}_k \middle| \mathbf{H}_1^{k-1}, \mathbf{S}_1^k, \mathbf{F}_1^k\right) && (337)\\
&\quad = I\left(\mathbf{H}_1^{k-1}; \mathbf{S}_k \middle| \mathbf{S}_1^{k-1}\right) + h\left(\mathbf{H}_k \middle| \mathbf{S}_1^k\right)\\
&\quad\quad - h\left(\mathbf{H}_k \middle| \mathbf{H}_1^{k-1}, \mathbf{S}_1^k\right) && (338)
\end{aligned}
$$

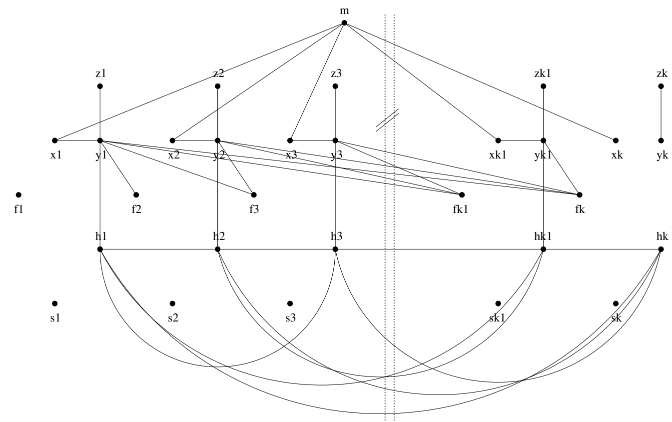

Fig. 12. The causally relevant subgraph of (303) showing the independence of $\mathbf{Y}_k$ and $\mathbf{H}_1^{k-1}$ when conditioned on $\left(X_k, \mathbf{H}_k, \mathbf{S}_1^k, \mathbf{F}_1^k\right)$.

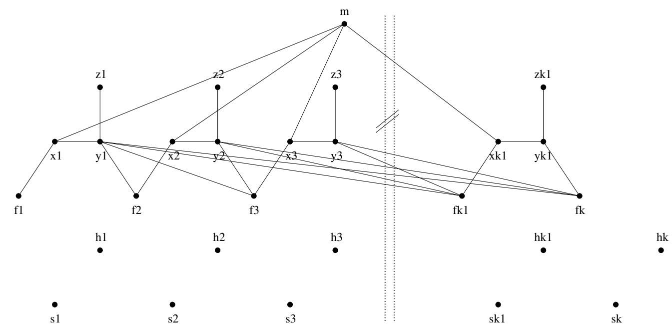

Fig. 13. The causally relevant subgraph of (303) showing the independence of $\mathbf{H}_k$ and $\mathbf{F}_1^k$ when conditioned on $\left(\mathbf{H}_1^{k-1}, \mathbf{S}_1^k\right)$.

$$
\begin{aligned}
&= I\left(\mathbf{H}_1^{k-1}; \mathbf{S}_k \middle| \mathbf{S}_1^{k-1}\right) + I\left(\mathbf{H}_1^{k-1}; \mathbf{H}_k \middle| \mathbf{S}_1^k\right) && (339)\\
&= I\left(\mathbf{H}_1^{k-1}; \mathbf{H}_k, \mathbf{S}_k \middle| \mathbf{S}_1^{k-1}\right) && (340)\\
&\leq I\left(\mathbf{H}_1^{k-1}, \mathbf{S}_1^{k-1}; \mathbf{H}_k, \mathbf{S}_k\right) && (341)
\end{aligned}
$$

where (338) is proven in Appendix A (see Fig. 13).

The third term in (335) is bounded as follows:

$$
\begin{aligned}
&I\left(X_k; \mathbf{Y}_k \middle| \mathbf{H}_k, \mathbf{S}_1^k, \mathbf{F}_1^k\right)\\
&\quad = h\left(\mathbf{Y}_k \middle| \mathbf{H}_k, \mathbf{S}_1^k, \mathbf{F}_1^k\right) - h\left(\mathbf{Y}_k \middle| X_k, \mathbf{H}_k, \mathbf{S}_1^k, \mathbf{F}_1^k\right) && (342)\\
&\quad = h\left(\mathbf{Y}_k \middle| \mathbf{H}_k, \mathbf{S}_1^k, \mathbf{F}_1^k\right) - h(\mathbf{Z}_k) && (343)\\
&\quad \leq \mathsf{E}\left[\sup_{X:\ \mathsf{E}[|X|^2] \leq \mathsf{E}_k} h(\mathbf{H}_k X + \mathbf{Z}_k | \mathbf{H}_k = \mathbf{h}_k)\right]\\
&\quad\quad - h(\mathbf{Z}_k) && (344)\\
&\quad = \mathsf{E}\left[\log\left((\pi e)^{n_\mathsf{R}} \det\left(\mathbf{H}_k \mathbf{H}_k^\dagger \mathsf{E}_k + \sigma^2 \mathsf{I}\right)\right)\right]\\
&\quad\quad - n_\mathsf{R} \log\left(\pi e \sigma^2\right) && (345)\\
&\quad \leq \mathsf{E}\left[\log \prod_{\ell=1}^{n_\mathsf{R}} \left(\left|H_k^{(\ell)}\right|^2 \mathsf{E}_k + \sigma^2\right)\right] - n_\mathsf{R} \log \sigma^2 && (346)\\
&\quad = \mathsf{E}\left[n_\mathsf{R} \frac{1}{n_\mathsf{R}} \sum_{\ell=1}^{n_\mathsf{R}} \log\left(1 + \left|H_k^{(\ell)}\right|^2 \frac{\mathsf{E}_k}{\sigma^2}\right)\right] && (347)\\
&\quad \leq \mathsf{E}\left[n_\mathsf{R} \log\left(1 + \frac{1}{n_\mathsf{R}} \|\mathbf{H}_k\|^2 \frac{\mathsf{E}_k}{\sigma^2}\right)\right] && (348)\\
&\quad \leq n_\mathsf{R} \log\left(1 + \frac{1}{n_\mathsf{R}} \mathsf{E}\left[\|\mathbf{H}_k\|^2\right] \frac{\mathsf{E}_k}{\sigma^2}\right). && (349)
\end{aligned}
$$

Here, in (344) we take the supremum over all input distributions with given second moment E_k, which is achieved by a Gaussian distribution; (346) follows from Hadamard's inequality [22]; and (348) and (349) follow from Jensen's inequality. Since $\mathsf{E}\left[\|\mathbf{H}_k\|^2\right] = \mathsf{E}\left[\|\mathbf{H}_0\|^2\right]$, this completes the proof.

APPENDIX D
DERIVATION OF $\delta_2(\xi_{\min},\kappa)$ AND $\delta_3(\xi_{\min},\kappa)$

We first show that

$$\beta_k\, I\big(\mathbf{Z}_{k-\kappa}^{k-1};\mathbf{Y}_k\big|\mathbf{Y}_{k-\kappa}^{k-1},\mathbf{S}_1^k,\mathcal{V}\big) \le \delta_2(\xi_{\min},\kappa) + \mathsf{H}_\mathsf{b}(\beta_k) \tag{350}$$

where

$$\lim_{\xi_{\min}\uparrow\infty} \delta_2(\xi_{\min},\kappa) = 0. \tag{351}$$

To that goal, we bound as follows:

$$\begin{aligned}
&\beta_k\, I\big(\mathbf{Z}_{k-\kappa}^{k-1};\mathbf{Y}_k\big|\mathbf{Y}_{k-\kappa}^{k-1},\mathbf{S}_1^k,\mathcal{V}\big)\\
&= \beta_k\, h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big|\mathbf{Y}_{k-\kappa}^{k-1},\mathbf{S}_1^k,\mathcal{V}\big) - \beta_k\, h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big|\mathbf{Y}_{k-\kappa}^{k},\mathbf{S}_1^k,\mathcal{V}\big) &(352)\\
&\le \beta_k\, h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big|\mathcal{V}\big)\\
&\quad - \beta_k\, h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big|\mathbf{Y}_{k-\kappa}^{k},X_{k-\kappa}^k,\mathbf{Z}_k,\mathbf{H}_1^{k-\kappa-1},\mathbf{S}_1^k,\mathcal{V}\big) &(353)\\
&= \beta_k\, h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big|\mathcal{V}\big) - \beta_k \mathsf{E}_{X_{k-\kappa}^k}\Big[h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big|\mathbf{Y}_{k-\kappa}^k,\\
&\qquad \{X_\ell = x_\ell\}_{\ell=k-\kappa}^k,\mathbf{Z}_k,\mathbf{H}_1^{k-\kappa-1},\mathbf{S}_1^k,\mathcal{V}\big)\Big] &(354)\\
&= \beta_k\, h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big|\mathcal{V}\big)\\
&\quad - \beta_k \mathsf{E}_{X_{k-\kappa}^k}\left[h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle|\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{X_\ell}\right\}_{\ell=k-\kappa}^{k-1},\mathbf{H}_k,\right.\right.\\
&\qquad\left.\left.\mathbf{H}_1^{k-\kappa-1},\mathbf{S}_1^k,\mathcal{V}\right)\right] &(355)\\
&\le \beta_k\, h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big|\mathcal{V}\big)\\
&\quad - \beta_k \inf_{\substack{x_\ell:\,|x_\ell|\ge\xi_{\min}\\ \ell=k-\kappa,\ldots,k}} h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle|\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{x_\ell}\right\}_{\ell=k-\kappa}^{k-1},\mathbf{H}_k,\right.\\
&\qquad\left.\mathbf{H}_1^{k-\kappa-1},\mathbf{S}_1^k,\mathcal{V}\right) &(356)\\
&= \beta_k\, h\big(\mathbf{Z}_{k-\kappa}^{k-1}\big|\mathcal{V}\big) - \beta_k\, h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle|\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=k-\kappa}^{k-1},\right.\\
&\qquad\left.\mathbf{H}_k,\mathbf{H}_1^{k-\kappa-1},\mathbf{S}_1^k,\mathcal{V}\right) &(357)\\
&= \beta_k\, I\left(\mathbf{Z}_{k-\kappa}^{k-1};\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=k-\kappa}^{k-1},\mathbf{H}_k,\mathbf{H}_1^{k-\kappa-1},\mathbf{S}_1^k\middle|\mathcal{V}\right) &(358)\\
&\le \beta_k\, I\left(\mathbf{Z}_{k-\kappa}^{k-1};\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=k-\kappa}^{k-1},\mathbf{H}_k,\mathbf{H}_1^{k-\kappa-1},\right.\\
&\qquad\left.\mathbf{S}_1^k\middle|A_k=0,B_k=1\right)\\
&\quad + (1-\beta_k)\, I\left(\mathbf{Z}_{k-\kappa}^{k-1};\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=k-\kappa}^{k-1},\mathbf{H}_k,\right.\\
&\qquad\left.\mathbf{H}_1^{k-\kappa-1},\mathbf{S}_1^k\middle|A_k=0,B_k=0\right) &(359)\\
&= I\left(\mathbf{Z}_{k-\kappa}^{k-1};\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=k-\kappa}^{k-1},\mathbf{H}_k,\right.\\
&\qquad\left.\mathbf{H}_1^{k-\kappa-1},\mathbf{S}_1^k\middle|A_k=0,B_k\right) &(360)\\
&\le I\left(\mathbf{Z}_{k-\kappa}^{k-1};\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=k-\kappa}^{k-1},\mathbf{H}_k,\right.\\
&\qquad\left.\mathbf{H}_1^{k-\kappa-1},\mathbf{S}_1^k,B_k\middle|A_k=0\right) &(361)\\
&= I\left(\mathbf{Z}_{k-\kappa}^{k-1};\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=k-\kappa}^{k-1},\mathbf{H}_k,\mathbf{H}_1^{k-\kappa-1},\right.\\
&\qquad\left.\mathbf{S}_1^k\middle|A_k=0\right)\\
&\quad + I\left(\mathbf{Z}_{k-\kappa}^{k-1};B_k\middle|\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=k-\kappa}^{k-1},\mathbf{H}_k,\right.\\
&\qquad\left.\mathbf{H}_1^{k-\kappa-1},\mathbf{S}_1^k,A_k=0\right) &(362)\\
&\le I\left(\mathbf{Z}_{k-\kappa}^{k-1};\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=k-\kappa}^{k-1},\mathbf{H}_k,\mathbf{H}_1^{k-\kappa-1},\right.\\
&\qquad\left.\mathbf{S}_1^k\middle|A_k=0\right) + \mathsf{H}(B_k|A_k=0) &(363)\\
&= I\left(\left\{\frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=-\kappa}^{-1};\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=-\kappa}^{-1},\mathbf{H}_0,\right.\\
&\qquad\left.\mathbf{H}_{-k+1}^{-\kappa-1},\mathbf{S}_{-k+1}^0\middle|A_0=0\right) + \mathsf{H}_\mathsf{b}(\beta_k) &(364)\\
&\le \lim_{j\uparrow\infty} I\left(\left\{\frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=-\kappa}^{-1};\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=-\kappa}^{-1},\mathbf{H}_0,\right.\\
&\qquad\left.\mathbf{H}_{-j}^{-\kappa-1},\mathbf{S}_{-j}^0\middle|A_0=0\right) + \mathsf{H}_\mathsf{b}(\beta_k) &(365)\\
&= \lim_{j\uparrow\infty} I\left(\left\{\frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=-\kappa}^{-1};\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=-\kappa}^{-1}\middle|\mathbf{H}_0,\right.\\
&\qquad\left.\mathbf{H}_{-j}^{-\kappa-1},\mathbf{S}_{-j}^0,A_0=0\right) + \mathsf{H}_\mathsf{b}(\beta_k) &(366)\\
&= \lim_{j\uparrow\infty} h\left(\left\{\mathbf{H}_\ell + \frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=-\kappa}^{-1}\middle|\mathbf{H}_0,\mathbf{H}_{-j}^{-\kappa-1},\mathbf{S}_{-j}^0,A_0=0\right)\\
&\quad - h\big(\mathbf{H}_{-\kappa}^{-1}\big|\mathbf{H}_0,\mathbf{H}_{-\infty}^{-\kappa-1},\mathbf{S}_{-\infty}^0,A_0=0\big) + \mathsf{H}_\mathsf{b}(\beta_k) &(367)\\
&\triangleq \delta_2(\xi_{\min},\kappa) + \mathsf{H}_\mathsf{b}(\beta_k). &(368)
\end{aligned}$$

Here, in (353) we add and remove conditioning random variables to differential entropy; (355) holds because conditional on $\big(Y_{k-\kappa}^k,\mathbf{H}_1^{k-\kappa-1}\big)$ the input is independent of $\mathbf{Z}_{k-\kappa}^{k-1}$;

(359) follows by adding a nonnegative term; (361) follows by moving B_k from the conditioning argument to a main argument of the mutual information; in (363) we drop any negative term and use conditioning that reduces entropy; in the subsequent equality (364) we use stationarity; then in (365) we add more arguments; and (366) follows from the independence of $\{\mathbf{H}_k, \mathbf{S}_k\}$ and $\{\mathbf{Z}_k\}$.

Note that (351) holds because

$$\begin{aligned}
&\lim_{\xi_{\min}\uparrow\infty}\lim_{j\uparrow\infty} h\left(\left\{\mathbf{H}_\ell+\frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=-\kappa}^{-1}\middle|\mathbf{H}_0,\mathbf{H}_{-j}^{-\kappa-1},\mathbf{S}_{-j}^0,A_0=0\right)\\
&= \lim_{j\uparrow\infty}\lim_{\xi_{\min}\uparrow\infty} h\left(\left\{\mathbf{H}_\ell+\frac{\mathbf{Z}_\ell}{\xi_{\min}}\right\}_{\ell=-\kappa}^{-1}\middle|\mathbf{H}_0,\mathbf{H}_{-j}^{-\kappa-1},\mathbf{S}_{-j}^0,A_0=0\right) && (369)\\
&= \lim_{j\uparrow\infty} h\left(\mathbf{H}_{-\kappa}^{-1}\middle|\mathbf{H}_0,\mathbf{H}_{-j}^{-\kappa-1},\mathbf{S}_{-j}^0,A_0=0\right) && (370)\\
&= h\left(\mathbf{H}_{-\kappa}^{-1}\middle|\mathbf{H}_0,\mathbf{H}_{-\infty}^{-\kappa-1},\mathbf{S}_{-\infty}^0,A_0=0\right). && (371)
\end{aligned}$$

Here, the exchange of the limits in (369) follows from [23, Theorem 7.11] (note that the differential entropy term is monotonically decreasing in j and bounded, i.e., it is uniformly converging), and (370) follows from a conditional version of [6, Lemma 6.11].

In a quite similar fashion, we next show that

$$\begin{aligned}
\beta_k\, I\left(\mathbf{Y}_{k-\kappa}^{k-1},\mathbf{Z}_{k-\kappa}^{k-1};\mathbf{Z}_k\middle|\mathbf{Y}_k,\mathbf{S}_1^k,\mathcal{V}\right)&\\
\le \delta_3(\xi_{\min},\kappa)+\mathsf{H}_\mathsf{b}(\beta_k)& && (372)
\end{aligned}$$

where

$$\lim_{\xi_{\min}\uparrow\infty}\delta_3(\xi_{\min},\kappa)=0. \qquad (373)$$

We have

$$\begin{aligned}
&\beta_k\, I\left(\mathbf{Y}_{k-\kappa}^{k-1},\mathbf{Z}_{k-\kappa}^{k-1};\mathbf{Z}_k\middle|\mathbf{Y}_k,\mathbf{S}_1^k,\mathcal{V}\right)\\
&= \beta_k\, h\left(\mathbf{Z}_k\middle|\mathbf{Y}_k,\mathbf{S}_1^k,\mathcal{V}\right)-\beta_k\, h\left(\mathbf{Z}_k\middle|\mathbf{Y}_{k-\kappa}^k,\mathbf{Z}_{k-\kappa}^{k-1},\mathbf{S}_1^k,\mathcal{V}\right) && (374)\\
&\le \beta_k\, h(\mathbf{Z}_k|\mathcal{V}) - \beta_k\, h\left(\mathbf{Z}_k\middle|\mathbf{Y}_{k-\kappa}^k,X_{k-\kappa}^k,\mathbf{Z}_{k-\kappa}^{k-1},\mathbf{H}_1^{k-\kappa-1},\mathbf{S}_1^k,\mathcal{V}\right) && (375)\\
&= \beta_k\, h(\mathbf{Z}_k|\mathcal{V}) - \beta_k \mathsf{E}_{X_{k-\kappa}^k}\left[h\left(\mathbf{Z}_k\middle|\mathbf{Y}_{k-\kappa}^k,\{X_\ell=x_\ell\}_{\ell=k-\kappa}^k,\mathbf{Z}_{k-\kappa}^{k-1},\mathbf{H}_1^{k-\kappa-1},\mathbf{S}_1^k,\mathcal{V}\right)\right] && (376)\\
&= \beta_k\, h(\mathbf{Z}_k|\mathcal{V}) - \beta_k\, \mathsf{E}_{X_{k-\kappa}^k}\left[h\left(\mathbf{Z}_k\middle|\mathbf{H}_{k-\kappa}^{k-1},\mathbf{H}_k+\frac{\mathbf{Z}_k}{X_k},\mathbf{H}_1^{k-\kappa-1},\mathbf{S}_1^k,\mathcal{V}\right)\right] && (377)\\
&\le \beta_k\, h(\mathbf{Z}_k|\mathcal{V}) - \beta_k \inf_{\substack{x_\ell:\,|x_\ell|\ge\xi_{\min}\\ \ell=k-\kappa,\ldots,k}} h\left(\mathbf{Z}_k\middle|\mathbf{H}_k+\frac{\mathbf{Z}_k}{x_k},\mathbf{H}_1^{k-1},\mathbf{S}_1^k,\mathcal{V}\right) && (378)\\
&= \beta_k\, h(\mathbf{Z}_k|\mathcal{V}) - \beta_k\, h\left(\mathbf{Z}_k\middle|\mathbf{H}_k+\frac{\mathbf{Z}_k}{\xi_{\min}},\mathbf{H}_1^{k-1},\mathbf{S}_1^k,\mathcal{V}\right) && (379)\\
&= \beta_k\, I\left(\mathbf{Z}_k;\mathbf{H}_k+\frac{\mathbf{Z}_k}{\xi_{\min}},\mathbf{H}_1^{k-1},\mathbf{S}_1^k\middle|\mathcal{V}\right) && (380)\\
&\le \beta_k\, I\left(\mathbf{Z}_k;\mathbf{H}_k+\frac{\mathbf{Z}_k}{\xi_{\min}},\mathbf{H}_1^{k-1},\mathbf{S}_1^k\middle|A_k=0,B_k=1\right)\\
&\quad + (1-\beta_k)\, I\left(\mathbf{Z}_k;\mathbf{H}_k+\frac{\mathbf{Z}_k}{\xi_{\min}},\mathbf{H}_1^{k-1},\mathbf{S}_1^k\middle|A_k=0,B_k=0\right) && (381)\\
&= I\left(\mathbf{Z}_k;\mathbf{H}_k+\frac{\mathbf{Z}_k}{\xi_{\min}},\mathbf{H}_1^{k-1},\mathbf{S}_1^k\middle|A_k=0,B_k\right) && (382)\\
&\le I\left(\mathbf{Z}_k;\mathbf{H}_k+\frac{\mathbf{Z}_k}{\xi_{\min}},\mathbf{H}_1^{k-1},\mathbf{S}_1^k,B_k\middle|A_k=0\right) && (383)\\
&\le I\left(\mathbf{Z}_k;\mathbf{H}_k+\frac{\mathbf{Z}_k}{\xi_{\min}},\mathbf{H}_1^{k-1},\mathbf{S}_1^k\middle|A_k=0\right)+\mathsf{H}_\mathsf{b}(\beta_k) && (384)\\
&= I\left(\frac{\mathbf{Z}_0}{\xi_{\min}};\mathbf{H}_0+\frac{\mathbf{Z}_0}{\xi_{\min}},\mathbf{H}_{-k+1}^{-1},\mathbf{S}_{-k+1}^0\middle|A_0=0\right)+\mathsf{H}_\mathsf{b}(\beta_k) && (385)\\
&\le \lim_{j\uparrow\infty} I\left(\frac{\mathbf{Z}_0}{\xi_{\min}};\mathbf{H}_0+\frac{\mathbf{Z}_0}{\xi_{\min}},\mathbf{H}_{-j}^{-1},\mathbf{S}_{-j}^0\middle|A_0=0\right)+\mathsf{H}_\mathsf{b}(\beta_k) && (386)\\
&= \lim_{j\uparrow\infty} I\left(\frac{\mathbf{Z}_0}{\xi_{\min}};\mathbf{H}_0+\frac{\mathbf{Z}_0}{\xi_{\min}}\middle|\mathbf{H}_{-j}^{-1},\mathbf{S}_{-j}^0,A_0=0\right)+\mathsf{H}_\mathsf{b}(\beta_k) && (387)\\
&= \lim_{j\uparrow\infty} h\left(\mathbf{H}_0+\frac{\mathbf{Z}_0}{\xi_{\min}}\middle|\mathbf{H}_{-j}^{-1},\mathbf{S}_{-j}^0,A_0=0\right) - h\left(\mathbf{H}_0\middle|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0,A_0=0\right)+\mathsf{H}_\mathsf{b}(\beta_k) && (388)\\
&\triangleq \delta_3(\xi_{\min},\kappa)+\mathsf{H}_\mathsf{b}(\beta_k). && (389)
\end{aligned}$$

Analogously to (369)–(371), one argues that

$$\begin{aligned}
&\lim_{\xi_{\min}\uparrow\infty}\lim_{j\uparrow\infty} h\left(\mathbf{H}_0+\frac{\mathbf{Z}_0}{\xi_{\min}}\middle|\mathbf{H}_{-j}^{-1},\mathbf{S}_{-j}^0,A_0=0\right)\\
&= h\left(\mathbf{H}_0\middle|\mathbf{H}_{-\infty}^{-1},\mathbf{S}_{-\infty}^0,A_0=0\right). && (390)
\end{aligned}$$

Appendix E
Derivation of Bound (213)

Similarly to (331)–(334), we bound as follows:

$$\begin{aligned}
&I\left(X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\middle|\mathbf{S}_1^k,A_k=0,B_k=0\right)\\
&\le I\left(X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k,\mathbf{H}_k\middle|\mathbf{S}_1^k,A_k=0,B_k=0\right) && (391)\\
&= I\left(\mathbf{H}_1^{k-1};\mathbf{H}_k\middle|\mathbf{S}_1^k,A_k=0,B_k=0\right)\\
&\quad + I\left(X_k;\mathbf{H}_k\middle|\mathbf{H}_1^{k-1},\mathbf{S}_1^k,A_k=0,B_k=0\right)\\
&\quad + I\left(X_k;\mathbf{Y}_k\middle|\mathbf{H}_k,\mathbf{S}_1^k,A_k=0,B_k=0\right)\\
&\quad + I\left(\mathbf{H}_1^{k-1};\mathbf{Y}_k\middle|\mathbf{H}_k,X_k,\mathbf{S}_1^k,A_k=0,B_k=0\right) && (392)\\
&= I\left(\mathbf{H}_1^{k-1};\mathbf{H}_k\middle|\mathbf{S}_1^k,A_k=0,B_k=0\right)\\
&\quad + I\left(X_k;\mathbf{Y}_k\middle|\mathbf{H}_k,\mathbf{S}_1^k,A_k=0,B_k=0\right) && (393)
\end{aligned}$$

where the last equality can be seen from Appendix A (see Figs. 5 and 6). Moreover, similarly to (342)–(349) we bound the second term in (393) as follows:

$$\begin{aligned}
&I\left(X_k;\mathbf{Y}_k\middle|\mathbf{H}_k,\mathbf{S}_1^k,A_k=0,B_k=0\right)\\
&= h\left(\mathbf{Y}_k\middle|\mathbf{H}_k,\mathbf{S}_1^k,A_k=0,B_k=0\right) - h\left(\mathbf{Y}_k\middle|X_k,\mathbf{H}_k,\mathbf{S}_1^k,A_k=0,B_k=0\right) && (394)\\
&= h\left(\mathbf{Y}_k\middle|\mathbf{H}_k,\mathbf{S}_1^k,A_k=0,B_k=0\right)-h(\mathbf{Z}_k) && (395)
\end{aligned}$$

$$\leq \mathsf{E}\bigg[\sup_{X:\ \mathsf{E}[|X|^2]\leq\xi^2_{\min}} h(\mathbf{H}_k X + \mathbf{Z}_k | \mathbf{H}_k = \mathbf{h}_k, A_k = 0, B_k = 0)\bigg| A_k = 0, B_k = 0\bigg] - h(\mathbf{Z}_k) \tag{396}$$

$$= \mathsf{E}\Big[\log\Big((\pi e)^{n_{\mathsf{R}}} \det\Big(\mathbf{H}_k\mathbf{H}_k^\dagger \xi^2_{\min} + \sigma^2 \mathsf{I}\Big)\Big)\Big| A_k = 0, B_k = 0\Big] - n_{\mathsf{R}}\log\big(\pi e\sigma^2\big) \tag{397}$$

$$\leq \mathsf{E}\left[\log\prod_{\ell=1}^{n_{\mathsf{R}}}\Big(\big|H_k^{(\ell)}\big|^2\xi^2_{\min} + \sigma^2\Big)\middle| A_k = 0, B_k = 0\right] - n_{\mathsf{R}}\log\sigma^2 \tag{398}$$

$$= \mathsf{E}\left[n_{\mathsf{R}}\,\frac{1}{n_{\mathsf{R}}}\sum_{\ell=1}^{n_{\mathsf{R}}}\log\left(1 + \big|H_k^{(\ell)}\big|^2\,\frac{\xi^2_{\min}}{\sigma^2}\right)\middle| A_k = 0, B_k = 0\right] \tag{399}$$

$$\leq \mathsf{E}\left[n_{\mathsf{R}}\log\left(1 + \frac{1}{n_{\mathsf{R}}}\,\|\mathbf{H}_k\|^2\,\frac{\xi^2_{\min}}{\sigma^2}\right)\middle| A_k = 0, B_k = 0\right] \tag{400}$$

$$\leq n_{\mathsf{R}}\log\left(1 + \frac{1}{n_{\mathsf{R}}}\,\mathsf{E}\big[\|\mathbf{H}_k\|^2\big| A_k = 0, B_k = 0\big]\,\frac{\xi^2_{\min}}{\sigma^2}\right) \tag{401}$$

$$\leq n_{\mathsf{R}}\log\left(1 + \frac{2\xi^2_{\min}}{n_{\mathsf{R}}\sigma^2}\,\frac{\mathsf{E}\big[\|\mathbf{H}_k\|^2\big]}{1-\beta_k}\right) \tag{402}$$

where the last inequality follows from Lemma 9 and (206). Thus,

$$\begin{aligned} &I\big(X_k, \mathbf{H}_1^{k-1}; \mathbf{Y}_k \big| \mathbf{S}_1^k, A_k = 0, B_k = 0\big) \\ &\quad\leq h\big(\mathbf{H}_k \big| \mathbf{S}_1^k, A_k = 0, B_k = 0\big) \\ &\qquad - h\big(\mathbf{H}_k \big| \mathbf{H}_1^{k-1}, \mathbf{S}_1^k, A_k = 0, B_k = 0\big) \\ &\qquad + n_{\mathsf{R}}\log\left(1 + \frac{2\xi^2_{\min}}{n_{\mathsf{R}}\sigma^2}\,\frac{\mathsf{E}\big[\|\mathbf{H}_k\|^2\big]}{1-\beta_k}\right). \end{aligned} \tag{403}$$

Next,

$$-h_\lambda\Big(\hat{\mathbf{H}}_k\,e^{\mathrm{i}\Theta_k}\Big|\big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}, \mathbf{S}_1^k, A_k = 0, B_k = 0\Big)$$

$$\leq -h_\lambda\Big(\hat{\mathbf{H}}_k\,e^{\mathrm{i}\Theta_k}\Big|\mathbf{H}_1^{k-1}, \big\{\hat{\mathbf{H}}_\ell\,e^{\mathrm{i}\Theta_\ell}\big\}_{\ell=k-\kappa}^{k-1}, \mathbf{S}_1^k, A_k = 0, B_k = 0\Big) \tag{404}$$

$$= -h_\lambda\Big(\hat{\mathbf{H}}_k\,e^{\mathrm{i}\Theta_k}\Big|\mathbf{H}_1^{k-1}, \mathbf{S}_1^k, A_k = 0\Big) \tag{405}$$

where we use that conditional on $\big(\mathbf{H}_1^{k-1}, \mathbf{S}_1^k\big)$, $\mathbf{H}_k$ is independent of the input (use Fig. 6 in Appendix A to see that X_1^k is independent of $\mathbf{H}_k$, when conditioned on $\big(\mathbf{H}_1^{k-1}, \mathbf{S}_1^k\big)$).

The bounding of the last term on the right-hand side of (213) is more elaborate. Note that the following derivation can be extended to a more general setup (compare with [6, Section 6]). We use the shorthand

$$\bar{\mathcal{V}} \triangleq \{A_k = 0, B_k = 0\} \tag{406}$$

and compute (using again $\mathscr{I}\{\cdot\}$ as the indicator function)

$$\begin{aligned} &-n_{\mathsf{R}}\,\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|\bar{\mathcal{V}}\big] \\ &\quad= -2n_{\mathsf{R}}\mathsf{E}\Big[\log\|\mathbf{H}_k\|\cdot\mathscr{I}\{\|\mathbf{H}_k\|\leq 1\} + \underbrace{\log\|\mathbf{H}_k\|\cdot\mathscr{I}\{\|\mathbf{H}_k\| > 1\}}_{\geq 0}\Big|\bar{\mathcal{V}}\Big] \end{aligned} \tag{407}$$

$$\leq -2n_{\mathsf{R}}\,\mathsf{E}\Big[\log\|\mathbf{H}_k\|\cdot\mathscr{I}\{\|\mathbf{H}_k\|\leq 1\}\Big|\bar{\mathcal{V}}\Big] \tag{408}$$

$$= 2n_{\mathsf{R}}\,\mathsf{E}\Big[\log\big(\|\mathbf{H}_k\|^{-1}\big)\cdot\mathscr{I}\{\|\mathbf{H}_k\|\leq 1\}\Big|\bar{\mathcal{V}}\Big] \tag{409}$$

$$= 2n_{\mathsf{R}}\int_{\|\mathbf{h}\|\leq 1} f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})\log\big(\|\mathbf{h}\|^{-1}\big)\,\mathrm{d}\mathbf{h} \tag{410}$$

$$= \underbrace{2n_{\mathsf{R}}\int_{\substack{\|\mathbf{h}\|\leq 1\\ f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})\leq\|\mathbf{h}\|^{-1/2}}} f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})\log\big(\|\mathbf{h}\|^{-1}\big)\,\mathrm{d}\mathbf{h}}_{\triangleq\,\mathsf{I}_1} + \underbrace{2n_{\mathsf{R}}\int_{\substack{\|\mathbf{h}\|\leq 1\\ f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})>\|\mathbf{h}\|^{-1/2}}} f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})\log\big(\|\mathbf{h}\|^{-1}\big)\,\mathrm{d}\mathbf{h}}_{\triangleq\,\mathsf{I}_2}. \tag{411}$$

The two integrals are now bounded separately. Using the surface area of the n_{R}-dimensional complex unit-sphere $\frac{2\pi^{n_{\mathsf{R}}}}{\Gamma(n_{\mathsf{R}})}$ and the standard relation between the PDF of $\mathbf{H}_k$ and the PDF of $\|\mathbf{H}_k\|$, we obtain

$$\mathsf{I}_1 = \int_{\substack{\|\mathbf{h}\|\leq 1\\ f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})\leq\|\mathbf{h}\|^{-1/2}}} f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})\underbrace{\log\big(\|\mathbf{h}\|^{-1}\big)}_{\geq 0}\,\mathrm{d}\mathbf{h} \tag{412}$$

$$\leq \int_{\|\mathbf{h}\|\leq 1}\|\mathbf{h}\|^{-\frac{1}{2}}\log\big(\|\mathbf{h}\|^{-1}\big)\,\mathrm{d}\mathbf{h} \tag{413}$$

$$= \frac{2\pi^{n_{\mathsf{R}}}}{\Gamma(n_{\mathsf{R}})}\int_{r\leq 1} r^{2n_{\mathsf{R}}-\frac{3}{2}}\log(r^{-1})\,\mathrm{d}r \tag{414}$$

$$= \frac{8\pi^{n_{\mathsf{R}}}}{\Gamma(n_{\mathsf{R}})(4n_{\mathsf{R}}-1)^2}. \tag{415}$$

The second integral can be bounded as follows:

$$\mathsf{I}_2 = 2\int_{\substack{\|\mathbf{h}\|^{-1/2}\geq 1\\ f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})>\|\mathbf{h}\|^{-1/2}}} f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})\log\big(\|\mathbf{h}\|^{-\frac{1}{2}}\big)\,\mathrm{d}\mathbf{h} \tag{416}$$

$$\leq 2\int_{\substack{\|\mathbf{h}\|^{-1/2}\geq 1\\ f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})>\|\mathbf{h}\|^{-1/2}}} f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})\log f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})\,\mathrm{d}\mathbf{h} \tag{417}$$

$$\leq 2\int_{f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})>1} f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})\log f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})\,\mathrm{d}\mathbf{h} \tag{418}$$

$$\triangleq 2h^-(\mathbf{H}_k|\bar{\mathcal{V}}). \tag{419}$$

Here the last step should be read as definition of $h^-(\mathbf{H}_k|\bar{\mathcal{V}})$. We have

$$h^-(\mathbf{H}_k|\bar{\mathcal{V}}) = h^+(\mathbf{H}_k|\bar{\mathcal{V}}) - h(\mathbf{H}_k|\bar{\mathcal{V}}) \tag{420}$$

where

$$h^+(\mathbf{H}_k|\bar{\mathcal{V}}) \triangleq \int_{0<f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})<1} f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})\log\frac{1}{f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})}\,\mathrm{d}\mathbf{h} \tag{421}$$

$$= p^+\int_{0<f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})<1}\frac{f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})}{p^+}\log\frac{1}{f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})/p^+}\,\mathrm{d}\mathbf{h} - p^+\log p^+ \tag{422}$$

with

$$p^+ \triangleq \Pr\big[0 < f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{H}_k) < 1\big]. \tag{423}$$

Note that the first term in (422) is a differential entropy of a PDF on the set $\{\mathbf{h}\colon 0 < f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h}) < 1\}$. We now bound this differential entropy by a standard bound on differential entropy for a given second moment:

$$h(\mathbf{U}) \leq n_{\mathrm{R}} \log\left(\frac{\pi e}{n_{\mathrm{R}}} \mathsf{E}\left[\|\mathbf{U}\|^2\right]\right) \tag{424}$$

(compare (343)–(349)). With

$$\begin{aligned} &\int_{0<f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})<1} \|\mathbf{h}\|^2 \, \frac{f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h})}{p^+} \, \mathrm{d}\mathbf{h} \\ &\qquad \leq \frac{1}{p^+} \int_{\mathbb{C}^{n_{\mathrm{R}}}} \|\mathbf{h}\|^2 \, f_{\mathbf{H}_k|\bar{\mathcal{V}}}(\mathbf{h}) \, \mathrm{d}\mathbf{h} &(425)\\ &\qquad = \frac{1}{p^+} \mathsf{E}\left[\|\mathbf{H}_k\|^2 \middle| \bar{\mathcal{V}}\right] &(426) \end{aligned}$$

we hence obtain

$$\begin{aligned} &h^+(\mathbf{H}_k|\bar{\mathcal{V}}) \\ &\leq p^+ n_{\mathrm{R}} \log\left(\frac{\pi e}{n_{\mathrm{R}}\, p^+} \mathsf{E}\left[\|\mathbf{H}_k\|^2 \middle| \bar{\mathcal{V}}\right]\right) - p^+ \log p^+ &(427)\\ &= p^+ n_{\mathrm{R}} \log\left(\frac{\pi e}{n_{\mathrm{R}}} \mathsf{E}\left[\|\mathbf{H}_k\|^2 \middle| \bar{\mathcal{V}}\right]\right) - (n_{\mathrm{R}}+1) p^+ \log p^+ &(428)\\ &\leq p^+ n_{\mathrm{R}} \log\left(1 + \frac{\pi e}{n_{\mathrm{R}}} \mathsf{E}\left[\|\mathbf{H}_k\|^2 \middle| \bar{\mathcal{V}}\right]\right) + (n_{\mathrm{R}}+1)\frac{1}{e} &(429)\\ &\leq n_{\mathrm{R}} \log\left(1 + \frac{\pi e}{n_{\mathrm{R}}} \mathsf{E}\left[\|\mathbf{H}_k\|^2 \middle| \bar{\mathcal{V}}\right]\right) + \frac{n_{\mathrm{R}}+1}{e}. &(430) \end{aligned}$$

Combining (411), (415), (419), (420), and (430) now yields

$$\begin{aligned} &-n_{\mathrm{R}} \mathsf{E}\left[\log \|\mathbf{H}_k\|^2 \middle| \bar{\mathcal{V}}\right] \\ &\quad \leq \frac{16 n_{\mathrm{R}} \pi^{n_{\mathrm{R}}}}{\Gamma(n_{\mathrm{R}})(4n_{\mathrm{R}}-1)^2} + 4n_{\mathrm{R}}^2 \log\left(1 + \frac{\pi e}{n_{\mathrm{R}}} \mathsf{E}\left[\|\mathbf{H}_k\|^2 \middle| \bar{\mathcal{V}}\right]\right) \\ &\qquad + \frac{4n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e} - 4n_{\mathrm{R}} h(\mathbf{H}_k|\bar{\mathcal{V}}) &(431)\\ &\quad \leq \frac{16 n_{\mathrm{R}} \pi^{n_{\mathrm{R}}}}{\Gamma(n_{\mathrm{R}})(4n_{\mathrm{R}}-1)^2} + 4n_{\mathrm{R}}^2 \log\left(1 + \frac{2\pi e}{n_{\mathrm{R}}} \frac{\mathsf{E}\left[\|\mathbf{H}_k\|^2\right]}{1-\beta_k}\right) \\ &\qquad + \frac{4n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e} - 4n_{\mathrm{R}} h\left(\mathbf{H}_k \middle| \mathbf{S}_1^k, \bar{\mathcal{V}}\right) &(432) \end{aligned}$$

where in the last step we have again made use of Lemma 9 and of (206), and we conditioned the last term on $\mathbf{S}_1^k$.

From (403), (405), and (432) we now obtain

$$\begin{aligned} &I\left(X_k, \mathbf{H}_1^{k-1}; \mathbf{Y}_k \middle| \mathbf{S}_1^k, A_k=0, B_k=0\right) \\ &- h_\lambda\left(\hat{\mathbf{H}}_k \, e^{\mathrm{i}\Theta_k} \middle| \left\{\hat{\mathbf{H}}_\ell \, e^{\mathrm{i}\Theta_\ell}\right\}_{\ell=k-\kappa}^{k-1}, \mathbf{S}_1^k, A_k=0, B_k=0\right) \\ &- n_{\mathrm{R}} \mathsf{E}\left[\log \|\mathbf{H}_k\|^2 \middle| A_k=0, B_k=0\right] \\ &\quad \leq -(4n_{\mathrm{R}}-1)\, h\left(\mathbf{H}_k \middle| \mathbf{S}_1^k, A_k=0, B_k=0\right) \\ &\qquad - h\left(\mathbf{H}_k \middle| \mathbf{H}_1^{k-1}, \mathbf{S}_1^k, A_k=0, B_k=0\right) \\ &\qquad + n_{\mathrm{R}} \log\left(1 + \frac{2\xi_{\min}^2}{n_{\mathrm{R}}\sigma^2} \frac{\mathsf{E}\left[\|\mathbf{H}_k\|^2\right]}{1-\beta_k}\right) \\ &\qquad - h_\lambda\left(\hat{\mathbf{H}}_k \, e^{\mathrm{i}\Theta_k} \middle| \mathbf{H}_1^{k-1}, \mathbf{S}_1^k, A_k=0\right) + \frac{16 n_{\mathrm{R}} \pi^{n_{\mathrm{R}}}}{\Gamma(n_{\mathrm{R}})(4n_{\mathrm{R}}-1)^2} \\ &\qquad + 4n_{\mathrm{R}}^2 \log\left(1 + \frac{2\pi e}{n_{\mathrm{R}}} \frac{\mathsf{E}\left[\|\mathbf{H}_k\|^2\right]}{1-\beta_k}\right) + \frac{4n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e} &(433)\\ &\quad \leq -4n_{\mathrm{R}}\, h\left(\mathbf{H}_k \middle| \mathbf{H}_1^{k-1}, \mathbf{S}_1^k, A_k=0, B_k=0\right) \\ &\qquad + \frac{4n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e} - h_\lambda\left(\hat{\mathbf{H}}_k \, e^{\mathrm{i}\Theta_k} \middle| \mathbf{H}_1^{k-1}, \mathbf{S}_1^k, A_k=0\right) \\ &\qquad + \frac{16 n_{\mathrm{R}} \pi^{n_{\mathrm{R}}}}{\Gamma(n_{\mathrm{R}})(4n_{\mathrm{R}}-1)^2} + n_{\mathrm{R}} \log\left(1 + \frac{2\xi_{\min}^2}{n_{\mathrm{R}}\sigma^2} \frac{\mathsf{E}\left[\|\mathbf{H}_k\|^2\right]}{1-\beta_k}\right) \\ &\qquad + 4n_{\mathrm{R}}^2 \log\left(1 + \frac{2\pi e}{n_{\mathrm{R}}} \frac{\mathsf{E}\left[\|\mathbf{H}_k\|^2\right]}{1-\beta_k}\right) &(434)\\ &\quad \leq -4n_{\mathrm{R}}\, h\left(\mathbf{H}_0 \middle| \mathbf{H}_{-\infty}^{-1}, \mathbf{S}_{-\infty}^0, A_0=0\right) + \frac{4n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e} \\ &\qquad - h_\lambda\left(\hat{\mathbf{H}}_0 \, e^{\mathrm{i}\Theta_0} \middle| \mathbf{H}_{-\infty}^{-1}, \mathbf{S}_{-\infty}^0, A_0=0\right) \\ &\qquad + \frac{16 n_{\mathrm{R}} \pi^{n_{\mathrm{R}}}}{\Gamma(n_{\mathrm{R}})(4n_{\mathrm{R}}-1)^2} + n_{\mathrm{R}} \log\left(1 + \frac{2\xi_{\min}^2}{n_{\mathrm{R}}\sigma^2} \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{1-\beta_k}\right) \\ &\qquad + 4n_{\mathrm{R}}^2 \log\left(1 + \frac{2\pi e}{n_{\mathrm{R}}} \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{1-\beta_k}\right). &(435) \end{aligned}$$

Here, in (434), we add more conditioning; and in the last inequality (435) we use the fact that conditional on $\left(\mathbf{H}_1^{k-1}, \mathbf{S}_1^k\right)$, $\mathbf{H}_k$ is independent of the input (use Fig. 6 in Appendix A to see that X_1^k is independent of $\mathbf{H}_k$ when conditioned on $\left(\mathbf{H}_1^{k-1}, \mathbf{S}_1^k\right)$), and then rely on stationarity and add some more additional conditioning.

Acknowledgments

This paper summarizes research performed over many years, and correspondingly, I am indebted to many people who contributed in some way. First of all, I would like to acknowledge Amos Lapidoth, under whose guidance the work for this project began and with whom some of the material was published quite some years back [17], [18]. In that context I also would like to thank Tobias Koch for many fruitful discussions.

As discussed in Section IV-E, our earlier versions of the proofs ([15], [17], [18]) were flawed. This was pointed out to us by Ashish Khisti, whom we therefore owe a very big thank you!

The generalization to SIMO was done jointly with my students Yuan-Zhu Guo [21] and Hsuan-Yin Lin. In that context, I also wish to thank Po-Ning Chen for his valuable input.

Finally, a very big thank you belongs to the associate editor Angel Lozano and to the two anonymous reviewers who provided very valuable input in extremely short time.

References

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

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

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

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

[5] G. Durisi, U. G. Schuster, H. Bölcskei, and S. Shamai (Shitz), "Noncoherent capacity of underspread fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 367–395, Jan. 2010.

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

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

[8] V. Sethuraman, L. Wang, B. Hajek, and A. Lapidoth, "Low-SNR capacity of noncoherent fading channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1555–1574, Apr. 2009.

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

[10] S. M. Moser, "The fading number of multiple-input multiple-output fading channels with memory," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2716–2755, Jun. 2009.

[11] J. L. Massey, "Causal interpretations of random variables," *Problemy Peredachi Informatsii (Problems of Inf. Transm.)*, vol. 32, no. 1, pp. 131–136, Jan.–Mar. 1996.

[12] J. L. Massey, "Correction to 'Causal interpretations of random variables'," *Problemy Peredachi Informatsii (Problems of Inf. Transm.)*, May 16, 1997.

[13] F. Topsøe, "An information theoretical identity and a problem involving capacity," *Studia Sci. Math. Hungar.*, vol. 2, pp. 291–292, 1967.

[14] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Budapest, Hungary: Academic Press, 1981.

[15] S. M. Moser, *Duality-Based Bounds on Channel Capacity*, ser. ETH Series in Information Theory and its Applications. Konstanz, Germany: Hartung-Gorre Verlag, Jan. 2005, vol. 1, edited by Amos Lapidoth. ISBN 3–89649–956–4. [Online]. Available: http://moser-isi.ethz.ch/publications.html

[16] G.-R. Lin and S. M. Moser, "The fading number of a multiple-access Rician fading channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4983–4991, Aug. 2011.

[17] A. Lapidoth and S. M. Moser, "Feedback increases neither the fading number nor the pre-log," in *Proc. 23rd IEEE Conv. El. & Electr. Eng. Israel*, Herzlia, Israel, Sept. 6–7, 2004, pp. 213–215.

[18] A. Lapidoth and S. M. Moser, "On non-coherent fading channels with feedback," in *Proc. Winter School Cod. and Inf. Theory*, Bratislava, Slovakia, Feb. 20–25, 2005, pp. 113–118.

[19] S. M. Moser, *Information Theory (Lecture Notes),* version 3, Signal and Inf. Proc. Lab., ETH Zürich, Switzerland, and Dept. El. & Comp. Eng., Nat. Chiao Tung Univ. (NCTU), Hsinchu, Taiwan, 2013. [Online]. Available: http://moser-isi.ethz.ch/scripts.html

[20] P. Billingsley, *Probability and Measure*, 3rd ed. Wiley, 1995.

[21] Y.-Z. Guo, "The asymptotic capacity of noncoherent single-input multiple-output fading channels with memory and feedback," Master's thesis, Inf. Th. Lab, Dept. El. & Comp. Eng., Nat. Chiao Tung Univ. (NCTU), Hsinchu, Taiwan, May 2013, supervised by Prof. Dr. Stefan M. Moser. [Online]. Available: http://moser-isi.ethz.ch/publications.html

[22] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985.

[23] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, 1976.

Stefan M. Moser (S'01–M'05–SM'10) received the diploma (M.Sc.) in electrical engineering in 1999, the M.Sc. degree in industrial management (M.B.A.) in 2003, and the Ph.D. degree (Dr. sc. techn.) in the field of information theory in 2004, all from ETH Zürich, Switzerland.

From 1999 to 2003, he was a Research and Teaching Assistant, and from 2004 to 2005, he was a Senior Research Assistant with the Signal and Information Processing Laboratory at ETH Zürich. From 2005 to 2013, he has been a Professor with the Department of Electrical and Computer Engineering at National Chiao Tung University (NCTU), Hsinchu, Taiwan. Currently he is a Senior Researcher with the Signal and Information Processing Laboratory at ETH Zürich and an Adjunct Professor with the Department of Electrical and Computer Engineering at National Chiao Tung University. His research interests are in information theory and digital communications.

Dr. Moser is recipient of the Wu Ta-You Memorial Award by the National Science Council of Taiwan in 2012, and the Best Paper Award for Young Scholars by the IEEE Communications Society Taipei and Tainan Chapters and the IEEE Information Theory Society Taipei Chapter in 2009. Further he received various awards from the National Chiao Tung University, so two awards for outstanding teaching in 2007 and 2012, and he was presented with the Willi Studer Award of ETH and the ETH Medal both in 1999, and with the Sandoz (Novartis) Basler Maturandenpreis in 1993.