Thesis for the degree of Licentiate of Engineering

The Information-Theoretic Cost of Learning Fading Channels

Wei Yang

CHALMERS

Communication Systems Group
Department of Signals and Systems
Chalmers University of Technology

Gothenburg, Sweden 2013

Yang, Wei
The Information-Theoretic Cost of Learning Fading Channels.

Department of Signals and Systems
Technical Report No. R018/2013
ISSN 1403-266X

Communication Systems Group
Department of Signals and Systems
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
Telephone: + 46 (0)31-772 1772
Email: ywei@chalmers.se

This thesis has been prepared using LaTeX.

Front Cover: A MIMO fading channel with no *a priori* channel state information.

Printed by Chalmers Reproservice,
Göteborg, Sweden, August 2013.

To my parents and Wanlu

"The art of doing mathematics consists in finding that special case which contains all the germs of generality."

– David Hilbert

Abstract

Recent results in communication theory suggest that significant throughput gains in wireless fading networks can be achieved by exploiting network coordination (e.g., CoMP, network MIMO, and interference alignment), provided that each node in the network has perfect channel knowledge. In practice, however, the channels are not known *a priori* and must be estimated. Lack of *a priori* channel knowledge determines a penalty on the throughput compared to the case where perfect channel knowledge is available.

In this thesis, we take a fresh look at the problem of learning fading channels. We characterize the cost of learning fading channels in an information-theoretic way by determining the maximal achievable rate over point-to-point fading channels under the assumption that neither the transmitter nor the receiver have *a priori* channel knowledge.

Paper A and Paper B characterizes the capacity of fading channels in the high signal-to-noise ratio (SNR) regime. Specifically, in paper A, we establish an upper bound on the capacity pre-log (i.e., the asymptotic ratio between capacity and the logarithm of SNR as SNR goes to infinity) of Rayleigh-fading correlated block-fading single-input multiple-output (SIMO) channels. The upper bound matches the lower bound reported in Riegler *et al.* (2011), and, hence, yields a complete characterization of the SIMO capacity pre-log, provided that the channel covariance matrix satisfies a mild technical condition. In Paper B, we characterize the capacity of Rayleigh-fading constant block-fading multiple-input multiple-output (MIMO) channels in the high SNR regime, and provide the input distribution that achieves capacity up to a term that vanishes as SNR grows large.

The insights gained from the high-SNR analysis allows us to obtain tight bounds on the maximal achievable rate $R^*(n,\epsilon)$ for a given blocklength n and block error probability ϵ at finite SNR. Specifically, in Paper C, we establish upper and lower bounds on $R^*(n,\epsilon)$ for a single-antenna Rayleigh-fading constant block-fading channel. Our results show that for a given block-length and error probability, $R^*(n,\epsilon)$ is not monotonic in the channel coherence time, but there exists a rate maximizing coherence time that optimally trades between diversity and cost of estimating the channel. Finally, in Paper D, we consider the special case when the fading channel does not vary over the transmission of each codeword (quasi-static fading channels). We characterize the channel dispersion in the SIMO setting, and provide tight bounds on $R^*(n,\epsilon)$ together with an easy-to-compute approximation.

Keywords: Shannon capacity, outage capacity, block-fading channel, finite blocklength, quasi-static fading, multiple antennas

List of Included Publications

The thesis is based on the following appended papers:

[A] W. Yang, G. Durisi, V. I. Morgenshtern, and E. Riegler, "Capacity pre-log of SIMO correlated block- fading channels," in *Proceedings of IEEE International Symposium on Wireless Communication Systems (ISWCS)*, Aachen, Germany, Nov. 2011, pp. 869-873.

[B] W. Yang, G. Durisi, and E. Riegler, "On the capacity of large-MIMO block-fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 117–132, Feb. 2013.

[C] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Diversity versus channel knowledge at finite block-length," in *Proceedings of IEEE Information Theory Workshop (ITW)*, Lausanne, Switzerland, Sep. 2012, pp. 572–576.

[D] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Quasi-static SIMO fading channels at finite blocklength," extended version of a paper in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, Jul. 2013.

Acknowledgements

First of all, I wish to express my sincere gratitude to my advisor Prof. Giuseppe Durisi, who introduced me into the area of information theory and worked with me closely during the last two and a half years. I'm very grateful for the enormous time he spent and effort he made to teach me how to give presentations, to write papers, and to do research. I'm also grateful to my co-supervisors: Prof. Thomas Eriksson and Prof. Erik Agrell for their kind support, guidance and for all the nice discussions.

I have been very fortunate to have the chance to collaborate with researchers outside Chalmers, including Prof. Yury Polyanskiy, Prof. Tobias Koch, Dr. Erwin Riegler and Dr. Veniamin I. Morgenshtern. I thank all of them for the stimulating discussions and fruitful collaborations. I also thank Yury for his kind hospitality at MIT where the last part of this research took place. Every time I discussed problems with him, I gained knowledge.

I would like to thank the current and former members of the communication systems group for creating a pleasant research atmosphere. In particular, I would like to thank Prof. Erik Ström for all the effort he made to make the group better and better. I would also like to thank Yutao for providing me various kinds of help, which makes my life much easier. I also want to thank my office-mate Naga for helping with my English, and Kasra for providing me all kinds of information. Many thanks to the administration staff at Department of Signals and Systems, especially Agneta, Natasha and Lars.

My Gothenburg experience was made enjoyable and colorful by all my Chinese friends: Zhennan, Xiaoming, Jingya, Tong, Songhe, Binru, and many others. I'm too lazy to list all of your names here, but I will always remember the time we spent together.

Finally, I would not have become a PhD student if I didn't have a family that attached great importance to education and encouraged its children to study. Therefore, I would like to thank my parents and my brother for their constant support, motivation and love. Most importantly, I would like to thank my wife Wanlu, who has been with me, believed in me, and encouraged me all along the way. To my parents and Wanlu, I dedicate this thesis.

Wei Yang
Gothenburg, August 2013

Acronyms

3GPP:	3rd Generation Partnership Project
AWGN:	Additive white Gaussian noise
BSTM:	Beta-variate space time modulation
cdf:	Cumulative distribution function
CoMP:	Coordinated multi-point
CSI:	Channel state information
DT:	Dependence testing
MIMO:	Multiple-input multiple-output
pdf:	Probability density function
SIMO:	Single input multiple output
SISO:	Single input single output
SNR:	Signal to noise ratio
SVD:	Singular value decomposition
USTM:	Unitary space time modulation

Contents

II Included papers 23

Part I

Overview

Chapter 1

Introduction

The surge in the use of broadband services in the public and private sectors poses high demands on future wireless communication networks, from the core (backhaul) to the periphery (cellular base stations). Ericsson, a wireless network manufacturer, predicts that mobile data traffic will have to double every year to match future demands [1]. At the same time, wireless networks also need to provide ubiquitous connectivity for the end user at a minimum guaranteed data rate. It is widely agreed that these throughput gains as well as the expectation of ubiquitous user experience can only be achieved by a dense and heterogeneous deployment of the wireless network infrastructure. Under this scenario, the availability of effective interference management schemes becomes crucial.

Recent results in communication theory suggest that throughput gains of up to two orders of magnitude are obtainable by exploiting interference management techniques such as CoMP [2], network MIMO [3], and interference alignment [4]. However, over-the-air trials have only demonstrated disappointingly low throughput improvements (typically not exceeding 30%) [2] or no improvements at all [5].

One possible explanation for this disconnect between theory and practice lies in the assumption of perfect channel state information (CSI) under which state-of-the-art interference management techniques are developed. This assumption is not realistic because, in most wireless communication networks, CSI is not known *a priori* at each node and must be estimated, for example through the transmission of pilot symbols. The time and/or bandwidth spent sending pilot symbols is not negligible, particularly in systems with a large number of nodes or antenna elements. Also, determining the amount of resources to be used for channel estimation is crucial in the design of wireless communication systems operating over fading channels. It is therefore of great importance to understand the cost of learning fading channels.

In this thesis, we take a fresh look at the problem of learning fading channels. Determining the Shannon capacity [6, Ch. 7] of fading channels under the assumption that neither the transmitter nor the receiver have *a priori* channel knowledge is a fundamental way to assess the cost of learning fading channels. We emphasize that in our setup of no *a priori* channel knowledge, the receiver is allowed to try and estimate the channel. The employed communication strategy does not necessarily need to be noncoherent. The typical approach of transmitting pilot symbols to estimate the channel is simply viewed as a specific form of coding. Hence, capacity in the absence of *a priori* CSI is the ultimate limit on the rate of reliable communication over fading channels, regardless of whether channel estimation is performed or not.

Unfortunately, capacity in the absence of *a priori* CSI is notoriously difficult to characterize analytically and no closed-form expressions are available to date, even for the simplest channel models [7, 8]. Most of the results in this area pertain to the asymptotic regime of low or high signal-to-noise ratio (SNR) [9–12].

In order to study channel capacity of fading channels with no *a priori* CSI, we need to specify a channel model that describes the channel variations in time. In this thesis, we consider two models, namely, the *constant* block-fading model and the *correlated* block-fading model. In the constant block-fading model, the fading process takes on independent realizations across blocks of a certain length, and remains constant within each block. This model is accurate for systems employing frequency hopping or time-division multiple access, and is widely used in the literature due to its simplicity [8, 9, 13]. The correlated block-fading model, which was first introduced in [14], generalizes the constant block-fading model by allowing more general forms of temporal correlation between different fading gains within each block. Hence, channel variations in time can be captured in a more accurate way than with the constant block-fading model. The correlated block-fading model arises naturally in the analysis of cyclic-prefix orthogonal frequency-division multiplexing (OFDM) systems operating over frequency-selective channels [12, 15].

Channel capacity is a relevant performance metric only when sufficiently long codewords are used. In emerging applications such as machine-type and vehicle-to-vehicle communications, the requirement of long codewords is often too stringent, and short codewords are needed to meet the delay constraints. Indeed, the 4G wireless standard LTE-Advanced employs codes with blocklengths as short as 100 symbols [16, Sec. 5.1.3]. For such short blocklengths, the fundamental limit of practical interest is the maximal achievable rate $R^*(n,\epsilon)$ for a given blocklength n and block error probability ϵ.

In the nonasymptotic regime of finite blocklength, there are no exact formulas for the maximal achievable rate as a function of blocklength and error probability. In fact, the exact computation of $R^*(n,\epsilon)$ is an NP-hard problem [17], and is prohibitive already for very short blocklengths. Nevertheless, computable upper and lower bounds on the maximal achievable rate are available in the literature [18–23]. Of particular interest for this thesis are the new bounds recently developed by Polyanskiy, Poor and Verdú [23] such as the meta-converse bound, the dependence-testing bound, and the $\kappa\beta$ bound. Using these bounds, they showed that for various channels with positive capacity C, the maximal achievable rate can be tightly approximated by

$$R^*(n,\epsilon) \approx C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon). \tag{1.1}$$

Here, $Q^{-1}(\cdot)$ denotes the inverse of the complementary cumulative distribution function (cdf) of a standard normal random variable, and V is the *channel dispersion* (see [23, Def. 1] or (2.17)).

1.1 Objectives

The aim of this thesis is to study—under the assumption of no *a priori* CSI—the capacity and the maximal achievable rate at finite blocklength of point-to-point fading channels. To this end, we first study the capacity in the high SNR regime. For the correlated block-fading model, we determine the *pre-log* (also known as *number of degrees of freedom* or *multiplexing gain*), which is defined as the asymptotic ratio between capacity and the logarithm of the SNR as SNR goes to infinity, in the single-input multiple-output (SIMO) setting. For the constant block-fading model, we characterize the capacity of multiple-input multiple-output (MIMO) channel in the high SNR regime and provide the input distribution that achieves capacity up to a term that vanishes as SNR grow large. The insights gained from the high-SNR analysis allows us to obtain tight bounds on the maximal achievable rate $R^*(n, \epsilon)$ at finite SNR and at finite blocklength. Specifically, we establish upper and lower bounds on $R^*(n, \epsilon)$ for a single-input single-output (SISO) constant block-fading channel at finite SNR and at finite blocklength. Finally, for the special case that the fading channel does not vary over the transmission of a codeword (quasi-static fading channels), we characterize the dispersion of the channel in the SIMO setting and provide tight bounds as well as an easy-to-compute approximation for $R^*(n, \epsilon)$.

1.2 Thesis Outline

The thesis is organized as follows: in Chapter 2, we introduce the fading channel models studied in this thesis, define the fundamental information-theoretic limits, and review previous results. In Chapter 3, we present the information-theoretic tools we shall use in this thesis. Finally, in Chapter 4, we summarize our contributions.

Chapter 2

Channel Model, Fundamental Limits, and Known Results

This thesis deals with the problem of characterizing the maximal achievable rate over point-to-point fading channels. In Section 2.1, we introduce the two channel models that will be studied in this thesis, namely, the constant block-fading model and the correlated block-fading model. In Section 2.2, we define the notion of channel capacity, outage capacity, and maximal achievable rate at finite blocklength, and review some known results.

2.1 Channel Model

The capacity of fading channels under no *a priori* CSI is known to be sensitive to the channel model used, especially in the high SNR regime [10, 24, 25]. In our quest for simplicity of exposition, we focus on block-fading channel models, which we will describe in the following two sections.

2.1.1 Constant Block-Fading Model

The *constant* block-fading model is perhaps the simplest model to capture channel variations. The input-output relation of a constant block-fading MIMO channel with T transmit antennas, R receive antennas within a block of N channel uses is [8, 9, 13]:

$$\boldsymbol{y}_r = \sqrt{\frac{\rho}{T}} \sum_{t=1}^{T} h_{r,t} \boldsymbol{x}_t + \boldsymbol{w}_r, \qquad r \in \{1, \dots, R\}. \tag{2.1}$$

Here, $\boldsymbol{x}_t \in \mathbb{C}^N$ is the signal vector originating from the tth transmit antenna over the block of N channel uses; $\boldsymbol{y}_r \in \mathbb{C}^N$ is the signal vector at the rth receive antenna; $h_{r,t}$ is the channel coefficient between receive antenna r and transmit antenna t; $\boldsymbol{w}_r \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_N)$ is the noise vector at the rth receive antenna; finally, $\rho \in \mathbb{R}_+$ denotes the SNR. In most of the thesis (with the exception of Paper D) we shall assume $h_{r,t} \sim \mathcal{CN}(0, 1)$, i.e., Rayleigh fading. In our setup, the transmitted signal vectors $\boldsymbol{x}_t$ do not depend on the channel coefficients $h_{r,t}$ and the noise vectors $\boldsymbol{w}_r$. The channel coefficients $h_{r,t}$ are assumed to be mutually independent and independent across $r \in \{1, \dots, R\}$ and $t \in \{1, \dots, T\}$, and to change in an independent fashion from block to block ("block-memoryless" assumption). We assume that neither the transmitter nor the receiver have *a priori* knowledge of the realizations of $\{h_{r,t}\}$ and $\{\boldsymbol{w}_r\}$, bur both know their statistics.

In most of the thesis, we shall consider an ergodic scenario where the coding is performed over sufficiently many blocks. A different scenario considered in Section 2.2.2 and Paper D is the so-called *quasi-static* setting, originally proposed in [26], where it is assumed that N is large and that coding is performed over one block only. In this scenario, the fading is a non-ergodic process.

2.1.2 Correlated Block-Fading Model

The *correlated* block-fading model extends the constant block-fading model to allow for the N fading coefficients in each block to be jointly Gaussian with a covariance matrix of rank $Q \geq 1$. The input-output relation of a correlated block-fading MIMO channel with T transmit antennas, R receive antennas within a block of N channel uses is [14]:

$$\boldsymbol{y}_r = \sum_{t=1}^{T} \text{diag}\{\boldsymbol{h}_{r,t}\}\boldsymbol{x}_t + \boldsymbol{w}_r, \quad r \in \{1, \dots, R\}. \tag{2.2}$$

Here, the vector $\boldsymbol{h}_{r,t} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{P}\boldsymbol{P}^{\text{H}})$ contains the fading coefficients between the receive antenna r and transmit antenna t, and its realizations are assumed to be unknown by the transmitter and the receiver. The $N \times Q$ matrix $\boldsymbol{P}$, which is deterministic and of full rank $1 \leq Q \leq N$, describes the correlation structure within a block and is assumed to be known perfectly to the transmitter and the receiver. We further assume that the rows of $\boldsymbol{P}$ have unit norm, and hence, the entries of $\boldsymbol{h}_{r,t}$ are identically distributed. The fading vectors $\{\boldsymbol{h}_{r,t}\}$ between all antenna pairs are assumed to be independent and identically distributed (i.i.d.),[1] and are independent across different blocks. The correlated block-fading model just described captures channel variation in time in a more accurate (yet simple) way than the constant block-fading model: large Q corresponds to fast channel variations. Furthermore, (2.2) models accurately the IO relation in the frequency domain of a cyclic-prefix OFDM system that operates over a multipath channel with Q uncorrelated taps. Note that (2.1) is a special case of (2.2) with $\boldsymbol{P} = [1, \dots, 1]^{\text{T}}$ and $Q = 1$.

Before ending this section, we remark that in both the constant block-fading model and the correlated block-fading model the fading is modeled as a non-stationary process. An alternative approach to capturing channel variations in time is to assume that the fading process is symbol-by-symbol *stationary.* The key differences between block-fading and stationary fading is discussed in [28]. The high-SNR capacity of the stationary channel with no *a priori* CSI is studied in [10, 24, 29, 30]. Somewhat surprisingly, this

[1] This assumption can be further relaxed by allowing for the fading vectors between different antenna pairs to have different covariance matrices of the same rank [27].

model can lead in some cases to capacity estimates that are drastically different from the ones obtained using the block-fading models.

2.2 Fundamental Limits and Known Results

2.2.1 Channel Capacity

To keep notation simple, we write $\boldsymbol{X} \triangleq [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T] \in \mathbb{C}^{N \times T}$ and $\boldsymbol{Y} \triangleq [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_R] \in \mathbb{C}^{N \times R}$. Because of the block-memoryless assumption, the ergodic capacity of the channel in (2.1) and (2.2) is given by

$$C(\rho) = \frac{1}{N} \sup I(\boldsymbol{X}; \boldsymbol{Y}). \tag{2.3}$$

Here, $I(\boldsymbol{X}; \boldsymbol{Y})$ denotes the mutual information [6, p. 251] and the supremum is taken over all input distributions on $\mathbb{C}^{N \times T}$ satisfying the average power-constraint

$$\sum_{t=1}^{T} \mathbb{E}\left[\|\boldsymbol{x}_t\|^2\right] \leq NT. \tag{2.4}$$

The capacity pre-log is defined as

$$\chi = \lim_{\rho \to \infty} \frac{C(\rho)}{\log \rho}. \tag{2.5}$$

The capacity of the constant block-fading MIMO channel (2.1) with no *a priori* CSI has been studied extensively in the literature for the Rayleigh fading case [8, 9, 13, 31]. However, no closed-form capacity expression is available to date. Marzetta and Hochwald [8] proved that the capacity-achieving input matrix $\boldsymbol{X}$ is of the form $\boldsymbol{X} = \boldsymbol{\Phi}\boldsymbol{D}$, where $\boldsymbol{\Phi}$ is an $N \times T$ isotropically distributed unitary matrix and $\boldsymbol{D}$ is a $T \times T$ nonnegative diagonal matrix independent of $\boldsymbol{\Phi}$. In the same paper, they also conjectured that allocating equal powers to different transmit antennas, i.e., choosing $\boldsymbol{D} = \sqrt{N}\boldsymbol{I}_T$, is optimal at high SNR. The corresponding probability distribution on $\boldsymbol{X}$ is sometimes referred to as *unitary space-time modulation (USTM)* [31–33]. Zheng and Tse [9] studied capacity of the channel (2.1) at high SNR and proved that the pre-log is given by

$$\chi = M^* \left(1 - \frac{M^*}{N}\right), \quad \text{with } M^* = \min\{T, R, \lfloor N/2 \rfloor\}. \tag{2.6}$$

For the case $N \geq T + R$, Zheng and Tse [9] proved that USTM is indeed optimal at high SNR, and that

$$C(\rho) = \chi \log \rho + c + o(1), \quad \rho \to \infty \tag{2.7}$$

where [9, Eq. (24)]

$$c = \frac{1}{N} \log \frac{\Gamma_T(T)}{\Gamma_T(N)} + T\left(1 - \frac{T}{N}\right) \log \frac{N}{Te} + \left(1 - \frac{T}{N}\right) \mathbb{E}\left[\log \det(\boldsymbol{H}\boldsymbol{H}^{\mathrm{H}})\right] \tag{2.8}$$

is a constant that does not depend on ρ, $\Gamma_m(a) = \pi^{m(m-1)/2} \prod_{k=1}^{m} \Gamma(a-k+1)$ with $\Gamma(\cdot)$ being the Gamma function [34, Eq. (6.1.1)], and $o(1)$ denotes a function of ρ that vanishes as $\rho \to \infty$. Differently from the pre-log result (2.6), the high-SNR expression (2.7) describes capacity accurately already at moderate SNR values [35].

For the correlated block-fading channel model, it was shown in [14] that the capacity pre-log for the SISO case is given by

$$\chi = \left(1 - \frac{Q}{N}\right). \tag{2.9}$$

A lower bound on the SIMO capacity pre-log is reported in [36, 37], and generalized in [27] to the MIMO case. Finding a capacity characterization that is tight up to $o(1)$ [cf. (2.7)] for the correlated block-fading channel is an open problem.

While all results reviewed in this section pertain to the Rayleigh fading case, it is shown in [38] that among all fading distributions whose law has no mass point at zero, the Rayleigh distribution gives rise to the smallest pre-log. So Rayleigh fading is the worst fading in terms of capacity pre-log.

2.2.2 Outage Capacity

For the quasi-static fading model introduced at the end of Section 2.1.1, the channel capacity is zero whenever the closure of the support of the fading distribution includes zero, because reliable communication cannot be guaranteed for any positive data rate. In this scenario, a more meaningful performance metric may be the *outage capacity* (also referred to as ϵ-capacity [39]). Let $\boldsymbol{H} \triangleq [h_{r,t}]_{1\le r\le R, 1\le t\le T} \in \mathbb{C}^{R\times T}$. The outage capacity of MIMO quasi-static fading channels is given by [40]

$$C_\epsilon = \sup\left\{\xi : P_{\text{out}}(\xi) \le \epsilon\right\} \tag{2.10}$$

where $P_{\text{out}}(\xi)$ denotes the outage probability[2] corresponding to a given rate ξ

$$P_{\text{out}}(\xi) = \inf_{\substack{\boldsymbol{Q}:\boldsymbol{Q}\succeq 0\\ \text{tr}\{\boldsymbol{Q}\}\le T}} \mathbb{P}\left[\log\det\left(\boldsymbol{I}_R + \sqrt{\frac{\rho}{T}}\boldsymbol{H}\boldsymbol{Q}\boldsymbol{H}^{\text{H}}\right) \le \xi\right]. \tag{2.11}$$

Note that (2.10) holds for the case when CSI is not available at the transmitter and regardless of whether CSI is available at the receiver or not. The matrix $\boldsymbol{Q}$ that minimizes the right-hand-side (RHS) of (2.11) is in general not known. For the Rayleigh-fading case, Telatar [40] conjectured that the optimal $\boldsymbol{Q}$ is of the form[3]

$$\frac{\rho}{T^*}\,\text{diag}\{\underbrace{1,\dots,1}_{T^*},\underbrace{0,\dots,0}_{T-T^*}\}, \qquad 1 \le T^* \le T \tag{2.12}$$

[2]Strictly speaking, (2.10) holds for all $\epsilon > 0$ for which C_ϵ is a continuous function of ϵ.
[3]This conjecture has recently been proved in [41] for the multiple-input single-output case.

and that for small ϵ values or for high SNR, all available transmit antennas should be used (see also [42, p. 367]).

When CSI is available at the transmitter, the well known water-filling power allocation achieves the infimum in (2.11).

2.2.3 Maximal Achievable Rate

We introduce the notion of the maximal achievable rate $R^*(n, \epsilon)$ for a given blocklength n and error probability ϵ. As in [23, 43], we define it for an abstract random transformation $P_{Y|X}$ with input and output sets A and B. An $(M, \epsilon)_{\max}$ code for the random transformation $P_{Y|X}$ is a pair of random transformations $P_{X|W} : \{1, \cdots, M\} \mapsto \mathcal{X}$ and $P_{\hat{W}|Y} : \mathcal{Y} \mapsto \{1, \cdots, M\}$ such that

$$\max_{1\leq j\leq M} \mathbb{P}[\hat{W} \neq W | W = j] = \max_{1\leq j\leq M} \Big(1 - P_{\hat{W}|W}(j|j)\Big) \leq \epsilon. \tag{2.13}$$

The random transformations $P_{X|W}$ and $P_{\hat{W}|Y}$ are called the encoder and the decoder of the code, respectively. An $(M, \epsilon)_{\text{avg}}$ code is defined in a similar way, except that (2.13) is replaced with the average probability of error criterion:

$$\mathbb{P}[\hat{W} \neq W] = \frac{1}{M} \sum_{j=1}^{M} \Big(1 - P_{\hat{W}|W}(j|j)\Big) \leq \epsilon \tag{2.14}$$

where W is equiprobable on $\{1, \ldots, M\}$.

A channel is a sequence of random transformations, $\{P_{Y^n|X^n}, n = 1, 2, \ldots\}$, parameterized by the blocklength n. An (M, ϵ) code for the n-th random transformation of the channel is called an (n, M, ϵ) code. The maximal achievable rate $R^*(n, \epsilon)$ is defined as[4]

$$R^*(n, \epsilon) \triangleq \sup\left\{\frac{\log M}{n} : \exists\, (n, M, \epsilon) \text{ code}\right\}. \tag{2.15}$$

As already mentioned, the capacity C and ϵ-capacity C_ϵ are the asymptotic limits of $R^*(n, \epsilon)$:

$$C = \lim_{\epsilon\to 0} \lim_{n\to\infty} R^*(n, \epsilon), \quad C_\epsilon = \lim_{n\to\infty} R^*(n, \epsilon). \tag{2.16}$$

As noted in Chapter 1, the exact value of $R^*(n, \epsilon)$ is unknown. In fact, the search for $R^*(n, \epsilon)$ in (2.15) is doubly exponential in the blocklength n. Nevertheless, bounds and easy-to-compute approximations for $R^*(n, \epsilon)$ are available in the literature [18–23]. In particular, Polyanskiy, Poor and Verdú [23] recently derived several new bounds on the maximal achievable rate, such as the meta-converse bound, the dependence-testing bound, and the $\kappa\beta$ bound. By analyzing these bounds using the central limit theorem, they obtained a simple closed-form approximation, given in (1.1), for the maximal achievable rate. To compute the approximation (1.1), one only needs to know one more

[4] In addition, one should also specify which probability of error criterion, (2.13) or (2.14), is used.

parameter of the channel (in addition to channel capacity), i.e., the channel dispersion [23, Def. 1], which is defined as

$$V = \lim_{\epsilon \to 0} \lim_{n \to \infty} n \left(\frac{C - R^*(n, \epsilon)}{Q^{-1}(\epsilon)} \right)^2 . \tag{2.17}$$

The rationale behind the approximation (1.1) and the definition (2.17) of channel dispersion is the following expansion [23]

$$R^*(n, \epsilon) = C - \sqrt{\frac{V}{n}} Q^{-1}(\epsilon) + \mathcal{O}\left(\frac{\log n}{n} \right) \tag{2.18}$$

which is valid for various channels with positive channel capacity. Here, the notation $f(x) = \mathcal{O}(g(x))$ means that $\limsup_{x \to \infty} |f(x)/g(x)| < \infty$.

Chapter 3

Information Theoretic Tools

In this chapter, we introduce the information theoretic tools we shall need for the analysis of the capacity and of the maximal achievable rate over block-memoryless fading channels.

3.1 The Duality Technique

For the convenience of notation, consider a memoryless channel $P_{Y|X} : \mathcal{X} \mapsto \mathcal{Y}$. Here, $\mathcal{X}$ and $\mathcal{Y}$ denote the input and output alphabets, respectively. Let P_Y be the probability distribution induced on Y by the input distribution P_X and by the channel $P_{Y|X}$. By the definition of mutual information [6, Sec. 8.5] [21, Eq. (1.8)], we have that

$$I(X;Y) \triangleq D(P_{Y|X} \| P_Y | P_X) \tag{3.1}$$
$$= \int dP_X(x) \int dP_{Y|X=x}(y) \log \frac{dP_{Y|X=x}}{dP_Y}(y) dy dx \tag{3.2}$$

where $D(\cdot\|\cdot)$ stands for the relative entropy between two probability distributions and

$$D(V\|W|P_X) \triangleq \mathbb{E}_{P_X}[D(V(\cdot|X)\|W(\cdot|X))] \tag{3.3}$$

denotes the conditional relative entropy [21, Eq. (2.4)]. The capacity of the channel $P_{Y|X}$ is given by

$$C = \sup_{P_X} D(P_{Y|X} \| P_Y | P_X) \tag{3.4}$$

where the supreme is taken over all probability measures P_X on $\mathcal{X}$ that satisfy the given cost constraint (e.g., an average power constraint).

Establishing lower bounds on channel capacity is relatively simple: every input distribution P_X leads to a lower bound on C

$$C \geq D(P_{Y|X} \| P_Y | P_X). \tag{3.5}$$

To obtain a tight lower bound, we need to choose P_X to be close to a capacity-achieving input distribution and such that $D(P_{Y|X} \| P_Y | P_X)$ is tractable.

Finding upper bounds on channel capacity is more difficult because we need to perform the maximization over all possible input distributions, and because the function we need to maximize depends on P_X in an intricate way (see (3.2)). The main idea behind the duality technique is the following dual expression for capacity

$$C = \inf_{Q_Y} \sup_{P_X} D(P_{Y|X} \| Q_Y | P_X). \tag{3.6}$$

We see from (3.6) that every probability distribution Q_Y on the output $\mathcal{Y}$ yields an upper bound on C

$$C \leq \sup_{P_X} D(P_{Y|X} \| Q_Y | P_X). \tag{3.7}$$

Comparing (3.4) and (3.7), we see that in the duality bound (3.7) the distribution P_Y is replaced by an axillary distribution Q_Y that does not depend on the input distribution P_X. This makes the RHS of (3.7) much easier to deal with than the RHS of (3.4). As illustrated in many examples (see, e.g., [10, 12] and Paper B), for an appropriate choice of Q_Y, it is possible to establish tight upper bounds on $D(P_{Y|X} \| Q_Y | P_X)$ that hold for every input distribution P_X. In other words, the maximization in (3.7) can be circumvented through appropriate bounding steps.

Obviously, the crucial step in deriving capacity upper bounds using the duality technique is to choose an appropriate output distribution Q_Y. In principle, one should choose Q_Y to be close to the capacity-achieving output distribution and so as to guarantee that the RHS of (3.7) is tractable. At high SNR, a typical choice of Q_Y is the output distribution induced by the input distribution without the additive Gaussian noise (see, e.g., [12]). In Paper A and Paper B, we used a more refined choice of Q_Y, which is motivated by geometric arguments.

3.2 Finite Blocklength Bounds

In this section, we review some of the bounds on the maximal achievable rate $R^*(n, \epsilon)$ developed in [23]. These results apply to arbitrary channels and arbitrary blocklengths. As in [23], we state them for an abstract random transformation $P_{Y|X}$ with input and output sets A and B.[1]

For a joint distribution P_{XY} on $\mathsf{A} \times \mathsf{B}$ we denote the information density by

$$i(x; y) = \log \frac{dP_{Y|X=x}}{dP_Y}(y). \tag{3.8}$$

The following result, referred to as the dependence testing (DT) bound, provides an achievability (lower) bound on $R^*(n, \epsilon)$.

[1] When we apply these results, we shall take A and B to be n-fold Cartesian products of some alphabets $\mathcal{X}$ and $\mathcal{Y}$. Thus, the element of A and B and the realizations of the random variables X and Y can be thought of as vectors of dimension equal to the blocklength n.

THEOREM 3.1 (DT bound, [23, Th. 17]) *For every distribution P_X on A, there exists a code with M codewords and average probability of error not exceeding*

$$\epsilon \leq \mathbb{E}\left[\exp\left\{-\left|i(X;Y) - \log\frac{M-1}{2}\right|^+\right\}\right] \tag{3.9}$$

where $|\cdot|^+ \triangleq \max\{0, \cdot\}$.

The DT bound uses a threshold decoder that sequentially tests all messages and returns the first message whose likelihood exceeds a pre-determined threshold. Applications of the DT bound to various channels (mainly channels without input constraints) can be found in [23] [44]. In Paper C, we use the DT bound to derive lower bounds on $R^*(n, \epsilon)$ for constant block-fading channels.

For channels with input constraints, an achievability bound, called $\kappa\beta$ bound, that is easier to analyze analytically than the DT bound is established in [23]. To state the $\kappa\beta$ bound, we need some definitions regarding the performance of hypothesis testing. Given two distributions P and Q on a common measurable space W, we define a randomized test between P and Q as a random transformation $P_{Z|W} : \mathsf{W} \mapsto \{0,1\}$ where 0 indicates that the test chooses Q. As performance metric for the test between P and Q, we take the Neyman-Pearson ROC curve [45]

$$\beta_\alpha(P, Q) = \min \int P_{Z|W}(1|w)Q(dw) \tag{3.10}$$

where the minimum is over all probability distributions $P_{Z|W}$ satisfying

$$\int P_{Z|W}(1|w)P(dw) \geq \alpha. \tag{3.11}$$

The minimum in (3.10) is guaranteed to be achieved by the Neyman-Pearson Lemma [46]. Each per-codeword cost constraint can be defined by specifying a subset $\mathsf{F} \subset \mathsf{A}$ of permissible inputs. For an arbitrary $\mathsf{F} \subset \mathsf{A}$, we define the following measure of performance for the composite hypothesis test between Q_Y and the collection $\{P_{Y|X=x}\}_{x\in\mathsf{F}}$:

$$\kappa_\tau(\mathsf{F}, Q_Y) = \inf \int P_{Z|Y}(1|y)Q_Y(dy). \tag{3.12}$$

Here, the infimum is over all probability distributions $P_{Z|Y}$ satisfying

$$\int P_{Z|Y}(1|y)P_{Y|X=x}(dy) \geq \tau, \quad \forall x \in \mathsf{F}. \tag{3.13}$$

We are now ready to state the $\kappa\beta$ bound.

THEOREM 3.2 ($\kappa\beta$ bound, [23, Th. 25]) *For every $0 < \epsilon < 1$, there exists an $(M, \epsilon)_{\max}$ code with codewords chosen from $\mathsf{F} \subset \mathsf{A}$, satisfying*

$$M \geq \sup_{0<\tau<\epsilon} \sup_{Q_Y} \frac{\kappa_\tau(\mathsf{F}, Q_Y)}{\sup_{x\in\mathsf{F}} \beta_{1-\epsilon+\tau}(P_{Y|X=x}, Q_Y)}. \tag{3.14}$$

The $\kappa\beta$ is used in [23] to derive an achievability bound on $R^*(n, \epsilon)$ for the AWGN channel and in [47] for point-to-point fading channels with perfect *a priori* CSI. In Paper D, we shall use a modified version of this bound to derive an achievability bound on $R^*(n, \epsilon)$ for quasi-static SIMO fading channels.

Next, we introduce a general converse bound called meta-converse bound.[2] The meta-converse bound allows one to establish a converse bound for the channel $P_{Y|X}$, provided that a converse bound for an auxiliary channel $Q_{Y|X}$ is available.

Theorem 3.3 (Meta-converse, [23, Th. 26]) *Let $Q_{Y|X} : \mathsf{A} \mapsto \mathsf{B}$ be an auxiliary channel. For a given code, let*

$$\begin{aligned} \epsilon &= \text{average error probability with } P_{Y|X} \\ \epsilon' &= \text{average error probability with } Q_{Y|X} \\ P_X &= \text{encoder output distribution with equiprobable codewords.} \end{aligned}$$

Then

$$\beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_{Y|X}) \leq 1 - \epsilon'. \tag{3.15}$$

Choosing $Q_{Y|X}$ in Theorem 3.3 to be independent of the input X, we get the following result.

Theorem 3.4 (Minmax converse, [23, Th. 27]) *Every $(M, \epsilon)_{\text{avg}}$ code with codewords belonging to A satisfies*

$$\log M \leq -\log \left\{ \inf_{P_X} \sup_{Q_Y} \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) \right\} \tag{3.16}$$

where P_X ranges over all probability measures on A, and Q_Y ranges over all probability measures on B.

Note that $\beta_\alpha(\cdot, \cdot)$ satisfies the following *saddle point* property [48] (under regularity conditions on the input set A)

$$\inf_{P_X} \sup_{Q_Y} \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) = \sup_{Q_Y} \inf_{P_X} \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y). \tag{3.17}$$

Hence, (3.16) can be written equivalently as

$$\log M \leq -\log \left\{ \sup_{Q_Y} \inf_{P_X} \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) \right\}. \tag{3.18}$$

The bound (3.18) can be further relaxed by fixing a convenient Q_Y as follows

$$\log M \leq -\log \left\{ \inf_{P_X} \beta_{1-\epsilon}(P_X P_{Y|X}, P_X Q_Y) \right\}. \tag{3.19}$$

[2] The maximal probability of error counterparts to Theorem 3.3 and Theorem 3.4 can be found in [23, Th. 30 and Th. 31].

As in the duality technique, a crucial step in using (3.19) to derive upper bounds on $R^*(n, \epsilon)$ is to choose an appropriate output distribution Q_Y. A natural choice for Q_Y is the capacity achieving output distribution (see, e.g., [23, p. 2324] and [47]). Other choices of Q_Y are motivated by symmetries in the channel input-output relations and geometry arguments.

It is interesting to compare the duality bound (3.7) on the channel capacity with the converse bound (3.19) on the maximal achievable rate at finite blocklength. The duality bound (3.7) can be seen as the asymptotic counterpart of the bound (3.19). In fact, the Chernoff-Stein Lemma [6, Th. 11.8.3] implies that

$$-\log \beta_{1-\epsilon}(P_{XY}^n, P_X^n Q_Y^n) = nD(P_{XY} \| P_X Q_Y) + o(n) \tag{3.20}$$

$$= nD\big(P_{Y|X} \| Q_Y \big| P_X\big) + o(n) \tag{3.21}$$

where P_{XY}^n, P_X^n and Q_Y^n are n i.i.d. copies of P_{XY}, P_X and Q_Y. In particular, we expect that the output distributions Q_Y that yield tight upper bounds on capacity via the duality bound perform also well at finite blocklength via the bound (3.19).

Chapter 4

Contributions

This thesis aims at studying—under the assumption of no *a priori* CSI—the capacity and the maximal achievable rate at finite blocklength of point-to-point fading channels. In Section 4.1, we list the papers that are appended to this thesis and summarize their contributions. Additional publications by the author, which are not included in this thesis, are listed in Section 4.2.

4.1 Included Publications

1. **Paper A: "Capacity pre-log of SIMO correlated block- fading channels"**
   We establish an upper bound on the noncoherent capacity pre-log of temporally correlated block-fading single-input multiple-output (SIMO) channels. The upper bound matches the lower bound recently reported in Riegler *et al.* (2011), and, hence, yields a complete characterization of the SIMO capacity pre-log, provided that the channel covariance matrix satisfies a mild technical condition. This result allows one to determine the optimal number of receive antennas to be used to maximize the capacity pre-log for a given block-length and a given rank of the channel covariance matrix.

2. **Paper B: "On the capacity of large-MIMO block-fading channels"**
   We characterize the capacity of Rayleigh block-fading multiple-input multiple-output (MIMO) channels in the noncoherent setting where transmitter and receiver have no *a priori* knowledge of the realizations of the fading channel. We prove that *unitary space-time modulation* (USTM) is not capacity-achieving in the high signal-to-noise ratio (SNR) regime when the total number of antennas exceeds the coherence time of the fading channel (expressed in multiples of the symbol duration), a situation that is relevant for MIMO systems with large antenna arrays (large-MIMO systems). This result settles a conjecture by Zheng & Tse (2002) in the affirmative. The capacity-achieving input signal, which we refer to as *Beta-variate space-time modulation* (BSTM), turns out to be the product of a unitary isotropically distributed random matrix, and a diagonal matrix whose nonzero entries are distributed as the square-root of the eigenvalues of a Beta-distributed random matrix of appropriate size. Numerical results

illustrate that using BSTM instead of USTM in large-MIMO systems yields a rate gain as large as 13% for SNR values of practical interest.

3. **Paper C: "Diversity versus channel knowledge at finite block-length"**
We study the maximal achievable rate $R^*(n, \epsilon)$ for a given block-length n and block error probability ϵ over Rayleigh block-fading channels in the noncoherent setting and in the finite block-length regime. Our results show that for a given block-length and error probability, $R^*(n, \epsilon)$ is not monotonic in the channel's coherence time, but there exists a rate maximizing coherence time that optimally trades between diversity and cost of estimating the channel.

4. **Paper D: "Quasi-static SIMO fading channels at finite blocklength"**
We investigate the maximal achievable rate for a given blocklength and error probability over quasi-static single-input multiple-output (SIMO) fading channels. Under mild conditions on the channel gains, it is shown that the channel dispersion is zero regardless of whether the fading realizations are available at the transmitter and/or the receiver. The result follows from computationally and analytically tractable converse and achievability bounds. Through numerical evaluation, we verify that, in some scenarios, zero dispersion indeed entails fast convergence to outage capacity as the blocklength increases. In the example of a particular 1×2 SIMO Rician channel, the blocklength required to achieve 90% of capacity is about an order of magnitude smaller compared to the blocklength required for an AWGN channel with the same capacity.

4.2 Publications Not Included

Publications by the author, which are not included in this thesis, are listed below.

[J1] V. I. Morgenshtern, E. Riegler, W. Yang, G. Durisi, S. Lin, B. Sturmfels, and H. Bölcskei, "Capacity pre-log of noncoherent SIMO channels via Hironaka's theorem," *IEEE Transactions on Information Theory*, vol. 59, no. 7, pp. 4213–4229, Jul. 2013.

[J2] Y. Sui, J. Vihrälä, A. Papadogiannis, M. Sternad, W. Yang, and T. Svensson, "Moving cells: a promising solution to boost performance for vehicular users," *IEEE Communications Magazine*, vol. 41, no. 6, pp. 62–68, Jun. 2013.

[C1] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Block-fading channels at finite blocklength," in *Proceedings of IEEE International Symposium on Wireless Communication Systems (ISWCS)*, Ilmenau, Germany, Aug. 2013, to appear.

[C2] E. Riegler, G. Koliander, W. Yang, and G. Durisi, "How costly is it to learn fading channels?" in *Proceedings of International Black Sea Conference on Communications and Networking (BlackSeaCom)*, Batumi, Georgia, Jul. 2013, to appear.

[C3] Yutao Sui, A. Papadogiannis, W. Yang, and T. Svensson, "Performance comparison of fixed and moving relays under co-channel interference," in *IEEE Globecom Workshops*, Anaheim, USA, Dec. 2012, pp. 574–579.

[C4] W. Yang, G. Durisi, and E. Riegler, "Unitary isotropically distributed inputs are not capacity-achieving for large-MIMO fading channels," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Boston, MA, U.S.A., Jul. 2012, pp. 1717–1721. (Student paper award).

References

[1] Ericsson, "Ericsson mobility report: on the pulse of the networked society," Nov. 2012. [Online]. Available: http://www.ericsson.com/res/docs/2012/ericsson-mobility-report-november-2012.pdf

[2] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[3] H. Huh, S.-H. Moon, Y.-T. Kim, I. Lee, and G. Caire, "Multi-cell MIMO downlink with cell cooperation and fair scheduling: A large-system limit analysis," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 7771–7786, Dec. 2011.

[4] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[5] A. Barbieri, P. Gaal, S. Geirhofer, T. Ji, D. Malladi, Y. Wei, and F. Xue, "Coordinate downlink multi-point communications in heterogeneous 4G cellular networks," in *Proc. Workshop on Inform. Theory and its Appl. (ITA)*, San Diego, CA, U.S.A., Feb. 2012, pp. 7–16.

[6] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New Jersey: Wiley, 2006.

[7] I. C. Abou-Faycal, M. D. Trott, and S. Shamai (Shitz), "The capacity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1301, May 2001.

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

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

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

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

[12] G. Durisi and H. Bölcskei, "High-SNR capacity of wireless communication channels in the noncoherent setting: A primer," *Int. J. Electron. Commun. (AEÜ)*, vol. 65, no. 8, pp. 707–712, Aug. 2011.

[13] B. M. Hochwald and T. L. Marzetta, "Unitary space–time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.

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

[15] A. Peled and A. Ruiz, "Frequency domain data transmission using reduced computational complexity algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 5, Denver, CO, USA, Apr. 1980, pp. 964–967.

[16] 3GPP TS 36.212, "Technical specification group radio access network; evolved universal terrestrial radio access (E-UTRA); multiplexing and channel coding (release 10)," Dec. 2012. [Online]. Available: http://www.3gpp.org/

[17] R. A. Costa, M. Langberg, and J. Barros, "One-shot capacity of discrete channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, USA, Jun. 2010, pp. 211–215.

[18] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inform. Theory*, vol. 4, no. 4, pp. 2–22, 1954.

[19] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, May 1959.

[20] R. L. Dobrushin, "Mathematical problems in the Shannon theory of optimal coding of information," in *Proc. 4th Berkeley Symp. Math. Statist. and Prob.*, vol. 1, 1961, pp. 211–252.

[21] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[22] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

[23] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

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

[25] G. Durisi, V. I. Morgenshtern, and H. Bölcskei, "On the sensitivity of continuous-time noncoherent fading channel capacity," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6372–6391, Oct. 2012.

[26] L. H. Ozarow, S. S. (Shitz), and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 359–378, May 1994.

[27] G. Koliander, E. Riegler, G. Durisi, V. I. Morgenshtern, and F. Hlawatsch, “A lower bound on the noncoherent capacity pre-log for the MIMO channel with temporally correlated fading,” in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Oct. 2012, pp. 1198–1205.

[28] A. Lapidoth, “On the high SNR capacity of stationary Gaussian fading channels,” in *Proc. Allerton Conf. Commun., Contr., Comput.*, Monticello, IL, Oct. 2003, pp. 410–419.

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

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

[31] B. Hassibi and T. L. Marzetta, “Multiple-antennas and isotropically random unitary inputs: The received signal density in closed form,” *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1473–1484, Jun. 2002.

[32] B. M. Hochwald, T. L. Marzetta, T. J. Richardson, W. Sweldens, and R. Urbanke, “Systematic design of unitary space-time constellations,” *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 1962–1973, Sep. 2000.

[33] A. Ashikhmin and R. Calderbank, “Grassmannian packings from operator Reed–Muller codes,” *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5689–5714, Nov. 2010.

[34] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed. New York: Dover: Government Printing Office, 1972.

[35] K. Takeuchi, R. R. Müller, M. Vehkaperä, and T. Tanaka, “On an achievable rate of large Rayleigh block-fading MIMO channels with no CSI,” Nov. 2011. [Online]. Available: http://arxiv.org/abs/1111.4626v1

[36] V. I. Morgenshtern, G. Durisi, and H. Bölcskei, “The SIMO pre-log can be larger than the SISO pre-log,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Austin, TX, U.S.A., Jun. 2010, pp. 320–324.

[37] E. Riegler, V. I. Morgenshtern, G. Durisi, S. Lin, B. Sturmfels, and H. Bölcskei, “Noncoherent SIMO pre-log via resolution of singularities,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Saint Petersburg, Russia, Aug. 2011, pp. 2149–2153.

[38] T. Koch and A. Lapidoth, “Gaussian fading is the worst fading,” *IEEE Trans. Inf. Theory*, vol. 3, no. 56, pp. 1158–1165, Mar. 2010.

[39] S. Verdú and T. S. Han, “A general formula for channel capacity,” *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[40] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[41] E. Abbe, S.-L. Huang, and İ. E. Telatar, "Proof of the outage probability conjecture for MISO channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2596–2602, May 2013.

[42] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication.* Cambridge, U.K.: Cambridge Univ. Press, 2005.

[43] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with non-vanishing error probability," *IEEE Trans. Inf. Theory*, Dec. 2012, submitted.

[44] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Dispersion of the Gilbert-Elliott channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1829–1848, Apr. 2011.

[45] Y. Polyanskiy, "Finite blocklength methods in channel coding," in *2013 IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, tutorial. [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/ISIT13_tutorial.pdf

[46] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philos. Trans. Roy. Soc. London A*, vol. 231, pp. 289–337, 1933.

[47] Y. Polyanskiy and S. Verdú, "Scalar coherent fading channel: dispersion analysis," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Saint Petersburg, Russia, Aug. 2011, pp. 2959–2963.

[48] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.