Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Signal and Information Processing Laboratory
Institut für Signal- und Informationsverarbeitung

Winter Semester 2003/2004 Prof. Dr. A. Lapidoth

Master's Thesis

On the Asymptotic Capacity of Multiple-Input Single-Output Fading Channels with Memory

Tobias Koch

Advisor: Prof. Amos Lapidoth

Co-Advisors: Stefan M. Moser and Daniel Hösli

aufgabenstellung1...

aufgabenstellung2...

Acknowledgments

I wish to thank my advisor Amos Lapidoth for his kind help during the last six months. The discussions with him led to many important results in this thesis. Additionally, I would like to thank Stefan Moser and Daniel Hösli for their friendly assistance and for supporting me at any time.

Zurich, April 23, 2004

Tobias Koch

Abstract

In this thesis we study the capacity of Gaussian fading channels with memory where neither transmitter nor receiver has access to the realization of the fading process, though both are fully cognizant of the fading laws. The emphasis is on the high signal-to-noise ratio (SNR) regime.

For regular fading processes, *i.e.*, the "present" fading cannot be predicted precisely from its "past", we derive an upper bound on the fading number of multiple-input single-output (MISO) fading channels, where the fading number is the second-order term in the high SNR expansion of capacity. We show that this bound is tight (*i.e.*, it coincides with a lower bound) if the channels are uncorrelated and the fading process is either zero-mean or its spectral density matrix contains identical entries. In the former case, the fading number can be achieved by transmitting from only one antenna, *i.e.*, the one that yields the smallest prediction error in predicting the "present" fading from its "past". In the latter case, the fading number can be achieved by beam forming.

For non-regular fading processes, *i.e.*, the "present" fading can be predicted precisely from its "past", we derive upper bounds on the capacity of MISO and multiple-input multiple-output (MIMO) fading channels. For cases where the channels are uncorrelated the bound on the MISO capacity is tight and we provide an expression for the pre-log, *i.e.*, the limiting ratio of the capacity to the logarithm of the SNR. Moreover, we show that this pre-log can be achieved by transmitting from only one antenna, *i.e.*, the one that yields the smallest prediction error in predicting the "present" fading from its "past". In addition, we present an improved lower bound on the capacity of single-input single-output (SISO) fading channels if channel capacity only grows double-logarithmically in the SNR. This allows for an expression of the pre-log-log, *i.e.*, the limiting ratio of the capacity to the logarithm of the logarithm of the SNR.

Keywords: channel capacity, fading channels, fading number, high SNR, multiplexing gain, non-coherent, pre-log, pre-log-log, regular and non-regular fading processes.

Contents

1 Introduction **1**
1.1 Motivation 1
1.2 Notation 3

2 Channel Capacity **5**

3 Linear Prediction **9**
3.1 Univariate Prediction 9
3.2 Multivariate Prediction 12

4 The Channel **15**
4.1 The Physical Description of the Channel 15
4.2 The Channel Model 19

5 Results **23**
5.1 Regular Processes 24
5.1.1 Previous Results 25
5.1.2 The Fading Number of MISO Fading Channels 26
5.2 Non-Regular Processes 29
5.2.1 Previous Results 30
5.2.2 The Pre-Log-Log of SISO Fading Channels 31
5.2.3 The Pre-Log of MISO Fading Channels 32
5.2.4 The Pre-Log of MIMO Fading Channels 33

6 Regular Processes **37**
6.1 A Proof of Theorem 5.5 37
6.2 A Proof of Corollary 5.6 41
6.3 A Proof of Corollary 5.7 43

7 Non-Regular Processes **47**
7.1 The Pre-Log-Log of SISO Fading Channels 47
7.1.1 A Proof of Theorem 5.10 48
7.1.2 A Proof of Corollary 5.11 51
7.2 The Pre-Log of MISO Fading Channels 54
7.2.1 A Proof of Theorem 5.12 54
7.2.2 A Proof of Corollary 5.13 58
7.3 The Pre-Log of MIMO Fading Channels 60
7.3.1 Notation 61
7.3.2 A Proof of Theorem 5.14 63
7.3.3 A Proof of Corollary 5.15 67

8 Discussion and Open Problems **71**
8.1 Discussion 71
8.1.1 SISO Fading Channels 72
8.1.2 MISO Fading Channels 72
8.1.3 MIMO Fading Channels 74
8.2 Open Problems 75

9 Summary and Conclusion **77**

A Proof of Lemma 3.1 **79**

Bibliography **83**

Chapter 1

Introduction

1.1 Motivation

Wireless communication systems have been investigated elaborately in the last few years as these systems offer high data rates at low implementation cost. In contrast to wired transmission through copper or optical fibers, a wireless link is highly affected by its environment and the interference therefrom. Additionally, due to movements of transmitter, receiver, or scatterers in the environment the link may vary over time. The resulting variation of the channel is called *fading* and we therefore refer to these channels as *fading channels*. Usually, the fading is modelled by a multiplicative noise term. In addition, some additive white Gaussian noise models the disturbance introduced by the receiver.

An important performance measure to analyze these channels is the *channel capacity*. The channel capacity is the largest rate at which reliable communication, *i.e.*, with arbitrarily small error-probability, is possible.

Recently, it was shown by Lapidoth and Moser [1] that if the law of the fading process is known to both, transmitter and receiver, but neither of them has knowledge of its realization and if the process is regular in the sense that the realization of the "present" fading cannot be predicted precisely from its "past", then capacity grows double-logarithmically in the signal-to-noise ratio (SNR). This result differs dramatically from the case where either transmitter, receiver, or both have access to the fading realization (perfect side information). Here, capacity increases logarithmically with the SNR.

In order to bridge the gap between the double-logarithmic and the logarithmic behaviors, Lapidoth extended the study in [1] to non-regular or deterministic fading processes, where the "present" fading can be predicted precisely from its "past" [2]. In particular, the capacity of single-input single-output (SISO) discrete-time Gaussian fading channels with memory was studied, where neither transmitter nor receiver has access to the realization of the fading process. It was demonstrated that for non-regular processes the increase of capacity with the SNR may be logarithmically, double-logarithmically or in between, *e.g.*, as a fractional power of the logarithm of the SNR. Additionally, an expression for the pre-log, *i.e.*, the limiting ratio of channel capacity to the logarithm of the SNR, was presented.

This thesis addresses the capacity of multiple-input single-output (MISO) Gaussian fading channels with memory. Both, regular and non-regular fading processes are considered. Again, neither transmitter nor receiver has access to the fading realization, though both are cognizant of the fading law. The emphasis is on the high SNR regime. In particular, for non-regular processes we determine the pre-log. In the case where the fading is regular, we give an expression for the fading number[1]. In addition, we derive an improved lower bound on the capacity of SISO fading channels as well as an upper bound on the capacity of multiple-input multiple-output (MIMO) fading channels where the fading process is in both cases non-regular.

This thesis is organized as follows: in Chapter 2 we provide a background in information theory. Chapter 3 studies the prediction theory of univariate and multivariate stochastic processes. In Chapter 4 we consider wireless communication links and establish the channel model that we address. The results obtained in this thesis are presented in Chapter 5. Chapters 6 and 7 show the derivations that yield those results. The results are discussed in Chapter 8; and Chapter 9 concludes this thesis with a brief summary.

[1]The fading number is the second-order term in the high SNR expansion of capacity. It was introduced in [1] by Lapidoth and Moser and can be viewed as an indication of the practical limiting rate for power-efficient communication over the channel.

1.2 Notation

In this section we establish the notation that we are using in this thesis. We have to differentiate between random and deterministic quantities as well as between scalars, vectors and matrices.

When dealing with scalars, we shall use upper case letters such as X for random quantities and lower case letters such as x for its realization. Random vectors are denoted with bold face upper case letters, *e.g.*, $\mathbf{X}$, and its realizations are written in bold lower case letters, *e.g.*, $\mathbf{x}$. For deterministic matrices, we use upper case letters of a special font, *e.g.*, H, for random matrices we use yet another font, *e.g*, $\mathbb{H}$.

In order to denote the entries of a matrix we shall use superscripts so that $H^{(r,t)}$ denotes the component of $\mathbb{H}$ that lies in row-r and column-t. We use r and t as indices because we think of r indexing the receive antennas and t indexing the transmit antennas. Consequently, we often use n_{R} to denote the number of rows and n_{T} to denote the number of columns.

Typically, subscripts are used for time indices. Thus, the matrix $\mathbb{H}$ at time k is denoted by $\mathbb{H}_k$. Sequences of random variables are denoted using a combination of superscript and subscript. So, if $X_1, X_2, \ldots$ is a sequence of random variables, X_k^n describes the sequence $X_k, \ldots, X_n$. If $k = 1$ we shall often write X^n instead of X_1^n.

With $\|\cdot\|$ we denote the Euclidean norm of vectors or the Euclidean operator norm of matrices, *i.e.*,

$$\|\mathbf{x}\| = \sqrt{\sum_{t=1}^{n_T} |x^{(t)}|^2}, \qquad \mathbf{x} \in \mathbb{C}^{n_{\mathrm{T}}}, \tag{1.1}$$

$$\|\mathsf{A}\| = \max_{\|\hat{\mathbf{w}}\|=1} \|\mathsf{A}\hat{\mathbf{w}}\|. \tag{1.2}$$

All rates given in this thesis are in nats per channel use. When writing $\log(\cdot)$ we shall mean the natural logarithm function.

The mean-μ variance-σ^2 univariate real Gaussian distribution is denoted by $\mathcal{N}_{\mathbb{R}}\left(\mu, \sigma^2\right)$. Similarly, $\mathcal{N}_{\mathbb{R}}\left(\boldsymbol{\mu}, \mathsf{K}\right)$ denominates the distribution of a real Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix K. With $\mathcal{N}_{\mathbb{C}}\left(\mu, \sigma^2\right)$ we shall mean the distribution of a complex Gaussian random variable with mean μ and variance σ^2 and where $X - \mu$ is circularly symmetric, *i.e.*, where the real and imaginary part of $X - \mu$ are independent $\mathcal{N}_{\mathbb{R}}\left(0, \sigma^2/2\right)$ random variables. The distribution of a complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix K and where $\mathbf{X} - \boldsymbol{\mu}$ is circularly symmetric is written as $\mathcal{N}_{\mathbb{C}}(\boldsymbol{\mu}, \mathsf{K})$.

Chapter 2

Channel Capacity

This chapter provides a background in information theory. In particular, we introduce the notion of channel capacity which is an important performance measure in communications.

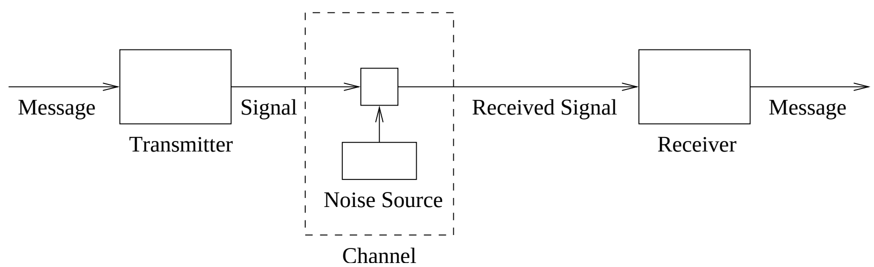

Figure 2.1: A communication system

We consider a communication system with a transmitter, a channel and a receiver, as shown in Figure 2.1[1]. The task of the communication system is to transmit a message from one point to another. The message could be a text or a picture that someone wants to send to a friend. The transmitter produces signals suitable for transmission over the channel. The channel is the medium used to transmit the signal from transmitter to receiver. It can be a pair of wires, a coaxial cable or air. Usually, the transmitted signal is perturbed by noise which is indicated in Figure 2.1 by the noise source. Based on the received signal, the receiver has to decide which message has been sent. If the decision differs from the original message an error occurs.

Figure 2.2 shows a mathematical model of the communication system shown in Figure 2.1. The message M is a random variable taken from the

[1]This communication system was originally introduced by Shannon in [3].

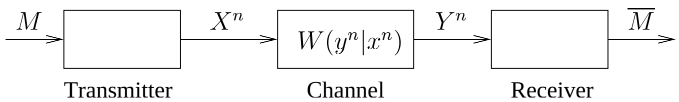

Figure 2.2: Mathematical description of a communication system

index set $\{1, 2, \ldots, |\mathcal{M}|\}$, where $|\mathcal{M}|$ denotes the number of elements in the set $\mathcal{M}$. Note that we can express any finite set $\mathcal{M}$ of messages in terms of such an index set by indexing all messages. The channel input X^n is a sequence of random variables X_k taking values in the set $\mathcal{X}$. Possibly, the input sequences have to fulfill an additional constraint (*e.g.*, power constraint, bandwidth constraint). We refer to a sequence x^n as a *codeword*. Usually, the alphabet of the codewords is restricted to a chosen codebook $\mathcal{C}$ satisfying

$$\mathcal{C} \subseteq \mathcal{X}^n \tag{2.1}$$

and $|\mathcal{C}| = |\mathcal{M}|$. The transmitter performs a mapping from the index set $\mathcal{M}$ to the set $\mathcal{C}$ referred to as *encoding*

$$\Psi_n(m) : \mathcal{M} \to \mathcal{C}. \tag{2.2}$$

The channel output Y^n is a sequence of random variables Y_k taking values in the set $\mathcal{Y}$. Based on the sequence y^n the receiver has to decide which message m has been sent. If the decision $\overline{m}$ differs from the original message m an error occurs. The decision process is referred to as *decoding* and can be described by the mapping—the so-called *decoding rule*—

$$\Phi_n(y^n) : \mathcal{Y} \to \mathcal{M}. \tag{2.3}$$

If a codebook $\mathcal{C}$ and a mapping $\Phi_n(y^n)$ can be chosen in such a way that for a sufficiently large n the probability of error can be made arbitrarily small, then we refer to this case as *reliable communication*.

The chosen codebook determines the rate defined as

$$\mathcal{R} = \frac{\log|\mathcal{C}|}{n}. \tag{2.4}$$

The rate is a measure of how many information symbols can be transmitted with a codeword of length n. The highest rate at which reliable communication is possible is called *channel capacity*. One can show that for the channel model that we are addressing in this thesis (see Section 4.2) the capacity is given by [4]

$$C = \lim_{n\to\infty} \frac{1}{n} \sup_{Q\in\mathcal{P}(\mathcal{X}^n)} I(X^n;Y^n), \tag{2.5}$$

where $\mathcal{P}(\mathcal{X}^n)$ is the set of all probability densities over $\mathcal{X}^n$ fulfilling the input constraint, and $I(X^n;Y^n)$ denotes the mutual information between the channel input sequence X^n and the output sequence Y^n given by

$$I(X^n;Y^n) = \int_{x^n\in\mathcal{X}^n}\int_{y^n\in\mathcal{Y}^n} Q(x^n)W(y^n|x^n)\log\frac{W(y^n|x^n)}{(QW)\,(y^n)}\,\mathrm{d}y^n\,\mathrm{d}x^n. \tag{2.6}$$

The density $Q(x^n)$ denotes the probability density of the input sequence X^n and $(QW)\,(y^n)$ denominates the probability density of the output sequence Y^n induced by the channel with law $W(y^n|x^n)$ and an input sequence according to the probability density $Q(x^n)$:

$$(QW)\,(y^n) = \int_{x'^n\in\mathcal{X}^n} Q(x'^n)W(y^n|x'^n)\,\mathrm{d}x'^n, \qquad y^n\in\mathcal{Y}^n. \tag{2.7}$$

The mutual information of random variables with probability densities can as well be expressed as

$$I(X^n;Y^n) = h(Y^n) - h(Y^n|X^n), \tag{2.8}$$

where $h(Y^n)$ denotes the differential entropy of the output given by

$$h(Y^n) = -\int_{y^n\in\mathcal{Y}^n}(QW)(y^n)\log(QW)(y^n)\,\mathrm{d}y^n \tag{2.9}$$

and $h(Y^n|X^n)$ denominates the conditional differential entropy of the channel

$$\begin{aligned} &h(Y^n|X^n)\\ &= \int_{x^n\in\mathcal{X}^n} Q(x^n)h(Y^n|X^n=x^n)\,\mathrm{d}x^n\\ &= -\int_{x^n\in\mathcal{X}^n} Q(x^n)\int_{y^n\in\mathcal{Y}^n} W(y^n|x^n)\log W(y^n|x^n)\,\mathrm{d}y^n\,\mathrm{d}x^n. \end{aligned} \tag{2.10}$$

For a description of the properties of mutual information and differential entropy see [5].

In general, the maximization over all input distributions in (2.5) can be very complicated. Therefore, it is often easier to find upper and lower bounds on the capacity. In the ideal case these bounds are tight (*i.e.*, they coincide) and one obtains an expression for the capacity. Good lower bounds can be found by choosing an input distribution $Q(x^n)$ close to the capacity-achieving input distribution that maximizes (2.5). This cannot be done in order to find upper bounds and, thus, good upper bounds are often more difficult to find. In Chapter 5 we present upper bounds on the capacity of fading channels.

Chapter 3

Linear Prediction

This chapter outlines the prediction theory of univariate and multivariate stochastic processes. In particular, we study the prediction error of linear predictors, which are optimal when the stochastic processes are Gaussian. This will be of importance when considering the capacity of channels with memory.

The chapter is divided into two parts. In Section 3.1 we consider univariate processes. For both regular and non-regular processes expressions are given that connect the minimum mean squared prediction error to the spectral density function of the process. In Section 3.2 multivariate processes are considered. We show for both regular and non-regular processes identities that link the determinant of the prediction error covariance matrix to the determinant of the matrix-valued spectral density function. Additionally, we show that if the entries in the random vectors of the stochastic process are uncorrelated, then the prediction error covariance matrix is diagonal and the diagonal entries correspond to the univariate prediction errors.

3.1 Univariate Prediction

In this section we glance at the prediction theory of univariate stochastic processes. In particular, we aim at an expression for the mean squared error in predicting a random variable A_0 from past values $A_{-1}, A_{-2}, \ldots$.

Consider a univariate zero-mean stationary stochastic process $\{A_k\}$ with *spectral distribution function* $\mathsf{F}(\lambda)$, $-1/2 \leq \lambda \leq 1/2$. Thus, $\mathsf{F}(\lambda)$ is a monotonically non-decreasing function on $[-1/2, 1/2]$ (see [2] and references therein) satisfying

$$\mathsf{E}\left[A_{k+m}A_k^*\right] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda}\,\mathrm{d}\mathsf{F}(\lambda), \qquad k, m \in \mathbb{Z}. \tag{3.1}$$

In the following we assume $\mathsf{F}(\lambda)$ to be absolutely continuous. Then, $\mathsf{F}'(\lambda)$ denotes its derivative and is referred to as *spectral density function*.

The task of prediction is now to estimate A_0 based on the past values $A_{-1}, A_{-2}, \dots$ in a way that the mean squared prediction error $\epsilon^2_{\mathrm{MSE}}$ between A_0 and its estimate $\overline{A}_0$, *i.e.*, $\mathsf{E}\left[|A_0 - \overline{A}_0|^2\right]$, is minimized. In general, the estimate that minimizes the mean squared error given the past values $a_{-1}, a_{-2}, \dots$ is given by

$$\overline{a}_0 = \mathsf{E}\left[A_0 \;\middle|\; A_{-\infty}^{-1}\right] \tag{3.2}$$

and may be very difficult to compute. However, if the stochastic process $\{A_k\}$ is Gaussian, then it is well known that $\overline{A}_0$ is of the form

$$\overline{A}_0 = \sum_{k=-\infty}^{-1} c_k A_k, \tag{3.3}$$

where the parameters c_k have to be chosen such that the prediction error is minimized. Thus, the estimate is a linear combination of the past values $A_{-1}, A_{-2}, \dots$ and we refer to this case as *linear prediction*.

It was shown that if one is restricted to a linear predictor and if the prediction error $\epsilon^2_{\mathrm{MSE}} > 0$, then there exists a formula that connects the minimum mean squared error $\epsilon^2_{\mathrm{MSE}}$ to the spectral density function $\mathsf{F}'(\lambda)$ (*e.g.*, [6]):

$$\epsilon^2_{\mathrm{MSE}} = \exp\left\{\int_{-1/2}^{1/2} \log \mathsf{F}'(\lambda)\,\mathrm{d}\lambda\right\}. \tag{3.4}$$

This formula is sometimes referred to as *Kolmogorov's formula*. It should be noted that this result holds for all kind of processes $\{A_k\}$, as long as linear prediction is performed. However, only if the process is Gaussian, then this error corresponds to the minimum mean squared prediction error among all predictors.

In [2] the term *regular* is used for processes with $\epsilon^2_{\mathrm{MSE}} > 0$ whereas processes for which $\epsilon^2_{\mathrm{MSE}} = 0$ are called *non-regular* or *deterministic*.

As mentioned above the expression for the prediction error (3.4) holds only if $\epsilon^2_{\mathrm{MSE}} > 0$. Thus, in cases where the process is non-regular (*i.e.*, $\epsilon^2_{\mathrm{MSE}} = 0$), we shall study the noisy prediction problem stated in [2] instead.

Let $\{W_k\}$ be a sequence of independent and identically distributed (i.i.d.) $\mathcal{N}\left(0, \delta^2\right)$ random variables, for a given δ^2. Furthermore, let the stochastic process $\{A_k\}$ be Gaussian and independent from the process $\{W_k\}$, *i.e.*, $A_k \perp\!\!\!\perp W_m$, $k, m \in \mathbb{Z}$. The noisy prediction problem is to predict A_0 based on the observations $A_{-1} + W_{-1}, A_{-2} + W_{-2}, \ldots$. It follows that in this case the minimum mean squared prediction error denoted by $\epsilon^2_{\mathrm{MSE}}(\delta^2)$ is given by

$$\epsilon^2_{\mathrm{MSE}}(\delta^2) = \exp\left\{\int_{-1/2}^{1/2} \log(\mathsf{F}'(\lambda) + \delta^2)\,\mathrm{d}\lambda\right\} - \delta^2. \tag{3.5}$$

This can be easily verified by noting that the conditional expectation of A_0 given the observations $A_{-1} + W_{-1}, A_{-2} + W_{-2}, \ldots$ is identical to the conditional expectation of $A_0 + W_0$ given those observations, *i.e.*,

$$\mathsf{E}\left[A_0 \;\middle|\; \{A_\nu + W_\nu\}_{\nu=-\infty}^{-1}\right] = \mathsf{E}\left[A_0 + W_0 \;\middle|\; \{A_\nu + W_\nu\}_{\nu=-\infty}^{-1}\right]. \tag{3.6}$$

Since W_0 is independent of A_0 and of the observations, it follows that the prediction error $\epsilon^2_{\mathrm{MSE}}(\delta^2)$ can be written as the prediction error of the process $\{A_k + W_k\}$ but with the variance of W_0 subtracted:

$$\begin{aligned}
\epsilon^2_{\mathrm{MSE}}(\delta^2) &= \mathsf{E}\left[|A_0 - \overline{A}_0|^2 \;\middle|\; \{A_\nu + W_\nu\}_{\nu=-\infty}^{-1}\right] \\
&= \mathsf{E}\left[|A_0 - \overline{A_0 + W_0}|^2 \;\middle|\; \{A_\nu + W_\nu\}_{\nu=-\infty}^{-1}\right] \\
&= \mathsf{E}\left[|A_0 - \overline{A_0 + W_0}|^2 + |W_0|^2 \;\middle|\; \{A_\nu + W_\nu\}_{\nu=-\infty}^{-1}\right] \\
&\quad - \mathsf{E}\left[|W_0|^2 \;\middle|\; \{A_\nu + W_\nu\}_{\nu=-\infty}^{-1}\right] \\
&= \mathsf{E}\left[|A_0 + W_0 - \overline{A_0 + W_0}|^2 \;\middle|\; \{A_\nu + W_\nu\}_{\nu=-\infty}^{-1}\right] - \mathsf{E}\left[|W_0|^2\right],
\end{aligned} \tag{3.7}$$

with $\overline{A}_0$ and $\overline{A_0 + W_0}$ being the estimates of A_0 and $A_0 + W_0$, respectively, given the observations. We then obtain (3.5) by noting that the density of $\{A_k + W_k\}$ is given by $\mathsf{F}'(\lambda) + \delta^2$.

At the end of this section, we recall some facts related to the prediction problem for circularly symmetric stationary Gaussian processes. We first note that if $\{A_k\}$ is a Gaussian process, then the random variable A_0 conditioned on $A_{-1}, A_{-2}, \ldots, A_{-n}$ has a Gaussian distribution with mean $\mathsf{E}\left[A_0 \;\middle|\; A_{-n}^{-1}\right]$ and variance $\epsilon_n^2 = \mathsf{E}\left[\left|A_0 - \mathsf{E}\left[A_0 \;\middle|\; A_{-n}^{-1}\right]\right|^2\right]$ that is independent of the realization of $A_{-1}, A_{-2}, \ldots, A_{-n}$. We denote the variance by ϵ_n^2 since it corresponds to the minimum mean squared error in predicting A_0 from past values $A_{-1}, A_{-2}, \ldots, A_{-n}$.

Furthermore, it follows by [2] and references therein, that if $\{A_k\}$ is additionally stationary, then the prediction error ϵ_n^2 is monotonically non-increasing in n and

$$\lim_{n\to\infty} \epsilon_n^2 = \epsilon_{\mathrm{MSE}}^2, \tag{3.8}$$

with $\epsilon_{\mathrm{MSE}}^2$ as in (3.4) and (3.5), respectively.

3.2 Multivariate Prediction

In the following we give an overview over the prediction theory of multivariate stochastic processes that is based on work by Wiener and Masani [7]. The concepts are similar to that in the univariate case.

We consider a multivariate zero-mean stationary stochastic process $\{\mathbf{A}_k\}$. Since $\mathbf{A}_k$ is a vector, the spectral distribution function $\mathsf{F}(\lambda)$ is now a matrix with diagonal entries that are real-valued and monotonically non-decreasing on $[-1/2, 1/2]$ [7]. The matrix-valued spectral distribution function satisfies

$$\mathsf{E}\left[\mathbf{A}_{k+m}\mathbf{A}_k^\dagger\right] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda}\,\mathrm{d}\mathsf{F}(\lambda), \qquad k, m \in \mathbb{Z}. \tag{3.9}$$

Thus, the (r,t)-th entry of the $n_\mathrm{R} \times n_\mathrm{R}$ covariance matrix $\mathsf{E}\left[\mathbf{A}_{k+m}\mathbf{A}_k^\dagger\right]$ is given by

$$\mathsf{E}\left[A_{k+m}^{(r)} A_k^{(t)*}\right] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda}\,\mathrm{d}\mathsf{F}^{(r,t)}(\lambda), \qquad k, m \in \mathbb{Z}. \tag{3.10}$$

In the following we assume that all entries in $\mathsf{F}(\lambda)$ are absolutely continuous. Then, like in the univariate case, $\mathsf{F}'(\lambda)$ denotes the derivative of $\mathsf{F}(\lambda)$ and is referred to as the (matrix-valued) spectral density function.

Notice that if the entries in $\mathbf{A}_k$ are uncorrelated, then the covariance matrix $\mathsf{E}\left[\mathbf{A}_{k+m}\mathbf{A}_k^\dagger\right]$ is diagonal for all $k, m \in \mathbb{Z}$ and it follows that the spectral density function $\mathsf{F}'(\lambda)$ is diagonal as well. This can be verified by noting that the covariance matrix and the spectral density function are Fourier pairs. Thus, if the (r,t)-th entry in the covariance matrix is zero for all $k, m \in \mathbb{Z}$, then the corresponding entry in the spectral density matrix must be zero as well.

In order to predict the random vector $\mathbf{A}_0$ from past values $\mathbf{A}_{-1}, \mathbf{A}_{-2}, \ldots$, we form an estimate $\overline{\mathbf{A}}_0$ of $\mathbf{A}_0$ such that the determinant of the prediction error covariance matrix Σ, given by

$$\Sigma = \mathsf{E}\left[(\mathbf{A}_0 - \overline{\mathbf{A}}_0)(\mathbf{A}_0 - \overline{\mathbf{A}}_0)^\dagger\right], \tag{3.11}$$

is minimized. In general, the estimate that minimizes the prediction error given the past values $\mathbf{a}_{-1}, \mathbf{a}_{-2}, \ldots$ is given by

$$\overline{\mathbf{a}}_0 = \mathsf{E}\left[\mathbf{A}_0 \mid \mathbf{A}_{-\infty}^{-1}\right]. \tag{3.12}$$

However, if the stochastic process $\{\mathbf{A}_k\}$ is Gaussian, then $\overline{\mathbf{A}}_0$ is of the form

$$\overline{\mathbf{A}}_0 = \sum_{k=-\infty}^{-1} \mathsf{C}_k \mathbf{A}_k, \tag{3.13}$$

where the matrix-valued parameters C_k have to be chosen such that the prediction error is minimized. Remember that we refer to this case as linear prediction.

Wiener and Masani showed that if one is restricted to a linear predictor and if the minimum prediction error covariance matrix Σ is nonsingular (*i.e.*, $\det \Sigma > 0$), then there exists a determinantal expression that connects Σ to the matrix-valued spectral density function $\mathsf{F}'(\lambda)$ [7]:

$$\det \Sigma = \exp\left\{\int_{-1/2}^{1/2} \log \det \mathsf{F}'(\lambda)\, \mathrm{d}\lambda\right\}. \tag{3.14}$$

Note, however, that the problem of expressing the covariance matrix Σ itself in terms of the spectral density function $\mathsf{F}'(\lambda)$ has not yet been solved.

As in the univariate case, we refer to processes with $\det \Sigma > 0$ as regular and to those for which $\det \Sigma = 0$ as non-regular or deterministic. Note that Wiener and Masani [7] use the term *non-deterministic* instead of regular.

As mentioned above the expression for the prediction error covariance matrix holds only if $\det \Sigma > 0$. In order to study non-regular processes (*i.e.*, $\det \Sigma = 0$), we shall extend the noisy prediction problem for univariate stochastic processes to the multivariate case.

Let $\{\mathbf{W}_k\}$ be a sequence of i.i.d. $\mathcal{N}\left(0, \delta^2 \mathsf{I}\right)$ random variables with I being the identity matrix. Furthermore, let the stochastic process $\{\mathbf{A}_k\}$ be Gaussian and independent from the process $\{\mathbf{W}_k\}$. The noisy prediction problem for multivariate processes is to predict $\mathbf{A}_0$ based on the observations $\mathbf{A}_{-1} + \mathbf{W}_{-1}, \mathbf{A}_{-2} + \mathbf{W}_{-2}, \ldots$. It follows that in this case the minimum prediction error covariance matrix Σ is connected to the spectral density function $\mathsf{F}'(\lambda)$ by the following expression:

$$\det\left(\Sigma + \delta^2 \mathsf{I}\right) = \exp\left\{\int_{-1/2}^{1/2} \log\det\left(\mathsf{F}'(\lambda) + \delta^2 \mathsf{I}\right)\,\mathrm{d}\lambda\right\}. \tag{3.15}$$

The derivation of this expression is analogous to that in the univariate case.

We shall often consider stochastic processes $\{\mathbf{A}_k\}$ where the entries $A_k^{(r)}$ are uncorrelated, *i.e.*,

$$\mathsf{E}\left[A_{k+m}^{(r)} A_k^{(t)*}\right] = 0 \qquad \text{for } r \neq t, \quad k, m \in \mathbb{Z}. \tag{3.16}$$

In this case we are able to express the prediction error covariance matrix Σ in terms of the density $\mathsf{F}'(\lambda)$. The result is stated in the following lemma.

Lemma 3.1 *Consider a multivariate zero-mean stationary stochastic process $\{\mathbf{A}_k\}$ with matrix-valued spectral distribution function $\mathsf{F}(\lambda)$. Furthermore, assume that the entries $A_k^{(r)}$ are uncorrelated,* i.e.*, (3.16) holds. Then, the prediction error covariance matrix Σ of the optimal linear predictor is diagonal. Moreover, the diagonal entries are given by*

$$\Sigma^{(r,r)} = \exp\left\{\int_{-1/2}^{1/2} \log \mathsf{F}'^{(r,r)}(\lambda)\,\mathrm{d}\lambda\right\}, \qquad 1 \leq r \leq n_{\mathrm{R}}, \tag{3.17}$$

where $\mathsf{F}'(\lambda)$ denotes the derivative of $\mathsf{F}(\lambda)$.

Proof: See Appendix A. □

Note that Lemma 3.1 is also of use for matrix-valued stochastic processes $\{\mathbb{A}_k\}$. Indeed, we can stack the components of $\mathbb{A}_k$ into one huge vector and consider the vector-valued case.

Chapter 4

The Channel

In this chapter we investigate the physical behavior of wireless links and derive the channel model that we will use in the following chapters.

The chapter is divided into two parts. In Section 4.1 we begin by describing the channel physically and show how this channel can be modeled mathematically. Section 4.2 shows the channel model that will be used in this thesis.

4.1 The Physical Description of the Channel

In this section, we study the transfer behavior of wireless transmission channels. We aim at deriving a mathematical channel model practical for the investigation of the capacity.

In contrast to wired transmission through copper or optical fibers, a wireless link is highly affected by its environment and the interference therefrom. The transmitted signal can be reflected by objects in the surrounding area (*e.g.*, buildings, mountains) or perturbed by atmospheric effects (*e.g.*, rain, snow, electromagnetic interference). Figure 4.1 shows a model of a wireless transmission channel. It illustrates the environmental influences represented by a mountain, a building and scatterers in the atmosphere. The arrows picture possible propagations of radiation referred to as *paths*. Note that transmitter and receiver are depicted by cars having in mind that both can be mobile.

The signal at the receiver is a superposition of signals corresponding to different paths. Depending on the environment, the paths will differ in length and occupy therefore different path delays. This affects the received

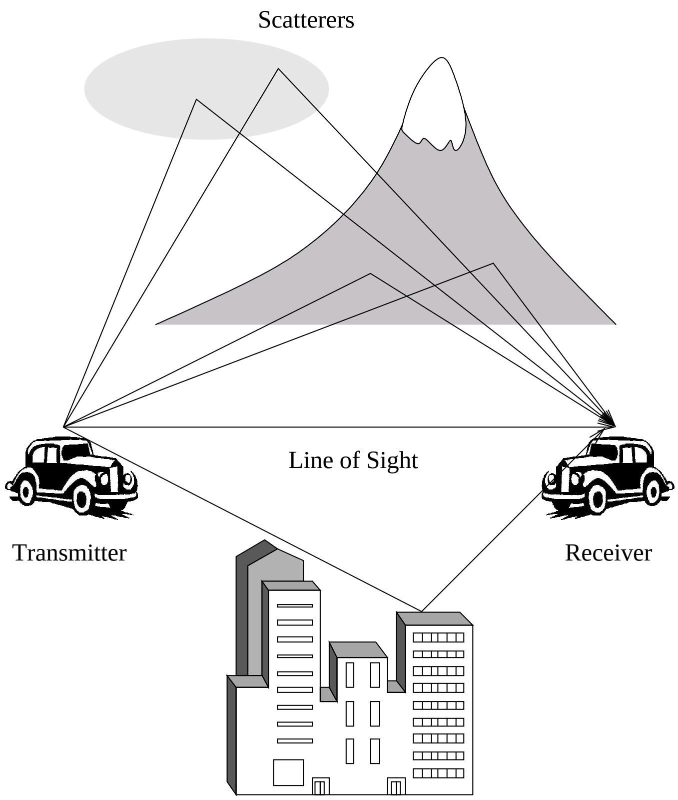

Figure 4.1: Physical channel model

signal in two ways: first of all, the signals arrive at different times which causes a temporal spread of the signal. Secondly, any path delay induces an appropriate phase-shift that leads to constructive and destructive interference[1] at the receiver.

Due to movements of transmitter, receiver, or the scatterers, the channel behavior may vary over time. This variation is called *fading*. It follows that a mathematical description of a fading channel must consider the time-dependency of the channel behavior. We may write the continuous-time signal $Y(t)$ at the receiver as

$$Y(t) = (h \star x)(t) + Z(t) = \int_{-\infty}^{\infty} h(t,\tau) x(t-\tau)\,\mathrm{d}\tau + Z(t), \qquad t \in \mathbb{R}, \quad (4.1)$$

where the zero-mean additive white Gaussian noise $Z(t)$ models the disturbance introduced by the receiver. Note that the impulse response $h(t,\tau)$ depends on the time index t.

So far, the channel impulse response $h(t,\tau)$ has been regarded as deterministic, since it can be determined for a given environment. However, the channel behavior depends on many parameters and, thus, a deterministic description of the channel is not feasible. Therefore, we resort to a stochastic description by trying to characterize the channel statistics. Then, $Y(t)$ is given by

$$Y(t) = \int_{-\infty}^{\infty} H(t,\tau) x(t-\tau)\,\mathrm{d}\tau + Z(t), \qquad t \in \mathbb{R} \quad (4.2)$$

with the random time-varying impulse response $H(t,\tau)$. We continue by replacing $H(t,\tau)$ and $x(t)$ with their Fourier transforms, *i.e.*,

$$L_H(t,f) = \int_{-\infty}^{\infty} H(t,\tau) e^{-i2\pi f\tau}\,\mathrm{d}\tau \quad (4.3)$$

and

$$\breve{x}(f) = \int_{-\infty}^{\infty} x(t) e^{-i2\pi ft}\,\mathrm{d}t, \qquad t, f \in \mathbb{R}, \quad (4.4)$$

where $L_H(t,f)$ is referred to as *time-varying transfer function* or *Weyl symbol*. It follows that

$$Y(t) = \int_{-\infty}^{\infty} L_H(t,f)\breve{x}(f) e^{i2\pi ft}\,\mathrm{d}f + Z(t), \qquad t \in \mathbb{R}. \quad (4.5)$$

[1] With interference we shall mean the superposition of several signals. If the signals differ in the phase by a multiple of 2π, then the interference is called constructive, if the phase difference can be written as $(2m+1)\pi$, $m \in \mathbb{Z}$, then the interference is referred to as destructive.

Under the condition that the bandwidth W of the signal $x(t)$ is sufficiently small we may assume that the transfer function $L_H(t,f)$ is constant on the interval $[-W,W]$, *i.e.*, $L_H(t,f) = L_H(t)$ for $f \in [-W,W]$, $t \in \mathbb{R}$. In this case the signal $Y(t)$ is given by

$$\begin{aligned} Y(t) &= \int_{-\infty}^{\infty} L_H(t,f)\check{x}(f)e^{i2\pi ft}\,\mathrm{d}f + Z(t) \\ &= L_H(t)\int_{-\infty}^{\infty} \check{x}(f)e^{i2\pi ft}\,\mathrm{d}f + Z(t) \\ &= H(t)x(t) + Z(t), \qquad t \in \mathbb{R}, \end{aligned} \tag{4.6}$$

where the last equality should be taken as a definition. We refer to this case as *flat fading* addressing the assumption of the transfer function being "flat" on the interval $[-W,W]$. It should be noted that the question whether the fading is flat or not depends on the signal-bandwidth as well as on the channel behavior.

For mathematical convenience people often use a discrete-time analog to (4.6), *i.e.*,

$$Y_k = H_k x_k + Z_k, \qquad k \in \mathbb{Z}. \tag{4.7}$$

Additionally, H_k is often chosen to be Gaussian distributed. This is justified by the assumption of having a large number of independent scatterers in the environment that affect H_k. The claim then follows from the central limit theorem.

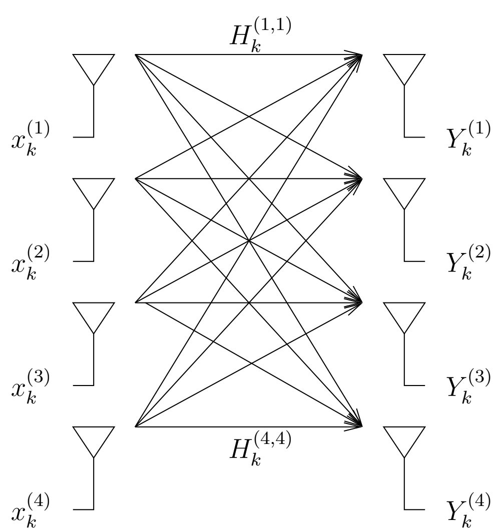

Figure 4.2: Multiple-input multiple-output channel

So far, we only considered a single-antenna system. However, the extension to multiple antennas is straightforward. Figure 4.2 shows a multiple-input multiple-output (MIMO) channel where any arrow depicts a single-input single-ouput (SISO) channel described by the channel model in (4.7). It follows that the time-k ouput $Y_k^{(r)} \in \mathbb{C}$ at the r-th receive antenna is given by

$$Y_k^{(r)} = \sum_{t=1}^{n_\mathrm{T}} H_k^{(r,t)} x_k^{(t)} + Z_k^{(r)}, \qquad 1 \leq r \leq n_\mathrm{R}, \tag{4.8}$$

where $H_k^{(r,t)}$ denotes the channel that connects the t-th transmit antenna with the r-th receive antenna. Using vectors and the fading matrix $\mathbb{H}_k$ we can write (4.8) as

$$\mathbf{Y}_k = \mathbb{H}_k \mathbf{x}_k + \mathbf{Z}_k. \tag{4.9}$$

In the following this channel model (4.9) will be used to describe the channel. The laws of $\mathbf{Y}_k$, $\mathbb{H}_k$ and $\mathbf{Z}_k$ are established in the next section.

4.2 The Channel Model

We consider a discrete-time MIMO channel whose time-k complex-valued output $\mathbf{Y}_k \in \mathbb{C}^{n_\mathrm{R}}$ is given by

$$\mathbf{Y}_k = \mathbb{H}_k \mathbf{x}_k + \mathbf{Z}_k, \tag{4.10}$$

where $\mathbf{x}_k \in \mathbb{C}^{n_\mathrm{T}}$ is the n_T-dimensional complex-valued input at time k; the complex process $\{\mathbb{H}_k\}$ with $\mathbb{H}_k \in \mathbb{C}^{n_\mathrm{R} \times n_\mathrm{T}}$, $k \in \mathbb{Z}$ models multiplicative noise; and the complex process $\{\mathbf{Z}_k\}$ models additive noise. The processes $\{\mathbb{H}_k\}$ and $\{\mathbf{Z}_k\}$ are assumed to be independent and of a joint law that does not depend on the input sequence $\{\mathbf{x}_k\}$.

We assume that $\{\mathbf{Z}_k\}$ is a sequence of independent and identically distributed (i.i.d.) circularly symmetric complex Gaussian random variables of zero-mean and variance σ^2, *i.e.*, $\mathbf{Z}_k \sim \mathcal{N}\left(0, \sigma^2 \mathsf{I}\right)$, where I denotes the identity matrix. The fading process $\{\mathbb{H}_k\}$ is assumed to be such that for a deterministic mean matrix $\mathsf{D} = \mathsf{E}[\mathbb{H}_k]$ the process $\{\mathbb{H}_k - \mathsf{D}\}$ is a zero-mean circularly symmetric stationary Gaussian process. Additionally, we assume that both $\{\mathbb{H}_k\}$ and $\{\mathbf{Z}_k\}$ have finite second moments, *i.e.*,

$$\mathsf{E}\left[\|\mathbb{H}_k\|^2\right], \mathsf{E}\left[\|\mathbf{Z}_k\|^2\right] < \infty, \qquad k \in \mathbb{Z}. \tag{4.11}$$

In this thesis we often consider channels with only one receive antenna. In this case, the fading can be described by a vector instead of a matrix and the channel output at time k is given by

$$Y_k = \mathbf{H}_k^{\mathsf{T}} \mathbf{x}_k + Z_k. \tag{4.12}$$

Note that we view channels with only one transmit antenna as a special case of the more general case with $n_{\mathrm{T}} \geq 1$.

The law of the fading process $\{\mathbf{H}_k\}$ can be described by the specular component $\mathbf{d} = \mathsf{E}[\mathbf{H}_k]$ and by the matrix-valued spectral distribution function $\mathsf{F}(\lambda)$, $-1/2 \leq \lambda \leq 1/2$. In general, $\mathsf{F}(\lambda)$ is such that

$$\mathsf{E}\Big[(\mathbf{H}_{k+m} - \mathbf{d})(\mathbf{H}_k - \mathbf{d})^{\dagger}\Big] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda}\,\mathrm{d}\mathsf{F}(\lambda) \tag{4.13}$$

and

$$\mathsf{E}\Big[(\mathbf{H}_k - \mathbf{d})(\mathbf{H}_k - \mathbf{d})^{\dagger}\Big] = \mathsf{K}, \qquad k, m \in \mathbb{Z} \tag{4.14}$$

with K being the covariance matrix of the fading process. In the following, we will assume that the spectral distribution function $\mathsf{F}(\lambda)$ is absolutely continuous and denote its derivative by $\mathsf{F}'(\lambda)$. Remember that if the channels are uncorrelated, *i.e.*, $\mathsf{E}\big[(\mathbf{H}_{k+m} - \mathbf{d})(\mathbf{H}_k - \mathbf{d})^{\dagger}\big]$ is diagonal, then the spectral density matrix $\mathsf{F}'(\lambda)$, $-1/2 \leq \lambda \leq 1/2$, is diagonal as well.

When considering MIMO channels, the covariance matrix has to be replaced by a tensor. Consequently, notation gets more laborious. However, if the entries in the fading matrix $\{\mathbb{H}_k\}$ are independent, *i.e.*,

$$\mathsf{E}\Big[(H_{k+m}^{(r,t)} - d^{(r,t)})(H_k^{(r',t')} - d^{(r',t')})^*\Big] = 0, \qquad \text{for } r \neq r' \text{ and } t \neq t', \quad k, m \in \mathbb{Z}, \tag{4.15}$$

then we shall define the matrix-valued spectral distribution function $\mathsf{F}(\lambda) \in \mathbb{R}^{n_{\mathrm{R}} \times n_{\mathrm{T}}}$ such that

$$\mathsf{E}\Big[(H_{k+m}^{(r,t)} - d^{(r,t)})(H_k^{(r,t)} - d^{(r,t)})^*\Big] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda}\,\mathrm{d}\mathsf{F}^{(r,t)}(\lambda), \qquad k, m \in \mathbb{Z}. \tag{4.16}$$

We will study both, regular and non-regular fading processes. In the case where the fading is regular, we consider an average-power constraint on the input, *i.e.*,

$$\frac{1}{n} \sum_{k=1}^{n} \mathsf{E}\big[\|\mathbf{X}_k\|^2\big] \leq \mathcal{E}_{\mathrm{s}}, \tag{4.17}$$

and we define the signal-to-noise ratio (SNR) by

$$\mathrm{SNR} = \frac{\mathcal{E}_{\mathrm{s}}}{\sigma^2}. \tag{4.18}$$

In the case of a non-regular fading process, we replace the average-power constraint by a peak-power constraint, *i.e.*,

$$\|\mathbf{x}_k\| \leq \mathsf{A}. \tag{4.19}$$

In this case, the SNR is defined by

$$\mathrm{SNR} = \frac{\mathsf{A}^2}{\sigma^2}. \tag{4.20}$$

Note that if the spectral density function is a diagonal matrix and its diagonal entries are constant on $[-1/2, 1/2]$, then $\{\mathbf{H}_k\}$ is a sequence of i.i.d. Gaussian random variables. Thus, our channel model includes as special cases the Rayleigh and Ricean channel models corresponding to zero-mean and non zero-mean i.i.d. fading.

Chapter 5

Results

This chapter gives an overview over all results obtained in this thesis. Additionally, we summarize previous work done by Lapidoth and Moser [1] and by Lapidoth [2], where we will focus on the results that are associated with our work.

The chapter is divided into two parts. The first part presents the results corresponding to fading channels with a regular fading process, *i.e.*, the present state of the channel cannot be estimated precisely from its past. Here, the emphasis is on the fading number, *i.e.*, the second order term in the high signal-to-noise ratio (SNR) expansion of the capacity, introduced by Lapidoth and Moser in [1]. After resuming some previous results we present an upper bound on the fading number of multiple-input single-output (MISO) Gaussian fading channels with memory. Moreover, we show that this bound is tight (*i.e.*, it coincides with a lower bound) in the case where the channels are uncorrelated and where the fading process is either zero-mean or its spectral density matrix contains identical entries. The derivations of these results can be found in Chapter 6.

In the second part, we consider fading channels with a non-regular fading process, *i.e.*, the present state of the channel can be estimated precisely from its past. Since the capacity of these channels can grow faster than double-logarithmically in the SNR, the fading number will be infinite in many cases and is thus not an appropriate performance measure anymore. Instead, we study the pre-log, *i.e.*, the limiting ratio of the capacity to the logarithm of the SNR. We present upper bounds on the pre-log of MISO and multiple-input multiple-output (MIMO) Gaussian fading channels with memory. In the case of MISO fading channels, we show that the upper bound is tight if the channels are uncorrelated. Additionally, we present a lower bound on the capacity of single-input single-output (SISO) channels that is tight when channel capacity only grows double-logarithmically in the SNR. The derivations of these results can be found in Chapter 7.

5.1 Regular Processes

In this section we study fading channels with a regular fading process, *i.e.*, the present state of the channel cannot be estimated precisely from its past. According to Section 4.2, we consider in this case an average-power constraint on the input (4.17) and it follows that the SNR is given by

$$\mathrm{SNR} = \frac{\mathcal{E}_\mathrm{s}}{\sigma^2}. \tag{5.1}$$

Lapidoth and Moser showed [1] that if the fading process is regular, channel capacity grows double-logarithmically in the SNR. Furthermore, they defined the *fading number* of a fading process $\{\mathbb{H}_k\}$ as

$$\chi(\{\mathbb{H}_k\}) = \varlimsup_{\mathcal{E}_\mathrm{s}\to\infty} \left\{ C(\mathcal{E}_\mathrm{s}) - \log\log\frac{\mathcal{E}_\mathrm{s}}{\sigma^2} \right\}, \tag{5.2}$$

where $C(\mathcal{E}_\mathrm{s})$ denotes the capacity of a channel under an average-power constraint $\mathcal{E}_\mathrm{s}$ on the input. The fading number has the following interpretation. The region where capacity only grows double-logarithmically in the SNR is very power-inefficient. So, in order to communicate power-efficiently, one should avoid this region and design the communication systems for lower rates. The fading number can be viewed as an indication of roughly how high the rate need be before one enters the double-logarithmic regime. In other words, at rates that are significantly higher than the fading number capacity grows only double-logarithmically in the SNR and communication is very power-inefficient.

In [1] several upper and lower bounds on the fading number are presented. We will show some of those in Section 5.1.1. In the subsequent section, we present an upper bound on the fading number of MISO Gaussian fading channels with memory. This upper bound is tight in the case where the channels are uncorrelated, *i.e.*, $\mathsf{E}\left[(\mathbf{H}_{k+m}-\mathbf{d})(\mathbf{H}_k-\mathbf{d})^\dagger\right]$ is diagonal, and where the fading process is either zero-mean or its spectral density matrix contains identical entries.

5.1.1 Previous Results

In the following we restate those results in [1] that we need to derive the upper bound presented in Section 5.1.2.

Lemma 5.1 *Consider a stationary fading process* $\{\mathbb{H}_k\}$ *with* n_{T} *transmit antennas and* n_{R} *receive antennas. Let* F *and* G *be nonsingular deterministic matrices of dimensions* $n_{\mathrm{T}} \times n_{\mathrm{T}}$ *and* $n_{\mathrm{R}} \times n_{\mathrm{R}}$*, respectively. Then*

$$\chi(\{\mathsf{G}\mathbb{H}_k\mathsf{F}\}) = \chi(\{\mathbb{H}_k\}). \tag{5.3}$$

Proof: See [1, Lemma 4.7]. □

The next corollary gives a formula for the fading number of memoryless MISO Gaussian fading channels. It is shown that the fading number is achievable by inputs of the form $X \cdot \hat{\mathbf{x}}$, where $\hat{\mathbf{x}}$ is a deterministic unit vector. We refer to this case as *beam forming*.

Corollary 5.2 *Consider a memoryless Gaussian MISO fading channel where the fading matrix is a row vector* $\mathbf{H}^{\mathsf{T}} \in \mathbb{C}^{n_{\mathrm{T}}}$*, where* $\mathbf{H} \sim \mathcal{N}(\mathbf{d}, \mathsf{K})$*,* $\det \mathsf{K} \neq 0$*. Then the fading number is given by*

$$\chi(\mathbf{H}^{\mathsf{T}}) = -1 + \log \mathbf{d}_*^2 - \mathrm{Ei}(-\mathbf{d}_*^2) \tag{5.4}$$

where

$$\mathbf{d}_* = \max_{\|\hat{\mathbf{x}}\|=1} \frac{|\mathsf{E}[\mathbf{H}^{\mathsf{T}}]\,\hat{\mathbf{x}}|}{\sqrt{\mathsf{Var}(\mathbf{H}^{\mathsf{T}}\hat{\mathbf{x}})}} \tag{5.5}$$

and $\mathrm{Ei}(-x)$ *denotes the exponential integral function defined as*

$$\mathrm{Ei}(-x) = -\int_x^\infty \frac{e^{-t}}{t}\,\mathrm{d}t, \qquad x > 0. \tag{5.6}$$

Proof: See [1, Corollary 4.28]. □

It should be noted that if the mean vector $\mathbf{d}$ is zero, then $\mathbf{d}_*$ is zero as well and the fading number is equal to $-1-\gamma$, where $\gamma \approx 0.577$ denotes Euler's constant. It is achievable by beam forming with an arbitrarily chosen direction.

Theorem 5.3 *Consider a MIMO fading channel*

$$\mathbf{Y}_k = \mathbb{H}_k\mathbf{x}_k + \mathbf{Z}_k \tag{5.7}$$

where the fading process $\{\mathbb{H}_k\}$ *is stationary and independent of the stationary additive noise process* $\{\mathbf{Z}_k\}$*. Assume further that* $\{\mathbf{Z}_k\}$ *is a sequence of independent and identically distributed (i.i.d.) random variables and the joint law* $(\{\mathbb{H}_k\}, \{\mathbf{Z}_k\})$ *does not depend on the input sequence* $\{\mathbf{x}_k\}$*. Then, the fading number* $\chi(\{\mathbb{H}_k\})$ *can be upper bounded by*

$$\begin{aligned}\chi(\{\mathbb{H}_k\}) &\leq \chi_{\text{i.i.d.}}(\mathbb{H}_1) + \lim_{n\to\infty} I(\mathbb{H}_n; \mathbb{H}^{n-1}) \\ &= \chi_{\text{i.i.d.}}(\mathbb{H}_0) + I(\mathbb{H}_0; \mathbb{H}_{-\infty}^{-1}),\end{aligned} \tag{5.8}$$

where $\chi_{\text{i.i.d.}}(\mathbb{H}_0)$ *denotes the fading number in the memoryless case with equal marginal.*

Proof: Follows directly from [1, Lemma 4.5]. □

For SISO systems, this bound is tight and we can give a formula for the fading number.

Corollary 5.4 *Consider a SISO fading process* $\{H_k\}$ *such that for some specular component* $d \in \mathbb{C}$ *the process* $\{H_k - d\}$ *is a zero-mean unit-variance circularly symmetric stationary complex Gaussian process whose spectrum is of continuous part* $\mathsf{F}'(\lambda)$*,* $-1/2 \leq \lambda \leq 1/2$*. Then*

$$\chi(\{H_k\}) = -1 + \log|d|^2 - \text{Ei}(-|d|^2) + \log\frac{1}{\epsilon^2_{\text{MSE}}} \tag{5.9}$$

where $\epsilon^2_{\text{MSE}} > 0$ *denotes the minimum mean squared error in predicting the present fading from its past*

$$\epsilon^2_{\text{MSE}} = \exp\left\{\int_{-1/2}^{1/2} \log \mathsf{F}'(\lambda)\,\mathrm{d}\lambda\right\}. \tag{5.10}$$

Proof: See [1, Corollary 4.42]. □

In the next section we give an improved upper bound on the fading number of MISO fading channels with memory.

5.1.2 The Fading Number of MISO Fading Channels

As stated before, the upper bound on the fading number given in Theorem 5.3 is only tight for SISO systems. Here, we present an improved upper bound on the fading number of MISO Gaussian fading channels with memory.

Theorem 5.5 *Consider a MISO Gaussian fading channel with fading process* $\{\mathbf{H}_k\}$ *such that for some specular component* $\mathbf{d} \in \mathbb{C}^{n_\mathrm{T}}$ *the process* $\{\mathbf{H}_k - \mathbf{d}\}$ *is a zero-mean circularly symmetric stationary Gaussian process with a matrix-valued spectral distribution function* $\mathsf{F}(\lambda)$ *such that*

$$\mathsf{E}\left[(\mathbf{H}_{k+m} - \mathbf{d})(\mathbf{H}_k - \mathbf{d})^\dagger\right] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda}\, \mathrm{d}\mathsf{F}(\lambda) \tag{5.11}$$

and

$$\det\left(\mathsf{E}\left[(\mathbf{H}_k - \mathbf{d})(\mathbf{H}_k - \mathbf{d})^\dagger\right]\right) \neq 0, \qquad k, m \in \mathbb{Z}. \tag{5.12}$$

Then, the fading number can be upper bounded by

$$\chi(\{\mathbf{H}_k^\mathsf{T}\}) \leq -1 + \log d_*^2 - \mathrm{Ei}(-d_*^2) + \log \frac{1}{\lambda_{\min}} \tag{5.13}$$

with

$$d_* = \max_{\|\hat{\mathbf{x}}\|=1} |\mathbf{d}^\mathsf{T}\hat{\mathbf{x}}| \tag{5.14}$$

and where $\lambda_{\min}$ *is the minimum eigenvalue of the prediction error covariance matrix* Σ *in predicting the present fading from its past defined as in (3.11).*

Proof: See Chapter 6. □

Note that since the fading process is regular, the present fading cannot be estimated precisely from its past and, therefore, $\lambda_{\min} > 0$. Furthermore, remember that we are not able to determine $\lambda_{\min}$ in the case where the channels are correlated due to the fact that in this case only a determinantal identity is known that connects the prediction error covariance matrix Σ to the spectral density $\mathsf{F}'(\lambda)$ (see Chapter 3).

In general, the upper bound given in Theorem 5.5 is not tight. However, we can show two special cases where this bound can be achieved. The corresponding fading numbers are presented in Corollaries 5.6 and 5.7.

Corollary 5.6 *Consider a MISO zero-mean circularly symmetric stationary Gaussian fading channel with a matrix-valued spectral distribution function* $\mathsf{F}(\lambda)$ *as in Theorem 5.5. Furthermore, let the channels be uncorrelated,* i.e., $\mathsf{F}'(\lambda)$ *is a diagonal matrix. Then*

$$\chi(\{\mathbf{H}_k^\mathsf{T}\}) = -1 - \gamma + \log \frac{1}{\epsilon_{\min}^2}, \tag{5.15}$$

where $\epsilon^2_{\min} > 0$ *is the minimum mean squared error in predicting that component of the present fading* $H_0^{(t)}$ *which leads to the smallest prediction error,* i.e.,

$$\epsilon^2_{\min} = \min_{1\le t\le n_\mathrm{T}} \exp\left\{\int_{-1/2}^{1/2} \log \mathsf{F}'^{(t,t)}(\lambda)\,\mathrm{d}\lambda\right\}, \tag{5.16}$$

Moreover, this fading number is achievable by transmitting from only one antenna, i.e., *the one that yields the smallest prediction error in predicting the fading from its past.*

Proof: See Chapter 6. □

Note that the fading number achieving strategy proposed in Corollary 5.6 can be viewed as a special case of beam forming (*i.e.*, the inputs are of the form $\mathbf{X}_k = \hat{\mathbf{x}}\widetilde{X}_k$), where the components of $\hat{\mathbf{x}}$ are given by

$$\hat{x}^{(t)} = \begin{cases} 1 & t = t_* \\ 0 & \text{otherwise} \end{cases} \tag{5.17}$$

and where t_* has to be chosen such that the minimum prediction error given in (5.16) is achieved.

Corollary 5.7 *Consider a MISO Gaussian fading channel with fading process* $\{\mathbf{H}_k\}$ *such that for some specular component* $\mathbf{d} \in \mathbb{C}^{n_\mathrm{T}}$ *the process* $\{\mathbf{H}_k - \mathbf{d}\}$ *is a zero-mean circularly symmetric stationary Gaussian process with a matrix-valued spectral distribution function* $\mathsf{F}(\lambda)$ *as in Theorem 5.5. Furthermore, assume that* $\mathsf{F}'(\lambda)$ *is a diagonal matrix with entries* $\mathsf{F}'^{(t,t)}(\lambda) = \mathsf{F}'(\lambda)$ *for* $1 \le t \le n_\mathrm{T}$*. Then*

$$\chi(\{\mathbf{H}_k^\mathsf{T}\}) = -1 + \log d_*^2 - \mathrm{Ei}(-d_*^2) + \log\frac{1}{\epsilon^2} \tag{5.18}$$

with

$$d_* = \max_{\|\hat{\mathbf{x}}\|=1} |\mathbf{d}^\mathsf{T}\hat{\mathbf{x}}| \tag{5.19}$$

and where $\epsilon^2 > 0$ *is the minimum mean squared error in predicting the present fading from its past,* i.e.,

$$\epsilon^2 = \exp\left\{\int_{-1/2}^{1/2} \log \mathsf{F}'(\lambda)\,\mathrm{d}\lambda\right\}. \tag{5.20}$$

Moreover, this fading number can be achieved by beam forming.

Proof: See Chapter 6. □

In general, the direction has to be chosen such that the resulting specular component $\mathbf{d}_*$ is maximized. However, if the specular component $\mathbf{d}$ is zero, then the fading number is equal to $-1-\gamma+\log\frac{1}{\epsilon^2}$ and can be achieved by beam forming with an arbitrarily chosen direction.

5.2 Non-Regular Processes

In this section we study fading channels with a non-regular fading process, *i.e.*, the present state of the channel can be predicted precisely from its past. According to Section 4.2, we consider in this case a peak-power constraint (4.19) instead of an average-power constraint and it follows that the SNR is given by

$$\text{SNR} = \frac{\mathsf{A}^2}{\sigma^2}. \tag{5.21}$$

The capacity of fading channels with a non-regular fading process can grow faster than double-logarithmically in the SNR. So, for most of these channels the fading number will be infinite and is thus not an appropriate performance measure anymore. Instead, we will consider the capacity pre-log in cases where capacity grows logarithmically, and the capacity pre-log-log in cases where capacity grows double-logarithmically in the SNR. The capacity pre-log Π is the limiting ratio of the capacity to the logarithm of the SNR, *i.e.*,

$$\Pi = \varlimsup_{\text{SNR}\to\infty} \frac{C(\text{SNR})}{\log \text{SNR}}, \tag{5.22}$$

where $C(\text{SNR})$ denotes capacity under a peak-power constraint on the input. Similarly, the pre-log-log Λ is defined as

$$\Lambda = \varlimsup_{\text{SNR}\to\infty} \frac{C(\text{SNR})}{\log\log \text{SNR}}. \tag{5.23}$$

Note that the capacity of a fading channel in the presence of perfect receiver side information, *i.e.*, when the receiver has perfect knowledge of the fading realization, is given by [8]

$$C_{\text{PSI}}(\text{SNR}) = \min\{n_{\text{T}}, n_{\text{R}}\} \cdot \log \text{SNR} + O(1), \tag{5.24}$$

where $O(1)$ is bounded by a constant and where n_{T} is the number of transmit and n_{R} the number of receive antennas. It follows by (5.24) and by noting that capacity in the absence of perfect receiver side information cannot be greater than capacity in its presence, that the pre-log can never be larger than $\min\{n_\mathrm{T}, n_\mathrm{R}\}$, *i.e.*, $\Pi \leq \min\{n_\mathrm{T}, n_\mathrm{R}\}$.

The capacity of SISO Gaussian fading channels where the fading process is non-regular was studied by Lapidoth in [2]. In the next section we show some of those results. Then, we present continuative results obtained in this thesis.

5.2.1 Previous Results

In this section we state some of the achievements in [2] needed to derive the results presented in Sections 5.2.2, 5.2.3 and 5.2.4.

Theorem 5.8 *Consider a SISO fading process $\{H_k\}$ such that for some specular component $d \in \mathbb{C}$ the process $\{H_k - d\}$ is a zero-mean unit-variance circularly symmetric stationary complex Gaussian process with spectral distribution function $\mathsf{F}(\lambda)$ and where the spectrum fulfills*

$$\mathsf{E}\left[(H_{k+m} - d)(H_k - d)^*\right] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda}\, \mathrm{d}\mathsf{F}(\lambda), \qquad k, m \in \mathbb{Z}. \tag{5.25}$$

Then, capacity can be upper bounded by

$$C(\mathrm{SNR}) \leq \log \frac{1}{\epsilon^2_{\mathrm{MSE}}(1/\mathrm{SNR})} + \log\log \mathrm{SNR} + O(1) \tag{5.26}$$

where $O(1)$ depends on d only and $\epsilon^2_{\mathrm{MSE}}(1/\mathrm{SNR})$ denotes the minimum mean squared error in predicting the present fading from a noisy observation of its past, i.e.,

$$\epsilon^2_{\mathrm{MSE}}(\delta^2) = \exp\left\{\int_{-1/2}^{1/2} \log(\mathsf{F}'(\lambda) + \delta^2)\, \mathrm{d}\lambda\right\} - \delta^2. \tag{5.27}$$

Additionally, capacity can be lower bounded by

$$C(\mathrm{SNR}) \geq \log \frac{1}{\epsilon^2_{\mathrm{MSE}}(4/\mathrm{SNR}) + \frac{2}{5} \cdot (4/\mathrm{SNR})} + O(1). \tag{5.28}$$

Proof: See [2]. □

Note that the prediction error $\epsilon^2_{\mathrm{MSE}}(1/\mathrm{SNR})$ goes to zero as the SNR tends to infinity and therefore capacity can increase faster than double-logarithmically with the SNR. Lapidoth derived an expression for the capacity pre-log when capacity grows logarithmically in the SNR.

Corollary 5.9 *Consider a SISO fading process* $\{H_k\}$ *as in Theorem 5.8. Then, the capacity pre-log* Π *is determined by the nulls of the spectral density*

$$\Pi = \mu\left(\left\{\lambda : \mathsf{F}'(\lambda) = 0\right\}\right), \tag{5.29}$$

where $\mu(\cdot)$ *denotes the Lebesgue measure on the interval* $[-1/2, 1/2]$.

Proof: See [2]. □

When capacity has a double-logarithmic increase with the SNR, *i.e.*,

$$\varlimsup_{\text{SNR}\to\infty} \frac{C(\text{SNR})}{\log\log \text{SNR}} < \infty, \tag{5.30}$$

then the lower bound (5.28) given in Theorem 5.8 is not tight. In the next section we present a tight lower bound on channel capacity for the case where (5.30) holds.

5.2.2 The Pre-Log-Log of SISO Fading Channels

In order to find an expression for the channel pre-log-log Λ, a tight lower bound on the capacity is required. The following lower bound is tight when channel capacity has a double-logarithmic increase with the SNR.

Theorem 5.10 *Consider a SISO fading process as in Theorem 5.8. Furthermore, assume that the capacity has a double-logarithmic increase with the SNR,* i.e.,

$$\varlimsup_{\text{SNR}\to\infty} \frac{C(\text{SNR})}{\log\log \text{SNR}} < \infty. \tag{5.31}$$

Then, channel capacity can be lower bounded by

$$C(\text{SNR}) \geq \log \left. \frac{1}{\epsilon^2_{\text{MSE}}(\delta^2) + \delta^2} \right|_{\delta^2 = \frac{\sigma^2}{\mathsf{A}^{2\alpha}}} + \log\log \text{SNR} + O(1), \tag{5.32}$$

where $0 < \alpha < 1$ *and where* $\epsilon^2_{\text{MSE}}(\delta^2)$ *is defined as in (5.27). The* $O(1)$ *term depends on* α *and on* d.

Proof: See Chapter 7. □

From Theorem 5.8 and 5.10 one can derive an expression for the channel pre-log-log.

Corollary 5.11 *Consider a SISO fading process $\{H_k\}$ as in Theorem 5.8. Furthermore, assume that channel capacity grows double-logarithmically in the SNR,* i.e.,

$$\overline{\lim_{\mathrm{SNR}\to\infty}} \frac{C(\mathrm{SNR})}{\log\log \mathrm{SNR}} < \infty. \tag{5.33}$$

Then, the capacity pre-log-log Λ is given by

$$\Lambda = 1 + \mathsf{K} \tag{5.34}$$

with

$$\mathsf{K} = \overline{\lim_{\delta^2 \downarrow 0}} \frac{-\int_{-1/2}^{1/2} \log\left(\mathsf{F}'(\lambda) + \delta^2\right)\, \mathrm{d}\lambda}{\log\log \frac{1}{\delta^2}}. \tag{5.35}$$

Proof: See Chapter 7. □

5.2.3 The Pre-Log of MISO Fading Channels

In this section we present an upper bound on the capacity of MISO fading channels. It can be shown that this upper bound is tight in the case where the channels are uncorrelated, *i.e.*, the covariance matrix $\mathsf{E}\left[(\mathbf{H}_{k+m} - \mathbf{d})(\mathbf{H}_k - \mathbf{d})^\dagger\right]$ is diagonal.

Theorem 5.12 *Consider a MISO fading process $\{\mathbf{H}_k\}$ such that for some specular component $\mathbf{d}$ the process $\{\mathbf{H}_k - \mathbf{d}\}$ is a zero-mean circularly symmetric stationary complex Gaussian process with matrix-valued spectral distribution function $\mathsf{F}(\lambda)$ such that*

$$\mathsf{E}\left[(\mathbf{H}_{k+m} - \mathbf{d})(\mathbf{H}_k - \mathbf{d})^\dagger\right] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda}\, \mathrm{d}\mathsf{F}(\lambda) \tag{5.36}$$

and

$$\det\left(\mathsf{E}\left[(\mathbf{H}_k - \mathbf{d})(\mathbf{H}_k - \mathbf{d})^\dagger\right]\right) \neq 0, \qquad k, m \in \mathbb{Z}. \tag{5.37}$$

Then, channel capacity can be upper bounded by

$$C(\mathrm{SNR}) \leq \log \frac{1}{\lambda_{\min}(1/\mathrm{SNR})} + \log\log \mathrm{SNR} + O(1), \tag{5.38}$$

where $\lambda_{\min}(1/\mathrm{SNR})$ denotes the minimum eigenvalue of the prediction error covariance matrix $\Sigma(1/\mathrm{SNR})$ in predicting the present fading vector from a noisy observation of its past.

Proof: See Chapter 7. □

Remember that when the channels are correlated, *i.e.*, $\mathsf{F}'(\lambda)$ is not a diagonal matrix, only a determinantal identity is known that connects the spectral density function $\mathsf{F}'(\lambda)$ to the prediction error covariance matrix $\Sigma(1/\mathrm{SNR})$ (see Chapter 3). However, if $\mathsf{F}'(\lambda)$ is diagonal , then $\Sigma(1/\mathrm{SNR})$ is diagonal as well and we can express the minimum eigenvalue $\lambda_{\min}(1/\mathrm{SNR})$ in terms of $\mathsf{F}'(\lambda)$. In this case, the pre-log is given as follows.

Corollary 5.13 *Consider a MISO fading process as in Theorem 5.12. Furthermore, assume that the channels are uncorrelated,* i.e.*,* $\mathsf{F}'(\lambda)$ *is a diagonal matrix. Then the pre-log* Π *is given by*

$$\Pi = \max_{1\le t\le n_{\mathrm{T}}} \mu\left(\left\{\mathsf{F}'^{(t,t)}(\lambda)=0\right\}\right). \tag{5.39}$$

Moreover, this pre-log is achievable by transmitting from only one antenna, i.e.*, the one that yields the smallest prediction error in predicting the present fading from its past.*

Proof: See Chapter 7. □

As commented in Section 5.1.2, transmitting from only one antenna can be viewed as a special case of beam forming.

5.2.4 The Pre-Log of MIMO Fading Channels

In the following we present an upper bound on the capacity of MIMO fading channels as well as an upper bound on the corresponding capacity pre-log. We can show that if the number of transmit antennas n_{T} is larger than the number of receive antennas n_{R}, this upper bound is at least as tight as the trivial upper bound $\Pi \le \min\{n_{\mathrm{T}}, n_{\mathrm{R}}\}$.

We assume that all entries in the $n_{\mathrm{R}} \times n_{\mathrm{T}}$ fading matrix $\mathbb{H}_k$ are independent, *i.e.*, for the deterministic $n_{\mathrm{R}} \times n_{\mathrm{T}}$ mean matrix D

$$\mathsf{E}\left[\left(H_{k+m}^{(r,t)} - d^{(r,t)}\right)\left(H_{k}^{(r',t')} - d^{(r',t')}\right)^{*}\right] = 0, \qquad \text{for } r \ne r' \text{ and } t \ne t', \quad k, m \in \mathbb{Z}. \tag{5.40}$$

We further assume that

$$\det\left(\mathsf{E}\left[(\mathbf{H}_k^{(r)} - \mathbf{d}^{(r)})(\mathbf{H}_k^{(r)} - \mathbf{d}^{(r)})^{\dagger}\right]\right) \ne 0, \quad 1 \le r \le n_{\mathrm{R}}, \quad k \in \mathbb{Z}, \tag{5.41}$$

where $\mathbf{H}_k^{(r)}$ and $\mathbf{d}^{(r)}$ denote the r-th row of the fading matrix $\mathbb{H}_k$ and the mean matrix D, respectively.

To simplify notation we define the matrix-valued spectral distribution function $\mathsf{F}(\lambda) \in \mathbb{R}^{n_\mathrm{R} \times n_\mathrm{T}}$ such that

$$\mathsf{E}\left[\left(H_{k+m}^{(r,t)} - d^{(r,t)}\right)\left(H_k^{(r,t)} - d^{(r,t)}\right)^*\right] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda}\,\mathrm{d}\mathsf{F}^{(r,t)}(\lambda), \quad k, m \in \mathbb{Z}. \tag{5.42}$$

Theorem 5.14 *Consider a MIMO fading process $\{\mathbb{H}_k\}$ such that for some $n_\mathrm{R} \times n_\mathrm{T}$ mean matrix D the process $\{\mathbb{H}_k - \mathsf{D}\}$ is a zero-mean stationary circularly symmetric complex Gaussian process with a matrix-valued spectral distribution function $\mathsf{F}(\lambda)$ fulfilling (5.42). Furthermore, assume that (5.41) and (5.40) hold. Then, capacity can be upper bounded by*

$$C(\mathrm{SNR}) \leq \sup_{\|\hat{\mathbf{x}}\|=1} \left\{\sum_{r=1}^{n_\mathrm{R}} \log \frac{1}{\hat{\mathbf{x}}^\mathsf{T}\Sigma_r(1/\mathrm{SNR})\hat{\mathbf{x}}^* + 1/\mathrm{SNR}}\right\} + \log\log\mathrm{SNR} + O(1), \tag{5.43}$$

where $\Sigma_r(1/\mathrm{SNR})$ is a diagonal $n_\mathrm{T} \times n_\mathrm{T}$ matrix with entries

$$\Sigma_r^{(t,t)}(\delta^2) = \exp\left\{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(r,t)}(\lambda) + \delta^2)\,\mathrm{d}\lambda\right\} - \delta^2. \tag{5.44}$$

Proof: See Chapter 7. □

With the aid of Theorem 5.14 we can upper bound the capacity pre-log Π.

Corollary 5.15 *Consider a MIMO fading process as in Theorem 5.14. Then, the capacity pre-log Π can be upper bounded by*

$$\Pi \leq \max_{1\leq t\leq n_\mathrm{T}} \sum_{r=1}^{n_\mathrm{R}} \mu\left(\left\{\lambda : \mathsf{F}'^{(r,t)}(\lambda) = 0\right\}\right). \tag{5.45}$$

Proof: See Chapter 7. □

Note that $\mu\left(\left\{\lambda : \mathsf{F}'^{(r,t)}(\lambda) = 0\right\}\right)$ can never be larger than 1 and, therefore, the pre-log Π is always bounded by n_R. So, if the number of transmit antennas n_T is larger than the number of receive antennas n_R, then

$$\Pi \leq \max_{1\leq t\leq n_\mathrm{T}} \sum_{r=1}^{n_\mathrm{R}} \mu\left(\left\{\lambda : \mathsf{F}'^{(r,t)}(\lambda) = 0\right\}\right) \leq n_\mathrm{R} = \min\{n_\mathrm{T}, n_\mathrm{R}\}, \tag{5.46}$$

and the upper bound given in Corollary 5.15 is at least as tight as the trivial upper bound $\Pi \leq \min\{n_{\mathrm{T}}, n_{\mathrm{R}}\}$.

Chapter 6

Regular Processes

In this chapter, we show the derivations of the results presented in Section 5.1, *i.e.*, where the fading process is regular. We consider channel capacity under an average-power constraint $\mathcal{E}_\text{s}$ on the inputs. The emphasis is on the high signal-to-noise ratio (SNR) regime, where the SNR is defined as $\mathcal{E}_\text{s}/\sigma^2$. In particular, bounds on the fading number are derived.

In Section 6.1 we derive an upper bound on the fading number of multiple-input single-output (MISO) fading channels. The result was stated in Theorem 5.5. Then, we present special cases where this bound is tight. In Section 6.2 an expression for the fading number of MISO zero-mean fading channels is deduced. This result was presented in Corollary 5.6. In Section 6.3 we derive an expression for the fading number, when the spectral density matrix of the fading process contains identical entries. This result was presented in Corollary 5.7.

6.1 A Proof of Theorem 5.5

We consider a MISO fading channel with a regular fading process. In the following we derive an upper bound on the corresponding fading number.

To upper bound the fading number we begin by noting that the channel output at time k is given by

$$Y_k = \mathbf{H}_k^\mathsf{T}\mathbf{x}_k + Z_k, \tag{6.1}$$

where $\mathbf{H}_k$, $\mathbf{x}_k$ and Z_k are as in Section 4.2. The fading vector $\{\mathbf{H}_k\}$ is a Gaussian regular process with mean $\mathbf{d}$ and covariance matrix

$$\mathsf{K} = \mathsf{E}\Big[(\mathbf{H}_k - \mathbf{d})(\mathbf{H}_k - \mathbf{d})^\dagger\Big], \qquad k \in \mathbb{Z} \tag{6.2}$$

with $\det \mathsf{K} \neq 0$. From Lemma 5.1 we know that the fading number is unchanged under multiplication of $\mathbf{H}_k$ by a nonsingular deterministic matrix. So, we might as well consider the case where the covariance matrix K is the identity I. This will simplify our derivations.

In order to find an upper bound on the fading number, we first upper bound channel capacity which is defined as

$$C(\mathrm{SNR}) = \lim_{n\to\infty} \frac{1}{n} \sup_{p_{\mathbf{X}^n}} I(\mathbf{X}^n; Y^n), \tag{6.3}$$

where the supremum is taken over all input distributions fulfilling the average-power constraint (4.17). Using the chain rule [5] we can write

$$I(\mathbf{X}^n; Y^n) = \sum_{k=1}^{n} I(Y_k; \mathbf{X}^n | Y^{k-1}). \tag{6.4}$$

We now proceed by upper bounding each of the terms in the sum as follows

$$\begin{aligned}
I(Y_k; \mathbf{X}^n | Y^{k-1}) &= I(Y_k; \mathbf{X}^n, Y^{k-1}) - I(Y_k; Y^{k-1}) \\
&\leq I(Y_k; \mathbf{X}^n, Y^{k-1}) \\
&= I(Y_k; \mathbf{X}^k, Y^{k-1}) \\
&= I(Y_k; \mathbf{X}_k) + I(Y_k; \mathbf{X}^{k-1}, Y^{k-1} | \mathbf{X}_k) \\
&\leq I(Y_k; \mathbf{X}_k) + I(Y_k; \mathbf{X}^{k-1}, Y^{k-1}, \mathbf{H}^{k-1} | \mathbf{X}_k) \\
&= I(Y_k; \mathbf{X}_k) + I(Y_k; \mathbf{H}^{k-1} | \mathbf{X}_k) \\
&\leq I(Y_0; \mathbf{X}_0) + I(Y_0; \mathbf{H}_{-\infty}^{-1} | \mathbf{X}_0),
\end{aligned} \tag{6.5}$$

where the first equality follows from the chain rule; the subsequent inequality from the non-negativity of mutual information; the next equality from the absence of feedback, which results in future inputs being independent of the present output given the present input and the past inputs and outputs. The following equality follows from the chain rule. In the next inequality we used the fact that adding information cannot reduce mutual information; the subsequent equality follows because the present output is independent of the past inputs and outputs given the present input and the past channel realizations; and the last inequality from stationarity and from adding information.

We now consider the maximization in (6.3) under an average-power constraint on the input sequence:

$$
\begin{aligned}
\frac{1}{n}\sup_{p_{\mathbf{X}^n}} I(\mathbf{X}^n;Y^n) &= \sup_{p_{\mathbf{X}^n}} \frac{1}{n}\sum_{k=1}^{n} I(Y_k;\mathbf{X}^n|Y^{k-1}) \\
&\le \frac{1}{n}\sum_{k=1}^{n}\sup_{p_{\mathbf{X}^n}} I(Y_k;\mathbf{X}^n|Y^{k-1}) \\
&= \sup_{p_{\mathbf{X}_0}} \left\{ I(Y_0;\mathbf{X}_0) + I(Y_0;\mathbf{H}_{-\infty}^{-1}|\mathbf{X}_0)\right\} \\
&\le \sup_{p_{\mathbf{X}_0}} I(Y_0;\mathbf{X}_0) + \sup_{p_{\mathbf{X}_0}} I(Y_0;\mathbf{H}_{-\infty}^{-1}|\mathbf{X}_0) \qquad (6.6)
\end{aligned}
$$

which follows by (6.5) and from splitting up the supremum.

We study the two terms on the RHS of (6.6) separately. The first term corresponds to the case of memoryless fading:

$$
\begin{aligned}
\sup_{p_{\mathbf{X}_0}} I(Y_0;\mathbf{X}_0) &= \log\log \mathrm{SNR} + \chi_{\mathrm{i.i.d}}(\mathbf{H}_0^{\mathsf{T}}) + o(1) \\
&= -1 + \log d_*^2 - \mathrm{Ei}(-d_*^2) + \log\log \mathrm{SNR} + o(1) \qquad (6.7)
\end{aligned}
$$

with

$$
d_* = \max_{\|\hat{\mathbf{x}}\|=1} \frac{|\mathsf{E}[\mathbf{H}_0^{\mathsf{T}}]\,\hat{\mathbf{x}}|}{\sqrt{\mathsf{Var}(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}})}}, \qquad (6.8)
$$

where we used the expression of the fading number (5.4) given in Corollary 5.2. The denominator in (6.8) can be simplified by choosing the covariance matrix K to be the identity matrix I. In this case,

$$
\mathsf{Var}(\mathbf{H}_0^{\mathsf{T}}\hat{\mathbf{x}}) = 1 \qquad (6.9)
$$

and the specular component d_* is given by

$$
d_* = \max_{\|\hat{\mathbf{x}}\|=1} |\mathbf{d}^{\mathsf{T}}\hat{\mathbf{x}}|, \qquad (6.10)
$$

with $\mathbf{d} = \mathsf{E}[\mathbf{H}_0]$.

To study the second term on the RHS of (6.6), we write $I(Y_0;\mathbf{H}_{-\infty}^{-1}|\mathbf{X}_0)$ as

$$
I(Y_0;\mathbf{H}_{-\infty}^{-1}|\mathbf{X}_0) = h(Y_0|\mathbf{X}_0) - h(Y_0|\mathbf{X}_0,\mathbf{H}_{-\infty}^{-1}) \qquad (6.11)
$$

and note that $(Y_0|\mathbf{X}_0=\mathbf{x}_0) \sim \mathcal{N}\left(\mathbf{d}^{\mathsf{T}}\mathbf{x}_0, \mathbf{x}_0^{\mathsf{T}}\mathsf{K}\mathbf{x}_0^* + \sigma^2\right)$. It follows that

$$
h(Y_0|\mathbf{X}_0=\mathbf{x}_0) = \log\pi + 1 + \log\left(\mathbf{x}_0^{\mathsf{T}}\mathsf{K}\mathbf{x}_0^* + \sigma^2\right), \qquad (6.12)
$$

where we used the expression for the differential entropy of a Gaussian random variable [5].

To compute the second term on the RHS of (6.11), we express the fading $\mathbf{H}_0$ as

$$\mathbf{H}_0 = \overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0, \tag{6.13}$$

where $\overline{\mathbf{H}}_0$ is the best estimate of the fading $\mathbf{H}_0$ given the past channel realizations. Thus, for given past values $\mathbf{h}_{-1}, \mathbf{h}_{-2}, \ldots$ the estimate $\overline{\mathbf{h}}_0$ is given by

$$\overline{\mathbf{h}}_0 = \mathsf{E}\left[\mathbf{H}_0 \;\middle|\; \mathbf{H}_{-\infty}^{-1}\right]. \tag{6.14}$$

Furthermore, we note that $\widetilde{\mathbf{H}}_0 \sim \mathcal{N}\;(0, \Sigma)$ where $\Sigma = \mathsf{E}\left[\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^\dagger\right]$ denotes the prediction error covariance matrix in predicting the present fading from its past. It follows that

$$\begin{aligned} h(Y_0|\mathbf{X}_0 = \mathbf{x}_0, \mathbf{H}_{-\infty}^{-1}) &= h((\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0)^\mathsf{T}\mathbf{X}_0 + Z_0|\mathbf{X}_0 = \mathbf{x}_0, \overline{\mathbf{H}}_0) \\ &= h(\widetilde{\mathbf{H}}_0^\mathsf{T}\mathbf{X}_0 + Z_0|\mathbf{X}_0 = \mathbf{x}_0) \\ &= \log \pi + 1 + \log\left(\mathbf{x}_0^\mathsf{T}\Sigma\mathbf{x}_0^* + \sigma^2\right), \end{aligned} \tag{6.15}$$

where the second equality follows from the fact that differential entropy is invariant under deterministic translation and the last equality from the expression for the differential entropy of a Gaussian random variable.

Combining (6.15), (6.12), and (6.11) and averaging over all realizations of $\mathbf{X}_0$ we get

$$\begin{aligned} I(Y_0; \mathbf{H}_{-\infty}^{-1}|\mathbf{X}_0) &= \mathsf{E}\left[\log \frac{\mathbf{X}_0^\mathsf{T}\mathsf{K}\mathbf{X}_0^* + \sigma^2}{\mathbf{X}_0^\mathsf{T}\Sigma\mathbf{X}_0^* + \sigma^2}\right] \\ &\leq \sup_{\mathbf{x}_0} \log \frac{\mathbf{x}_0^\mathsf{T}\mathsf{K}\mathbf{x}_0^* + \sigma^2}{\mathbf{x}_0^\mathsf{T}\Sigma\mathbf{x}_0^* + \sigma^2} \\ &= \sup_{\mathbf{x}_0} \log \frac{\hat{\mathbf{x}}_0^\mathsf{T}\mathsf{K}\hat{\mathbf{x}}_0^*\|\mathbf{x}_0\|^2 + \sigma^2}{\hat{\mathbf{x}}_0^\mathsf{T}\Sigma\hat{\mathbf{x}}_0^*\|\mathbf{x}_0\|^2 + \sigma^2} \\ &= \sup_{\|\hat{\mathbf{x}}_0\|=1} \log \frac{\hat{\mathbf{x}}_0^\mathsf{T}\mathsf{K}\hat{\mathbf{x}}_0^*}{\hat{\mathbf{x}}_0^\mathsf{T}\Sigma\hat{\mathbf{x}}_0^*} \\ &= \sup_{\|\hat{\mathbf{x}}_0\|=1} \log \frac{1}{\hat{\mathbf{x}}_0^\mathsf{T}\Sigma\hat{\mathbf{x}}_0^*}, \end{aligned} \tag{6.16}$$

where the inequality follows from the fact that the supremum is always larger than the expectation; the subsequent equality by expressing the input vector $\mathbf{x}_0$ in terms of its magnitude $\|\mathbf{x}_0\|$ and its direction $\hat{\mathbf{x}}_0 = \frac{\mathbf{x}_0}{\|\mathbf{x}_0\|}$. The next equality follows by noting that the minimum prediction error in predicting the fading $H_0^{(t)}$ from its past cannot be greater than its variance $\mathsf{E}\left[|H_0^{(t)} - d^{(t)}|^2\right]$. Thus, the entries in the prediction error covariance matrix

Σ are always smaller than the corresponding entries in K and, consequently, $\hat{\mathbf{x}}_0^\mathsf{T}\Sigma\hat{\mathbf{x}}_0^* \leq \hat{\mathbf{x}}_0^\mathsf{T}\mathsf{K}\hat{\mathbf{x}}_0^*$ in which case the mutual information is monotonically increasing in $\|\mathbf{x}_0\|^2$. The equality follows then by letting $\|\mathbf{x}_0\|^2$ go to infinity. Finally, the last equality follows by replacing K with I.

To maximize the RHS in (6.16) we have to minimize $\hat{\mathbf{x}}_0^\mathsf{T}\Sigma\hat{\mathbf{x}}_0^*$ over all unit vectors $\hat{\mathbf{x}}_0$. It follows then by the Rayleigh-Ritz Theorem [9, Theorem 4.2.2] that

$$\min_{\|\hat{\mathbf{x}}\|=1} \hat{\mathbf{x}}^\mathsf{T}\Sigma\hat{\mathbf{x}}^* = \lambda_{\min}, \tag{6.17}$$

where $\lambda_{\min}$ is the minimum eigenvalue of the prediction error covariance matrix Σ. Thus, $I(Y_0; \mathbf{H}_{-\infty}^{-1}|\mathbf{X}_0)$ can be upper bounded by

$$I(Y_0; \mathbf{H}_{-\infty}^{-1}|\mathbf{X}_0) \leq \log\frac{1}{\lambda_{\min}}. \tag{6.18}$$

Combining (6.18), (6.7), and (6.6) and using the expressions for channel capacity and the fading number, we obtain the following upper bound on the fading number:

$$\chi(\{\mathbf{H}_k^\mathsf{T}\}) \leq -1 + \log d_*^2 - \mathrm{Ei}(-d_*^2) + \log\frac{1}{\lambda_{\min}} \tag{6.19}$$

with

$$d_* = \max_{\|\hat{\mathbf{x}}\|=1} |\mathbf{d}^\mathsf{T}\hat{\mathbf{x}}|. \tag{6.20}$$

This concludes the proof.

Note that if the channels are correlated, we are not able to express the minimum eigenvalue $\lambda_{\min}$ in terms of the spectral distribution function $\mathsf{F}(\lambda)$, since in this case there exists only a determinantal identity that connects $\lambda_{\min}$ with the spectrum.

When the channels are uncorrelated and the fading process is either zero-mean or its spectral density matrix contains identical entries, then we can find a lower bound on the fading number that is tight, *i.e.*, it coincides with the upper bound (6.19). The corresponding fading numbers are given in Corollaries 5.6 and 5.7, which are proven in the next two sections.

6.2 A Proof of Corollary 5.6

In this section we give an expression for the fading number of MISO zero-mean fading channels, where the channels are uncorrelated, *i.e.*, $\mathsf{E}\left[\mathbf{H}_{k+m}\mathbf{H}_k^\dagger\right]$ is a diagonal matrix.

We first use (6.19) to derive an upper bound on the fading number. The fading is zero-mean, thus $d_* = 0$ and we have [1]

$$\log d_*^2 - \mathrm{Ei}(-d_*^2) = -\gamma, \tag{6.21}$$

where $\gamma \approx 0.577$ denotes Euler's constant. Thus, the fading number can be upper bounded by

$$\chi(\{\mathbf{H}_k^{\mathsf{T}}\}) \leq -1 - \gamma + \log \frac{1}{\lambda_{\min}}. \tag{6.22}$$

To express $\lambda_{\min}$ in terms of the spectral distribution function $\mathsf{F}(\lambda)$, we begin by noting that if the channels are uncorrelated, then due to Lemma 3.1 the prediction error covariance matrix Σ is diagonal and the diagonal entries are given by

$$\Sigma^{(t,t)} = \exp\left\{ \int_{-1/2}^{1/2} \log \mathsf{F}'^{(t,t)}(\lambda)\,\mathrm{d}\lambda \right\}, \qquad 1 \leq t \leq n_{\mathrm{T}}. \tag{6.23}$$

It follows that the minimum eigenvalue $\lambda_{\min}$ is just the smallest entry in Σ, *i.e.*,

$$\lambda_{\min} = \epsilon_{\min}^2 = \min_{1 \leq t \leq n_{\mathrm{T}}} \exp\left\{ \int_{-1/2}^{1/2} \log \mathsf{F}'^{(t,t)}(\lambda)\,\mathrm{d}\lambda \right\}. \tag{6.24}$$

Combining (6.24) and (6.22) we get the following upper bound:

$$\chi(\{\mathbf{H}_k^{\mathsf{T}}\}) \leq -1 - \gamma + \log \frac{1}{\epsilon_{\min}^2}. \tag{6.25}$$

In the following we show that this fading number can be achieved by transmitting from only one antenna. We note that transmitting from one antenna is equivalent to choosing the inputs to be of the form $\mathbf{X}_k = \hat{\mathbf{x}} \cdot \widetilde{X}_k$, where $\hat{\mathbf{x}}$ is a deterministic unit vector with components

$$x^{(t)} = \begin{cases} 1 & t = t_* \\ 0 & \text{otherwise} \end{cases}, \tag{6.26}$$

for a given t_*. We notice that this restriction on the inputs yields a lower bound on the fading number. In the following we show that the lower bound is tight, *i.e.*, it coincides with the upper bound (6.25).

We use the expression for the fading number of SISO fading channels with memory (5.9) given in Corollary 5.4 and lower bound the fading number by

$$\chi(\{\mathbf{H}_k^{\mathsf{T}}\}) \geq -1 - \gamma + \log \frac{1}{\epsilon_{t_*}^2} \tag{6.27}$$

with

$$\epsilon_{t_*}^2 = \exp\left\{\int_{-1/2}^{1/2} \log \mathsf{F}'^{(t_*,t_*)}(\lambda)\,\mathrm{d}\lambda\right\}. \tag{6.28}$$

By maximizing (6.27) over all possible choices of t_*, we get

$$\chi(\{\mathbf{H}_k^{\mathsf{T}}\}) \geq -1 - \gamma + \log \frac{1}{\epsilon_{\min}^2} \tag{6.29}$$

with

$$\epsilon_{\min}^2 = \min_{1\leq t\leq n_{\mathrm{T}}} \exp\left\{\int_{-1/2}^{1/2} \log \mathsf{F}'^{(t,t)}(\lambda)\,\mathrm{d}\lambda\right\} \tag{6.30}$$

which coincides with the upper bound given in (6.25).

6.3 A Proof of Corollary 5.7

Another special case, where the upper bound on the fading number given in Theorem 5.5 is tight, is the case where the channels are uncorrelated and the spectral density matrix of the fading process contains identical entries, *i.e.*, $\mathsf{F}'(\lambda)$ is a diagonal matrix with entries

$$\mathsf{F}'^{(t,t)}(\lambda) = \mathsf{F}'(\lambda), \qquad \text{for } 1 \leq t \leq n_{\mathrm{T}}. \tag{6.31}$$

To upper bound the fading number, we use the general upper bound (6.19):

$$\chi(\{\mathbf{H}_k^{\mathsf{T}}\}) \leq -1 + \log d_*^2 - \mathrm{Ei}(-d_*^2) + \log \frac{1}{\lambda_{\min}} \tag{6.32}$$

with

$$d_* = \max_{\|\hat{\mathbf{x}}\|=1} |\mathbf{d}^{\mathsf{T}}\hat{\mathbf{x}}|. \tag{6.33}$$

In our case, $\lambda_{\min}$ can be expressed as

$$\begin{aligned}\lambda_{\min} &= \min_{1\leq t\leq n_{\mathrm{T}}} \exp\left\{\int_{-1/2}^{1/2} \log \mathsf{F}'^{(t,t)}(\lambda)\,\mathrm{d}\lambda\right\} \\ &= \exp\left\{\int_{-1/2}^{1/2} \log \mathsf{F}'(\lambda)\,\mathrm{d}\lambda\right\},\end{aligned} \tag{6.34}$$

where the first equality follows from the fact that the channels are uncorrelated and by Lemma 3.1, and the second equality from (6.31).

In the following we derive a lower bound on the fading number and show that this bound is tight, *i.e.*, it coincides with the upper bound (6.32).

We obtain a lower bound by restricting ourselves to inputs of the form $\mathbf{X}_k = \hat{\mathbf{x}} \cdot \widetilde{X}_k$ where the deterministic unit vector $\hat{\mathbf{x}}$ is chosen such that the maximum in (6.33) is achieved. We express the channel output at time k as

$$\begin{aligned} Y_k &= \mathbf{H}_k^\mathsf{T} \hat{\mathbf{x}} \widetilde{X}_k + Z_k \\ &= (\mathbf{H}_k^\mathsf{T} \hat{\mathbf{x}}) \widetilde{X}_k + Z_k \\ &= \widetilde{H}_k \widetilde{X}_k + Z_k, \end{aligned} \tag{6.35}$$

where the last equality should be taken as a definition.

In order to determine the spectral distribution function $\widetilde{\mathsf{F}}(\lambda)$ of the fading process $\{\widetilde{H}_k\}$ we begin by noting that

$$\widetilde{d} = \mathsf{E}\left[\widetilde{H}_k\right] = \mathbf{d}^\mathsf{T} \hat{\mathbf{x}}. \tag{6.36}$$

Then, it follows

$$\begin{aligned} \int_{-1/2}^{1/2} e^{i2\pi m\lambda} \widetilde{\mathsf{F}}'(\lambda)\,\mathrm{d}\lambda &= \mathsf{E}\left[(\widetilde{H}_{k+m} - \widetilde{d})(\widetilde{H}_k - \widetilde{d})^*\right] \\ &= \mathsf{E}\left[(\hat{\mathbf{x}}^\mathsf{T}\mathbf{H}_{k+m} - \hat{\mathbf{x}}^\mathsf{T}\mathbf{d})(\hat{\mathbf{x}}^\mathsf{T}\mathbf{H}_k - \hat{\mathbf{x}}^\mathsf{T}\mathbf{d})^*\right] \\ &= \hat{\mathbf{x}}^\mathsf{T} \mathsf{E}\left[(\mathbf{H}_{k+m} - \mathbf{d})(\mathbf{H}_k - \mathbf{d})^\dagger\right] \hat{\mathbf{x}}^* \\ &= \hat{\mathbf{x}}^\mathsf{T} \left(\int_{-1/2}^{1/2} e^{i2\pi m\lambda} \mathsf{F}'(\lambda)\,\mathrm{d}\lambda \right) \hat{\mathbf{x}}^* \\ &= \int_{-1/2}^{1/2} e^{i2\pi m\lambda} \hat{\mathbf{x}}^\mathsf{T} \mathsf{F}'(\lambda) \hat{\mathbf{x}}^*\,\mathrm{d}\lambda, \end{aligned} \tag{6.37}$$

where the first equality follows from the definition of the spectral distribution function $\widetilde{\mathsf{F}}(\lambda)$; the second equality from (6.35) and (6.36); the subsequent equality from the fact, that $\hat{\mathbf{x}}$ is deterministic; the next inequality from the definition of the spectral distribution function $\mathsf{F}(\lambda)$; and the last equality from the linearity of the integral.

From (6.37) we obtain an expression for the spectrum $\widetilde{\mathsf{F}}(\lambda)$:

$$\widetilde{\mathsf{F}}'(\lambda) = \hat{\mathbf{x}}^\mathsf{T} \mathsf{F}'(\lambda) \hat{\mathbf{x}}^*. \tag{6.38}$$

The channels are uncorrelated, *i.e.*, the covariance matrix $\mathsf{E}\left[(\mathbf{H}_{k+m} - \mathbf{d})(\mathbf{H}_k - \mathbf{d})^\dagger\right]$ is diagonal and, consequently, $\mathsf{F}'(\lambda)$ is diagonal as well. It follows that

$$\begin{aligned}\widetilde{\mathsf{F}}'(\lambda) &= \sum_{t=1}^{n_\mathrm{T}} |\hat{x}^{(t)}|^2 \mathsf{F}'^{(t,t)}(\lambda) \\ &= \mathsf{F}'(\lambda) \sum_{t=1}^{n_\mathrm{T}} |\hat{x}^{(t)}|^2 \\ &= \mathsf{F}'(\lambda), \end{aligned} \tag{6.39}$$

where the first equality follows from the fact that $\mathsf{F}'(\lambda)$ is a diagonal matrix; the second equality from (6.31); and the last equality because $\hat{\mathbf{x}}$ is a unit vector. It should be noted that the spectral density $\widetilde{\mathsf{F}}'(\lambda)$ does not depend on the choice of $\hat{\mathbf{x}}$.

Using the expression for the fading number of SISO fading channels with memory given in Corollary 5.4, the fading number can be lower bounded by

$$\chi(\{\mathbf{H}_k^\mathsf{T}\}) \geq -1 + \log d_*^2 - \mathrm{Ei}(-d_*^2) + \log \frac{1}{\widetilde{\epsilon}^2} \tag{6.40}$$

with

$$d_* = \max_{\|\hat{\mathbf{x}}\|=1} |\mathbf{d}^\mathsf{T}\hat{\mathbf{x}}| \tag{6.41}$$

and

$$\begin{aligned}\widetilde{\epsilon}^2 &= \exp\left\{\int_{-1/2}^{1/2} \log \widetilde{\mathsf{F}}'(\lambda)\,\mathrm{d}\lambda\right\} \\ &= \exp\left\{\int_{-1/2}^{1/2} \log \mathsf{F}'(\lambda)\,\mathrm{d}\lambda\right\}, \end{aligned} \tag{6.42}$$

where the last equality follows from (6.39). Comparing (6.40) and (6.32) as well as (6.42) and (6.34) proves the claim.

Chapter 7

Non-Regular Processes

In this chapter, we show the derivations of the results presented in Section 5.2, *i.e.*, where the fading process is non-regular. In contrast to the analysis in Chapter 6, we now consider channel capacity under a peak-power constraint A on the inputs. The emphasis is on the high signal-to-noise ratio (SNR) regime, where the SNR is defined as A^2/σ^2. In particular, bounds on the capacity pre-log are derived in the cases, where capacity grows logarithmically in the SNR. In the case, where capacity only increases double-logarithmically with the SNR, the capacity pre-log-log is considered.

In Section 7.1 we discuss a single-input single-output (SISO) fading channel with a double-logarithmic increase with the SNR. In particular, we derive a lower bound on channel capacity which allows us to formulate an expression for the capacity pre-log-log. In Section 7.2 an upper bound on the capacity of multiple-input single-output (MISO) fading channels is derived that is tight in the case where the channels are uncorrelated. Section 7.3 studies the capacity of multiple-input multiple-output (MIMO) fading channels, where the channels are uncorrelated. We derive an upper bound on the capacity as well as an upper bound on the capacity pre-log.

7.1 The Pre-Log-Log of SISO Fading Channels

In this section we consider a SISO fading channel where the corresponding channel capacity only grows double-logarithmically in the SNR, *i.e.*,

$$\overline{\lim_{\mathrm{SNR}\to\infty}} \frac{C(\mathrm{SNR})}{\log\log \mathrm{SNR}} < \infty. \tag{7.1}$$

In this case, the capacity pre-log is zero and it is more interesting to study the capacity pre-log-log. However, when (7.1) holds, the lower bound on the capacity (5.28) given in Theorem 5.8 is not tight. In the next section we derive a lower bound that is tight for this case. An expression for the capacity pre-log-log is derived in Section 7.1.2.

7.1.1 A Proof of Theorem 5.10

To derive a lower bound on channel capacity we consider circularly symmetric inputs $\{X_k\}$ that are independent and identically distributed (i.i.d.) with $\log|X_k|^2$ uniformly distributed over the interval $[\alpha \log \mathsf{A}^2, \log \mathsf{A}^2]$ for $0 < \alpha < 1$.

We begin by considering the expression for channel capacity, *i.e.*,

$$C(\mathrm{SNR}) = \lim_{n\to\infty} \frac{1}{n} \sup_{p_{X^n}} I(X^n; Y^n) \tag{7.2}$$

and using the chain rule [5]:

$$I(X^n; Y^n) = \sum_{k=1}^{n} I(X_k; Y^n | X^{k-1}). \tag{7.3}$$

Then, channel capacity can be lower bounded by

$$\begin{aligned} C(\mathrm{SNR}) &\geq \lim_{n\to\infty} \frac{1}{n} I(X^n; Y^n) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I(X_k; Y^n | X^{k-1}) \\ &\geq \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} I(X_k; Y^k | X^{k-1}) \\ &\geq \varliminf_{k\to\infty} I(X_k; Y^k | X^{k-1}), \end{aligned} \tag{7.4}$$

where the first inequality follows from the probably suboptimal choice of the input distribution; the next equality from (7.3); the subsequent inequality from removing information; and the last inequality by a Cesàro-type theorem [5, Theorem 4.2.3].

Making use of the properties of the chosen input distribution, *i.e.*, $\{X_k\}$ i.i.d. and satisfying $|X_k|^2 \geq \mathsf{A}^{2\alpha}$, we can further lower bound (7.4) by

$$I(X_k; Y^k | X^{k-1}) = I(X_k; Y_k, Y^{k-1}, X^{k-1})$$

$$
\begin{aligned}
&= I\left(X_k; Y_k, \left\{\frac{Y_\nu}{X_\nu}\right\}_{\nu=1}^{k-1}, X_{k-1}\right) \\
&= I\left(X_k; Y_k, \left\{H_\nu + \frac{Z_\nu}{X_\nu}\right\}_{\nu=1}^{k-1} \middle| X^{k-1}\right) \\
&\geq I(X_k; Y_k, \{H_\nu + W'_\nu\}_{\nu=1}^{k-1}) \\
&= I(X_k; Y_k | \{H_\nu + W'_\nu\}_{\nu=1}^{k-1}),
\end{aligned} \tag{7.5}
$$

with $\{W'_\nu\}$ i.i.d. $\mathcal{N}\left(0, \frac{\sigma^2}{\mathsf{A}^{2\alpha}}\right)$. The last equality follows from the fact that the present input X_k is independent of the channel realization.

To lower bound $I(X_k; Y_k|\{H_\nu + W'_\nu\}_{\nu=1}^{k-1})$ we express the fading at time k as

$$H_k = \overline{H}_k + \widetilde{H}_k, \tag{7.6}$$

where $\overline{H}_k$ is the best estimate of H_k given a noisy observation of its past. Thus, for given past values $h_{k-1}+w_{k-1}, \ldots, h_1+w_1$ the estimate $\overline{h}_k$ is given by

$$\overline{h}_k = \mathsf{E}\left[H_k \,\middle|\, \{H_\nu + W'_\nu\}_{\nu=1}^{k-1}\right]. \tag{7.7}$$

Moreover, we note that $\widetilde{H}_k \sim \mathcal{N}\left(0, \epsilon_k^2\right)$ where ϵ_k^2 is the minimum mean squared error in predicting the present fading H_k from a noisy observation of its past $\{H_\nu + W'_\nu\}_{\nu=1}^{k-1}$. Additionally, it follows from [2] and references therein

$$\lim_{k\to\infty} \epsilon_k^2 = \epsilon_{\mathrm{MSE}}^2(\delta^2)\Big|_{\delta^2=\frac{\sigma^2}{\mathsf{A}^{2\alpha}}} \tag{7.8}$$

with

$$\epsilon_{\mathrm{MSE}}^2(\delta^2) = \exp\left\{\int_{-1/2}^{1/2} \log(\mathsf{F}'(\lambda) + \delta^2)\,\mathrm{d}\lambda\right\} - \delta^2. \tag{7.9}$$

We continue (7.5)

$$
\begin{aligned}
&I(X_k; Y_k|\{H_\nu + W'_\nu\}_{\nu=1}^{k-1}) \\
&\quad = I(X_k; (\overline{H}_k + \widetilde{H}_k)X_k + Z_k | \overline{H}_k) \\
&\quad = h\left((\overline{H}_k + \widetilde{H}_k)X_k + Z_k \middle| \overline{H}_k\right) - h\left((\overline{H}_k + \widetilde{H}_k)X_k + Z_k \middle| X_k, \overline{H}_k\right) \\
&\quad = h\left((\overline{H}_k + \widetilde{H}_k)X_k + Z_k \middle| \overline{H}_k\right) - h\left(\widetilde{H}_k X_k + Z_k \middle| X_k, \overline{H}_k\right) \\
&\quad = h\left((\overline{H}_k + \widetilde{H}_k)X_k + Z_k \middle| \overline{H}_k\right) - h\left(\widetilde{H}_k + \frac{Z_k}{X_k} \middle| X_k\right) - \mathsf{E}\left[\log|X_k|^2\right] \\
&\quad \geq h\left((\overline{H}_k + \widetilde{H}_k)X_k + Z_k \middle| \overline{H}_k\right) - h\left(\widetilde{H}_k + W'_k\right) - \mathsf{E}\left[\log|X_k|^2\right] \qquad (7.10) \\
&\quad \geq h\left((\overline{H}_k + \widetilde{H}_k)X_k + Z_k \middle| \overline{H}_k, \widetilde{H}_k, Z_k\right) - h(\widetilde{H}_k + W'_k) - \mathsf{E}\left[\log|X_k|^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&= h(H_k X_k | H_k) - h(\widetilde{H}_k + W'_k) - \mathsf{E}\left[\log |X_k|^2\right] \\
&= \mathsf{E}\left[\log |H_k|^2\right] + h(X_k) - h(\widetilde{H}_k + W'_k) - \mathsf{E}\left[\log |X_k|^2\right] \\
&= \log 2\pi + h(|X_k|) - \mathsf{E}[\log |X_k|] + \mathsf{E}\left[\log |H_k|^2\right] - h(\widetilde{H}_k + W'_k) \qquad (7.11) \\
&= \log 2\pi + h(\log |X_k|) + \mathsf{E}\left[\log |H_k|^2\right] - h(\widetilde{H}_k + W'_k) \\
&= \log \pi + h(\log |X_k|^2) + \mathsf{E}\left[\log |H_k|^2\right] - h(\widetilde{H}_k + W'_k) \\
&= \log \pi + \log(\log \mathsf{A}^2 - \alpha \log \mathsf{A}^2) + \mathsf{E}\left[\log |H_k|^2\right] - h(\widetilde{H}_k + W'_k) \\
&= \log \pi + \log\log \mathsf{A}^2 + \log(1-\alpha) + \mathsf{E}\left[\log |H_k|^2\right] - h(\widetilde{H}_k + W'_k) \\
&= \log \pi + \log\log \frac{\mathsf{A}^2}{\sigma^2} + \log(1-\alpha) + \mathsf{E}\left[\log |H_k|^2\right] - h(\widetilde{H}_k + W'_k) + o(1),
\end{aligned}
$$
$$(7.12)$$

where $o(1)$ tends to zero as the SNR goes to infinity and $\{W'_\nu\}$ is as in (7.5). Here, the first equality follows by (7.6); the subsequent equality from the definition of mutual information; the next equality because differential entropy is invariant under deterministic translation; and the subsequent equality from the behavior under scaling of the differential entropy of random variables [5].

Inequality (7.10) follows from the fact that $|X_k|^2 \geq \mathsf{A}^{2\alpha}$; the next inequality because conditioning reduces entropy; the subsequent equality from the invariance of differential entropy under deterministic translation; and the following equality from the behavior under scaling of the differential entropy of random variables.

The next equality (7.11) follows by [1, Lemma 6.16] for the differential entropy of circularly symmetric random variables; the subsequent equality by relating the differential entropy of a positive random variable to that of its logarithm as in [1, Lemma 6.15]; the following equality from the behavior under scaling of the differential entropy of random variables; the following equality by evaluating the differential entropy for the chosen input distribution; the subsequent equality by analyzing $\log(\log \mathsf{A}^2 - \alpha \log \mathsf{A}^2)$; and the last equality by the limiting behavior of the $\log\log(\cdot)$-function under scaling, *i.e.*,

$$\lim_{\mathrm{SNR}\to\infty} \{\log\log(\alpha \mathrm{SNR}) - \log\log \mathrm{SNR}\} = 0, \qquad \alpha > 0. \qquad (7.13)$$

We proceed to evaluate the terms $\mathsf{E}\left[\log |H_k|^2\right]$ and $h(\widetilde{H}_k + W'_k)$ on the RHS of (7.12). Using the expression for the expectation of the logarithm of a noncentral chi-square distributed random variable [1, Appendix X], the first term can be written as

$$\mathsf{E}\left[\log|H_k|^2\right] = \log|d|^2 - \mathrm{Ei}(-|d|^2). \tag{7.14}$$

The second term can be computed by noting that $(\widetilde{H}_k + W'_k) \sim \mathcal{N}\left(0, \epsilon_k^2 + \frac{\sigma^2}{\mathsf{A}^{2\alpha}}\right)$ and using the expression for the differential entropy of a Gaussian random variable:

$$h(\widetilde{H}_k + W'_k) = \log\pi + 1 - \log \left.\frac{1}{\epsilon_k^2 + \delta^2}\right|_{\delta^2 = \frac{\sigma^2}{\mathsf{A}^{2\alpha}}}. \tag{7.15}$$

Combining (7.15), (7.14), (7.12), (7.8), (7.5), and (7.4) we can lower bound channel capacity by

$$\begin{aligned} C(\mathrm{SNR}) &\geq \varliminf_{k\to\infty} \left\{ \log\log\mathrm{SNR} + \log \left.\frac{1}{\epsilon_k^2 + \delta^2}\right|_{\delta^2 = \frac{\sigma^2}{\mathsf{A}^{2\alpha}}} + O(1) \right\} \\ &= \varliminf_{k\to\infty} \log \left.\frac{1}{\epsilon_k^2 + \delta^2}\right|_{\delta^2 = \frac{\sigma^2}{\mathsf{A}^{2\alpha}}} + \log\log\mathrm{SNR} + O(1) \\ &= \log \left.\frac{1}{\epsilon_{\mathrm{MSE}}^2(\delta^2) + \delta^2}\right|_{\delta^2 = \frac{\sigma^2}{\mathsf{A}^{2\alpha}}} + \log\log\mathrm{SNR} + O(1), \end{aligned} \tag{7.16}$$

where $O(1)$ only depends on d and on α.

7.1.2 A Proof of Corollary 5.11

In the following we derive an expression for the capacity pre-log-log. We begin by using Theorem 5.8 to upper bound the capacity pre-log-log Λ:

$$C(\mathrm{SNR}) \leq \log\frac{1}{\epsilon_{\mathrm{MSE}}^2(1/\mathrm{SNR})} + \log\log\mathrm{SNR} + O(1) \tag{7.17}$$

and, consequently,

$$\begin{aligned} \Lambda &= \varlimsup_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log\log\mathrm{SNR}} \\ &\leq 1 + \varlimsup_{\mathrm{SNR}\to\infty} \frac{\log\frac{1}{\epsilon_{\mathrm{MSE}}^2(1/\mathrm{SNR})}}{\log\log\mathrm{SNR}}. \end{aligned} \tag{7.18}$$

To derive a lower bound on the capacity pre-log-log we use (7.16)

$$\begin{aligned} \Lambda &= \varlimsup_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log\log\mathrm{SNR}} \\ &\geq 1 + \varlimsup_{\mathsf{A}\to\infty} \frac{\log\left.\frac{1}{\epsilon_{\mathrm{MSE}}^2(\delta^2)+\delta^2}\right|_{\delta^2=\frac{\sigma^2}{\mathsf{A}^{2\alpha}}}}{\log\log\frac{\mathsf{A}^2}{\sigma^2}} \end{aligned} \tag{7.19}$$

for $0 < \alpha < 1$.

Now, we show that the lower bound on the pre-log-log in (7.19) coincides with the upper bound in (7.18):

$$\begin{aligned}
&\varlimsup_{\mathsf{A}\to\infty} \frac{\log \frac{1}{\epsilon^2_{\mathrm{MSE}}(\delta^2)+\delta^2}\Big|_{\delta^2=\frac{\sigma^2}{\mathsf{A}^{2\alpha}}}}{\log\log\frac{\mathsf{A}^2}{\sigma^2}} \\
&= \varlimsup_{\mathsf{A}\to\infty} \left\{ \frac{\log \frac{1}{\epsilon^2_{\mathrm{MSE}}(\delta^2)+\delta^2}\Big|_{\delta^2=\frac{\sigma^2}{\mathsf{A}^{2\alpha}}}}{\log\log\frac{\mathsf{A}^{2\alpha}}{\sigma^2}} \cdot \frac{\log\log\frac{\mathsf{A}^{2\alpha}}{\sigma^2}}{\log\log\frac{\mathsf{A}^2}{\sigma^2}} \right\} \\
&= \varlimsup_{\mathsf{A}\to\infty} \frac{\log \frac{1}{\epsilon^2_{\mathrm{MSE}}(\delta^2)+\delta^2}\Big|_{\delta^2=\frac{\sigma^2}{\mathsf{A}^{2\alpha}}}}{\log\log\frac{\mathsf{A}^{2\alpha}}{\sigma^2}} \cdot \lim_{\mathsf{A}\to\infty} \frac{\log\log\frac{\mathsf{A}^{2\alpha}}{\sigma^2}}{\log\log\frac{\mathsf{A}^2}{\sigma^2}} \\
&= \varlimsup_{\mathrm{SNR}'\to\infty} \frac{\log \frac{1}{\epsilon^2_{\mathrm{MSE}}(1/\mathrm{SNR}')+1/\mathrm{SNR}'}}{\log\log \mathrm{SNR}'} \cdot \lim_{\mathsf{A}\to\infty} \frac{\log\log\frac{\mathsf{A}^{2\alpha}}{\sigma^2}}{\log\log\frac{\mathsf{A}^2}{\sigma^2}}
\end{aligned} \tag{7.20}$$

with $\mathrm{SNR}' = \frac{\mathsf{A}^{2\alpha}}{\sigma^2}$. To evaluate the RHS of (7.20), we first note that

$$\begin{aligned}
\lim_{\mathsf{A}\to\infty} \frac{\log\log\frac{\mathsf{A}^{2\alpha}}{\sigma^2}}{\log\log\frac{\mathsf{A}^2}{\sigma^2}} &= \lim_{\mathsf{A}\to\infty} \frac{\log(\alpha\log\mathsf{A}^2 - \log\sigma^2)}{\log\log\frac{\mathsf{A}^2}{\sigma^2}} \\
&= \lim_{\mathsf{A}\to\infty} \frac{\log(\alpha\log\frac{\mathsf{A}^2}{\sigma^2} + (\alpha-1)\log\sigma^2)}{\log\log\frac{\mathsf{A}^2}{\sigma^2}} \\
&= \lim_{\mathsf{A}\to\infty} \frac{\log\left(\alpha\log\frac{\mathsf{A}^2}{\sigma^2}\right)}{\log\log\frac{\mathsf{A}^2}{\sigma^2}} \\
&= \lim_{\mathsf{A}\to\infty} \frac{\log\alpha + \log\log\frac{\mathsf{A}^2}{\sigma^2}}{\log\log\frac{\mathsf{A}^2}{\sigma^2}} \\
&= 1.
\end{aligned} \tag{7.21}$$

The other term on the RHS of (7.20) can be evaluated by noting that

$$\varlimsup_{\delta\downarrow 0} \frac{\epsilon^2_{\mathrm{MSE}}(\delta^2)}{\delta^2} = \infty, \tag{7.22}$$

since otherwise

$$\varlimsup_{\mathrm{SNR}'\to\infty} \frac{\log \frac{1}{\epsilon^2_{\mathrm{MSE}}(1/\mathrm{SNR}')+1/\mathrm{SNR}'}}{\log\log \mathrm{SNR}'} = \infty, \tag{7.23}$$

and, consequently, the capacity pre-log-log Λ would be infinite, which contradicts (7.1). It follows that

$$\varlimsup_{\mathrm{SNR}'\to\infty} \frac{\log \frac{1}{\epsilon^2_{\mathrm{MSE}}(1/\mathrm{SNR}')+1/\mathrm{SNR}'}}{\log\log \mathrm{SNR}'} = \varlimsup_{\mathrm{SNR}'\to\infty} \frac{\log \frac{1}{\epsilon^2_{\mathrm{MSE}}(1/\mathrm{SNR}')}}{\log\log \mathrm{SNR}'}$$
$$= \varlimsup_{\mathrm{SNR}\to\infty} \frac{\log \frac{1}{\epsilon^2_{\mathrm{MSE}}(1/\mathrm{SNR})}}{\log\log \mathrm{SNR}}. \tag{7.24}$$

Combining (7.24), (7.21), (7.20), and (7.19) we obtain the following lower bound on the capacity pre-log-log Λ:

$$\Lambda \geq 1 + \varlimsup_{\mathrm{SNR}\to\infty} \frac{\log \frac{1}{\epsilon^2_{\mathrm{MSE}}(1/\mathrm{SNR})}}{\log\log \mathrm{SNR}}, \tag{7.25}$$

which coincides with the upper bound in (7.18).

We now show that (7.25) is identical to the expression for the pre-log-log (5.35) given in Corollary 5.11:

$$\begin{aligned}
\Lambda &= 1 + \varlimsup_{\mathrm{SNR}\to\infty} \frac{\log \frac{1}{\epsilon^2_{\mathrm{MSE}}(1/\mathrm{SNR})}}{\log\log \mathrm{SNR}} \\
&= 1 + \varlimsup_{\delta^2\downarrow 0} \frac{\log \frac{1}{\epsilon^2_{\mathrm{MSE}}(\delta^2)}}{\log\log \frac{1}{\delta^2}} \\
&= 1 + \varlimsup_{\delta^2\downarrow 0} \frac{\log \frac{1}{\epsilon^2_{\mathrm{MSE}}(\delta^2)+\delta^2}}{\log\log \frac{1}{\delta^2}} \\
&= 1 + \varlimsup_{\delta^2\downarrow 0} \frac{-\log\left(\epsilon^2_{\mathrm{MSE}}(\delta^2)+\delta^2\right)}{\log\log \frac{1}{\delta^2}} \\
&= 1 + \varlimsup_{\delta^2\downarrow 0} \frac{-\log\left(\exp\left\{\int_{-1/2}^{1/2} \log\left(\mathsf{F}'(\lambda)+\delta^2\right)\,\mathrm{d}\lambda\right\}\right)}{\log\log \frac{1}{\delta^2}} \\
&= 1 + \varlimsup_{\delta^2\downarrow 0} \frac{-\int_{-1/2}^{1/2} \log\left(\mathsf{F}'(\lambda)+\delta^2\right)\,\mathrm{d}\lambda}{\log\log \frac{1}{\delta^2}},
\end{aligned} \tag{7.26}$$

where the first equality follows by (7.25) and (7.18); the subsequent equality by substituting $\delta^2 = 1/\mathrm{SNR}$; the next equality by (7.22); the following equality from the behavior of the logarithm function; the subsequent equality from the expression for the minimum mean squared error in predicting the present fading from a noisy observation of its past; and the last equality by taking the logarithm of the exponential function.

7.2 The Pre-Log of MISO Fading Channels

In this section we consider a MISO fading channel with a non-regular fading process. An upper bound on capacity of such channels is derived in the next section. In Section 7.2.2 we deduce from this upper bound an expression for the capacity pre-log in the case where the channels are uncorrelated.

7.2.1 A Proof of Theorem 5.12

Channel capacity is defined as

$$C(\mathrm{SNR}) = \lim_{n\to\infty} \frac{1}{n} \sup_{p_{\mathbf{X}^n}} I(\mathbf{X}^n; Y^n), \tag{7.27}$$

where the supremum is taken over all input distributions fulfilling the peak-power constraint (4.19). To upper bound the capacity we begin by using the chain rule [5]

$$I(\mathbf{X}^n; Y^n) = \sum_{k=1}^{n} I(Y_k; \mathbf{X}^n | Y^{k-1}) \tag{7.28}$$

and upper bounding each of the terms in the sum by

$$\begin{aligned} I(Y_k; \mathbf{X}^n | Y^{k-1}) &= I(Y_k; \mathbf{X}^n, Y^{k-1}) - I(Y_k; Y^{k-1}) \\ &\leq I(Y_k; \mathbf{X}^n, Y^{k-1}) \\ &= I(Y_k; \mathbf{X}^k, Y^{k-1}) \\ &= I(Y_k; \mathbf{X}_k) + I(Y_k; \mathbf{X}^{k-1}, Y^{k-1} | \mathbf{X}_k) \\ &\leq \sup_{p_{\mathbf{X}^k}} \Big\{ I(Y_k; \mathbf{X}_k) + I(Y_k; \mathbf{X}^{k-1}, Y^{k-1} | \mathbf{X}_k) \Big\} \\ &\leq \sup_{p_{\mathbf{X}_k}} I(Y_k; \mathbf{X}_k) + \sup_{p_{\mathbf{X}^k}} I(Y_k; \mathbf{X}^{k-1}, Y^{k-1} | \mathbf{X}_k) \end{aligned} \tag{7.29}$$

where the first equality follows from the chain rule; the following inequality from non-negativity of mutual information; the subsequent equality from the absence of feedback, which results in future inputs being independent of the present output given the present input and the past inputs and outputs; the next equality follows from the chain rule; the following inequality by maximizing over all possible input distributions satisfying the peak-power constraint; and the last inequality from splitting up the supremum.

The first term on the RHS of (7.29) corresponds to the capacity of a memoryless MISO fading channel and is given by

$$\sup_{p_{\mathbf{X}_k}} I(Y_k;\mathbf{X}_k) = \chi_{\text{i.i.d.}}(\mathbf{H}_0^{\mathsf{T}}) + \log\log \text{SNR} + o(1), \tag{7.30}$$

where $\chi_{\text{i.i.d.}}(\mathbf{H}_0^{\mathsf{T}})$ denotes the fading number of MISO fading channels (5.4) given in Corollary 5.2. Note that the expression for the fading number was derived under an average-power constraint on the inputs. However, it can be shown [1] that this expression holds also if the average-power constraint is replaced with a peak-power constraint.

To evaluate the second term on the RHS of (7.29) we note that the channel output Y_k at time k is given by

$$Y_k = \mathbf{H}_k^{\mathsf{T}}\mathbf{X}_k + Z_k. \tag{7.31}$$

Now, we proceed by noting that if $\mathbf{X}_k$ is given, then $\mathbf{X}^{k-1}$ and Y^{k-1} influence the mutual information $I(Y_k;\mathbf{X}^{k-1},Y^{k-1}|\mathbf{X}_k)$ only through the information they convey on the fading realizations $\mathbf{H}^{k-1}$. If the number of transmit antennas is larger than one, then the pair $(\mathbf{X}_k, Y_k)$ does not provide information about the fading vector $\mathbf{H}_k$, but only about the projection of $\mathbf{H}_k$ onto the input vector $\mathbf{X}_k$. However, we will assume that $(\mathbf{X}_k, Y_k)$ gives us information about the whole vector $\mathbf{H}_k$, which yields an upper bound. It follows

$$\begin{aligned}
&\sup_{p_{\mathbf{X}^k}} I(Y_k;\mathbf{X}^{k-1},Y^{k-1}|\mathbf{X}_k)\\
&\quad\le \sup_{p_{\mathbf{X}^k}} I\left(Y_k;\mathbf{X}^{k-1},\{H_\nu^{(1)}X_\nu^{(1)}+Z_\nu\}_{\nu=1}^{k-1},\ldots,\{H_\nu^{(n_{\mathrm{T}})}X_\nu^{(n_{\mathrm{T}})}+Z_\nu\}_{\nu=1}^{k-1}\,\middle|\,\mathbf{X}_k\right)\\
&\quad= \sup_{p_{\mathbf{X}^k}} I\left(Y_k;\mathbf{X}^{k-1},\left\{H_\nu^{(1)}+\frac{Z_\nu}{X_\nu^{(1)}}\right\}_{\nu=1}^{k-1},\ldots,\left\{H_\nu^{(n_{\mathrm{T}})}+\frac{Z_\nu}{X_\nu^{(n_{\mathrm{T}})}}\right\}_{\nu=1}^{k-1}\,\middle|\,\mathbf{X}_k\right)\\
&\quad\le \sup_{p_{\mathbf{X}_k}} I\left(Y_k;\{H_\nu^{(1)}+W_\nu^{(1)}\}_{\nu=1}^{k-1},\ldots,\{H_\nu^{(n_{\mathrm{T}})}+W_\nu^{(n_{\mathrm{T}})}\}_{\nu=1}^{k-1}\,\middle|\,\mathbf{X}_k\right)\\
&\quad= \sup_{p_{\mathbf{X}_k}} I\left(Y_k;\{\mathbf{H}_\nu+\mathbf{W}_\nu\}_{\nu=1}^{k-1}\,\middle|\,\mathbf{X}_k\right)
\end{aligned} \tag{7.32}$$

with

$$\{\mathbf{W}_\nu\} \text{ i.i.d.} \sim \mathcal{N}\left(0,\frac{\sigma^2}{\mathsf{A}^2}\mathsf{I}\right). \tag{7.33}$$

Here, the first inequality follows by assuming that $(\mathbf{X}_\nu, Y_\nu)$ provides information about the whole fading vector $\mathbf{H}_\nu$; and the following equality because dividing by $X_\nu^{(i)}$ does not influence the mutual information. The subsequent inequality follows by recalling that the variables $H_\nu^{(t)} + \frac{Z_\nu}{X_\nu^{(t)}}$ influence the mutual information only through the information they give about the fading $\mathbf{H}_\nu$. This information is maximized, if the inputs $X_\nu^{(t)}$ are as large as possible since then the additive noise term is minimized. Furthermore, we choose all inputs $X_\nu^{(t)}$ to be of maximum magnitude A, which violates the peak-power constraint, *i.e.*, $\|\mathbf{X}_\nu\| \leq \mathsf{A}$, and yields an upper bound on the mutual information.

We can further upper bound the RHS of (7.32) by

$$\begin{aligned}
&\sup_{p_{\mathbf{X}_k}} I\left(Y_k; \{\mathbf{H}_\nu + \mathbf{W}_\nu\}_{\nu=1}^{k-1} \,\middle|\, \mathbf{X}_k\right) \\
&\quad \leq \sup_{p_{\mathbf{X}_0}} I\left(Y_0; \{\mathbf{H}_\nu + \mathbf{W}_\nu\}_{\nu=-\infty}^{-1} \,\middle|\, \mathbf{X}_0\right) \\
&\quad \leq \sup_{\|\mathbf{x}_0\| \leq \mathsf{A}} I\left(Y_0; \{\mathbf{H}_\nu + \mathbf{W}_\nu\}_{\nu=-\infty}^{-1} \,\middle|\, \mathbf{X}_0 = \mathbf{x}_0\right),
\end{aligned} \tag{7.34}$$

where the first inequality follows from the stationarity of the channel and from adding information; and the second inequality from the expression for the conditional mutual information and from the fact that the expectation of a random variable can never be larger than its largest value.

To evaluate (7.34) we express the fading as

$$\mathbf{H}_0 = \overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0, \tag{7.35}$$

where $\overline{\mathbf{H}}_0$ is the best estimate of the fading $\mathbf{H}_0$ given a noisy observation of its past. Thus, for given past values $\mathbf{h}_{-1} + \mathbf{w}_{-1}, \mathbf{h}_{-2} + \mathbf{w}_{-2}, \ldots$ the estimate $\overline{\mathbf{h}}_0$ is given by

$$\overline{\mathbf{h}}_0 = \mathsf{E}\left[\mathbf{H}_0 \,\middle|\, \{\mathbf{H}_\nu + \mathbf{W}_\nu\}_{\nu=-\infty}^{-1}\right]. \tag{7.36}$$

Furthermore, note that $\widetilde{\mathbf{H}}_0 \sim \mathcal{N}\,(0, \Sigma(1/\mathrm{SNR}))$ with the prediction error covariance matrix $\Sigma(1/\mathrm{SNR})$. It follows that

$$\begin{aligned}
&I\left(Y_0; \{\mathbf{H}_\nu + \mathbf{W}_\nu\}_{\nu=-\infty}^{-1} \,\middle|\, \mathbf{X}_0 = \mathbf{x}_0\right) \\
&\quad = I\left((\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0)^\mathsf{T}\mathbf{X}_0 + Z_0; \overline{\mathbf{H}}_0 \,\middle|\, \mathbf{X}_0 = \mathbf{x}_0\right) \\
&\quad = h\left((\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0)^\mathsf{T}\mathbf{X}_0 + Z_0 \,\middle|\, \mathbf{X}_0 = \mathbf{x}_0\right) \\
&\qquad - h\left((\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0)^\mathsf{T}\mathbf{X}_0 + Z_0 \,\middle|\, \overline{\mathbf{H}}_0, \mathbf{X}_0 = \mathbf{x}_0\right) \\
&\quad = h\left((\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0)^\mathsf{T}\mathbf{X}_0 + Z_0 \,\middle|\, \mathbf{X}_0 = \mathbf{x}_0\right) - h\left(\widetilde{\mathbf{H}}_0^\mathsf{T}\mathbf{X}_0 + Z_0 \,\middle|\, \mathbf{X}_0 = \mathbf{x}_0\right),
\end{aligned} \tag{7.37}$$

where the first equality follows by (7.35) and by noting that $\widetilde{\mathbf{H}}_0$ is independent from the variables $\{\mathbf{H}_\nu + \mathbf{W}_\nu\}_{\nu=-\infty}^{-1}$ which results in the fact that $\{\mathbf{H}_\nu + \mathbf{W}_\nu\}_{\nu=-\infty}^{-1}$ influence the mutual information only through the information they convey on $\overline{\mathbf{H}}_0$; the subsequent equality follows from the definition of mutual information; and the last equality from the invariance of differential entropy under deterministic translation.

To evaluate the first term on the RHS of (7.37) we note that $\left((\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0)\mathbf{X}_0 + Z_0 \middle| \mathbf{X}_0 = \mathbf{x}_0\right) \sim \mathcal{N}\left(\mathbf{d}^\mathsf{T}\mathbf{x}_0, \mathbf{x}_0^\mathsf{T}\mathsf{K}\mathbf{x}_0^* + \sigma^2\right)$ with $\mathbf{d} = \mathsf{E}[\mathbf{H}_0]$ and $\mathsf{K} = \mathsf{E}\left[(\mathbf{H}_0 - \mathbf{d})(\mathbf{H}_0 - \mathbf{d})^\dagger\right]$. It follows that

$$h((\overline{\mathbf{H}}_0 + \widetilde{\mathbf{H}}_0)^\mathsf{T}\mathbf{X}_0 + Z_0|\mathbf{X}_0 = \mathbf{x}_0) = \log\pi + 1 + \log(\mathbf{x}_0^\mathsf{T}\mathsf{K}\mathbf{x}_0^* + \sigma^2), \tag{7.38}$$

where we used the expression for the differential entropy of a Gaussian random variable.

To express the second term on the RHS of (7.37) we note that $\left(\widetilde{\mathbf{H}}_0^\mathsf{T}\mathbf{X}_0 + Z_0 \middle| \mathbf{X}_0 = \mathbf{x}_0\right) \sim \mathcal{N}\left(0, \mathbf{x}_0^\mathsf{T}\Sigma(1/\mathrm{SNR})\mathbf{x}_0^* + \sigma^2\right)$ and it follows

$$h(\widetilde{\mathbf{H}}_0^\mathsf{T}\mathbf{X}_0 + Z_0|\mathbf{X}_0 = \mathbf{x}_0) = \log\pi + 1 + \log(\mathbf{x}_0^\mathsf{T}\Sigma(1/\mathrm{SNR})\mathbf{x}_0^* + \sigma^2). \tag{7.39}$$

Combining (7.39), (7.38), and (7.37) we obtain

$$\begin{aligned}
&\sup_{\|\mathbf{x}_0\|\leq \mathsf{A}} I\left(Y_0; \{\mathbf{H}_\nu + \mathbf{W}_\nu\}_{\nu=-\infty}^{-1} | \mathbf{X}_0 = \mathbf{x}_0\right) \\
&= \sup_{\|\mathbf{x}_0\|\leq \mathsf{A}} \log \frac{\mathbf{x}_0^\mathsf{T}\mathsf{K}\mathbf{x}_0^* + \sigma^2}{\mathbf{x}_0^\mathsf{T}\Sigma(1/\mathrm{SNR})\mathbf{x}_0^* + \sigma^2} \\
&= \sup_{\|\mathbf{x}_0\|\leq \mathsf{A}} \log \frac{\hat{\mathbf{x}}_0^\mathsf{T}\mathsf{K}\hat{\mathbf{x}}_0^*\|\mathbf{x}_0\|^2 + \sigma^2}{\hat{\mathbf{x}}_0^\mathsf{T}\Sigma(1/\mathrm{SNR})\hat{\mathbf{x}}_0^*\|\mathbf{x}_0\|^2 + \sigma^2} \\
&\leq \sup_{\|\hat{\mathbf{x}}_0\|=1} \log \frac{\hat{\mathbf{x}}_0^\mathsf{T}\mathsf{K}\hat{\mathbf{x}}_0^*}{\hat{\mathbf{x}}_0^\mathsf{T}\Sigma(1/\mathrm{SNR})\hat{\mathbf{x}}_0^*} \\
&\leq \sup_{\|\hat{\mathbf{x}}_0\|=1} \log \frac{\|\mathsf{K}\|}{\hat{\mathbf{x}}_0^\mathsf{T}\Sigma(1/\mathrm{SNR})\hat{\mathbf{x}}_0^*} \\
&= \log \frac{\|\mathsf{K}\|}{\lambda_{\min}(1/\mathrm{SNR})},
\end{aligned} \tag{7.40}$$

where $\lambda_{\min}(1/\mathrm{SNR})$ denotes the minimum eigenvalue of $\Sigma(1/\mathrm{SNR})$. Here, the second equality follows by expressing the input vector $\mathbf{x}_0$ in terms of its magnitude $\|\mathbf{x}_0\|$ and its direction $\hat{\mathbf{x}}_0 = \frac{\mathbf{x}_0}{\|\mathbf{x}_0\|}$; the following inequality by noting that[1] $\hat{\mathbf{x}}_0^\mathsf{T}\mathsf{K}\hat{\mathbf{x}}_0^* \geq \hat{\mathbf{x}}_0^\mathsf{T}\Sigma(1/\mathrm{SNR})\hat{\mathbf{x}}_0^*$ in which case the mutual information

[1] The reasoning is identical to that in Section 6.1 when considering regular fading processes. We therefore refer to Section 6.1 for a more elaborate explanation.

is monotonically increasing in $\|\mathbf{x}_0\|^2$, and by letting $\|\mathbf{x}_0\|^2$ go to infinity; the following inequality by the Rayleigh-Ritz Theorem [9] and the definition of the Euclidean operator norm of matrices; and the last equality by minimizing $\hat{\mathbf{x}}_0^{\mathsf{T}}\Sigma(1/\mathrm{SNR})\hat{\mathbf{x}}_0^*$ over all unit vectors $\hat{\mathbf{x}}_0$ (Rayleigh-Ritz Theorem [9]).

Combining (7.40), (7.34), (7.30), (7.29), and (7.28), and using the expression for channel capacity, we obtain the upper bound

$$C(\mathrm{SNR}) \leq \log \frac{1}{\lambda_{\min}(1/\mathrm{SNR})} + \log\log \mathrm{SNR} + O(1). \tag{7.41}$$

This concludes the proof.

As mentioned in Section 5.2.3, we are only able to express $\lambda_{\min}(1/\mathrm{SNR})$, when the channels are uncorrelated, *i.e.*, $\mathsf{E}\left[(\mathbf{H}_{k+m}-\mathbf{d})(\mathbf{H}_k-\mathbf{d})^\dagger\right]$ is a diagonal matrix.

7.2.2 A Proof of Corollary 5.13

In the following we derive an expression for the capacity pre-log of MISO fading channels, when the channels are uncorrelated, *i.e.*, $\mathsf{E}\left[(\mathbf{H}_{k+m}-\mathbf{d})(\mathbf{H}_k-\mathbf{d})^\dagger\right]$ is a diagonal matrix.

We begin by noting that if the channels are uncorrelated, then due to Lemma 3.1 the $n_\mathrm{T} \times n_\mathrm{T}$ prediction error covariance matrix $\Sigma(1/\mathrm{SNR})$ is diagonal and the diagonal entries are given by

$$\Sigma^{(t,t)}(1/\mathrm{SNR}) = \exp\left\{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(t,t)}(\lambda) + 1/\mathrm{SNR})\,\mathrm{d}\lambda\right\} - 1/\mathrm{SNR}, \quad 1 \leq t \leq n_\mathrm{T}. \tag{7.42}$$

In this case, the minimum eigenvalue of $\Sigma(1/\mathrm{SNR})$ is just the smallest entry, *i.e.*,

$$\begin{aligned}\lambda_{\min}(1/\mathrm{SNR}) &= \epsilon^2_{\min}(1/\mathrm{SNR}) \\ &= \min_{1\leq t\leq n_\mathrm{T}} \exp\left\{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(t,t)}(\lambda) + 1/\mathrm{SNR})\,\mathrm{d}\lambda\right\} - 1/\mathrm{SNR}.\end{aligned} \tag{7.43}$$

Combining (7.43) and (7.41) we obtain the upper bound

$$C(\mathrm{SNR}) \leq \log \frac{1}{\epsilon^2_{\min}(1/\mathrm{SNR})} + \log\log \mathrm{SNR} + O(1). \tag{7.44}$$

Note that this bound is identical to the upper bound on the capacity of SISO channels where the fading process has the spectral distribution function

$$\mathsf{F}(\lambda) = \mathsf{F}^{(t_*,t_*)}(\lambda) \tag{7.45}$$

with

$$t_* = \arg\min_{1\leq t\leq n_\mathrm{T}} \exp\left\{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(t,t)}(\lambda) + 1/\mathrm{SNR})\,\mathrm{d}\lambda\right\}. \tag{7.46}$$

It follows by Corollary 5.9 that the pre-log Π can be upper bounded by

$$\begin{aligned}\Pi &\leq \mu\left(\left\{\lambda : \mathsf{F}'^{(t_*,t_*)}(\lambda) = 0\right\}\right) &(7.47)\\ &\leq \max_{1\leq t\leq n_\mathrm{T}} \mu\left(\left\{\lambda : \mathsf{F}'^{(t,t)}(\lambda) = 0\right\}\right), &(7.48)\end{aligned}$$

where the last inequality follows by maximizing over all indices t. Note that there is, in fact, equality in (7.48). Indeed, the index t_* that minimizes (7.43) maximizes the pre-log as well.

In the following we show that this upper bound can be achieved by transmitting from only one antenna. We begin by noting that transmitting from one antenna is equivalent to choosing the inputs to be of the form $\mathbf{X}_k = \hat{\mathbf{x}} \cdot \widetilde{X}_k$, where $\hat{\mathbf{x}}$ is a deterministic unit vector with entries

$$\hat{x}^{(t)} = \begin{cases} 1 & t = t_* \\ 0 & \text{otherwise} \end{cases}, \tag{7.49}$$

for a given t_*. We notice that this restriction on the inputs yields a lower bound on the capacity. In the following we show that the lower bound is tight.

For the above-mentioned choice of inputs the channel output at time k can be written as

$$\begin{aligned}Y_k &= \mathbf{H}_k^\mathsf{T}\hat{\mathbf{x}}\widetilde{X}_k + Z_k\\ &= H_k^{(t_*)}\widetilde{X}_k + Z_k, &(7.50)\end{aligned}$$

where t_* is defined as in (7.49). We observe that (7.50) describes the output of a SISO fading channel with spectral distribution function

$$\mathsf{F}(\lambda) = \mathsf{F}^{(t_*,t_*)}(\lambda). \tag{7.51}$$

Therefore, we can make use of Theorem 5.8 to further lower bound the capacity by

$$C(\text{SNR}) \geq \log \frac{1}{\epsilon_{t_*}^2(4/\text{SNR}) + \frac{2}{5}\cdot(4/\text{SNR})} + O(1) \tag{7.52}$$

with

$$\epsilon_{t_*}^2(\delta^2) = \exp\left\{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(t_*,t_*)}(\delta^2) + \delta^2)\,\mathrm{d}\lambda\right\} - \delta^2. \tag{7.53}$$

It follows again by Corollary 5.9 that the pre-log can be lower bounded by

$$\Pi \geq \mu\left(\left\{\lambda : \mathsf{F}'^{(t_*,t_*)}(\lambda) = 0\right\}\right). \tag{7.54}$$

By maximizing (7.54) over all indices $1 \leq t_* \leq n_\text{T}$ we obtain the lower bound

$$\Pi \geq \max_{1\leq t\leq n_\text{T}} \mu\left(\left\{\lambda : \mathsf{F}'^{(t,t)}(\lambda) = 0\right\}\right) \tag{7.55}$$

which coincides with the upper bound in (7.48).

7.3 The Pre-Log of MIMO Fading Channels

In this section we derive an upper bound on the capacity of MIMO fading channels. All entries in the $n_\text{R} \times n_\text{T}$ fading matrix $\mathbb{H}_k$ are assumed to be independent, *i.e.*, for the deterministic $n_\text{R} \times n_\text{T}$ mean matrix D

$$\mathsf{E}\left[(H_{k+m}^{(r,t)} - d^{(r,t)})(H_k^{(r',t')} - d^{(r',t')})^*\right] = 0, \qquad \text{for } r \neq r' \text{ and } t \neq t',\quad k,m \in \mathbb{Z}. \tag{7.56}$$

We further assume that

$$\det\left(\mathsf{E}\left[(\mathbf{H}_k^{(r)} - \mathbf{d}^{(r)})(\mathbf{H}_k^{(r)} - \mathbf{d}^{(r)})^\dagger\right]\right) \neq 0, \quad 1 \leq r \leq n_\text{R}, \quad k \in \mathbb{Z}, \tag{7.57}$$

where $\mathbf{H}_k^{(r)}$ and $\mathbf{d}^{(r)}$ denote the r-th row of the fading matrix $\mathbb{H}_k$ and the mean matrix D, respectively.

To simplify notation we define the matrix-valued spectral distribution function $\mathsf{F}(\lambda) \in \mathbb{R}^{n_\text{R}\times n_\text{T}}$ such that

$$\mathsf{E}\left[(H_{k+m}^{(r,t)} - d^{(r,t)})(H_k^{(r,t)} - d^{(r,t)})^*\right] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda}\,\mathrm{d}\mathsf{F}^{(r,t)}, \qquad k,m \in \mathbb{Z}. \tag{7.58}$$

An upper bound on the capacity of MIMO fading channels under the conditions above is presented in Theorem 5.14. In Section 7.3.2 we show the corresponding proof. Preliminarily, we introduce some notations that, hopefully, help us to make the derivations more concise.

7.3.1 Notation

In contrast to the MISO case studied in Section 7.2, here the fading at time k is expressed by a matrix $\mathbb{H}_k$ and, consequently, notation gets more laborious. Therefore, we introduce some abbreviations that.

First of all, we will write the fading matrix $\mathbb{H}_0$ as

$$\mathbb{H}_0 = \begin{pmatrix} \mathbf{H}_0^{(1)\mathsf{T}} \\ \vdots \\ \mathbf{H}_0^{(n_\mathrm{R})\mathsf{T}} \end{pmatrix}, \tag{7.59}$$

where $\mathbf{H}_0^{(r)}$ is an n_T-dimensional random vector.

Furthermore, we introduce the vector

$$\widetilde{\mathbf{Y}}_0 = (\mathbb{H}_0 - \mathsf{D})\mathbf{x}_0 \tag{7.60}$$

with

$$\widetilde{Y}_0^{(r)} = (\mathbf{H}_0^{(r)} - \mathbf{d}^{(r)})^\mathsf{T}\mathbf{x}_0, \tag{7.61}$$

where $\mathbf{d}^{(r)}$ is the mean vector corresponding to $\mathbf{H}_0^{(r)}$, *i.e.*, $\mathbf{d} = \mathsf{E}\Big[\mathbf{H}_0^{(r)}\Big]$. It follows that the entries in the covariance matrix of $\widetilde{\mathbf{Y}}_0$ are given by

$$\begin{aligned} \mathsf{E}\Big[\widetilde{Y}_0^{(r)}\widetilde{Y}_0^{(l)*} \Big| \mathbf{X}_0\Big] &= \mathbf{x}_0^\mathsf{T}\mathsf{E}\Big[(\mathbf{H}_0^{(r)} - \mathbf{d}^{(r)})(\mathbf{H}_0^{(l)} - \mathbf{d}^{(l)})^\dagger\Big]\mathbf{x}_0^* \\ &= \|\mathbf{x}_0\|^2\hat{\mathbf{x}}_0^\mathsf{T}\mathsf{E}\Big[(\mathbf{H}_0^{(r)} - \mathbf{d}^{(r)})(\mathbf{H}_0^{(l)} - \mathbf{d}^{(l)})^\dagger\Big]\hat{\mathbf{x}}_0^* \end{aligned} \tag{7.62}$$

with $\|\hat{\mathbf{x}}_0\| = 1$. We define the $n_\mathrm{R} \times n_\mathrm{R}$ matrix $\mathsf{R}_{\widetilde{Y}\widetilde{Y}}$ such that

$$\mathsf{R}_{\widetilde{Y}\widetilde{Y}}^{(r,l)} = \frac{\mathsf{E}\Big[\widetilde{Y}_0^{(r)}\widetilde{Y}_0^{(l)*} \Big| \mathbf{X}_0\Big]}{\|\mathbf{x}_0\|^2} = \hat{\mathbf{x}}_0^\mathsf{T}\mathsf{E}\Big[(\mathbf{H}_0^{(r)} - \mathbf{d}^{(r)})(\mathbf{H}_0^{(l)} - \mathbf{d}^{(l)})^\dagger\Big]\hat{\mathbf{x}}_0^*. \tag{7.63}$$

Note that since the channels are uncorrelated the matrix $\mathsf{R}_{\widetilde{Y}\widetilde{Y}}$ is diagonal. We further note that the covariance matrix $\mathsf{E}\Big[(\mathbf{H}_0^{(r)} - \mathbf{d}^{(r)})(\mathbf{H}_0^{(r)} - \mathbf{d}^{(r)})^\dagger\Big]$ is always non-negative definite for $1 \leq r \leq n_\mathrm{R}$. It follows by (7.57) that $\mathsf{E}\Big[(\mathbf{H}_0^{(r)} - \mathbf{d}^{(r)})(\mathbf{H}_0^{(r)} - \mathbf{d}^{(r)})^\dagger\Big]$ is even positive definite and the diagonal entries in $\mathsf{R}_{\widetilde{Y}\widetilde{Y}}$ are strictly larger than zero.

As in Section 7.2 when dealing with MISO channels, we may express the fading matrix $\mathbb{H}_0$ as

$$\mathbb{H}_0 = \overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0, \tag{7.64}$$

where $\overline{\mathbb{H}}_0$ is the best estimate of the fading matrix given a noisy observation of its past. For given past values $\mathsf{H}_{-1}+\mathsf{W}_{-1}, \mathsf{H}_{-2}+\mathsf{W}_{-2}, \ldots$ the estimate $\overline{\mathsf{H}}_0$ is given by

$$\overline{\mathsf{H}}_0 = \mathsf{E}\left[\mathbb{H}_0 \,\middle|\, \{\mathbb{H}_\nu + \mathbb{W}_\nu\}_{\nu=-\infty}^{-1}\right], \tag{7.65}$$

with

$$W_\nu^{(r,t)} \perp\!\!\!\perp W_\nu^{(r',t')}, \qquad \text{for } r \neq r' \text{ and } t \neq t', \qquad \forall \nu \tag{7.66}$$

and $W_\nu^{(r,t)}$ i.i.d. $\mathcal{N}\left(0, \delta^2\right)$, for a given δ^2. The entries of the matrix $\widetilde{\mathbb{H}}_0$ are independently distributed with $\widetilde{H}_0^{(r,t)} \sim \mathcal{N}\left(0, \epsilon^2_{(r,t)}(\delta^2)\right)$, where

$$\epsilon^2_{(r,t)}(\delta^2) = \exp\left\{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(r,t)}(\lambda) + \delta^2)\, \mathrm{d}\lambda\right\} - \delta^2, 1 \leq r \leq n_\mathrm{R}, \quad 1 \leq t \leq n_\mathrm{T} \tag{7.67}$$

is the minimum mean squared error in prediction the fading $H_0^{(r,t)}$ based on a noisy observation of its past. This follows from Lemma 3.1 by stacking the $n_\mathrm{R} \times n_\mathrm{T}$ components of $\mathbb{H}_0$ into one vector.

To express the covariance matrix of $\widetilde{\mathbb{H}}_0 \mathbf{x}_0$ we may write $\widetilde{\mathbb{H}}_0$ in terms of n_R n_T-dimensional vectors, *i.e.*,

$$\widetilde{\mathbb{H}}_0 = \begin{pmatrix} \widetilde{\mathbf{H}}_0^{(1)\mathsf{T}} \\ \vdots \\ \widetilde{\mathbf{H}}_0^{(n_\mathrm{R})\mathsf{T}} \end{pmatrix}. \tag{7.68}$$

Then it follows

$$\begin{aligned} &\mathsf{E}\left[(\widetilde{\mathbf{H}}^{(r)\mathsf{T}}\mathbf{X}_0)(\widetilde{\mathbf{H}}^{(l)\mathsf{T}}\mathbf{X}_0)^* \,\middle|\, \mathbf{X}_0\right] \\ &\quad = \mathbf{x}_0^\mathsf{T} \mathsf{E}\left[\widetilde{\mathbf{H}}_0^{(r)} \widetilde{\mathbf{H}}_0^{(l)\dagger}\right] \mathbf{x}_0^* \\ &\quad = \|\mathbf{x}_0\|^2 \hat{\mathbf{x}}_0^\mathsf{T} \mathsf{E}\left[\widetilde{\mathbf{H}}_0^{(r)} \widetilde{\mathbf{H}}_0^{(l)\dagger}\right] \hat{\mathbf{x}}_0^* \end{aligned} \tag{7.69}$$

with $\|\hat{\mathbf{x}}_0\| = 1$. We define the $n_\mathrm{R} \times n_\mathrm{R}$ matrix $\mathsf{R}_{\epsilon\epsilon}$ such that

$$\mathsf{R}_{\epsilon\epsilon}^{(r,l)} = \frac{\mathsf{E}\left[(\widetilde{\mathbf{H}}^{(r)\mathsf{T}}\mathbf{X}_0)(\widetilde{\mathbf{H}}^{(l)\mathsf{T}}\mathbf{X}_0)^* \,\middle|\, \mathbf{X}_0\right]}{\|\mathbf{x}_0\|^2} = \hat{\mathbf{x}}_0^\mathsf{T} \mathsf{E}\left[\widetilde{\mathbf{H}}^{(r)} \widetilde{\mathbf{H}}^{(l)\dagger}\right] \hat{\mathbf{x}}_0^*. \tag{7.70}$$

Since the channels are uncorrelated, *i.e.*,

$$\mathsf{E}\left[\widetilde{\mathbf{H}}_0^{(r)} \widetilde{\mathbf{H}}_0^{(l)\dagger}\right] = 0, \qquad \text{for } r \neq l, \tag{7.71}$$

the matrix $\mathsf{R}_{\epsilon\epsilon}$ is diagonal. Note that $\mathsf{E}\left[\widetilde{\mathbf{H}}_0^{(r)} \widetilde{\mathbf{H}}_0^{(r)\dagger}\right]$ corresponds to the prediction error covariance matrix of a MISO channel arising from the MIMO channel by considering only the r-th receive antenna. Therefore, we may often write $\Sigma_r(\delta^2)$ instead of $\mathsf{E}\left[\widetilde{\mathbf{H}}^{(r)}\widetilde{\mathbf{H}}^{(r)\dagger}\right]$, where δ^2 is as above. Furthermore, note that $\Sigma_r(\delta^2)$ is a diagonal matrix with diagonal entries

$$\Sigma_r^{(t,t)}(\delta^2) = \exp\left\{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(r,t)}(\lambda)+\delta^2)\,\mathrm{d}\lambda\right\} - \delta^2, 1 \le r \le n_{\mathrm{R}}, \quad 1 \le t \le n_{\mathrm{T}}. \tag{7.72}$$

7.3.2 A Proof of Theorem 5.14

In the following we derive an upper bound on the capacity of MIMO fading channels. The derivation is very similar to that in the MISO case.

We begin by using the chain rule

$$I(\mathbf{X}^n;\mathbf{Y}^n) = \sum_{k=1}^{n} I(\mathbf{Y}_k;\mathbf{X}^n|\mathbf{Y}^{k-1}) \tag{7.73}$$

and upper bounding each of the terms in the sum by

$$\begin{aligned}
&I(\mathbf{Y}_k;\mathbf{X}^n|\mathbf{Y}^{k-1}) \\
&\quad = I(\mathbf{Y}_k;\mathbf{X}^n,\mathbf{Y}^{k-1}) - I(\mathbf{Y}_k;\mathbf{Y}^{k-1}) \\
&\quad \le I(\mathbf{Y}_k;\mathbf{X}^n,\mathbf{Y}^{k-1}) \\
&\quad = I(\mathbf{Y}_k;\mathbf{X}^k,\mathbf{Y}^{k-1}) \\
&\quad = I(\mathbf{Y}_k;\mathbf{X}_k) + I(\mathbf{Y}_k;\mathbf{X}^{k-1},\mathbf{Y}^{k-1}|\mathbf{X}_k) \\
&\quad \le \sup_{p_{\mathbf{X}_k}} I(\mathbf{Y}_k;\mathbf{X}_k) + \sup_{p_{\mathbf{X}^k}} I(\mathbf{Y}_k;\mathbf{X}^{k-1},\mathbf{Y}^{k-1}|\mathbf{X}_k) \\
&\quad \le \sup_{p_{\mathbf{X}_k}} I(\mathbf{Y}_k;\mathbf{X}_k) + \sup_{p_{\mathbf{X}^k}} I\left(\mathbf{Y}_k;\{\mathbb{H}_\nu+\mathbb{W}_\nu\}_{\nu=1}^{k-1}\middle|\,\mathbf{X}_k\right) && (7.74)\\
&\quad \le \sup_{p_{\mathbf{X}_0}} I(\mathbf{Y}_0;\mathbf{X}_0) + \sup_{p_{\mathbf{X}_0}} I\left(\mathbf{Y}_0;\{\mathbb{H}_\nu+\mathbb{W}_\nu\}_{\nu=-\infty}^{-1}\middle|\,\mathbf{X}_0\right) \\
&\quad \le \sup_{p_{\mathbf{X}_0}} I(\mathbf{Y}_0;\mathbf{X}_0) + \sup_{\|\mathbf{x}_0\|\le\mathsf{A}} I\left(\mathbf{Y}_0;\{\mathbb{H}_\nu+\mathbb{W}_\nu\}_{\nu=-\infty}^{-1}\middle|\,\mathbf{X}_0=\mathbf{x}_0\right) && (7.75)
\end{aligned}$$

with

$$W_\nu^{(r,t)} \perp\!\!\!\perp W_\nu^{(r',t')}, \qquad \text{for } r \ne r' \text{ and } t \ne t' \tag{7.76}$$

and

$$W_\nu^{(r,t)} \sim \mathcal{N}\left(0, \frac{\sigma^2}{\mathsf{A}^2}\right), \qquad \forall \nu. \tag{7.77}$$

Here, the first equality follows from the chain rule; the subsequent inequality from the non-negativity of mutual information; the next equality from the absence of feedback which results in future inputs being independent of the present output given the present input and the past inputs and outputs; the subsequent equality from the chain rule; and the following inequality from maximizing over all input distributions $p_{\mathbf{X}^k}$ satisfying $\|\mathbf{X}\| \leq \mathsf{A}$.

Inequality (7.74) follows by assuming that the pairs $(\mathbf{X}^{k-1}, \mathbf{Y}^{k-1})$ provide information about the entire fading matrix $\mathbb{H}^{k-1}$ and by violating the peak-power constraint, which both yield an upper bound[2]; the subsequent inequality follows from stationarity and by adding information; and the last inequality from the expression for the conditional mutual information and from the fact that the expectation of a random variable can never be larger than its largest value.

To evaluate the RHS of (7.75) we first note that the first term corresponds to the memoryless case and is given by [1]

$$\sup_{p_{\mathbf{X}_0}} I(\mathbf{Y}_0; \mathbf{X}_0) = \chi_{\text{i.i.d.}}(\mathbb{H}_0) + \log\log \text{SNR} + o(1), \tag{7.78}$$

where $\chi_{\text{i.i.d.}}(\mathbb{H}_0)$ denotes the fading number. Remember that the expression for the fading number holds for both average-power and peak-power constraint [1].

To evaluate the second term on the RHS of (7.75), we express the fading $\mathbb{H}_0$ as

$$\mathbb{H}_0 = \overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0, \tag{7.79}$$

where $\overline{\mathbb{H}}_0$ is the best estimate of the fading $\mathbb{H}_0$ (7.65). We obtain

$$\begin{aligned} I &\left(\mathbf{Y}_0; \{\mathbb{H}_\nu + \mathbb{W}_\nu\}_{\nu=-\infty}^{-1} \middle| \mathbf{X}_0 = \mathbf{x}_0\right) \\ &= I\left((\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0)\mathbf{X}_0 + \mathbf{Z}_0; \{\mathbb{H}_\nu + \mathbb{W}_\nu\}_{\nu=-\infty}^{-1} \middle| \mathbf{X}_0 = \mathbf{x}_0\right) \\ &= h\left((\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0)\mathbf{X}_0 + \mathbf{Z}_0 \middle| \mathbf{X}_0 = \mathbf{x}_0\right) \\ &\quad - h\left((\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0)\mathbf{X}_0 + \mathbf{Z}_0 \middle| \{\mathbb{H}_\nu + \mathbb{W}_\nu\}_{\nu=-\infty}^{-1}, \mathbf{X}_0 = \mathbf{x}_0\right), \end{aligned} \tag{7.80}$$

where the first equality follows by expressing the channel output $\mathbf{Y}_0$ in terms of the input $\mathbf{x}_0$, the fading $\mathbb{H}_0$, and the noise $\mathbf{Z}_0$; and the subsequent equality from the definition of the mutual information.

To compute the first term on the RHS of (7.80) we note that the random variable $\left((\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}})\mathbf{X}_0 + Z_0 \middle| \mathbf{X}_0 = \mathbf{x}_0\right)$ has a Gaussian distribution

[2] The reasoning is identical to that in Section 7.2 when dealing with MISO fading channels. For convenience, we are doing here without any details and refer to Section 7.2 for a more elaborate explanation.

with mean

$$\mathsf{E}\Big[(\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0)\mathbf{x}_0 + \mathbf{Z}_0\Big] = \mathsf{D}\mathbf{x}_0 \tag{7.81}$$

and covariance matrix

$$\mathsf{K} = \|\mathbf{x}_0\|^2 \mathsf{R}_{\widetilde{Y}\widetilde{Y}} + \sigma^2 \mathsf{I}. \tag{7.82}$$

This follows from the definition of $\mathsf{R}_{\widetilde{Y}\widetilde{Y}}$ (7.63) and from the expression for the covariance matrix. Indeed, the (r, l)-th entry of the covariance matrix K is given by

$$\begin{aligned}
&\mathsf{E}\Big[\Big((\overline{\mathbf{H}}_0^{(r)} + \widetilde{\mathbf{H}}_0^{(r)})^{\mathsf{T}}\mathbf{x}_0 + Z_0^{(r)} - \mathbf{d}^{(r)\mathsf{T}}\mathbf{x}_0\Big) \\
&\quad \cdot \Big((\overline{\mathbf{H}}_0^{(l)} + \widetilde{\mathbf{H}}_0^{(l)})^{\mathsf{T}}\mathbf{x}_0 + Z_0^{(l)} - \mathbf{d}^{(l)\mathsf{T}}\mathbf{x}_0\Big)^*\Big] \\
&\quad = \mathsf{E}\Big[\Big((\overline{\mathbf{H}}_0^{(r)} + \widetilde{\mathbf{H}}_0^{(r)} - \mathbf{d}^{(r)})^{\mathsf{T}}\mathbf{x}_0\Big)\Big((\overline{\mathbf{H}}_0^{(l)} + \widetilde{\mathbf{H}}_0^{(l)} - \mathbf{d}^{(l)})^{\mathsf{T}}\mathbf{x}_0\Big)^*\Big] \\
&\qquad + \mathsf{E}\Big[Z_0^{(r)} Z_0^{(l)*}\Big] \\
&\quad = \mathbf{x}_0^{\mathsf{T}}\mathsf{E}\Big[\Big(\mathbf{H}_0^{(r)} - \mathbf{d}^{(r)}\Big)\Big(\mathbf{H}_0^{(l)} - \mathbf{d}^{(l)}\Big)^{\dagger}\Big]\mathbf{x}_0^* + \mathsf{E}\Big[Z_0^{(r)} Z_0^{(l)*}\Big] \\
&\quad = \|\mathbf{x}_0\|^2 \mathcal{R}_{\widetilde{Y}\widetilde{Y}}^{(r,l)} + \sigma^2 \delta_{r-l},
\end{aligned} \tag{7.83}$$

where $\mathbf{d}^{(r)}$ denotes the mean vector corresponding to $\mathbf{H}_0^{(r)}$ and δ_{r-l} denominates the Kronecker delta. Here, the first equality follows from the fact that the additive noise $\mathbf{Z}_0$ is zero-mean and independent from the fading matrix $\mathbb{H}_0$.

Using the expression for differential entropy of a multivariate Gaussian random variable [5] we obtain

$$\begin{aligned}
&h\Big((\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0)\mathbf{X}_0 + \mathbf{Z}_0 \Big| \mathbf{X}_0 = \mathbf{x}_0\Big) \\
&\quad = n_{\mathrm{R}} \log \pi + n_{\mathrm{R}} + \log\det\left(\|\mathbf{x}_0\|^2 \mathsf{R}_{\widetilde{Y}\widetilde{Y}} + \sigma^2 \mathsf{I}\right) \\
&\quad = n_{\mathrm{R}} \log \pi + n_{\mathrm{R}} + \sum_{r=1}^{n_{\mathrm{R}}} \log\Big(\|\mathbf{x}_0\|^2 \mathcal{R}_{\widetilde{Y}\widetilde{Y}}^{(r,r)} + \sigma^2\Big),
\end{aligned} \tag{7.84}$$

where the second equality follows by noting that $\mathsf{R}_{\widetilde{Y}\widetilde{Y}}$ is a diagonal matrix (see Section 7.3.1).

To compute the second term on the RHS of (7.80) we note that for given values $\mathbf{x}_0$ and $\{\mathsf{H}_\nu + \mathsf{W}_\nu\}_{\nu=-\infty}^{-1}$ the random variable $\Big((\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0)\mathbf{X}_0 + \mathbf{Z}_0 \Big| \mathbb{H}_{-1} + \mathbb{W}_{-1} = \mathsf{H}_{-1} + \mathsf{W}_{-1}, \ldots, \mathbf{X}_0 = \mathbf{x}_0\Big)$ has a Gaussian distribution with mean

$$\mathsf{E}\Big[(\overline{\mathsf{H}}_0 + \widetilde{\mathbb{H}}_0)\mathbf{X}_0 + \mathbf{Z}_0 \;\Big|\; \{\mathbb{H}_\nu + \mathbb{W}_\nu\}_{\nu=-\infty}^{-1}, \mathbf{X}_0\Big] = \big(\mathsf{D} + \overline{\mathsf{H}}_0\big)\,\mathbf{x}_0 \tag{7.85}$$

and covariance matrix

$$\mathsf{K} = \|\mathbf{x}_0\|^2 \mathsf{R}_{\epsilon\epsilon} + \sigma^2 \mathsf{I}. \tag{7.86}$$

This can be easily verified by an analysis similar to that in (7.83).

Using the expression for differential entropy of a multivariate Gaussian random variable we obtain

$$\begin{aligned} h&\left((\overline{\mathbb{H}}_0 + \widetilde{\mathbb{H}}_0)\mathbf{X}_0 + \mathbf{Z}_0 \middle| \{\mathbb{H}_\nu + \mathbb{W}_\nu\}_{\nu=-\infty}^{-1}, \mathbf{X}_0 = \mathbf{x}_0 \right) \\ &= n_\mathrm{R} \log \pi + n_\mathrm{R} + \log\det\left(\|\mathbf{x}_0\|^2 \mathsf{R}_{\epsilon\epsilon} + \sigma^2 \mathsf{I}\right) \\ &= n_\mathrm{R} \log \pi + n_\mathrm{R} + \sum_{r=1}^{n_\mathrm{R}} \log\left(\|\mathbf{x}_0\|^2 \mathcal{R}_{\epsilon\epsilon}^{(r,r)} + \sigma^2\right), \end{aligned} \tag{7.87}$$

where we used the fact, that $\mathsf{R}_{\epsilon\epsilon}$ is a diagonal matrix.

Combining (7.87), (7.84), (7.80), (7.78), and (7.75) and using the expression for channel capacity, we obtain the upper bound

$$\begin{aligned} C(\mathrm{SNR}) &\le \sup_{\|\mathbf{x}_0\| \le \mathsf{A}} \left\{ \sum_{r=1}^{n_\mathrm{R}} \log \frac{\|\mathbf{x}_0\|^2 \mathcal{R}_{\widetilde{Y}\widetilde{Y}}^{(r,r)} + \sigma^2}{\|\mathbf{x}_0\|^2 \mathcal{R}_{\epsilon\epsilon}^{(r,r)} + \sigma^2} \right\} + \log\log \mathrm{SNR} + O(1) \\ &\le \sup_{\|\hat{\mathbf{x}}\|=1} \left\{ \sum_{r=1}^{n_\mathrm{R}} \log \frac{\mathcal{R}_{\widetilde{Y}\widetilde{Y}}^{(r,r)} + \frac{\sigma^2}{\mathsf{A}^2}}{\mathcal{R}_{\epsilon\epsilon}^{(r,r)} + \frac{\sigma^2}{\mathsf{A}^2}} \right\} + \log\log \mathrm{SNR} + O(1), \end{aligned} \tag{7.88}$$

where the last inequality follows by noting that[3] $\mathsf{R}_{\widetilde{Y}\widetilde{Y}}^{(r,r)} > \mathsf{R}_{\epsilon\epsilon}^{(r,r)}$ in which case the RHS of (7.88) is monotonically increasing in $\|\mathbf{x}_0\|^2$, and by choosing $\|\mathbf{x}_0\|^2$ as large as possible, *i.e.*, $\|\mathbf{x}_0\|^2 = \mathsf{A}^2$. It should be noted that $\mathsf{R}_{\widetilde{Y}\widetilde{Y}}$ as well as $\mathsf{R}_{\epsilon\epsilon}$ depend on $\mathbf{x}_0$.

As mentioned in Section 7.3.1, we may write $\mathcal{R}_{\epsilon\epsilon}^{(r,r)}$ as

$$\mathcal{R}_{\epsilon\epsilon}^{(r,r)} = \hat{\mathbf{x}}^\mathsf{T} \Sigma_r(1/\mathrm{SNR}) \hat{\mathbf{x}}^*, \tag{7.89}$$

where $\Sigma_r(1/\mathrm{SNR})$ is an $n_\mathrm{T} \times n_\mathrm{T}$ diagonal matrix with entries

$$\Sigma_r^{(t,t)}(1/\mathrm{SNR}) = \exp\left\{ \int_{-1/2}^{1/2} \log(\mathsf{F}'^{(r,t)}(\lambda) + 1/\mathrm{SNR})\,\mathrm{d}\lambda \right\} - 1/\mathrm{SNR}. \tag{7.90}$$

It follows that

$$C(\mathrm{SNR}) \le \sup_{\|\hat{\mathbf{x}}\|=1} \left\{ \sum_{r=1}^{n_\mathrm{R}} \log \frac{1}{\hat{\mathbf{x}}^\mathsf{T} \Sigma_r(1/\mathrm{SNR}) \hat{\mathbf{x}}^* + 1/\mathrm{SNR}} \right\} + \log\log \mathrm{SNR} + O(1). \tag{7.91}$$

This concludes the proof.

[3] The reasoning is identical to that in Section 6.1 when considering regular fading processes. We therefore refer to Section 6.1 for a more elaborate explanation.

7.3.3 A Proof of Corollary 5.15

To derive an upper bound on the capacity pre-log we begin by considering (7.91) and minimizing

$$\sum_{r=1}^{n_{\mathrm{R}}} \log\left(\hat{\mathbf{x}}^{\mathsf{T}}\Sigma_r(1/\mathrm{SNR})\hat{\mathbf{x}}^{*} + 1/\mathrm{SNR}\right) \tag{7.92}$$

over all unit-vectors $\hat{\mathbf{x}}$. We lower bound each of the terms in the sum by

$$\begin{aligned}
&\log\left(\hat{\mathbf{x}}^{\mathsf{T}}\Sigma_r(\delta^2)\hat{\mathbf{x}}^{*} + \delta^2\right)\\
&\quad = \log\left(\sum_{t=1}^{n_{\mathrm{T}}} |\hat{x}^{(t)}|^2 \Sigma_r^{(t,t)}(\delta^2) + \delta^2\right)\\
&\quad = \log\left(\sum_{t=1}^{n_{\mathrm{T}}} |\hat{x}^{(t)}|^2 \left(\Sigma_r^{(t,t)}(\delta^2) + \delta^2\right)\right)\\
&\quad = \log\left(\sum_{t=1}^{n_{\mathrm{T}}} |\hat{x}^{(t)}|^2 \exp\left\{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(r,t)}(\lambda) + \delta^2)\,\mathrm{d}\lambda\right\}\right)\\
&\quad \geq \sum_{t=1}^{n_{\mathrm{T}}} |\hat{x}^{(t)}|^2 \log\left(\exp\left\{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(r,t)}(\lambda) + \delta^2)\,\mathrm{d}\lambda\right\}\right)\\
&\quad = \sum_{t=1}^{n_{\mathrm{T}}} \left(|\hat{x}^{(t)}|^2 \int_{-1/2}^{1/2} \log(\mathsf{F}'^{(r,t)}(\lambda) + \delta^2)\,\mathrm{d}\lambda\right),
\end{aligned} \tag{7.93}$$

with $\delta^2 = 1/\mathrm{SNR}$. Here, the first equality follows because $\Sigma_r(\delta^2)$ is a diagonal matrix; the subsequent equality from the fact that $\hat{\mathbf{x}}$ is a unit-vector; the next equality from (7.90); the following inequality by noting that $\hat{\mathbf{x}}$ is a unit vector and using Jensen's inequality; and the last equality by taking the logarithm of the exponential function.

Then, the capacity pre-log Π can be upper bounded by

$$\begin{aligned}
\Pi &= \varlimsup_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log \mathrm{SNR}}\\
&\leq \varlimsup_{\mathrm{SNR}\to\infty} \sup_{\|\hat{\mathbf{x}}\|=1} \sum_{r=1}^{n_{\mathrm{R}}} \frac{\log \frac{1}{\hat{\mathbf{x}}^{\mathsf{T}}\Sigma_r(1/\mathrm{SNR})\hat{\mathbf{x}}^{*}+1/\mathrm{SNR}}}{\log \mathrm{SNR}}\\
&= \varlimsup_{\delta^2\downarrow 0} \sup_{\|\hat{\mathbf{x}}\|=1} \sum_{r=1}^{n_{\mathrm{R}}} \frac{\log \frac{1}{\hat{\mathbf{x}}^{\mathsf{T}}\Sigma_r(\delta^2)\hat{\mathbf{x}}^{*}+\delta^2}}{\log \frac{1}{\delta^2}}\\
&= \varlimsup_{\delta^2\downarrow 0} \sup_{\|\hat{\mathbf{x}}\|=1} \sum_{r=1}^{n_{\mathrm{R}}} \frac{\log\left(\hat{\mathbf{x}}^{\mathsf{T}}\Sigma_r(\delta^2)\hat{\mathbf{x}}^{*}+\delta^2\right)}{\log \delta^2}\\
&\leq \lim_{\delta^2\downarrow 0} \sup_{\|\hat{\mathbf{x}}\|=1} \sum_{r=1}^{n_{\mathrm{R}}} \sum_{t=1}^{n_{\mathrm{T}}} |\hat{x}^{(t)}|^2 \frac{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(r,t)}(\lambda) + \delta^2)\,\mathrm{d}\lambda}{\log \delta^2}
\end{aligned}$$

$$
\begin{aligned}
&= \lim_{\delta^2 \downarrow 0} \sup_{\|\hat{\mathbf{x}}\|=1} \sum_{t=1}^{n_\mathrm{T}} |\hat{x}^{(t)}|^2 \sum_{r=1}^{n_\mathrm{R}} \frac{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(r,t)}(\lambda) + \delta^2)\,\mathrm{d}\lambda}{\log \delta^2} \\
&= \lim_{\delta^2 \downarrow 0} \max_{1 \le t \le n_\mathrm{T}} \sum_{r=1}^{n_\mathrm{R}} \frac{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(r,t)}(\lambda) + \delta^2)\,\mathrm{d}\lambda}{\log \delta^2} \\
&= \max_{1 \le t \le n_\mathrm{T}} \lim_{\delta^2 \downarrow 0} \sum_{r=1}^{n_\mathrm{R}} \frac{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(r,t)}(\lambda) + \delta^2)\,\mathrm{d}\lambda}{\log \delta^2} \qquad (7.94) \\
&= \max_{1 \le t \le n_\mathrm{T}} \sum_{r=1}^{n_\mathrm{R}} \lim_{\delta^2 \downarrow 0} \frac{\int_{-1/2}^{1/2} \log(\mathsf{F}'^{(r,t)}(\lambda) + \delta^2)\,\mathrm{d}\lambda}{\log \delta^2} \\
&= \max_{1 \le t \le n_\mathrm{T}} \sum_{r=1}^{n_\mathrm{R}} \mu\left(\left\{\lambda : \mathsf{F}'^{(r,t)}(\lambda) = 0\right\}\right). \qquad (7.95)
\end{aligned}
$$

Here, the first equality follows from the definition of the capacity pre-log; the subsequent inequality by upper bounding the capacity by (7.91); the next equality by substituting $\delta^2 = 1/\mathrm{SNR}$; the following equality from the behavior of the logarithm; the subsequent inequality from (7.93); the following equality by interchanging both sums; and the next equality because $\hat{\mathbf{x}}$ is a unit vector and therefore the upper bound is maximized by choosing just the largest term in the sum.

Equality (7.94) follows by noting that for each index t the sum on the RHS of (7.94) converges to $\sum_{r=1}^{n_\mathrm{R}} \mu\left(\left\{\lambda : \mathsf{F}'^{(r,t)}(\lambda) = 0\right\}\right)$ as δ^2 tends to zero. Thus, it converges pointwise on the set $\mathcal{T} = \{1, \ldots, n_\mathrm{T}\}$. It follows that since the set $\mathcal{T}$ is finite pointwise convergence implies uniform convergence in which case we are allowed to interchange the maximization with the limit [10].

The last two equalities follow by taking the limit into the sum; and from the expression for the capacity pre-log of SISO channels [2]. This concludes the proof.

Note that $\mu(\cdot)$ cannot be larger than 1. Consequently, the upper bound on the capacity pre-log given in (7.95) is always upper bounded by the number of receive antennas n_R, *i.e.*,

$$
\Pi \le \max_{1 \le t \le n_\mathrm{T}} \sum_{r=1}^{n_\mathrm{R}} \mu\left(\left\{\lambda : \mathsf{F}'^{(r,t)}(\lambda) = 0\right\}\right) = \Pi_\mathrm{U} \le n_\mathrm{R}, \qquad (7.96)
$$

where the equality should be taken as a definition. It follows that if the number of transmit antennas n_T is larger than the number of receive antennas n_R, then the upper bound Π_U obtained in this thesis is always upper bounded by the trivial upper bound $\min\{n_{\mathrm{T}}, n_{\mathrm{R}}\}$ arising from the expression for channel capacity in presence of perfect side information at the receiver, *i.e.*,

$$\Pi_{\mathrm{U}} \leq n_{\mathrm{R}} = \min\{n_{\mathrm{T}}, n_{\mathrm{R}}\}. \tag{7.97}$$

Chapter 8

Discussion and Open Problems

In this chapter the results presented in Chapter 5 are discussed. To the best of our knowledge, the only studies that adress our channel model without any simplifications are by Lapidoth and Moser [1] and by Lapidoth [2]. Thus, we relate the results obtained in this thesis mostly to the studies in [1] and [2]. Additionally, we compare our results with results corresponding to the case, where the receiver has knowledge of the channel realization. This is particularly interesting, if the fading process is non-regular, *i.e.*, the channel state can be predicted precisely from its past.

The chapter concludes with a listing of open questions emanating from the study in this thesis.

8.1 Discussion

In the following, we discuss the results obtained in this thesis and relate them to previous work. In contrast to Chapters 5, 6, and 7 we do not discuss regular and non-regular processes separately. Instead, it seems more reasonable to divide the discussion into the following three parts: in Section 8.1.1 the results corresponding to single-input single-output (SISO) fading channels are discussed; in Section 8.1.2 we study the multiple-input single-output (MISO) case; and in Section 8.1.3 the multiple-input multiple-output (MIMO) case.

8.1.1 SISO Fading Channels

In order to show the relevance of the obtained results, we give a review of related studies.

The capacity of fading channels where neither transmitter nor receiver has access to the fading process, though both are fully cognizant of the fading laws, was recently studied by Lapidoth and Moser [1]. They show that if the fading process is regular, then capacity increases double-logarithmically with the signal-to-noise ratio (SNR), *i.e.*,

$$C(\mathrm{SNR}) = \log\log \mathrm{SNR} + \chi(\{\mathbb{H}_k\}) + o(1), \tag{8.1}$$

where $\chi(\{\mathbb{H}_k\})$ denotes the fading number. This is in stark contrast to the case where the receiver has perfect side information, in which case capacity is given by [11]

$$C_{\mathrm{PSI}}(\mathrm{SNR}) = \log \mathrm{SNR} - \gamma + o(1). \tag{8.2}$$

To bridge the gap between the double-logarithmic and the logarithmic behavior Lapidoth also considered non-regular fading processes [2]. He showed that in this case the asymptotic dependence of channel capacity on the SNR can be logarithmically, double-logarithmically, or in between, *e.g.*, as a fractional power of the logarithm of the SNR. However, Lapidoth did not present a lower bound that is tight, when capacity grows double-logarithmically in the SNR.

The lower bound presented in Section 5.2.2 is tight for this case and allows for an analysis of the capacity pre-log-log. Thus, it completes the study in [2] and may help us to better understand channel capacity at high SNR.

8.1.2 MISO Fading Channels

The capacity of MISO fading channels with memory was studied in this thesis for both, regular and non-regular fading processes. For regular fading, an upper bound on the fading number was derived. This bound is tight if the channels are uncorrelated and if the fading is either zero-mean or its spectral density matrix contains identical entries.

In the following we compare the upper bound obtained in this thesis with the upper bound presented in [1], *i.e.*,

$$\chi(\{\mathbf{H}_k\}) \leq \chi_{\mathrm{i.i.d}}(\mathbf{H}_0) + I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) = \chi_{\mathrm{LM}}, \tag{8.3}$$

where the equality should be taken as a definition. The fading number is invariant under multiplication of $\mathbf{H}_k$ with a deterministic nonsingular matrix. Therefore, we can assume without loss of generality that the fading has the covariance matrix $\mathsf{K} = \mathsf{I}$, where I is the identity matrix. Then, the second term in (8.3) is given by

$$I(\mathbf{H}_0; \mathbf{H}_{-\infty}^{-1}) = \log \frac{1}{\det \Sigma} = \sum_{t=1}^{n_{\mathrm{T}}} \log \frac{1}{\lambda_t}, \tag{8.4}$$

where Σ denotes the prediction error covariance matrix in predicting the present fading from its past and λ_t denominates the t-th eigenvalue of Σ. Furthermore, we can write our upper bound χ_{U} given in Theorem 5.5 as

$$\chi(\{\mathbf{H}_k\}) \leq \chi_{\mathrm{i.i.d}}(\mathbf{H}_0) + \log \frac{1}{\lambda_{\min}} = \chi_{\mathrm{U}}. \tag{8.5}$$

It follows by noting that $\lambda_t > 0$, $1 \leq t \leq n_{\mathrm{T}}$,

$$\chi_{\mathrm{U}} \leq \chi_{\mathrm{LM}}, \tag{8.6}$$

thus, the upper bound on the fading number obtained in this thesis is at least as tight as the upper bound presented in [1].

For the case of non-regular fading processes, we derived an upper bound on channel capacity that is tight when the channels are uncorrelated. It can be viewed as an extension to the study in [2].

We note that the capacities on which we were able to find tight upper bounds can be achieved by beam forming, *i.e.*, by inputs of the form $\mathbf{X}_k = \hat{\mathbf{x}} \cdot \widetilde{X}_k$, where $\hat{\mathbf{x}}$ is a deterministic unit vector. The optimal beam direction $\hat{\mathbf{x}}$ is either a function of the specular component $\mathbf{d}$ or a function of the spectrum, depending on the channel characteristic.

In general, the capacity of MISO fading channels depends on the specular component $\widetilde{d} = \mathbf{d}^{\mathsf{T}}\hat{\mathbf{x}}$ as well as on the prediction error $\epsilon^2_{\mathrm{MSE}}$ in predicting the present fading from its past. Both parameters $\widetilde{d}$ and $\epsilon^2_{\mathrm{MSE}}$ are influenced by the choice of the beam direction $\hat{\mathbf{x}}$. Thus, when choosing $\hat{\mathbf{x}}$ we have to trade off between maximizing $\widetilde{d}$ and minimizing $\epsilon^2_{\mathrm{MSE}}$.

We note that $\epsilon^2_{\mathrm{MSE}}$ is minimized by transmitting from only one antenna. Thus, we transmit from one antenna if the fading is non-regular, since in this case the asymptotic behavior depends highly on the decrease of the estimation error with the SNR. Transmitting from one antenna is also optimal if the fading process is regular and zero-mean, in which case the resulting specular component $\widetilde{d}$ is zero for any choice of $\hat{\mathbf{x}}$ and the best one can do is to minimize the prediction error.

On the other hand, maximizing the specular component $\widetilde{d}$ is optimal, if an estimate of the channel is either not available (if the fading is memoryless), or the channel estimate does not depend on the beam direction $\hat{\mathbf{x}}$ (if the spectral density matrix of the regular fading process contains identical entries).

We observe that in all cases where we could prove beam forming to be optimal, one has to optimize only one of these two parameters. It is an open question if beam forming is also optimal when one has to trade off between both parameters.

Beam forming need not always be optimal for MISO fading channels. It was shown by Jafar and Goldsmith [12] that if the receiver has perfect side information the question whether beam forming is optimal or not depends on the specular component $\mathbf{d}$ as well as on the SNR. Moreover, it can be shown that asymptotically if the SNR tends to infinity beam forming is never optimal.

8.1.3 MIMO Fading Channels

In order to discuss the results corresponding to MIMO fading channels, we first note that channel capacity in the presence of perfect side information at the receiver is given by [8]

$$C_{\text{PSI}}(\text{SNR}) = \min\{n_{\text{T}}, n_{\text{R}}\} \log \text{SNR} + O(1). \tag{8.7}$$

Thus, the use of multiple antennas at the transmitter and receiver offers a $\min\{n_{\text{T}}, n_{\text{R}}\}$-fold increase in capacity compared to the capacity of SISO channels. The capacity pre-log $\min\{n_{\text{T}}, n_{\text{R}}\}$ is often referred to as *multiplexing gain*.

However, in the absence of perfect side information the situation may be different. It was shown by Lapidoth and Moser [1] that if the fading process is regular, then capacity is given by

$$C(\text{SNR}) = \log\log \text{SNR} + O(1), \tag{8.8}$$

where only the $O(1)$ term depends on the number of antennas. Thus, in this case the use of multiple antennas has only a small impact on channel capacity in the high SNR regime.

It remains the question how big the multiplexing gain is if one is able to predict the channel state perfectly (*i.e.*, if the fading process is non-regular). We can find a trivial upper bound on the pre-log by noting that capacity in the absence of perfect side information can never be larger than in its presence. Therefore, the pre-log (or multiplexing gain) Π is always upper bounded by

$$\Pi \leq \min\{n_{\mathrm{T}}, n_{\mathrm{R}}\} = \Pi_{\mathrm{PSI}}. \tag{8.9}$$

A more elaborate upper bound is stated in Corollary 5.15. It is given by

$$\Pi \leq \max_{1 \leq t \leq n_{\mathrm{T}}} \sum_{t=1}^{n_{\mathrm{R}}} \mu\left(\left\{\lambda : \mathsf{F}'^{(r,t)}(\lambda) = 0\right\}\right) = \Pi_{\mathrm{U}}. \tag{8.10}$$

Note that $\mu(\cdot)$ cannot be larger than one, *i.e.*, $\Pi_{\mathrm{U}} \leq n_{\mathrm{R}}$. In the case where the number of transmit antennas is larger than the number of receive antennas, it follows that

$$\Pi_{\mathrm{U}} \leq n_{\mathrm{R}} = \Pi_{\mathrm{PSI}}, \qquad \text{for } n_{\mathrm{T}} \geq n_{\mathrm{R}}. \tag{8.11}$$

Thus, the upper bound derived in this thesis is at least as tight as the trivial upper bound corresponding to capacity in the presence of perfect side information.

However, if the number of transmit antennas is smaller than the number of receive antennas, then situations may exist where the trivial upper bound yields better results.

8.2 Open Problems

In the previous section several results were discussed, addressing the asymptotic behavior of channel capacity of SISO, MISO, and MIMO fading channels. However, some questions remain.

In the study of MISO fading channels with memory we could only find tight upper bounds on channel capacity when the channels are uncorrelated. Moreover, if the fading process is regular, then the upper bound is tight in only two special cases: if the fading process is zero-mean or its spectral density matrix contains identical entries. So, it will be interesting to investigate those cases where we could not find tight upper bounds and see whether beam forming is optimal or not.

Another interesting topic, not covered in this thesis, is the analysis of single-input multiple-output (SIMO) fading channels where the fading process is non-regular. We conjecture that the study of such channels may give important indications about the behavior of MIMO channels.

And last but not least, MIMO fading channels with a non-regular fading process need to be investigated in more detail. In particular, it will be interesting to study the capacity in the absence of perfect side information and see whether the use of multiple antennas at the transmitter and receiver has a similarly beneficial effect on the capacity pre-log as in the presence of perfect side information.

Chapter 9

Summary and Conclusion

In this thesis the capacity of Gaussian fading channels with memory was studied where neither transmitter nor receiver has access to the realization of the fading, though both are fully cognizant of the fading laws.

Based on previous work by Lapidoth and Moser [1] and by Lapidoth [2], channels with both regular (*i.e.*, where the channel realization cannot be predicted precisely from its past) and non-regular (*i.e.*, where the channel realization can be predicted precisely) fading processes were investigated. In both cases the emphasis was on the high signal-to-noise ratio (SNR) regime. In particular, in cases where the fading process is regular, the fading number, *i.e.*, the second order term in the high SNR expansion of capacity, was studied. In cases where the fading process is non-regular, the capacity pre-log was considered, *i.e.*, the limiting ratio of the capacity to the logarithm of the SNR.

Addressing the asymptotic analysis of channels with a regular fading process, we derived an upper bound on the fading number of multiple-input single-output (MISO) fading channels with memory. Moreover, we could show that the upper bound is tight (*i.e.*, it coincides with a lower bound) in cases where the channels are uncorrelated and the fading process is either zero-mean or its spectral density matrix contains identical entries. In the former case, the fading number can be achieved by transmitting from only one antenna, *i.e.*, the one that allows for the smallest mean squared error in predicting the channel from its past. In the latter case, the fading number can be achieved by beam forming.

Investigating the capacity of channels with non-regular fading processes, we presented upper bounds on the capacity of MISO as well as multiple-input multiple-output (MIMO) fading channels. We could show that the upper bound on the capacity of MISO fading channels is tight in the case where the channels are uncorrelated. Moreover, we presented an expression for the capacity pre-log and demonstrated that this pre-log can be achieved by transmitting from only one antenna, *i.e.*, the one that yields the smallest mean squared error in predicting the present fading from its past. In addition, a lower bound on capacity of single-input single-output (SISO) channels was derived which allowed us to express the capacity pre-log-log (*i.e.,* the limiting ratio of the logarithm to the logarithm of the SNR) in cases where channel capacity only grows double-logarithmically in the SNR.

It should be noted that in the case where the fading process is regular we considered an average-power constraint, whereas in the case where the fading is non-regular, the average-power constraint was replaced, for mathematical convenience, by a peak-power constraint. We suspect, however, that both constraints lead to very similar results in the high SNR regime. Indeed, if the fading is regular, both constraints yield the same fading number [1].

Most of the results in this thesis are new and cannot be compared to previous work. In these cases where the problem has already been studied we could find better upper bounds.

However, some open questions remain. When considering MISO fading channels, we could only find tight upper bounds for some special cases. Thus, it would be interesting to study all the other cases to find out whether beam forming is optimal or not. Moreover, the capacity of single-input multiple-output (SIMO) channels with memory has not been investigated in this thesis. We conjecture, however, that this may give important indications about the behavior of MIMO channels. And last but not least, it will be interesting to study the capacity pre-log of MIMO fading channels with memory, to explore the impact of using multiple antennas on channel capacity.

While the behavior of channel capacity in the presence of perfect side information is well understood, there remain many open questions addressing channel capacity in its absence. The study in this thesis answers some of them and has a stake in understanding the asymptotic behavior of the capacity of fading channels.

Appendix A

Proof of Lemma 3.1

We consider a multivariate zero-mean stationary stochastic process $\{\mathbf{A}_k\}$ with matrix-valued distribution function $\mathsf{F}(\lambda)$. Furthermore, we assume that the entries in $\mathbf{A}_k$ are uncorrelated, *i.e.*, the $n_\mathrm{R} \times n_\mathrm{R}$ covariance matrix $\mathsf{E}\left[\mathbf{A}_{k+m}\mathbf{A}_k^\dagger\right]$ is diagonal.

Since we are restricted to linear predictors, the estimate $\overline{\mathbf{A}}_0$ of $\mathbf{A}_0$ is of the form

$$\overline{\mathbf{A}}_0 = \sum_{k=-\infty}^{-1} \mathsf{C}_k \mathbf{A}_k. \tag{A.1}$$

In order to minimize the error in predicting $\mathbf{A}_0$ from the past values $\mathbf{A}_{-1}, \mathbf{A}_{-2}, \ldots$, we have to choose the parameters C_k according to the orthogonality principle

$$\mathsf{E}\left[\left(\mathbf{A}_0 - \overline{\mathbf{A}}_0\right)\mathbf{A}_l^\dagger\right] = 0, \qquad l = -\infty, \ldots, -1. \tag{A.2}$$

Combining (A.2) and (A.1) and using the fact that the parameters C_k are deterministic, we obtain

$$\mathsf{E}\left[\mathbf{A}_0\mathbf{A}_l^\dagger\right] = \mathsf{E}\left[\overline{\mathbf{A}}_0\mathbf{A}_l^\dagger\right] = \sum_{k=-\infty}^{-1} \mathsf{C}_k \mathsf{E}\left[\mathbf{A}_k\mathbf{A}_l^\dagger\right], \qquad l = -\infty, \ldots, -1. \tag{A.3}$$

We now rewrite (A.3) using the fact that the entries in $\mathbf{A}_k$ are uncorrelated, *i.e.*, $\mathsf{E}\left[\mathbf{A}_k\mathbf{A}_l^\dagger\right]$ is diagonal for all $k, l \in \mathbb{Z}$, and consider the diagonal and the off-diagonal elements of $\mathsf{E}\left[\mathbf{A}_0\mathbf{A}_l^\dagger\right]$ separately. It follows

$$0 = \sum_{k=-\infty}^{-1} c_k^{(r,t)} \mathsf{E}\left[A_k^{(t)} A_l^{(t)*}\right], \qquad l = -\infty, \ldots, -1, \quad r \neq t \tag{A.4}$$

and

$$\mathsf{E}\left[A_0^{(r)}A_l^{(r)*}\right] = \sum_{k=-\infty}^{-1} c_k^{(r,r)}\mathsf{E}\left[A_k^{(r)}A_l^{(r)*}\right], \qquad l=-\infty,\dots,-1$$
$$1 \le r,t \le n_{\mathrm{R}}. \qquad \text{(A.5)}$$

We note that the off-diagonal elements of C_k only appear in (A.4), whereas the diagonal elements only appear in (A.5). Thus, we can consider (A.4) and (A.5) separately. Since the choice

$$c_k^{(r,t)} = 0, \qquad k \in \mathbb{Z}, \quad r \neq t, \quad 1 \le r,t \le n_{\mathrm{T}} \qquad \text{(A.6)}$$

solves the equations in (A.4), we can reason that there is no loss in optimality in choosing the matrices C_k to be diagonal.

Then, the prediction error covariance matrix Σ is given by

$$\begin{aligned}
\Sigma &= \mathsf{E}\left[\left(\mathbf{A}_0 - \sum_{k=-\infty}^{-1}\mathsf{C}_k\mathbf{A}_k\right)\left(\mathbf{A}_0 - \sum_{m=-\infty}^{-1}\mathsf{C}_m\mathbf{A}_m\right)^\dagger\right]\\
&= \mathsf{E}\left[\mathbf{A}_0\mathbf{A}_0^\dagger\right] - \mathsf{E}\left[\sum_{k=-\infty}^{-1}\mathsf{C}_k\mathbf{A}_k\mathbf{A}_0^\dagger\right] - \mathsf{E}\left[\mathbf{A}_0\left(\sum_{m=-\infty}^{-1}\mathsf{C}_m\mathbf{A}_m\right)^\dagger\right]\\
&\quad + \mathsf{E}\left[\left(\sum_{k=-\infty}^{-1}\mathsf{C}_k\mathbf{A}_k\right)\left(\sum_{m=-\infty}^{-1}\mathsf{C}_m\mathbf{A}_m\right)^\dagger\right]\\
&= \mathsf{E}\left[\mathbf{A}_0\mathbf{A}_0^\dagger\right] - \mathsf{E}\left[\sum_{k=-\infty}^{-1}\mathsf{C}_k\mathbf{A}_k\mathbf{A}_0^\dagger\right] - \mathsf{E}\left[\mathbf{A}_0\sum_{m=-\infty}^{-1}\mathbf{A}_m^\dagger\mathsf{C}_m^\dagger\right]\\
&\quad + \mathsf{E}\left[\sum_{k=-\infty}^{-1}\mathsf{C}_k\mathbf{A}_k\sum_{m=-\infty}^{-1}\mathbf{A}_m^\dagger\mathsf{C}_m^\dagger\right]\\
&= \mathsf{E}\left[\mathbf{A}_0\mathbf{A}_0^\dagger\right] - \sum_{k=-\infty}^{-1}\mathsf{C}_k\mathsf{E}\left[\mathbf{A}_k\mathbf{A}_0^\dagger\right] - \sum_{m=-\infty}^{-1}\mathsf{E}\left[\mathbf{A}_0\mathbf{A}_m^\dagger\right]\mathsf{C}_m^\dagger\\
&\quad + \sum_{k=-\infty}^{-1}\mathsf{C}_k\sum_{m=-\infty}^{-1}\mathsf{E}\left[\mathbf{A}_k\mathbf{A}_m^\dagger\right]\mathsf{C}_m^\dagger\\
&= \mathsf{E}\left[\mathbf{A}_0\mathbf{A}_0^\dagger\right] - \sum_{k=-\infty}^{-1}\mathsf{C}_k\mathsf{E}\left[\mathbf{A}_k\mathbf{A}_0^\dagger\right] - \sum_{m=-\infty}^{-1}\left(\mathsf{C}_m\mathsf{E}\left[\mathbf{A}_m\mathbf{A}_0^\dagger\right]\right)^\dagger\\
&\quad + \sum_{k=-\infty}^{-1}\mathsf{C}_k\left(\sum_{m=-\infty}^{-1}\mathsf{C}_m\mathsf{E}\left[\mathbf{A}_m\mathbf{A}_k^\dagger\right]\right)^\dagger\\
&= \mathsf{E}\left[\mathbf{A}_0\mathbf{A}_0^\dagger\right] - \sum_{k=-\infty}^{-1}\mathsf{C}_k\mathsf{E}\left[\mathbf{A}_k\mathbf{A}_0^\dagger\right] - \sum_{m=-\infty}^{-1}\left(\mathsf{C}_m\mathsf{E}\left[\mathbf{A}_m\mathbf{A}_0^\dagger\right]\right)^\dagger
\end{aligned}$$

$$+\sum_{k=-\infty}^{-1} \mathsf{C}_k \left(\mathsf{E}\left[\mathbf{A}_0\mathbf{A}_k^\dagger\right]\right)^\dagger \tag{A.7}$$

$$= \mathsf{E}\left[\mathbf{A}_0\mathbf{A}_0^\dagger\right] - \sum_{k=-\infty}^{-1} \mathsf{C}_k \mathsf{E}\left[\mathbf{A}_k\mathbf{A}_0^\dagger\right] - \sum_{m=-\infty}^{-1} \left(\mathsf{C}_m \mathsf{E}\left[\mathbf{A}_m\mathbf{A}_0^\dagger\right]\right)^\dagger$$

$$+\sum_{k=-\infty}^{-1} \mathsf{C}_k \mathsf{E}\left[\mathbf{A}_k\mathbf{A}_0^\dagger\right]$$

$$= \mathsf{E}\left[\mathbf{A}_0\mathbf{A}_0^\dagger\right] - \sum_{m=-\infty}^{-1} \left(\mathsf{C}_m \mathsf{E}\left[\mathbf{A}_m\mathbf{A}_0^\dagger\right]\right)^\dagger. \tag{A.8}$$

Here, (A.7) follows from (A.3). We first note that since $\mathsf{E}\left[\mathbf{A}_0\mathbf{A}_0^\dagger\right]$, C_m, and $\mathsf{E}\left[\mathbf{A}_m\mathbf{A}_0^\dagger\right]$ are all diagonal for $m \in \mathbb{Z}$, the prediction error covariance matrix Σ is diagonal as well. This proves the first part of Lemma 3.1.

To prove the second part, we consider (A.8) and express each of the diagonal elements of Σ as

$$\Sigma^{(r,r)} = \mathsf{E}\left[A_0^{(r)}A_0^{(r)*}\right] - \sum_{m=-\infty}^{-1} c_m^{(r,r)*}\mathsf{E}\left[A_0^{(r)}A_m^{(r)*}\right], \qquad 1 \le r \le n_{\mathrm{R}}. \tag{A.9}$$

We note that for any index r, the diagonal element $\Sigma^{(r,r)}$ only depends on the parameters $c_m^{(r,r)}$, $m = -1, -2, \ldots$. Thus, in order to minimize the prediction error covariance matrix Σ over all matrices C_m, we can minimize each diagonal element $\Sigma^{(r,r)}$ over all scalars $c_m^{(r,r)}$. Hence, we can reduce the multivariate optimization problem to n_{R} univariate optimization problems. We know that in the univariate case, the minimum prediction error is given by

$$\epsilon_{\mathrm{MSE}}^2 = \exp\left\{\int_{-1/2}^{1/2} \log \mathsf{F}'(\lambda)\,\mathrm{d}\lambda\right\} \tag{A.10}$$

and we obtain, consequently, the following expression for the diagonal entries of Σ:

$$\Sigma^{(r,r)} = \exp\left\{\int_{-1/2}^{1/2} \log \mathsf{F}'^{(r,r)}(\lambda)\,\mathrm{d}\lambda\right\}, \qquad 1 \le r \le n_{\mathrm{R}}. \tag{A.11}$$

This concludes the proof.

Bibliography

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

[2] Amos Lapidoth. On the asymptotic capacity of fading channels. Submitted, available at `<http://www.isi.ee.ethz.ch/~lapidoth>`, 2003.

[3] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423 and 623–656, July and October 1948.

[4] Sergio Verdú and Te Sun Han. A general formula for channel capacity. *IEEE Transactions on Information Theory*, 40(4), July 1994.

[5] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 1991.

[6] Joseph L. Doob. *Stochastic Processes*. John Wiley & Sons, 1990.

[7] Norbert Wiener and Pesi R. Masani. The prediction theory of multivariate stochastic processes I. *Acta Mathematica*, 98:111–150, November 1957.

[8] İ. Emre Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10(6):585–595, 1999.

[9] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[10] Hilary A. Priestley. *Introduction to Integration*. Oxford University Press, 1997.

[11] Thomas H. E. Ericson. A Gaussian channel with slow fading. *IEEE Transactions on Information Theory*, 16(3):353–355, May 1970.

[12] Syed Ali Jafar and Andrea Goldsmith. On optimality of beamforming of multiple antenna systems with imperfect feedback. In *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Washington DC, USA, June 24–29, 2001.