# 國立交通大學

# 電信工程研究所

# 碩士論文

## 非同調單輸入多輸出
## 回授記憶性衰減通道之漸近通道容量

## The Asymptotic Capacity of Noncoherent Single-Input Multiple-Output Fading Channels with Memory and Feedback

研究生：郭沅竹

指導教授：Prof. Stefan M. Moser

陳柏寧 教授

中華民國　一百零三　年　七　月

# 非同調單輸入多輸出
# 回授記憶性衰減通道之漸近通道容量
# The Asymptotic Capacity of Noncoherent Single-Input Multiple-Output Fading Channels with Memory and Feedback

研究生：郭沅竹
指導教授：莫詩台方 教授
陳柏寧 教授

Student: Yuan-Zhu Guo
Advisors: Prof. Stefan M. Moser
Prof. Po-Ning Chen

國立交通大學
電信工程研究所
碩士論文

A Thesis
Submited to Institute of Communication Engineering
College of Electrical Engineering and Computer Science
National Chiao Tung University
in partial Fulfillment of the Requirements
for the Degree of
Master of Science
in Communication Engineering
July 2014
Hsinchu, Taiwan, Republic of China

中華民國一百零三年七月

**Information Theory Laboratory**

Institute of Communication Engineering
National Chiao Tung University

Master Project

# The Asymptotic Capacity of Noncoherent Single-Input Multiple-Output Fading Channels with Memory and Feedback

郭沅竹
Yuan-Zhu Guo

| | |
|---|---|
| Advisor: | Prof. Dr. Stefan M. Moser<br>National Chiao Tung University, Taiwan |
| Co-Advisor: | Prof. Dr. Po-Ning Chen<br>National Chiao Tung University, Taiwan |
| Graduation Committee: | Prof. Dr. Hsiao-Feng Lu<br>National Chiao Tung University, Taiwan |
| | Prof. Dr. Ying Li<br>Yuan Ze University, Taiwan |

# 非同調單輸入多輸出
# 回授記憶性衰減通道之漸近通道容量

研究生：郭沅竹　　　　　　　　　　指導教授：莫詩台方　博士

國立交通大學電信工程研究所碩士班

## 中文摘要

在本篇論文中，我們針對一個廣義非同調規律、單輸入多輸出、且具有迴授系統的記憶性衰減通道，作漸近通道容量(asymptotic channel capacity)的探討。我們假設通道的衰減過程可以是任意地穩定態且均勻遍歷(ergodic)的隨機過程，同時此隨機過程的能量與微分熵量比率(differential entropy rate)皆是有限的。而對於迴授系統的通道部份，則假設其是無任何雜訊影響的，即有無限的通道容量，但是具有因果關係的(causal)。

研究結果顯示，具有迴授系統的漸近通道容量依然隨著能量呈雙對數(double-logarithmically) 的速度成長。此外，我們也證明，在漸進通道容量展開式中的第二項，通稱爲衰減數(fading number)，與沒有迴授系統時，兩者的衰減數是一樣的。

# The Asymptotic Capacity of Noncoherent Single-Input Multiple-Output Fading Channels with Memory and Feedback

Student: Guo Yuan-Zhu Advisor: Prof. Stefan M. Moser

Institute of Communication Engineering
National Chiao Tung University

# Abstract

In this thesis, the channel capacity of a general noncoherent regular single-input multiple-output fading channel with memory and with feedback is investigated. The fading process is assumed to be a general stationary and ergodic random process of finite energy and finite differential entropy rate. The feedback is assumed to be noisefree (i.e., it is of infinite capacity), but causal.

We show that the asymptotic capacity grows double-logarithmically in the power and that the second term in the asymptotic expansion, the *fading number*, is unchanged with respect to the same channel without feedback.

# Acknowledgments

This work was from Stefan, my advisor; I was there like an observer. But the opportunity being able to observe closely how a great researcher worked is the best gift from this two and half year's studies. Stefan have shown me how to live and work happily and energetically, not only for himself, but also for the people around him. He let me know even the smartest person has to work diligently. There is no free lunch in research and life. He also gave me free space to do whatever I liked, letting me recover from biased learning attitude and realized that: learning is not about showing off, learning is not about comparison, learning is about fun and happiness.

My lab members also helped me a lot. They let me feel at home when I am in the lab. Especially thanks to Hsuan-Yin, for all the technical support.

My roommates: 大卡, 小花, and 阿將, gave me the best memory in NCTU.

Finally, thanks to my family, for providing me a place to refresh and recover when I was exhausted.

Hsinchu, 9 July 2014

Guo Yuan-Zhu

# Contents

# Chapter 1

# Introduction

Noncoherent multiple-antenna fading channel models have attracted a lot of attention for quite some years because they realistically describe the omnipresent mobile wireless communication channels. Here, *noncoherent* refers to the fundamental assumption that transmitter and receiver only have knowledge about the distribution of the fading process, but have no direct access to the current realization. Hence, the communication system needs to provide some means of measuring the current channel state, thereby using part of the available bandwidth, power, and computational efforts for the channel state estimation.

This is in stark contrast to the coherent fading models where it is assumed that the receiver has *free* and *noiseless* access to the current fading realization [1]. It is particularly the latter assumption of *perfect* knowledge of the fading realization that leads to overly optimistic capacity results for coherent channel models with respect to what can be expected to be seen in practice.

The noncoherent channel models can be split into different families. For so-called *underspread fading channels*, it is assumed that the fading process is wide-sense stationary and uncorrelated in the delay, where the product of the delay and Doppler spread is small (for more details, see [2] and references therein). The *block-fading* models assume that for a certain time, the fading realization remains unchanged before a new (potentially dependent) value is taken on [3], [4], [5]. In *nonregular* fading, the fading process is assumed to be stationary with strong memory that permits a quite precise prediction of the present fading values from the past [6], [7]. It might be even the case that one can perfectly compute the current values from the infinite past with a zero prediction error. Note, however, that due to the noncoherence assumption and due to the additive noise, the receiver never has access to the exact past fading values, but only to a noisy observation of them.

In this thesis we investigate the family of noncoherent *regular* fading channels. In contrast to nonregular fading, here it is assumed that the stationary fading process has a finite differential entropy rate. In [8] it has been shown that the capacity of multiple-antenna regular fading channels only grows *double-logarithmically* in the available power at high signal-to-noise ratios (SNR). This is much slower than the common logarithmic growth, e.g., of coherent fading channels, and it persists independently of the number of antennas used at transmitter and receiver and independently of the memory in the fading process.

For a more precise description of this phenomena, [8] defined the *fading number* $\chi$ as the second term in the high-SNR asymptotic expansion of the channel capacity:

$$\chi(\{\mathbb{H}_k\}) \triangleq \varlimsup_{\mathcal{E}\uparrow\infty} \{\mathsf{C}(\mathcal{E}) - \log\log\mathcal{E}\}. \tag{1.1}$$

An analytic expression for its value for general multiple-input multiple-output fading channels with memory has been derived in [8], [9].

While the assumption of a noncoherent communication system is realistic, we also should take into account that many practical communication systems are bidirectional allowing to send feedback from the receiver back to the transmitter. Such a feedback link will help to simplify the necessary coding scheme and it even has the potential to increase the channel capacity. In this thesis, we investigate the impact of feedback in the situation of a general regular single-input multiple-output (SIMO) fading channel with memory. We do not restrict the exact distribution of the fading process, apart from it being stationary and ergodic. Concerning the feedback, we assume the rather unrealistic situation of a feedback link that has infinite capacity. This will lead to an upper bound on the capacity in the presence of any practical type of feedback. The only constraint we make is *causality*, i.e., the feedback will arrive at the transmitter delayed by one time-step.

The structure of this thesis is as follows: In the remainder of this chapter we will shortly describe our notation. In Chapter 2 we will specify the channel model in detail. In Chapter 3, we will show some mathematical tools that are related to our analysis. In Chapter 4 summarizes the results for the channel model without feedback including some required definitions and some explanations about the meaning of the fading number. The main result, i.e., the exact asymptotic capacity of SIMO fading channels with noiseless feedback, is then presented in Chapter 5. In Chapter 6 we give the detailed derivation of our result, and Chapter 7 contains some concluding remarks.

In order to make this thesis easier to read, we attempt to use a consistent and precise notation. For random quantities, we use upper-case letters such as $X$ to denote scalar random variables, and their realizations are written in lower-case, e.g., $x$. For random vectors we use bold-face capitals, e.g., $\mathbf{X}$ and bold lower-case for their realization.

Some exceptions that are widely used in literature and therefore kept in their customary shape are as follows:

- $h(\cdot)$ denotes the differential entropy of a continuous random variable.
- $I(\cdot;\cdot)$ denotes the mutual information.

The letter $\mathsf{C}$ denotes the channel capacity. The energy per symbol is denoted by $\mathcal{E}$. Also note that we use $\log(\cdot)$ to denote the natural logarithmic function and all rates are specified in nats.

# Chapter 2

# Channel Model

We consider a communication system as shown in Figure 2.1. A message $M$ is transmitted over a SIMO fading channel with memory where the transmitter has one antenna and the receiver has $n_{\mathrm{R}}$ antennas. The channel output vector $\mathbf{Y}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ at time $k$ is given by

$$\mathbf{Y}_k = \mathbf{H}_k x_k + \mathbf{Z}_k, \tag{2.1}$$

where $x_k \in \mathbb{C}$ denotes the time-$k$ channel input; the random vector $\mathbf{H}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ denotes the time-$k$ fading vector with $n_{\mathrm{R}}$ components corresponding to the $n_{\mathrm{R}}$ antennas at the receiver; and the random vector $\mathbf{Z}_k \in \mathbb{C}^{n_{\mathrm{R}}}$ models additive noise.

We assume that the additive noise process $\{\mathbf{Z}_k\}$ is spatially and temporally independent and identically distributed (IID), circularly-symmetric, and complex Gaussian with zero mean and with variance $\sigma^2 > 0$:

$$\{\mathbf{Z}_k\} \text{ IID} \sim \mathcal{N}_{\mathbb{C}}\big(\mathbf{0}, \sigma^2 \mathsf{I}_{n_{\mathrm{R}}}\big). \tag{2.2}$$

Here, $\mathsf{I}_{n_{\mathrm{R}}}$ denotes the $n_{\mathrm{R}} \times n_{\mathrm{R}}$ identity matrix.

The fading process $\{\mathbf{H}_k\}$ is statistically independent of $\{\mathbf{Z}_k\}$ and is assumed to be stationary, ergodic, of finite energy $\mathsf{E}\big[\|\mathbf{H}_k\|^2\big] < \infty$, and of finite differential entropy rate

$$h(\{\mathbf{H}_k\}) > -\infty. \tag{2.3}$$

A random process satisfying this later condition (2.3) is usually called *regular*. Note that we do not make any further assumptions about $\{\mathbf{H}_k\}$, i.e., we do not assume a particular law (like, e.g., a Gaussian distribution). In particular we do allow for arbitrary dependences between the different components $\big\{H_k^{(j)}\big\}$ corresponding to the different antennas (spatial memory) and over time (temporal memory).

We assume noncoherent communication, i.e., neither transmitter nor receiver know the realization of $\{\mathbf{H}_k\}$, they only know its law.

From the receiver to the transmitter we have a noiseless feedback link (i.e., the link has infinite capacity and allows the receiver to send everything it knows back to the transmitter). However, to preserve causality of the system, we require the feedback to

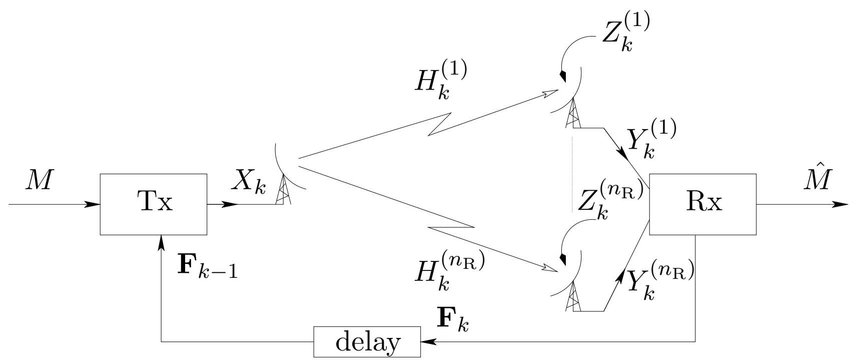


Figure 2.1: Regular SIMO fading channel with $n_R$ antennas and with noiseless causal feedback.

be delayed by one discrete time-step. So the feedback vector $\mathbf{F}_k$ that is available at the transmitter at time $k$ consists of all past channel output vectors:

$$\mathbf{F}_k = \mathbf{Y}_1^{k-1}. \tag{2.4}$$

The channel input $x_k$ at time $k$ therefore is a deterministic function of the message $M$ and the feedback $\mathbf{Y}_1^{k-1}$. Note that we assume $M$ to be uniformly distributed.

We consider two types of power constraints: an average-power constraint and a peak-power constraint. Under the former we require that for every message $m$

$$\frac{1}{n}\sum_{k=1}^{n} \mathsf{E}\left[\left|X_k\left(m, \mathbf{Y}_1^{k-1}\right)\right|^2\right] \leq \mathcal{E}, \tag{2.5}$$

where $n$ denotes the blocklength. Under the peak-power constraint we replace (2.5) with the almost-sure constraint that for every message $m$

$$\left|X_k\left(m, \mathbf{Y}_1^{k-1}\right)\right|^2 \leq \mathcal{E}, \quad \text{a.s.,} \ k = 1, \ldots, n. \tag{2.6}$$

To clarify notation we will use a subscript "FB" whenever feedback is available, while the subscript "IID" refers to a situation without memory or feedback. RHS stands for 'right-hand side'.

# Chapter 3

# Mathematical Preliminaries

In this chapter, we show some mathematical tools that will be used in our proof.

## 3.1 Differential Entropy

### 3.1.1 $h^+(\mathbf{X})$

The differntial entropy $h(\mathbf{X})$ of an n-dimensional real random vector $\mathbf{X}$ is defined if the density $p_{\mathbf{x}}(\mathbf{x})$ (with respect to the Lebesque measure on $\mathbb{R}^n$) is defined and if at least one of the integrals

$$h^+(\mathbf{X}) \triangleq \int_{\{\mathbf{x}\in\mathbb{R}^n : 0<p_{\mathbf{x}}(\mathbf{x})<1\}} p_{\mathbf{x}}(\mathbf{x}) \log \frac{1}{p_{\mathbf{x}}(\mathbf{x})} \, \mathrm{d}\mathbf{x} \tag{3.1}$$

$$h^-(\mathbf{X}) \triangleq \int_{\{\mathbf{x}\in\mathbb{R}^n : p_{\mathbf{x}}(\mathbf{x})>1\}} p_{\mathbf{x}}(\mathbf{x}) \log p_{\mathbf{x}}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \tag{3.2}$$

is finite. In this case, $h(\mathbf{X})$ is defined as the difference between the two nonnegative integrals,

$$h(\mathbf{X}) \triangleq h^+(\mathbf{X}) - h^-(\mathbf{X}) \tag{3.3}$$

where we use the rules $+\infty - a = +\infty$ and $a - \infty = -\infty$ for all $a \in \mathbb{R}$. This is written as

$$h(\mathbf{X}) = \int_{\mathbb{R}^n} p_{\mathbf{x}}(\mathbf{x}) \log \frac{1}{p_{\mathbf{x}}(\mathbf{x})} \, \mathrm{d}\mathbf{x}. \tag{3.4}$$

The differential entropy of an $n$-dimensional *complex* random variable is defined as the differential entropy of the $2n$-dimensional real vector comprising of the real and imaginary parts of each of its components. Finally, the differential entropy $h(\mathbb{H})$ of a random matrix $\mathbb{H}$ is the differential entropy of the vector comprising of its entries.

### 3.1.2 $h_\lambda(\cdot)$

Let $\hat{\mathbf{X}}_k$ denote the unit vector

$$\hat{\mathbf{X}}_k \triangleq \frac{\mathbf{X}_k}{\|\mathbf{X}_k\|}. \tag{3.5}$$

Because the unit vectors $\hat{\mathbf{X}}_k$ only take value on the unit sphere in $\mathbb{C}^{n_\mathrm{R}}$ and since the surface of this unit sphere has zero measure over $\mathbb{C}^{n_\mathrm{R}}$, we define a differential entropy-like quantity $h_\lambda(\cdot)$ that only lives on the surface of the unit sphere in $\mathbb{C}^{n_\mathrm{R}}$: for $\mathbf{V} \in \mathbb{C}^{n_\mathrm{R}}$ and $\hat{\mathbf{V}} \triangleq \frac{\mathbf{V}}{\|\mathbf{V}\|}$, we define

$$h_\lambda(\hat{\mathbf{V}}) \triangleq \mathsf{E}\left[-\log p^\lambda_{\hat{\mathbf{V}}}(\hat{\mathbf{V}})\right] = -\int p^\lambda_{\hat{\mathbf{V}}}(\hat{\mathbf{v}}) \log p^\lambda_{\hat{\mathbf{V}}}(\hat{\mathbf{v}}) d\hat{\mathbf{v}}, \tag{3.6}$$

if the expectation exists. Here $p^\lambda_{\hat{\mathbf{V}}}(\hat{\mathbf{v}})$ denotes the PDF of the random unit-vector $\hat{\mathbf{V}}$ with respect to the $\mathbb{C}^{n_\mathrm{R}}$-surface measure $\lambda$. Note that $p^\lambda_{\hat{\mathbf{V}}}(\hat{\mathbf{v}})$ is implicitly defined by the PDF of $\mathbf{V}$, $p^\lambda_{\mathbf{V}}(\mathbf{v})$. For more details we refer to [9, Sec. II].

**Lemma 3.1** *Let* $\mathbf{V}$ *be a complex random vector taking value in* $\mathbb{C}^m$ *and of differntial entropy* $h(\mathbf{V})$*. Then*

$$h(\mathbf{V}) = h(\|\mathbf{V}\|) + h_\lambda(\hat{\mathbf{V}}|\|\mathbf{V}\|) + (2m-1)\mathsf{E}[\log\|\mathbf{V}\|] \tag{3.7}$$

*whenever all the quantities in* (3.7) *are defined. Here the first term on the right is the differential entropy of* $\|\mathbf{V}\|$ *when viewed as a real (scalar) random variable.*

Note that it is a conditional version of $h_\lambda$.

### 3.1.3 Differential Entropy and Expectation of Logarithms

**Lemma 3.2** *Let* $\mathbf{X}$ *be an* $n$*-dimensional complex random vector of density* $p_{\mathbf{X}}(\mathbf{x})$*. Then the following relationship between differential entropy and the expected log-norm hold: If* $h^-(\mathbf{X}) < \infty$*, then for any* $0 < \alpha < n$ *there exists some finite number* $\Delta(n, \alpha)$ *(not depending on the law of* $\mathbf{X}$*) such that*

$$\mathsf{E}[\log\|\mathbf{X}\|] \geq -\frac{1}{\alpha}h^-\big(\mathbf{X}\big) - \Delta(n,\alpha) \tag{3.8}$$

*Proof:* See [10, Appendix. A.4.4]. □

## 3.2 Markov's Inequality

**Lemma 3.3 (Markov's Inequality)** *For any non-negative random variable* $V$ *and any constant* $\delta > 0$*,*

$$\Pr[V \geq \delta] \leq \frac{\mathsf{E}[V]}{\delta} \tag{3.9}$$

*Proof:* See for example [11]. □

## 3.3 Causal Interpretation

Massey [12], [13] shows a way of graphically determining independence of random variables based on *causal interpretations*. A causal interpretation is an ordered list of random variables. The idea behind a specific choice of order lies in the causality of the system. Loosely speaking in an engineering way of thinking, we would like to think of some random variables being generated "first" and some "later based on" the generation of the others. Note that *a priori* every ordered list is a valid causal interpretation, but some choices will be more useful keeping the engineering idea in mind.

As an example consider the vector

$$\mathbf{V} = (M, X_1^k, \mathbf{Y}_1^k, \mathbf{H}_1^k, \mathbf{Z}_1^k, \mathbf{F}_1^k), \tag{3.10}$$

where all components are random variables defined in Chapter 2.

For simplicity assume for the moment that all components take value in discrete alphabets[1]. We choose the following causal interpretation:

$$(M, \mathbf{H}_1, ..., \mathbf{H}_k, \mathbf{Z}_1, ..., \mathbf{Z}_k, \mathbf{F}_1, X_1, \mathbf{Y}_1, \mathbf{F}_2, X_2, \mathbf{Y}_2, ..., \mathbf{F}_k, X_k, \mathbf{Y}_k). \tag{3.11}$$

If we consider now the entropy of $\mathbf{V}$ and write it as a sum using the chain rule

$$H(\mathbf{V}) = \sum_j H\big(V^{(j)}\big|V^{(1)}, ..., V^{(j-1)}\big), \tag{3.12}$$

then we see that our choice of a causal interpretation for $\mathbf{V}$ simplifies the expression for the entropy significantly:

$$\begin{aligned} H(\mathbf{V}) = H(M) &+ H(\mathbf{H}_1) + H(\mathbf{H}_2|\mathbf{H}_1) + \cdots + H(\mathbf{H}_k|\mathbf{H}_1^k) \\ &+ H(\mathbf{Z}_1) + \cdots + H(\mathbf{Z}_k) + H(\mathbf{F}_1) \\ &+ H(X_1|\mathbf{F}_1, M) + H(\mathbf{Y}_1|X_1, \mathbf{H}_1, \mathbf{Z}_1) + H(\mathbf{F}_2|\mathbf{Y}_1) \\ &+ H(X_2|\mathbf{F}_1^2, M) + H(\mathbf{Y}_2|X_2, \mathbf{H}_2, \mathbf{Z}_2) + \cdots + H(\mathbf{F}_k|\mathbf{Y}_1^{k-1}) \\ &+ H(X_k|\mathbf{F}_1^k, M) + H(\mathbf{Y}_k|X_k, \mathbf{H}_k, \mathbf{Z}_k) \end{aligned} \tag{3.13}$$

Massey calls this a *causal-order expansion* of $H(\mathbf{V})$. It can easily be depicted graphically in a *causality graph*, which is a directed graph with an edge from vertex $\mathbf{V}^{(j_1)}$ to $\mathbf{V}^{(j_2)}$ if and only if $\mathbf{V}^{(j_1)}$ is in the conditioning expression for $H\left(V^{(j_2)}\middle|V^{(1)}, ..., V^{(j_2-1)}\right)$. We shall say that a vertex $\mathbf{V}^{(j_1)}$ is *causally prior* to vertex $\mathbf{V}^{(j_2)}$ if there is a directed path from $\mathbf{V}^{(j_1)}$ to $\mathbf{V}^{(j_2)}$.

In our case the corresponding graph of (3.11) is shown in Figure 3.2.

Note that once we have established the graph, we do not consider the entropy anymore. We only used the entropy in order to be able to invoke the chain rule in establishing the "dependencies" between the different components.

[1] We will drop this assumption soon again, however, here it simplifies notation considerably because we need not worry about differential entropy. In the end, we are not interested in the entropy at all, but in the "dependencies" between the components.

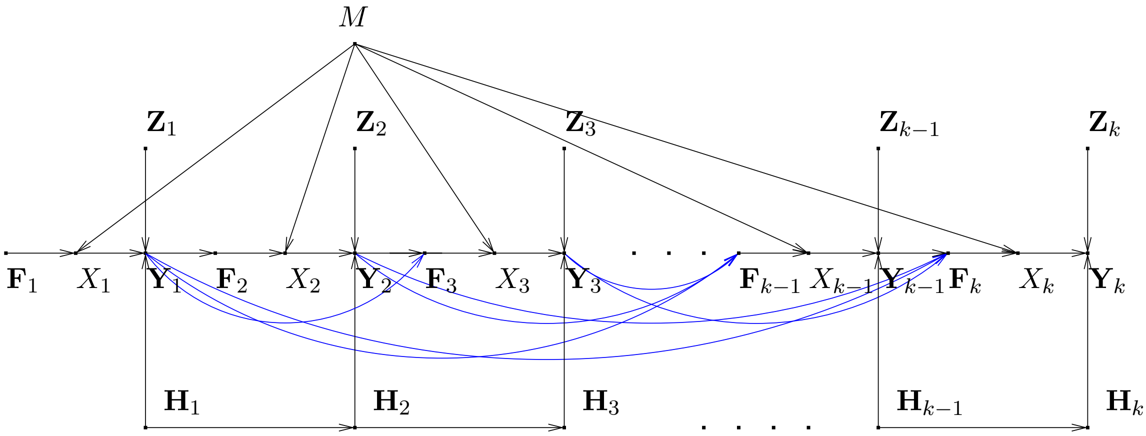


Figure 3.2: The causality graph of our model.

A causality graph is very useful when determining the statistical dependence between two groups of random variables possibly conditioned on a third group.

To state this property in more clarity let $\mathcal{A}, \mathcal{B}, \mathcal{C} \subset \{1, ..., \text{length}(\mathbf{V})\}$ be three index sets. Let $\mathbf{V}^{(\mathcal{A})}$ denote a vector containing as components of all components of $\mathbf{V}$ whose indices are in $\mathcal{A}$, similarly, define $\mathbf{V}^{(\mathcal{B})}$ and $\mathbf{V}^{(\mathcal{C})}$.

Any causality graph of $\mathbf{V}$ can now be used in order to investigate the independence of $\mathbf{V}^{(\mathcal{A})}$ and $\mathbf{V}^{(\mathcal{B})}$ when conditioned on $\mathbf{V}^{(\mathcal{C})}$. To that goal consider the following procedure:

- from the specify causality graph take the *subgraph causally relevant*[2] to $\mathbf{V}^{(\mathcal{A}\cup\mathcal{B}\cup\mathcal{C})}$;
- delete all edges *leaving* any component of $\mathbf{V}^{(\mathcal{C})}$;
- drop all directions of the remaining edges;
- if now all components of $\mathbf{V}^{(\mathcal{A})}$ are unconnected to the components of $\mathbf{V}^{(\mathcal{B})}$, then $\mathbf{V}^{(\mathcal{A})}$ is statistically independent of $\mathbf{V}^{(\mathcal{B})}$ when conditioned on $\mathbf{V}^{(\mathcal{C})}$.

Note that using this procedure we only make statements about the independence, but not about possible dependences, i.e., if the components of $\mathbf{V}^{(\mathcal{A})}$ and $\mathbf{V}^{(\mathcal{B})}$ are not disconnected, then they might be statistically dependent or independent.

[2] A subgraph causally relevant to some $\tilde{\mathbf{V}}$ consists of all those vertices that are either components of $\tilde{\mathbf{V}}$ or causally prior to $\tilde{\mathbf{V}}$ in the given causal interpretation, together with the edges connecting these vertices.

# Chapter 4

# Capacity and Fading Number without Feedback

It has been shown in [8] that the capacity of general regular SIMO fading channels under either an average-power constraint or a peak-power constraint is

$$\mathsf{C}(\mathcal{E}) = \log(1 + \log(1 + \mathcal{E})) + \chi(\{\mathbf{H}_k\}) + o(1), \tag{4.1}$$

where $o(1)$ denotes terms that tend to zero as $\mathcal{E}$ tends to infinity, and

$$\overline{\lim_{\mathcal{E}\uparrow\infty}} \{\mathsf{C}(\mathcal{E}) - \log\log\mathcal{E}\} < 0. \tag{4.2}$$

Therefore, we can define

$$\chi(\{\mathbf{H}_k\}) \triangleq \overline{\lim_{\mathcal{E}\uparrow\infty}} \{\mathsf{C}(\mathcal{E}) - \log\log\mathcal{E}\} \tag{4.3}$$

$$= h_\lambda\left(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0} \middle| \{\hat{\mathbf{H}}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\infty}^{-1}\right) - \log 2 + n_\mathrm{R}\mathsf{E}\left[\log\|\mathbf{H}_0\|^2\right] - h\left(\mathbf{H}_0 \middle| \mathbf{H}_{-\infty}^{-1}\right) \tag{4.4}$$

where the second equality is given by [8] and $h_\lambda(\cdot)$ is defined in 3.1.2. Here, $\{\Theta_k\}$ is IID $\sim \mathcal{U}((-\pi, \pi])$ and independent of $\{\mathbf{H}_k\}$.

From (4.1) it is obvious that the capacity of the fading channel (2.1) grows extremely slowly at large power. Indeed, $\log(1 + \log(1 + \mathcal{E}))$ grows so slowly that, for the smallest values of $\mathcal{E}$ for which $o(1) \approx 0$, the (constant!) fading number $\chi$ usually is much larger than $\log(1 + \log(1 + \mathcal{E}))$. Hence, the threshold between the low-power regime and the capacity-inefficient high-power regime is strongly related to the fading number: the larger the fading number is, the higher the rate can be chosen without operating the system in the inefficient double-logarithmic regime.

Also note that even though the double-logarithmic term on the RHS of (4.1) does not depend on $\{\mathbf{H}_k\}$ or, particularly, on the number of antennas $n_\mathrm{R}$, it is still beneficial to have multiple antennas because the fading number $\chi$ does depend strongly on the fading process and the number of antennas.

From (4.4) one also sees that in the case of a memoryless SIMO fading channel, the fading number is given by

$$\chi_{\text{IID}}(\mathbf{H}) = h_\lambda\bigl(\hat{\mathbf{H}}e^{\mathrm{i}\Theta}\bigr) - \log 2 + n_{\text{R}}\mathsf{E}\bigl[\log\|\mathbf{H}\|^2\bigr] - h(\mathbf{H}), \tag{4.5}$$

and that therefore the fading number in (4.4) also can be written as

$$\chi(\{\mathbf{H}_k\}) = \chi_{\text{IID}}(\mathbf{H}_0) + I\left(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\right) - I\left(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0};\bigl\{\hat{\mathbf{H}}_\ell e^{\mathrm{i}\Theta_\ell}\bigr\}_{\ell=-\infty}^{-1}\right). \tag{4.6}$$

In [8], it has also been shown that for an arbitrary value of the power $\mathcal{E}$, the channel capacity can be bounded as follows:

$$\mathsf{C}(\mathcal{E}) \leq \mathsf{C}_{\text{IID}}(\mathcal{E}) + I\bigl(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\bigr), \quad \mathcal{E} \geq 0. \tag{4.7}$$

From (4.6) we see that this upper bound may not be tight. In particular, asymptotically for $\mathcal{E} \to \infty$ it is strictly loose.

# Chapter 5

# Capacity and Fading Number with Feedback

While it is well-known that feedback has no effect on the capacity of a memoryless channel, in general feedback does increase capacity for channels with memory. The reason for this is that the combination of feedback and memory allows the transmitter to predict the current channel state and thereby adapt to it. Unfortunately, for regular fading channels this increase in capacity due to the feedback turns out to be very limited.

**Theorem 5.1 (Capacity Increase by Feedback is Bounded by a Constant)** *Let a general SIMO fading channel with memory be defined as in* (2.1) *and consider a noiseless causal feedback link as described in* (2.4) *(see Figure 2.1). Then the channel capacity under either an average-power constraint* (2.5) *or a peak-power constraint* (2.6) *is upper-bounded as follows:*

$$\mathsf{C}_{\text{FB}}(\mathcal{E}) \leq \mathsf{C}_{\text{IID}}(\mathcal{E}) + I\big(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}\big), \quad \mathcal{E} \geq 0. \tag{5.1}$$

*Proof:* When we defined channel capacity, we relied on a result by Dobrusin [14] which shows that for information stable channels the capacity is given by

$$\mathsf{C} = \lim_{n\to\infty} \frac{1}{n} \sup_{Q\in\mathcal{P}(\mathcal{X}^n)} I(X_1^n; Y_1^n), \tag{5.2}$$

where $\mathcal{P}(\mathcal{X}^n)$ is the set of all probability measures $Q$ over $\mathcal{X}^n$ satisfying the given input constraints.

Here, however, we have feedback and can therefore not rely on the above result, but have to derive a new converse to the coding theorem for the new situation.

Note that since the channel capacity under a peak-power constraint $\mathcal{E}$ cannot be larger than the capacity under an average-power constraint $\mathcal{E}$, all upper bounds that are based on an average-power constraint are also valid for the situation with a peak-power constraint. We will therefore in the following only consider an average-power constraint.

Hence, assume that there is a sequence of code schemes with $\lfloor e^{n\mathsf{R}_{\text{FB}}}\rfloor$ codewords of blocklength $n$—i.e., for each $n$ the rate of the code is not larger that $\mathsf{R}_{\text{FB}}$—such that the probability of error

$$P_e^{(n)} \triangleq \Pr[\hat{M} \neq M] \tag{5.3}$$

tends to zero as $n$ tends to infinity. Then

$$H(M) = \log\lfloor e^{n\mathsf{R}_{\text{FB}}}\rfloor \geq \log(e^{n\mathsf{R}_{\text{FB}}} - 1) = n\mathsf{R}_{\text{FB}} - \epsilon_n, \tag{5.4}$$

where $\epsilon_n \to 0$ as $n \to \infty$. Therefore,

$$\begin{aligned} \mathsf{R}_{\text{FB}} &\leq \frac{1}{n}H(M) + \frac{\epsilon_n}{n} && (5.5)\\ &= \frac{1}{n}I(M;\mathbf{Y}_1^n) + \frac{1}{n}H(M|\mathbf{Y}_1^n) + \frac{\epsilon_n}{n} && (5.6)\\ &= \frac{1}{n}I(M;\mathbf{Y}_1^n) + \frac{\log 2 + P_e^{(n)}\log\lfloor e^{n\mathsf{R}_{\text{FB}}}\rfloor}{n} + \frac{\epsilon_n}{n} && (5.7)\\ &= \frac{1}{n}I(M;\mathbf{Y}_1^n) + \frac{\log 2}{n} + \frac{P_e^{(n)}n\mathsf{R}_{\text{FB}}}{n} + \frac{\epsilon_n}{n} && (5.8)\\ &= \frac{1}{n}I(M;\mathbf{Y}_1^n) + \frac{\log 2}{n} + P_e^{(n)}\mathsf{R}_{\text{FB}} + \frac{\epsilon_n}{n} && (5.9) \end{aligned}$$

Here (5.6) follows from that definition of mutual information, and the subsequent inequality from Fano's inequality.

Therefore, for $n \to \infty$ we must have

$$\mathsf{R}_{\text{FB}} \leq \lim_{n\to\infty} \frac{1}{n}I(M;\mathbf{Y}_1^n). \tag{5.10}$$

Hence, any upper bound on the RHS of (5.10) will yield an upper bound on channel capacity in presence of feedback. We will therefore continue with bounding $I(M;\mathbf{Y}_1^n)$:

$$\begin{aligned} \frac{1}{n}I(M;\mathbf{Y}_1^n) &= \frac{1}{n}\sum_{k=1}^{n} I(M;\mathbf{Y}_k|\mathbf{Y}_1^{k-1}) && (5.11)\\ &= \frac{1}{n}\sum_{k=1}^{n}\Big(I(M,\mathbf{Y}_1^{k-1};\mathbf{Y}_k) - I(\mathbf{Y}_1^{k-1};\mathbf{Y}_k)\Big) && (5.12)\\ &\leq \frac{1}{n}\sum_{k=1}^{n} I(M,\mathbf{Y}_1^{k-1};\mathbf{Y}_k) && (5.13)\\ &\leq \frac{1}{n}\sum_{k=1}^{n} I(M,\mathbf{Y}_1^{k-1},\mathbf{H}_1^{k-1};\mathbf{Y}_k) && (5.14)\\ &= \frac{1}{n}\sum_{k=1}^{n} I(M,\mathbf{Y}_1^{k-1},\mathbf{H}_1^{k-1},X_k;\mathbf{Y}_k) && (5.15)\\ &= \frac{1}{n}\sum_{k=1}^{n}\Big(I(\mathbf{H}_1^{k-1},X_k;\mathbf{Y}_k) + I(M,\mathbf{Y}_1^{k-1};\mathbf{Y}_k|\mathbf{H}_1^{k-1},X_k)\Big) && (5.16)\\ &= \frac{1}{n}\sum_{k=1}^{n} I(\mathbf{H}_1^{k-1},X_k;\mathbf{Y}_k) && (5.17)\\ &= \frac{1}{n}\sum_{k=1}^{n}\Big(I(X_k;\mathbf{Y}_k) + I(\mathbf{H}_1^{k-1},\mathbf{Y}_k|X_k)\Big). && (5.18) \end{aligned}$$

Here the first two equalities follow from the chain rule; the subsequent inequality from the non-negativity of mutual information; the following inequality from adding more terms; the subsequent equality follows since $X_k$ is a deterministic function of $M$ and $\mathbf{Y}_1^{k-1}$ (and hypothetically also $\mathbf{H}_1^{k-1}$); then we have used the chain rule again; (5.17) follows since[3]

$$I(M, \mathbf{Y}_1^{k-1}; \mathbf{Y}_k | \mathbf{H}_1^{k-1}, X_k) = 0 \tag{5.19}$$

and finally we have used the chain rule once more.

We have to take into account that $X_k$ depends on past outputs via the feedback in the next step.

$$\begin{aligned}
I(\mathbf{H}_1^{k-1}; \mathbf{Y}_k | X_k) &\leq I(\mathbf{H}_1^{k-1}; \mathbf{Y}_k, \mathbf{H}_k | X_k) && (5.20)\\
&= I(\mathbf{H}_1^{k-1}; \mathbf{H}_k | X_k) + I(\mathbf{H}_1^{k-1}; \mathbf{Y}_k | X_k, \mathbf{H}_k) && (5.21)\\
&= I(\mathbf{H}_1^{k-1}; \mathbf{H}_k | X_k) && (5.22)\\
&= h(\mathbf{H}_k | X_k) - h(\mathbf{H}_k | \mathbf{H}_1^{k-1}, X_k) && (5.23)\\
&\leq h(\mathbf{H}_k) - h(\mathbf{H}_k | \mathbf{H}_1^{k-1}, X_k) && (5.24)\\
&= h(\mathbf{H}_k) - h(\mathbf{H}_k | \mathbf{H}_1^{k-1}) && (5.25)\\
&= I(\mathbf{H}_k; \mathbf{H}_1^{k-1}) && (5.26)
\end{aligned}$$

Here the first inequality follows from adding one more term; the subsequent equality follows from the chain rule; (5.22) follows since

$$I(\mathbf{H}_1^{k-1}; \mathbf{Y}_k | X_k, \mathbf{H}_k) = 0 \tag{5.27}$$

which can be seen similarly to (5.19); (5.24) is due to conditioning that reduces entropy; and the subsequent equality holds since conditional on $\mathbf{H}_1^{k-1}$, $X_k$ and $\mathbf{H}_k$ are independent.

Together with (5.18) this yields

$$\begin{aligned}
\frac{1}{n} I(M; \mathbf{Y}_1^n) &\leq \frac{1}{n} \sum_{k=1}^{n} \Big( I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_k; \mathbf{H}_1^{k-1}) \Big) && (5.28)\\
&\leq \frac{1}{n} \sum_{k=1}^{n} \Big( I(X_k; \mathbf{Y}_k) + I(\mathbf{H}_k; \mathbf{H}_{-\infty}^{k-1}) \Big) && (5.29)\\
&\leq \frac{1}{n} \sum_{k=1}^{n} \mathrm{C}_{\mathrm{IID}}(\mathcal{E}_k) + I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) && (5.30)
\end{aligned}$$

where in the last inequality we have used the stationarity of $\{\mathbf{H}_k\}$ and used $\mathrm{C}_{\mathrm{IID}}(\mathcal{E}_k)$ to denote the capacity without feedback or memory for a given power $\mathcal{E}_k$. Note that the

[3] To see this keep in mind that $\mathbf{Y}_k$ is fully determined by $\mathbf{Z}_k$, $\mathbf{H}_k$, and $X_k$. The noise $\mathbf{Z}_k$ is independent of everything else and can therefore not be estimated from any other random variable; $X_k$ is given; only $\mathbf{H}_k$ is not known. However, it can be approximated using the past $\mathbf{H}_1^{k-1}$ which again are given. Therefore, conditional on $\mathbf{H}_1^{k-1}$ and $X_k$, $M$ and $\mathbf{Y}_1^{k-1}$ are independent of $\mathbf{Y}_k$. This statement can also be proven graphically using a technique based on *causal interpretations*, see Section 3.3.

power allocation depends on the feedback. However, due to (2.5) $\mathcal{E}_k$ must satisfy

$$\frac{1}{n}\sum_{k=1}^{n}\mathcal{E}_k \le \mathcal{E}. \tag{5.31}$$

Using this together with Jensen's inequality relying on the concavity of channel capacity in the power, we get

$$\frac{1}{n}I(M;\mathbf{Y}_1^n) \le \mathsf{C}_{\text{IID}}\left(\frac{1}{n}\sum_{k=1}^{n}\mathcal{E}_k\right) + I(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}) \tag{5.32}$$

$$\le \mathsf{C}_{\text{IID}}(\mathcal{E}) + I(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}) \tag{5.33}$$

where in the second inequality we used the fact that $\mathsf{C}_{\text{IID}}(\cdot)$ is nondecreasing.

Therefore,

$$\mathsf{R}_{\text{FB}}(\mathcal{E}) \le \lim_{n\to\infty}\frac{1}{n}I(M;\mathbf{Y}_1^n) \le \mathsf{C}_{\text{IID}}(\mathcal{E}) + I(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}) \tag{5.34}$$

which proves (5.1).

□

We note that the RHS of (5.1) is identical to the RHS of (4.7). Hence the same (alas potentially loose) bound holds both for the channel capacity with and without feedback. Moreover, also note that $\mathsf{C}(\mathcal{E})$ trivially is a lower bound to $\mathsf{C}_{\text{FB}}(\mathcal{E})$ since the transmitter can simply ignore the feedback and achieve the same results as without feedback.

An immediate consequence of Theorem 5.1 is that $\mathsf{C}_{\text{FB}}(\mathcal{E})$ only grows double-logarithmically in the power at high power and therefore there exists a fading number $\chi_{\text{FB}}(\{\mathbf{H}_k\})$ with a definition as follows:

**Corollary 5.2** *Because*

$$\varlimsup_{\mathcal{E}\uparrow\infty}\{\mathsf{C}_{\text{FB}}(\mathcal{E}) - \log\log\mathcal{E}\} < 0, \tag{5.35}$$

*we define*

$$\chi_{\text{FB}}(\{\mathbf{H}_k\}) \triangleq \varlimsup_{\mathcal{E}\uparrow\infty}\{\mathsf{C}_{\text{FB}}(\mathcal{E}) - \log\log\mathcal{E}\} \tag{5.36}$$

$$\tag{5.37}$$

Theorem 5.1 can then be applied to $\chi_{\text{FB}}(\{\mathbf{H}_k\})$.

**Corollary 5.3** *Using the same result as in Theorem 5.1, we learn*

$$\chi_{\text{FB}}(\{\mathbf{H}_k\}) \le \chi_{\text{IID}}(\{\mathbf{H}_k\}) + I\big(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\big). \tag{5.38}$$

Next, we state a stronger statement.

**Theorem 5.4 (SIMO Fading Number with Feedback)** *Let a general SIMO fading channel with memory be defined as in* (2.1) *and consider a noiseless causal feedback link as described in* (2.4) *(see Figure 2.1). Then the asymptotic channel capacity under either an average-power constraint* (2.5) *or a peak-power constraint* (2.6) *is identical to the asymptotic channel capacity for the channel without feedback:*

$$\mathsf{C}_{\text{FB}}(\mathcal{E}) = \log(1+\log(1+\mathcal{E})) + \chi_{\text{FB}}(\{\mathbf{H}_k\}) + o(1) \tag{5.39}$$

*where the fading number is*

$$\begin{aligned}\chi_{\text{FB}}(\{\mathbf{H}_k\}) &= \chi(\{\mathbf{H}_k\}) \\ &= h_\lambda\left(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\,\middle|\,\{\hat{\mathbf{H}}_\ell e^{\mathrm{i}\Theta_\ell}\}_{\ell=-\infty}^{-1}\right) - \log 2 \\ &\quad + n_{\text{R}}\mathsf{E}\left[\log\|\mathbf{H}_0\|^2\right] - h\left(\mathbf{H}_0\,\middle|\,\mathbf{H}_{-\infty}^{-1}\right).\end{aligned} \tag{5.40}$$

We would like to point out that this result even holds in the (hypothetical) case when the feedback is improved in the sense that in addition to the past channel outputs the transmitter also is informed about the past fading realizations $\mathbf{H}_1^{k-1}$. Note further that since we have assumed the most optimistic form of causal feedback, any type of realistic feedback will yield the same result.

We would like to give a hand-waving explanation of this behavior. 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 as can be seen as follows: for any constant $\beta > 0$ ($\beta$ can be chosen arbitrarily large!),

$$\varlimsup_{\mathcal{E}\uparrow\infty}\{\log\log\beta\mathcal{E} - \log\log\mathcal{E}\} = \varlimsup_{\mathcal{E}\uparrow\infty}\{\log(\log\beta + \log\mathcal{E}) - \log\log\mathcal{E}\} \tag{5.41}$$

$$= \varlimsup_{\mathcal{E}\uparrow\infty}\{\log(\log\mathcal{E}) - \log\log\mathcal{E}\} \tag{5.42}$$

$$= 0. \tag{5.43}$$

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

# Chapter 6

# Proof of Theorem 5.4

## 6.1 Main Line Through the Proof

Since the channel capacity of the system without feedback trivially is a lower bound on the channel capacity with feedback, and since the capacity under a peak-power constraint is a lower bound on the capacity with an average-power constraint, it is sufficient to derive an upper bound on $\chi_{\text{FB}}(\{\mathbf{H}_k\})$ under the assumption of the average-power constraint (2.5) and to show that it is identical to the fading number without feedback and under the assumption of a peak-power constraint.

The proof is very lengthy and we therefore outline the main ideas in the beginning. The basic structure follows the proof of the general fading number of MIMO fading channels with memory given in [9]. However, there are many details that need to be adapted and taken care of. Particularly, we have to consider the following challenges:

- Due to the feedback, the channel input, the fading, and the additive noise become dependent.
- We cannot rely on the important auxiliary result given in [9, Th. 3] that shows that the optimal input is stationary.
- We cannot rely on the important auxiliary result given in [15, Th. 8] that shows that the capacity-achieving input distribution *escapes to infinity*.

To handle the first challenge, we often rely on the concept of *causal interpretations*, which is introduced in Chapter 3.3, [12], [13]. This is a tool that allows to graphically proof the independence of random variables when conditioned on certain other random variables.

The missing auxiliary result concerning the capacity-achieving input distribution escaping to infinity can be proven indirectly inside of the derivation.

The biggest difficulty is caused by the nonstationarity of the channel input that is inherent to the given context because the transmitter continuously learns more about the fading process through the feedback and thereby changes the optimal distribution of the input.

The proof starts with Fano's inequality (see (5.9)), which states that any given sequence of communication systems with rate $\mathsf{R}_{\text{FB}}$ and power $\mathcal{E}$ must satisfy

$$\begin{aligned}
\mathsf{R}_{\text{FB}}(\mathcal{E}) &\leq \frac{1}{n} I\big(M;\mathbf{Y}_1^n\big) + \frac{\log 2}{n} + P_{\text{e}}^{(n)}\mathsf{R}_{\text{FB}}(\mathcal{E}) + \frac{\epsilon_n}{n} && (6.1)\\
&= \frac{1}{n}\sum_{k=1}^{n} I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}\big) + \frac{\log 2}{n} + P_{\text{e}}^{(n)}\mathsf{R}_{\text{FB}}(\mathcal{E}) + \frac{\epsilon_n}{n} && (6.2)\\
&= \frac{1}{n}\sum_{k=1}^{\kappa} I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}\big) + \frac{n-\kappa}{n-\kappa}\frac{1}{n}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}\big)\\
&\quad + \frac{\log 2}{n} + P_{\text{e}}^{(n)}\mathsf{R}_{\text{FB}}(\mathcal{E}) + \frac{\epsilon_n}{n} && (6.3)\\
&\leq \frac{1}{n}\sum_{k=1}^{\kappa}\Big(\mathsf{C}_{\text{IID}}(\mathcal{E}_k) + I\big(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\big)\Big) + \frac{n-\kappa}{n}\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}\big)\\
&\quad + \frac{\log 2}{n} + P_{\text{e}}^{(n)}\mathsf{R}_{\text{FB}}(\mathcal{E}) + \frac{\epsilon_n}{n} && (6.4)
\end{aligned}$$

In (6.3), we separate the sum into two parts. The first part, $1 \leq k \leq \kappa$, can be considered as transient state. Since $\kappa$ is a constant, it is bounded anyway and in (6.4) we bound the mutual information term in the sum as in (5.30). Using Jensen's inequality relying on the concavity of channel capacity in the power, we get

$$\begin{aligned}
\mathsf{R}_{\text{FB}}(\mathcal{E}) &\leq \frac{\kappa}{n}\mathsf{C}_{\text{IID}}\left(\frac{1}{\kappa}\sum_{k=1}^{\kappa}\mathcal{E}_k\right) + \frac{\kappa}{n} I\big(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\big) + \frac{n-\kappa}{n}\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}\big)\\
&\quad + \frac{\log 2}{n} + P_{\text{e}}^{(n)}\mathsf{R}_{\text{FB}}(\mathcal{E}) + \frac{\epsilon_n}{n}. && (6.5)
\end{aligned}$$

Next we focus on $\kappa+1 \leq k \leq n$ and bound as follows:

$$\begin{aligned}
I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}\big) &\leq I\big(M;\mathbf{Y}_k, G_k\big|\mathbf{Y}_1^{k-1}\big) && (6.6)\\
&= I\big(M;G_k\big|\mathbf{Y}_1^{k-1}\big) + I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}, G_k\big) && (6.7)\\
&= H\big(G_k\big|Y_1^{k-1}\big) - \underbrace{H\big(G_k\big|M, Y_1^{k-1}\big)}_{\geq 0} + \gamma_k I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}, G_k = 1\big)\\
&\quad + (1-\gamma_k) I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}, G_k = 0\big). && (6.8)
\end{aligned}$$

Here in (6.6), we add the indicator random variable $G_k$ that is defined as

$$G_k \triangleq \begin{cases} 1 & \text{if } \|\mathbf{H}_k\|^2 \geq t,\\ 0 & \text{otherwise,}\end{cases} \tag{6.9}$$

for some given $t > 0$. We will choose $t$ large such that

$$\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t} \leq 0.5 \tag{6.10}$$

Moreover, we define

$$\gamma_k \triangleq \Pr[G_k = 1] = \Pr\big[\|\mathbf{H}_k\|^2 \geq t\big], \tag{6.11}$$

and note that by the Markov inequality (Lemma 3.3),

$$\gamma_k = \Pr\big[\|\mathbf{H}_k\|^2 \geq t\big] \leq \frac{\mathsf{E}\big[\|\mathbf{H}_k\|^2\big]}{t}. \tag{6.12}$$

Because

$$H\big(G_k\big|\mathbf{Y}_1^{k-1}\big) \leq H(G_k) = H_\mathrm{b}(\gamma_k) \leq H_\mathrm{b}\left(\frac{\mathsf{E}\big[\|\mathbf{H}_k\|^2\big]}{t}\right) \tag{6.13}$$

where $H_\mathrm{b}(\cdot)$ denoting the binary entropy function, with $t$ large enough and by using (5.30) again with conditioning on $G_k = 1$, we further bound (6.8) as follows:

$$\begin{aligned}
&I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}\big)\\
&\leq H_\mathrm{b}(\gamma_k) + \gamma_k \mathsf{C}_\mathrm{IID}(\mathcal{E}_k|G_k=1) + \gamma_k I\big(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\big|G_0=1\big)\\
&\quad + \underbrace{(1-\gamma_k)}_{\leq 1} I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}, G_k=0\big) \qquad (6.14)\\
&\leq H_\mathrm{b}\left(\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\right) + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\tilde{\mathsf{C}}_\mathrm{IID}(t\mathcal{E}_k) + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t} I\big(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\big|G_0=1\big)\\
&\quad + I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}, G_k=0\big). \qquad (6.15)
\end{aligned}$$

Here in (6.15), $\tilde{\mathsf{C}}_\mathrm{IID}(\cdot)$ is the capacity of a channel

$$\tilde{\mathbf{Y}}_k = \frac{\mathbf{Y}_k}{t} = \frac{\mathbf{H}_k}{t}x_k + \frac{\mathbf{Z}_k}{t} \triangleq \tilde{\mathbf{H}}_k x_k + \tilde{\mathbf{Z}}_k \tag{6.16}$$

where we condition on the event that $G_k = 1$, i.e., $\|\mathbf{H}_k\|^2 > t$. This is a different regular fading channel, for which we know

$$\lim_{\tilde{\mathcal{E}}\to\infty} \{\tilde{\mathsf{C}}_\mathrm{IID}(\tilde{\mathcal{E}}) - \log\log\tilde{\mathcal{E}}\} \leq \infty. \tag{6.17}$$

Hence,

$$\frac{1}{n}\sum_{k=\kappa+1}^{n} \tilde{\mathsf{C}}_\mathrm{IID}(t\cdot\mathcal{E}_k) \leq \frac{1}{n}\sum_{k=1}^{n} \tilde{\mathsf{C}}_\mathrm{IID}(t\cdot\mathcal{E}_k) \leq \tilde{\mathsf{C}}_\mathrm{IID}\left(t\cdot\frac{1}{n}\sum_{k=1}^{n}\mathcal{E}_k\right)$$

where the last inequality follows by Jensen's inequality relying on the concavity of channel capacity in the power again. Putting everything back into (6.5), we get

$$\begin{aligned}
\mathsf{R}_\mathrm{FB}(\mathcal{E}) &\leq \frac{\kappa}{n}\mathsf{C}_\mathrm{IID}\left(\frac{1}{\kappa}\sum_{k=1}^{\kappa}\mathcal{E}_k\right) + \frac{\kappa}{n} I\big(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\big)\\
&\quad + \frac{n-\kappa}{n}\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}, G_k=0\big) + \frac{n-\kappa}{n} H_\mathrm{b}\left(\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\right)\\
&\quad + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t}\tilde{\mathsf{C}}_\mathrm{IID}\left(t\cdot\frac{1}{n}\sum_{k=1}^{n}\mathcal{E}_k\right) + \frac{n-\kappa}{n}\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{t} I\big(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\big|G_0=1\big)\\
&\quad + \frac{\log 2}{n} + P_\mathrm{e}^{(n)}\mathsf{R}_\mathrm{FB}(\mathcal{E}) + \frac{\epsilon_n}{n}. \qquad (6.18)
\end{aligned}$$

The third term in (6.18) is then bounded as follows:

$$
\begin{aligned}
&I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1},G_k=0\big)\\
&\quad= I\big(M,\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big|G_k=0\big) - I\big(\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big|G_k=0\big) &(6.19)\\
&\quad\le I\big(M,\mathbf{Y}_1^{k-1},X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|G_k=0\big) - I\big(\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big|G_k=0\big) &(6.20)\\
&\quad= I\big(X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|G_k=0\big) + \underbrace{I\big(M,\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big|X_k,\mathbf{H}_1^{k-1},G_k=0\big)}_{=0,\text{ see Appendix D}}\\
&\qquad - I\big(\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big|G_k=0\big) &(6.21)\\
&\quad\le I\big(E_k,X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|G_k=0\big) - I\big(\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big|G_k=0\big) &(6.22)\\
&\quad= I\big(E_k;\mathbf{Y}_k\big|G_k=0\big) + I(X_k;\mathbf{Y}_k|E_k,G_k=0)\\
&\qquad + I\big(\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|X_k,E_k,G_k=0\big) - I\big(\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big|G_k=0\big) &(6.23)\\
&\quad\le H_\mathrm{b}(\beta_k) + \beta_k I(X_k;\mathbf{Y}_k|E_k=1,G_k=0)\\
&\qquad + \beta_k I\big(\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|X_k,E_k=1,G_k=0\big) - I\big(\mathbf{Y}_1^{k-1};\mathbf{Y}_k,G_k=0\big)\\
&\qquad + (1-\beta_k) I\big(X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|E_k=0,G_k=0\big). &(6.24)
\end{aligned}
$$

Here in (6.20) the current input $X_k$ and the past fading values $\mathbf{H}_1^{k-1}$ are added. In (6.22) we add the indicator random variable $E_k$ that is defined as

$$
E_k \triangleq \begin{cases} 1 & \text{if } |X_\ell| \ge \xi_{\min},\ \forall\,\ell=1,\ldots,k,\\ 0 & \text{otherwise,}\end{cases} \tag{6.25}
$$

for some given $\xi_{\min} \ge 0$. Moreover, $\beta_k \triangleq \Pr[E_k=1|G_k=0]$. Finally, (6.24) follows because we bound

$$
I(E_k;\mathbf{Y}_k\big|G_k=0) = H(E_k\big|G_k=0) - \underbrace{H(E_k|\mathbf{Y}_k,G_k=0)}_{\ge 0} \le H(E_k) = H_\mathrm{b}(\beta_k). \tag{6.26}
$$

Note that the three middle terms on the RHS of (6.24) correspond to a memoryless term, a term with memory, and a correction term, respectively. We will show in Section 6.2 that the second term on the RHS of (6.24) can be bounded as follows:

$$
\begin{aligned}
&I\left(X_k;\mathbf{Y}_k|E_k=1,G_k=0\right)\\
&= h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\Big|E_k=1,G_k=0\right) - h(\mathbf{H}_k\big|X_k,E_k=1,G_k=0) - \log 2\\
&\quad + n_\mathrm{R}\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|E_k=1,G_k=0\big]\\
&\quad + \mu\Big(\log\eta - \mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|E_k=1,G_k=0\big] - \mathsf{E}\big[\log|X_k|^2\big|E_k=1,G_k=0\big]\Big)\\
&\quad + \log\Gamma\left(\mu,\frac{\nu}{\eta}\right) + \epsilon_{\nu,k} + \frac{1}{\eta}\mathsf{E}\big[\|\mathbf{H}_k^2\||X_k|^2\big|E_k=1,G_k=0\big] + \frac{\nu}{\eta},
\end{aligned} \tag{6.27}
$$

the third term on the right hand side of (6.24) as

$$
\begin{aligned}
&I\big(\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|X_k,E_k=1,G_k=0\big)\\
&\quad\le h(\mathbf{H}_k\big|X_k,E_k=1,G_k=0) - h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},G_k=0\big),
\end{aligned} \tag{6.28}
$$

and the fourth term on the right hand side of (6.24) as

$$\begin{aligned}
&I\big(\mathbf{Y}_1^{k-1};\mathbf{Y}_k\big|G_k=0\big)\\
&\quad\geq \beta_k h_\lambda\big(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\big|E_k=1,G_k=0\big) - h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, G_k=0\right)\\
&\qquad + (1-\beta_k)h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{k-\kappa}^{k-1}, E_k=0, G_k=0\right)\\
&\qquad - 3H_{\mathrm{b}}(\beta_k) - \delta_1(\kappa,\xi_{\min}) - \delta_2(\kappa,\xi_{\min}).
\end{aligned} \tag{6.29}$$

Plugging (6.27), (6.28), and (6.29) back to (6.24), we get

$$\begin{aligned}
&I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1},G_k=0\big)\\
&\leq H_{\mathrm{b}}(\beta_k) + \underline{\beta_k h_\lambda\big(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}|E_k=1,G_k=0\big)} - \underline{\beta_k h(\mathbf{H}_k|X_k,E_k=1,G_k=0)} - \beta_k\log 2\\
&\quad + \beta_k n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|E_k=1,G_k=0\big]\\
&\quad + \beta_k\mu\Big(\log\eta - \mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|E_k=1,G_k=0\big] - \mathsf{E}\big[\log|X_k|^2|E_k=1,G_k=0\big]\Big)\\
&\quad + \beta_k\log\Gamma\left(\mu,\frac{\nu}{\eta}\right) + \beta_k\epsilon_{\nu,k} + \beta_k\frac{1}{\eta}\mathsf{E}\big[\|\mathbf{H}_k^2\||X_k|^2\big|E_k=1,G_k=0\big] + \beta_k\frac{\nu}{\eta}\\
&\quad + \underline{\beta_k h(\mathbf{H}_k|X_k,E_k=1,G_k=0)} - \beta_k h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},G_k=0\big)\\
&\quad - \underline{\beta_k h_\lambda\big(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}|E_k=1,G_k=0\big)} + h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, G_k=0\right)\\
&\quad - (1-\beta_k)h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, E_k=0,G_k=0\right)\\
&\quad + 3H_{\mathrm{b}}(\beta_k) + \delta_1(\kappa,\xi_{\min}) + \delta_2(\kappa,\xi_{\min})\\
&\quad + (1-\beta_k)I\big(X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|E_k=0,G_k=0\big).
\end{aligned} \tag{6.30}$$

Note that the four underlining terms in (6.30) cancel each other, and that

$$\begin{aligned}
&\beta_k n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|E_k=1,G_k=0\big]\\
&\quad = n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|G_k=0\big] - (1-\beta_k)n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|E_k=0,G_k=0\big],\\
&-\mu\beta_k\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|E_k=1,G_k=0\big]\\
&\quad = -\mu\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|G_k=0\big] + \mu(1-\beta_k)\mathsf{E}\big[\log\|\mathbf{H}_k\|^2\big|E_k=0,G_k=0\big].
\end{aligned}$$

Moreover,

$$\mathsf{E}\big[\log|X_k|^2|E_k=1,G_k=0\big] \geq \log\xi_{\min}^2,$$

and by (B.10) and (B.11)

$$\begin{aligned}
\beta_k\epsilon_{\nu,k} &= \sup_{\gamma\geq\xi_{\min}}\Big\{\beta_k\mathsf{E}\big[\log(\|\mathbf{H}_k\|^2\gamma^2+\nu)\big|E_k=1,G_k=0\big]\\
&\qquad\qquad - \beta_k\mathsf{E}\big[\log(\|\mathbf{H}_k\|^2\gamma^2)\big|E_k=1,G_k=0\big]\Big\} \qquad (6.31)\\
&\leq \sup_{\gamma\geq\xi_{\min}}\Big\{\beta_k\mathsf{E}\big[\log(\|\mathbf{H}_k\|^2\gamma^2+\nu)\big|E_k=1,G_k=0\big]\\
&\qquad\qquad - \beta_k\mathsf{E}\big[\log(\|\mathbf{H}_k\|^2\gamma^2)\big|E_k=1,G_k=0\big]
\end{aligned}$$

$$
\begin{aligned}
&\qquad + (1-\beta_k)\mathsf{E}\left[\log(\|\mathbf{H}_k\|^2\gamma^2+\nu)\middle|E_k=0,G_k=0\right] \\
&\qquad - (1-\beta_k)\mathsf{E}\left[\log(\|\mathbf{H}_k\|^2\gamma^2)\middle|E_k=0,G_k=0\right]\Big\} \\
&= \sup_{\gamma\geq\xi_{\min}} \left\{\mathsf{E}\left[\log(\|\mathbf{H}_0\|^2\gamma^2+\nu)\middle|G_k=0\right] - \mathsf{E}\left[\log(\|\mathbf{H}_0\|^2\gamma^2)\middle|G_k=0\right]\right\} \qquad (6.32)\\
&= \epsilon_\nu \qquad (6.33)
\end{aligned}
$$

where (6.32) follows because we add something positive ($\nu \geq 0$, chosen freely). Therefore, (6.30) becomes

$$
\begin{aligned}
&I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1},G_k=0\big) \\
&\quad\leq H_\mathrm{b}(\beta_k) - \beta_k\log 2 + n_\mathrm{R}\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|G_k=0\right] \\
&\qquad - (1-\beta_k)\,n_\mathrm{R}\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|E_k=0,G_k=0\right] + \beta_k\mu\log\eta - \mu\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|G_k=0\right] \\
&\qquad + \mu(1-\beta_k)\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|E_k=0,G_k=0\right] - \beta_k\mu\log\xi_{\min}^2 \\
&\qquad + \beta_k\log\Gamma\left(\mu,\frac{\nu}{\eta}\right) + \epsilon_\nu + \beta_k\frac{1}{\eta}\mathsf{E}\left[\|\mathbf{H}_k^2\||X_k|^2\middle|E_k=1,G_k=0\right] + \beta_k\frac{\nu}{\eta} \\
&\qquad - \beta_k h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},G_k=0\big) + h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1},G_k=0\right) \\
&\qquad - (1-\beta_k)h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1},E_k=0,G_k=0\right) \\
&\qquad + 3H_\mathrm{b}(\beta_k) + \delta_1(\kappa,\xi_{\min}) + \delta_2(\kappa,\xi_{\min}) \\
&\qquad + (1-\beta_k)I\big(X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|E_k=0,G_k=0\big) \qquad (6.34)\\
&\quad= 4H_\mathrm{b}(\beta_k) + \delta_1(\kappa,\xi_{\min}) + \delta_2(\kappa,\xi_{\min}) + \epsilon_\nu \\
&\qquad + h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1},G_k=0\right) - \beta_k h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},G_k=0\big) \\
&\qquad + n_\mathrm{R}\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|G_k=0\right] \\
&\qquad + \mu\Big(\beta_k\log\eta - \mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|G_k=0\right] \\
&\qquad\qquad + (1-\beta_k)\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|E_k=0,G_k=0\right] - \beta_k\log\xi_{\min}^2\Big) \\
&\qquad + \beta_k\log\Gamma\left(\mu,\frac{\nu}{\eta}\right) + \frac{\beta_k}{\eta}\mathsf{E}\left[\|\mathbf{H}_k\|^2|X_k|^2|E_k=1,G_k=0\right] + \beta_k\frac{\nu}{\eta} - \beta_k\log 2 \\
&\qquad + (1-\beta_k)\Bigg(I\big(X_k,\mathbf{H}_1^{k-1};\mathbf{Y}_k\big|E_k=0,G_k=0\big) \\
&\qquad\qquad - h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1},E_k=0,G_k=0\right) \\
&\qquad\qquad - n_\mathrm{R}\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|E_k=0,G_k=0\right]\Bigg). \qquad (6.35)
\end{aligned}
$$

Here, in (6.35) we arithmetically rearrange the terms. We further bound the last term in (6.35) as follows:

$$
\begin{aligned}
&(1-\beta_k)\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|E_k=0,G_k=0\right] \\
&\quad\leq (1-\beta_k)\log\mathsf{E}\left[\|\mathbf{H}_k\|^2\middle|E_k=0,G_k=0\right] \qquad (6.36)
\end{aligned}
$$

$$\leq (1-\beta_k)\log \frac{\mathsf{E}\left[\|\mathbf{H}_k\|^2 \middle| G_k = 0\right]}{(1-\beta_k)} \tag{6.37}$$

$$= (1-\beta_k)\log \mathsf{E}\left[\|\mathbf{H}_0\|^2 \middle| G_0 = 0\right] - (1-\beta_k)\log(1-\beta_k) \tag{6.38}$$

$$\leq (1-\beta_k)\log \mathsf{E}\left[\|\mathbf{H}_0\|^2 \middle| G_0 = 0\right] - e^{-1}\log e^{-1} \tag{6.39}$$

$$= (1-\beta_k)\log \mathsf{E}\left[\|\mathbf{H}_0\|^2 \middle| G_0 = 0\right] + \frac{1}{e}, \tag{6.40}$$

where (6.36) follows from Jensen's inequality; (6.37) follows because

$$\begin{aligned}\mathsf{E}\left[\|\mathbf{H}_k\|^2 \middle| G_k = 0\right] &= \beta_k \mathsf{E}\left[\|\mathbf{H}_k\|^2 \middle| E_k = 1, G_k = 0\right] \\ &\quad + (1-\beta_k)\mathsf{E}\left[\|\mathbf{H}_k\|^2 \middle| E_k = 0, G_k = 0\right] &(6.41)\\ &\geq (1-\beta_k)\mathsf{E}\left[\|\mathbf{H}_k\|^2 \middle| E_k = 0, G_k = 0\right]; &(6.42)\end{aligned}$$

(6.38) follows because $\{\mathbf{H}_k\}$ is a stationary process; and (6.39) follows because the function $x \mapsto x\log x$ has its minimum when $x = e$. Putting (6.40) back into (6.35) and using the stationarity property of $\{\mathbf{H}_k\}$ again, we get

$$\begin{aligned}&I\left(M; \mathbf{Y}_k \middle| \mathbf{Y}_1^{k-1}, G_k = 0\right) \\ &\quad \leq 4H_\mathrm{b}(\beta_k) + \delta_1(\kappa, \xi_{\min}) + \delta_2(\kappa, \xi_{\min}) + \epsilon_\nu \\ &\qquad + h_\lambda\left(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0} \middle| \{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\}_{l=-\kappa}^{-1}, G_0 = 0\right) - \beta_k h\left(\mathbf{H}_0 \middle| \mathbf{H}_{-k+1}^{-1}, G_0 = 0\right) \\ &\qquad + n_\mathrm{R}\mathsf{E}\left[\log \|\mathbf{H}_0\|^2 \middle| G_0 = 0\right] \\ &\qquad + \mu\Bigg(\beta_k \log\eta - \mathsf{E}\left[\log\|\mathbf{H}_0\|^2 \middle| G_0 = 0\right] \\ &\qquad\qquad + (1-\beta_k)\log \mathsf{E}\left[\|\mathbf{H}_0\|^2 \middle| G_0 = 0\right] + \frac{1}{e} - \beta_k \log \xi_{\min}^2\Bigg) \\ &\qquad + \beta_k \log\Gamma\left(\mu, \frac{\nu}{\eta}\right) + \frac{\beta_k}{\eta}\mathsf{E}\left[\|\mathbf{H}_k\|^2|X_k|^2 \middle| E_k = 1, G_k = 0\right] + \beta_k\frac{\nu}{\eta} - \beta_k \log 2 \\ &\qquad + (1-\beta_k)\Bigg( I\left(X_k, \mathbf{H}_1^{k-1}; \mathbf{Y}_k \middle| E_k = 0, G_k = 0\right) \\ &\qquad\qquad - h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k} \middle| \left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, E_k = 0, G_k = 0\right) \\ &\qquad\qquad - n_\mathrm{R}\mathsf{E}\left[\log\|\mathbf{H}_k\|^2 \middle| E_k = 0, G_k = 0\right]\Bigg).\end{aligned} \tag{6.43}$$

Next note that

$$\begin{aligned}&\frac{\beta_k}{\eta}\mathsf{E}\left[\|\mathbf{H}_k\|^2|X_k|^2 \middle| E_k = 1, G_k = 0\right] \\ &\quad = \frac{1}{\eta}\mathsf{E}\left[\|\mathbf{H}_k\|^2|X_k|^2 \middle| G_k = 0\right] - \frac{1-\beta_k}{\eta}\mathsf{E}\left[\|\mathbf{H}_k\|^2|X_k|^2 \middle| E_k = 0, G_k = 0\right] &(6.44)\\ &\quad \leq \frac{1}{\eta}\mathsf{E}\left[\|\mathbf{H}_k\|^2|X_k|^2 \middle| G_k = 0\right]. &(6.45)\end{aligned}$$

Moreover in Appendix A, we further bound the last part of (6.43) as follows:

$$\begin{aligned}
&I\big(X_k, \mathbf{H}_1^{k-1}; \mathbf{Y}_k \big| E_k = 0, G_k = 0\big) - h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k} \middle| \left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, E_k = 0, G_k = 0\right) \\
&-n_\mathrm{R}\mathsf{E}\big[\log\|\mathbf{H}_k\|^2 \big| E_k = 0, G_k = 0\big] \\
&\quad \le \mathcal{C}_\mathrm{IID}(\xi_\mathrm{min}\big|G_k = 0) - (n_\mathrm{R}-1)h(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, G_0 = 0) \\
&\qquad - h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\mathbf{H}_{-\infty}^{-1}, G_0 = 0\big) - h(\mathbf{H}_0\big|\mathbf{H}_{-k+1}^{-1}, G_0 = 0) \\
&\qquad + \frac{n_\mathrm{R}(n_\mathrm{R}+1)}{e} + n_\mathrm{R}^2\log^+\left(\frac{\pi e}{n_\mathrm{R}}\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2|G_0 = 0\big]}{1-\beta_k}\right) + n_\mathrm{R}\Delta(n_\mathrm{R}, 1).
\end{aligned} \tag{6.46}$$

where

$$\log^+(x) \triangleq \max\{0, \log(x)\} \tag{6.47}$$

and $\Delta(n_\mathrm{R}, 1)$ is some finite number. Therefore, we get

$$\begin{aligned}
&I\big(M; \mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}, G_k = 0\big) \\
&\quad \le 4H_\mathrm{b}(\beta_k) + \delta_1(\kappa, \xi_\mathrm{min}) + \delta_2(\kappa, \xi_\mathrm{min}) + \epsilon_\nu \\
&\qquad + h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\}_{l=-\kappa}^{-1}, G_0 = 0\big) - \underline{\beta_k h\big(\mathbf{H}_0\big|\mathbf{H}_{-k+1}^{-1}, G_0 = 0\big)} \\
&\qquad + n_\mathrm{R}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0 = 0\big] \\
&\qquad + \mu\Big(\beta_k\log\eta - \mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0 = 0\big] \\
&\qquad\qquad + (1-\beta_k)\log\mathsf{E}\big[\|\mathbf{H}_0\|^2\big|G_0 = 0\big] + \frac{1}{e} - \beta_k\log\xi_\mathrm{min}^2\Big) \\
&\qquad + \beta_k\log\Gamma\left(\mu, \frac{\nu}{\eta}\right) + \frac{1}{\eta}\mathsf{E}\big[\|\mathbf{H}_k\|^2|X_k|^2\big|G_k = 0\big] + \beta_k\frac{\nu}{\eta} - \beta_k\log 2 \\
&\qquad + (1-\beta_k)\bigg(\mathcal{C}_\mathrm{IID}(\xi_\mathrm{min}\big|G_k = 0) - (n_\mathrm{R}-1)h(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, G_0 = 0) \\
&\qquad\qquad - h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\mathbf{H}_{-\infty}^{-1}, G_k = 0\big) - \underline{h(\mathbf{H}_0\big|\mathbf{H}_{-k+1}^{-1}, G_0 = 0)} \\
&\qquad\qquad + \frac{n_\mathrm{R}(n_\mathrm{R}+1)}{e} + n_\mathrm{R}^2\log^+\left(\frac{\pi e}{n_\mathrm{R}}\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2|G_0 = 0\big]}{1-\beta_k}\right) + n_\mathrm{R}\Delta(n_\mathrm{R}, 1)\bigg)
\end{aligned} \tag{6.48}$$

$$\begin{aligned}
&\quad \le 4H_\mathrm{b}(\beta_k) + \delta_1(\kappa, \xi_\mathrm{min}) + \delta_2(\kappa, \xi_\mathrm{min}) + \epsilon_\nu \\
&\qquad + h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\}_{l=-\kappa}^{-1}, G_0 = 0\big) - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, G_0 = 0\big) \\
&\qquad + n_\mathrm{R}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0 = 0\big] \\
&\qquad + \mu\Big(\beta_k\log\eta - \mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0 = 0\big] \\
&\qquad\qquad + (1-\beta_k)\log\mathsf{E}\big[\|\mathbf{H}_0\|^2\big|G_0 = 0\big] + \frac{1}{e} - \beta_k\log\xi_\mathrm{min}^2\Big) \\
&\qquad + \beta_k\log\Gamma\left(\mu, \frac{\nu}{\eta}\right) + \frac{1}{\eta}\mathsf{E}\big[\|\mathbf{H}_k\|^2|X_k|^2\big|G_k = 0\big] + \beta_k\frac{\nu}{\eta} - \beta_k\log 2 \\
&\qquad + (1-\beta_k)\bigg(\mathcal{C}_\mathrm{IID}(\xi_\mathrm{min}\big|G_k = 0) - (n_\mathrm{R}-1)h(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, G_0 = 0)
\end{aligned}$$

$$
\begin{aligned}
&- h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\mathbf{H}_{-\infty}^{-1}, G_k = 0\big) \\
&+ \frac{n_\mathrm{R}(n_\mathrm{R}+1)}{e} + n_\mathrm{R}^2 \log^+\left(\frac{\pi e}{n_\mathrm{R}} \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2|G_0=0\right]}{1-\beta_k}\right) + n_\mathrm{R}\Delta(n_\mathrm{R},1)\Bigg),
\end{aligned}
\tag{6.49}
$$

where (6.49) follows because two underlining terms in (6.48) combine to $h\big(\mathbf{H}_0\big|\mathbf{H}_{-k+1}^{-1}, G_0 = 0\big)$.

Defining

$$
\beta \triangleq \frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n} \beta_k, \tag{6.50}
$$

using Jensen's inequality for the binary entropy function, adding the sum from (6.18) in front of (6.49), we get

$$
\begin{aligned}
&\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1}, G_k=0\big) \\
&\quad\le 4H_\mathrm{b}(\beta) + \delta_1(\kappa,\xi_{\min}) + \delta_2(\kappa,\xi_{\min}) + \epsilon_\nu + h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\}_{l=-\kappa}^{-1}, G_0=0\big) \\
&\qquad - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, G_0=0\big) + n_\mathrm{R}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2, G_0=0\big] \\
&\qquad + \mu\big(\beta\log\eta - \mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0=0\big] + (1-\beta)\log\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0=0\big] \\
&\qquad\qquad + \frac{1}{e} - \beta\log\xi_{\min}^2\big) \\
&\qquad + \beta\log\Gamma\left(\mu,\frac{\nu}{\eta}\right) + \frac{1}{\eta}\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n}\mathsf{E}\big[\|\mathbf{H}_k\|^2|X_k|^2\big|G_k=0\big] + \beta\frac{\nu}{\eta} - \beta\log 2 \\
&\qquad + (1-\beta)\Big(\mathcal{C}_\mathrm{IID}(\xi_{\min}\big|G_0=0) - (n_\mathrm{R}-1)h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, G_0=0\big) \\
&\qquad\qquad - h_\lambda(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\mathbf{H}_{-\infty}^{-1}, G_0=0) + \frac{n_\mathrm{R}(n_\mathrm{R}+1)}{e} + n_\mathrm{R}\Delta(n_\mathrm{R},1)\Big) \\
&\qquad + \frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n}(1-\beta_k)n_\mathrm{R}^2\log^+\left(\frac{\pi e}{n_\mathrm{R}}\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big|G_0=0\big]}{1-\beta_k}\right).
\end{aligned}
\tag{6.51}
$$

Because for $\beta_k \ge 1 - \frac{\pi e \mathsf{E}[\|\mathbf{H}_0\|^2|G_0=0]}{n_\mathrm{R}}$, $(1-\beta_k)\log^+\left(\frac{\pi e}{n_\mathrm{R}}\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\middle|G_0=0\right]}{1-\beta_k}\right)$ is concave in $\beta_k$, the last term in (6.51) can be further bounded as

$$
\begin{aligned}
&\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n}(1-\beta_k)n_\mathrm{R}^2\log^+\left(\frac{\pi e}{n_\mathrm{R}}\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big|G_0=0\big]}{1-\beta_k}\right) \\
&\qquad\le (1-\beta)n_\mathrm{R}^2\log^+\left(\frac{\pi e}{n_\mathrm{R}}\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big|G_0=0\big]}{1-\beta}\right).
\end{aligned}
\tag{6.52}
$$

Moreover, in order to get rid of the dependence on the input, $\{X_k\}$ (note that $\beta$ depends on $\{X_k\}$), we add a supremum over $\beta$:

$$
\begin{aligned}
&\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1},G_k=0\big)\\
&\quad\le h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\}_{l=-\kappa}^{-1},G_0=0\big)-h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},G_0=0\big)\\
&\qquad-\log 2+n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0=0\big]\\
&\qquad+\sup_{0\le\beta\le1}\Bigg\{4H_{\mathrm{b}}(\beta)+\delta_1(\kappa,\xi_{\min})+\delta_2(\kappa,\xi_{\min})+\epsilon_\nu\\
&\qquad\qquad+\mu\big(\beta\log\eta-\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0=0\big]+(1-\beta)\log\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0=0\big]\\
&\qquad\qquad\quad+e^{-1}-\beta\log\xi_{\min}^2\big)\\
&\qquad\qquad+\beta\log\Gamma\left(\mu,\frac{\nu}{\eta}\right)+\frac{1}{\eta}\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n}\mathsf{E}\big[\|\mathbf{H}_k\|^2|X_k|^2\big|G_k=0\big]+\beta\frac{\nu}{\eta}\\
&\qquad\qquad+(1-\beta)\Big(\mathcal{C}_{\mathrm{IID}}(\xi_{\min}\big|G_0=0)-(n_{\mathrm{R}}-1)h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},G_k=0\big)+\log 2\\
&\qquad\qquad\qquad-h_\lambda(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\mathbf{H}_{-\infty}^{-1},G_0=0)+\frac{n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e}+n_{\mathrm{R}}\Delta(n_{\mathrm{R}},1)\Big)\\
&\qquad\qquad+(1-\beta)n_{\mathrm{R}}^2\log^+\left(\frac{\pi e}{n_{\mathrm{R}}}\frac{E[\|\mathbf{H}_0\|^2\big|G_0=0]}{1-\beta}\right)\Bigg\}.
\end{aligned}\tag{6.53}
$$

Because

$$\frac{1}{\eta}\mathsf{E}\big[\|\mathbf{H}_k\|^2|X_k|^2\big|G_k=0\big]\le\frac{1}{\eta}\mathsf{E}\big[t\cdot|X_k|^2\big|G_k=0\big]\le\frac{t}{\eta}\mathcal{E}_k\tag{6.54}$$

and

$$\frac{t}{\eta}\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n}\mathcal{E}_k\le\frac{t}{\eta}\frac{n}{n-\kappa}\underbrace{\frac{1}{n}\sum_{k=1}^{n}\mathcal{E}_k}_{\le\mathcal{E}}\le\frac{t}{\eta}\frac{n}{n-\kappa}\mathcal{E},\tag{6.55}$$

(6.53) becomes

$$
\begin{aligned}
&\frac{1}{n-\kappa}\sum_{k=\kappa+1}^{n} I\big(M;\mathbf{Y}_k\big|\mathbf{Y}_1^{k-1},G_k=0\big)\\
&\quad\le h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\}_{l=-\kappa}^{-1},G_0=0\big)-h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},G_0=0\big)\\
&\qquad-\log 2+n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0=0\big]\\
&\qquad+\sup_{0\le\beta\le1}\Bigg\{4H_{\mathrm{b}}(\beta)+\delta_1(\kappa,\xi_{\min})+\delta_2(\kappa,\xi_{\min})+\epsilon_\nu\\
&\qquad\qquad+\mu\big(\beta\log\eta-\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0=0\big]+(1-\beta)\log\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0=0\big]\\
&\qquad\qquad\quad+\mathrm{e}^{-1}-\beta\log\xi_{\min}^2\big)\\
&\qquad\qquad+\beta\log\Gamma\left(\mu,\frac{\nu}{\eta}\right)+\frac{t}{\eta}\frac{n}{n-\kappa}\mathcal{E}+\beta\frac{\nu}{\eta}\\
&\qquad\qquad+(1-\beta)\Big(\mathcal{C}_{\mathrm{IID}}(\xi_{\min}\big|G_0=0)-(n_{\mathrm{R}}-1)h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1},G_k=0\big)+\log 2\\
&\qquad\qquad\qquad-h_\lambda(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\mathbf{H}_{-\infty}^{-1},G_0=0)+\frac{n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e}+n_{\mathrm{R}}\Delta(n_{\mathrm{R}},1)\Big)
\end{aligned}
$$

$$+ (1-\beta)n_{\mathrm{R}}^2 \log^+ \left( \frac{\pi e}{n_{\mathrm{R}}} \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2 \middle| G_0 = 0\right]}{1-\beta} \right) \Bigg\}. \tag{6.56}$$

Let $n \to \infty$ and choose

$$\mu = \frac{\nu}{\log \mathcal{E}} \tag{6.57}$$

$$\eta = \frac{\mathcal{E} \log \mathcal{E}}{\nu} \tag{6.58}$$

$$t = \log \mathcal{E}. \tag{6.59}$$

Then the bound on the capacity with feedback in (6.18) becomes

$$\begin{aligned}
\mathsf{R}_{\mathrm{FB}}(\mathcal{E}) \le\ & h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0} \big| \{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\}_{l=-\kappa}^{-1}, G_0 = 0\big) - h\big(\mathbf{H}_0 \big| \mathbf{H}_{-\infty}^{-1}, G_0 = 0\big) \\
& - \log 2 + n_{\mathrm{R}} \mathsf{E}\left[\log \|\mathbf{H}_0\|^2 \middle| G_0 = 0\right] \\
& + \sup_{0 \le \beta \le 1} \Bigg\{ 4H_{\mathrm{b}}(\beta) + \delta_1(\kappa, \xi_{\min}) + \delta_2(\kappa, \xi_{\min}) + \epsilon_\nu \\
& \qquad + \frac{\nu}{\log \mathcal{E}} \Bigg( \beta \log \frac{\mathcal{E} \log \mathcal{E}}{\nu} - \mathsf{E}\left[\log \|\mathbf{H}_0\|^2 \middle| G_0 = 0\right] \\
& \qquad\qquad + (1-\beta) \log \mathsf{E}\left[\log \|\mathbf{H}_0\|^2 \middle| G_0 = 0\right] + e^{-1} - \beta \log \xi_{\min}^2 \Bigg) \\
& \qquad + \beta \log \Gamma \left( \frac{\nu}{\log \mathcal{E}}, \frac{\nu^2}{\mathcal{E} \log \mathcal{E}} \right) + \nu + \frac{\beta \nu^2}{\mathcal{E} \log \mathcal{E}} \\
& \qquad + (1-\beta) \Bigg( \mathcal{C}_{\mathrm{IID}}(\xi_{\min} \big| G_0 = 0) - (n_{\mathrm{R}} - 1) h\big(\mathbf{H}_0 \big| \mathbf{H}_{-\infty}^{-1}, G_0 = 0\big) \\
& \qquad\qquad - h_\lambda(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0} \big| \mathbf{H}_{-\infty}^{-1}, G_0 = 0) + \frac{n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e} \\
& \qquad\qquad + n_{\mathrm{R}} \Delta(n_{\mathrm{R}}, 1) \Bigg) \\
& \qquad + (1-\beta) n_{\mathrm{R}}^2 \log^+ \left( \frac{\pi e}{n_{\mathrm{R}}} \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2 \middle| G_0 = 0\right]}{1-\beta} \right) \Bigg\} \\
& + H_{\mathrm{b}} \left( \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{\log \mathcal{E}} \right) + \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{\log \mathcal{E}} \tilde{\mathsf{C}}_{\mathrm{IID}}(\mathcal{E} \log \mathcal{E}) \\
& + \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\right]}{\log \mathcal{E}} I\big(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1} \big| G_0 = 1\big) \qquad\qquad (6.60) \\
=\ & h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0} \big| \{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\}_{l=-\kappa}^{-1}, G_0 = 0\big) - h\big(\mathbf{H}_0 \big| \mathbf{H}_{-\infty}^{-1}, G_0 = 0\big) \\
& - \log 2 + n_{\mathrm{R}} \mathsf{E}\left[\log \|\mathbf{H}_0\|^2 \middle| G_0 = 0\right] \\
& + \sup_{0 \le \beta \le 1} \Bigg\{ 4H_{\mathrm{b}}(\beta) + \delta_1(\kappa, \xi_{\min}) + \delta_2(\kappa, \xi_{\min}) + \epsilon_\nu \\
& \qquad + \frac{\nu\beta}{\log \mathcal{E}} (\log \mathcal{E} + \log\log \mathcal{E} - \log \nu) - \frac{\nu \mathsf{E}\left[\log \|\mathbf{H}_0\|^2 \middle| G_0 = 0\right]}{\log \mathcal{E}} \\
& \qquad + \frac{(1-\beta)\nu \log \mathsf{E}\left[\log \|\mathbf{H}_0\|^2 \middle| G_0 = 0\right]}{\log \mathcal{E}} + \frac{\nu}{e \log \mathcal{E}} - \frac{\nu\beta \log \xi_{\min}^2}{\log \mathcal{E}}
\end{aligned}$$

$$
\begin{aligned}
&\qquad + \beta \log \Gamma\left(\frac{\nu}{\log \mathcal{E}}, \frac{\nu^2}{\mathcal{E}\log\mathcal{E}}\right) + \nu + \frac{\beta\nu^2}{\mathcal{E}\log\mathcal{E}} \\
&\qquad + (1-\beta)\Big(\mathcal{C}_{\mathrm{IID}}(\xi_{\min}\big|G_0=0) - (n_{\mathrm{R}}-1)h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, G_0=0\big) + \log 2 \\
&\qquad\qquad - h_\lambda(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\mathbf{H}_{-\infty}^{-1}, G_0=0) + \frac{n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e} \\
&\qquad\qquad + n_{\mathrm{R}}\Delta(n_{\mathrm{R}},1)\Big) \\
&\qquad + (1-\beta)n_{\mathrm{R}}^2 \log^+\left(\frac{\pi e}{n_{\mathrm{R}}}\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big|G_0=0\big]}{1-\beta}\right)\Bigg\} \\
&+ H_{\mathrm{b}}\left(\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{\log\mathcal{E}}\right) + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{\log\mathcal{E}}\tilde{\mathsf{C}}_{\mathrm{IID}}(\mathcal{E}\log\mathcal{E}) \\
&+ \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{\log\mathcal{E}} I\big(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\big|G_0=1\big).
\end{aligned} \tag{6.61}
$$

Note that this bound holds for any system, hence also for a capacity-achieving system. Therefore we can use (6.61) to upper-bound $\mathsf{C}_{\mathrm{FB}}(\mathcal{E})$:

$$
\chi_{\mathrm{FB}}(\{\mathbf{H}_k\}) = \varlimsup_{\mathcal{E}\uparrow\infty}\{\mathsf{C}_{\mathrm{FB}}(\mathcal{E}) - \log\log\mathcal{E}\} \tag{6.62}
$$

$$
\begin{aligned}
&\le \varlimsup_{\mathcal{E}\uparrow\infty}\Bigg\{ h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\}_{l=-\kappa}^{-1}, G_0=0\big) - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, G_0=0\big) \\
&\qquad - \log 2 + n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0=0\big] \\
&\qquad + \sup_{0\le\beta\le1}\Bigg\{ 4H_{\mathrm{b}}(\beta) + \delta_1(\kappa,\xi_{\min}) + \delta_2(\kappa,\xi_{\min}) + \epsilon_\nu \\
&\qquad\qquad + \frac{\nu\beta}{\log\mathcal{E}}(\log\mathcal{E} + \log\log\mathcal{E} - \log\nu) - \frac{\nu\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0=0\big]}{\log\mathcal{E}} \\
&\qquad\qquad + \frac{(1-\beta)\nu\log\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0=0\big]}{\log\mathcal{E}} + \frac{\nu}{e\log\mathcal{E}} - \frac{\nu\beta\log\xi_{\min}^2}{\log\mathcal{E}} \\
&\qquad\qquad + \beta\log\Gamma\left(\frac{\nu}{\log\mathcal{E}}, \frac{\nu^2}{\mathcal{E}\log\mathcal{E}}\right) + \nu + \frac{\beta\nu^2}{\mathcal{E}\log\mathcal{E}} \\
&\qquad\qquad + (1-\beta)\Big(\mathcal{C}_{\mathrm{IID}}(\xi_{\min}\big|G_0=0) - (n_{\mathrm{R}}-1)h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, G_0=0\big) \\
&\qquad\qquad\qquad - h_\lambda(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\mathbf{H}_{-\infty}^{-1}, G_0=0) + \frac{n_{\mathrm{R}}(n_{\mathrm{R}}+1)}{e} \\
&\qquad\qquad\qquad + n_{\mathrm{R}}\Delta(n_{\mathrm{R}},1) + \log 2\Big) \\
&\qquad\qquad + (1-\beta)n_{\mathrm{R}}^2\log^+\left(\frac{\pi e}{n_{\mathrm{R}}}\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big|G_0=0\big]}{1-\beta}\right)\Bigg\} \\
&\qquad + H_{\mathrm{b}}\left(\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{\log\mathcal{E}}\right) + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{\log\mathcal{E}}\tilde{\mathsf{C}}_{\mathrm{IID}}(\mathcal{E}\log\mathcal{E}) \\
&\qquad + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{\log\mathcal{E}} I\big(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\big|G_0=1\big) - \log\log\mathcal{E}\Bigg\}
\end{aligned} \tag{6.63}
$$

$$\begin{aligned}
\leq \varlimsup_{\mathcal{E}\uparrow\infty} \Big\{ & h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\}_{l=-\kappa}^{-1}, G_0=0\big) - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, G_0=0\big) \\
& - \log 2 + n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0=0\big] + \delta_1(\kappa,\xi_{\min}) + \delta_2(\kappa,\xi_{\min}) + \epsilon_\nu \\
& + \frac{\nu}{\log\mathcal{E}}\left(\log\mathcal{E} + \log\log\mathcal{E} - \log\nu\right) - \frac{\nu\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big|G_0=0\big]}{\log\mathcal{E}} \\
& + \frac{\nu}{e\log\mathcal{E}} - \frac{\nu\log\xi_{\min}^2}{\log\mathcal{E}} + \log\Gamma\left(\frac{\nu}{\log\mathcal{E}}, \frac{\nu^2}{\mathcal{E}\log\mathcal{E}}\right) + \nu + \frac{\nu^2}{\mathcal{E}\log\mathcal{E}} \\
& + H_{\mathrm{b}}\left(\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{\log\mathcal{E}}\right) + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{\log\mathcal{E}}\tilde{\mathsf{C}}_{\mathrm{IID}}(\mathcal{E}\log\mathcal{E}) \\
& + \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{\log\mathcal{E}} I\big(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\big|G_0=1\big) - \log\log\mathcal{E}\Big\}
\end{aligned} \tag{6.64}$$

Here, in (6.64), we try to find the value of $\beta$ that achieves the supremum: note that we found that first, $H_{\mathrm{b}}(\beta)$ and those terms with $1-\beta$ are constant with respect to $\mathcal{E}$; second, the remaining terms do not grow with $\mathcal{E}$ except $\log\Gamma\big(\frac{\nu}{\log\mathcal{E}}, \frac{\nu^2}{\mathcal{E}\log\mathcal{E}}\big)$ since

$$\varlimsup_{\mathcal{E}\to\infty}\left\{\log\Gamma\left(\frac{\nu}{\log\mathcal{E}}, \frac{\nu^2}{\mathcal{E}\log\mathcal{E}}\right) - \log\log\mathcal{E}\right\} = \log(1-e^{-\nu}) - \log\nu, \tag{6.65}$$

which means $\log\Gamma(\cdot)$ grows as fast as $\log\log\mathcal{E}$. So $\log\Gamma(\cdot)$ is the only term inside the sup that grows with $\mathcal{E}$. Therefore, the supremum is achieved if $\beta = 1$. Actually, this is related to the property called "escaping to infinity" (see [10, Corollary 2.8]).

Next, note that

$$\begin{aligned}
\varlimsup_{\mathcal{E}\to\infty} \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{\log\mathcal{E}}\tilde{\mathsf{C}}_{\mathrm{IID}}(\mathcal{E}\log\mathcal{E}) &= \varlimsup_{\mathcal{E}\to\infty}\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{\log\mathcal{E}}\big(\log\log(\mathcal{E}\log\mathcal{E}) + \text{const}\big) \\
&= \varlimsup_{\mathcal{E}\to\infty}\frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2\big]}{\log\mathcal{E}}\big(\log(\log\mathcal{E} + \log\log\mathcal{E}) + \text{const}\big) \\
&= 0,
\end{aligned} \tag{6.66}$$

and

$$I\big(\mathbf{H}_0;\mathbf{H}_{-\infty}^{-1}\big|G_0=1\big) = h\big(\mathbf{H}_0\big|G_0=1\big) - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}, G_0=1\big) < \infty. \tag{6.67}$$

Moreover, we drop $G_0 = 0$ because as $\mathcal{E}\to\infty$ and $t = \log\mathcal{E}$, the conditioning on $G_0 = 0$ is implicitly satisfied. As the result, (6.64) becomes

$$\begin{aligned}
\chi_{\mathrm{FB}}(\{\mathbf{H}_k\}) = {} & h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\}_{l=-\kappa}^{-1}\big) - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}\big) - \log 2 + n_{\mathrm{R}}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big] \\
& + \delta_1(\kappa,\xi_{\min}) + \delta_2(\kappa,\xi_{\min}) + \epsilon_\nu + \nu + \log(1-e^{-\nu}) - \log\nu + \nu
\end{aligned} \tag{6.68}$$

In a next step, we let $\nu$ go to zero. Note that $\epsilon_\nu \to 0$ as $\nu \to 0$ as can be seen from the definition of $\epsilon_\nu$ in Appendix B. Note further that

$$\lim_{\nu\to 0}\big\{\log(1-e^{-\nu}) - \log\nu\big\} = \lim_{\nu\to 0}\left\{\log\frac{(1-e^{-\nu})}{\nu}\right\} = \log\lim_{\nu\to 0}\left\{\frac{(1-e^{-\nu})}{\nu}\right\} = 0 \tag{6.69}$$

Therefore, we get

$$\chi_{\text{FB}}(\{\mathbf{H}_k\}) \leq h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\}_{l=-\kappa}^{-1}\big) - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}\big) \\ - \log 2 + n_{\text{R}}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big] + \delta_1(\kappa,\xi_{\min}) + \delta_2(\kappa,\xi_{\min}). \tag{6.70}$$

Next, we let $\xi_{\min}$ tend to infinity, then it is shown in Appendix C that $\delta_1(\kappa,\xi_{\min}) \to 0$ and $\delta_2(\kappa,\xi_{\min}) \to 0$. Finally, we let $\kappa$ tend to infinity and the fading number without feedback becomes

$$\chi_{\text{FB}}(\{\mathbf{H}_k\}) \leq h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0}\big|\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\}_{l=-\infty}^{-1}\big) - h\big(\mathbf{H}_0\big|\mathbf{H}_{-\infty}^{-1}\big) \\ - \log 2 + n_{\text{R}}\mathsf{E}\big[\log\|\mathbf{H}_0\|^2\big]. \tag{6.71}$$

## 6.2 Detailed Derivations for Three Terms in (6.24)

### 6.2.1 First Term

The second term on the RHS of (6.24) is bounded as follows:

$$\begin{aligned}
&I(X_k;\mathbf{Y}_k\big|E_k=1,G_k=0) \\
&\quad\leq I\big(X_k;\mathbf{Y}_k,\mathbf{H}_kX_k\big|E_k=1,G_k=0\big) && (6.72)\\
&\quad= I\big(X_k;\mathbf{H}_kX_k\big|E_k=1,G_k=0\big) \\
&\qquad+ \underbrace{I\big(X_k;\mathbf{H}_kX_k+\mathbf{Z}_k\big|\mathbf{H}_kX_k,E_k=1,G_k=0\big)}_{=0 \text{ see Appendix D}} && (6.73)\\
&\quad= I\left(X_k;\|\mathbf{H}_kX_k\|,\frac{\mathbf{H}_kX_k}{\|\mathbf{H}_kX_k\|}\middle|E_k=1,G_k=0\right) && (6.74)\\
&\quad= I\left(X_k;\|\mathbf{H}_k\||X_k|,\frac{\mathbf{H}_kX_k}{\|\mathbf{H}_k\||X_k|}\middle|E_k=1,G_k=0\right) && (6.75)\\
&\quad= I\left(X_k;\|\mathbf{H}_k\||X_k|,\hat{\mathbf{H}}_ke^{\mathrm{i}\Phi_k}\middle|E_k=1,G_k=0\right) && (6.76)\\
&\quad= I\left(X_k;\|\mathbf{H}_k\||X_k|,\hat{\mathbf{H}}_ke^{\mathrm{i}\Phi_k},e^{\mathrm{i}\Theta_k}\middle|E_k=1,G_k=0\right) && (6.77)\\
&\quad= I\left(X_k;\|\mathbf{H}_k\||X_k|e^{i\Theta_k},\hat{\mathbf{H}}_ke^{\mathrm{i}(\Phi_k+\Theta_k)},e^{\mathrm{i}\Theta_k}\middle|E_k=1,G_k=0\right) && (6.78)\\
&\quad= I\left(X_k;\|\mathbf{H}_k\||X_k|e^{i\Theta_k},\hat{\mathbf{H}}_ke^{\mathrm{i}(\Phi_k+\Theta_k)}\middle|E_k=1,G_k=0\right) && (6.79)\\
&\quad= I\left(X_k;\|\mathbf{H}_k\||X_k|e^{i\Theta_k}\middle|E_k=1,G_k=0\right) \\
&\qquad+ I\left(X_k;\hat{\mathbf{H}}_ke^{\mathrm{i}(\Phi_k+\Theta_k)}\middle|\|\mathbf{H}_k\||X_k|,e^{\mathrm{i}\Theta_k},E_k=1,G_k=0\right) && (6.80)
\end{aligned}$$

In (6.76), $\hat{\mathbf{H}}_k \triangleq \frac{\mathbf{H}_k}{\|\mathbf{H}_k\|}$, and $\Phi_k$ denotes the phase of $X_k$; in (6.77), $\{\Theta_k\}$ is IID $\sim \mathcal{U}((-\pi,\pi])$ and independent of $\{\mathbf{H}_k\}$ and $\{X_k\}$; (6.79) follows because we can get back $\Theta_k$ from $\|\mathbf{H}_k\||X_k|e^{i\Theta_k}$; (6.80) follows because of the chain rule.

We continue to bound (6.80) using a duality-based bound, for detail we refer to Appendix B:

$$\begin{aligned}
&I\left(X_k; \|\mathbf{H}_k\||X_k|e^{i\Theta_k}\middle|E_k=1, G_k=0\right)\\
&\quad + I\left(X_k; \hat{\mathbf{H}}_k e^{\mathrm{i}(\Phi_k+\Theta_k)}\middle|\|\mathbf{H}_k\||X_k|, e^{\mathrm{i}\Theta_k}, E_k=1, G_k=0\right)\\
&\leq -\log 2 - h(\mathbf{H}_k|X_k, E_k=1, G_k=0) + (2n_{\mathrm{R}}-1)\mathsf{E}[\log\|\mathbf{H}_k\||E_k=1, G_k=0]\\
&\quad - \mathsf{E}[\log\|\mathbf{H}_k\||E_k=1, G_k=0] + \mu\log\eta + \log\Gamma\left(\mu, \frac{\nu}{\eta}\right)\\
&\quad + (1-\mu)\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|E_k=1, G_k=0\right] - \mu\mathsf{E}\left[\log|X_k|^2\middle|E_k=1, G_k=0\right] + \epsilon_{\nu,k}\\
&\quad + \frac{1}{\eta}\mathsf{E}\left[\|\mathbf{H}_k\|^2|X_k|^2\middle|E_k=1, G_k=0\right] + \frac{\nu}{\eta} + h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|E_k=1, G_k=0\right)
\end{aligned} \tag{6.81}$$

Arithmetically rearranging the terms in (6.81), we have the second term on the RHS of (6.24) be bounded as follows:

$$\begin{aligned}
&I(X_k; \mathbf{Y}_k|E_k=1, G_k=0)\\
&\leq h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|E_k=1, G_k=0\right) - h(\mathbf{H}_k|X_k, E_k=1, G_k=0) - \log 2\\
&\quad + n_{\mathrm{R}}\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|E_k=1, G_k=0\right]\\
&\quad + \mu\left(\log\eta - \mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|E_k=1, G_k=0\right] - \mathsf{E}\left[\log|X_k|^2\middle|E_k=1, G_k=0\right]\right)\\
&\quad + \log\Gamma\left(\mu, \frac{\nu}{\eta}\right) + \epsilon_{\nu,k} + \frac{1}{\eta}\mathsf{E}\left[\|\mathbf{H}_k^2\||X_k|^2\middle|E_k=1, G_k=0\right] + \frac{\nu}{\eta}.
\end{aligned} \tag{6.82}$$

### 6.2.2 Second Term

The third term on the RHS of (6.24) is bounded as follows:

$$\begin{aligned}
&I\left(\mathbf{H}_1^{k-1}; \mathbf{Y}_k\middle|X_k, E_k=1, G_k=0\right)\\
&\leq I\left(\mathbf{H}_1^{k-1}; \mathbf{Y}_k, \mathbf{H}_k\middle|X_k, E_k=1, G_k=0\right) &(6.83)\\
&= I\left(\mathbf{H}_1^{k-1}; \mathbf{H}_k\middle|X_k, E_k=1, G_k=0\right)\\
&\quad + \underbrace{I\left(\mathbf{H}_1^{k-1}; \mathbf{Y}_k\middle|\mathbf{H}_k, X_k, E_k=1, G_k=0\right)}_{=0 \text{ see Appendix D}} &(6.84)\\
&= h(\mathbf{H}_k|X_k, E_k=1, G_k=0) - h\left(\mathbf{H}_k\middle|\mathbf{H}_1^{k-1}, X_k, E_k=1, G_k=0\right) &(6.85)\\
&= h(\mathbf{H}_k|X_k, E_k=1, G_k=0) - h\left(\mathbf{H}_k\middle|\mathbf{H}_1^{k-1}, G_k=0\right), &(6.86)
\end{aligned}$$

where the last step follows because conditional on $G_k = 0$ and all the past values $\mathbf{H}_1^{k-1}$ of $\{\mathbf{H}_k\}$, $\mathbf{H}_k$ is independent of $X_k$ and $E_k$.

### 6.2.3 Third Term

Recalling the definition of $E_k$ in (6.25), we lower-bound the fourth term on the RHS of (6.24) as follows:

$$\begin{aligned}
&I\left(\mathbf{Y}_1^{k-1}; \mathbf{Y}_k\middle|G_k=0\right)\\
&\geq I\left(\mathbf{Y}_{k-\kappa}^{k-1}; \mathbf{Y}_k\middle|G_k=0\right)
\end{aligned} \tag{6.87}$$

$$
\begin{aligned}
&= I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1},E_k\big|G_k=0\big)-I\big(\mathbf{Y}_k;E_k\big|\mathbf{Y}_{k-\kappa}^{k-1},G_k=0\big) && (6.88)\\
&= \underbrace{I\left(\mathbf{Y}_k;E_k|G_k=0\right)}_{\geq 0}+I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k,G_k=0\big)\\
&\quad-\underbrace{H(E_k|\mathbf{Y}_{k-\kappa}^{k-1},G_k=0)}_{\leq H_\mathrm{b}(E_k)}+\underbrace{H(E_k|\mathbf{Y}_{k-\kappa}^{k},G_k=0)}_{\geq 0} && (6.89)\\
&\geq \beta_k I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k=1,G_k=0\big)\\
&\quad+(1-\beta_k)\underbrace{I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k=0,G_k=0\big)}_{\geq 0}-H_\mathrm{b}(\beta_k) && (6.90)\\
&\geq \beta_k I\big(\mathbf{Y}_k;\mathbf{Y}_{k-\kappa}^{k-1}\big|E_k=1,G_k=0\big)-H_\mathrm{b}(\beta_k) && (6.91)\\
&= \beta_k I\left(\mathbf{Y}_k,e^{\mathrm{i}\Theta_k};\mathbf{Y}_{k-\kappa}^{k-1},\{e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1}\middle|E_k=1,G_k=0\right)-H_\mathrm{b}(\beta_k) && (6.92)\\
&= \beta_k I\left(\mathbf{H}_kX_ke^{\mathrm{i}\Theta_k}+\mathbf{Z}_k,e^{\mathrm{i}\Theta_k};\{\mathbf{H}_lX_le^{\mathrm{i}\Theta_l}+\mathbf{Z}_l\}_{l=k-\kappa}^{k-1},\{e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1}\middle|E_k=1,G_k=0\right)\\
&\quad-H_\mathrm{b}(\beta_k) && (6.93)\\
&\geq \beta_k I\left(\mathbf{H}_ke^{\mathrm{i}\Theta_k}X_k+\mathbf{Z}_k;\{\mathbf{H}_lX_le^{\mathrm{i}\Theta_l}+\mathbf{Z}_l\}_{l=k-\kappa}^{k-1}\middle|E_k=1,G_k=0\right)-H_\mathrm{b}(\beta_k) && (6.94)\\
&= \beta_k I\left(\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k}+\mathbf{Z}_k;\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}+\mathbf{Z}_l\}_{l=k-\kappa}^{k-1}\middle|E_k=1,G_k=0\right)-H_\mathrm{b}(\beta_k) && (6.95)\\
&= \beta_k I\left(\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k}+\mathbf{Z}_k;\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}+\mathbf{Z}_l\}_{l=k-\kappa}^{k-1},\mathbf{Z}_{k-\kappa}^{k-1}\middle|E_k=1,G_k=0\right)-H_\mathrm{b}(\beta_k)\\
&\quad-\beta_k I\left(\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k}+\mathbf{Z}_k;\mathbf{Z}_{k-\kappa}^{k-1}\,\middle|\,\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}+\mathbf{Z}_l\}_{l=k-\kappa}^{k-1},E_k=1,G_k=0\right). && (6.96)
\end{aligned}
$$

Here, (6.90) follows because the first term and last term in (6.89) are equal or greater than zero and $H_\mathrm{b}(E_k|\mathbf{Y}_{k-\kappa}^{k-1},G_k=0)\leq H_\mathrm{b}(E_k)=H_\mathrm{b}(\beta_k)$; in (6.92), we add $\{\Theta_k\}$, which is IID $\sim\mathcal{U}((-\pi,\pi])$ and independent of $\mathbf{Y}_k$. Because $\{\Theta_k\}$ is uniformly distributed, it destroys the phase of $\{\mathbf{H}_k\}$ and let $\{\mathbf{H}_ke^{\mathrm{i}\Theta_k}\}$ becomes circularly symmetric. (6.94) follows because we drop $e^{\Theta_k}$ on both side of mutual information.

By Appendix C.1, we have

$$
\begin{aligned}
&\beta_k I\left(\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k}+\mathbf{Z}_k;\mathbf{Z}_{k-\kappa}^{k-1}\,\middle|\,\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}+\mathbf{Z}_l\}_{l=k-\kappa}^{k-1},E_k=1,G_k=0\right)\\
&\quad\leq \delta_1(\kappa,\xi_{\min})+H_\mathrm{b}(\beta_k), && (6.97)
\end{aligned}
$$

so we further bound (6.96) as follows:

$$
\begin{aligned}
&I\left(\mathbf{Y}_1^{k-1};\mathbf{Y}_k\middle|G_k=0\right)\\
&\quad\geq \beta_k I\left(\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k}+\mathbf{Z}_k;\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}+\mathbf{Z}_l\}_{l=k-\kappa}^{k-1},\mathbf{Z}_{k-\kappa}^{k-1}\middle|E_k=1,G_k=0\right)\\
&\qquad-2H_\mathrm{b}(\beta_k)-\delta_1(\kappa,\xi_{\min}) && (6.98)\\
&\quad= \beta_k I\left(\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k}+\mathbf{Z}_k;\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1},\mathbf{Z}_{k-\kappa}^{k-1}\middle|E_k=1,G_k=0\right)\\
&\qquad-2H_\mathrm{b}(\beta_k)-\delta_1(\kappa,\xi_{\min}) && (6.99)\\
&\quad\geq \beta_k I\left(\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k}+\mathbf{Z}_k;\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1}\middle|E_k=1,G_k=0\right)\\
&\qquad-2H_\mathrm{b}(\beta_k)-\delta_1(\kappa,\xi_{\min}) && (6.100)\\
&\quad= \beta_k I\left(\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k}+\mathbf{Z}_k,\mathbf{Z}_k;\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1}\middle|E_k=1,G_k=0\right)
\end{aligned}
$$

$$
\begin{aligned}
&- \beta_k I\left(\mathbf{Z}_k; \{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1}\middle|\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k} + \mathbf{Z}_k, E_k = 1, G_k = 0\right)\\
&- 2H_b(\beta_k) - \delta_1(\kappa, \xi_{\min}). \qquad (6.101)
\end{aligned}
$$

Here, in (6.100), we drop $\{\mathbf{Z}_{k-\kappa}^{k-1}\}$, so the mutual information becomes smaller.

By Appendix C.2, we have

$$
\begin{aligned}
&\beta_k I\left(\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}\}_{k-\kappa}^{k-1}; \mathbf{Z}_k \,\middle|\, \mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k} + \mathbf{Z}_k, E_k = 1, G_k = 0\right)\\
&\quad \leq \delta_2(\kappa, \xi_{\min}) + H_{\mathrm{b}}(\beta_k), \qquad (6.102)
\end{aligned}
$$

and we further bound (6.101) as follows:

$$
\begin{aligned}
&I\left(\mathbf{Y}_1^{k-1}; \mathbf{Y}_k\middle|G_k = 0\right)\\
&\geq \beta_k I\left(\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k} + \mathbf{Z}_k, \mathbf{Z}_k; \{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1}\middle|E_k = 1, G_k = 0\right)\\
&\quad - 3H_{\mathrm{b}}(\beta_k) - \delta_1(\kappa, \xi_{\min}) - \delta_2(\kappa, \xi_{\min}) \qquad (6.103)\\
&\geq \beta_k I\left(\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k}; \{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1}\middle|E_k = 1, G_k = 0\right)\\
&\quad - 3H_{\mathrm{b}}(\beta_k) - \delta_1(\kappa, \xi_{\min}) - \delta_2(\kappa, \xi_{\min}) \qquad (6.104)\\
&= \beta_k I\left(\|\mathbf{H}_k\||X_k|, \hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}; \{\|\mathbf{H}_l\||X_l|\}_{l=k-\kappa}^{k-1}, \left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}\middle|E_k = 1, G_k = 0\right)\\
&\quad - 3H_{\mathrm{b}}(\beta_k) - \delta_1(\kappa, \xi_{\min}) - \delta_2(\kappa, \xi_{\min}) \qquad (6.105)\\
&\geq \beta_k I\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}; \left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}\middle|E_k = 1, G_k = 0\right)\\
&\quad - 3H_{\mathrm{b}}(\beta_k) - \delta_1(\kappa, \xi_{\min}) - \delta_2(\kappa, \xi_{\min}) \qquad (6.106)\\
&= \beta_k h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|E_k = 1, G_k = 0\right) - \beta_k h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, E_k = 1, G_k = 0\right)\\
&\quad - 3H_{\mathrm{b}}(\beta_k) - \delta_1(\kappa, \xi_{\min}) - \delta_2(\kappa, \xi_{\min}) \qquad (6.107)\\
&= \beta_k h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|E_k = 1, G_k = 0\right) - \beta_k h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, E_k = 1, G_k = 0\right)\\
&\quad - (1-\beta_k) h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, E_k = 0, G_k = 0\right)\\
&\quad + (1-\beta_k) h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, E_k = 0, G_k = 0\right)\\
&\quad - 3H_{\mathrm{b}}(\beta_k) - \delta_1(\kappa, \xi_{\min}) - \delta_2(\kappa, \xi_{\min}) \qquad (6.108)\\
&= \beta_k h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|E_k = 1, G_k = 0\right) - h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, E_k, G_k = 0\right)\\
&\quad + (1-\beta_k) h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, E_k = 0, G_k = 0\right)\\
&\quad - 3H_{\mathrm{b}}(\beta_k) - \delta_1(\kappa, \xi_{\min}) - \delta_2(\kappa, \xi_{\min}) \qquad (6.109)\\
&\geq \beta_k h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|E_k = 1, G_k = 0\right) - h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, G_k = 0\right)\\
&\quad + (1-\beta_k) h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, E_k = 0, G_k = 0\right)\\
&\quad - 3H_{\mathrm{b}}(\beta_k) - \delta_1(\kappa, \xi_{\min}) - \delta_2(\kappa, \xi_{\min}) \qquad (6.110)
\end{aligned}
$$

Here, (6.105) follows from taking the magnitude from $\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k}$; (6.106) follows because we drop some terms in mutual information; (6.107) follows from the definition of differential entropy for unit vectors (see Section 3.1.2); (6.110) follows because dropping conditioning increases entropy.

# Chapter 7

# Discussion and Conclusion

In this thesis, we have shown that the asymptotic capacity of general regular SIMO fading channels with memory remains unchanged even if one allows causal noiseless feedback. This once again shows the extremely unattractive behavior of regular fading channels at high SNR: besides the double-logarithmic growth [8] and the very poor performance in a multiple-user setup (where the maximum sum-rate only can be achieved if all users apart from one *always* remain switched off [16]), we now see that any type of feedback does not increase capacity in spite of memory in the channel.

Possible future works for the general regular fading channels with memory and feedback might include the following:

- Considering the case with multiple-input single-output, i.e., having several mobile phones (each having one antenna) communicating with one base station (having only one antenna). The difficulties for this case lies in the fact that now we not only need to optimize the phase and magnitude of the inputs, but also the direction of them.
- Considering the case with multiple-input multiple-output.
- The situation where both transmitter and receiver have access to causal partial side-information $\mathbf{S}_k$ about the fading, where by *partial* we mean that

$$\lim_{n\to\infty} \frac{1}{n} I\big(\mathbf{S}_1^n; \mathbf{H}_1^n\big) < \infty. \tag{7.1}$$

# Appendix A

# Upper Bound (6.46)

In this appendix, we derive the following upper bound:

$$
\begin{aligned}
&I\big(X_k, \mathbf{H}_1^{k-1}; \mathbf{Y}_k \big| E_k = 0, G_k = 0\big) - h_\lambda \left( \hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k} \middle| \left\{ \hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l} \right\}_{l=k-\kappa}^{k-1}, E_k = 0, G_k = 0 \right) \\
&-n_\mathrm{R} \mathsf{E}\big[\log \|\mathbf{H}_k\|^2 \big| E_k = 0, G_k = 0\big] \\
&\quad \le \mathcal{C}_\mathrm{IID}(\xi_\mathrm{min} \big| G_k = 0) - (n_\mathrm{R} - 1) h(\mathbf{H}_0 \big| \mathbf{H}_{-\infty}^{-1}, G_0 = 0) \\
&\qquad - h_\lambda\big(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0} \big| \mathbf{H}_{-\infty}^{-1}, G_0 = 0\big) - h(\mathbf{H}_0 \big| \mathbf{H}_{-k+1}^{-1}, G_0 = 0) \\
&\qquad + \frac{n_\mathrm{R}(n_\mathrm{R}+1)}{e} + n_\mathrm{R}^2 \log^+ \left( \frac{\pi e}{n_\mathrm{R}} \frac{\mathsf{E}\big[\|\mathbf{H}_0\|^2 | G_0 = 0\big]}{1 - \beta_k} \right) + n_\mathrm{R} \Delta(n_\mathrm{R}, 1). \qquad \text{(A.1)}
\end{aligned}
$$

We bound the first term as follows:

$$
\begin{aligned}
&I\big(X_k, \mathbf{H}_1^{k-1}; \mathbf{Y}_k \big| E_k = 0, G_k = 0\big) \\
&\quad = I\big(X_k; \mathbf{Y}_k \big| E_k = 0, G_k = 0\big) + I\big(\mathbf{H}_1^{k-1}; \mathbf{Y}_k \big| X_k, E_k = 0, G_k = 0\big) && \text{(A.2)} \\
&\quad \le I\big(X_k; \mathbf{Y}_k \big| E_k = 0, G_k = 0\big) + I\big(\mathbf{H}_1^{k-1}; \mathbf{Y}_k, \mathbf{H}_k \big| X_k, E_k = 0, G_k = 0\big) && \text{(A.3)} \\
&\quad = I\big(X_k; \mathbf{Y}_k \big| E_k = 0, G_k = 0\big) + I\big(\mathbf{H}_1^{k-1}; \mathbf{H}_k \big| X_k, E_k = 0, G_k = 0\big) \\
&\qquad + \underbrace{I\big(\mathbf{H}_1^{k-1}; \mathbf{Y}_k \big| \mathbf{H}_k, X_k, E_k = 0, G_k = 0\big)}_{= 0 \text{ see Appendix D}} && \text{(A.4)} \\
&\quad = I\big(X_k; \mathbf{Y}_k \big| E_k = 0, G_k = 0\big) + I\big(\mathbf{H}_1^{k-1}; \mathbf{H}_k \big| X_k, E_k = 0, G_k = 0\big) && \text{(A.5)} \\
&\quad = I\big(X_k; \mathbf{Y}_k \big| E_k = 0, G_k = 0\big) + h\big(\mathbf{H}_k \big| X_k, E_k = 0, G_k = 0\big) \\
&\qquad - h\big(\mathbf{H}_k \big| \mathbf{H}_1^{k-1}, X_k, E_k = 0, G_k = 0\big) && \text{(A.6)} \\
&\quad \le \mathcal{C}_\mathrm{IID}(\xi_\mathrm{min} \big| G_k = 0) + h\big(\mathbf{H}_k \big| X_k, E_k = 0, G_k = 0\big) - h\big(\mathbf{H}_k \big| \mathbf{H}_1^{k-1}, G_k = 0\big), && \text{(A.7)}
\end{aligned}
$$

where in (A.7), $\mathcal{C}_\mathrm{IID}(\cdot)$ denotes the capacity without feedback or memory for a given power. Because $\mathcal{C}_\mathrm{IID}(\cdot)$ is nondecreasing, and under the condition that $E_k = 0$, i.e., $|X_k| \le \xi_\mathrm{min}$, $\mathcal{C}_\mathrm{IID}(\xi_\mathrm{min} \big| G_k = 0)$ is the upper bound. Therefore, we get

$$
I\big(X_k, \mathbf{H}_1^{k-1}; \mathbf{Y}_k \big| E_k = 0, G_k = 0\big) - h_\lambda \left( \hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k} \middle| \left\{ \hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l} \right\}_{l=k-\kappa}^{k-1}, E_k = 0, G_k = 0 \right)
$$

$$
\begin{aligned}
&-n_{\mathrm{R}}\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\big|E_k=0,G_k=0\right]\\
&\quad\le \mathcal{C}_{\mathrm{IID}}(\xi_{\min}\big|G_k=0)+h\big(\mathbf{H}_k\big|X_k,E_k=0,G_k=0\big)\\
&\qquad-h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1},E_k=0,G_k=0\right)\\
&\qquad-n_{\mathrm{R}}\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\big|E_k=0,G_k=0\right]-h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},G_k=0\big) &\text{(A.8)}\\
&\quad\le \mathcal{C}_{\mathrm{IID}}(\xi_{\min}\big|G_k=0)+h\big(\mathbf{H}_k\big|E_k=0,G_k=0\big)\\
&\qquad-h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1},\mathbf{H}_1^{k-1},E_k=0,G_k=0\right)\\
&\qquad-n_{\mathrm{R}}\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\big|E_k=0,G_k=0\right]-h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},G_k=0\big) &\text{(A.9)}\\
&\quad= \mathcal{C}_{\mathrm{IID}}(\xi_{\min}\big|G_k=0)+h\big(\mathbf{H}_k\big|E_k=0,G_k=0\big)-h_\lambda\big(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\big|\mathbf{H}_1^{k-1},G_k=0\big)\\
&\qquad-n_{\mathrm{R}}\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\big|E_k=0,G_k=0\right]-h\big(\mathbf{H}_k\big|\mathbf{H}_1^{k-1},G_k=0\big). &\text{(A.10)}
\end{aligned}
$$

Here, (A.9) follows from conditioning that reduces entropy; and (A.10) follows because conditional on $\mathbf{H}_1^{k-1}$, $\mathbf{H}_k$ is independent of $\left\{\hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}$ and $E_k$.

Next we will bound the term $\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\big|E_k=0,G_k=0\right]$. We first have the following inequality:

$$
\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\big|E_k=0,G_k=0\right]\ge-\frac{1}{\xi}h^-\big(\mathbf{H}_k\big|E_k=0,G_k=0\big)-\Delta(n_{\mathrm{R}},\xi) \tag{A.11}
$$

by Lemma 3.2 where $h^-(\cdot)$, $\xi$, and $\Delta(n_{\mathrm{R}},\xi)$ are defined in Section 3.1. Because

$$
h\big(\mathbf{H}_k\big|E_k=0,G_k=0\big)=h^+\big(\mathbf{H}_k\big|E_k=0,G_k=0\big)-h^-\big(\mathbf{H}_k\big|E_k=0,G_k=0\big), \tag{A.12}
$$

where both $h^+(\cdot)$ and $h^-(\cdot)$ are nonnegative (see Section 3.1.1), we further bound the first term in (A.11) as follows:

$$
\begin{aligned}
&-\frac{1}{\xi}h^-\big(\mathbf{H}_k\big|E_k=0,G_k=0\big)\\
&\quad=\frac{1}{\xi}h\big(\mathbf{H}_k\big|E_k=0,G_k=0\big)-\frac{1}{\xi}h^+\big(\mathbf{H}_k\big|E_k=0,G_k=0\big) &\text{(A.13)}\\
&\quad\ge\frac{1}{\xi}h\big(\mathbf{H}_k\big|E_k=0,G_k=0\big)-\frac{1}{\xi}\frac{n_{\mathrm{R}}+1}{e}\\
&\qquad-\frac{n_{\mathrm{R}}}{\xi}\log^+\left(\frac{\pi e}{n_{\mathrm{R}}}\mathsf{E}\left[\|\mathbf{H}_k\|^2\big|E_k=0,G_k=0\right]\right) &\text{(A.14)}\\
&\quad\ge\frac{1}{\xi}h\big(\mathbf{H}_k\big|E_k=0,G_k=0\big)-\frac{1}{\xi}\frac{n_{\mathrm{R}}+1}{e}-\frac{n_{\mathrm{R}}}{\xi}\log^+\left(\frac{\pi e}{n_{\mathrm{R}}}\frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2\big|G_0=0\right]}{1-\beta_k}\right). &\text{(A.15)}
\end{aligned}
$$

Here, (A.14) follows from Lemma A.12 in [10, Appendix A.4.2]; and (A.15) follows because

$$
\begin{aligned}
&\mathsf{E}\left[\|\mathbf{H}_k\|^2\big|E_k=0,G_k=0\right]\\
&\quad=\frac{1}{1-\beta_k}\left(\mathsf{E}\left[\|\mathbf{H}_k\|^2\big|G_k=0\right]-\beta_k\mathsf{E}\left[\|\mathbf{H}_k\|^2\big|E_k=1,G_k=0\right]\right) &\text{(A.16)}\\
&\quad\le\frac{1}{1-\beta_k}\mathsf{E}\left[\|\mathbf{H}_k\|^2\big|G_k=0\right]. &\text{(A.17)}
\end{aligned}
$$

Choosing $\xi = 1$, we then get from (A.15) and (A.11)

$$\begin{aligned}
&\mathsf{E}\left[\log \|\mathbf{H}_k\|^2 \middle| E_k = 0, G_k = 0\right] \\
&\geq h\left(\mathbf{H}_k \middle| E_k = 0, G_k = 0\right) - \frac{n_\mathrm{R}+1}{e} - n_\mathrm{R} \log^+\left(\frac{\pi e}{n_\mathrm{R}} \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2 \middle| G_0 = 0\right]}{1-\beta_k}\right) - \Delta(n_\mathrm{R}, 1) \quad \text{(A.18)}
\end{aligned}$$

We put this back into (A.10), and get

$$\begin{aligned}
&I\left(X_k, \mathbf{H}_1^{k-1}; \mathbf{Y}_k \middle| E_k = 0, G_k = 0\right) - h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k} \middle| \hat{\mathbf{H}}_l e^{\mathrm{i}\Theta_l}, E_k = 0, G_k = 0\right) \\
&-n_\mathrm{R}\mathsf{E}\left[\log \|\mathbf{H}_k\|^2 \middle| E_k = 0, G_k = 0\right] \\
&\quad \leq \mathcal{C}_\mathrm{IID}(\xi_\mathrm{min} \middle| G_k = 0) + h\left(\mathbf{H}_k \middle| E_k = 0, G_k = 0\right) - h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k} \middle| \mathbf{H}_1^{k-1}, G_k = 0\right) \\
&\qquad - h(\mathbf{H}_k \middle| \mathbf{H}_1^{k-1}, G_k = 0) - n_\mathrm{R} h(\mathbf{H}_k \middle| E_k = 0, G_k = 0) + \frac{n_\mathrm{R}(n_\mathrm{R}+1)}{e} \\
&\qquad + n_\mathrm{R}^2 \log^+\left(\frac{\pi e}{n_\mathrm{R}} \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2 | G_0 = 0\right]}{1-\beta_k}\right) + n_\mathrm{R}\Delta(n_\mathrm{R}, 1) \qquad \text{(A.19)} \\
&\quad \leq \mathcal{C}_\mathrm{IID}(\xi_\mathrm{min} \middle| G_0 = 0) - (n_\mathrm{R}-1) h(\mathbf{H}_k \middle| E_k = 0, G_k = 0) - h_\lambda\left(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0} \middle| \mathbf{H}_{-\infty}^{-1}, G_0 = 0\right) \\
&\qquad - h(\mathbf{H}_k \middle| \mathbf{H}_1^{k-1}, G_k = 0) + \frac{n_\mathrm{R}(n_\mathrm{R}+1)}{e} + n_\mathrm{R}^2 \log^+\left(\frac{\pi e}{n_\mathrm{R}} \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2 | G_0 = 0\right]}{1-\beta_k}\right) \\
&\qquad + n_\mathrm{R}\Delta(n_\mathrm{R}, 1) \qquad \text{(A.20)} \\
&\quad \leq \mathcal{C}_\mathrm{IID}(\xi_\mathrm{min} \middle| G_0 = 0) - (n_\mathrm{R}-1) h(\mathbf{H}_0 | \mathbf{H}_{-\infty}^{-1}, G_0 = 0) - h_\lambda\left(\hat{\mathbf{H}}_0 e^{\mathrm{i}\Theta_0} \middle| \mathbf{H}_{-\infty}^{-1}, G_0 = 0\right) \\
&\qquad - h(\mathbf{H}_0 \middle| \mathbf{H}_{1-k}^{-1}, G_0 = 0) + \frac{n_\mathrm{R}(n_\mathrm{R}+1)}{e} + n_\mathrm{R}^2 \log^+\left(\frac{\pi e}{n_\mathrm{R}} \frac{\mathsf{E}\left[\|\mathbf{H}_0\|^2 | G_0 = 0\right]}{1-\beta_k}\right) \\
&\qquad + n_\mathrm{R}\Delta(n_\mathrm{R}, 1) \qquad \text{(A.21)}
\end{aligned}$$

where (A.20) follows because we shift the time index in $h_\lambda(\cdot)$ by $k$ using the stationarity of $\{\mathbf{H}_k\}$, and add more terms to it; we also shift $G_k$ to $G_0$ in $\mathcal{C}_\mathrm{IID}(\cdot)$ since $\mathcal{C}_\mathrm{IID}(\cdot)$ is IID. (A.21) follows because $\{\mathbf{H}_k\}$ is a stationary process, $h(\mathbf{H}_k | \mathbf{H}_1^{k-1})$ is nonincreasing in $k$, and therefore, we have

$$\begin{aligned}
h(\mathbf{H}_k | E_k = 0, G_k = 0) &\geq h(\mathbf{H}_k | \mathbf{H}_1^{k-1}, E_k = 0, G_k = 0) && \text{(A.22)} \\
&= h(\mathbf{H}_k | \mathbf{H}_1^{k-1}, G_k = 0) && \text{(A.23)} \\
&\geq h(\mathbf{H}_k | \mathbf{H}_{-\infty}^{k-1}, G_k = 0) && \text{(A.24)} \\
&= h(\mathbf{H}_0 | \mathbf{H}_{-\infty}^{-1}, G_0 = 0). && \text{(A.25)}
\end{aligned}$$

# Appendix B

# Upper Bound (6.81)

In this appendix, we derive the following upper bound:

$$\begin{aligned}
&I\left(X_k; \|\mathbf{H}_k\||X_k|e^{i\Theta_k}\middle|E_k=1, G_k=0\right) \\
&\quad + I\left(X_k; \hat{\mathbf{H}}_k e^{\mathrm{i}(\Phi_k+\Theta_k)}\middle|\|\mathbf{H}_k\||X_k|, e^{\mathrm{i}\Theta_k}, E_k=1, G_k=0\right) \qquad \text{(B.1)}\\
&\le -\log 2 - h(\mathbf{H}_k\big|X_k, E_k=1, G_k=0) + (2n_\mathrm{R}-1)\mathsf{E}[\log\|\mathbf{H}_k\||E_k=1, G_k=0] \\
&\quad - \mathsf{E}[\log\|\mathbf{H}_k\||E_k=1, G_k=0] + \mu\log\eta + \log\Gamma\left(\mu, \frac{\nu}{\eta}\right) \\
&\quad + (1-\mu)\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|E_k=1, G_k=0\right] - \mu\mathsf{E}\left[\log|X_k|^2\middle|E_k=1, G_k=0\right] + \epsilon_{\nu,k} \\
&\quad + \frac{1}{\eta}\mathsf{E}\left[\|\mathbf{H}_k\|^2|X_k|^2\middle|E_k=1, G_k=0\right] + \frac{\nu}{\eta} + h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k}\middle|E_k=1, G_k=0\right), \qquad \text{(B.2)}
\end{aligned}$$

using a similar approach as in [9, Appendix D].

First, we apply Lemma 11 in [9] to the first term in (B.1), i.e., we choose $\mathbf{S} = X_k$ and $T = \|\mathbf{H}_k\||X_k|e^{i\Theta_k}$. Note that we need to condition everything on the events $E_k = 1$ and $G_k = 0$.

$$\begin{aligned}
&I\big(X_k; \|\mathbf{H}_k\||X_k|e^{i\Theta_k}\big|E_k=1, G_k=0\big) \\
&\quad \le -h\big(\|\mathbf{H}_k\||X_k|e^{\mathrm{i}\Theta_k}\big|X_k, E_k=1, G_k=0\big) + \log\pi + \mu\log\eta + \log\Gamma\left(\mu, \frac{\nu}{\eta}\right) \\
&\qquad + (1-\mu)\mathsf{E}\left[\log(\|\mathbf{H}_k\|^2|X_k|^2+\nu)\middle|E_k=1, G_k=0\right] \\
&\qquad + \frac{1}{\eta}\mathsf{E}\left[\|\mathbf{H}_k\|^2|X_k|^2\middle|E_k=1, G_k=0\right] + \frac{\nu}{\eta} \qquad \text{(B.3)}
\end{aligned}$$

where $\mu, \eta > 0$, and $\nu \ge 0$ can be chosen freely. Note that from a conditional version of Lemma 2 in [9] with $m = 1$ follows that

$$\begin{aligned}
&h\big(\|\mathbf{H}_k\||X_k|e^{\mathrm{i}\Theta_k}\big|X_k=x_k, E_k=1, G_k=0\big) \\
&\quad = h\big(\Theta_k\big|X_k=x_k, E_k=1, G_k=0\big) + h\big(\|\mathbf{H}_k\||X_k|\big|e^{\mathrm{i}\Theta_k}, X_k=x_k, E_k=1, G_k=0\big) \\
&\qquad + \mathsf{E}\left[\log\|\mathbf{H}_k\||X_k|\middle|X_k=x_k, E_k=1, G_k=0\right] \qquad \text{(B.4)}\\
&\quad = \log 2\pi + h(\|\mathbf{H}_k\||X_k|\big|X_k=x_k, E_k=1, G_k=0) \\
&\qquad + \mathsf{E}\left[\log\|\mathbf{H}_k\||X_k|\middle|X_k=x_k, E_k=1, G_k=0\right], \qquad \text{(B.5)}
\end{aligned}$$

where we have used that $\Theta_k$ is independent of all other random quantities and uniformly distributed on the unit circle. Taking the expectation over $X_k$ conditional on $E_k = 1, G_k = 0$ and noting that by the law of total expectation

$$\begin{aligned}
&\mathsf{E}_{X_k}\left[\mathsf{E}\left[\log\|\mathbf{H}_k\||X_k|\middle|X_k = x_k, E_k = 1, G_k = 0\right]\middle|E_k = 1, G_k = 0\right] \\
&\quad = \mathsf{E}\left[\log\|\mathbf{H}_k\||X_k|\middle|E_k = 1, G_k = 0\right], \qquad \text{(B.6)}
\end{aligned}$$

we then get

$$\begin{aligned}
&h\big(\|\mathbf{H}_k\||X_k|e^{\mathrm{i}\Theta_k}\big|X_k, E_k = 1, G_k = 0\big) \\
&\quad = \log 2\pi + h(\|\mathbf{H}_k\||X_k|\big|X_k, E_k = 1, G_k = 0) + \mathsf{E}\left[\log\|\mathbf{H}_k\||X_k|\middle|E_k = 1, G_k = 0\right] \qquad \text{(B.7)} \\
&\quad = \log 2\pi + h(\|\mathbf{H}_k\|\big|X_k, E_k = 1, G_k = 0) \\
&\qquad + \mathsf{E}\left[\log|X_k|\middle|E_k = 1, G_k = 0\right] + \mathsf{E}\left[\log\|\mathbf{H}_k\|\middle|E_k = 1, G_k = 0\right] \\
&\qquad + \mathsf{E}\left[\log|X_k|\middle|E_k = 1, G_k = 0\right] \qquad \text{(B.8)} \\
&\quad = \log 2\pi + h(\|\mathbf{H}_k\|\big|X_k, E_k = 1, G_k = 0) \\
&\qquad + 2\mathsf{E}\left[\log|X_k|\middle|E_k = 1, G_k = 0\right] + \mathsf{E}\left[\log\|\mathbf{H}_k\|\middle|E_k = 1, G_k = 0\right] \qquad \text{(B.9)}
\end{aligned}$$

where (B.8) follows from the scaling property of entropy with a real argument.

We choose $0 < \mu < 1$ (recall that $\mu$ is a free parameter!) such that $1 - \mu > 0$. Then we define

$$\begin{aligned}
\epsilon_{\nu,k} &\triangleq \sup_{\gamma \geq \xi_{\min}} \Big\{\mathsf{E}\left[\log(\|\mathbf{H}_k\|^2\gamma^2 + \nu)\middle|E_k = 1, G_k = 0\right] \\
&\qquad\qquad - \mathsf{E}\left[\log(\|\mathbf{H}_k\|^2\gamma^2)\middle|E_k = 1, G_k = 0\right]\Big\}, \qquad \text{(B.10)} \\
\epsilon_{\nu} &\triangleq \sup_{\gamma \geq \xi_{\min}} \left\{\mathsf{E}\left[\log(\|\mathbf{H}_0\|^2\gamma^2 + \nu)\middle|G_0 = 0\right] - \mathsf{E}\left[\log(\|\mathbf{H}_0\|^2\gamma^2)\middle|G_0 = 0\right]\right\}, \qquad \text{(B.11)}
\end{aligned}$$

such that

$$\begin{aligned}
&(1-\mu)\mathsf{E}\left[\log(\|\mathbf{H}_k\|^2|X_k|^2 + \nu)\middle|E_k = 1, G_k = 0\right] \\
&\quad = (1-\mu)\mathsf{E}\left[\log(\|\mathbf{H}_k\|^2|X_k|^2)\middle|E_k = 1, G_k = 0\right] \\
&\qquad + (1-\mu)\mathsf{E}\left[\log(\|\mathbf{H}_k\|^2|X_k|^2 + \nu)\middle|E_k = 1, G_k = 0\right] \\
&\qquad - (1-\mu)\mathsf{E}\left[\log(\|\mathbf{H}_k\|^2|X_k|^2)\middle|E_k = 1, G_k = 0\right] \qquad \text{(B.12)} \\
&\quad = (1-\mu)\mathsf{E}\left[\log(\|\mathbf{H}_k\|^2)\middle|E_k = 1, G_k = 0\right] + (1-\mu)\mathsf{E}\left[\log(|X_k|^2)\middle|E_k = 1, G_k = 0\right] \\
&\qquad + (1-\mu)\Big(\mathsf{E}\left[\log(\|\mathbf{H}_k\|^2|X_k|^2 + \nu)\middle|E_k = 1, G_k = 0\right] \\
&\qquad\qquad - \mathsf{E}\left[\log(\|\mathbf{H}_k\|^2|X_k|^2)\middle|E_k = 1, G_k = 0\right]\Big) \qquad \text{(B.13)} \\
&\quad \leq (1-\mu)\mathsf{E}\left[\log(\|\mathbf{H}_k\|^2)\middle|E_k = 1, G_k = 0\right] + (1-\mu)\mathsf{E}\left[\log(|X_k|^2)\middle|E_k = 1, G_k = 0\right] \\
&\qquad + (1-\mu)\sup_{|x_k| \geq \xi_{\min}} \Big\{\mathsf{E}\left[\log(\|\mathbf{H}_k\|^2|x_k|^2 + \nu)\middle|E_k = 1, G_k = 0\right] \\
&\qquad\qquad - \mathsf{E}\left[\log(\|\mathbf{H}_k\|^2|X_k|^2)\middle|E_k = 1, G_k = 0\right]\Big\} \qquad \text{(B.14)} \\
&\quad = (1-\mu)\mathsf{E}\left[\log\|\mathbf{H}_k\|^2\middle|E_k = 1, G_k = 0\right] + (1-\mu)\mathsf{E}\left[\log|X_k|^2\middle|E_k = 1, G_k = 0\right]
\end{aligned}$$

$$
\begin{aligned}
&\quad + (1-\mu)\epsilon_{\nu,k} && \text{(B.15)}\\
&\leq (1-\mu)\mathsf{E}\big[\log \|\mathbf{H}_k\|^2 \big| E_k = 1, G_k = 0\big] + (1-\mu)\mathsf{E}\big[\log |X_k|^2 \big| E_k = 1, G_k = 0\big]\\
&\quad + \epsilon_{\nu,k}\,. && \text{(B.16)}
\end{aligned}
$$

Note that in (B.14) we use our knowledge $E_k = 1$, i.e., $|X_k| \geq \xi_{\min}$. Plugging (B.9) and (B.16) into (B.3) yields

$$
\begin{aligned}
& I\big(X_k; \|\mathbf{H}_k\||X_k|e^{i\Theta_k} \big| E_k = 1, G_k = 0\big)\\
&\quad \leq -\log 2 - h(\|\mathbf{H}_k\| \big| X_k, E_k = 1, G_k = 0) - \mathsf{E}\big[\log |X_k|^2 \big| E_k = 1, G_k = 0\big]\\
&\qquad - \mathsf{E}\big[\log \|\mathbf{H}_k\| \big| E_k = 1, G_k = 0\big] + \mu \log \eta + \log \Gamma\left(\mu, \frac{\nu}{\eta}\right)\\
&\qquad + (1-\mu)\mathsf{E}\big[\log \|\mathbf{H}_k\|^2 \big| E_k = 1, G_k = 0\big] + (1-\mu)\mathsf{E}\big[\log |X_k|^2 \big| E_k = 1, G_k = 0\big] + \epsilon_{\nu,k}\\
&\qquad + \frac{1}{\eta}\mathsf{E}\big[\|\mathbf{H}_k\|^2 |X_k|^2 \big| E_k = 1, G_k = 0\big] + \frac{\nu}{\eta} && \text{(B.17)}\\
&\quad = -\log 2 - h(\|\mathbf{H}_k\| \big| X_k, E_k = 1, G_k = 0)\\
&\qquad - \mathsf{E}\big[\log \|\mathbf{H}_k\| \big| E_k = 1, G_k = 0\big] + \mu \log \eta + \log \Gamma\left(\mu, \frac{\nu}{\eta}\right)\\
&\qquad + (1-\mu)\mathsf{E}\big[\log \|\mathbf{H}_k\|^2 \big| E_k = 1, G_k = 0\big] - \mu\mathsf{E}\big[\log |X_k|^2 \big| E_k = 1, G_k = 0\big] + \epsilon_{\nu,k}\\
&\qquad + \frac{1}{\eta}\mathsf{E}\big[\|\mathbf{H}_k\|^2 |X_k|^2 \big| E_k = 1, G_k = 0\big] + \frac{\nu}{\eta}. && \text{(B.18)}
\end{aligned}
$$

Next, we continue with the second term in (B.1):

$$
\begin{aligned}
& I\big(X_k; \hat{\mathbf{H}}_k e^{\mathrm{i}(\Phi_k+\Theta_k)} \big| \|\mathbf{H}_k\||X_k|, e^{\mathrm{i}\Theta_k}, E_k = 1, G_k = 0\big)\\
&\quad = h_\lambda\big(\hat{\mathbf{H}}_k e^{\mathrm{i}(\Phi_k+\Theta_k)} \big| \|\mathbf{H}_k\||X_k|, e^{\mathrm{i}\Theta_k}, E_k = 1, G_k = 0\big)\\
&\qquad - h_\lambda\big(\hat{\mathbf{H}}_k e^{\mathrm{i}(\Phi_k+\Theta_k)} \big| \|\mathbf{H}_k\||X_k|, e^{\mathrm{i}\Theta_k}, X_k, E_k = 1, G_k = 0\big) && \text{(B.19)}\\
&\quad \leq h_\lambda\big(\hat{\mathbf{H}}_k e^{\mathrm{i}(\Phi_k+\Theta_k)} \big| E_k = 1, G_k = 0\big) - h_\lambda\big(\hat{\mathbf{H}}_k \big| \|\mathbf{H}_k\|, X_k, E_k = 1, G_k = 0\big) && \text{(B.20)}\\
&\quad = h_\lambda\big(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k} \big| E_k = 1, G_k = 0\big) - h_\lambda\big(\hat{\mathbf{H}}_k \big| \|\mathbf{H}_k\|, X_k, E_k = 1, G_k = 0\big). && \text{(B.21)}
\end{aligned}
$$

Hence, using (B.21) and (B.18) we get the following upper bound for (B.1):

$$
\begin{aligned}
& I\big(X_k; \|\mathbf{H}_k\||X_k|e^{i\Theta_k} \big| E_k = 1, G_k = 0\big)\\
& + I\big(X_k; \hat{\mathbf{H}}_k e^{\mathrm{i}(\Phi_k+\Theta_k)} \big| \|\mathbf{H}_k\||X_k|, e^{\mathrm{i}\Theta_k}, E_k = 1, G_k = 0\big)\\
&\quad \leq -\log 2 - h(\|\mathbf{H}_k\| \big| X_k, E_k = 1, G_k = 0)\\
&\qquad - \mathsf{E}\big[\log \|\mathbf{H}_k\| \big| E_k = 1, G_k = 0\big] + \mu \log \eta + \log \Gamma\left(\mu, \frac{\nu}{\eta}\right)\\
&\qquad + (1-\mu)\mathsf{E}\big[\log \|\mathbf{H}_k\|^2 \big| E_k = 1, G_k = 0\big] - \mu\mathsf{E}\big[\log |X_k|^2 \big| E_k = 1, G_k = 0\big] + \epsilon_{\nu,k}\\
&\qquad + \frac{1}{\eta}\mathsf{E}\big[\|\mathbf{H}_k\|^2 |X_k|^2 \big| E_k = 1, G_k = 0\big] + \frac{\nu}{\eta}\\
&\qquad + h_\lambda\big(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k} \big| E_k = 1, G_k = 0\big) - h_\lambda\big(\hat{\mathbf{H}}_k \big| \|\mathbf{H}_k\|, X_k, E_k = 1, G_k = 0\big) && \text{(B.22)}\\
&\quad = -\log 2 - h(\mathbf{H}_k \big| X_k, E_k = 1, G_k = 0) + (2n_\mathrm{R} - 1)\mathsf{E}\big[\log \|\mathbf{H}_k\| \big| E_k = 1, G_k = 0\big]\\
&\qquad - \mathsf{E}\big[\log \|\mathbf{H}_k\| \big| E_k = 1, G_k = 0\big] + \mu \log \eta + \log \Gamma\left(\mu, \frac{\nu}{\eta}\right)
\end{aligned}
$$

$$
+ (1-\mu)\mathsf{E}\left[\log \|\mathbf{H}_k\|^2 \middle| E_k = 1, G_k = 0\right] - \mu \mathsf{E}\left[\log |X_k|^2 \middle| E_k = 1\right] + \epsilon_{\nu,k}
+ \frac{1}{\eta}\mathsf{E}\left[\|\mathbf{H}_k\|^2 |X_k|^2 \middle| E_k = 1\right] + \frac{\nu}{\eta} + h_\lambda\left(\hat{\mathbf{H}}_k e^{\mathrm{i}\Theta_k} \middle| E_k = 1, G_k = 0\right) \tag{B.23}
$$

Here, (B.23) follows from a conditional version of Lemma 2 in [9] similar to (B.4)–(B.9) which allows us to combine the second and the last term in (B.22).

# Appendix C

# Upper Bounds (6.97) and (6.102)

In this appendix, we will find bounds (6.97) and (6.102).

## C.1 $\delta_1(\kappa, \xi_{\min})$

We first derive upper bound (6.97):

$$\beta_k I\left(\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k}+\mathbf{Z}_k;\mathbf{Z}_{k-\kappa}^{k-1}\,\middle|\,\left\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}+\mathbf{Z}_l\right\}_{l=k-\kappa}^{k-1}, E_k=1, G_k=0\right)$$
$$\leq \delta_1(\kappa,\xi_{\min})+H_{\mathrm{b}}(\beta_k). \tag{C.1}$$

We start as follows:

$$\beta_k I\left(\mathbf{H}_k|X_k|e^{\mathrm{i}\Theta_k}+\mathbf{Z}_k;\mathbf{Z}_{k-\kappa}^{k-1}\middle|\left\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}+\mathbf{Z}_l\right\}_{l=k-\kappa}^{k-1}, E_k=1, G_k=0\right)$$
$$=\beta_k h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle|\left\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}+\mathbf{Z}_l\right\}_{l=k-\kappa}^{k-1}, E_k=1, G_k=0\right)$$
$$-\beta_k h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle|\left\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}+\mathbf{Z}_l\right\}_{l=k-\kappa}^{k}, E_k=1, G_k=0\right) \tag{C.2}$$
$$\leq \beta_k h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle|E_k=1, G_k=0\right)$$
$$-\beta_k h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle|\left\{\mathbf{H}_l|X_l|e^{\mathrm{i}\Theta_l}+\mathbf{Z}_l\right\}_{l=k-\kappa}^{k}, |X_l|_{l=k-\kappa}^{k}, \mathbf{Z}_k, E_k=1, G_k=0\right) \tag{C.3}$$
$$=\beta_k h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle|E_k=1, G_k=0\right)$$
$$-\beta_k h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle|\left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}+\frac{\mathbf{Z}_l}{|X_l|}\right\}_{l=k-\kappa}^{k-1}, \mathbf{H}_k e^{\mathrm{i}\Theta_k}, |X_l|_{l=k-\kappa}^{k}, E_k=1, G_k=0\right). \tag{C.4}$$

Here (C.3) follows from conditioning that reduces entropy. The reason why we do not drop $E_k$ is because we have $\beta_k$ in front of the mutual information, if we drop $E_k$ now, we will not be able to get rid of $\beta_k$ later. In (C.4) we drop $Z_k$ since $\{\mathbf{Z}_k\}$ are IID. In order to get rid of the dependence on input, we take an infimum:

$$\beta_k h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle|E_k=1, G_k=0\right)$$
$$-\beta_k h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle|\left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}+\frac{\mathbf{Z}_l}{|X_l|}\right\}_{l=k-\kappa}^{k-1}, \mathbf{H}_k e^{\mathrm{i}\Theta_k}, |X_l|_{l=k-\kappa}^{k}, E_k=1, G_k=0\right)$$

$$\leq \beta_k h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle| E_k = 1, G_k = 0\right)$$
$$- \beta_k \inf_{\gamma_l \geq \xi_{\min}} h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle|\left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\gamma_l}\right\}_{l=k-\kappa}^{k-1}, \mathbf{H}_k e^{\mathrm{i}\Theta_k}, E_k = 1, G_k = 0\right) \tag{C.5}$$
$$= \beta_k h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle| E_k = 1, G_k = 0\right)$$
$$- \beta_k h\left(\mathbf{Z}_{k-\kappa}^{k-1}\middle|\left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}, \mathbf{H}_k e^{\mathrm{i}\Theta_k}, E_k = 1, G_k = 0\right) \tag{C.6}$$
$$= \beta_k I\left(\mathbf{Z}_{k-\kappa}^{k-1}; \left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}, \mathbf{H}_k e^{\mathrm{i}\Theta_k}\middle| E_k = 1, G_k = 0\right). \tag{C.7}$$

Here (C.6) follows because the smaller $\gamma_l$ is, the more $\mathbf{Z}_{k-\kappa}^{k-1}$ can reflect in $\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\gamma_l}$, thus the smaller the entropy of $\mathbf{Z}_{k-\kappa}^{k-1}$ would be. From this stage, the dependence on input inside mutual information is gone (except $E_k$), but we still have $\beta_k$ in front of mutual information, therefore we add $1 - \beta_k$ to get rid of $\beta_k$ as follows:

$$\beta_k I\left(\mathbf{Z}_{k-\kappa}^{k-1}; \left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}, \mathbf{H}_k e^{\mathrm{i}\Theta_k}\middle| E_k = 1, G_k = 0\right)$$
$$\leq \beta_k I\left(\left\{\frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}; \left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}, \mathbf{H}_k e^{\mathrm{i}\Theta_k}\middle| E_k = 1, G_k = 0\right)$$
$$+ (1-\beta_k) I\left(\left\{\frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}; \left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}, \mathbf{H}_k e^{\mathrm{i}\Theta_k}\middle| E_k = 0, G_k = 0\right) \tag{C.8}$$
$$= I\left(\left\{\frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}; \left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}, \mathbf{H}_k e^{\mathrm{i}\Theta_k}\middle| E_k, G_k = 0\right) \tag{C.9}$$

Now it is time to deal with $E_k$:

$$I\left(\left\{\frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}; \left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}, \mathbf{H}_k e^{\mathrm{i}\Theta_k}\middle| E_k, G_k = 0\right)$$
$$\leq I\left(\left\{\frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}; \left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}, \mathbf{H}_k e^{\mathrm{i}\Theta_k}, E_k\middle| G_k = 0\right) \tag{C.10}$$
$$= I\left(\left\{\frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}; \left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}\middle| G_k = 0\right)$$
$$+ I\left(\left\{\frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}; \mathbf{H}_k e^{\mathrm{i}\Theta_k}\middle| \mathbf{H}_l + \frac{\mathbf{Z}_l}{\xi_{\min}}, G_k = 0\right)$$
$$+ I\left(\left\{\frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}; E_k\middle|\left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=k-\kappa}^{k-1}, \mathbf{H}_k e^{\mathrm{i}\Theta_k}, G_k = 0\right) \tag{C.11}$$
$$\leq h\left(\left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=-\kappa}^{-1}\middle| G_0 = 0\right) - h\left(\left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\right\}_{l=-\kappa}^{-1}\middle| G_0 = 0\right)$$

$$+ h\left(\mathbf{H}_0 e^{\mathrm{i}\Theta_0}\middle|\left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=-\kappa}^{-1}, G_0 = 0\right)$$
$$- h\left(\mathbf{H}_0 e^{\mathrm{i}\Theta_0}\middle|\left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\right\}_{l=-\kappa}^{-1}, G_0 = 0\right) + H_{\mathrm{b}}(\beta_k) \tag{C.12}$$
$$\triangleq \delta_1(\kappa, \xi_{\min}) + H_{\mathrm{b}}(\beta_k). \tag{C.13}$$

Here, (C.12) follows because the last term in (C.11) is smaller than $H_{\mathrm{b}}(\beta_k)$ and $\{\mathbf{H}_k\}$ and $\{\mathbf{Z}_k\}$ are stationary processes.

If $\xi_{\min}$ goes to infinity, $h\left(\left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=-\kappa}^{-1}\middle|G_0 = 0\right)$ converges to $h\left(\left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\right\}_{l=-\kappa}^{-1}\middle|G_0 = 0\right)$ and $h\left(\mathbf{H}_0 e^{\mathrm{i}\Theta_0}\middle|\left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l} + \frac{\mathbf{Z}_l}{\xi_{\min}}\right\}_{l=-\kappa}^{-1}, G_0 = 0\right)$ converges to $h\left(\mathbf{H}_0 e^{\mathrm{i}\Theta_0}\middle|\left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\right\}_{l=-\kappa}^{-1}, G_0 = 0\right)$, therefore, $\delta_1(\kappa, \xi_{\min}) \to 0$.

## C.2 $\delta_2(\kappa, \xi_{\min})$

Next, we derive upper bound (6.102):

$$\beta_k I\left(\left\{\mathbf{H}_l |X_l| e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}; \mathbf{Z}_k \,\middle|\, \mathbf{H}_k |X_k| e^{\mathrm{i}\Theta_k} + \mathbf{Z}_k, E_k = 1, G_k = 0\right)$$
$$\leq \delta_2(\kappa, \xi_{\min}) + H_{\mathrm{b}}(\beta_k). \tag{C.14}$$

The derivation is similar to (C.2)–(C.13).

$$\beta_k I\left(\left\{\mathbf{H}_l |X_l| e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}; \mathbf{Z}_k \,\middle|\, \mathbf{H}_k |X_k| e^{\mathrm{i}\Theta_k} + \mathbf{Z}_k, E_k = 1, G_k = 0\right)$$
$$= \beta_k h\left(\mathbf{Z}_k \,\middle|\, \mathbf{H}_k |X_k| e^{\mathrm{i}\Theta_k} + \mathbf{Z}_k, E_k = 1, G_k = 0\right)$$
$$- \beta_k h\left(\mathbf{Z}_k \,\middle|\, \mathbf{H}_k |X_k| e^{\mathrm{i}\Theta_k} + \mathbf{Z}_k, \left\{\mathbf{H}_l |X_l| e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, E_k = 1, G_k = 0\right) \tag{C.15}$$
$$\leq \beta_k h\left(\mathbf{Z}_k | E_k = 1, G_k = 0\right)$$
$$- \beta_k h\left(\mathbf{Z}_k \middle| \mathbf{H}_k |X_k| e^{\mathrm{i}\Theta_k} + \mathbf{Z}_k, \left\{\mathbf{H}_l |X_l| e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, |X_{k-\kappa}^k|, E_k = 1, G_k = 0\right) \tag{C.16}$$
$$= \beta_k h\left(\mathbf{Z}_k | E_k = 1, G_k = 0\right)$$
$$- \beta_k h\left(\mathbf{Z}_k \middle| \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{|X_k|}, \left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, |X_{k-\kappa}^k|, E_k = 1, G_k = 0\right) \tag{C.17}$$

Here (C.16) follows from conditioning that reduces entropy and for the same reason as in Section C.1, we keep $E_k = 1$. In order to get rid of the dependence on input, we take an infimum:

$$\beta_k h\left(\mathbf{Z}_k | E_k = 1, G_k = 0\right) - \beta_k h\left(\mathbf{Z}_k \middle| \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{|X_k|}, \left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, |X_{k-\kappa}^k|, E_k = 1, G_k = 0\right)$$
$$\leq \beta_k h\left(\mathbf{Z}_k | E_k = 1, G_k = 0\right)$$
$$- \beta_k \inf_{\gamma_k \geq \xi_{\min}} h\left(\mathbf{Z}_k \middle| \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\gamma_k}, \left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, E_k = 1, G_k = 0\right) \tag{C.18}$$
$$= \beta_k h\left(\mathbf{Z}_k | E_k = 1, G_k = 0\right)$$
$$- \beta_k h\left(\mathbf{Z}_k \middle| \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\xi_{\min}}, \left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1}, E_k = 1, G_k = 0\right) \tag{C.19}$$
$$= \beta_k I\left(\mathbf{Z}_k; \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\xi_{\min}}, \left\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\right\}_{l=k-\kappa}^{k-1} \middle| E_k = 1, G_k = 0\right). \tag{C.20}$$

Here, (C.19) follows because the smaller $\gamma_k$ is, the more $\mathbf{Z}_k$ can reflect in $\mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\gamma_k}$, thus the smaller the entropy of $\mathbf{Z}_k$ would be. Next, we want to get rid of $\beta_k$ as follows:

$$
\begin{aligned}
&\beta_k I\left(\mathbf{Z}_k; \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\xi_{\min}}, \{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1} \middle| E_k = 1, G_k = 0\right) \\
&\quad \le \beta_k I\left(\frac{\mathbf{Z}_k}{\xi_{\min}}; \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\xi_{\min}}, \{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1} \middle| E_k = 1, G_k = 0\right) \\
&\qquad + (1-\beta_k) I\left(\frac{\mathbf{Z}_k}{\xi_{\min}}; \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\xi_{\min}}, \{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1} \middle| E_k = 0, G_k = 0\right) && \text{(C.21)} \\
&\quad = I\left(\frac{\mathbf{Z}_k}{\xi_{\min}}; \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\xi_{\min}}, \{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1} \middle| E_k, G_k = 0\right) && \text{(C.22)} \\
&\quad \le I\left(\frac{\mathbf{Z}_k}{\xi_{\min}}; \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\xi_{\min}}, \{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1}, E_k \middle| G_k = 0\right) && \text{(C.23)} \\
&\quad = I\left(\frac{\mathbf{Z}_k}{\xi_{\min}}; \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\xi_{\min}} \middle| G_k = 0\right) + I\left(\frac{\mathbf{Z}_k}{\xi_{\min}}; \{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1} \middle| \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\xi_{\min}}, G_k = 0\right) \\
&\qquad + I\left(\frac{\mathbf{Z}_k}{\xi_{\min}}; E_k \middle| \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\xi_{\min}}, \{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1}, G_k = 0\right) && \text{(C.24)} \\
&\quad \le h\left(\mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\xi_{\min}} \middle| G_k = 0\right) - h(\mathbf{H}_k e^{\mathrm{i}\Theta_k} | G_k = 0) \\
&\qquad + h\left(\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1} \middle| \mathbf{H}_k e^{\mathrm{i}\Theta_k} + \frac{\mathbf{Z}_k}{\xi_{\min}}, G_k = 0\right) - h\left(\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\}_{l=k-\kappa}^{k-1} \middle| \mathbf{H}_k e^{\mathrm{i}\Theta_k}, G_k = 0\right) \\
&\qquad + H_{\mathrm{b}}(\beta_k) && \text{(C.25)} \\
&\quad = h\left(\mathbf{H}_0 e^{\mathrm{i}\Theta_0} + \frac{\mathbf{Z}_0}{\xi_{\min}} \middle| G_0 = 0\right) - h\left(\mathbf{H}_0 e^{\mathrm{i}\Theta_0} | G_0 = 0\right) \\
&\qquad + h\left(\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\}_{-\kappa}^{-1} \middle| \mathbf{H}_0 e^{\mathrm{i}\Theta_0} + \frac{\mathbf{Z}_0}{\xi_{\min}}, G_0 = 0\right) - h\left(\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\}_{-\kappa}^{-1} \middle| \mathbf{H}_0 e^{\mathrm{i}\Theta_0}, G_0 = 0\right) \\
&\qquad + H_{\mathrm{b}}(\beta_k) && \text{(C.26)} \\
&\quad \triangleq \delta_2(\kappa, \xi_{\min}) + H_{\mathrm{b}}(\beta_k). && \text{(C.27)}
\end{aligned}
$$

Here, (C.25) follows because the last term in (C.24) is smaller than $H_{\mathrm{b}}(\beta_k)$ and $\{\mathbf{H}_k\}$ and $\{\mathbf{Z}_k\}$ are independent of each other; (C.26) follows because $\{\mathbf{H}_k\}$ and $\{\mathbf{Z}_k\}$ are stationary processes.

If $\xi_{\min}$ goes to infinity, $h\left(\mathbf{H}_0 e^{\mathrm{i}\Theta_0} + \frac{\mathbf{Z}_0}{\xi_{\min}} \middle| G_0 = 0\right)$ converges to $h(\mathbf{H}_0 e^{\mathrm{i}\Theta_0} | G_0 = 0)$ and $h\left(\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\}_{-\kappa}^{-1} \middle| \mathbf{H}_0 e^{\mathrm{i}\Theta_0} + \frac{\mathbf{Z}_0}{\xi_{\min}}, G_0 = 0\right)$ converges to $h\left(\{\mathbf{H}_l e^{\mathrm{i}\Theta_l}\}_{-\kappa}^{-1} \middle| \mathbf{H}_0 e^{\mathrm{i}\Theta_0}, G_0 = 0\right)$, therefore, $\delta_2(\kappa, \xi_{\min}) \to 0$.

# Appendix D

# Causal Interpretations for Independence

In Figures D.3–D.5 we prove the following independence claims used in (6.21), (6.73), and (6.84).

- $(M, \mathbf{Y}_1^{k-1}) \perp\!\!\!\perp \mathbf{Y}_k$ when conditioned on $(X_k, \mathbf{H}_1^{k-1})$;
- $X_k \perp\!\!\!\perp \mathbf{Y}_k$ when conditioned on $(\mathbf{H}_k X_k)$;
- $\mathbf{H}_1^{k-1} \perp\!\!\!\perp \mathbf{Y}_k$ when conditioned on $(X_k, \mathbf{H}_k)$.

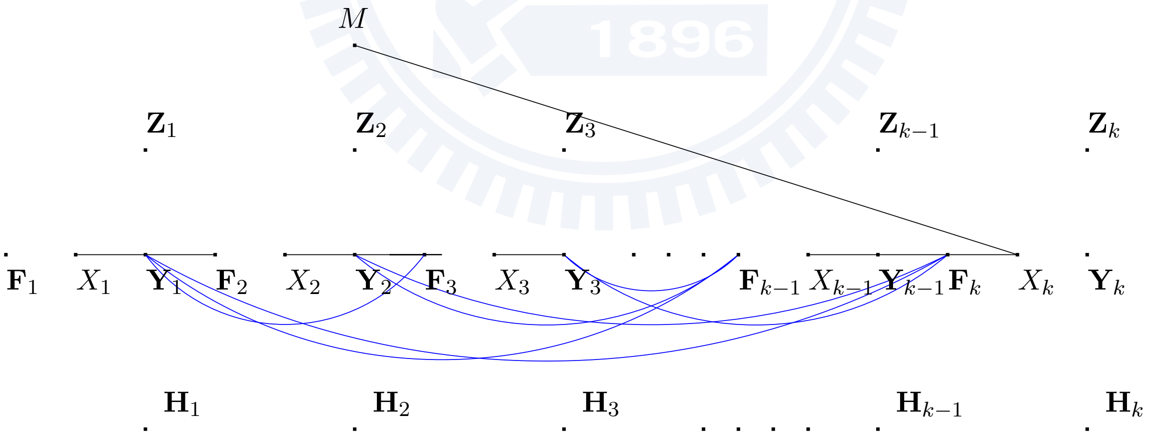


Figure D.3: The relevant subgraph of $\mathbf{V}$ showing the independence of $(M, \mathbf{Y}_1^{k-1})$ and $\mathbf{Y}_k$ when conditioned on $(X_k, \mathbf{H}_1^{k-1})$.

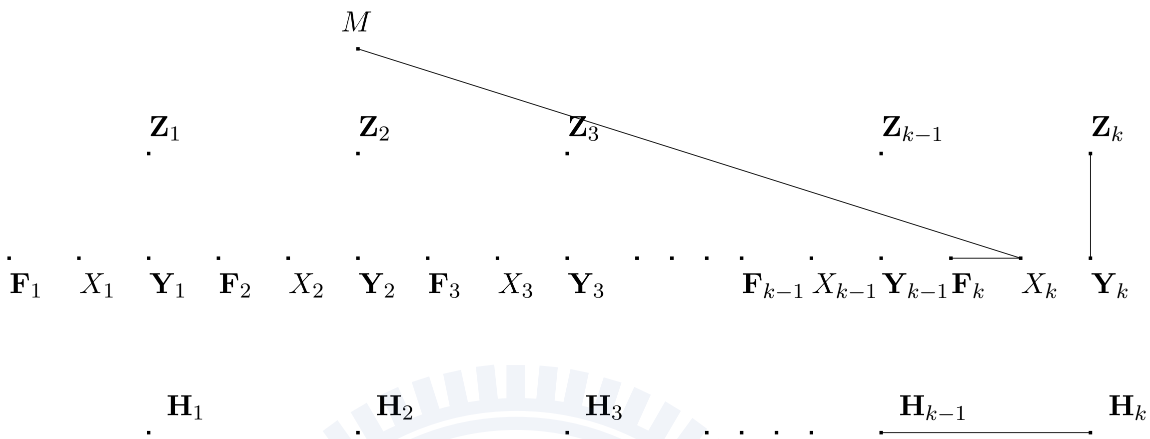


Figure D.4: The relevant subgraph of $\mathbf{V}$ showing the independence of $X_k$ and $\mathbf{Y}_k$ when conditioned on $(\mathbf{H}_k X_k)$.

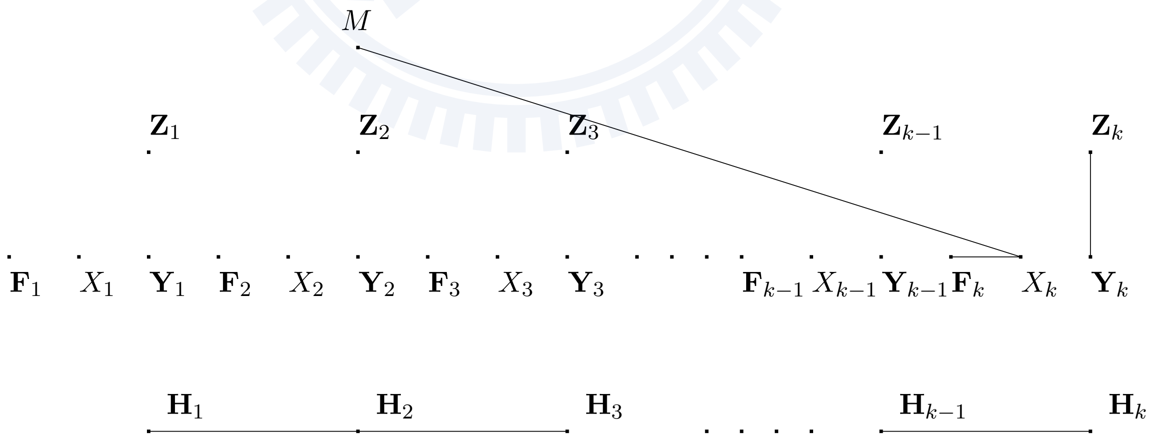


Figure D.5: The relevant subgraph of $\mathbf{V}$ showing the independence of $\mathbf{H}_1^{k-1}$ and $\mathbf{Y}_k$ when conditioned on $(X_k, \mathbf{H}_k)$.

# Bibliography

[1] İ. E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Transactions on Telecommunications*, vol. 10, no. 6, pp. 585–595, November–December 1999.

[2] G. Durisi, U. G. Schuster, H. Bölcskei, and S. Shamai (Shitz), “Noncoherent capacity of underspread fading channels,” *IEEE Transactions on Information Theory*, vol. 56, no. 1, pp. 367–395, January 2010.

[3] T. L. Marzetta and B. M. Hochwald, “Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading,” *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, January 1999.

[4] L. Zheng and D. N. C. Tse, “Communicating on the Grassmann manifold: a geometric approach to the non coherent multiple-antenna channel,” *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, February 2002.

[5] Y. Liang and V. V. Veeravalli, “Capacity of noncoherent time-selective Rayleigh-fading channels,” *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3095–3110, December 2004.

[6] T. Koch and A. Lapidoth, “Degrees of freedom in non-coherent stationary MIMO fading channels,” in *Proceedings Winter School on Coding and Information Theory*, Bratislava, Slovakia, February 20–25, 2005, pp. 91–97.

[7] ——, “The fading number and degrees of freedom in non-coherent MIMO fading channels: a peace pipe,” in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, September 4–9, 2005, pp. 661–665.

[8] A. Lapidoth and S. M. Moser, “Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels,” *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2426–2467, October 2003.

[9] S. M. Moser, “The fading number of multiple-input multiple-output fading channels with memory,” *IEEE Transactions on Information Theory*, vol. 55, no. 6, pp. 2716–2755, June 2009.

[10] ——, "Duality-based bounds on channel capacity," Ph.D. dissertation, ETH Zurich, October 2004, Diss. ETH No. 15769. [Online]. Available: http://moser.cm.nctu.edu.tw/publications.html

[11] R. G. Gallager, *Low Density Parity Check Codes.* Cambridge, MA: MIT Press, 1962.

[12] J. L. Massey, "Causal interpretations of random variables," *Problemy Peredachi Informatsii (Problems of Information Transmission)*, vol. 32, no. 1, pp. 131–136, January–March 1996.

[13] ——, "Correction to 'Causal interpretations of random variables'," *Problemy Peredachi Informatsii (Problems of Information Transmission)*, May 16, 1997.

[14] R. L. Dobrušin, "General formulation of Shannon's main theorem in information theory," in *Proceedings American Mathematical Society Translations*, ser. 2, 1963, vol. 33, pp. 323–438.

[15] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 437–453, February 2006.

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