

Optimal Designs for Decentralized Transmission with Asymmetric Channel State Information

Dissertation

submitted to

Sorbonne Université

*in partial fulfillment of the requirements for the degree of
Doctor of Philosophy*

Author:

Lorenzo MIRETTI

Successfully defended on the 24th of September 2021, before a committee composed of:

Reviewer	Prof. Erik G. LARSSON	<i>Linköping University, Sweden</i>
Reviewer	Prof. Serdar YÜKSEL	<i>Queen's University, Canada</i>
Examiner	Prof. Mari KOBAYASHI	<i>CentraleSupélec, France</i>
Examiner	Prof. Luca SANGUINETTI	<i>University of Pisa, Italy</i>
Examiner	Prof. Petros ELIA	<i>EURECOM, France</i>
Thesis Director	Prof. David GESBERT	<i>EURECOM, France</i>



Conceptions Optimales pour la Transmission Décentralisée avec Information Asymétrique sur l'État du Canal

Thèse

soumise à

Sorbonne Université

pour l'obtention du Grade de Docteur

Présenté par:

Lorenzo MIRETTI

Soutenance de thèse effectuée le 24 Septembre 2021, devant une jury composé de:

Rapporteur	Prof. Erik G. LARSSON	<i>Linköping University, Sweden</i>
Rapporteur	Prof. Serdar YÜKSEL	<i>Queen's University, Canada</i>
Examineur	Prof. Mari KOBAYASHI	<i>CentraleSupélec, France</i>
Examineur	Prof. Luca SANGUINETTI	<i>University of Pisa, Italy</i>
Examineur	Prof. Petros ELIA	<i>EURECOM, France</i>
Directeur de Thèse	Prof. David GESBERT	<i>EURECOM, France</i>



Abstract

The performance of wireless networks can be substantially enhanced by allowing cooperation among geographically distributed transmitters. However, in order to capitalize on transmitter cooperation in a scalable and cost-effective manner, future generation networks are expected to decentralize an increasing number of operations which were originally conceived for centrally controlled systems. On the physical layer side, decentralizing the transmission opens a Pandora's box of research problems dealing with the possibly limited sharing of crucial control information, e.g., about the channel state. Unfortunately, as of today, very little is known on how to optimally design decentralized transmission techniques operating under asymmetry of information. Therefore, for transmitter cooperation to find its rightful place within future wireless standards, it is imperative to place the understanding of these techniques on a more solid ground.

Starting from rigorous information theoretical models, the first part of this thesis extends known coding theorems for centralized systems to decentralized transmission with *distributed CSIT*, that is, by assuming that encoding is done on the basis of transmitter-specific channel state information. With this background at hand, as a first main contribution, we show that *distributed precoding* of Gaussian codewords achieves the capacity of a decentralized MIMO fading channel towards a single receiver, where the CSI is acquired through asymmetric feedback links. Surprisingly, we demonstrate that it may be suboptimal to send a number of data streams bounded by the number of transmit antennas as typically considered in a centralized setup.

As a second main contribution, we then move to the problem of distributed precoding design for systems with multiple receivers. Motivated by duality arguments, we introduce a novel scheme, coined *team minimum mean-square error* (TMMSE) precoding, which rigorously generalizes classical centralized MMSE precoding to distributed CSIT. Building on the so-called *theory of teams*, we derive a set of necessary and sufficient conditions for optimal TMMSE precoding, in the form of an infinite dimensional linear system of equations. These optimality conditions are further specialized to cell-free massive MIMO networks, and explicitly solved for two important examples, i.e., the classical case of local CSI and the case of unidirectional CSI sharing along a serial fronthaul. In both cases, our optimal distributed design outperforms the available methods heuristically adapted from centralized transmission theory.

Résumé

Les performances des réseaux sans fil peuvent être considérablement améliorées en permettant la coopération entre des émetteurs géographiquement distribués. Toutefois, afin de profiter de la coopération entre émetteurs de manière évolutive et avec un bon rapport coût-efficacité, les réseaux de la future génération devraient décentraliser un nombre croissant d'opérations conçues à l'origine pour des systèmes à contrôle centralisé. Du côté de la couche physique, la décentralisation de la transmission ouvre une boîte de Pandore pleine de problèmes de recherche concernant le partage potentiellement limité d'informations de contrôle cruciales, par exemple sur l'état du canal. Malheureusement, à ce jour, on sait très peu sur comment concevoir de manière optimale des techniques de transmission décentralisées opérant avec asymétrie d'information. Par conséquent, pour que la coopération entre émetteurs trouve la place qui lui revient dans les futures spécifications des réseaux sans fil, il est impératif de donner à la compréhension de ces techniques une base plus solide.

En se servant de modèles rigoureux fondés sur la théorie de l'information, la première partie de cette thèse étend les théorèmes de codage centralisés déjà existants à la transmission décentralisée avec *CSIT distribué*, c'est-à-dire, en supposant que le codage est effectué sur la base d'informations sur l'état du canal spécifiques à chaque émetteur. Fort de ce cadre, nous montrons, comme première contribution principale, que le *précodage distribué* des codes Gaussiens atteint la capacité d'un canal MIMO décentralisé à évanouissement vers un récepteur unique, où le CSIT est acquis par des liens de rétroaction asymétriques. Curieusement, nous démontrons qu'il peut être sous-optimal d'envoyer un nombre de flux de données limité par le nombre d'antennes de transmission, comme il est généralement envisagé dans une configuration centralisée.

Comme deuxième contribution principale, nous abordons ensuite le problème de la conception du précodage distribué pour les systèmes à récepteurs multiples. Motivés par des arguments de dualité, nous introduisons un nouveau schéma, appelé précodage *team minimum mean-square error (TMMSE)*, qui généralise rigoureusement le précodage *MMSE* centralisé classique au CSIT distribué. Sur la base de la *théorie des équipes*, nous dérivons un ensemble de conditions nécessaires et suffisantes pour un précodage TMMSE optimal, sous la forme d'un système d'équations linéaire de dimension infinie. Ces conditions d'optimalité sont ensuite adaptées aux réseaux MIMO massifs *cell-free*, et résolues explicitement pour deux exemples importants, notamment le cas classique du CSIT local, et le cas du partage unidirectionnel du CSIT le long d'un fronthaul série. Dans les deux cas, notre conception distribuée optimale surclasse les méthodes disponibles adaptées heuristiquement de la théorie de la transmission centralisée.

Acknowledgements

There are probably too many persons and institutions that I should acknowledge for their contributions to this doctoral thesis, since it is the product of decades of top-quality public education in Italy, France, Germany, and the European Union in general. Needless to say, the constant support of my family and loved ones has also been absolutely crucial. In a world of unequal opportunities, I feel extremely grateful and privileged to have reached this milestone.

However, this is an engineering thesis, and engineers do not like complexity, so let us try to narrow the acknowledgements down to the essentials. First and foremost, a big share of merit goes to my supervisor David Gesbert. In addition to his proven talent as both a researcher and a manager, I am particularly impressed by his persistence in pursuing the visions he is convinced of, such as the one behind the topic of my thesis. I am quite confident that this work will repay his persistence, and help transform his vision into a new and exciting field of research. Second, special thanks go to Paul de Kerret, Antonio Bazco, and Renato Cavalcante, for guiding me in my first steps into the wild world of scientific research. Third, I would like to thank my coauthors Mary Kobayashi and Emil Björnson for their enthusiasm and technical help in developing the content of this thesis. Fourth, another good share of merit goes to the reviewers and to the members of the jury, who, directly or indirectly, considerably contributed to the results of this thesis.

Last but not least, I definitely need to thank my girlfriend Verjina and all the amazing friends I met on the beautiful Côte d'Azur. Without you, my journey towards the Ph.D. would have seemed much more like a lonely time than the real fun and socially fulfilling experience it actually was.

Lorenzo Miretti

Contents

1	Introduction	1
1.1	Cellular networks and beyond	1
1.2	Distributed CSIT	3
1.3	Contributions and thesis outline	5
1.4	List of publications	8
2	Information Theoretic Models	9
2.1	Decentralized transmission with distributed CSIT	9
2.1.1	System model	9
2.1.2	Problem statement	11
2.2	Decentralized MIMO fading channels	11
2.3	Extensions to partial message sharing	12
2.3.1	Multiple-access channels with common message	13
2.3.2	User-centric cooperation clusters	14
3	Coding Theorems for Decentralized Transmission	15
3.1	Decentralized transmission towards a single RX	15
3.1.1	Capacity and Shannon strategies	15
3.1.2	Example: channel with additive binary inputs and state	18
3.1.3	Stationary and ergodic state processes	19
3.1.4	Multiple-access channels	21
3.2	Decentralized transmission towards multiple RXs	23
3.2.1	Broadcast channels and Shannon strategies	23
3.2.2	Distributed linear precoding	26
4	Ergodic Capacity under Asymmetric Feedback	29
4.1	Optimality of distributed linear precoding with an unconventional number of data streams	29
4.2	Capacity computation via convex optimization	32
4.2.1	Example: asymmetric feedback rates	34
4.3	Further comments on the optimal number of data streams	36

4.3.1	Example: asymmetric feedback rates (cont.)	38
4.4	Extension to multiple-access channels	38
4.5	Joint precoding and feedback design	40
4.5.1	Example: asymmetric feedback rates (cont.)	41
5	Team Theory for Distributed Precoding Design	44
5.1	Distributed precoding design	44
5.1.1	A challenging problem	44
5.1.2	Definitions and assumptions	45
5.2	Team MMSE precoding	47
5.2.1	Achievable rates via uplink-downlink duality	47
5.2.2	Quadratic teams for distributed precoding design	49
5.3	Applications to cell-free massive MIMO	51
5.3.1	No CSIT sharing	53
5.3.2	Unidirectional CSIT sharing	54
5.3.3	Asymptotic results and competing unidirectional recursive schemes	57
5.4	Performance evaluation	58
5.4.1	Simulation setup	58
5.4.2	Comparison among different CSIT configurations	59
5.4.3	Comparison among local precoding schemes	60
5.4.4	Comparison among unidirectional recursive schemes	62
5.5	Extensions	64
5.5.1	Partial message sharing	64
5.5.2	Per-TX power constraint	66
5.5.3	Performance evaluation (cont.)	67
6	Conclusion	70
6.1	Lessons learned	70
6.2	Lessons yet to be learned	70
A	Team Decision Theory	72
B	Collection of Proofs	75

Mathematical Notation

\mathbb{R}	set of real numbers
\mathbb{C}	set of complex numbers
\mathbb{N}	set of natural numbers
\mathbb{R}_+	set of non-negative real numbers
\mathbb{S}^n	set of Hermitian symmetric matrices of dimension n
\mathbb{S}_+^n	set of Hermitian positive-semidefinite matrices of dimension n

We reserve calligraphic letters (e.g., \mathcal{A}) for sets, and italic letters (e.g., a , A) for scalars and elements of generic sets. An n -sequence $(a_1, \dots, a_n) \in \mathcal{A}^n$ is denoted by a^n . We use lower case boldface letters (e.g., \mathbf{a}) for column vectors, and upper case boldface letters (e.g., \mathbf{A}) for matrices. Random variables are typographically distinguished from their realizations as follows:

A	random variable taking values in a generic set \mathcal{A}
\mathbf{a}	random column vector taking values in $\mathbb{C}^{n \times 1}$
\mathbf{A}	random matrix taking values in $\mathbb{C}^{n \times m}$
a	realization of A
\mathbf{a}	realization of \mathbf{a}
\mathbf{A}	realization of \mathbf{A}

Functions of random variables are equivalently denoted using the two notations for random and deterministic quantities (e.g., $F(X)$ or $f(X)$), since a function of a random variable is itself a random variable.

\mathbf{a}^H	Hermitian transpose of \mathbf{a}
\mathbf{a}^T	transpose of \mathbf{a}
$\ \mathbf{a}\ $	Euclidean norm of \mathbf{a}
$\ \mathbf{A}\ _F$	Frobenius norm of \mathbf{A}
$[\mathbf{A}]_{i,j}$	(i, j) -th entry of \mathbf{A}
\mathbf{I}_n	identity matrix of dimension n ; the subscript may be omitted when no confusion arises
$\text{tr}(\mathbf{A})$	trace of \mathbf{A}
$\Re(\mathbf{A})$	real part of \mathbf{A}
$\Im(\mathbf{A})$	imaginary part of \mathbf{A}
\mathbf{e}_i	standard column selector, i.e., the i -th column of \mathbf{I}
$\mathbf{0}_{n \times m}$	$n \times m$ matrix of all zeros; the subscript may be omitted when no confusion arises
\preceq, \succeq	generalized inequalities w.r.t. the cone of nonnegative Hermitian matrices.
\prec, \succ	strict generalized inequalities w.r.t. the cone of nonnegative Hermitian matrices.
$:=, =:$	definitions from the right and from the left
$\prod_{i=l'}^l \mathbf{A}_i$	for integers $l \geq l' \geq 1$, <i>left</i> product chain, $\mathbf{A}_l \mathbf{A}_{l-1} \dots \mathbf{A}_{l'}$ of $l - l' + 1$ ordered matrices of compatible dimension. We adopt the convention $\prod_{i=l'}^l \mathbf{A}_i = \mathbf{I}$ for $l < l'$.
$\text{diag}(\mathbf{A}_1, \dots, \mathbf{A}_n)$	block-diagonal matrix with the matrices $\mathbf{A}_1, \dots, \mathbf{A}_n$ on its diagonal.
$\text{vec}(\mathbf{A})$	column vector obtained by stacking column-wise the elements of \mathbf{A}
$E[A]$	expectation of A
$\text{Var}[A]$	variance of A
$p(a)$	probability mass (density) function of a discrete (continuous) random variable A
$\Pr[\mathcal{E}]$	probability of an event \mathcal{E}
$H(A)$	entropy of A
$h(A)$	differential entropy of A
$I(A; B)$	mutual information between A and B
$\mathcal{CN}(\boldsymbol{\mu}, \mathbf{K})$	Circularly symmetric complex Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix \mathbf{K}

Acronyms and Abbreviations

TX	transmitter
RX	receiver
DL	downlink
UL	uplink
TDD	time division duplex
FDD	frequency division duplex
MIMO	multiple-input and multiple-output
CSI	channel state information
CSIT	channel state information at the transmitter
CSIR	channel state information at the receiver
MSE	mean-square error
MMSE	minimum mean-square error
WMMSE	weighted minimum mean-square error
TMMSE	team minimum mean-square error
MRT	maximum-ratio transmission
ZF	zero-forcing
LTE	long term evolution
5G	5-th generation cellular networks
3GPP	3-rd generation partnership project
MAC	multiple-access channel
BC	broadcast channel
TIN	treating interference as noise
DoF	degrees-of-freedom
SNR	signal-to-noise ration
SINR	signal-to-interference-plus-noise ratio
LoS	line-of-sight
NLoS	non line-of-sight
UC	user-centric
SGD	stochastic gradient descent
DM	decision maker
OFDM	orthogonal frequency division multiplexing
RAN	radio access network
C-RAN	cloud radio access network
CoMP	coordinated multi-point

w.r.t. with respect to
w.l.o.g. without loss of generality
a.s. almost surely
i.i.d. independent and identically distributed

Chapter 1

Introduction

Wireless communications networks have become an essential component of modern societies. From Guglielmo Marconi's first commercially successful radio transmission system, to modern 5G cellular networks, the worldwide mobile data traffic has been increasing at an exponential pace, reaching 49 exabyte exchanged monthly at the end of 2020 [1]. This trend shows no sign of slowing down, as we keep advancing towards fully networked societies where ultra-broadband data services need to be accessible ubiquitously, reliably, and efficiently. A crucial quest for engineers and researchers is therefore the identification of technologies able to meet these increasingly demanding requirements. The goal of this thesis is to improve and deepen the understanding of one of the candidate technologies, described next.

1.1 Cellular networks and beyond

The cellular paradigm Traditionally, wireless communication systems have been designed by following a *cellular* paradigm, i.e., by clustering the users into disjoint cells, typically corresponding to specific portions of the coverage area, with every cell being served by a single base station. In order to increase capacity, every new generation of these systems has been generally characterized by higher power consumption, denser cell deployment, and higher spectral efficiency due to sensible software and hardware improvements [2]. From a technological point of view, the outstanding evolution experienced by radio access networks (RAN) in the last two decades has been essentially driven by: (i) the revolutions in *channel coding* theory [3], marked by the discovery of powerful codes finally delivering the data rates promised by Claude Shannon's *information* theory [4]; (ii) significant signal processing advances, e.g., in *multi-carrier* modulation formats and in *multiple-input multiple-output* (MIMO) antenna techniques [5], which allow to successfully mitigate the detrimental characteristics of the wireless channel. Remarkably, MIMO

technologies are also a key innovation of the current 5G standard, where the new *massive* MIMO equipments [6] are finally turning into a commercially attractive reality the academic concept of *multi-user* communications [7].

Beyond the cellular paradigm The cellular paradigm presents inherent limitations which makes it unsuitable for meeting the expected traffic demand and quality of service requirements of future societies [8], and many current urban deployments are already operating close to their ultimate limits. In fact, relying on aggressive spectrum reuse and network densification to increase capacity will inevitably face *inter-cell* interference as a major limiting factor. Despite the aforementioned technological advances improve the performance within each cell, they are essentially unable to resolve this limitation, unless a profound rethinking of the cellular network architecture is performed. In this regard, future generation systems are expected to implement advanced cooperative communication techniques, in particular by letting geographically distributed base stations jointly serve their users. The main benefit of cooperation is to sensibly improve the performance of *cell-edge* users, the ones which are most impacted by inter-cell interference, ideally by completely removing the concept of cell-edge. This idea has been pioneered by the academia since the early '90s [9, 10], where cooperation is introduced to model the uplink (UL) and downlink (DL) of the entire network as a single MIMO *multiple-access channel* (MAC) and *broadcast channel* (BC) [7], respectively. It has been then extensively explored and enriched over the last decades, under the names of *network MIMO* [11, 12], *cloud RAN* (C-RAN) [13, 14], and, more recently, *cell-free massive MIMO* [15, 8]. This nomenclature essentially reflects differences in the implementation and sophistication of the underlying MIMO technology, and in the definition of cell or service area.

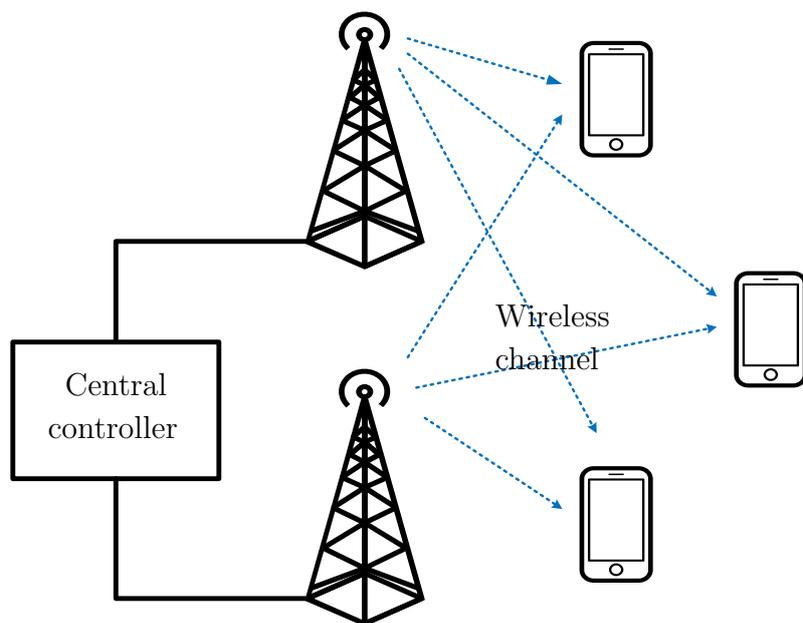
Decentralized transmission The development of cooperative wireless networks has been already attempted in the past, but with very little or no commercial success. Major evidences of this failure are the disappointing adoption of the *coordinated multi-point* (CoMP) techniques standardized in LTE-Advanced networks, and the complete dominance of cellular massive MIMO technologies in the 5G standardization effort. The most important reason for this lack of success is that a commercially attractive deployment of cooperative wireless networks is currently prevented by the severe scalability issue arising from network-wide processing [16]. Specifically, the excessive amount of data and channel state information (CSI), needed to be timely shared for implementing centrally controlled and fully cooperative networks such as in the original network MIMO or C-RAN concepts, often becomes the main bottleneck when practical fronthaul capacity constraints are introduced. Studying decentralized network architectures with more realistic

cooperation regimes, entailing limited data and CSI sharing, is hence of fundamental importance for making network cooperation an attractive technology for next generation systems [16, 17]. In this context, this thesis explores the problem of optimal decentralized transmission under asymmetric information, whose main underlying assumption is depicted in Figure 1.1 and outlined in the next section.

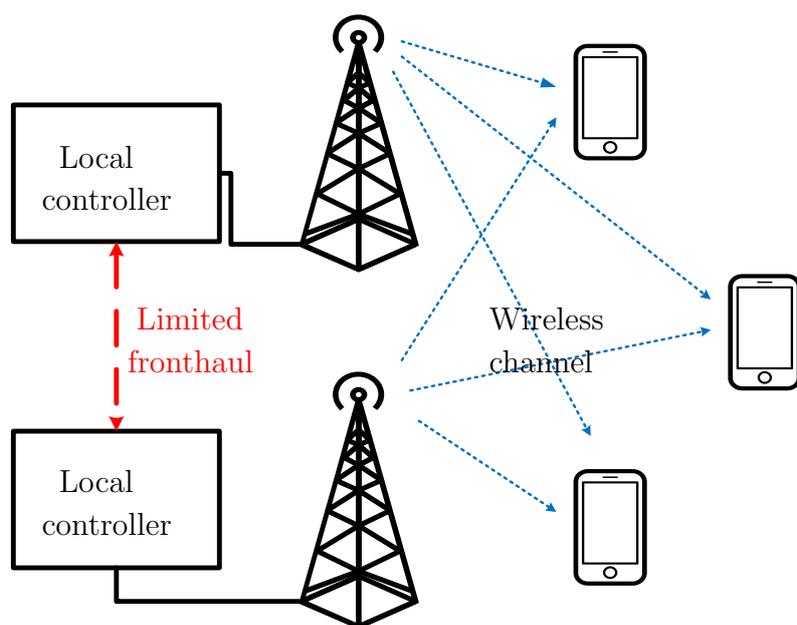
1.2 Distributed CSIT

We consider a DL setup with a set of transmitters (TXs) jointly communicating with a set of receivers (RXs) through a wireless channel governed by a random fading state. In most parts of this thesis, we assume a cooperation regime with full message sharing and general *distributed* CSI at the transmitters (CSIT) [18, 19, 20], that is, we let each TX operate on the basis of possibly different estimates of the *global* channel state obtained through some arbitrary CSIT acquisition and sharing mechanism. This assumption is relevant, e.g., for all service situations where the CSIT sharing burden dominates the fronthaul overhead. For instance, it is suitable in case of rapidly varying channels due to user mobility, or when delay-tolerant data is proactively made available at the TXs using caching techniques (see [20] and references therein for a detailed discussion). As an extreme example, a cooperation regime with full message sharing and no CSIT sharing (a configuration here referred to as *local* CSIT) is perhaps the leading motivation behind the early development of the now popular cell-free massive MIMO paradigm [15]. This paradigm combines the benefits of ultra-dense networks with simple yet effective TX cooperation schemes, and emerged as a promising evolution of the network MIMO and C-RAN concepts. The distributed CSIT assumption also covers extensions of [15] to more complex setups ranging from partial to full CSIT sharing (see, e.g., [21] [22]).

Clearly, these cooperation regimes are still far from being scalable, since they all assume network-wide message sharing. Splitting the network into clusters of cooperating TXs [23, 24, 22], possibly dynamically and with a user-centric approach [25, 26, 27, 28], and applying similar transmission techniques assuming full message sharing within each cluster, emerged as a viable solution for implementing scalable cooperation regimes in practical systems. All the discussed approaches do not consider complementary service situations where CSIT can be more easily shared than messages, hence entering the realm of interference *coordination* or *alignment* [16, 29, 30]. This thesis will also discuss partial message sharing, but only on simplified setups without interference, or by following the aforementioned network clustering approach. Nevertheless, the results presented in this thesis can be seen as a first step towards a more general theory jointly covering partial message and CSIT sharing.



(a) Centralized transmission



(b) Decentralized transmission

Figure 1.1: Pictorial representation of a cooperative wireless communication system using (a) centralized transmission, and (b) decentralized transmission. In (a), each TX is connected through infinite capacity links to the same controller. In (b), the TXs are connected to local controllers coordinating the transmission on the basis of locally available information, partially shared through a fronthaul of limited capacity.

Although the importance of decentralized transmission has been acknowledged in the literature, a satisfactory understanding of systems with distributed CSIT is still missing. Most of the available information theoretical results rely on asymptotic signal-to-noise ratio (SNR) tools [18, 20]. These results concisely identify challenges and opportunities behind the design of decentralized transmission schemes that cater for distributed CSIT, but they are based on ideal CSIT models and do not typically lead to practical schemes for many network setups. On the other hand, most of the available practical schemes are essentially based on heuristic adaptations of known centralized precoding designs such as maximum-ratio transmission (MRT), zero-forcing (ZF), or minimum mean-square error (MMSE) precoding [15, 21, 28].

1.3 Contributions and thesis outline

Given the setup and assumptions outlined above, this thesis makes considerable progress in both the theoretical understanding and in the practical design of decentralized transmission techniques. Overall, the findings of this thesis suggest that, while certain aspects of centralized transmission theory carry over to decentralized setups, other aspects should be handled with special care. An informal summary of the main contributions supporting the above conclusion is given in what follows.

Information theoretic foundations We first present in Chapter 2 and Chapter 3 some necessary fundamentals which, starting from practically relevant information theoretic models, revisits classical centralized channel coding theorems (see, e.g., [7]) in light of the distributed CSIT assumption. In some cases, we prove novel channel capacity theorems and identify known design choices which optimally carry over to decentralized setups. The most important one is perhaps given by the following capacity formula:

$$C = \max_{\substack{p(u) \\ x_l = f_l(u, s_l)}} I(U; Y), \quad (1.1)$$

which extends the classical result of Shannon [31] on coding with causal CSIT to a decentralized memoryless channel with output Y , inputs (X_1, \dots, X_L) , and where each of the input must be formed on the basis of a shared message W and the corresponding entry of the distributed CSIT tuple (S_1, \dots, S_L) . Shannon's formula (i.e., (1.1) with $L = 1$) forms the theoretical basis for important tools in wireless communications such as the concepts of linear precoding and ergodic capacity. Although simple, its decentralized version (1.1) was not available in the literature.

Ergodic capacity under asymmetric feedback As a first main contribution of this thesis, in Chapter 4 we establish the capacity of a single-RX decentralized MIMO fading channel where each TX acquires its CSIT through asymmetric feedback links. In particular, we prove that a decentralized version of classical linear precoding, called *distributed precoding* [19], is capacity achieving. We here informally illustrate the result by focusing for simplicity on $L = 2$ single antenna TXs communicating over a fading channel

$$Y = \mathbf{g}^H \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \text{Gaussian noise}, \quad (1.2)$$

on the basis of quantized measurements $S_l = q_l(\mathbf{g})$ of the random fading state \mathbf{g} , obtained via explicit feedback from a single antenna RX, and subject to a per-TX power constraint $\mathbb{E}[|X_l|^2] \leq P_l$. We show that the capacity formula in (1.1) can be simplified to

$$C = \max_{\substack{\mathbb{f}_l(S_l) \in \mathbb{C}^d \\ \mathbb{E}[\|\mathbb{f}_l(S_l)\|^2] \leq P_l}} \mathbb{E} \left[\log \left(1 + \mathbf{g}^H \boldsymbol{\Sigma}(S_1, S_2) \mathbf{g} \right) \right], \quad (1.3)$$

where

$$\boldsymbol{\Sigma}(S_1, S_2) = \begin{bmatrix} \mathbb{f}_1^H(S_1) \\ \mathbb{f}_2^H(S_2) \end{bmatrix} \begin{bmatrix} \mathbb{f}_1(S_1) & \mathbb{f}_2(S_2) \end{bmatrix} \quad (1.4)$$

is the conditional input covariance induced by a linear precoding scheme

$$X_l = \mathbb{f}_l^H(S_l) \mathbf{u} \quad (1.5)$$

applied at each TX l to a shared data-bearing vector $\mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ of dimension d . As a key element for the achievability proof, we let d in (1.3) be an optimization variable upper bounded by the CSIT cardinalities as

$$d \leq D = |\mathcal{S}_1| + |\mathcal{S}_2|. \quad (1.6)$$

Furthermore, and in surprising contrast to a centralized CSIT setup, we show that the traditional design choice (see, e.g., [32]) of bounding the number of precoded data streams by the number of TX antennas (i.e., $d \leq 2$) may be strictly suboptimal. As an interesting byproduct, we then show that increasing the number of data streams up to D allows to transform the otherwise difficult capacity computation problem into a convex form. This technique is also applied to the related problem of joint precoding and feedback design under asymmetric feedback rate constraints. Finally, we consider partial message sharing and similarly derive the capacity region of a fading multiple-access channel with asymmetric feedback.

Team theory for distributed precoding design In Chapter 5 we depart from information theoretical optimality, and consider the difficult problem of optimal single-stream distributed precoding design for decentralized MIMO fading channels towards multiple RXs. We propose a novel scheme called *team* MMSE (TMMSE) precoding, which rigorously generalizes centralized MMSE precoding to distributed CSIT. Leveraging UL-DL duality results, the proposed scheme is shown to span the Pareto boundary of the achievable rate region predicted by the popular *hardening* bound [6], assuming a sum power constraint. The second main contribution of this thesis is the derivation of a useful set of necessary and sufficient conditions for optimal TMMSE precoding design in the form of an infinite dimensional linear system of equations. The key novelty lies in the exploitation of selected elements from the *theory of teams*, a mathematical framework for multi-agent coordinated decision making under asymmetry of information [33, 34, 35]. These optimality conditions are further applied to cell-free massive MIMO networks, where each TX l forms its CSIT S_l using local measurements of its local state \mathbb{H}_l , which models the fading channel from its N antennas to all RXs, and information about other channels \mathbb{H}_j , $j \neq l$, obtained via the fronthaul according to an arbitrary CSIT sharing pattern. For illustration purposes, we here informally present the results by focusing on an arbitrary Pareto optimal point, and by assuming that each TX l has perfect knowledge of its local channel \mathbb{H}_l . In this case, the optimal (unnormalized) TMMSE precoder $\mathfrak{t}_{l,k}(S_l)$, applied at the N antennas of TX l to the data stream encoding the message for RX k , must satisfy the equations

$$\mathfrak{t}_{l,k}(S_l) = \mathbb{T}_l \left(\mathbf{e}_k - \sum_{j \neq l} \mathbf{E} \left[\mathbb{H}_j \mathfrak{t}_{j,k}(S_j) \middle| S_l \right] \right) \quad \text{a.s.}, \quad l = 1, \dots, L, \quad (1.7)$$

where $\mathbb{T}_l = (\mathbb{H}_l^H \mathbb{H}_l + P^{-1} \mathbf{I})^{-1} \mathbb{H}_l^H$ is a *local* MMSE precoder with nominal SNR parameter P , and where the term inside the brackets takes into account the effects of the other TXs on the basis of the available CSIT S_l and statistical information. By solving (1.7) explicitly, we derive the optimal TMMSE precoders assuming no CSIT sharing

$$\mathfrak{t}_{l,k}(S_l) = \mathbb{T}_l \mathbf{C}_l \mathbf{e}_k, \quad l = 1, \dots, L, \quad (1.8)$$

where $\mathbf{C}_l \in \mathbb{C}^{K \times K}$ are statistical stages solving $\mathbf{C}_l + \sum_{j \neq l} \mathbf{E} [\mathbb{H}_l \mathbb{T}_l] \mathbf{C}_j = \mathbf{I} \forall l$. We show that, by optimally exploiting channel statistics, (1.8) considerably improves upon competing local precoding schemes reviewed in [8]. Furthermore, we derive the optimal TMMSE precoders by assuming that CSIT is shared unidirectionally along a serial fronthaul, an architecture also known as a *radio stripe* [27]. The proposed scheme takes the following multi-stage form

$$\mathfrak{t}_{l,k}(S_l) = \mathbb{T}_l \mathbb{V}_l \bar{\mathbb{V}}_{l-1} \dots \bar{\mathbb{V}}_1 \mathbf{e}_k, \quad l = 1, \dots, L, \quad (1.9)$$

where $(\mathbb{V}_l, \bar{\mathbb{V}}_l)$ depends only on the local CSIT \mathbb{H}_l unidirectionally shared from TX l to TX L . Interestingly, the scheme in (1.9) can be efficiently implemented in a sequential fashion, an idea that has been explored already in [27, 36] for uplink processing, and in [37] under a different cellular context. Finally, we discuss the extension to partial message sharing via user-centric network clustering [28], and the use of more restrictive power constraints.

1.4 List of publications

The work behind this thesis resulted in the following publications:

- [J1] L. Miretti, M. Kobayashi, D. Gesbert and P. De Kerret, “Cooperative multiple-access channels with distributed state information,” *IEEE Transactions on Information Theory*, 2021.
- [J2] L. Miretti, E. Björnson and D. Gesbert, “Team MMSE precoding with applications to cell-free massive MIMO,” submitted, *IEEE Transactions on Wireless Communications*, 2021.
- [C1] L. Miretti, P. De Kerret and D. Gesbert, “On the fundamental limits of cooperative multiple-access channels with distributed CSIT,” *IEEE Information Theory Workshop (ITW)*, 2019.
- [C2] L. Miretti, M. Kobayashi and D. Gesbert, “Precoding for cooperative MIMO channels with asymmetric feedback,” *IEEE International Conference on Communications (ICC)*, 2020.
- [C3] L. Miretti and D. Gesbert, “Interference mitigation for cooperative MIMO channels with asymmetric feedback,” *54th Asilomar Conference on Signals, Systems, and Computers*, 2020.
- [C4] L. Miretti, E. Björnson and D. Gesbert, “Precoding for scalable cell-free massive MIMO with radio stripes,” *22nd IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2021.

Chapter 2 and 3 are mostly based on [J1], the introductory part of [J2], and unpublished results. Chapter 4 is based on [J1, C2]. Chapter 5 is based on [J2] and unpublished results. References [C1] and [C4] are preliminary versions of [J1] and [J2], respectively. The preliminary results of [C3] are only briefly mentioned since they have been overshadowed by the results in [C4, J2].

Chapter 2

Information Theoretic Models

This chapter details the models and performance metrics adopted throughout this thesis. Specifically, we introduce formal information theoretical tools for modelling and evaluating the performance of a communication system based on decentralized transmission with distributed CSIT, and present their application to modern wireless networks.

2.1 Decentralized transmission with distributed CSIT

2.1.1 System model

Channel Model We consider the state-dependent channel of Figure 2.1 with L TXs indexed by $\mathcal{L} := \{1, \dots, L\}$, K RXs indexed by $\mathcal{K} := \{1, \dots, K\}$, inputs $(X_1, \dots, X_L) \in \prod_{l=1}^L \mathcal{X}_l$, outputs $(Y_1, \dots, Y_K) \in \prod_{k=1}^K \mathcal{Y}_k$, state $S \in \mathcal{S}$, and distributed CSIT $(S_1, \dots, S_L) \in \prod_{l=1}^L \mathcal{S}_l$. We let an n -sequence of inputs, outputs, states, and CSITs be governed by the memoryless law

$$p(y_1^n, \dots, y_K^n | x_1^n, \dots, x_L^n, s^n) = \prod_{i=1}^n p(y_{1,i}, \dots, y_{K,i} | x_{1,i}, \dots, x_{L,i}, s_i) \quad (2.1)$$

$$p(s^n, s_1^n, \dots, s_L^n) = \prod_{i=1}^n p(s_i, s_{1,i}, \dots, s_{L,i}), \quad (2.2)$$

where $p(y_1, \dots, y_K | x_1, \dots, x_L, s)$ is a given state-dependent input-output distribution and $p(s, s_1, \dots, s_L)$ is a given state and CSIT distribution. We finally let the messages (W_1, \dots, W_K) be independently and uniformly distributed over the sets $\mathcal{W}_k := \{1, \dots, 2^{\lceil nR_k \rceil}\}$, $k \in \mathcal{K}$, where $R_k \geq 0$ is the rate of the message W_k intended

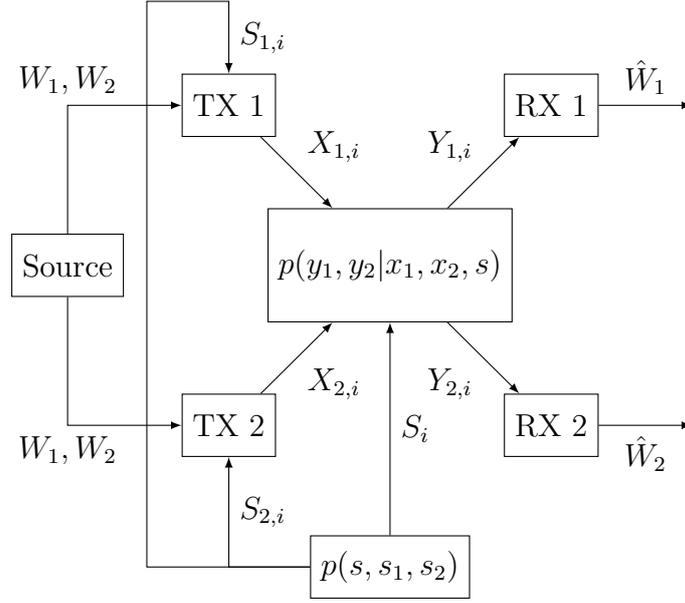


Figure 2.1: State-dependent channel with $L = 2$ cooperating TXs, $K = 2$ RXs, and causal distributed CSIT.

for RX k . Although not explicitly, the above model covers systems with arbitrary CSI at the RX (CSIR), which can be interpreted as a part of the outputs Y_k .

Encoding and Decoding A $(2^{nR_1}, \dots, 2^{nR_K}, n)$ block code of length n with causal and distributed CSIT is defined by a set of encoding functions

$$\phi_{l,i} : \prod_{k=1}^K \mathcal{W}_k \times \mathcal{S}_l^i \rightarrow \mathcal{X}_l, \quad l \in \mathcal{L}, \quad i = 1, \dots, n, \quad (2.3)$$

yielding the transmitted symbols $x_{l,i} = \phi_{l,i}(w_1, \dots, w_K, s_l^i)$, as well as a set of decoding functions

$$\psi_k : \mathcal{Y}_k^n \rightarrow \mathcal{W}_k, \quad k \in \mathcal{K}, \quad (2.4)$$

yielding the decoded messages $\hat{w}_k = \psi_k(y_k^n)$. Each encoder $l \in \mathcal{L}$ is subject to an average input cost constraint

$$\mathbf{E} [\eta_l^n(X_l^n)] \leq P_l, \quad \eta_l^n(x_l^n) := \frac{1}{n} \sum_{i=1}^n \eta_l(x_{l,i}), \quad (2.5)$$

where $\eta_l : \mathcal{X}_l \rightarrow \mathbb{R}_+$ is a single-letter cost function. For a given cost tuple (P_1, \dots, P_L) , a rate tuple (R_1, \dots, R_K) is said to be achievable if, for the considered channel, there exists a family of block codes defined as before such that

the probability of error $P_e^{(n)} := \Pr \left[\cup_{k \in \mathcal{K}} (\hat{W}_k \neq W_k) \right]$ satisfies

$$\lim_{n \rightarrow \infty} P_e^{(n)} = 0. \quad (2.6)$$

2.1.2 Problem statement

For a given cost tuple (P_1, \dots, P_L) , the closure of the set of all achievable rates (R_1, \dots, R_K) is the capacity-cost region $\mathcal{C}(P_1, \dots, P_L)$ of the considered channel. We also define the capacity $C_k(P_1, \dots, P_L)$ of RX k as the supremum of all achievable rates R_k . The optimistic goal of this thesis is to characterize the largest possible achievable region $\mathcal{R}(P_1, \dots, P_L) \subseteq \mathcal{C}(P_1, \dots, P_L)$ and the corresponding decentralized transmission schemes. When this implies the use of overly complicated techniques, we consider the alternative goal of characterizing the largest possible achievable region $\mathcal{R}(P_1, \dots, P_L)$ while restricting ourselves to more practical decentralized transmission schemes.

2.2 Decentralized MIMO fading channels

The most important application of the general model in Section 2.1 is to wireless multi-antenna fading channels, described in what follows. We assume that each TX and RX is equipped with N and M antennas respectively, and let an arbitrary channel use be governed by the MIMO channel law

$$\mathbf{y}_k = \sum_{l=1}^L \mathbb{H}_{k,l} \mathbf{x}_l + \mathbf{n}_k, \quad \forall k \in \mathcal{K}, \quad (2.7)$$

where $\mathbf{y}_k \in \mathbb{C}^K$ is the channel output at RX k , $\mathbb{H}_{k,l} \in \mathbb{C}^{M \times N}$ is a random matrix modelling the fading between TX l and RX k , $\mathbf{x}_l \in \mathbb{C}^N$ is the channel input at TX l , and $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is the independent noise at RX k . Furthermore, given the n -sequence $\mathbf{x}_l^n = (\mathbf{x}_{l,1}, \dots, \mathbf{x}_{l,n})$ of random inputs at each TX $l \in \mathcal{L}$, we assume an input cost constraint

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} [\|\mathbf{x}_{l,i}\|^2] \leq P_l, \quad (2.8)$$

which can be interpreted as a per-TX long-term average power constraint.

Fading State We map the random channel state $S \in \mathcal{S}$ in Section 2.1 to the *global* fading matrix $\mathbb{H} \in \mathbb{C}^{KM \times LN}$, with realizations given by

$$\mathbb{H} := \begin{bmatrix} \mathbb{H}_{1,1} & \dots & \mathbb{H}_{1,L} \\ \vdots & \ddots & \vdots \\ \mathbb{H}_{K,1} & \dots & \mathbb{H}_{K,L} \end{bmatrix}. \quad (2.9)$$

For most parts of this work, we do not specify the distribution of \mathbb{H} . However, we reasonably assume the channel submatrices corresponding to different TX-RX pairs to be mutually independent, and finite fading power $\mathbb{E}[\|\mathbb{H}\|_{\mathbb{F}}^2] < \infty$. For convenience, we also define the matrix

$$\mathbb{H}_l := \begin{bmatrix} \mathbb{H}_{1,l} \\ \vdots \\ \mathbb{H}_{K,l} \end{bmatrix}, \quad \forall l \in \mathcal{L}, \quad (2.10)$$

collecting the channel coefficients from TX l to all RXs, the *dual* channel matrices $\mathbb{G}_{k,l}^{\mathbb{H}} := \mathbb{H}_{k,l}$ and $\mathbb{G}^{\mathbb{H}} := \mathbb{H}$, and the matrix

$$\mathbb{G}_k^{\mathbb{H}} := [\mathbb{G}_{k,1}^{\mathbb{H}} \quad \dots \quad \mathbb{G}_{k,L}^{\mathbb{H}}], \quad \forall k \in \mathcal{K}, \quad (2.11)$$

collecting the channel coefficients from all TXs to RX k . Whenever appropriate, we use the vector notation in place of the matrix notation, e.g., by replacing $\mathbb{G}_{k,l}$ with $\mathbf{g}_{k,l}$ if RX k is equipped with a single antenna.

Ergodic Rates Together with the assumptions in Section 2.1, the above model assumes memoryless channels and that the transmission takes place over $n \gg 1$ i.i.d. realizations of the fading process \mathbb{H}^n . This setup is typically referred to as the *fast-fading* regime, and the resulting rates (R_1, \dots, R_K) are called *ergodic* rates [5].

Remark 2.1 (Correlated fading). *In modern wireless systems the fading process \mathbb{H}^n is often correlated, i.e., the memoryless assumption (2.2) does not hold. However, if \mathbb{H}^n is stationary and ergodic, this correlation can be neglected through ideal interleaving while achieving the same rates as in the i.i.d. fading case (see Section 3.1.3 for additional details specific to our setup). Although correlation may be exploited to achieve higher rates, in some cases the above operation incurs no loss of optimality [38, 39].*

Overall, the model and performance metric presented here are relevant, e.g., for wideband OFDM systems [5] targeting ultra-broadband, spectrally efficient, and ubiquitous connectivity without stringent latency constraints.

2.3 Extensions to partial message sharing

Although the main focus of this thesis is the study of the effect of partial CSIT sharing, some of the presented results can be extended to networks with partial message sharing. Specifically, we will also cover the following extensions of the model in Section 2.1, focusing on different aspects of the partial message sharing problem.

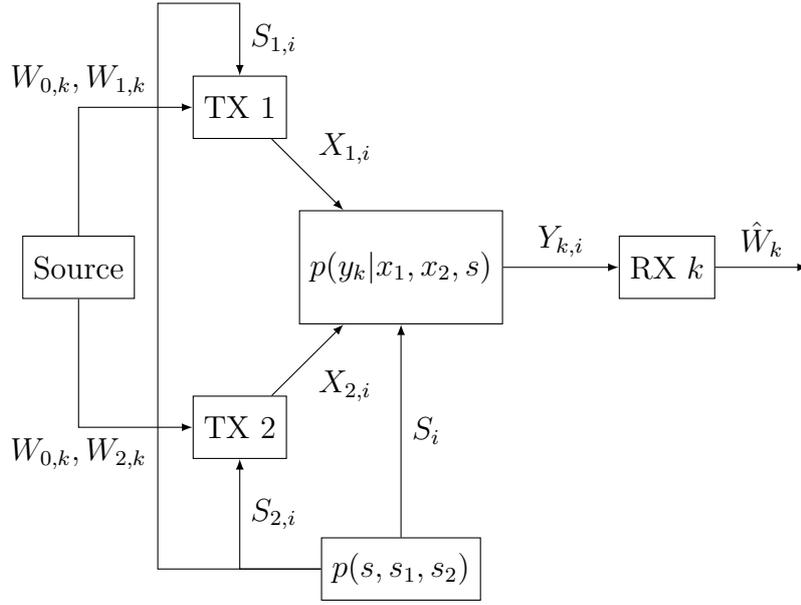


Figure 2.2: State-dependent MAC with common message and causal distributed CSIT.

2.3.1 Multiple-access channels with common message

We modify the model in Section 2.1 by considering transmission towards a single arbitrary RX k , $L = 2$ TXs, and by splitting the message W_k into

$$W_k = (W_{0,k}, W_{1,k}, W_{2,k}) \in \mathcal{W}_{0,k} \times \mathcal{W}_{1,k} \times \mathcal{W}_{2,k}, \quad (2.12)$$

where $W_{0,k}$ is the portion of W_k available at both TXs, while $W_{1,k}$ is available only at TX l . We assume that the three messages $W_{0,k}, W_{1,k}, W_{2,k}$ are independently and uniformly distributed over the sets $\mathcal{W}_{j,k} := \{1, \dots, 2^{\lceil nR_{j,k} \rceil}\}$, $j = 0, 1, 2$, where $R_{j,k} \geq 0$ is the rate of the message $W_{j,k}$. The family of encoding functions are then modified to

$$\phi_{l,i} : \mathcal{W}_{0,k} \times \mathcal{W}_{l,k} \times \mathcal{S}_l^i \rightarrow \mathcal{X}_l, \quad l \in \mathcal{L}, \quad i = 1, \dots, n. \quad (2.13)$$

The above model identifies a so-called state-dependent *multiple-access channel* (MAC) with common message and distributed CSIT, and it is depicted in Figure 2.2. For a given pair (P_1, P_2) , the closure of the set of all achievable rates $(R_{0,k}, R_{1,k}, R_{2,k})$ is the capacity-cost region $\mathcal{C}_k^{\text{MAC}}(P_1, P_2)$ of the considered MAC.

Remark 2.2. *The sum capacity*

$$C_{\text{sum},k}(P_1, P_2) := \max\{R_{\text{sum},k} : (R_{0,k}, R_{1,k}, R_{2,k}) \in \mathcal{C}_k^{\text{MAC}}(P_1, P_2)\}, \quad (2.14)$$

where $R_{\text{sum},k} := R_{0,k} + R_{1,k} + R_{2,k}$ is the rate of the aggregate message W_k , is equal to the capacity $C_k(P_1, P_2)$ of RX k as defined in Section 2.1.

Remark 2.3. $\mathcal{C}_k^{\text{MAC}}$ has a very different operational meaning than the region \mathcal{C} given in Section 2.1. Specifically, the former assumes transmission towards a single RX and describes different message cooperation regimes at the TX side; in contrast, the latter assumes full message sharing at the TXs and describes different multi-user multiplexing regimes at the RX side.

Remark 2.4. The considered small scale setup ($L = 2$ TXs) is enough to capture the main issues behind partial message and CSIT sharing when transmitting towards a single RX. For the purposes of this thesis, considering $L > 2$ TXs with arbitrary message sets (see, e.g., [40]) is cumbersome and does not provide additional insight.

2.3.2 User-centric cooperation clusters

We modify the model in Section 2.1 by limiting the message sharing according to the following *user-centric cooperation clustering* approach [25, 26, 28]. The name of this approach comes from allocating network resources such that each RX k is served only by a subset of the TXs, hence grouping the TXs from a user-centric perspective. Specifically, we here assume the message W_k to be available only at a subset $\mathcal{K}_k \subseteq \mathcal{K}$ of the TXs, and modify the family of encoding functions to

$$\phi_{l,i} : \prod_{k \in \mathcal{K}_l} \mathcal{W}_k \times \mathcal{S}_l^i \rightarrow \mathcal{X}_l, \quad l \in \mathcal{L}, \quad i = 1, \dots, n, \quad (2.15)$$

where $\mathcal{K}_l \subseteq \mathcal{K}$ is the set of RXs served by TX l , i.e.,

$$\mathcal{K}_l := \{k \in \mathcal{K} \text{ s.t. } l \in \mathcal{L}_k\}, \quad l \in \mathcal{L}. \quad (2.16)$$

Remark 2.5. In contrast to the MAC model in Section 2.3.1, the model of this section does not consider for simplicity an arbitrary TX to have partial knowledge of a message W_k ; here, W_k is either fully available or not available.

Chapter 3

Coding Theorems for Decentralized Transmission

This chapter revisits known coding theorems for centralized transmission extended to decentralized transmission with distributed CSIT. In some cases, simple yet novel results on optimal transmission, in the form of capacity theorems, are provided. Most importantly, the coding theorems presented here form the theoretical basis for the main results of this thesis.

3.1 Decentralized transmission towards a single RX

3.1.1 Capacity and Shannon strategies

In this section, we study the maximum achievable rate towards an arbitrary RX $k \in \mathcal{K}$, i.e., we focus on the capacity C_k of the channel towards RX k .

Remark 3.1. *Since the choice of k plays essentially no role for the purposes of this section, in the following we omit its subscript and consider the transmission of a message W of rate R over a channel with output Y and capacity C .*

The following theorem extends the fundamental result on optimal transmission over state-dependent point-to-point channels (i.e., $L = 1$ TX) with causal state information given by [31, 41] to distributed transmission:

Theorem 3.1. *The capacity of the channel at hand is given by*

$$C(P_1, \dots, P_L) = \max_{\substack{p(u) \\ x_t = f_t(u, s_t) \\ \mathbb{E}[\eta_t(X_t)] \leq P_t}} I(U; Y), \quad (3.1)$$

where the maximum is taken over an auxiliary random variable $U \in \mathcal{U}$, independent of (S, S_1, \dots, S_L) , and L functions $f_l : \mathcal{U} \times \mathcal{S}_l \rightarrow \mathcal{X}_l$ for $l \in \mathcal{L}$.

Proof. Assume for simplicity that all alphabets have finite cardinality¹. The achievability part can be sketched by a simple modification of Shannon's *physical device* argument, used in [31, 7] for $L = 1$ TX: an achievable scheme for the channel at hand can be obtained by attaching a deterministic device $f_l(u, s_l)$ with inputs (U, S_l) and output X_l in front of each of the actual channel inputs, inducing a stateless channel with input U , output Y , and memoryless law

$$p(y|u) = \sum_{s, s_1, \dots, s_L} p(y|f_1(u, s_1), \dots, f_L(u, s_L), s)p(s, s_1, \dots, s_L), \quad (3.2)$$

over which $R = I(U, Y)$ is achievable [7].

For the converse, we define $U_i = (W, S_1^{i-1}, \dots, S_L^{i-1})$ and assume that past CSIT realizations $(S_1^{i-1}, \dots, S_L^{i-1})$ are available at all encoders. Hence, we assume that $X_{l,i}$ is a function of $(U_i, S_{l,i}) \forall l \in \mathcal{L}$. Note that U_i is independent of $(S_i, S_{1,i}, \dots, S_{L,i})$. We then have:

$$nR = H(W) \quad (3.3)$$

$$= I(W; Y^n) + H(W|Y^n) \quad (3.4)$$

$$\leq I(W; Y^n) + n\epsilon_n \quad (3.5)$$

$$= \sum_{i=1}^n I(W; Y_i | Y^{i-1}) + n\epsilon_n \quad (3.6)$$

$$\leq \sum_{i=1}^n I(U_i; Y_i | Y^{i-1}) + n\epsilon_n \quad (3.7)$$

$$\leq \sum_{i=1}^n I(U_i, Y^{i-1}; Y_i) + n\epsilon_n \quad (3.8)$$

$$= \sum_{i=1}^n I(U_i; Y_i) + n\epsilon_n, \quad (3.9)$$

where (3.5) follows from Fano's inequality ($\lim_{n \rightarrow \infty} \epsilon_n = 0$), and (3.9) follows since the channel is memoryless, which implies the Markov chain $Y^{i-1} \rightarrow U_i \rightarrow Y_i$. The code must also satisfy the input cost constraints

$$P_l \geq \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \eta_l(X_{l,i}) \right], \quad \forall l \in \mathcal{L}. \quad (3.10)$$

¹The extension to infinite dimensional alphabets as in the MIMO fading channel model of Section 2.2 can be obtained via standard techniques as in [7].

We combine the bounds in (3.9) and (3.10) by means of a time-sharing variable Q uniformly distributed in $\{1, \dots, n\}$ and independent of everything else, and by letting $U := (U_Q, Q)$, $X_l := X_{l,Q}$, $Y := Y_Q$, $S := S_Q$, $S_l := S_{l,Q}$. Note that the joint pmf of $(Y, X_1, \dots, X_L, S, S_1, \dots, S_L, U)$ factors as required. With these identifications, we finally obtain

$$R \leq I(U_Q; Y|Q) + \epsilon_n \leq I(U; Y) + \epsilon_n, \quad P_l \geq \mathbb{E}[\eta_l(X_l)]. \quad (3.11)$$

The finite cardinality of U follows directly by Shannon's argument [7], which corresponds to coding over an augmented input alphabet of tuples of functions $\mathcal{S}_l \rightarrow \mathcal{X}_l$ of size $\prod_{l=1}^L |\mathcal{X}_l|^{|\mathcal{S}_l|}$, indexed by U . Therefore, we can restrict $|\mathcal{U}| \leq \prod_{l=1}^L |\mathcal{X}_l|^{|\mathcal{S}_l|}$. \square

The main finding of Theorem 3.1 is that the single RX capacity is achieved by using so-called *Shannon strategies*, i.e., by the following simple transmission scheme:

- Encode message W over a data-bearing signal U^n , shared by all TXs.
- At channel use i , each TX l forms its input $X_{l,i} = f_l(U_i, S_{l,i})$ as a function of U_i and the current CSIT $S_{l,i}$ only. Throughout this work, we refer to (f_1, \dots, f_L) as a tuple of *distributed precoding* functions.

The interesting feature of the above scheme is that optimal transmission can be performed by neglecting the past CSIT sequence. Even more, the converse proof also shows that providing the strictly causal sequence $(S_1^{i-1}, \dots, S_L^{i-1})$ to all TXs does not increase C , meaning that there is no capacity gain in letting the TXs to retrospectively share their current CSIT.

It is also worth emphasizing the difference with respect to the (virtually) centralized CSIT case where all TXs share $S_l =: S_T \forall l \in \mathcal{L}$. In such a case, by omitting for simplicity the input cost constraints, we recover the result of [31, 41]

$$C = \max_{\substack{p(u) \\ (x_1, \dots, x_L) = f(u, s_T)}} I(U; Y). \quad (3.12)$$

Although Shannon strategies are optimal in both the distributed and centralized CSIT cases, the distributed CSIT assumption generally imposes the design of L distributed precoding functions (f_1, \dots, f_L) depending on the local CSIT only each, rather than a single precoding function $f : \mathcal{U} \times \mathcal{S}_T \rightarrow \prod_{l=1}^L \mathcal{X}_l$ as in the centralized case. This poses considerable optimization issues, as will be better clarified throughout this thesis.

Remark 3.2 (CSIR). *The above result readily generalizes to arbitrary CSIR S_R at RX k by replacing S with an augmented state $\tilde{S} := (S, S_R)$, and Y with an*

augmented output $\tilde{Y} := (Y, S_R)$. With this substitutions, the mutual information in (3.1) can be rewritten as

$$I(U; Y, S_R) = I(U; Y|S_R) + I(U; S_R) = I(U; Y|S_R). \quad (3.13)$$

Remark 3.3 (Feedback). As a particular case of (3.13), we let $S_l = q_l(S_R)$ for L deterministic functions $q_l : \mathcal{S}_R \rightarrow \mathcal{S}_l$. This can be interpreted, e.g., as the CSIT acquired through feedback links from the RX, as in frequency-division duplex (FDD) systems. In this case, (3.13) can be rewritten as

$$I(U; Y|S_R) = I(X_1, \dots, X_L, U; Y|S_R) = I(X_1, \dots, X_L; Y|S_R), \quad (3.14)$$

where the first equality follows from X_l being a function of (U, S_R) , and the second equality from the Markov chain $U \rightarrow (X_1, \dots, X_L, S_R) \rightarrow Y$. An advantage of the above rewriting is that optimal codes can be constructed directly on the input alphabets, since (3.1) becomes equivalent to

$$C(P_1, \dots, P_L) = \max_{\substack{p(x_l|s_l) \\ \mathbb{E}[\eta_l(X_l)] \leq P_l}} I(X_1, \dots, X_L; Y|S_R). \quad (3.15)$$

3.1.2 Example: channel with additive binary inputs and state

As a simple example, we consider the following channel

$$Y = X_1 + X_2 + S, \quad (3.16)$$

with binary inputs and state, i.e. $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{S} = \{0, 1\}$, and where $\mathcal{Y} = \{0, 1, 2, 3\}$. We do not consider power constraints. We further assume $S \sim \text{Bernoulli}(q)$, and distributed CSIT $p(s_1, s_2|s) = p(s_1|s)p(s_2|s)$, where $p(s_l|s)$ is a binary symmetric channel with transition probability $\epsilon_l \in [0, 0.5]$.

A (non-scalable) method for computing the capacity C in (3.1) is to adapt to the considered distributed setting the interpretation of Shannon strategies as encoding over an alphabet of functions [31, 7], combined with classical results on the computation of the capacity of point-to-point channels [4]. More precisely, we proceed as follows:

1. We build the alphabet of Shannon strategies by enumerating all the functions

$$t_u = (t_{1,u}, t_{2,u}), \quad t_{l,u} : \mathcal{S}_l \rightarrow \mathcal{X}_l, \quad (3.17)$$

where each function is indexed by U . There are $|\mathcal{U}| = |\mathcal{X}_1|^{|\mathcal{S}_1|} |\mathcal{X}_2|^{|\mathcal{S}_2|} = 16$ such functions.

2. We set $x_l = f_l(u, s_l) = t_{l,u}(s_l)$ and compute the equivalent stateless point-to-point channel law $p(y|u)$.
3. We run the Blahut-Arimoto algorithm for computing the capacity of the equivalent channel $p(y|u)$ [4].

Note that the above procedure is similar to the one outlined in [42] for centralized settings.

In Figure 3.1 we plot the capacity C versus the CSIT quality at TX 2, for various choices of CSIT quality at TX 1, and for $q = 0.5$. Note that $\epsilon_l = 0$ and $\epsilon_l = 0.5$ model respectively perfect and no CSIT at the l -th TX. Interestingly, the capacity of the system decreases with ϵ_2 down to a flat regime in which any further decrease in quality does not matter, and the turning point depends on ϵ_1 . This can be interpreted as a regime in which the quality at one TX is so degraded that, although some CSIT is available, it does not allow for proper coordination with the better informed TX. Intuitively, it is important for the less informed TX to not act as unknown noise for the other TX. In fact, in the aforementioned regime it turns out that the optimal scheme at the less informed TX is to throw away completely its CSIT, making its behaviour not adaptive to the channel conditions but completely predictable by the more informed TX.

3.1.3 Stationary and ergodic state processes

In this section we briefly mention how to adapt the coding technique achieving the capacity of memoryless channels given by Theorem 3.1 to channels governed by stationary and ergodic state processes. Specifically, we modify the channel model by assuming that the process $(S, S_1, \dots, S_L)^n$ is jointly stationary and ergodic, i.e., the channel is not memoryless. Using Shannon strategies $X = f_l(U, S_l)$, we obtain a new channel with input U and output Y . Achievable rates can be given by [43]

$$R = \lim_{n \rightarrow \infty} \frac{1}{n} I(U^n, Y^n), \quad (3.18)$$

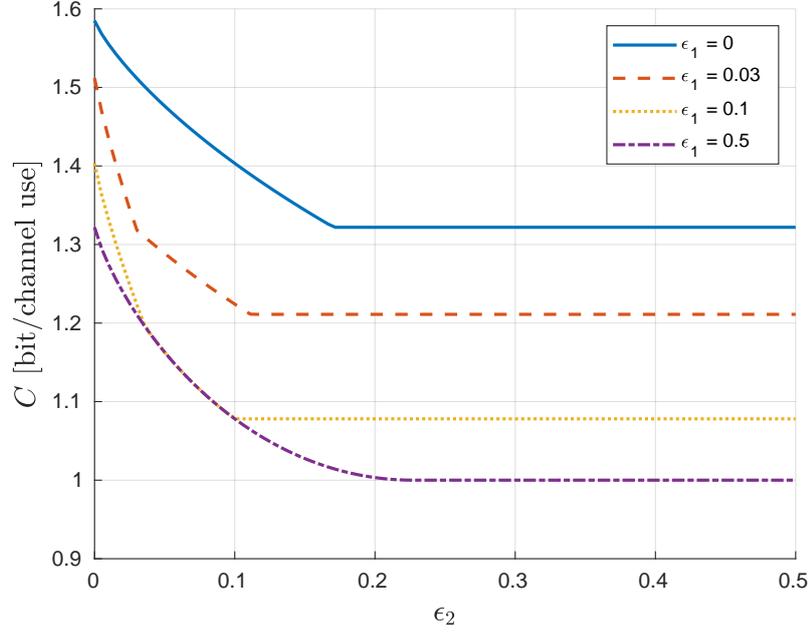


Figure 3.1: Capacity vs. CSIT distortion ϵ_2 at TX 2, for various choices of CSIT distortion ϵ_1 at TX 1.

if U^n is chosen such that $(U, Y)^n$ is stationary and ergodic. Picking an i.i.d. input distribution $U^n \sim \prod_{i=1}^n p(u_i)$ for some $p(u)$, we obtain

$$I(U^n, Y^n) = H(U^n) - H(U^n|Y^n) \quad (3.19)$$

$$\geq H(U^n) - \sum_{i=1}^n H(U_i|Y^n) \quad (3.20)$$

$$\geq H(U^n) - \sum_{i=1}^n H(U_i|Y_i) \quad (3.21)$$

$$= \sum_{i=1}^n H(U_i) - H(U_i|Y_i) \quad (3.22)$$

$$= nI(U, Y), \quad (3.23)$$

where the last equality comes from $(U, Y)^n$ being jointly stationary. Therefore, the rate $R = C$ is achievable, where C is the capacity of the corresponding memoryless channel with the same first order distribution.

Remark 3.4. *In contrast to memoryless channels, past CSIT can be very useful. A simple and popular way of exploiting the past CSIT sequence is to use*

the past sequence S_l^{i-1} and fresh information $S_{l,i}$ to form an enhanced estimate $\tilde{S}_{l,i}$ of the i -th realization of the state S . Then, if the new state and CSIT sequence $(S, \tilde{S}_1, \dots, \tilde{S}_L)^n$ is also jointly stationary and ergodic, we can achieve possibly higher rates $I(U, Y)$ with $X = f(U, S_l)$ replaced by $X = f(U, \tilde{S}_l)$.

3.1.4 Multiple-access channels

In this section we release the full message sharing assumption and discuss the application of Shannon strategies to the MAC model in Section 2.3.1, i.e., we focus on the capacity region $\mathcal{C}_k^{\text{MAC}}(P_1, P_2)$ of the MAC towards RX k .

Remark 3.5. *As in Section 3.1, the choice of k plays essentially no role for the purposes of this section, hence its subscript is omitted. However, special care must be taken not to confuse the components of $(R_0, R_1, R_2) \in \mathcal{C}^{\text{MAC}}$ with the rates $(R_1, \dots, R_K) \in \mathcal{C}$ studied in the multi-RX setups.*

Similarly to the achievability part of Theorem 3.1, and focusing again on finite dimensional alphabets for simplicity, we obtain the following achievable region by combining Slepian-Wolf coding [44] for the stateless MAC with common message and Shannon strategies:

Lemma 3.1 (Slepian-Wolf coding). *For the channel in Figure 2.2, $\mathcal{C}^{\text{MAC}}(P_1, P_2)$ includes the convex hull of all rate triples (R_0, R_1, R_2) such that*

$$R_1 \leq I(U_1; Y | U_2, U_0), \quad (3.24)$$

$$R_2 \leq I(U_2; Y | U_1, U_0), \quad (3.25)$$

$$R_1 + R_2 \leq I(U_1, U_2; Y | U_0), \quad (3.26)$$

$$R_0 + R_1 + R_2 \leq I(U_1, U_2; Y), \quad (3.27)$$

for some auxiliary variables $(U_0, U_1, U_2) \in \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2$ independent of (S, S_1, S_2) , with pmf factorizing as $p(u_0)p(u_1|u_0)p(u_2|u_0)$, and L functions $f_l : \mathcal{U}_l \times \mathcal{S}_l \rightarrow \mathcal{X}_l$, $x_l = f_l(u_l, s_l)$, satisfying $\mathbb{E}[\eta_l(X_l)] \leq P_l, \forall l \in \mathcal{L}$.

Proof. Let $x_{l,i} = f_l(u_{l,i}, s_{l,i})$, and apply the scheme in [44] achieving the capacity region of a stateless MAC with inputs U_1, U_2 and output Y . \square

In the above inner bound, U_0 and (U_1, U_2) represent the data bearing signals over which the shared message W_0 and the individual messages (W_1, W_2) are respectively encoded. It is well known that Shannon strategies, i.e., the scheme of Lemma 3.1, fail to achieve all the rate tuples in $\mathcal{C}^{\text{MAC}}(P_1, P_2)$ for general $p(s, s_1, s_2)$, as observed for a special case of a common CSIT $S_1 = S_2 = S$ in [45]. This is because block-Markov encoding enables the two TXs to compress past state information and send it as a common message to provide possibly useful CSIR to

the RX. Nevertheless, Theorem 3.1 shows that such a scheme based on block-Markov encoding is not necessary for achieving the sum capacity of the considered setup. Namely, provided that R_0 is large enough (in the worst case, equal to $C_{\text{sum}}(P_1, P_2) = C(P_1, P_2)$), the sum capacity is indeed achievable by the scheme in Lemma 3.1.

In the following we focus on the particular case where CSIT is a deterministic function of some CSIR S_R available at the RX, as already discussed in Remark 3.3. As a straightforward extension of [46, Theorem 5] and [47, Theorem 4] restricted to the MAC with no common message², we characterize the capacity region as follows.

Theorem 3.2. *Consider the channel in Figure 2.2 with S replaced by $\tilde{S} := (S, S_R)$, and Y by $\tilde{Y} := (Y, S_R)$. Assume $S_l = q_l(S_R)$ for two deterministic functions $q_l : \mathcal{S}_R \rightarrow \mathcal{S}_l$. Then, the capacity region $\mathcal{C}^{\text{MAC}}(P_1, P_2)$ is given by the convex hull of all rate triples (R_0, R_1, R_2) satisfying*

$$R_1 \leq I(X_1; Y | X_2, U, S_R), \quad (3.28)$$

$$R_2 \leq I(X_2; Y | X_1, U, S_R), \quad (3.29)$$

$$R_1 + R_2 \leq I(X_1, X_2; Y | U, S_R), \quad (3.30)$$

$$R_0 + R_1 + R_2 \leq I(X_1, X_2; Y | S_R), \quad (3.31)$$

for some pmf $p(x_1|s_1, u)p(x_2|s_2, u)p(u)$, where $U \in \mathcal{U}$ is an auxiliary variable independent of (S, S_1, S_2, S_R) , satisfying $\mathbb{E}[\eta_l(X_l)] \leq P_l$ for $l \in \mathcal{L}$.

Proof. The proof is given in Appendix B.1.1. □

The main difference between Theorem 3.2 and related results lies in its converse proof, which solves the issue highlighted in [46] through an appropriate identification of the auxiliary variable U , and simplifies the less traditional yet insightful converse proof given by [47]. In contrast to the general case discussed before, Theorem 3.2 refers to a setup where the RX is fully informed about (S_1, S_2) , hence there is no need for the TXs to convey (S_1, S_2) through block-Markov schemes as in [45, 48]. Furthermore, note that, similarly to Remark 3.3, optimal codes for the individual messages (W_1, W_2) can be constructed directly over the input alphabets.

We conclude this section by providing the following outer bound for a relaxation of the setup in Theorem 3.2 covering, e.g., CSIT acquired via over-the-uplink signalling in time-division duplex (TDD) systems, or via noisy feedback from the RX.

²[47, Theorem 4] generalized the case of independent states (S_1, S_2) in [46, Theorem 5] to arbitrary joint pmfs on (S_1, S_2) .

Lemma 3.2. *Consider the channel in Figure 2.2 with S replaced by $\tilde{S} := (S, S_R)$, and Y by $\tilde{Y} := (Y, S_R)$. Assume that $S_1 \rightarrow S_R \rightarrow S_2$ forms a Markov chain. Then, $\mathcal{C}^{\text{MAC}}(P_1, P_2)$ is included in the convex hull of all rate triples (R_0, R_1, R_2) satisfying*

$$R_1 + R_2 \leq I(U_1, U_2; Y|U_0, S_R), \quad (3.32)$$

$$R_0 + R_1 + R_2 \leq I(U_1, U_2; Y|S_R), \quad (3.33)$$

for some auxiliary variables $(U_0, U_1, U_2) \in \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2$ independent of (S, S_1, S_2, S_R) , with pmf factorizing as $p(u_0)p(u_1|u_0)p(u_2|u_0)$, and L functions $f_l : \mathcal{U}_l \times \mathcal{S}_l \rightarrow \mathcal{X}_l$, $x_l = f_l(u_l, s_l)$, satisfying $\mathbb{E}[\eta_l(X_l)] \leq P_l, \forall l \in \mathcal{L}$.

Proof. The proof is given in Appendix B.1.2. □

Lemma 3.2 shows that, for (S_1, S_2) conditionally independent given the CSIR S_R , Shannon strategies may be sub-optimal only in terms of individual rates (indeed, [48] proves that higher individual rates are achievable for some channels with independent (S_1, S_2) and $S_R = \emptyset$). This extends [46, Theorem 4], which considered independent (S_1, S_2) and no common message. The bound on $R_1 + R_2$ was already reported in [49] and references therein by using the same technique as [47]; similarly to Theorem 3.2, it is here derived using a simplified proof.

3.2 Decentralized transmission towards multiple RXs

3.2.1 Broadcast channels and Shannon strategies

A complete characterization of $\mathcal{C}(P_1, \dots, P_L)$ is not surprisingly hard to obtain. One of the main issues is that the characterization of the capacity of the broadcast channel (BC), which is a particular case of the considered setup, is itself a notorious and long lasting open problem. However, the best known coding schemes for the BC are known to perform well, or even optimally, in many setups. Therefore, inspired by the results on decentralized transmission towards a single RX, in what follows we revisit these schemes transposed to distributed CSIT by means of Shannon strategies.

Specifically, we readily obtain multi-RX decentralized transmission techniques by letting $X_l = f_l(U, S_l)$ for some auxiliary variable $U \in \mathcal{U}$, L distributed precoding functions $f_l : \mathcal{U} \times \mathcal{S}_l \rightarrow \mathcal{X}_l$, and by applying known coding schemes for the BC without state, input U , and K outputs Y_k . The resulting achievable schemes are conceptually identical to their centralized counterparts, and relegate the impact of

the distributed CSIT assumption to the design of the distributed precoding functions f_l . Starting from the popular *treating interference as noise* (TIN) approach, we obtain the following inner bound:

Lemma 3.3 (TIN). *A rate tuple (R_1, \dots, R_K) is achievable if*

$$R_k \leq I(U_k; Y_k), \quad \forall k \in \mathcal{K}, \quad (3.34)$$

for K auxiliary variables $(U_1, \dots, U_K) \in \prod_{k=1}^K \mathcal{U}_k =: \mathcal{U}$, with pmf $\prod_{k=1}^K p(u_k)$, independent of (S, S_1, \dots, S_L) , and L functions $f_l: \mathcal{U} \times \mathcal{S}_l \rightarrow \mathcal{X}_l$, $x_l = f_l(u_1, \dots, u_K, s_l)$, satisfying $\mathbb{E}[\eta_l(X_l)] \leq P_l$ for $l \in \mathcal{L}$.

Proof. The proof follows by the physical device argument used in Theorem 3.1 and from the first part of [7, Section 8.3]. \square

In (3.34), each of the auxiliary variables U_k represents the data-bearing signal over which W_k is independently encoded. Although generally suboptimal, TIN offers an excellent performance / complexity compromise in many practical cases; its main feature is that it keeps both the TX and RX architectures simple, and mostly relies on a good design of the distributed precoding functions f_l for mitigating the interference. Therefore, intuitively, TIN performs well whenever these functions are able to drive the interference down to the noise floor; for example, it is close-to-optimal for centralized MIMO fading channels with sufficiently high quality CSIT [50, 2].

A second more complex approach is to let the RXs decode (some of) the interference. Focusing on $K = 2$ RXs for simplicity, this idea leads to the following inner bound:

Lemma 3.4 (Superposition coding). *A rate pair (R_1, R_2) is achievable if*

$$R_1 \leq I(U_1; Y_1|U_0) + I(U_0; Y_1), \quad (3.35)$$

$$R_2 \leq I(U_2; Y_2|U_0) + I(U_0; Y_2), \quad (3.36)$$

$$R_1 + R_2 \leq I(U_1; Y_1|U_0) + I(U_2; Y_2|U_0) + \min\{I(U_0; Y_1), I(U_0; Y_2)\}, \quad (3.37)$$

for some auxiliary variables $(U_0, U_1, U_2) \in \mathcal{U}_0 \times \mathcal{U}_1 \times \mathcal{U}_2 =: \mathcal{U}$ independent of (S, S_1, \dots, S_L) , with pmf factorizing as $p(u_0)p(u_1|u_0)p(u_2|u_0)$, and L functions $f_l: \mathcal{U} \times \mathcal{S}_l \rightarrow \mathcal{X}_l$, $x_l = f_l(u_0, u_1, u_2, s_l)$, satisfying $\mathbb{E}[\eta_l(X_l)] \leq P_l, \forall l \in \mathcal{L}$.

Proof. The proof follows, e.g., from a particular case of [7, Proposition 8.1]. \square

The *superposition coding* inner bound improves upon TIN by allowing part of (W_1, W_2) to be encoded over a data-bearing signal U_0 to be decoded by both RXs. This is known to be capacity achieving for the degraded BC without state [7], and for the degraded BC with causal CSIT [51]. The following theorem proves that this holds also for decentralized transmission.

Theorem 3.3. Consider physically degraded channels, i.e., such that

$$p(y_1, y_2|x_1, x_2, s) = p(y_1|x_1, x_2, s)p(y_2|y_1), \quad (3.38)$$

or, more in general, stochastically degraded channels, i.e., such that there is a physically degraded channel $p'(y_1, y_2|x_1, x_2, s) = p(y_1|x_1, x_2, s)p'(y_2|y_1)$ with the same marginal $p(y_2|x_1, x_2, s)$. Then, $\mathcal{C}(P_1, \dots, P_L)$ is given by the convex hull of the set of all rate pairs (R_1, R_2) satisfying

$$R_1 \leq I(U_1; Y_1|U_2), \quad (3.39)$$

$$R_2 \leq I(U_2; Y_2), \quad (3.40)$$

for some auxiliary variables $(U_1, U_2) \in \mathcal{U}_1 \times \mathcal{U}_2 =: \mathcal{U}$ independent from (S, S_1, S_2) , and L functions $f_l : \mathcal{U} \times \mathcal{S}_l \rightarrow \mathcal{X}_l$, $x_l = f_l(u_1, u_2, s_l)$, satisfying $\mathbf{E}[\eta_l(X_l)] \leq P_l$, $\forall l \in \mathcal{L}$.

Proof. The proof is given in Appendix B.1.3. □

The above theorem extends the result in [51] to the considered network with distributed CSIT. The setup in [51] is recovered for the centralized and perfect CSIT case $S_1 = S_2 = S$. The optimal transmission scheme exploits the degradedness property which ensures that RX 1 can always decode the full message W_2 encoded over $U_0 = U_2$. Unfortunately, MIMO fading channels are typically not degraded. However, we remark that superposition coding can still be very useful for MIMO fading channels (where it is often referred to as *rate splitting* or *non-orthogonal multiple access*), especially when the available CSIT does not allow to bring the interference down to the noise floor, making TIN highly suboptimal [52]. Remarkably, the scheme in Lemma 3.4 is shown in [20] to achieve the optimal throughput scaling law (the so-called *sum DoF*) for some decentralized MIMO fading channels with non-trivial distributed CSIT configurations.

A third approach is to interpret the interference to be generated as a known non-casual state, whose impact can be preventively mitigated at the encoder side. By focusing again on $K = 2$ RXs for simplicity, this idea leads to the following inner bound:

Lemma 3.5 (Marton's inner bound). A rate pair (R_1, R_2) is achievable if

$$R_1 \leq I(U_1; Y_1), \quad (3.41)$$

$$R_2 \leq I(U_2; Y_2), \quad (3.42)$$

$$R_1 + R_2 \leq I(U_1; Y_1) + I(U_2; Y_2) - I(U_1, U_2), \quad (3.43)$$

for some auxiliary variables $(U_1, U_2) \in \mathcal{U}_1 \times \mathcal{U}_2 =: \mathcal{U}$ independent from (S, S_1, \dots, S_L) , and L functions $f_l : \mathcal{U} \times \mathcal{S}_l \rightarrow \mathcal{X}_l$, $x_l = f_l(u_1, u_2, s_l)$, satisfying $\mathbf{E}[\eta_l(X_l)] \leq P_l$, $\forall l \in \mathcal{L}$.

Proof. The proof follows from [7, Theorem 8.3]. \square

The main benefit of *Marton's* inner bound w.r.t. the TIN bound is that the former allows for arbitrary dependence between the data-bearing signals U_1 and U_2 . This, however, comes at a penalty on the achievable sum rate. Remarkably, a particular case of Lemma 3.5, known as *dirty paper* coding [53], is known to achieve the capacity region of MIMO fading channels with perfect CSIT and CSIR [54]. Its main drawbacks are the considerable increase of complexity at the TX end, and the stringent CSIT quality requirements for the interference *precancelation* idea to be effective.

Remark 3.6 (Marton's inner bound with superposition coding). *Marton's coding can be readily combined with superposition coding, leading to the largest known single-letter achievable region for the BC [7, Proposition 8.1]. Therefore, the techniques presented in this section give the largest known single-letter achievable region $\mathcal{R}(P_1, \dots, P_L) \subseteq \mathcal{C}(P_1, \dots, P_L)$ for the considered decentralized network, if we restrict the encoders to neglect the past CSIT sequences (i.e., to use Shannon strategies). As already mentioned, under this restriction the difference between decentralized and centralized transmission is confined to the design of good distributed precoding functions f_l .*

3.2.2 Distributed linear precoding

As typically done in practical centralized systems, a simple yet effective design choice is to restrict the distributed precoding functions f_l to the class of linear functions of the data-bearing signals. Under this restriction, we obtain the following extension of the classical *linear precoding* idea to decentralized MIMO fading channels with distributed CSIT.

Specifically, we consider the channel model in Section 2.2, and focus for simplicity on the TIN bound (3.34) with input distribution restricted as follows. We take the K auxiliary variables as $U_k = \mathbf{u}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{d_k})$ for some $d_k \in \mathbb{N}$, and let each TX l forms its transmit signal according to the following *distributed* linear precoding scheme [19]:

$$\mathbf{x}_l = \sum_{k=1}^K \mathbb{F}_{l,k}^H(S_l) \mathbf{u}_k, \quad (3.44)$$

where $\mathbb{F}_{l,k}^H(S_l) \in \mathbb{C}^{N \times d_k}$ is a linear precoder applied at TX l to the shared data-bearing signal \mathbf{u}_k based only on the local information S_l . With this choice, the L power constraints become $\sum_{k=1}^K \mathbb{E}[\|\mathbb{F}_{l,k}(S_l)\|_F^2] \leq P_l$.

The parameter d_k can be interpreted as the number of independent data streams over which the message W_k is encoded. Traditional design choices for

centralized CSIT, as well as the distributed precoding schemes available in the literature, consider (see, e.g., [32, 19, 15])

$$d_k \leq LN, \quad (3.45)$$

i.e., bound the number of data streams by the total number of TX antennas. As we will review later on, this choice is capacity achieving for centralized channels with CSIT as a function of perfect CSIR³. However, in Chapter 4, we remove the constraint (3.45) and show that $d_k > LN$ may be beneficial under distributed CSIT.

Evaluating the rate bounds $I(\mathbf{u}_k; \mathbf{y}_k | \mathbb{H}, S_1, \dots, S_L)$ in (3.34) under perfect CSIR $(\mathbb{H}, S_1, \dots, S_L)$ readily gives the achievable rates ($\forall k \in \mathcal{K}$)

$$R_k \leq R_k^{\text{CSIR}} := \mathbb{E} \left[\log \frac{\det \left(\mathbf{I}_M + \mathbf{G}_k^H \sum_{j=1}^K \boldsymbol{\Sigma}_j(S_1, \dots, S_L) \mathbf{G}_k \right)}{\det \left(\mathbf{I}_M + \mathbf{G}_k^H \sum_{j \neq k} \boldsymbol{\Sigma}_j(S_1, \dots, S_L) \mathbf{G}_k \right)} \right], \quad (3.46)$$

where we define the conditional input covariances

$$\boldsymbol{\Sigma}_k(S_1, \dots, S_L) := \begin{bmatrix} \mathbb{F}_{1,k}^H(S_1) \\ \vdots \\ \mathbb{F}_{L,k}^H(S_L) \end{bmatrix} \begin{bmatrix} \mathbb{F}_{1,k}(S_1) & \dots & \mathbb{F}_{L,k}(S_L) \end{bmatrix}. \quad (3.47)$$

In fact, for the rate in (3.46) to be achievable, it is sufficient that RX k knows its local channel after precoding

$$\tilde{\mathbf{G}}_k := \begin{bmatrix} \mathbb{F}_{1,1}(S_1) & \dots & \mathbb{F}_{L,1}(S_L) \\ \vdots & \ddots & \vdots \\ \mathbb{F}_{1,K}(S_1) & \dots & \mathbb{F}_{L,K}(S_L) \end{bmatrix} \mathbf{G}_k \in \mathbb{C}^{\sum_{k=1}^K d_k \times M}. \quad (3.48)$$

Unfortunately, the evaluation of (3.34) in absence of CSIR $I(\mathbf{u}_k; \mathbf{y}_k)$ is not as straightforward; however, for the case of $M = 1$ RX antennas and $d_k = 1$ data stream per RX, we can consider a popular lower bound on $I(U_k; \mathbf{y}_k)$, known as the *hardening* bound [6, 2], which proves the achievability of all rate tuples (R_1, \dots, R_K) such that

$$R_k \leq R_k^{\text{hard}} := \log \left(1 + \frac{p_k \mathbb{E}[|\mathbf{g}_k^H \mathbf{t}_k|^2]}{p_k \text{Var}[\mathbf{g}_k^H \mathbf{t}_k] + \sum_{j \neq k} p_j \mathbb{E}[|\mathbf{g}_k^H \mathbf{t}_j|^2] + 1} \right), \quad (3.49)$$

³In fact, motivated by DoF arguments, $d_k \leq \min(M, LN)$ is typically assumed; however, forcing $d_k \leq M$ may not be capacity achieving under imperfect CSIT.

where, for reasons that will be clarified later in the thesis, we define the K power scaling factors $(p_1, \dots, p_K) \in \mathbb{R}_+$, and the K precoding vectors for message W_k

$$\mathfrak{t}_k := [\mathfrak{t}_{1,k}^H \ \dots \ \mathfrak{t}_{L,k}^H]^H, \quad p_k \mathfrak{t}_{l,k} := \mathbb{F}_{l,k}^H(S_l) \in \mathbb{C}^N. \quad (3.50)$$

The bound in (3.49) satisfies

$$R_k^{\text{hard}} \leq I(U_k; \mathfrak{y}_k) \leq I(U_k; \mathfrak{y}_k | \mathbb{H}, S_1, \dots, S_L) = R_k^{\text{CSIR}}. \quad (3.51)$$

In Chapter 5, we consider R_k^{hard} instead of R_k^{CSIR} because of the less stringent CSIR requirements and, perhaps most importantly, for treatability reasons. Due to its name and historical use, it is sometimes believed that the tightness of R_k^{hard} w.r.t. R_k^{CSIR} relies on the *channel hardening* effect arising in massive MIMO systems [2]. Although this is correct for some precoding design such as MRT, we remark that (3.49) can be tight also in absence of channel hardening. For instance, R_k^{hard} and R_k^{CSIR} coincide under a ZF scheme putting $\mathfrak{g}_k^H \mathfrak{t}_j = 0$ for $j \neq k$ and $\mathfrak{g}_k^H \mathfrak{t}_k = \mu_k$ a.s., for some deterministic constants $\{\mu_k\}_{k=1}^K$ such that the long term power constraints are satisfied.

Remark 3.7 (Drawbacks of the hardening bound). *On a high level, the inequalities in (3.51) can be quite tight for every scheme exploiting CSIT to make the resulting channel coefficients after precoding as deterministic as possible. However, forcing this condition may produce undesirable peak power fluctuations, and generally reduces the possibility of allocating power over time to, e.g., circumvent deep fade events. Nevertheless, allowing power allocation over many TX antennas (as in massive MIMO systems) helps reducing the above detrimental effects.*

Chapter 4

Ergodic Capacity under Asymmetric Feedback

In this chapter, we particularize the capacity theorem of Section 3.1 to decentralized wireless networks operating in FDD mode. Specifically, we establish the capacity C_k of the channel in Figure 4.1, where quantized measurements of the fading matrix \mathbb{G}_k are fed back from the RX through asymmetric feedback links. The goal of this chapter is to exploit this relatively simple setup to unveil novel insights on optimal decentralized transmission.

4.1 Optimality of distributed linear precoding with an unconventional number of data streams

We consider an instance of the MIMO fading channel model in Section 2.2, where we assume perfect CSIR \mathbb{G}_k , and let the CSIT be a quantized version of it, i.e.,

$$S_l = q_l(\mathbb{G}_k), \quad l \in \mathcal{L}, \quad (4.1)$$

$$q_l : \mathbb{C}^{LN \times M} \rightarrow \mathcal{S}_l := \{1, \dots, |\mathcal{S}_l|\}, \quad |\mathcal{S}_l| < \infty. \quad (4.2)$$

In what follows, we establish the capacity C_k of the considered channel, and show that the *distributed linear precoding* scheme presented in Section 3.2.2 is optimal. As we will see, the key novelty lies in splitting the encoding of the message W_k over an unconventional number of independent data streams. For ease of exposition, we present the main findings by focusing on $L = 2$ TXs with $N = 1$ antenna each; the extension to general (L, N) is straightforward and will be only briefly discussed. Furthermore, we follow the lighter notation of Section 3.1 and omit the RX subscript k everywhere.

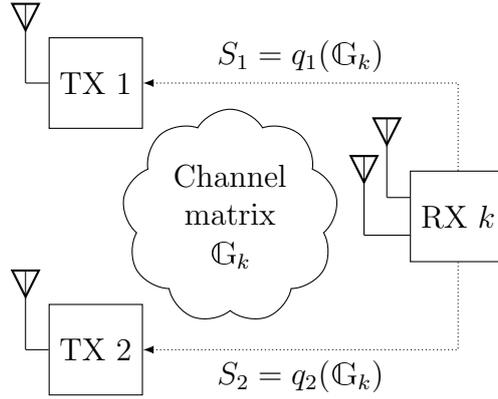


Figure 4.1: Illustration of a cooperative MIMO channel with fading state matrix \mathbb{G}_k known at the RX and distributed CSIT (S_1, S_2) obtained from asymmetric feedback links.

Theorem 4.1. *The capacity of the considered decentralized MIMO fading channel with asymmetric feedback is given by*

$$C(P_1, P_2) = \max_{\substack{\mathbb{f}_l(S_l) \in \mathbb{C}^d \\ \mathbb{E}[\|\mathbb{f}_l(S_l)\|^2] \leq P_l}} \mathbb{E} \left[\log \det (\mathbf{I}_M + \mathbb{G}^H \boldsymbol{\Sigma}(S_1, S_2) \mathbb{G}) \right], \quad (4.3)$$

where

$$\boldsymbol{\Sigma}(S_1, S_2) := \begin{bmatrix} \mathbb{f}_1^H(S_1) \\ \mathbb{f}_2^H(S_2) \end{bmatrix} [\mathbb{f}_1(S_1) \quad \mathbb{f}_2(S_2)], \quad (4.4)$$

and where $d \in \mathbb{N}$ is an optimization variable satisfying

$$d \leq D := |\mathcal{S}_1| + |\mathcal{S}_2|. \quad (4.5)$$

Furthermore $C(P_1, P_2)$ is achieved by letting

$$X_l = \mathbb{f}_l^H(S_l) \mathfrak{u}, \quad \mathfrak{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_d), \quad \forall l \in \mathcal{L}, \quad (4.6)$$

where \mathfrak{u} is a shared data-bearing signal over which W_k is encoded.

Proof. We use the capacity formula

$$C(P_1, P_2) = \max_{\substack{p(\mathfrak{u}) \\ \mathfrak{x}_l = \mathbb{f}_l(\mathfrak{u}, S_l) \\ \mathbb{E}[\eta(X_l)] \leq P_l \quad \forall l \in \mathcal{L}}} I(\mathfrak{x}_1, \dots, \mathfrak{x}_L; \mathfrak{y} | \mathbb{G}) \quad (4.7)$$

given by (3.1) and (3.14), specialized to the consider channel model. We first apply the well-known *maximum differential entropy lemma* [7, p. 21] to the mutual information using the conditional input covariance $\boldsymbol{\Sigma}(S_1, S_2) = \mathbb{E}[\mathfrak{x}\mathfrak{x}^H | S_1, S_2]$, where

$\mathbf{x} := \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ collects all channel inputs, obtaining the upper bound

$$R \leq \mathbb{E} \left[\log \det \left(\mathbf{I}_M + \mathbf{G}^H \boldsymbol{\Sigma}(S_1, S_2) \mathbf{G} \right) \right]. \quad (4.8)$$

For the achievability part, it suffices to show that every feasible $\boldsymbol{\Sigma}(S_1, S_2)$ can be obtained via distributed linear precoders of dimension $D = |\mathcal{S}_1| + |\mathcal{S}_2|$. To this end, we first define the random vectors

$$\tilde{\mathbf{x}}_l := \begin{bmatrix} f_l(U, 1) \\ \vdots \\ f_l(U, |\mathcal{S}_l|) \end{bmatrix} \in \mathbb{C}^{|\mathcal{S}_l|}, \quad l \in \mathcal{L}, \quad (4.9)$$

collecting the random inputs X_l conditioned on each of the $|\mathcal{S}_l|$ realizations of S_l , and the covariance matrix

$$\mathbf{Q} := \mathbb{E} \left[\begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_1^H & \tilde{\mathbf{x}}_2^H \end{bmatrix} \right] \in \mathbb{S}_+^D. \quad (4.10)$$

It is easy to see that \mathbf{Q} contains all the realizations of $\boldsymbol{\Sigma}(S_1, S_2)$. Note that, due to the power constraint $\sum_{s_l \in \mathcal{S}_l} \mathbb{E}[|f_l(U, s_l)|^2] p(s_l) \leq P_l < \infty$, any feasible input distribution must satisfy $\mathbb{E}[|f_l(U, s_l)|^2] < \infty \forall s_l \in \mathcal{S}_l$, hence \mathbf{Q} has finite entries. Since \mathbf{Q} is Hermitian positive semi-definite, there exists a matrix $\mathbf{F} \in \mathbb{C}^{D \times D}$ such that $\mathbf{F}^H \mathbf{F} = \mathbf{Q}$. We denote its column vectors by

$$\mathbf{F} =: [\mathbf{f}_1(1) \ \dots \ \mathbf{f}_1(|\mathcal{S}_1|) \ \mathbf{f}_2(1) \ \dots \ \mathbf{f}_2(|\mathcal{S}_2|)], \quad (4.11)$$

and define the random vectors $\mathbb{f}_l(S_l)$ with realizations $\mathbf{f}_l(s_l)$. Finally, simple calculations show that the scheme in (4.6) with distributed linear precoders designed using the above procedure, i.e., selected from (4.11), preserves the desired $\boldsymbol{\Sigma}(S_1, S_2)$, and attains the maximum entropy bound. \square

The main result of Theorem 4.1 is that the distributed linear precoding scheme presented in Section 3.2.2 achieves the capacity of the considered decentralized MIMO fading channel with asymmetric feedback. However, as a sufficient condition to prove achievability, Theorem 4.1 considers the transmission of possibly $D = |\mathcal{S}_1| + |\mathcal{S}_2|$ independent data streams. This unconventional design choice appears to be in sharp contrast with the (virtually) centralized configuration $S_1 = S_2 =: S_T$, where the capacity of the $2 \times M$ MIMO channel is achieved by using $d \leq 2$ streams. In this latter case, by considering a per-antenna power constraint, the capacity takes the well-known expression (see, e.g., [32])

$$C(P_1, P_2) = \max_{\substack{\boldsymbol{\Sigma}(S_T) \in \mathbb{S}_+^2 \\ \mathbb{E}[\boldsymbol{\Sigma}(S_T)_{l,l}] \leq P_l}} \mathbb{E} \left[\log \det \left(\mathbf{I}_M + \mathbf{G}^H \boldsymbol{\Sigma}(S_T) \mathbf{G} \right) \right], \quad (4.12)$$

where $\Sigma(S_T) := \mathbf{E}[\mathbf{x}\mathbf{x}^H|S_T]$, $\mathbf{x} := \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, is the conditional input covariance. The capacity in (4.12) can be simply achieved by taking the matrix square-root $\mathbb{F}(S_T) := \Sigma^{\frac{1}{2}}(S_T) \in \mathbb{C}^{2 \times 2}$ and by letting

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \mathbb{F}^H(S_T)\mathbf{u}, \quad \mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_2), \quad (4.13)$$

or, in other words, by precoding $d = LN$ data streams only, where $LN = 2$ is the total number of TX antennas. Such an approach cannot be used for general distributed CSIT, as it may lead to unfeasible linear precoders violating the functional dependencies $X_l = f_l(\mathbf{u}, S_l)$. The proof of Theorem 4.1 addresses this issue by *increasing* the size d of the linear precoders up to $D \gg 2$, i.e., beyond conventional design choices. A more detailed analysis of the role played by d in terms of optimal distributed precoding is provided in the following sections.

Remark 4.1 (Extension to arbitrary (L, N)). *Theorem 4.1 can be readily extended to L TXs with N antennas as follows:*

$$C(P_1, \dots, P_L) = \max_{\substack{\mathbb{F}_l(S_L) \in \mathbb{C}^{d \times N} \\ \mathbf{E}[\|\mathbb{F}_l(S_L)\|_{\mathbb{F}}^2] \leq P_l}} \mathbf{E} \left[\log \det (\mathbf{I} + \mathbf{G}^H \Sigma(S_1, \dots, S_L) \mathbf{G}) \right], \quad (4.14)$$

where

$$\Sigma(S_1, \dots, S_L) := \begin{bmatrix} \mathbb{F}_1^H(S_1) \\ \vdots \\ \mathbb{F}_L^H(S_L) \end{bmatrix} \begin{bmatrix} \mathbb{F}_1(S_1) & \dots & \mathbb{F}_L(S_L) \end{bmatrix}, \quad (4.15)$$

and where $d \leq D := N \sum_{l=1}^L |\mathcal{S}_l|$. Furthermore, the capacity is achievable by

$$\mathbf{x}_l = \mathbb{F}_l^H(S_l)\mathbf{u}, \quad \mathbf{u} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_d), \quad \forall l \in \mathcal{L}. \quad (4.16)$$

The proof follows the same lines of Theorem 4.1, hence it is omitted.

4.2 Capacity computation via convex optimization

The distributed precoding design problem (4.3) belongs to the class of *static team decision* problems [34, 33, 35, 19], for which efficient optimal solution methods are not known in general¹. However, in contrast to the traditional precoding design with $d \leq LN = 2$, by letting $d \leq D$ we are able to recast the optimal precoding design problem (4.3) into a convex form.

¹This class of problems will be treated in more details in Chapter 5, which deals with distributed precoding design for a multi-RX setup.

Lemma 4.1. *Problem (4.3) is equivalent to the following convex problem*

$$\begin{aligned} \underset{\mathbf{Q} \in \mathcal{S}_+^D}{\text{maximize}} \quad & \mathbb{E} \left[\log \det \left(\mathbf{I}_M + \mathbb{G}_{\text{eq}}^H \mathbf{Q} \mathbb{G}_{\text{eq}} \right) \right] \quad \text{subject to} \\ & \sum_{i=1}^{|\mathcal{S}_1|} [\mathbf{Q}]_{i,i} \Pr(S_1 = i) \leq P_1, \\ & \sum_{j=|\mathcal{S}_1|+1}^{|\mathcal{S}_1|+|\mathcal{S}_2|} [\mathbf{Q}]_{j,j} \Pr(S_2 = j - |\mathcal{S}_1|) \leq P_2, \end{aligned} \quad (4.17)$$

where we defined $\mathbb{G}_{\text{eq}} := \mathbf{E}(S_1, S_2)\mathbb{G} \in \mathbb{C}^{D \times M}$, and

$$\mathbf{E}(S_1, S_2) := \begin{bmatrix} \mathbf{e}_{S_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{e}_{S_2} \end{bmatrix} \in \mathbb{C}^{D \times 2}, \quad (4.18)$$

where $\mathbf{e}_{S_1} \in \{0, 1\}^{|\mathcal{S}_1|}$ and $\mathbf{e}_{S_2} \in \{0, 1\}^{|\mathcal{S}_2|}$ are standard column selectors.

Proof. The proof follows by simply rewriting (4.3) in light of the technique used for the proof of Theorem 4.1. Specifically, we let $d = D$ and use

$$\mathbb{G}^H \boldsymbol{\Sigma}(S_1, S_2) \mathbb{G} = \mathbb{G}^H \mathbf{E}^H(S_1, S_2) \mathbf{F}^H \mathbf{F} \mathbf{E}(S_1, S_2) \mathbb{G} = \mathbb{G}_{\text{eq}}^H \mathbf{Q} \mathbb{G}_{\text{eq}}, \quad (4.19)$$

where \mathbf{F} is given by (4.11) and $\mathbf{Q} = \mathbf{F}^H \mathbf{F}$. The power constraints correspond to linear constraints on the diagonal elements of \mathbf{Q} . \square

Problem (4.17) corresponds to the capacity of a centralized $D \times M$ MIMO fading channel with state \mathbb{G}_{eq} , perfect CSIR, no CSIT, and fixed transmit covariance \mathbf{Q} . The capacity achieving distributed precoders for the original channel can be then designed from the optimal \mathbf{Q}^* in Problem (4.17) as follows:

$$\left[\mathbb{f}_1^*(S_1) \quad \mathbb{f}_2^*(S_2) \right] = (\mathbf{Q}^*)^{\frac{1}{2}} \mathbf{E}(S_1, S_2) \in \mathbb{C}^{D \times 2}. \quad (4.20)$$

Remark 4.2. *An important remark here is that if the constraint $d \leq D$ of Problem (4.3) is replaced by $d < D$, the technique of Proposition 4.1 does not lead to a convex reformulation. This is because the $D \times D$ matrix \mathbf{F} is replaced by a $d \times D$ matrix \mathbf{F}' , hence introducing a non-convex constraint $\text{rank}(\mathbf{Q}) \leq d < D$ to Problem (4.17). However, note that if the optimal \mathbf{Q}^* for the unconstrained problem has rank $r < D$, then we can reduce with no loss of optimality the dimensionality of $\mathbb{f}_l^*(S_l)$ in (4.20) down to $d = r$.*

Problem (4.17) can be solved numerically via known convex optimization tools. A comprehensive discussion on the efficiency of various competing approaches is out of the scope of this thesis. Here, we point out two critical issues that should be

taken into account in a practical system design. First, advanced stochastic optimization techniques may be required if the fading distribution $p(\mathbf{G})$ is continuous². Second, classical second-order methods as interior-point methods for semi-definite optimization typically scale badly with the dimension D of \mathbf{Q} . Hence, first-order methods may be more suitable whenever the cardinality of the CSIT alphabets $|\mathcal{S}_l|$ is large. As a result of the algorithmic complexity stemming out of the above considerations (which are still very active research topics), we envision that feasible implementations of the proposed distributed precoding design should operate in an offline fashion. Specifically, Problem (4.17) could be solved in a preliminary codebook design phase, while in the data transmission phase TX l selects the precoder from the pre-designed codebook based on the received CSIT index S_l .

4.2.1 Example: asymmetric feedback rates

In this section we simulate a decentralized MIMO fading channel with CSIT feedback links of asymmetric rate. For simplicity, we let each entry of \mathbb{G} to be i.i.d. $\mathcal{CN}(0, 1)$, and we set $P_1 = P_2 := \text{SNR}$. The distributed CSIT configuration $p(\mathbf{G}, s_1, s_2)$ is given by two random quantizers with different rates β_1 and β_2 bits. More precisely, we let $\mathcal{S}_l = \{1, \dots, 2^{\beta_l}\}$ be the index set of a codebook $\{\hat{\mathbf{G}}_{l,i}\}_{i=1}^{2^{\beta_l}}$ of randomly and independently generated codewords distributed as $p(\mathbf{G})$. We then let $q_l(\mathbf{G})$ to be a simple nearest neighbour vector quantizers in the Frobenius norm, i.e., $q_l(\mathbf{G}) \in \arg \min_{i \in \mathcal{S}_l} \|\mathbf{G} - \hat{\mathbf{G}}_{l,i}\|_F$. This scenario corresponds to a zero-delay and error-free feedback link from the RX to the l -th TX with a limited rate of β_l bits per channel use. We set $\beta_1 = 4$ and $\beta_2 = 3$, which implies $d = |\mathcal{S}_1| + |\mathcal{S}_2| = 24$. We recall that the RX is assumed to have perfect CSIR.

We approximately solve Problem (4.17) through an off-the-shelf numerical solver for convex problems, by substituting $p(\mathbf{G}, s_1, s_2)$ with its empirical distribution $\hat{p}(\mathbf{G}, s_1, s_2) := \frac{1}{N_{\text{sim}}} \sum_{n=1}^{N_{\text{sim}}} \mathbb{1}[(\mathbf{G}, s_1, s_2) = (\mathbf{G}_n, q_1(\mathbf{G}_n), q_2(\mathbf{G}_n))]$ obtained from $N_{\text{sim}} = 1000$ i.i.d. samples $\{\mathbf{G}_n\}_{n=1}^{N_{\text{sim}}}$. This allows us to replace the expectation in (4.17) with a finite sum of N_{sim} convex functions. The capacity obtained is exact for a channel with state distribution equal to the empirical distribution $\hat{p}(\mathbf{G}, s_1, s_2)$, and approximates the capacity for $p(\mathbf{G}, s_1, s_2)$ as N_{sim} grows large. We repeat the above experiment for $M = 2$ and $M = 1$ RX antennas, two settings called, respectively, 2×2 *decentralized MIMO* and 2×1 *decentralized MISO*.

In Figure 4.2 we plot the resulting capacity versus the SNR. For comparison,

²Note that all the results presented in this section do not require the fading alphabet \mathcal{S} to be a discrete set, but only $|\mathcal{S}_l| < \infty$ for $l \in \mathcal{L}$.

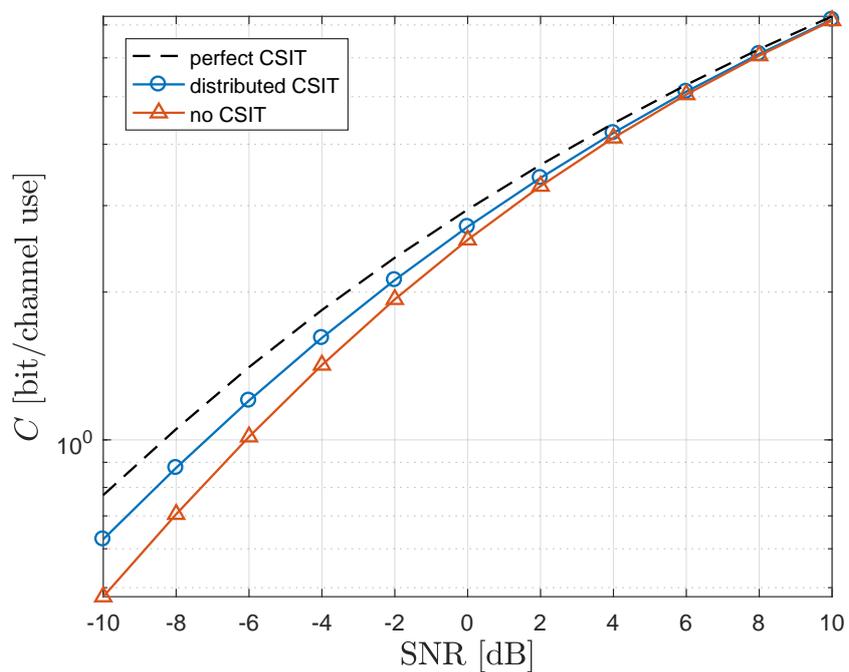
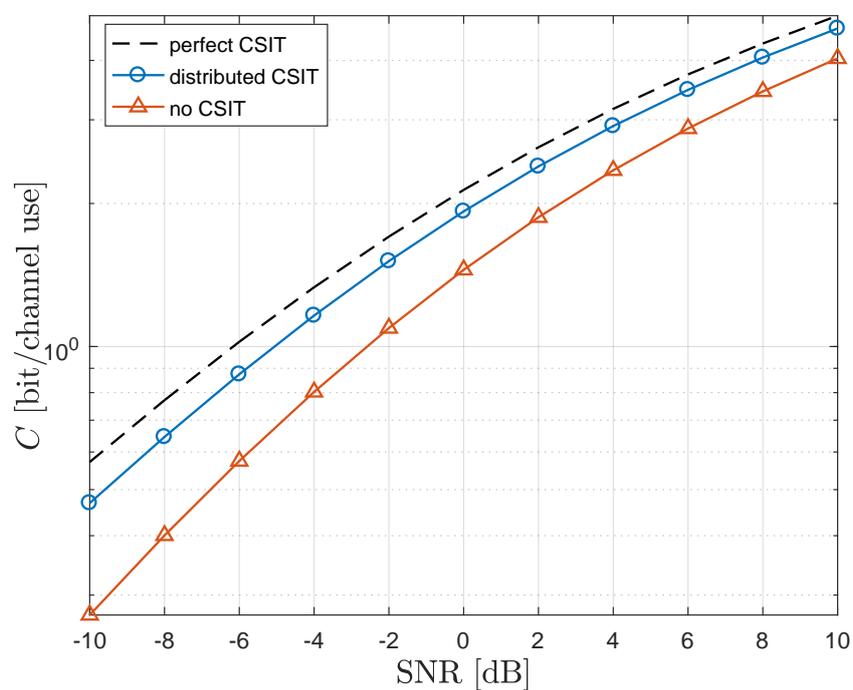
(a) 2×2 decentralized MIMO(b) 2×1 decentralized MISO

Figure 4.2: Capacity vs the SNR of the considered decentralized multi-antenna fading channel with asymmetric CSI feedback rates, $L = 2$ single-antenna TXs, and (a) $M = 2$ or (b) $M = 1$ RX antennas.

we also plot the capacity for perfect CSIT at both TXs

$$C^{(\text{perf. CSIT})} = \max_{\substack{\boldsymbol{\Sigma}(\mathbf{G}) \in \mathcal{S}_+^2: \\ \mathbb{E}[\boldsymbol{\Sigma}(\mathbf{G})]_{l,l} \leq P_l}} \mathbb{E} \left[\log \det (\mathbf{I} + \mathbf{G}^H \boldsymbol{\Sigma}(\mathbf{G}) \mathbf{G}) \right], \quad (4.21)$$

and for no CSIT

$$C^{(\text{no CSIT})} = \max_{\boldsymbol{\Sigma} \in \mathcal{S}_+^2: [\boldsymbol{\Sigma}]_{l,l} \leq P_l} \mathbb{E} \left[\log \det (\mathbf{I} + \mathbf{G}^H \boldsymbol{\Sigma} \mathbf{G}) \right]. \quad (4.22)$$

We recall that these CSIT configurations are equivalent to a centralized MIMO system. For a fair comparison, these capacities are computed over the same empirical marginal distribution $\hat{p}(\mathbf{G})$. As expected, for the MIMO case, the capacity gain given by distributed CSIT w.r.t no CSIT follows the well-known *beamforming gain* trend of the perfect CSIT case, i.e., it vanishes in the high SNR regime. Similarly, for the MISO case, this gain converges to a constant power offset.

4.3 Further comments on the optimal number of data streams

In this section we further elaborate on the optimal number of data streams d by addressing two important questions left open by Section 4.1. Theorem 4.1 shows that using a number of precoded data streams $d = D$ is a sufficient condition for achieving the capacity $C(P_1, P_2)$. However, we know that for some CSIT configurations (e.g., for centralized CSIT, as already discussed), $d > 2$ is not necessary. A first crucial question is whether there exists some distributed CSIT configuration for which such a condition is indeed necessary. In the next proposition we answer positively to this question.

Lemma 4.2. *For some $p(\mathbf{G}, s_2, s_2)$ and power constraints (P_1, P_2) , restricting $d \leq LT$ in Problem (4.3), where $LT = 2$ is the total number of TX antennas, leads to strictly suboptimal rates.*

Proof. The proof is given in Appendix B.2.1, by showing the existence of a CSI distribution $p(\mathbf{G}, s_2, s_2)$ with binary CSIT $|\mathcal{S}_1| = |\mathcal{S}_2| = 2$ such that $d \geq 3$ is necessary for achieving $C(P_1, P_2)$. \square

A second natural question is whether the developed upper bound $d \leq D = |\mathcal{S}_1| + |\mathcal{S}_2|$ is tight, for some $p(\mathbf{G}, s_1, s_2)$. In the following we answer negatively to this question, by showing that indeed we can consider a slightly tighter upper bound. However, we firstly remark that obtaining tighter bounds is not trivial and is in fact related to the well-known low-rank matrix completion problem [55]. Let

us consider the matrix $\mathbf{Q} \in \mathbb{S}_+^D$ defined in the proof of Theorem 4.1, or equivalently in Lemma 4.1, and its partition into blocks

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_{12} \\ \mathbf{Q}_{12}^H & \mathbf{Q}_2 \end{bmatrix}, \quad \mathbf{Q}_l \in \mathbb{S}_+^{|\mathcal{S}_l|}. \quad (4.23)$$

Informally, we recall that \mathbf{Q} collects the elements of the conditional input covariances $\Sigma(S_1, S_2)$ for all realizations of (S_1, S_2) . By direct inspection of the capacity expression (4.3), or equivalently of the objective in Problem (4.17), we observe that the off-diagonal elements of the sub-matrices \mathbf{Q}_l do not contribute to the achievable rate, since they do not correspond to any element of any realization of $\Sigma(S_1, S_2)$. Hence, by letting $\tilde{\mathbf{Q}}$ be any optimal solution of (4.17), the solution \mathbf{Q}^* of the (non-convex) problem

$$\begin{aligned} & \underset{\mathbf{Q} \in \mathbb{S}_+^D}{\text{minimize}} && \text{rank}(\mathbf{Q}) \\ & \text{subject to} && \mathbf{Q}_{12} = \tilde{\mathbf{Q}}_{12} \\ & && [\mathbf{Q}]_{i,i} = [\tilde{\mathbf{Q}}]_{i,i}, \quad i = 1, \dots, D \end{aligned} \quad (4.24)$$

is also an optimal solution of (4.17), but where the off-diagonal elements of \mathbf{Q}_l have been optimized such that the rank is minimized. Since we have seen that the rank $r \leq D$ of \mathbf{Q}^* corresponds to the dimension d of optimal distributed precoders (see Section 4.2), establishing a tighter upper-bound on d can be cast into finding an upper-bound on the solution of (4.24), which is an instance of a low-rank (semi-definite) matrix completion problem (see, e.g., [55]).

To the best of the authors knowledge, non-trivial upper-bounds to problems of the type (4.24) remain elusive. Nevertheless, in the following proposition we provide a simple result showing the existence of a tighter upper-bound than $D = |\mathcal{S}_1| + |\mathcal{S}_2|$.

Lemma 4.3. *The capacity $C(P_1, P_2)$ given by Theorem 4.1 is also achievable by letting*

$$d \leq |\mathcal{S}_1| + |\mathcal{S}_2| - 1.$$

Proof. The proof follows by manipulating an off-diagonal entry of the sub-matrices $\mathbf{Q}_1, \mathbf{Q}_2$ given by (4.23) until \mathbf{Q} becomes rank-deficient while still maintaining its positive semi-definiteness. Consider the symmetric matrix

$$\tilde{\mathbf{Q}}(t) := \mathbf{Q} + t \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad t \in \mathbb{R}, \quad \mathbf{Q} \in \mathbb{S}_+^D. \quad (4.25)$$

Let $\lambda_{\min} : \mathbb{S}^D \rightarrow \mathbb{R}$ be the minimum eigenvalue of a Hermitian symmetric matrix (not necessarily positive semi-definite). By definition, $\lambda_{\min}(\tilde{\mathbf{Q}}(0)) \geq 0$. In addition,

$\exists t_1 > 0$ such that $\lambda_{\min}(\tilde{\mathbf{Q}}(t_1)) < 0$. Furthermore, by the continuity of the map λ_{\min} in the matrix entries [56, Theorem 5.2], $\lambda_{\min}(\tilde{\mathbf{Q}}(t))$ is a continuous function of t . Hence, by the intermediate value theorem, $\exists t_0 \in [0, t_1]$ such that $\lambda_{\min}(\tilde{\mathbf{Q}}(t_0)) = 0$, i.e. such that $\tilde{\mathbf{Q}}(t_0)$ is positive semi-definite, low rank, and achieves the same rates as \mathbf{Q} . \square

The above result is by no means satisfactory, since the dimensionality reduction is marginal for large CSIT alphabets. Informally, the main limitation of the above bound is that the proof optimizes only one of the (coupled) variables in (4.24). However, note that the above bound is tight for the toy example considered in the proof of Proposition 4.2.

4.3.1 Example: asymmetric feedback rates (cont.)

We continue the example of Section 4.2.1, and compare the obtained capacities with the rates achieved by a simple distributed precoding scheme entailing the transmission of $d = 2$ data streams only, hence satisfying the classical design choice $d \leq LN$. Specifically, we use the (robust) *naïve* precoding scheme given in [19], which consists in optimizing the precoder of each TX by assuming that the other TX shares the same CSIT. More precisely, we let each TX l solve

$$\max_{\substack{\boldsymbol{\Sigma}(S_l) \in \mathfrak{S}_+^2 \\ \mathbb{E}[\boldsymbol{\Sigma}(S_l)_{j,j}] \leq P_j}} \mathbb{E} [\log \det (\mathbf{I}_M + \mathbf{G}^H \boldsymbol{\Sigma}(S_l) \mathbf{G})], \quad (4.26)$$

and take $\mathbf{f}_l(S_l) = \boldsymbol{\Sigma}^{\frac{1}{2}}(S_l) \mathbf{e}_l$, i.e., each TX l computes both precoders based on S_l , but only keeps its own relevant vector. Figure 4.3 plots the ergodic rates (3.46) achieved by naïve precoding over the considered decentralized MIMO and decentralized MISO channels, as well as their respective capacities. The simulation results show that the benefits of transmitting $d > 2$ data streams may be in fact rather marginal, since a much simpler scheme entailing the transmission of $d = 2$ data streams performs very close to capacity.

4.4 Extension to multiple-access channels

In this section we show that the unconventional achievability proof presented in Section 4.1 can be applied to extend Theorem 4.1 to the capacity region $\mathcal{C}_k^{\text{MAC}}$ of the MAC towards an arbitrary RX k . Specifically, we consider the same MIMO fading model with asymmetric feedback of Section 4.1, together with the coding theorems of Section 3.1.4, and omit as usual the RX subscript k everywhere. We then obtain:

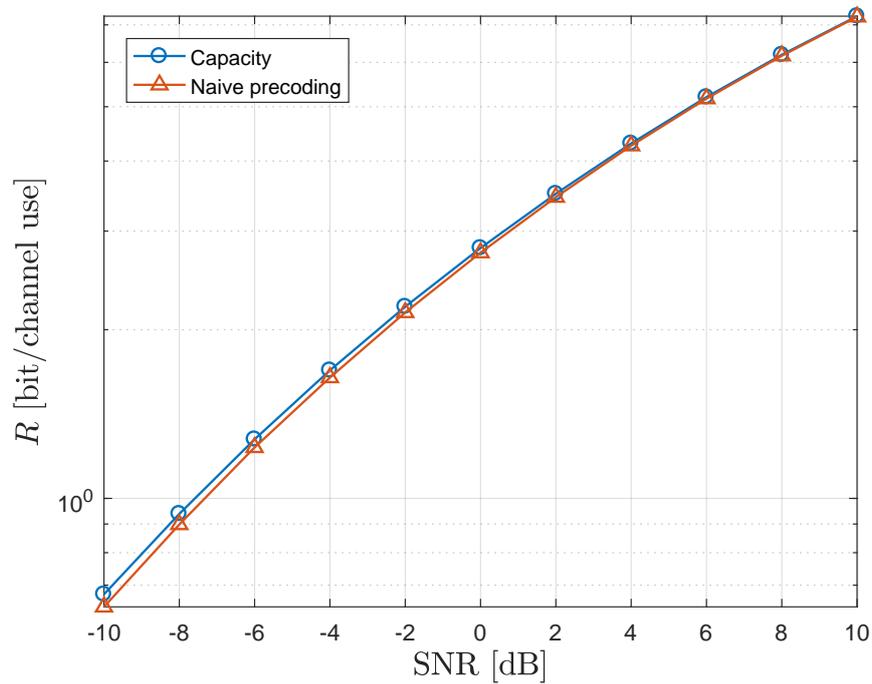
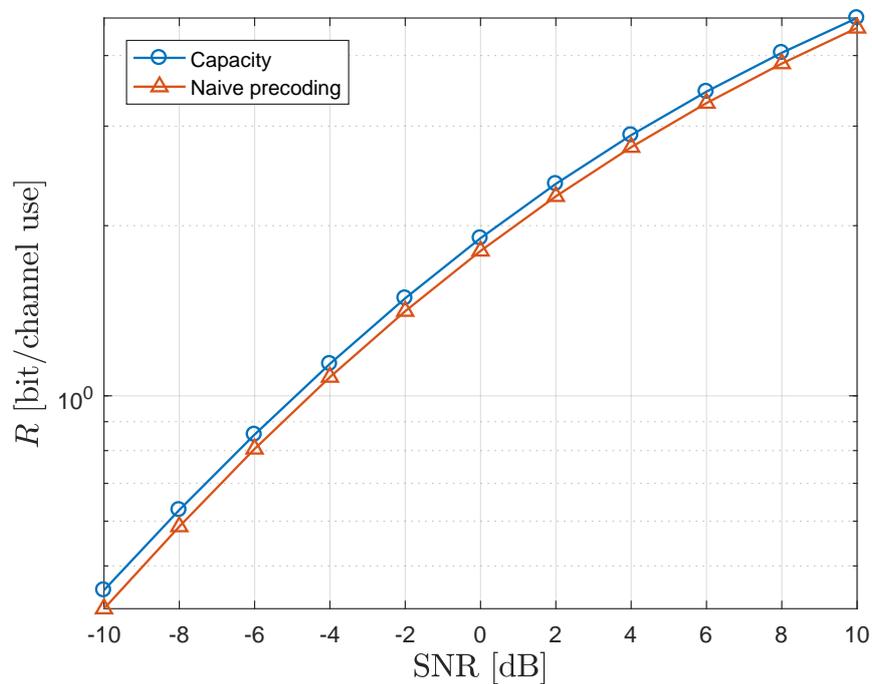
(a) 2×2 decentralized MIMO(b) 2×1 decentralized MISO

Figure 4.3: Comparison between the capacity of the considered channel and the rate achieved by a suboptimal distributed precoding scheme entailing the transmission of only $d = 2 \ll D$ data streams.

Theorem 4.2. *The capacity region \mathcal{C}^{MAC} of the considered channel is given by the union of all rate triples (R_0, R_1, R_2) such that*

$$R_1 \leq \mathbf{E} [\log (1 + \gamma_1(S_1) \|\mathbf{g}_1\|^2)] \quad (4.27)$$

$$R_2 \leq \mathbf{E} [\log (1 + \gamma_2(S_2) \|\mathbf{g}_2\|^2)] \quad (4.28)$$

$$R_1 + R_2 \leq \mathbf{E} [\log \det (\mathbf{I} + \mathbf{G}^{\text{H}} \text{diag}(\gamma_1(S_1), \gamma_2(S_2)) \mathbf{G})] \quad (4.29)$$

$$R_0 + R_1 + R_2 \leq \mathbf{E} [\log \det (\mathbf{I} + \mathbf{G}^{\text{H}} \boldsymbol{\Sigma}(S_1, S_2) \mathbf{G})] \quad (4.30)$$

where

$$\boldsymbol{\Sigma}(S_1, S_2) = \begin{bmatrix} \mathbb{f}_1^{\text{H}}(S_1) \\ \mathbb{f}_2^{\text{H}}(S_2) \end{bmatrix} [\mathbb{f}_1(S_1) \ \mathbb{f}_2(S_2)] + \begin{bmatrix} \gamma_1(S_1) & 0 \\ 0 & \gamma_2(S_2) \end{bmatrix}, \quad (4.31)$$

for some $\mathbb{f}_l : \mathcal{S}_l \rightarrow \mathbb{C}^d$ and $\gamma_l : \mathcal{S}_l \rightarrow \mathbb{R}_+$ such that

$$d \leq D := |\mathcal{S}_1| + |\mathcal{S}_2|, \quad (4.32)$$

and $\mathbf{E} [\|\mathbb{f}_l(S_l)\|^2] + \mathbf{E} [\gamma_l(S_l)] \leq P_l$ for $l = 1, 2$. Furthermore any rate triple in $\mathcal{C}^{\text{MAC}}(P_1, P_2)$ can be achieved by letting

$$X_l = \mathbb{f}_l^{\text{H}}(S_l) \mathbf{u} + \sqrt{\gamma_l(S_l)} V_l, \quad (4.33)$$

where $(\mathbf{u}, V_1, V_2) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{D+2})$ carry the common and individual messages.

Proof. The proof is given in Appendix B.2.2. \square

Theorem 4.2 shows that superposition of jointly and independently encoded Gaussian codewords achieves the capacity region. However, while encoding of the individual messages W_1, W_2 follows traditional approaches (one power-controlled stream per TX), the joint encoding of W_0 may require a larger number of precoded data streams (beyond two streams in our case) as already seen for Theorem 4.1.

4.5 Joint precoding and feedback design

In this section we exploit the main results of this chapter to address the related problem of optimal CSIT feedback design given asymmetric constraints on the feedback rates. Ideally, we are interested in solving the following optimization problem:

$$\max_{\substack{q_l(\mathbf{G}) \in \mathcal{S}_l, \\ |\mathcal{S}_l| = 2^{\beta_l}}} C(P_1, P_2), \quad (4.34)$$

where the channel capacity is maximized by optimizing the quantizers $q_l(\mathbf{G})$ under feedback rate constraints (β_1, β_2) . Inspired by the method developed in [57] for centralized channels, we propose a suboptimal approach based on the *generalized Lloyd algorithm* [58] for vector quantization, which alternates between precoding design and feedback design as follows:

1. Quantizers update step:

$$(q_1^*(\mathbb{G}), q_2^*(\mathbb{G})) \in \arg \max_{(s_1, s_2) \in \mathcal{S}_1 \times \mathcal{S}_2} r(\mathbb{G}, \mathbf{Q}), \quad (4.35)$$

2. Precoders update step:

$$\mathbf{Q}^* = \arg \max_{\substack{\mathbf{Q} \in \mathfrak{S}_+^D \\ [\mathbf{Q}]_{i,i} \leq P_1, i=1, \dots, |\mathcal{S}_1| \\ [\mathbf{Q}]_{j,j} \leq P_2, j=(|\mathcal{S}_1|+1), \dots, D}} \mathbb{E} [r(\mathbb{G}, \mathbf{Q})], \quad (4.36)$$

where we defined

$$r(\mathbb{G}, \mathbf{Q}) := \log \det (\mathbf{I} + \mathbb{G}_{\text{eq}}^H \mathbf{Q} \mathbb{G}_{\text{eq}}),$$

with $\mathbb{G}_{\text{eq}} = \mathbf{E}^H(S_1, S_2)\mathbb{G}$ given by Lemma 4.1. Note that the precoders update step is a convex problem and corresponds to Problem (4.17) slightly modified such that the average power constraint is replaced by an instantaneous power constraint for each realization of S_l . This modification avoids the quantizers update step to produce unfeasible solutions.

The two above steps are respectively similar to the *Voronoi regions* update step and the *centroid* update step of the Lloyd algorithm, under a modified distortion measure $-r(\mathbb{G}, \mathbf{Q})$. However, there is a non-trivial difference between the proposed algorithm and the ones in [57, 58]. In particular, because of the distributed CSIT assumption, the centroids must be jointly optimized in a unique step, and not disjointly as in the classical Lloyd algorithm.

4.5.1 Example: asymmetric feedback rates (cont.)

We continue the example of Section 4.2.1, by replacing the random vector quantizers in the Frobenius norm with (4.35), where \mathbf{Q} is designed according to the proposed Lloyd-type alternating optimization algorithm. Note that, for an arbitrary initialization, this method converges to a local optimum of (4.34). Classical multi-start methods may be applied to avoid bad local optima, but this step is here omitted. The distributed precoders \mathbf{Q} are optimized using $N_{\text{sim}} = 100$ training samples, and the performance are evaluated over a different test set of $N_{\text{test}} = 10000$ samples.

For comparison, we consider the capacity assuming no CSIT (4.22), a centralized method similar to [57] assuming symmetric feedback of rate $\min(\beta_1, \beta_2)$ or $\max(\beta_1, \beta_2)$, and the following suboptimal scheme adapting [57] to asymmetric feedback following again the naïve precoding idea of Section 4.3.1. More precisely, for the last of the aforementioned schemes, we run an alternating optimization procedure composed by the following two steps:

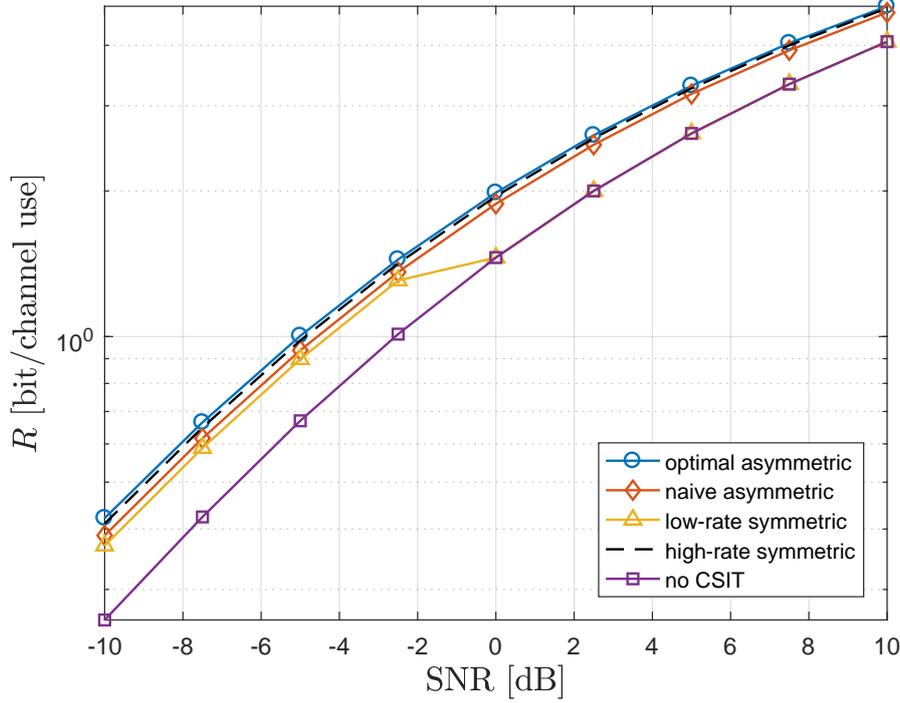


Figure 4.4: Performance comparison between the proposed joint precoding and feedback design and other suboptimal or unfeasible methods, for a 2×1 decentralized MISO channel with feedback rates $(\beta_1, \beta_2) = (2, 1)$ bits.

1. $q_l(\mathbf{G}) \in \arg \max_{s_l \in \mathcal{S}_l} \log \det (\mathbf{I} + \mathbf{G}^H \Sigma(S_l) \mathbf{G})$,
2. $\Sigma(S_l) \in \arg \max_{\substack{\Sigma \in \mathcal{S}_+^2 \\ [\Sigma]_{l,l} \leq P_l}} \mathbb{E} [\log \det (\mathbf{I} + \mathbf{G}^H \Sigma \mathbf{G}) | S_l]$,

and finally let $\mathbb{f}_l(S_l) = \Sigma^{\frac{1}{2}}(S_l) \mathbf{e}_l$. This method is termed *naïve asymmetric*. For a fair comparison, its performance is evaluated using the rate optimal feedback strategy (4.35). We remark that the aforementioned centralized methods, termed *low-rate symmetric* and *high-rate symmetric*, correspond to the above naïve procedure when both TXs have the same feedback rate $\min(\beta_1, \beta_2)$ or $\max(\beta_1, \beta_2)$, respectively.

Figure 4.4 plots the performance of the considered schemes for the 2×1 decentralized MISO setup, demonstrating the gains of distributed precoding and feedback design, especially in the high SNR regime. In particular, the low-rate symmetric scheme exploits only $\min(\beta_1, \beta_2) = 1$ bit of feedback and its rate converges to the no CSIT cases in the high SNR regime. In contrast, the naïve asym-

metric and the proposed (locally) optimal scheme show that performance gains can be obtained by allocating additional feedback bits to only one TX. Interestingly, we also observe that the high-rate symmetric scheme, which is unfeasible since it assumes symmetric feedback rate constraints of 2 bits, performs worse than the proposed scheme. This can be explained by the fact that the optimal asymmetric feedback scheme conveys $\beta_1 + \beta_2 = 3$ bits of information, which, although distributed across the TXs, is more than the $\max(\beta_1, \beta_2) = 2$ bits conveyed by the high-rate symmetric scheme. This also suggests that, to fully exploit the available feedback opportunities, symmetric feedback schemes should be avoided even if the feedback rate constraints are symmetric.

Chapter 5

Team Theory for Distributed Precoding Design

It is well-known that linear precoding and TIN offer an excellent compromise between performance and complexity in many wireless communication systems, such as in massive MIMO networks [2]. However, for this to hold, the linear precoders must be properly designed. The main goal of this chapter is to develop guidelines for optimal precoding design in decentralized networks, for which no practical method is currently available.

5.1 Distributed precoding design

5.1.1 A challenging problem

Ideally, in this chapter we aim at designing distributed linear precoders spanning the largest possible subset of $\mathcal{C}(P_1, \dots, P_L)$. However, in many setups, one quickly realizes that the already challenging issues behind centralized precoding design are greatly exacerbated by the information constraints induced by the distributed CSIT assumption. For example, let us consider the following distributed variation of the classical sum-rate maximization problem

$$\underset{\substack{\mathbb{F}_{l,k}(S_l) \in \mathbb{C}^{d_k \times N} \\ \sum_{k=1}^K \mathbb{E}[\|\mathbb{F}_{l,k}(S_l)\|_{\mathbb{F}}^2] \leq P_l}}{\text{maximize}} \sum_{k=1}^K R_k^{\text{CSIR}}, \quad (5.1)$$

where R_k^{CSIR} is the achievable rate bound under perfect CSIR given by (3.46), and where the optimization is over the space of distributed linear precoders as defined in Section 3.2.2.

Centralized CSIT Mostly because of the non-convexity of the utility function, Problem (5.1) is known to be challenging even by assuming centralized CSIT, and hence requires simplifications and/or suboptimal methods. The best known approach for solving (5.1) under centralized CSIT not requiring prohibitively complex global optimization techniques is perhaps the so-called *weighted-MMSE* (WMMSE) method introduced in [59] for the MIMO BC with perfect CSIT, and subsequently adapted to other setups including imperfect (yet centralized) CSIT [60, 52]. Another popular approach is to exploit the so-called *UL-DL duality* principle to cast the precoding design problem as an equivalent linear combining design problem in a virtual dual UL channel [61, 25]. Remarkably, both approaches reveal a MMSE-type structure of the optimal centralized precoders. The latter approach is slightly more complex than the former, but it shows the optimality of MMSE-type precoders for a more general class of network utilities than (5.1) or its weighted variation.

Distributed CSIT Moving to the distributed CSIT assumption, we observe that Problem (5.1) is also an instance of a *team decision* problem [34, 35], a class of multi-agent optimization problems dealing with decision making under asymmetry of information. This class of problems is exposed to an additional layer of considerable difficulty stemming from the (non-trivial) information constraint. Unfortunately, the available team theoretical results provide little help in dealing with complicated cost functions as in Problem (5.1). The early attempts on distributed precoding design, surveyed in [19], mostly rely on the suboptimal notion of *person-by-person optimality* (see Appendix A) to directly tackle Problem 5.1 via an alternating sequence of subproblems for each TX l , with centralized information S_l . However, each subproblem is still very challenging due to the non-convexity of the utility, but also for the difficulties in being robust against the generally unknown effect of the other TXs. Therefore, most of the available methods resort to rather poor or impractical finite dimensional representations of the state and/or the optimization space (e.g., using neural networks).

Instead of (directly) addressing difficult problems as Problem 5.1, in this chapter we follow the UL-DL duality approach and rigorously motivate a novel distributed version of the MMSE precoding design criterion. As we will show, the major advantage of this approach is that it leads to one of the few team theoretical formulations for which solid solution approaches are available.

5.1.2 Definitions and assumptions

We specialize the distributed linear precoding scheme defined in Section 3.2.2 by taking the following assumptions. We assume $M = 1$ antennas at each RX, and

adopt the traditional design choice $d_k = 1$ data stream per RX. Furthermore, we focus on $N < K$ antennas per TX, that is, on the regime where TX cooperation is particularly important for effective spatial interference management [16]. Overall, the above assumptions fit the typical architecture of current wireless networks, which lets the RXs perform simple single-stream decoding in Gaussian noise, and leaves most of the interference management burden to the TXs.

To allow for a rigorous treatment of the information constraint, we use a more formal measure theoretical definition of the random linear precoder $\mathfrak{t}_{l,k}$ (see Appendix A). Specifically, we let $(\Omega, \Sigma_0, \mathbf{P})$ be the underlying probability space over which all random quantities are defined, and let

$$\mathfrak{t}_{l,k} = \mathfrak{t}_{l,k}(S_l) \quad (5.2)$$

be an element of the vector space \mathcal{T}_l of square-integrable Σ_l -measurable functions¹ $\Omega \rightarrow \mathbb{C}^N$, where $\Sigma_l \subseteq \Sigma_0$ denotes the sub- σ -algebra generated by S_l on Ω , called the *information subfield* of TX l [34, 35]. This assumption² rigorously describes the functional dependency of $\mathfrak{t}_{l,k}$ on the portion S_l of the overall system randomness, and includes a reasonable finiteness constraint $\mathbf{E}[\|\mathfrak{t}_{l,k}\|^2] < \infty$ on precoders power. Then, the full precoding vector $\mathfrak{t}_k^T = [\mathfrak{t}_{1,k}^T \ \dots \ \mathfrak{t}_{L,k}^T]^T$ for message W_k must satisfy $\mathfrak{t}_k \in \mathcal{T} := \prod_{l=1}^L \mathcal{T}_l$.

Because of the less stringent CSIR requirements, and, most importantly, for treatability reason, we consider achievable rates measured by the hardening bound (3.49)

$$R_k^{\text{hard}} = \log \left(1 + \frac{p_k \mathbf{E}[\mathbf{g}_k^H \mathfrak{t}_k]^2}{p_k \text{Var}[\mathbf{g}_k^H \mathfrak{t}_k] + \sum_{j \neq k} p_j \mathbf{E}[\|\mathbf{g}_k^H \mathfrak{t}_j\|^2] + 1} \right), \quad \forall k \in \mathcal{K}, \quad (5.3)$$

where $\mathbf{E}[\|\mathfrak{x}_l\|^2] = \sum_{k=1}^K p_k \mathbf{E}[\|\mathfrak{t}_{l,k}\|^2]$ is the power used by TX l . We then let $\mathcal{R}^{\text{hard}}$ be the union of all rate tuples $(R_1, \dots, R_K) \in \mathbb{R}^K$ such that $R_k \leq R_k^{\text{hard}} \ \forall k \in \mathcal{K}$ for some set of distributed precoders $\{\mathfrak{t}_k\}_{k=1}^K$ and power allocation policy $\{p_k\}_{k=1}^K$ satisfying

$$\sum_{l=1}^L \mathbf{E}[\|\mathfrak{x}_l\|^2] \leq \sum_{l=1}^L P_l =: P_{\text{sum}} < \infty. \quad (5.4)$$

The set $\mathcal{R}^{\text{hard}}$ is an inner bound for the capacity region of the considered network with distributed CSIT, no CSIR, and subject to a long-term sum power constraint

¹Differently than Appendix A, we here use \mathcal{T}_l to denote the function set, and not the alphabet \mathbb{C}^N of $\mathfrak{t}_{l,k}$.

²The measure theoretical formulation presented above is necessary for establishing Theorem 5.3. However, the rest of this thesis does not require any particular measure theoretical background.

P_{sum} , that is

$$\mathcal{R}^{\text{hard}} \subseteq \bigcup_{\sum_{l=1}^L P_l \leq P_{\text{sum}}} \mathcal{C}(P_1, \dots, P_L). \quad (5.5)$$

The long-term sum power constraint is chosen because it allows for strong analytical results and simplifies system design. This constraint may be directly relevant for systems such as the radio stripes, treated in Section 5.3.2, where all the TXs share the same power supply [27]. However, note that many simple heuristic methods, such as power scaling factors, can be applied to adapt systems designed under a long-term sum power constraint to more restrictive cases (see Section 5.5.2 for additional details). Finally, due to its importance in system design and resource allocation, we consider the notion of (weak) Pareto optimality on $\mathcal{R}^{\text{hard}}$ and we mostly focus on the (weak) Pareto boundary of $\mathcal{R}^{\text{hard}}$, denoted by $\partial\mathcal{R}^{\text{hard}}$ [25]. As a final remark, note that, by the Cauchy–Schwarz inequality and the assumption on finite fading power, we always have $p_k |\mathbf{E}[\mathbf{g}_k^H \mathbf{t}_k]|^2 < \infty$, hence $\partial\mathcal{R}^{\text{hard}}$ is finite.

5.2 Team MMSE precoding

We study the following novel *team* MMSE precoding design criterion: given a vector of nonnegative weights $\mathbf{w} := [w_1, \dots, w_K]^\top$ belonging to the simplex $\mathbb{R}_\Delta := \{\mathbf{w} \in \mathbb{R}_+^K \mid \sum_{k=1}^K w_k = K\}$, we consider for each $k \in \mathcal{K}$ the problem

$$\underset{\mathbf{t}_k \in \mathcal{T}}{\text{minimize}} \text{MSE}_k(\mathbf{t}_k) := \mathbf{E} \left[\left\| \mathbf{W}^{\frac{1}{2}} \mathbb{H} \mathbf{t}_k - \mathbf{e}_k \right\|^2 + \frac{\|\mathbf{t}_k\|^2}{P} \right], \quad (5.6)$$

where $\mathbf{W} := \text{diag}(w_1, \dots, w_K)$, \mathbf{e}_k is the k -th column of \mathbf{I}_K , and $P := P_{\text{sum}}/K$. A solution to the above problem can be recognized as a distributed version of the classical centralized MMSE precoding design [2]. For $P \rightarrow \infty$, it can be interpreted as the ‘closest’ distributed approximation of the ZF solution. By means of team theoretical arguments, this section provides fundamental properties of the optimal solution to Problem (5.6). Before providing the main results of this section, we also revisit the effectiveness of the MSE criterion in terms of network performance, which is well-known for centralized precoding.

Remark 5.1. *Hereafter, with the exception of Section 5.2.1, we consider w.l.o.g. $\mathbf{W} = \mathbf{I}$. The general case will readily follow by replacing \mathbb{H}_l with $\mathbf{W}^{\frac{1}{2}} \mathbb{H}_l$ everywhere.*

5.2.1 Achievable rates via uplink-downlink duality

This section discusses the formal connection between the objective of Problem (5.6) and $\mathcal{R}^{\text{hard}}$ by revisiting UL-DL duality under the distributed CSIT assumption.

Theorem 5.1. Consider an arbitrary set of distributed precoders $\{\mathfrak{t}_k\}_{k=1}^K$ and weights $\mathbf{w} \in \mathbb{R}_\Delta$. Then, any rate tuple $(R_1, \dots, R_K) \in \mathbb{R}^K$ such that

$$R_k \leq \log(\text{MSE}_k(\mathfrak{t}_k))^{-1} \quad (5.7)$$

belongs to $\mathcal{R}^{\text{hard}}$. The power allocation policy $\{p_k\}_{k=1}^K$ achieving the above inner bound is given in Appendix B.3.1, and satisfies $\sum_{l=1}^L \mathbb{E}[\|\mathfrak{x}_l\|^2] = P_{\text{sum}}$. Furthermore, if \mathfrak{t}_k solves Problem (5.6) $\forall k \in \mathcal{K}$, then (R_1, \dots, R_K) with $R_k = \log(\text{MSE}_k(\mathfrak{t}_k))^{-1}$ is Pareto optimal, and every rate tuple in $\partial\mathcal{R}^{\text{hard}}$ is obtained for some $\mathbf{w} \in \mathbb{R}_\Delta$.

Proof. The proof is based on connecting the solution to Problem (5.6) and the solution to the problem of ergodic rate maximization in a dual UL channel, where \mathbf{w} is a dual UL power allocation vector, and where achievable rates are measured by using the so-called *use-and-then-forget* (UatF) bound [2, Theorem 4.4]. The details are given in Appendix B.3.1. \square

Theorem 5.1 states that the Pareto boundary of $\mathcal{R}^{\text{hard}}$ can be parametrized by $K - 1$ nonnegative real parameters, i.e., by the weights $\mathbf{w} \in \mathbb{R}_\Delta$. A similar parametrization was already known for deterministic channels (see, e.g., [25]), or, equivalently, for fading channels with perfect CSIT and CSIR. This work extends the aforementioned results to imperfect and possibly distributed CSIT, and no CSIR. In theory, \mathbf{w} should be selected according to some network utility (e.g., the sum rate or the max-min rate). In practice, \mathbf{w} is often fixed heuristically (e.g., from the real UL powers), while the network utility is optimized a posteriori by varying the DL power allocation policy $\{p_k\}_{k=1}^K$.

From a precoding design point of view, Theorem 5.1 generalizes the duality-based argument behind classical MMSE precoding given by [2]. While [2] motivates the MMSE solution as the optimal combiner maximizing a dual UL ergodic rate bound based on coherent decoding, the proof of Theorem 5.1 directly relates the MSE criterion to the more conservative UatF bound. This last point is particularly relevant under distributed CSIT, where an optimal solution to the coherent ergodic rate maximization problem is not known in general. We recall that, in turn, [2] generalizes classical duality-based arguments for deterministic channels to fading channels. As a concluding remark, we stress that the inner bound (5.7) should not be confused with the well-known inner bound based on the notion of *expected WMSE* given in [60], which does not rely on UL-DL duality and hence the precoders for *all* messages contribute to each rate bound.

5.2.2 Quadratic teams for distributed precoding design

Problem (5.6) is an instance of a *team decision* problem, as defined in Appendix A. In particular, by rewriting the objective as

$$\text{MSE}_k(\mathbf{t}_k) = \mathbb{E}[c_k(\mathbb{H}, \mathbf{t}_{1,k}, \dots, \mathbf{t}_{L,k})], \quad (5.8)$$

$$c_k(\mathbf{H}, \mathbf{t}_{1,k}, \dots, \mathbf{t}_{L,k}) := \mathbf{t}_k^H \mathbf{O} \mathbf{t}_k - 2\Re(\mathbf{g}_k^H \mathbf{t}_k) + 1, \quad (5.9)$$

where

$$\mathbf{O} := \mathbb{H}^H \mathbb{H} + \frac{1}{P} \mathbf{I}, \quad \mathbf{g}_k = \mathbb{H}^H \mathbf{e}_k, \quad (5.10)$$

and by noticing that $\mathbf{O} \succ \mathbf{0}$ a.s., we recognize that Problem (5.6) belongs to the class of *quadratic teams* as defined in [34, Sect. 4]. This class exhibits strong structural properties, in particular related to the *stationary* solution concept given by Definition A.3 in Appendix A. By applying standard results on differentiation of real-valued quadratic forms over a complex domain, the stationary conditions (A.5) can be here expressed as

Definition 5.1 (Stationary TMMSE solution). *A tuple $\mathbf{t}_k^* := (\mathbf{t}_{1,k}^*, \dots, \mathbf{t}_{L,k}^*) \in \mathcal{T}$ is a stationary solution for Problem (5.6) if $\text{MSE}_k(\mathbf{t}_k^*) < \infty$ and if the following set of equalities hold*

$$\mathbb{E}[\mathbf{O}_{l,l} | S_l] \mathbf{t}_{l,k}^*(S_l) + \sum_{j \neq l} \mathbb{E}[\mathbf{O}_{l,j} \mathbf{t}_{j,k}^* | S_l] - \mathbb{E}[\mathbf{g}_{k,l} | S_l] = \mathbf{0}_N \quad \text{a.s.}, \quad \forall l \in \mathcal{L}, \quad (5.11)$$

where $\mathbf{O}_{l,l} := \mathbb{H}_l^H \mathbb{H}_l + \frac{1}{P} \mathbf{I}$, $\mathbf{O}_{l,j} := \mathbb{H}_l^H \mathbb{H}_j$ for $j \neq l$, and $\mathbf{g}_{k,l} = \mathbb{H}_l^H \mathbf{e}_k$, provided that all expectations are finite.

As explained in Appendix A, each of the stationary conditions (5.11) can be interpreted as a first order optimality condition for optimizing $\mathbf{t}_{l,k}$ while keeping the others $\mathbf{t}_{j,k}^*$ fixed $\forall j \neq l$. This is reminiscent of the game theoretical notion of *Nash equilibrium*, with the difference that here all the TXs share the same objective, and hence they act as a team. Similarly to Nash equilibria, stationary solutions may be in general inefficient, i.e., lead to a local optimum. However, a stronger result holds for quadratic teams:

Theorem 5.2. *If \mathbf{O} is uniformly bounded above, i.e., there exists a positive scalar $B < \infty$ such that $\mathbf{O} \prec B \mathbf{I}$ a.s., then Problem (5.6) admits a unique optimal solution, which is also the unique stationary solution satisfying (5.11).*

Proof. Theorem 2.6.6 of [35]. □

The assumption of the above theorem is essentially used to ensure the existence of all the expectations in the steps of the proof, and is satisfied for any fading distribution with bounded support. However, it is not satisfied for the classical Gaussian fading model $\text{vec}(\mathbb{H}) \sim \mathcal{CN}(\boldsymbol{\mu}, \mathbf{K})$. Despite being unrealistic, since physically consistent fading distributions cannot have unbounded support, Gaussian fading is a very common model due to its analytical treatability, for example in deriving effective channel estimation error models. Furthermore, except for the tails of the distribution, it usually fits measurements well. Therefore, in the following we derive optimality results covering this case.

Theorem 5.3. *If $\mathbb{E}[\|\mathbb{O}\|_{\mathbb{F}}^2] < \infty$, then Problem (5.6) admits a unique optimal solution, which is also the unique stationary solution satisfying (5.11).*

Proof. The proof is given in Appendix B.3.2. □

The condition $\mathbb{E}[\|\mathbb{O}\|_{\mathbb{F}}^2] < \infty$ is stronger than the assumption $\mathbb{E}[\|\mathbb{H}\|_{\mathbb{F}}^2] < \infty$ given in Section 2.2. However, it is satisfied for Gaussian fading and any realistic distribution with bounded support. To summarize, we have seen that, under mild regularity requirements, solving Problem (5.6) is equivalent to finding the unique $\mathfrak{t}_k^* \in \mathcal{T}$ satisfying (5.11), which can be compactly rewritten as the following infinite dimensional linear feasibility problem [35]:

$$\text{find } \mathfrak{t}_k \in \mathcal{T} \quad \text{s.t.} \quad \mathbf{A}[\mathfrak{t}_k] = \hat{\mathbf{g}}_k \text{ a.s.}, \quad (5.12)$$

where

$$\hat{\mathbf{g}}_k \in \mathcal{T}, \quad \hat{\mathbf{g}}_k(S_1, \dots, S_L) := \begin{bmatrix} \mathbb{E}[\mathfrak{g}_{k,1}|S_1] \\ \vdots \\ \mathbb{E}[\mathfrak{g}_{k,L}|S_L] \end{bmatrix}, \quad (5.13)$$

and where $\mathbf{A} : \mathcal{T} \rightarrow \mathcal{T}$ is a linear operator given by

$$\mathbf{A} := \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_L \end{bmatrix}, \quad \mathbf{A}_l[\mathfrak{t}_k](S_l) := \mathbb{E} \left[\sum_{j=1}^L \mathbb{O}_{l,j} \mathfrak{t}_{j,k} \middle| S_l \right]. \quad (5.14)$$

It is generally difficult to solve (5.12) in closed form. However, the optimal TMMSE precoders may be approached via one of the many approximation methods available in the literature, for example the methods surveyed in [35] based on interpreting the solution to (5.12) as the unique fixed point of a linear map. Other promising methods may also include finite dimensional approximations of (5.12) obtained, e.g., by sampling the CSIT process (S_1, \dots, S_L) and by interpreting the sampled version of (5.12) as a classical function interpolation problem from a finite set of linear measurements [62]. Further discussions on approximate solution

methods are left for future work, and most of the parts of this chapter will focus on special cases where (5.12) can be solved explicitly.

Nevertheless, it turns out that the optimality conditions (5.11) are not only useful to characterize the optimal TMMSE solution, but also to evaluate the suboptimality of its approximations. This can be done via an appropriate measure of violation of the optimality conditions. Specifically, we have the following result:

Lemma 5.1. *Suppose that the assumption of Theorem 5.3 holds, and let \mathfrak{t}_k^* be the unique TMMSE solution to Problem (5.6). Furthermore, let $\mathfrak{z}_k := [\mathfrak{z}_{1,k}^\top, \dots, \mathfrak{z}_{L,k}^\top]^\top$ where $\mathfrak{z}_{l,k} = \mathfrak{z}_{l,k}(S_l)$ is given by the left-hand side of (5.11) with \mathfrak{t}^* replaced by an arbitrary $\mathfrak{t}_k \in \mathcal{T}$. If $\mathfrak{z}_k \in \mathcal{T}$, i.e., if $\mathbb{E}[\|\mathfrak{z}_{l,k}(S_l)\|^2] < \infty$, the following optimality bounds hold:*

$$\text{MSE}_k(\mathfrak{t}_k) - \text{MSE}_k(\mathfrak{t}_k^*) \leq \mathbb{E} \left[\|\mathbb{O}^{-\frac{1}{2}} \mathfrak{z}_k\|^2 \right] \quad (5.15)$$

$$\leq P\mathbb{E} \left[\|\mathfrak{z}_k\|^2 \right], \quad (5.16)$$

Proof. The proof is given in Appendix B.3.3. \square

Clearly, $\mathfrak{z}_k = \mathbf{0}$ a.s. gives the optimality conditions (5.11) or (5.12), and in fact it corresponds to a zero optimality gap in (5.15) and (5.16). Intuitively, the bounds in (5.15) and (5.16) can be quite tight if $\mathfrak{z}_k \approx \mathbf{0}$ with high probability. However, if this is not satisfied, we remark that both bounds can be looser than other trivial bounds obtained, e.g., by assuming a centralized information structure, or even output negative estimates of $\text{MSE}_k(\mathfrak{t}_k^*)$. As already mentioned, we leave further studies on suboptimal solutions for future work, and we use (5.16) only in Section 5.3.3 for getting analytical insights into a particular setup.

5.3 Applications to cell-free massive MIMO

In this section we specialize the theory of Section 5.2 to cell-free massive MIMO networks, and explicitly derive optimal TMMSE precoders for two practical examples. In the scope of this study, the important feature of the cell-free massive MIMO paradigm is the exploitation of time division duplex operations and channel reciprocity to efficiently acquire measurements $\hat{\mathbb{H}}_l$ of the local channel \mathbb{H}_l at each TX l via over-the-uplink training [15]. These measurements may be subsequently shared through the fronthaul according to some predefined CSIT sharing mechanism, forming at each TX l a side information about the global channel \mathbb{H} of the type

$$S_l := \left(\hat{\mathbb{H}}_l, \bar{S}_l \right) \quad (5.17)$$

where \bar{S}_l denotes the side information about the other channels $\{\mathbb{H}_j\}_{j \neq l}$ collected at TX l . Depending on the CSIT sharing mechanism, \bar{S}_l may be a function of the

other local channel estimates $\{\hat{\mathbb{H}}_j\}_{j \neq l}$ (e.g., in case of error-free digital signalling), or include additional noise (e.g., in case of random events such as protocol delays). Consistently with the above discussion, we consider the following assumptions:

Assumption 5.1 (Local channel estimation). *For every $l \in \mathcal{L}$, let $\mathbb{E}_l := \mathbb{H}_l - \hat{\mathbb{H}}_l$ be the local channel estimation error for the local channel. Assume that $\hat{\mathbb{H}}_l$ and \mathbb{E}_l are independent. Furthermore, assume $\mathbf{E}[\mathbb{E}_l] = \mathbf{0}$, and that $\mathbf{E}[\mathbb{E}_l^H \mathbb{E}_l] =: \boldsymbol{\Sigma}_l$ has finite elements. Finally, assume that $(\hat{\mathbb{H}}_l, \mathbb{E}_l)$ and $(\hat{\mathbb{H}}_j, \mathbb{E}_j)$ are independent for $l \neq j$.*

Assumption 5.2 (CSIT sharing mechanism). *For every $(l, j) \in \mathcal{L}^2$ s.t. $l \neq j$, assume the following Markov chain: $\mathbb{H}_l \rightarrow \hat{\mathbb{H}}_l \rightarrow S_l \rightarrow S_j \rightarrow \hat{\mathbb{H}}_j \rightarrow \mathbb{H}_j$.*

Assumption 5.1 is widely used in the wireless communication literature and it holds, e.g., for pilot-based MMSE estimates of Gaussian channels [15, 28]. Assumption 5.2 essentially states that all the available information about \mathbb{H}_l is fully contained in $\hat{\mathbb{H}}_l$ at TX l , and that TX j can only obtain a degraded version of it. We now rewrite the optimality conditions (5.11) or (5.12) in light of the considered model:

Lemma 5.2. *Suppose that Assumption 5.1, Assumption 5.2, and the assumption of Theorem 5.3 hold. Then, the unique TMMSE solution to Problem (5.6) is given by the unique $\mathfrak{t}_k^* \in \mathcal{T}$ satisfying*

$$\mathfrak{t}_{l,k}^*(S_l) = \mathbb{T}_l \left(\mathbf{e}_k - \sum_{j \neq l} \mathbf{E} \left[\hat{\mathbb{H}}_j \mathfrak{t}_{j,k}^* \middle| S_l \right] \right) \quad \text{a.s.}, \quad \forall l \in \mathcal{L}, \quad (5.18)$$

where $\mathbb{T}_l := \left(\hat{\mathbb{H}}_l^H \hat{\mathbb{H}}_l + \boldsymbol{\Psi}_l + P^{-1} \mathbf{I} \right)^{-1} \hat{\mathbb{H}}_l^H$.

Proof. The first term of the stationarity conditions (5.11) is evaluated by letting

$$\begin{aligned} \mathbf{E}[\mathbb{O}_{l,l} | S_l] &= \mathbf{E}[\mathbb{H}_l^H \mathbb{H}_l | S_l] + P^{-1} \mathbf{I} \\ &= \mathbf{E}[(\hat{\mathbb{H}}_l + \mathbb{E}_l)^H (\hat{\mathbb{H}}_l + \mathbb{E}_l) | \hat{\mathbb{H}}_l] + P^{-1} \mathbf{I} \\ &= \hat{\mathbb{H}}_l^H \hat{\mathbb{H}}_l + \boldsymbol{\Psi}_l + P^{-1} \mathbf{I}, \end{aligned} \quad (5.19)$$

where we used the Markov chain $\mathbb{H}_l \rightarrow \hat{\mathbb{H}}_l \rightarrow \hat{\mathbb{H}}_l$ and Assumption 5.1.

Then, for $j \neq l$:

$$\mathbf{E}[\mathbb{O}_{l,j} \mathfrak{t}_j^*(S_j) | \hat{\mathbb{H}}_l] = \mathbf{E}[\mathbb{H}_l^H \mathbb{H}_j \mathfrak{t}_{j,k}^* | S_l] \quad (5.20)$$

$$= \mathbf{E} \left[\mathbf{E}[\mathbb{H}_l^H \mathbb{H}_j | S_l, S_j] \mathfrak{t}_{j,k}^* \middle| S_l \right] \quad (5.21)$$

$$= \mathbf{E} \left[\mathbf{E}[\mathbb{H}_l^H | S_l, S_j] \mathbf{E}[\mathbb{H}_j | \hat{\mathbb{H}}_l, \hat{\mathbb{H}}_j] \mathfrak{t}_{j,k}^* \middle| S_l \right] \quad (5.22)$$

$$= \mathbf{E} \left[\mathbf{E}[\mathbb{H}_l^H | \hat{\mathbb{H}}_l] \mathbf{E}[\mathbb{H}_j | \hat{\mathbb{H}}_j] \mathfrak{t}_{j,k}^* \middle| S_l \right] \quad (5.23)$$

$$= \hat{\mathbb{H}}_l^H \mathbf{E}[\hat{\mathbb{H}}_j \mathfrak{t}_{j,k}^* | S_l], \quad (5.24)$$

where (5.21) follows from the law of total expectation, (5.22) from the Markov chain $\mathbb{H}_l \rightarrow (S_l, S_j) \rightarrow \mathbb{H}_j$, (5.23) from the Markov chain $\mathbb{H}_l \rightarrow \hat{\mathbb{H}}_l \rightarrow (S_l, S_j)$, and (5.24) from Assumption 5.1. Note that all the aforementioned Markov chains are implied by Assumption 5.2. The proof is concluded by letting $\mathbb{E}[\mathbf{g}_{l,k}|S_l] = \mathbb{E}[\mathbb{H}_l|\hat{\mathbb{H}}_l]\mathbf{e}_k = \hat{\mathbb{H}}_l\mathbf{e}_k$ and by rearranging the terms. \square

The above lemma reveals the following interpretation of the optimal TMMSE solution: the matrix \mathbb{T}_l can be recognized as a *local* MMSE precoding stage (studied, e.g., in [28]), that is, a centralized MMSE solution [2] assuming that there are no other TXs than TX l ; the remaining part can be then interpreted as a ‘corrective’ stage which takes into account the effect of the other TXs based on the available CSIT and long-term statistical information.

5.3.1 No CSIT sharing

As an important example, we assume that no local channel measurement is shared along the fronthaul. This corresponds to the original cell-free massive MIMO setup studied in [15], which is relevant for any fronthaul architecture. Specifically, $\forall l \in \mathcal{L}$, we let $\hat{\mathbb{H}}_l$ as in Assumption 5.1 and

$$S_l = \hat{\mathbb{H}}_l, \quad \forall l \in \mathcal{L}. \quad (5.25)$$

Theorem 5.4. *The TMMSE precoders solving (5.18) under no CSIT sharing (5.25) are given by*

$$\mathbf{t}_{l,k}^*(S_l) = \mathbb{T}_l \mathbf{C}_l \mathbf{e}_k, \quad \forall l \in \mathcal{L}, \quad (5.26)$$

for some matrices of coefficients $\mathbf{C}_l \in \mathbb{C}^{K \times K}$. Furthermore, the optimal \mathbf{C}_l are given by the unique solution of the linear system $\mathbf{C}_l + \sum_{j \neq l} \mathbf{\Pi}_j \mathbf{C}_j = \mathbf{I}$, $\forall l \in \mathcal{L}$, where $\mathbf{\Pi}_l := \mathbb{E}[\hat{\mathbb{H}}_l \mathbb{T}_l]$.

Proof. Substituting (5.26) into the optimality conditions (5.18), we need to show that

$$\hat{\mathbb{H}}_l^H \left(\mathbf{C}_l + \sum_{j \neq l} \mathbb{E}[\hat{\mathbb{H}}_j \mathbb{T}_j \mathbf{C}_j | S_l] - \mathbf{I} \right) \mathbf{e}_k = \mathbf{0}_N \quad \text{a.s.}, \quad \forall l \in \mathcal{L}. \quad (5.27)$$

By the independence between $\hat{\mathbb{H}}_l$ and $\hat{\mathbb{H}}_j$, we can drop the conditioning on S_l and obtain $\hat{\mathbb{H}}_l^H \left(\mathbf{C}_l + \sum_{j \neq l} \mathbf{\Pi}_j \mathbf{C}_j - \mathbf{I} \right) \mathbf{e}_k = \mathbf{0}$ a.s., $\forall l \in \mathcal{L}$. The proof is concluded by observing that $\mathbf{C}_l + \sum_{j \neq l} \mathbf{\Pi}_j \mathbf{C}_j = \mathbf{I}$, $\forall l \in \mathcal{L}$, always has a unique solution, as shown in Appendix B.3.4. \square

The optimal solution (5.26) corresponds to a two-stage precoding scheme composed by a local MMSE precoding stage \mathbb{T}_l preceded by a statistical precoding stage \mathbf{C}_l . By letting the rows $\hat{\mathbf{g}}_{l,k}^H$ of $\hat{\mathbb{H}}_l$ to be independent and distributed as $\mathcal{CN}(\mathbf{0}, \mathbf{K}_{l,k})$, corresponding for instance to a non-line-of-sight (NLoS) scenario with no pilot contamination [28], it can be shown that the matrices $\mathbf{\Pi}_l$ are diagonal. Hence, (5.26) takes the simpler form

$$\mathfrak{t}_{l,k}^*(S_l) = c_{l,k} \mathbb{T}_l \mathbf{e}_k, \quad (5.28)$$

which, by mapping the optimal $c_{l,k}$ to the optimal large-scale fading decoding coefficients in a dual UL channel, was already studied in [63]. However, if the channels have non-zero mean, such as in line-of-sight (LoS) models, (5.26) may provide significantly higher rates than (5.28). To see this, let $\hat{\mathbb{H}}_l \approx \bar{\mathbb{H}}_l$ for some fixed matrix $\bar{\mathbb{H}}_l, \forall l \in \mathcal{L}$. Then, since $\bar{\mathbb{H}}_l$ is statistical information known to all TXs, the TMMSE precoders should take a form similar to a ‘long-term’ centralized MMSE solution, which cannot be implemented using (5.28). Finally, we point out that a suboptimal variation of (5.26) called *optimal bilinear equalizer* (OBE), with \mathbb{T}_l replaced by $\hat{\mathbb{H}}_l^H$, was already proposed in [64] as a low-complexity alternative to centralized MMSE precoding which maintains robustness against pilot contamination.

5.3.2 Unidirectional CSIT sharing

We now consider a more involved example and let the local channel measurements be shared unidirectionally along a serial fronthaul. This setup is relevant, e.g., for the cell-free massive MIMO network in Figure 5.1, where CSIT, messages, and power are distributed along a serial fronthaul from and/or towards a CPU located at one edge, an architecture also known as a *radio stripe* [27, 36].

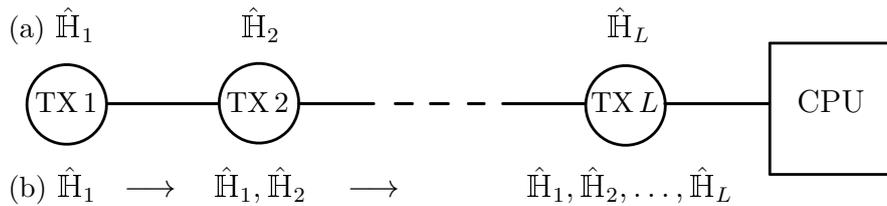


Figure 5.1: Pictorial representation of a radio stripe with (a) no CSIT sharing, and (b) unidirectional CSIT sharing.

Specifically, $\forall l \in \mathcal{L}$, we let $\hat{\mathbb{H}}_l$ as in Assumption 5.1 and

$$S_l = (\hat{\mathbb{H}}_1, \dots, \hat{\mathbb{H}}_l), \quad \forall l \in \mathcal{L}. \quad (5.29)$$

This particular information structure can be interpreted as the CSIT which is accumulated at every TX during the first phase of a centralized precoding scheme for radio stripes, where the CPU collects the $K \times LN$ channel matrix \mathbb{H} through the serial fronthaul.

Theorem 5.5. *The TMMSE precoders solving (5.18) under unidirectional CSIT sharing (5.29) are given by*

$$\mathbf{t}_{l,k}^*(S_l) = \mathbb{T}_l \mathbb{V}_l \left[\prod_{i=1}^{l-1} \bar{\mathbb{V}}_i \right] \mathbf{e}_k, \quad \forall l \in \mathcal{L}, \quad (5.30)$$

where we use the following short-hands:

- $\mathbb{V}_l := (\mathbf{I} - \mathbf{\Pi}_l \mathbb{P}_l)^{-1} (\mathbf{I} - \mathbf{\Pi}_l)$;
- $\bar{\mathbb{V}}_l := \mathbf{I} - \mathbb{P}_l \mathbb{V}_l$;
- $\mathbb{P}_l := \hat{\mathbb{H}}_l \mathbb{T}_l$;
- $\mathbf{\Pi}_l := \mathbb{E}[\mathbb{P}_{l+1} \mathbb{V}_{l+1}] + \mathbf{\Pi}_{l+1} \mathbb{E}[\bar{\mathbb{V}}_{l+1}]$, $\mathbf{\Pi}_L := \mathbf{0}_{K \times K}$.

Proof. We first assume that all the matrix inverses involved in the following steps exist. Substituting (5.30) into (5.18), we need to show that

$$\hat{\mathbb{H}}_l^H \left(\mathbb{V}_l \prod_{i=1}^{l-1} \bar{\mathbb{V}}_i + \sum_{j \neq l} \mathbb{E} \left[\mathbb{P}_j \mathbb{V}_j \prod_{i=1}^{j-1} \bar{\mathbb{V}}_i \middle| S_l \right] - \mathbf{I} \right) \mathbf{e}_k = \mathbf{0} \quad \text{a.s.}, \quad \forall l \in \mathcal{L}. \quad (5.31)$$

To verify the above statement, we rewrite the first two terms inside the outer brackets as:

$$\left(\mathbb{V}_l + \sum_{j>l} \mathbb{E} \left[\mathbb{P}_j \mathbb{V}_j \prod_{i=l+1}^{j-1} \bar{\mathbb{V}}_i \right] \bar{\mathbb{V}}_l \right) \prod_{i=1}^{l-1} \bar{\mathbb{V}}_i + \sum_{j<l} \mathbb{P}_j \mathbb{V}_j \prod_{i=1}^{j-1} \bar{\mathbb{V}}_i, \quad (5.32)$$

where we use the fact that \mathbb{P}_j , \mathbb{V}_j , and $\bar{\mathbb{V}}_j$ are deterministic functions of $\hat{\mathbb{H}}_j$ only, hence they are independent from S_l for $j > l$, while they are deterministic functions of S_l otherwise. Furthermore, since \mathbb{P}_j , \mathbb{V}_j , and $\bar{\mathbb{V}}_j$ are independent from \mathbb{P}_i , \mathbb{V}_i ,

and $\bar{\mathbf{V}}_i \forall i \neq j$, we have

$$\begin{aligned}
\sum_{j>l} \mathbf{E} \left[\mathbb{P}_j \mathbb{V}_j \prod_{i=l+1}^{j-1} \bar{\mathbf{V}}_i \right] &= \sum_{j>l} \mathbf{E} [\mathbb{P}_j \mathbb{V}_j] \prod_{i=l+1}^{j-1} \mathbf{E} [\bar{\mathbf{V}}_i] \\
&= \mathbf{E} [\mathbb{P}_{l+1} \mathbb{V}_{l+1}] + \sum_{j>l+1} \mathbf{E} [\mathbb{P}_j \mathbb{V}_j] \prod_{i=l+1}^{j-1} \mathbf{E} [\bar{\mathbf{V}}_i] \\
&= \mathbf{E} [\mathbb{P}_{l+1} \mathbb{V}_{l+1}] + \left(\sum_{j>l+1} \mathbf{E} [\mathbb{P}_j \mathbb{V}_j] \prod_{i=l+2}^{j-1} \mathbf{E} [\bar{\mathbf{V}}_i] \right) \mathbf{E} [\bar{\mathbf{V}}_{l+1}].
\end{aligned} \tag{5.33}$$

The second and last term of the above chain of equalities define a recursion terminating with $\mathbf{E} [\mathbb{P}_L \mathbb{V}_L] + \mathbf{0} \mathbf{E} [\bar{\mathbf{V}}_L] = \mathbf{\Pi}_{L-1}$. This recursion gives precisely $\sum_{j>l} \mathbf{E} [\mathbb{P}_j \mathbb{V}_j \prod_{i=1}^{j-1} \bar{\mathbf{V}}_i] = \mathbf{\Pi}_l$. Together with the property $\mathbb{V}_l + \mathbf{\Pi}_l \bar{\mathbf{V}}_l = \mathbf{I}$, (5.32) simplifies to

$$\begin{aligned}
\prod_{i=1}^{l-1} \bar{\mathbf{V}}_i + \sum_{j<l} \mathbb{P}_j \mathbb{V}_j \prod_{i=1}^{j-1} \bar{\mathbf{V}}_i &= (\bar{\mathbf{V}}_{l-1} + \mathbb{P}_{l-1} \mathbb{V}_{l-1}) \prod_{i=1}^{l-2} \bar{\mathbf{V}}_i + \sum_{j<l-1} \mathbb{P}_j \mathbb{V}_j \prod_{i=1}^{j-1} \bar{\mathbf{V}}_i \\
&= \prod_{i=1}^{l-2} \bar{\mathbf{V}}_i + \sum_{j<l-1} \mathbb{P}_j \mathbb{V}_j \prod_{i=1}^{j-1} \bar{\mathbf{V}}_i,
\end{aligned} \tag{5.34}$$

where the last equation follows from the definition of $\bar{\mathbf{V}}_l$, and where we identify another recursive structure among the remaining terms. By continuing until termination, we finally obtain $\prod_{i=1}^{l-1} \bar{\mathbf{V}}_i + \sum_{j<l} \mathbb{P}_j \mathbb{V}_j \prod_{i=1}^{j-1} \bar{\mathbf{V}}_i = \mathbf{I}$, which proves the main statement under the assumption that all the matrix inverses involved exist. This assumption is indeed always satisfied, as shown in Appendix B.3.5. \square

By locally computing precoders based on S_l only, and at the expense of some performance loss, the scheme in (5.30) eliminates the additional overhead required by centralized precoding to share back the computed $K \times LN$ precoding matrix from the CPU to the TXs. Furthermore, inspired by the schemes proposed in [27, 63, 36] for UL processing exploiting the peculiarity of a serial fronthaul, the CSIT sharing overhead can be further reduced as follows:

Remark 5.2. *The scheme in (5.30) can be alternatively implemented via a recursive algorithm involving a $K \times K$ aggregate information matrix $\prod_{i=1}^{l-1} \bar{\mathbf{V}}_i$ which is sequentially processed and forwarded in the direction from TX 1 to TX L. Therefore, the capacity of the serial fronthaul can be made independent from L, which is typically larger than K.*

Furthermore, if message sharing is implemented through the sequential forwarding of a vector $\mathbf{u} := [U_1, \dots, U_K]^\top \in \mathbb{C}^K$ of coded and modulated I/Q symbols originating from a CPU placed next to TX 1, then this can be replaced by the forwarding of a sequentially precoded K -dimensional vector $\prod_{i=1}^{l-1} \bar{\mathbf{V}}_i \mathbf{u}$, thus eliminating the CSIT sharing overhead.

We conclude this section by providing the following corollary to Theorem 5.5.

Corollary 5.1. *An alternative expression for centralized MMSE precoding [2, 28], or equivalently, for the TMMSE solution under full CSIT sharing $S_l = (\hat{\mathbf{H}}_1, \dots, \hat{\mathbf{H}}_L)$ $\forall l \in \mathcal{L}$, is given by (5.30) with $\mathbf{\Pi}_l$ replaced by $\bar{\mathbf{P}}_l := \mathbb{P}_{l+1} \bar{\mathbf{V}}_{l+1} + \bar{\mathbf{P}}_{l+1} \bar{\mathbf{V}}_{l+1}$, $\bar{\mathbf{P}}_L := \mathbf{0}$.*

Proof. Since all random quantities become deterministic after conditioning on S_l , the proof of Theorem 5.5 can be repeated by removing $\mathbf{E}[\cdot]$ everywhere. \square

Without getting into details, we remark that the expression in Corollary 5.1 can be alternatively derived by applying recursively known block-matrix inversion lemmas to the original centralized MMSE precoding expression [2]

$$\mathfrak{t}_k^{\text{MMSE}} := \left(\hat{\mathbf{H}}^H \hat{\mathbf{H}} + \mathbf{\Psi} + P^{-1} \mathbf{I} \right)^{-1} \hat{\mathbf{H}}^H \mathbf{e}_k, \quad (5.35)$$

where $\hat{\mathbf{H}} := [\mathbb{H}_1, \dots, \mathbb{H}_L]$ and $\mathbf{\Psi} = \text{diag}(\mathbf{\Psi}_1, \dots, \mathbf{\Psi}_L)$. Similarly to the implementation of (5.30) described in Remark 5.2, Corollary 5.1 provides a novel distributed and recursive implementation of centralized MMSE precoding (5.35). The main difference is that, in contrast to $\mathbf{\Pi}_l$ which can be computed offline, the computation of $\bar{\mathbf{P}}_l$ entails another sequential procedure in the reverse direction, thus doubling the overhead.

5.3.3 Asymptotic results and competing unidirectional recursive schemes

The idea of designing recursive precoding schemes exploiting the opportunities of a serial connection between antenna elements has been also explored by [37]. Motivated by the need of reducing hardware complexity of a massive MIMO cellular base station, and by focusing on $N = 1$ and no channel estimation error, the authors of [37] propose the following so-called *SGD* precoding scheme:

$$T_{l,k}(S_l) = \mu_{l,k} \mathbf{h}_l^H \left(\mathbf{e}_k - \sum_{j=1}^{l-1} \mathbf{h}_j T_{j,k}(S_j) \right), \quad \forall l \in \mathcal{L}, \quad (5.36)$$

where S_l is given by (5.29) assuming unidirectional CSIT sharing, and $\mu_{l,k} \in \mathbb{R}$ are tunable step-sizes of a stochastic gradient descent algorithm. The choice

$\mu_{l,k} = \|\mathbb{h}_l\|^{-2}$ is motivated by [37] as a good solution for i.i.d. Rayleigh fading and high SNR. Furthermore, to cope with finite SNR, [37] suggests to take $\mu_{l,k} = \mu_k \|\mathbb{h}_l\|^{-2}$ for a single scalar $\mu_k \in \mathbb{R}$ per RX to be optimized, e.g., using line search. Interestingly, the SGD scheme with $\mu_{l,k} = \|\mathbb{h}_l\|^{-2}$ can be also derived from team theoretical arguments, as a particular case of the following asymptotic result considering $N \geq 1$:

Lemma 5.3. *Assume $\text{vec}(\mathbb{H}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, no measurement noise $\hat{\mathbb{H}}_l = \mathbb{H}_l \forall l \in \mathcal{L}$, unidirectional CSIT sharing (5.29), and let \mathfrak{t}_k^* be the optimal TMMSE solution of Problem (5.6). Then,*

$$R_k = \log(\text{MSE}_k(\mathfrak{t}_k^*))^{-1} \leq L \log\left(\frac{K}{K-N}\right), \quad (5.37)$$

with equality attained as $P \rightarrow \infty$ by

$$\mathfrak{t}_{l,k}(S_l) = (\mathbb{H}_l^H \mathbb{H}_l)^{-1} \mathbb{H}_l^H \left(\mathbf{e}_k - \sum_{j=1}^{l-1} \mathbb{H}_j \mathfrak{t}_{j,k}(S_j) \right), \quad \forall l \in \mathcal{L}. \quad (5.38)$$

Proof. The proof is given in Appendix B.3.6. □

5.4 Performance evaluation

5.4.1 Simulation setup

Inspired by the “football arena” [36] or “outdoor piazza” [27] scenarios, we simulate the network in Figure 5.2, composed by a radio stripe of $L = 30$ equally spaced TXs with $N = 2$ antennas each wrapped around a circular area of radius $r_1 = 60$ m, and $K = 7$ RXs independently and uniformly drawn within a concentric circular area of radius $r_2 = 50$ m. We let the channel coefficient $H_{l,k,n}$ between the n -th antenna of TX l and RX k be independently distributed as $H_{l,k,n} \sim \mathcal{CN}(0, \rho_{l,k}^2)$, where $\rho_{l,k}^2$ denotes the channel gain between TX l and RX k . We follow the 3GPP NLoS Urban Microcell path-loss model [65, Table B.1.2.1-1]

$$\text{PL}_{l,k} = 36.7 \log_{10}\left(\frac{\text{dist}_{l,k}}{1 \text{ m}}\right) + 22.7 + 26 \log_{10}\left(\frac{f_c}{1 \text{ GHz}}\right) \quad [\text{dB}], \quad (5.39)$$

where $f_c = 2$ GHz is the carrier frequency, and $\text{dist}_{l,k}$ is the distance between TX l and RX k including a difference in height of 10 m. We let the noise power at all RXs be given by

$$P_{\text{noise}} = -174 + 10 \log_{10}(B_w/1 \text{ Hz}) + F_{\text{noise}} \quad [\text{dBm}], \quad (5.40)$$

where $B_w = 20$ MHz is the system bandwidth, and $F_{\text{noise}} = 7$ dB is the noise figure. Finally, we let $\rho_{l,k}^2 := 10^{-\frac{\text{PL}_{l,k} + P_{\text{noise}}}{10}}$ mW⁻¹, and, leveraging the short distances, we assume a relatively low total radiated power $P_{\text{sum}} = 100$ mW.

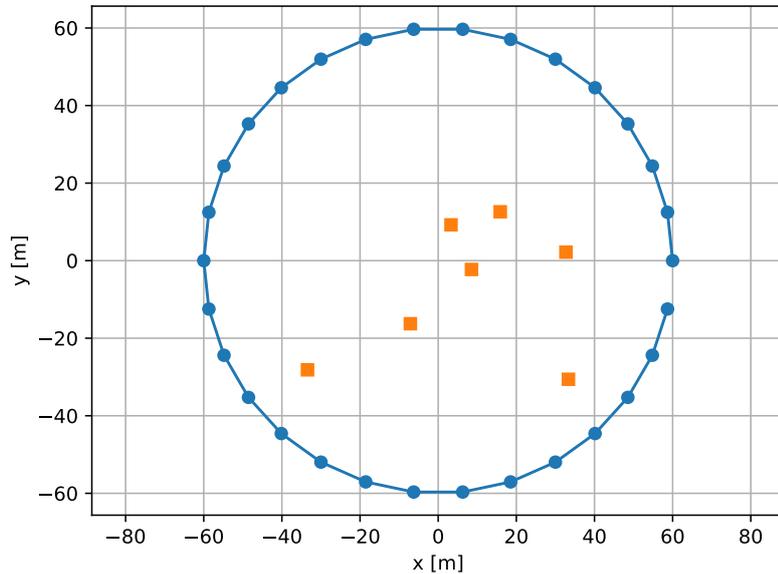


Figure 5.2: Pictorial representation of the simulation setup, with $K = 7$ RXs uniformly distributed within a circular service area, and a radio stripe of $L = 30$ TXs wrapped around.

5.4.2 Comparison among different CSIT configurations

We numerically evaluate the Pareto optimal rates $R_k = -\log(\text{MSE}_k(\mathfrak{t}_k^*))$, where \mathfrak{t}_k^* denotes the optimal solution of Problem (5.6), under the following CSIT configurations: (i) no CSIT sharing (5.25), (ii) unidirectional CSIT sharing (5.29), and (iii) full CSIT sharing as in Corollary 5.1. The resulting optimal precoding schemes are respectively termed as (i) local TMMSE, (ii) unidirectional TMMSE, and (iii) centralized MMSE. We assume for simplicity $\hat{\mathbb{H}}_l = \mathbb{H}_l$ to study the impact of the different CSIT configurations in absence of measurement noise, and focus on the Pareto optimal point parametrized by $w_k = 1 \forall k \in \mathcal{K}$.

Figure 5.4.2 reports the empirical cumulative distribution function (CDF) of R_k for multiple i.i.d. realizations of the RX locations. As expected, adding information constraints on the CSIT configuration leads to performance degradation. However, the degradation is less pronounced from centralized to unidirectional MMSE precoding, showing that unidirectional CSIT sharing does not prevent effective forms of network-wide interference management. Therefore, the unidirectional team MMSE scheme appears as a promising intermediate solution whenever centralized MMSE precoding becomes too costly, e.g., when the CSIT sharing overhead becomes problematic due to high RXs mobility. Quantifying the savings in

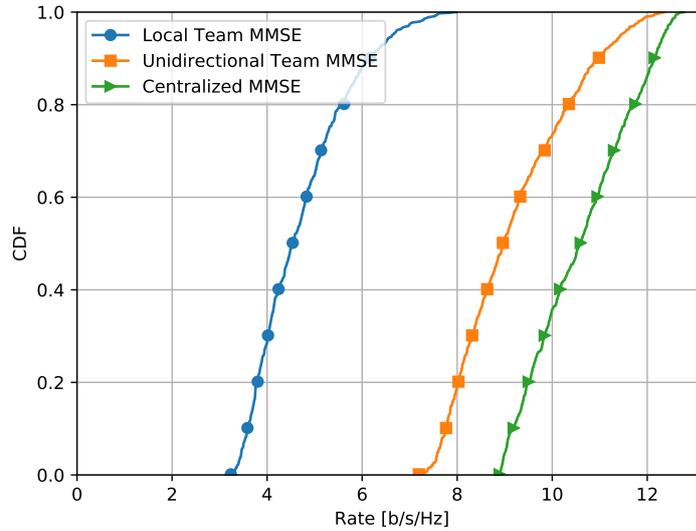
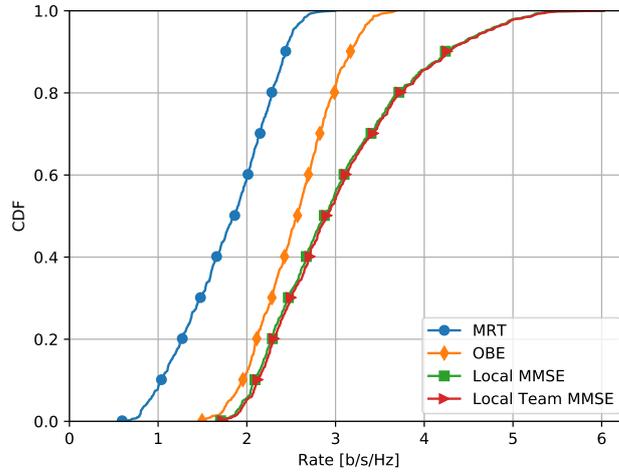


Figure 5.3: Comparison among different CSIT configuration: Empirical CDF of the optimal per-RX achievable rates. Unidirectional TMMSE is a promising intermediate solution for supporting network-wide interference management when centralized precoding becomes too costly.

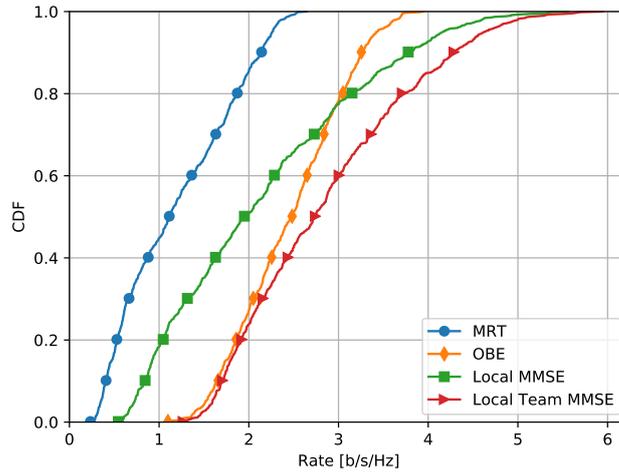
terms of CSIT sharing overhead is an interesting open problem which depends on many implementation details. For instance, if computational complexity is not an issue and the message sharing is implemented through the forwarding of high-precision I/Q symbols, unidirectional TMMSE precoding may have the same overhead as local TMMSE precoding, owing to the sequential implementation outlined in Remark 5.2. If this is not possible, then the savings may become less prominent, e.g., down to a factor 2.

5.4.3 Comparison among local precoding schemes

In this section, we compare the optimal local TMMSE solution against classical MRT, the OBE method [64], and local MMSE precoding (5.28) with optimal large-scale fading coefficients $c_{l,k}$ computed using the method in [63]. Since the bound in (5.7) may be overly pessimistic for suboptimal schemes, for a fair comparison we compute the DL rates $R_k = R_k^{\text{hard}}$ by means of their dual UL rates R_k^{UatF} as defined in the proof of Theorem 5.1, using the same dual UL power allocation $w_k = 1 \forall k \in \mathcal{K}$. One of the major weaknesses of MRT and local MMSE precoding is that they do not exploit channel mean information, typically arising from LoS components. To study this effect, we modify our simulation setup by letting $N = 1$



(a)



(b)

Figure 5.4: Empirical CDF of the per-RX achievable rates for different local precoding schemes, under Ricean factor (a) $\kappa = 0$, and (b) $\kappa = 1$. In contrast to previously known heuristics, the team MMSE approach optimally exploits statistical information such as the channel mean and always exhibits superior performance. Furthermore, consistently with our theoretical results, local MMSE precoding [63] is optimal in case (a). However, as expected, and in contrast to the team MMSE approach and the OBE method [64], it may not handle well the interference originating from the channel mean, as shown by the poor performance of the weaker RXs in case (b).

and by considering a simple Ricean fading model $H_{l,k,1} \sim \mathcal{CN}\left(\sqrt{\frac{\kappa}{\kappa+1}}\rho_{l,k}, \frac{1}{\kappa+1}\rho_{l,k}^2\right)$ for some $\kappa \geq 0$, and consider again no measurement noise $\hat{\mathbb{H}}_l = \mathbb{H}_l$. Figure 5.4 confirms the above observation: while, as expected, local MMSE precoding is optimal for a NLoS setup ($\kappa = 0$), it may incur significant performance loss w.r.t. local TMMSE precoding and the OBE method even in case of relatively weak LoS components ($\kappa = 1$).

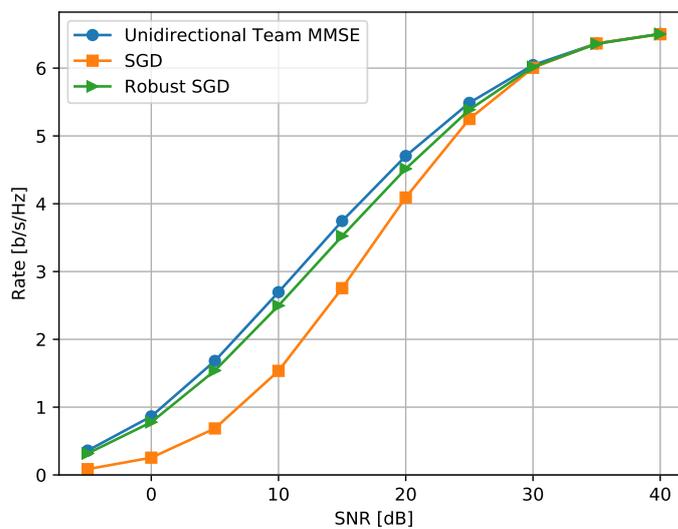
5.4.4 Comparison among unidirectional recursive schemes

In this section, we compare the unidirectional TMMSE solution (5.29) for $N = 1$ against the suboptimal SGD scheme (5.36) proposed in [37] for $\mu_k = 1$ and its robust version obtained by optimizing μ_k via line search. Figure 5.5 plots the rate $R_1 = R_1^{\text{hard}}$ of the first RX (measured via its dual UL rate as in Section 5.4.3) versus the SNR $:= P \sum_l \rho_{l,1}^2$ for a single realization of the simulation setup, and by focusing on the following aspects:

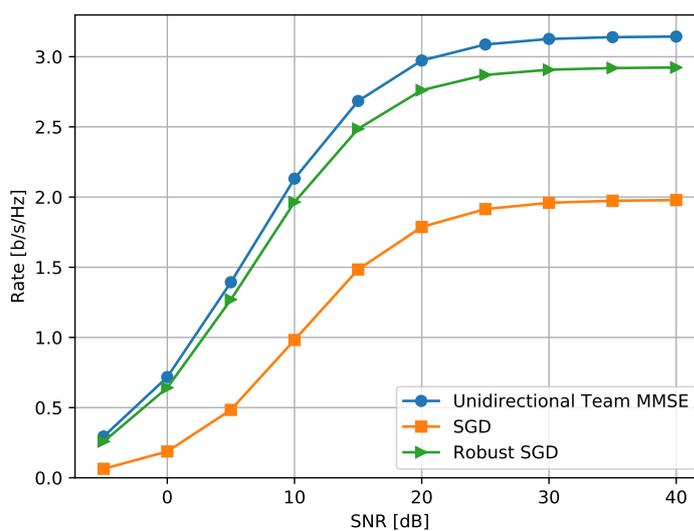
- (a) Equal path-loss, i.e., $r_2 = 0$ (all RXs colocated at the center of the circular service area), and no measurement noise, i.e., $\hat{\mathbb{H}}_l = \mathbb{H}_l$;
- (b) Equal path-loss, and measurement noise, i.e., we let $\mathbb{E}_l \sim \mathcal{CN}(\mathbf{0}, \epsilon \mathbf{R}_l)$ and $\hat{\mathbb{H}}_l \sim \mathcal{CN}(\mathbf{0}, (1 - \epsilon) \mathbf{R}_l)$, where $\mathbf{R}_l := \text{diag}(\rho_{l,1}, \dots, \rho_{l,K})$ and $\epsilon = 0.2$;
- (c) Realistic path-loss, i.e., $r_2 = 50$ m (single realization), and no measurement noise.

Although the SGD scheme assumes no measurement noise, in the above experiments we adapt (5.36) to case (b) by replacing \mathbb{h}_l with $\hat{\mathbb{H}}_l$ everywhere. As expected, from Figure 5.5a we observe that the SGD scheme is asymptotically optimal in case (a), but its performance degrades for low SNR, or in the presence of channel estimation error noise and/or realistic path-loss as shown in Figure 5.5b and 5.5c. In contrast, its robust version seems sufficient to recover most of the loss due to finite SNR and channel estimation errors. However, Figure 5.5c shows that the (robust) SGD scheme may not handle more realistic path-loss configurations.

The main advantage of the (robust) SGD scheme over optimal unidirectional TMMSE precoding is that the former does not perform any $K \times K$ matrix inversion. Furthermore, as described in [37], it can be similarly implemented by sequentially updating and forwarding a $K \times K$ matrix. Therefore, it may be considered as a low-complexity alternative to unidirectional TMMSE precoding. However, further research is needed in particular regarding the choice of the parameters $\mu_{l,k}$ and the support for $N > 1$ TX antennas.

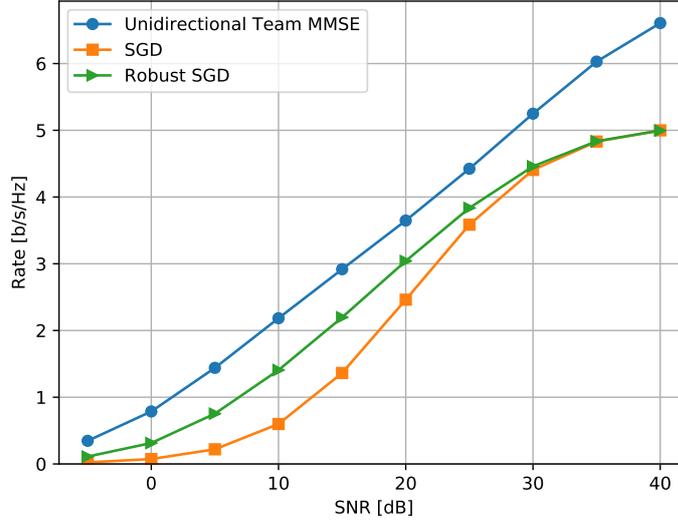


(a)



(b)

Figure 5.5: Rate vs SNR for RX 1 under: (a) equal path-loss and no channel estimation errors; (b) equal path-loss and channel estimation errors; (c) realistic path-loss and no channel estimation errors. In contrast to the (robust) SGD scheme [37], the team MMSE approach optimally exploits the path loss information and hence exhibits superior performance in case (c).



(c)

Figure 5.5: Rate vs SNR for RX 1 under: (a) equal path-loss and no channel estimation errors; (b) equal path-loss and channel estimation errors; (c) realistic path-loss and no channel estimation errors. In contrast to the (robust) SGD scheme [37], the team MMSE approach optimally exploits the path loss information and hence exhibits superior performance in case (c).

5.5 Extensions

5.5.1 Partial message sharing

In this section we discuss how TMMSE precoding, illustrated by assuming for simplicity full message sharing, can be readily applied to networks where the messages are partially shared following a user-centric clustering approach. Let $\mathcal{L}_k \subseteq \mathcal{L}$ be the subset of TXs serving a given RX k as defined in Section 2.3.2, and consider the sets of functions

$$\mathcal{T}_l^{(k)} := \begin{cases} \mathcal{T}_l & \text{if } l \in \mathcal{L}_k \\ \{\mathfrak{t}_{l,k}(S_l) = \mathbf{0} \text{ a.s.}\} & \text{otherwise} \end{cases}. \quad (5.41)$$

The set of admissible distributed precoders $\mathcal{T}^{(k)} := \prod_{l=1} \mathcal{T}_l^{(k)} \subseteq \mathcal{T}$ is then modified by forcing to zero the precoder $\mathfrak{t}_{l,k}$ if TX l does not know the message for RX k . It can be easily shown that the following user-centric variation of the TMMSE

precoding design problem

$$\underset{\mathbf{t}_k \in \mathcal{T}^{(k)}}{\text{minimize}} \text{MSE}_k(\mathbf{t}_k) := \mathbb{E} \left[\left\| \mathbf{W}^{\frac{1}{2}} \mathbb{H} \mathbf{t}_k - \mathbf{e}_k \right\|^2 + \frac{\|\mathbf{t}_k\|^2}{P} \right] \quad (5.42)$$

maintains all the properties of the original formulation (5.6). In particular, the optimality conditions (5.11) and their specialization to cell-free massive MIMO networks (5.18) hold exactly by replacing \mathcal{T} with $\mathcal{T}^{(k)}$, hence reducing to, respectively,

$$\mathbb{E}[\mathbb{O}_{l,l}|S_l] \mathbf{t}_{l,k}^*(S_l) + \sum_{j \in \mathcal{L}_k \setminus l} \mathbb{E}[\mathbb{O}_{l,j} \mathbf{t}_{j,k}^* | S_l] - \mathbb{E}[\mathbf{g}_{k,l} | S_l] = \mathbf{0}_N \quad \text{a.s.}, \quad \forall l \in \mathcal{L}_k, \quad (5.43)$$

and

$$\mathbf{t}_{l,k}^*(S_l) = \mathbb{T}_l \left(\mathbf{e}_k - \sum_{j \in \mathcal{L}_k \setminus l} \mathbb{E} \left[\hat{\mathbb{H}}_j \mathbf{t}_{j,k}^* | S_l \right] \right) \quad \text{a.s.}, \quad \forall l \in \mathcal{L}_k. \quad (5.44)$$

This can be easily seen by observing that (5.42) can be rewritten in the same form as the original problem (5.6) by replacing \mathbb{H} with a reduced channel matrix $\mathbb{H}^{(k)}$ containing only the columns of the TXs in \mathcal{L}_k , and \mathbf{t}_k with a corresponding reduced distributed precoding vector $\mathbf{t}_k^{(k)} \in \prod_{l \in \mathcal{L}_k} \mathcal{T}_l$. Furthermore, Theorem 5.1 relating TMMSE precoding with the Pareto optimal rates

$$R_k = \log(\text{MSE}_k(\mathbf{t}_k^*))^{-1}, \quad (5.45)$$

is also readily extended by replacing $\mathbf{t}_k \in \mathcal{T}$ with $\mathbf{t}_k \in \mathcal{T}^{(k)} \forall k \in \mathcal{K}$, and by modifying the definition of $\mathcal{R}^{\text{hard}}$ accordingly. This is because no part of the proof of Theorem 5.1 (and in particular, the UL-DL duality principle) is impacted by the above change.

The optimal local TMMSE precoders (5.28) and unidirectional TMMSE precoders (5.29) for an arbitrary RX k can be finally extended to user-centric partial message sharing by removing in their equations all the TXs not participating to the transmission towards RX k . For instance, the local TMMSE solution takes the form

$$\mathbf{t}_{l,k}^*(S_l) = \mathbb{T}_l \mathbf{c}_{l,k}, \quad \forall l \in \mathcal{L}_k, \quad (5.46)$$

for some vectors of coefficients $\mathbf{c}_{l,k} \in \mathbb{C}^K$ solving $\mathbf{c}_{l,k} + \sum_{j \in \mathcal{L}_k \setminus l} \mathbf{\Pi}_j \mathbf{c}_{j,k} = \mathbf{e}_k, \forall l \in \mathcal{L}_k$, where $\mathbf{\Pi}_l$ is defined as in Theorem 5.4.

Remark 5.3. *Following the user-centric clustering approach in [28] which also encompass the allocation of pilot sequences for channel estimation, it is reasonable*

to assume that each TX l knows only the channels of the RXs it is actually serving, i.e.,

$$\hat{\mathbb{H}}_l =: \begin{bmatrix} \hat{\mathbf{g}}_{1,l}^H \\ \vdots \\ \hat{\mathbf{g}}_{K,l}^H \end{bmatrix}, \quad \hat{\mathbf{g}}_{k,l} = \mathbf{E}[\mathbf{g}_{k,l}] \quad \forall k \notin \mathcal{K}_l. \quad (5.47)$$

However, the theory of this section is not restricted to this assumption.

5.5.2 Per-TX power constraint

The main drawback of the proposed TMMSE precoding method is that its optimality in terms of achievable rates have been proven only under a long-term sum-power constraint. The key ingredient of the proof is the UL-DL duality principle for fading channels given by [2], which holds only under a long-term sum power constraint.

However, for deterministic channels, a weaker form of the UL-DL duality principle and of the Pareto boundary parametrization exists also under a per-TX power constraint (see, e.g., [66, 25]). Although a formal attempt of establishing a similar result for fading channels is left for future work, inspired by [66, 25], we here postulate that an optimal distributed precoding design under a long-term per-TX power constraint $\mathbf{E}[\|\mathbf{x}_l\|^2] \leq P_l \forall l \in \mathcal{L}$ may be given by the solution of

$$\underset{\mathbf{t}_k \in \mathcal{T}}{\text{minimize}} \text{MSE}_k(\mathbf{t}_k) := \mathbf{E} \left[\left\| \mathbf{W}^{\frac{1}{2}} \mathbb{H} \mathbf{t}_k - \mathbf{e}_k \right\|^2 + \sum_{l=1}^L \lambda_l \|\mathbf{t}_{l,k}\|^2 \right], \quad (5.48)$$

for a set of L non-negative coefficients $\{\lambda_l\}_{l=1}^L$ properly penalizing the individual TX powers. If $\lambda_l > 0 \forall l \in \mathcal{L}$, the TMMSE optimality conditions and closed form solutions can be readily adapted by replacing $\frac{1}{P} \mathbf{I}$ with positive diagonal matrices $\lambda_l \mathbf{I}$. If one or more λ_l is set to zero, additional care must be taken since the quadratic form in (5.48) may not be positive definite as required by the proofs of Theorem 5.2 and Theorem 5.3; however, we note that, due to the assumption $N < K$, the obtained closed form TMMSE expressions keep on existing if we take some $\lambda_l \rightarrow 0$, hence at least stationarity is preserved. In this case, since the quadratic form is positive semidefinite, we can use weaker optimality results such as [35, Theorem 2.6.5] to conclude that stationarity implies optimality, although the converse statement and the uniqueness property do not necessarily hold.

As a much simpler yet clearly suboptimal approach, suppose \mathbf{x}_l^* is the l -th TX signal obtained using optimal TMMSE precoding under a long-term sum power constraint $\sum_{l=1}^L P_l = P_{\text{sum}}$, and consider the following trivial adaptation

$$\mathbf{x}_l = \frac{1}{\nu} \mathbf{x}_l^* \quad \forall l \in \mathcal{L}, \quad \nu^2 = \max \left(1, \frac{\mathbf{E}[\|\mathbf{x}_1^*\|^2]}{P_1}, \dots, \frac{\mathbf{E}[\|\mathbf{x}_L^*\|^2]}{P_L} \right), \quad (5.49)$$

which scales everything down until the per-TX power constraint is satisfied. In terms of performance, if $\{\mathfrak{t}_k^*\}_{k=1}^K$ is the set of TMMSE precoders with corresponding power allocation $\{p_k^*\}_{k=1}^K$ generating \mathfrak{x}_l^* , the rates achieved by using the above simple method are

$$R_k^{\text{hard}} = \log \left(1 + \frac{p_k^* |\mathbb{E}[\mathfrak{g}_k^H \mathfrak{t}_k^*]|^2}{p_k^* \text{Var}[\mathfrak{g}_k^H \mathfrak{t}_k^*] + \sum_{j \neq k} p_j^* \mathbb{E}[|\mathfrak{g}_k^H \mathfrak{t}_j^*|^2] + \nu^2} \right), \quad \forall k \in \mathcal{K}, \quad (5.50)$$

that is, compared to the optimal solution assuming a sum power constraint, there is a SNR loss proportional to the largest violation of the TX power constraints. Nevertheless, this loss may be marginal if the interference terms are still dominating the denominator in the rate expressions. Therefore, we argue that the simple power scaling factor described above may be particularly suitable for the setup of this thesis, where partial CSIT sharing indeed often prevents the interference to be driven down to the noise floor.

Other variations may be also considered, for example we can repeat the TMMSE precoders computation for decreasing values of P_{sum} until the per-TX power constraint is satisfied. Furthermore, power scaling factors coupled with clipping techniques can be also used to adapt long-term power constraints to short-term power constraints of the type $\|\mathfrak{x}_l\|^2 \leq P_l$ almost surely. We leave further discussions on power constraints for future work.

5.5.3 Performance evaluation (cont.)

In this section we illustrate the discussed extensions of TMMSE precoding to partial message sharing and per-TX power constraints on a modified version of the simulation setup used in Section 5.4. Specifically we modify the topology and consider the network in Figure 5.6, where $L = 50$ TXs with $N = 2$ antennas each are placed on a grid covering a 100×100 m² squared service area with $K = 20$ uniformly distributed RXs. Following a simple user-centric clustering rule, we let RX k being served only by the set \mathcal{L}_k of its $L_c = 10$ closest TXs. Furthermore, in the same spirit of Remark 5.3, we assume that each TX l acquires only (yet perfectly) the channels of the RXs it is serving. We then focus on the optimal user-centric local TMMSE precoders (UC-LTMMSE) in (5.46) under a sum power constraint P_{sum} , and its simple adaptation to a per-TX power constraint $P_l = P_{\text{sum}}/L \forall l \in \mathcal{L}$ using the scaling factor in (5.49). For comparison, we also consider the case with clusters size $L_c = L$, that is, by using the original LTMMSE precoders with full message sharing. The empirical CDFs of the achievable rates are reported in Figure 5.7 for two different values of P_{sum} . Not surprisingly, we observe that the use of UC-LTMMSE for limiting the message sharing burden incurs some performance loss w.r.t. to the full message sharing case. Furthermore, we observe that the

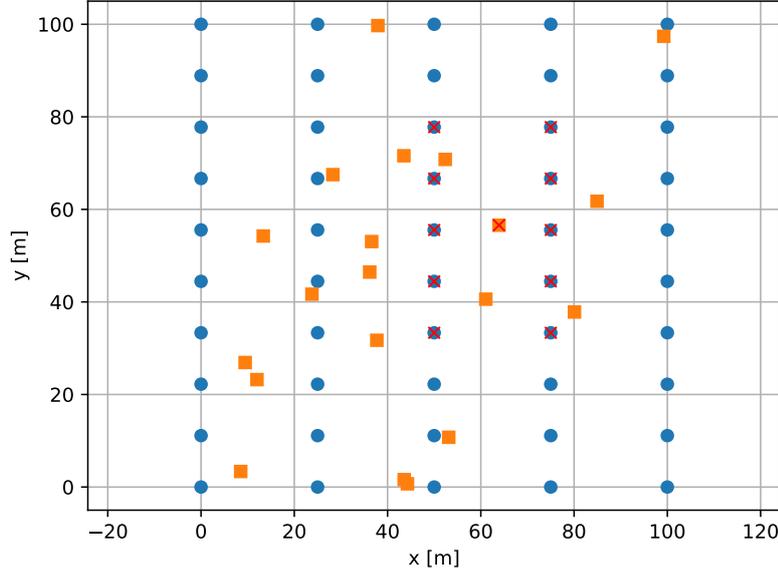
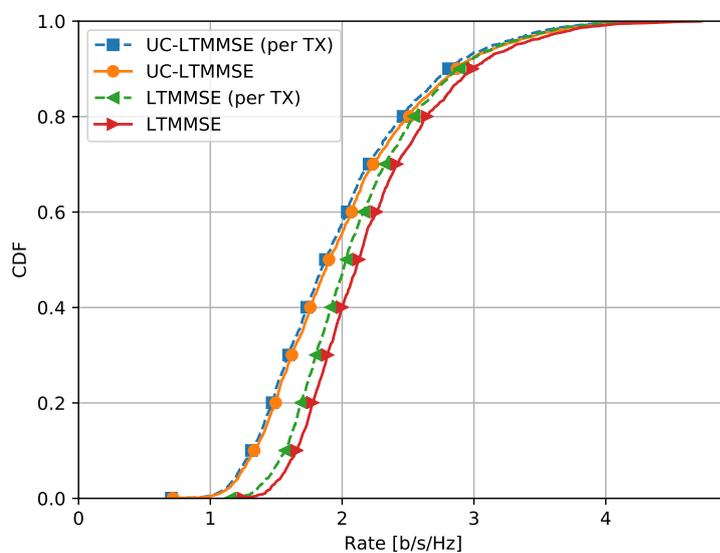


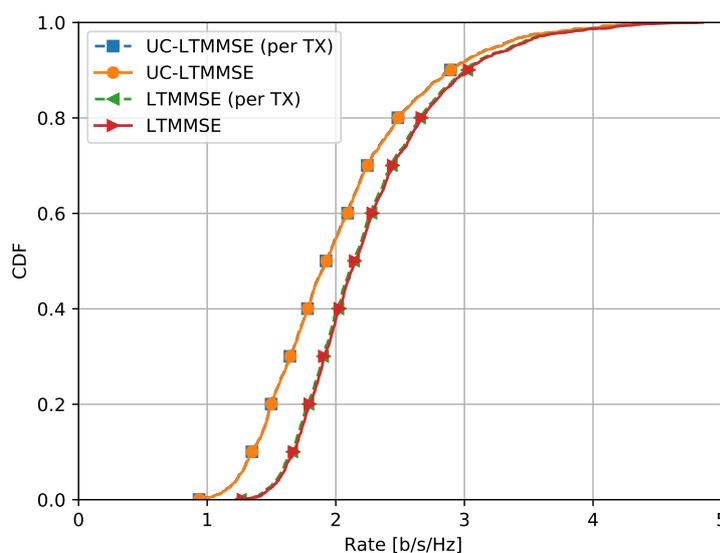
Figure 5.6: Pictorial representation of the simulation setup, with $K = 20$ RXs uniformly distributed within a squared service area, and a grid of $L = 50$ TXs. The red crosses identify the user-centric cluster of an arbitrary RX obtained from its $L_c = 10$ closest TXs.

simple suboptimal scheme satisfying the per-TX power constraint performs quite close to the sum power constraint upper bound, since the system is operating close to the interference limited regime. In the considered setup, this effect is particularly visible when using $P_{\text{sum}} = 100$ mW, but also when the interference mitigation capabilities are reduced by the partial message sharing.

Remark 5.4. *The produced plots are only for illustration purposes, since practical systems should not waste power operating in the interference limited regime as in (b). Once the interference limited regime is reached, more advanced interference management schemes such as superposition coding and/or Marton's coding (see Section 3.2.1) should be included to resolve the saturating behaviour of TIN as P_{sum} grows, but we leave further discussions on this point for future work.*



(a)



(b)

Figure 5.7: Empirical CDF of the per-RX rates achieved by optimal LTMMSE precoding, optimal UC-LTMMSE precoding, and their suboptimal adaptations to a per-TX power constraint, by assuming (a) $P_{\text{sum}} = 10$ mW, and (b) $P_{\text{sum}} = 100$ mW. As expected, the suboptimal adaptations to a per-TX power constraint become close-to-optimal in interference limited regimes such as in (b) or for the UC-LTMMSE case in (a).

Chapter 6

Conclusion

6.1 Lessons learned

The main take-away message of this thesis is that, despite being challenging, the accurate design of decentralized transmission techniques natively taking into account the distributed nature of the problem can be quite rewarding, as showcased by the remarkable performance of the TMMSE precoding method developed in Chapter 5. We believe that the results given by this thesis should motivate more researchers to further investigate this exciting topic.

From an information theoretical point of view, we have seen in Chapter 3 that when the messages are fully shared among the TXs, the major difference between decentralized and centralized transmission lies indeed in the accurate design of precoding functions mapping the codewords into channel inputs, which in the former case must act on the basis of distributed CSIT.

The results of Chapter 4 are of some theoretical importance, since they show that distributed linear precoding and Gaussian codes are information theoretically optimal at least for single-RX decentralized MIMO fading channels with CSIT taken as a quantized version of the CSIR, mirroring similar results for centralized point-to-point fading channels. However, we have observed the surprising phenomenon that, in some cases, and in contrast to centralized systems, the number of precoded data streams should not be restricted by the usual bound given by the number of antennas.

6.2 Lessons yet to be learned

How many data streams do we need? The main question left open by this thesis is whether the use of additional data streams as in Chapter 4 can significantly enhance the performance of distributed linear precoding. Essentially, the

advantage of using additional data streams is that, under distributed CSIT, the set of feasible conditional input covariances

$$\Sigma(S_1, \dots, S_L) = \mathbf{E}[\mathbf{x}\mathbf{x}^H | S_1, \dots, S_L] \quad (6.1)$$

can be larger than what can be implemented using a classical number of data streams. Although allowing to achieve its capacity, the benefits for the single-RX channel of Chapter 4 seem rather marginal, making the use of complex multi-stream decoders not worthwhile in practice. Moving to multi-RX systems, it is not clear to what extent this larger set of conditional input covariances can be exploited for interference mitigation. Preliminary negative results are reported in [C3], where the use of additional data streams is shown to provide little help in mitigating the interference over a simple $L = 2$ TXs, $K = 2$ RXs channel with asymmetric feedback. However, on top of its perhaps overly simplified setup, [C3] is based on a heuristic precoding design, and does not cover the more recent theoretical results given by [C4, J2]. Therefore, we think that further research is needed in this direction, as the available preliminary results might lead to misleading conclusions.

Evolving the TMMSE precoding framework Without going into the difficult discussion on multi-stream transmission, the TMMSE precoding framework alone seems to have a fairly big potential for evolution. We here list only a subset of the many possible future research directions:

- Find closed form solutions for more involved CSIT sharing patterns as, e.g., tree patterns, or mixtures of the available patterns.
- Explore good approximation methods, possibly data-driven, for solving the TMMSE optimality conditions in general cases.
- Revisit the performance of local precoding for ultra-dense deployments, typically characterized by more frequent LoS components which can be now better exploited. Similarly, revisit the impact of pilot contamination.
- Revisit power allocation algorithms for network utility optimization in light of the TMMSE precoding framework.
- Explore the impact of limited CSIT/message sharing in practical networks by explicitly modelling the fronthaul constraint.
- Extend to superposition coding and/or Marton's coding.

As a final important remark, note that, although developed by focusing on the DL, the proposed framework is also immediately applicable to derive optimal distributed combining schemes in the UL (in fact, it is derived using the UL-DL duality principle).

Appendix A

Team Decision Theory

This appendix recalls some basic definitions and properties from team decision theory, which is at the core of Chapter 5. Team decision theory essentially extends classical decision theory, dealing with decision making under uncertainty on the system state, to a multi-agent setup where coordinated decisions must be taken on the basis of distributed state information.

The following definitions follow the modern subject review given by [35], and can be traced back to the seminal work of Radner [34]. Among other contributions, one of the key points of [34] is the rigorous formalization of the distributed state information constraint, at the core of team decision problems, using sub- σ -algebras and measure theoretical arguments. This level of rigour enables the derivation of important structural results, such as the ones exploited and extended in Chapter 5.

A static *team* of L decision makers (DMs), where each DM is indexed by $l \in \mathcal{L} := \{1, \dots, L\}$, is here rigorously defined by the following components:

- A probability space $(\Omega, \Sigma_0, \mathbb{P})$ over which all random quantities are defined.
- A collection of Borel measurable spaces $\{(\mathcal{S}, \Sigma), (\mathcal{S}_l, \Sigma_l), (\mathcal{T}_l, \mathbb{T}_l), l \in \mathcal{L}\}$. The pair (\mathcal{S}, Σ) is called the *state space*. The pairs $(\mathcal{S}_l, \Sigma_l)$ and $(\mathcal{T}_l, \mathbb{T}_l)$ are called respectively the *observation* and *action* spaces for DM l .
- A system *state* S , where $S : (\Omega, \Sigma_0) \rightarrow (\mathcal{S}, \Sigma)$ is a measurable function mapping an element $\omega \in \Omega$ into a state realization $s := S(\omega)$.
- An *information structure* (S_1, \dots, S_L) , where $S_l : (\Omega, \Sigma_0) \rightarrow (\mathcal{S}_l, \Sigma_l)$ is a measurable function mapping an element $\omega \in \Omega$ into an information signal $s_l := S_l(\omega)$ available at DM l .
- A *team policy* $T := (T_1, \dots, T_L)$, where $T_l : (\mathcal{S}_l, \Sigma_l) \rightarrow (\mathcal{T}_l, \mathbb{T}_l)$ is a measurable function mapping the information signal $s_l \in \mathcal{S}_l$ into an action $t_l := T_l(s_l)$ performed by DM l . Let \mathcal{H}_l be the set of measurable functions $(\mathcal{S}_l, \Sigma_l) \rightarrow$

(\mathcal{T}_l, T_l) optionally satisfying some feasibility constraint, and denote by $\mathcal{H} := \prod_{l=1}^L \mathcal{H}_l$ the set of feasible team policies.

- A Borel measurable function $c : \mathcal{S} \times \prod_{l=1}^L \mathcal{T}_l \rightarrow \mathbb{R}_+$, mapping a state $s \in \mathcal{S}$ and a tuple of actions $t = (t_1, \dots, t_L) \in \prod_{l=1}^L \mathcal{T}_l$ into a cost $c(s, t)$.

Note that \mathcal{H}_l can be equivalently interpreted as the set of Σ_l -measurable functions $(\Omega, \Sigma_0) \rightarrow (\mathcal{T}_l, T_l)$ mapping an element $\omega \in \Omega$ into $t_l = T_l(S_l(\omega))$, where $\Sigma_l \subseteq \Sigma_0$ is the sub- σ -algebra generated by S_l on Ω , known as the *information subfield* of DM l .

We now define the following *team decision* problem:

$$\underset{T=(T_1, \dots, T_L) \in \mathcal{H}}{\text{minimize}} \quad J(T) := \mathbb{E}[c(S, T_1, \dots, T_L)], \quad (\text{A.1})$$

which can be interpreted as the DMs collectively optimizing a common objective $J(T)$, hence acting as a team. The most desirable solutions obviously satisfy the following definition:

Definition A.1 (Team optimal solution [35]). *A team policy $T^* \in \mathcal{H}$ is team optimal for Problem (A.1) if*

$$J(T^*) = \inf_{T \in \mathcal{T}} J(T) < \infty. \quad (\text{A.2})$$

Unfortunately, for non-trivial information constraints, team optimal solutions are typically not easy to find. When this is indeed the case, another approach is to use the following weaker notion of optimality:

Definition A.2 (Person-by-person optimal solution [35]). *A team policy $T^* := (T_1^*, \dots, T_L^*) \in \mathcal{H}$ is person-by-person optimal for Problem (A.1) if the following set of inequalities hold*

$$J(T^*) \leq J(T_{-l}^*, T_l), \quad \forall T_l \in \mathcal{H}_l, \quad \forall l \in \mathcal{L}, \quad (\text{A.3})$$

where we use the shorthand $(T_{-l}, T_l) := (T_1, \dots, T_{l-1}, T_l, T_{l+1}, \dots, T_L)$.

The conditions in (A.3) are reminiscent of the game theoretical notion of *Nash equilibrium*, with the difference that here all the DMs share the same objective. Person-by-person optimality is a weaker notion of optimality than team optimality, because the conditions in (A.3) are necessary but in general not sufficient conditions for a tuple T^* to be team optimal [34, 35].

Suppose now that no additional feasibility constraint is imposed on \mathcal{H} . By definition, if T^* is person by person optimal, then it also given by

$$T_l^*(S_l) \in \arg \min_{t_l \in \mathcal{T}_l} \mathbb{E}[c(S, T_{-l}, t_l) | S_l] \quad \text{a.s.}, \quad \forall l \in \mathcal{L}, \quad (\text{A.4})$$

corresponding to a set of optimization problems in the action spaces \mathcal{T}_l , which are generally easier to handle than optimization problems over the spaces of functions \mathcal{H}_l . This last point often motivates the development of iterative methods alternating between problems of the type in (A.4), until convergence to a person-by-person optimal solution (i.e., a local optimum).

Intuitively, a special class of problems for which (A.4) is easy to handle is the one where the cost c is convex and differentiable in each of the t_l . Following this intuition, we can define the following solution concept:

Definition A.3 (Stationary solution [35]). *A team policy $T^* := (T_1^*, \dots, T_L^*) \in \mathcal{H}$ is a stationary solution to Problem (A.1) if $J(T^*) < \infty$ and if the following set of equalities hold*

$$\nabla_{t_l} \mathbf{E} [c(S, T_{-l}^*, t_l) | S_l] \Big|_{t_l = T_l^*(S_l)} = 0 \quad \text{a.s.,} \quad \forall l \in \mathcal{L}, \quad (\text{A.5})$$

where ∇_{t_l} denotes the gradient operator (note: t_l may be multidimensional).

Under some technical conditions (convexity and differentiability are not enough), a stationary solution is also person-by-person optimal. Furthermore, under some stronger technical conditions, these two notions also lead to team optimality. Two relevant classes of problems for which the stationarity conditions (A.5) produce team optimal solutions are the class of *quadratic* teams [34], and the class of *exponentiated* teams [67].

This thesis mostly exploits selected elements from the theory of quadratic teams, which will be recalled in the main body. We refer to [35] for more details on the content of this appendix.

Appendix B

Collection of Proofs

B.1 Proofs for Chapter 3

B.1.1 Proof of Theorem 3.2

Achievability builds on Lemma 3.1, where we rewrite the mutual information terms similarly to Remark 3.2 and Remark 3.3. By focusing on the bound on the individual rate R_1 , we have

$$I(U_1; \tilde{Y}|U_2, U_0) = I(U_1; Y|U_2, U_0, S_R) \quad (\text{B.1})$$

$$= I(X_1, U_1; Y|X_2, U_2, U_0, S_R) \quad (\text{B.2})$$

$$= H(Y|X_2, U_2, U_0, S_R) - H(Y|X_1, X_2, U_1, U_2, U_0, S_R) \quad (\text{B.3})$$

$$= I(X_1; Y|X_2, U_0, S_R), \quad (\text{B.4})$$

where (B.1) follows from $I(S_R; U_1|U_2, U_0) = 0$ and the chain rule for the mutual information, (B.2) from X_l being a function of (U_l, S_R) , and (B.4) from the Markov chains $U_2 \rightarrow (X_2, U_0, S_R) \rightarrow Y$ and $(U_1, U_2) \rightarrow (X_1, X_2, U_0, S_R) \rightarrow Y$. The other rate bounds are obtained similarly. Then, by the functional representation lemma [7, Appendix B] and since U_1 and U_2 do not appear in the mutual information terms rewritten as above, designing the functions $f_l(u_l, s_l)$ and the pmf $p(u_0)p(u_1|u_0)p(u_2|u_0)$ is equivalent to designing $p(u)p(x_1|s_1, u)p(x_2|s_2, u)$, where U_0 is replaced by U .

For the converse, we define $U_i = (W_0, S_1^{i-1}, S_2^{i-1})$ and construct an outer bound by assuming that past CSIT realizations (S_1^{i-1}, S_2^{i-1}) are available at both encoders. Hence, we assume that $X_{1,i}$ and $X_{2,i}$ are functions of $(W_1, U_i, S_{1,i})$ and $(W_2, U_i, S_{2,i})$

respectively. We then bound

$$nR_1 \leq I(W_1; Y^n, S_R^n | W_0, W_2) + n\epsilon_n \quad (\text{B.5})$$

$$= I(W_1; Y^n | W_0, W_2, S_R^n) + n\epsilon_n \quad (\text{B.6})$$

$$= \sum_{i=1}^n I(W_1; Y_i | Y^{i-1}, W_0, W_2, S_R^n) + n\epsilon_n \quad (\text{B.7})$$

$$= \sum_{i=1}^n H(Y_i | W_2, X_{2,i}, U_i, Y^{i-1}, S_R^n) \quad (\text{B.8})$$

$$- H(Y_i | W_1, W_2, X_{1,i}, X_{2,i}, U_i, Y^{i-1}, S_R^n) + n\epsilon_n$$

$$= \sum_{i=1}^n H(Y_i | X_{2,i}, U_i, Y^{i-1}, S_R^n) - H(Y_i | X_{1,i}, X_{2,i}, U_i, S_{R,i}) + n\epsilon_n \quad (\text{B.9})$$

$$\leq \sum_{i=1}^n H(Y_i | X_{2,i}, S_{R,i}) - H(Y_i, S_{R,i} | X_{1,i}, X_{2,i}, U_i, S_{R,i}) + n\epsilon_n \quad (\text{B.10})$$

$$= \sum_{i=1}^n I(X_{1,i}; Y_i | X_{2,i}, U_i, S_{R,i}) + n\epsilon_n, \quad (\text{B.11})$$

where (B.5) follows from Fano's inequality ($\lim_{n \rightarrow \infty} \epsilon_n = 0$), (B.6) from the independence of W_1 and S_R^n , (B.8) from $(S_{i,1}, S_{2,i})$ being a function of S_R^n , and (B.9) from the Markov chain $(W_1, W_2, Y^{i-1}, \{S_{R,j}\}_{j \neq i}) \rightarrow (X_{1,i}, X_{2,i}, U_i, S_{R,i}) \rightarrow Y_i$. Similarly, we have

$$nR_2 \leq \sum_{i=1}^n I(X_{2,i}; Y_i | X_{1,i}, U_i, S_{R,i}) + n\epsilon_n, \quad (\text{B.12})$$

$$n(R_1 + R_2) \leq \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i | U_i, S_{R,i}) + n\epsilon_n, \quad (\text{B.13})$$

$$n(R_0 + R_1 + R_2) \leq \sum_{i=1}^n I(X_{1,i}, X_{2,i}; Y_i | S_{R,i}) + n\epsilon_n. \quad (\text{B.14})$$

The code must also satisfy the input cost constraints

$$P_1 \geq \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \eta_1(X_{1,i}) \right], \quad P_2 \geq \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \eta_2(X_{2,i}) \right]. \quad (\text{B.15})$$

We combine all the bounds by means of a time-sharing variable Q uniformly distributed in $\{1, \dots, n\}$ and independent of everything else, and by letting $U := (U_Q, Q)$, $X_1 := X_{1,Q}$, $X_2 := X_{2,Q}$, $Y := Y_Q$, $S := S_Q$, $S_1 := S_{1,Q}$, $S_2 := S_{2,Q}$,

$S_R = S_{R,Q}$. Note that the joint pmf on $(Y, X_1, X_2, S, S_1, S_2, S_R, U)$ factors as required. With these identifications, we readily obtain

$$R_1 \leq I(X_1; Y | X_2, U_0, S_R) + \epsilon_n, \quad (\text{B.16})$$

$$R_2 \leq I(X_2; Y | X_1, U_0, S_R) + \epsilon_n, \quad (\text{B.17})$$

$$R_1 + R_2 \leq I(X_1, X_2; Y | U_0, S_R) + \epsilon_n \quad (\text{B.18})$$

$$R_0 + R_1 + R_2 \leq I(X_1, X_2; Y | S_R, Q) + \epsilon_n \leq I(X_1, X_2; Y | S_R) + \epsilon_n, \quad (\text{B.19})$$

and $P_1 \geq \mathbb{E}[\eta_1(X_1)]$, $P_2 \geq \mathbb{E}[\eta_2(X_2)]$.

B.1.2 Proof of Lemma 3.2

Let us define $U_{0,i} := (W_0, S_R^{i-1})$, and $U_{l,i} := (W_l, S_l^{i-1}, U_{0,i})$ for $l = 1, 2$. Note that $X_{1,i}$ and $X_{2,i}$ are functions of $(U_{1,i}, S_{1,i})$ and $(U_{2,i}, S_{2,i})$ respectively, and, due to the Markov chain $S_{1,i} \rightarrow S_{R,i} \rightarrow S_{2,i}$, we also have $U_{1,i} \rightarrow U_{0,i} \rightarrow U_{2,i}$ as required. By Fano's inequality ($\lim_{n \rightarrow \infty} \epsilon_n = 0$), and by following similar steps as in Appendix B.1.1, we obtain

$$n(R_1 + R_2) \leq I(W_1, W_2; Y^n, S_R^n | W_0) + n\epsilon_n \quad (\text{B.20})$$

$$= I(W_1, W_2; Y^n | W_0, S_R^n) + n\epsilon_n \quad (\text{B.21})$$

$$= \sum_{i=1}^n I(W_1, W_2; Y_i | Y^{i-1}, W_0, S_R^n) + n\epsilon_n \quad (\text{B.22})$$

$$\leq \sum_{i=1}^n I(U_{1,i}, U_{2,i}; Y_i | Y^{i-1}, W_0, S_R^n) + n\epsilon_n \quad (\text{B.23})$$

$$= \sum_{i=1}^n I(U_{1,i}, U_{2,i}; Y_i | Y^{i-1}, U_{0,i}, S_R^n) + n\epsilon_n \quad (\text{B.24})$$

$$\leq \sum_{i=1}^n I(U_{1,i}, U_{2,i}; Y_i | U_{0,i}, S_{R,i}) + n\epsilon_n, \quad (\text{B.25})$$

where the last inequality comes from the memoryless property of the channel. Following similar lines one can prove

$$n(R_0 + R_1 + R_2) \leq \sum_{i=1}^n I(U_{1,i}, U_{2,i}; Y_i | S_{R,i}) + n\epsilon_n, \quad (\text{B.26})$$

which can be combined with the bound on $R_1 + R_2$ and the power constraints by means of a final time-sharing step.

B.1.3 Proof of Theorem 3.3

Achievability follows from Lemma 3.4 with $U_0 = U_2$, and by eliminating the redundant rate bounds exploiting the property $I(U_0; Y_2) \leq I(U_0; Y_1)$ implied by the degradedness assumption. For the converse, we define $U_{1,i} = (W_1, S_1^{i-1}, S_2^{i-1})$, $U_{2,i} = (W_2, Y_1^{i-1})$, and construct an upper-bound by assuming that output feedback Y_1^{i-1} as well as past CSIT realizations (S_1^{i-1}, S_2^{i-1}) are available at both encoders. Hence, we assume that $X_{1,i}$ and $X_{2,i}$ are functions of $(W_1, W_2, S_1^i, Y_1^{i-1}, S_2^{i-1}) = (U_{1,i}, U_{2,i}, S_{1,i})$ and $(W_1, W_2, S_2^i, Y_1^{i-1}, S_1^{i-1}) = (U_{1,i}, U_{2,i}, S_{2,i})$ respectively. Note that $U_{1,i}$ and $U_{2,i}$ are independent from $(S_{1,i}, S_{2,i})$.

By Fano's inequality ($\lim_{n \rightarrow \infty} \epsilon_n = 0$), we have

$$nR_1 \leq I(W_1; Y_1^n | W_2) + n\epsilon_n \quad (\text{B.27})$$

$$= \sum_{i=1}^n I(W_1; Y_{1,i} | Y_1^{i-1}, W_2) + n\epsilon_n \quad (\text{B.28})$$

$$\leq \sum_{i=1}^n I(W_1, S_1^{i-1}, S_2^{i-1}; Y_{1,i} | Y_1^{i-1}, W_2) + n\epsilon_n \quad (\text{B.29})$$

$$= \sum_{i=1}^n I(U_{1,i}; Y_{1,i} | U_{2,i}) + n\epsilon_n. \quad (\text{B.30})$$

For the second rate bound, we have

$$nR_2 \leq I(W_2; Y_2^n) + n\epsilon_n \quad (\text{B.31})$$

$$= \sum_{i=1}^n I(W_2; Y_{2,i} | Y_2^{i-1}) + n\epsilon_n \quad (\text{B.32})$$

$$\leq \sum_{i=1}^n I(W_2, Y_2^{i-1}; Y_{2,i}) + n\epsilon_n \quad (\text{B.33})$$

$$\leq \sum_{i=1}^n I(W_2, Y_2^{i-1}, Y_1^{i-1}; Y_{2,i}) + n\epsilon_n \quad (\text{B.34})$$

$$\leq \sum_{i=1}^n I(U_{2,i}, Y_2^{i-1}; Y_{2,i}) + n\epsilon_n. \quad (\text{B.35})$$

$$\leq \sum_{i=1}^n I(U_{2,i}; Y_{2,i}) + n\epsilon_n, \quad (\text{B.36})$$

where the last inequality comes from the Markov chain $Y_2^{i-1} \rightarrow (W_2, Y_1^{i-1}) \rightarrow Y_{2,i}$, which is a direct consequence of the degradedness assumption. A standard final time sharing step concludes the proof.

B.2 Proofs for Chapter 4

B.2.1 Proof of Lemma 4.2

The proof is split in the following three steps:

1. We fix a specific conditional input covariance matrix $\Sigma^*(S_1, S_2)$, and we show that it is achievable via distributed linear precoding if and only if $d > 2$.
2. We construct a specific $p(\mathbf{G}, s_1, s_2)$ such that $\Sigma^*(S_1, S_2)$ is the unique optimal solution to Problem (4.3).
3. We combine the above steps to show that there exists a channel for which $d \leq 2$ leads to strictly suboptimal rates.

Step 1: Consider binary CSIT alphabets, i.e., $\mathcal{S}_1 = \mathcal{S}_2 = \{0, 1\}$, and let $\Sigma^*(S_1, S_2)$ be given by

$$\begin{aligned} \Sigma^*(0, 0) &= \mathbf{I}, & \Sigma^*(1, 0) &= \mathbf{I}, \\ \Sigma^*(0, 1) &= \begin{bmatrix} 1 & 0.6 \\ 0.6 & 1 \end{bmatrix}, & \Sigma^*(1, 1) &= \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}. \end{aligned} \quad (\text{B.37})$$

Define the set $\mathcal{F}(D')$ of conditional input covariance matrices $\Sigma(S_1, S_2)$ which are achievable via distributed linear precoders of maximal dimension D' , i.e.,

$$\mathcal{F}(D') := \begin{cases} \Sigma(S_1, S_2) \in \mathbb{S}_+^2 & \text{s.t.} \\ \Sigma(S_1, S_2) = \begin{bmatrix} \mathbf{f}_1^H(S_1) \\ \mathbf{f}_2^H(S_2) \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(S_1) & \mathbf{f}_2(S_2) \end{bmatrix}, \\ \text{for some } \mathbf{f}_l : \mathcal{S}_l \rightarrow \mathbb{C}^d, d \leq D'. \end{cases} \quad (\text{B.38})$$

Clearly, $\mathcal{F}(D') \subseteq \mathcal{F}(D'')$, for $D' \leq D''$. The following lemma holds:

Lemma B.1. $\Sigma^*(S_1, S_2) \in \mathcal{F}(3)$, and $\Sigma^*(S_1, S_2) \notin \mathcal{F}(2)$.

Proof. For $\Sigma^*(S_1, S_2)$ to be achievable, we need to find precoders $\mathbf{f}_l(S_l)$ s.t.

$$\begin{cases} \mathbf{f}_1^H(0)\mathbf{f}_2(0) = 0, & \mathbf{f}_1^H(0)\mathbf{f}_2(1) = 0.6, \\ \mathbf{f}_1^H(1)\mathbf{f}_2(0) = 0, & \mathbf{f}_1^H(1)\mathbf{f}_2(1) = 0.8, \\ \|\mathbf{f}_1(0)\| = \|\mathbf{f}_1(1)\| = \|\mathbf{f}_2(0)\| = \|\mathbf{f}_2(1)\| = 1. \end{cases} \quad (\text{B.39})$$

For $\mathbf{f}_l(S_l)$ of dimension $d = 2$, the above system has no solution. In fact, we need to simultaneously satisfy

$$\begin{cases} \mathbf{f}_1(0) \perp \mathbf{f}_2(0), \\ \mathbf{f}_1(1) \perp \mathbf{f}_2(0), \\ \|\mathbf{f}_1(0)\| = \|\mathbf{f}_1(1)\| = 1, \end{cases} \quad (\text{B.40})$$

which, for $d = 2$, implies $\mathbf{f}_1(0) = \pm\mathbf{f}_1(1)$, and hence leads to the following contradiction $0.6 = \mathbf{f}_1^H(0)\mathbf{f}_2(1) = \pm\mathbf{f}_1^H(1)\mathbf{f}_2(1) = \pm 0.8$. Instead, one can check that $\Sigma^*(S_1, S_2)$ is readily obtained by letting $d = 3$ and

$$\begin{aligned} \mathbf{f}_1(0) &= [1 \ 0 \ 0], & \mathbf{f}_1(1) &= [0 \ 1 \ 0], \\ \mathbf{f}_2(0) &= [0 \ 0 \ 1], & \mathbf{f}_2(1) &= [0.6 \ 0.8 \ 0]. \end{aligned} \quad (\text{B.41})$$

□

Step 2: Consider the following rewriting of Problem (4.3), by letting again $\mathcal{S}_1 = \mathcal{S}_2 = \{0, 1\}$ (hence $D = 4$), and unitary power constraint $P_1 = P_2 = 1$:

$$C = \max_{\Sigma(S_1, S_2) \in \mathcal{P} \cap \mathcal{F}(4)} \mathbb{E} [\log \det (\mathbf{I} + \mathbb{G}^H \Sigma(S_1, S_2) \mathbb{G})], \quad (\text{B.42})$$

where $\mathcal{F}(4)$ is given by (B.38), and where

$$\mathcal{P} := \left\{ \Sigma(S_1, S_2) \in \mathbb{S}_+^2 \text{ s.t. } \mathbb{E} [\Sigma(S_1, S_2)_{l,l}] \leq 1, l = 1, 2 \right\} \quad (\text{B.43})$$

is the per-TX power constraint. Note that $\Sigma^*(S_1, S_2)$ belongs to the feasible set, i.e. $\Sigma^*(S_1, S_2) \in \mathcal{P} \cap \mathcal{F}(4)$.

Lemma B.2. *There exist some $p(\mathbf{S}, s_1, s_2)$ such that $\Sigma^*(S_1, S_2)$ given by (B.37) is the unique optimal solution for problem (B.42).*

Proof. The main idea is to build such CSI distribution by “reversing” a spatio-temporal water-filling algorithm which gives as unique optimal solution the conditional input covariance $\Sigma^*(S_1, S_2)$. We now provide the details.

Define a uniformly distributed random state \mathbb{G} taking values in the finite alphabet $\mathcal{S} = \{\mathbf{G}_1, \mathbf{G}_2, \mathbf{G}_3, \mathbf{G}_4\}$, and the CSIT be given by the functions

$$s_1 = q_1(\mathbf{G}) = \begin{cases} 0 & \text{for } \mathbf{G} \in \{\mathbf{G}_1, \mathbf{G}_2\} \\ 1 & \text{otherwise} \end{cases} \quad (\text{B.44})$$

$$s_2 = q_2(\mathbf{G}) = \begin{cases} 0 & \text{for } \mathbf{G} \in \{\mathbf{G}_1, \mathbf{G}_3\} \\ 1 & \text{otherwise} \end{cases} \quad (\text{B.45})$$

The capacity of such a channel can be upper bounded by

$$C = \max_{\Sigma(S_1, S_2) \in \mathcal{P} \cap \mathcal{F}(4)} \mathbb{E} [\log \det (\mathbf{I} + \mathbb{G}^H \Sigma(S_1, S_2) \mathbb{G})], \quad (\text{B.46})$$

$$\leq \max_{\Sigma(S_1, S_2) \in \mathcal{P}} \mathbb{E} [\log \det (\mathbf{I} + \mathbb{G}^H \Sigma(S_1, S_2) \mathbb{G})], \quad (\text{B.47})$$

$$\leq \max_{\Sigma(S_1, S_2) \in \mathcal{P}'} \mathbb{E} [\log \det (\mathbf{I} + \mathbb{G}^H \Sigma(S_1, S_2) \mathbb{G})], \quad (\text{B.48})$$

where

$$\mathcal{P}' := \{\boldsymbol{\Sigma}(S_1, S_2) \in \mathfrak{S}_+^2 \mid \text{tr}\{\mathbf{E}[\boldsymbol{\Sigma}(S_1, S_2)]\} \leq 2\}, \quad (\text{B.49})$$

is the set obtained by relaxing the per-TX power constraint \mathcal{P} to a total power constraint ($\mathcal{P} \subseteq \mathcal{P}'$). Inequalities (B.47) and (B.48) are obtained respectively by relaxing the achievability via distributed linear precoding and the power constraint.

Problem (B.48) turns out to be an instance of a classical (centralized) MIMO capacity problem, where the optimal solution is given by the well-known spatio-temporal water-filling algorithm. More precisely, let us rewrite (B.48) as

$$\begin{aligned} & \max_{\boldsymbol{\Sigma}(S_1, S_2) \in \mathcal{P}'} \sum_{i=1}^4 p(\mathbf{G}_i) \log \det (\mathbf{I} + \mathbf{G}_i^H \boldsymbol{\Sigma}(q_1(\mathbf{G}_i), q_2(\mathbf{G}_i)) \mathbf{G}_i) \\ &= \max_{\{\boldsymbol{\Sigma}_i\} \in \tilde{\mathcal{P}}'} \frac{1}{4} \sum_{i=1}^4 \log \det (\mathbf{I} + \mathbf{G}_i^H \boldsymbol{\Sigma}_i \mathbf{G}_i), \end{aligned} \quad (\text{B.50})$$

where we defined $\boldsymbol{\Sigma}_i := \boldsymbol{\Sigma}(q_1(\mathbf{G}_i), q_2(\mathbf{G}_i))$, and where

$$\tilde{\mathcal{P}}' = \left\{ \boldsymbol{\Sigma}_i \in \mathfrak{S}_+^2 \mid \frac{1}{4} \sum_{i=1}^4 \text{tr}\{\boldsymbol{\Sigma}_i\} \leq 2 \right\}. \quad (\text{B.51})$$

A well-known application of the Hadamard's inequality gives the following upper bound in terms of the channel eigen-decompositions $\mathbf{G}_i \mathbf{G}_i^H = \mathbf{V}_i \boldsymbol{\Lambda}_i \mathbf{V}_i^H$, $\boldsymbol{\Lambda}_i = \text{diag}(\lambda_{i,1}, \lambda_{i,2})$

$$\max_{\{\boldsymbol{\Sigma}_i\} \in \tilde{\mathcal{P}}'} \frac{1}{4} \sum_{i=1}^4 \log \det (\mathbf{I} + \mathbf{G}_i^H \boldsymbol{\Sigma}_i \mathbf{G}_i) \leq \max_{\substack{\xi_{i,l} \geq 0 \\ \frac{1}{4} \sum_{i,l} \xi_{i,l} \leq 2}} \frac{1}{4} \sum_{i=1}^4 \sum_{l=1}^2 \log(1 + \lambda_{i,l} \xi_{i,l}), \quad (\text{B.52})$$

where the optimal $\xi_{i,l}$ are given by the water-filling conditions

$$\xi_{i,l} = \max \left\{ \nu - \frac{1}{\lambda_{i,l}}, 0 \right\}, \quad i = 1, \dots, 4, \quad l = 1, 2, \quad (\text{B.53})$$

$$\sum_{i,l} \max \left\{ \nu - \frac{1}{\lambda_{i,l}}, 0 \right\} = 8, \quad (\text{B.54})$$

and where equality is achieved for

$$\boldsymbol{\Sigma}_i = \mathbf{V}_i \boldsymbol{\Xi}_i \mathbf{V}_i^H, \quad \boldsymbol{\Xi}_i = \text{diag}(\xi_{i,1}, \xi_{i,2}), \quad i = 1, \dots, 4. \quad (\text{B.55})$$

Consider the conditional covariance $\boldsymbol{\Sigma}^*(S_1, S_2)$ given by (B.37). Note that $\boldsymbol{\Sigma}^*(S_1, S_2) \in \mathcal{P}'$, i.e., it satisfies the total power constraint. We wish to construct

$\mathcal{S} = \{\mathbf{G}_i\}$ such that $\boldsymbol{\Sigma}^*(S_1, S_2)$ is the unique optimal solution for (B.48). This can be done by “reversing” the MIMO water-filling algorithm described above. More precisely, let us consider $\boldsymbol{\Sigma}_i^* := \boldsymbol{\Sigma}^*(q_1(\mathbf{G}_i), q_2(\mathbf{G}_i))$ and their eigen-decompositions

$$\boldsymbol{\Sigma}_i^* = \mathbf{V}_i^* \boldsymbol{\Xi}_i^* \mathbf{V}_i^{*\text{H}}, \quad \boldsymbol{\Xi}_i^* = \text{diag}(\xi_{i,1}^*, \xi_{i,2}^*). \quad (\text{B.56})$$

We construct now \mathcal{S} by letting

$$\mathbf{G}_i = (\mathbf{V}_i^* \boldsymbol{\Lambda}_i^* \mathbf{V}_i^{*\text{H}})^{\frac{1}{2}}, \quad i = 1, \dots, 4 \quad (\text{B.57})$$

where the eigenvalues $\boldsymbol{\Lambda}_i^* = \text{diag}(\lambda_{i,1}^*, \lambda_{i,2}^*)$ are given by

$$\lambda_{i,l}^* = \frac{1}{\nu^* - \xi_{i,l}^*}, \quad i = 1, \dots, 4, \quad l = 1, 2, \quad (\text{B.58})$$

and any choice of $\infty > \nu^* > \max_{i,l} \xi_{i,l}^* = 1.8$.

By construction, $\boldsymbol{\Sigma}^*(S_1, S_2)$ is an optimal solution for (B.48). Uniqueness of the solution can be proven by contradiction as in [68, Section III.A], or directly by the strict concavity of $\sum_{i=1}^4 \log \det(\mathbf{I} + \mathbf{G}_i^{\text{H}} \boldsymbol{\Sigma}_i \mathbf{G}_i)$ in $\{\boldsymbol{\Sigma}_i \succeq \mathbf{0}\}$, which is a direct consequence of the strict concavity of $\log \det(\mathbf{A})$ in $\mathbf{A} \succ \mathbf{0}$ and of the positive definiteness of \mathbf{G}_i by construction. Finally, since $\boldsymbol{\Sigma}^*(S_1, S_2) \in \mathcal{P} \cap \mathcal{F}(4)$, (B.48) and (B.47) are satisfied with equality. \square

Step 3: The proof is now concluded by combining Lemma B.1 and Lemma B.2, yielding $\exists p(\mathbf{G}, s_1, s_2)$ such that

$$\arg \max_{\boldsymbol{\Sigma}(S_1, S_2) \in \mathcal{P} \cap \mathcal{F}(4)} \mathbb{E} [\log \det(\mathbf{I} + \mathbf{G}^{\text{H}} \boldsymbol{\Sigma}(S_1, S_2) \mathbf{G})] \stackrel{(a)}{=} \{\boldsymbol{\Sigma}^*(S_1, S_2)\} \stackrel{(b)}{\notin} \mathcal{F}(2), \quad (\text{B.59})$$

where (a) follows from Lemma B.2, and (b) from Lemma B.1, which implies that $\exists p(\mathbf{G}, s_1, s_2)$ such that

$$\begin{aligned} & \max_{\boldsymbol{\Sigma}(S_1, S_2) \in \mathcal{P} \cap \mathcal{F}(4)} \mathbb{E} [\log \det(\mathbf{I} + \mathbf{G}^{\text{H}} \boldsymbol{\Sigma}(S_1, S_2) \mathbf{G})] \\ & > \max_{\boldsymbol{\Sigma}(S_1, S_2) \in \mathcal{P} \cap \mathcal{F}(2)} \mathbb{E} [\log \det(\mathbf{I} + \mathbf{G}^{\text{H}} \boldsymbol{\Sigma}(S_1, S_2) \mathbf{G})]. \end{aligned} \quad (\text{B.60})$$

B.2.2 Proof of Theorem 4.2

We construct an outer bound $\mathcal{C}_o^{\text{MAC}}(P_1, P_2)$ by following similar steps as in [69], but starting from the single-letter formulation of $\mathcal{C}^{\text{MAC}}(P_1, P_2)$ given by Theorem 3.2 extended to continuous alphabets similarly to [7, 41, 32]. We consider the following

applications of the *maximum differential entropy lemma* [7, p. 21], adapted to the complex field. From the tightest bound in [7, Eq. 2.6], we have

$$h(\mathbf{y}|\mathbb{G} = \mathbf{G}) \leq \log \left((\pi e)^2 \det \left(\mathbf{E} [\mathbf{y}\mathbf{y}^H | \mathbb{G} = \mathbf{G}] \right) \right) \quad (\text{B.61})$$

$$= \log \left((\pi e)^2 \det \left(\mathbf{G}^H \boldsymbol{\Sigma} (q_1(\mathbf{G}), q_2(\mathbf{G})) \mathbf{G} + \mathbf{I} \right) \right), \quad (\text{B.62})$$

where $\boldsymbol{\Sigma}(S_1, S_2) = \mathbf{E} [\mathbf{x}\mathbf{x}^H | S_1, S_2]$, $\mathbf{x} := \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$. Then, from the bound in [7, Eq. 2.7], we have

$$h(\mathbf{y}|U, \mathbb{G} = \mathbf{G}) \leq \log \left((\pi e)^2 \det \left(\mathbf{E} \left[(\mathbf{y} - \mathbf{E}[\mathbf{y}|U, \mathbb{G} = \mathbf{G}]) \right. \right. \right. \\ \left. \left. \left. \times (\mathbf{y} - \mathbf{E}[\mathbf{y}|U, \mathbb{G} = \mathbf{G}])^H | \mathbb{G} = \mathbf{G} \right] \right) \right) \quad (\text{B.63})$$

$$= \log (\pi e)^2 \det \left(\mathbf{G}^H \boldsymbol{\Gamma} (q_1(\mathbf{G}), q_2(\mathbf{G})) \mathbf{G} + \mathbf{I} \right), \quad (\text{B.64})$$

where

$$\boldsymbol{\Gamma}(S_1, S_2) = \boldsymbol{\Sigma}(S_1, S_2) - \mathbf{E} \left[\mathbf{E}[\mathbf{x}|U, S_1, S_2] \mathbf{E}[\mathbf{x}^H | U, S_1, S_2] | S_1, S_2 \right]. \quad (\text{B.65})$$

By the structure of the input distribution, we now observe that $\boldsymbol{\Gamma}(S_1, S_2) = \text{diag}(\gamma_1(S_1), \gamma_2(S_2))$ and that, $\forall (s_1, s_2) \in \mathcal{S}_1 \times \mathcal{S}_2$,

$$\boldsymbol{\Sigma}(s_1, s_2) = \begin{bmatrix} \gamma_1(s_1) & 0 \\ 0 & \gamma_2(s_2) \end{bmatrix} + \begin{bmatrix} \mathbf{E}[|\mu_1(U, s_1)|^2] & \mathbf{E}[\mu_1(U, s_1)\mu_2^*(U, s_2)] \\ \mathbf{E}[\mu_2(U, s_2)\mu_1^*(U, s_1)] & \mathbf{E}[|\mu_2(U, s_2)|^2] \end{bmatrix}, \quad (\text{B.66})$$

where we define the functions

$$\mu_l(U, S_l) := \mathbf{E}[X_l | U, S_l]$$

$$\gamma_l(S_l) := \mathbf{E} [|X_l|^2 | S_l] - \mathbf{E} [|\mu_l(U, S_l)|^2 | S_l] \geq 0.$$

By following similar steps for $h(\mathbf{y}|X_1, U, \mathbb{G} = \mathbf{G})$ and $h(\mathbf{y}|X_2, U, \mathbb{G} = \mathbf{G})$, and by applying the resulting bounds to the mutual information terms in Theorem 3.2, we obtain

$$R_1 \leq \mathbf{E} \left[\log \left(1 + \gamma_1(S_1) \|\mathbf{g}_1\|^2 \right) \right], \quad (\text{B.67})$$

$$R_2 \leq \mathbf{E} \left[\log \left(1 + \gamma_2(S_2) \|\mathbf{g}_2\|^2 \right) \right], \quad (\text{B.68})$$

$$R_1 + R_2 \leq \mathbf{E} \left[\log \det \left(\mathbf{I} + \mathbf{G}^H \text{diag}(\gamma_1(S_1), \gamma_2(S_2)) \mathbf{G} \right) \right], \quad (\text{B.69})$$

$$R_0 + R_1 + R_2 \leq \mathbf{E} \left[\log \det \left(\mathbf{I} + \mathbf{G}^H \boldsymbol{\Sigma}(S_1, S_2) \mathbf{G} \right) \right]. \quad (\text{B.70})$$

The outer bound $\mathcal{C}_o^{\text{MAC}}(P_1, P_2)$ is then established by taking the convex hull of the union of all rate triples (R_0, R_1, R_2) satisfying (B.67), (B.68), (B.69), and (B.70) for some $p(x_1|s_1, u)p(x_2|s_2, u)p(u)$ such that $\mathbf{E}[|X_l|^2] \leq P_l$.

Similarly to the proof of Theorem 4.1, it can be now shown that every $\Sigma(S_1, S_2)$ as in (B.66) induced by any $p(x_1|s_1, u)p(x_2|s_2, u)p(u)$ can also be obtained by the scheme in (4.33), i.e., via superposition of linearly precoded Gaussian codes. The key point is showing that the second term in the right-hand side of (B.66) can be obtained via distributed linear precoders of dimension D . This follows by the same technique used in the proof of Theorem 4.1, by simply replacing the functions $f_l(u, s_l)$ with $\mu_l(u, s_l)$. Finally, since the inputs are conditionally Gaussian, standard arguments [7] show that (4.33) attains $\mathcal{C}_o^{\text{MAC}}(P_1, P_2)$, without time sharing (i.e., we can omit the convex hull).

B.3 Proofs for Chapter 5

B.3.1 Proof of Theorem 5.1

Consider a dual UL network with K single-antenna TXs and L cooperating RXs equipped with N antennas each, governed by the MIMO fading channel law $\mathbf{y} = \sum_{k=1}^K \mathbf{g}_k X_k + \mathbf{n}$, where $\mathbf{y} \in \mathbb{C}^{LN}$ is the received signal at all RXs, $[\mathbf{g}_1, \dots, \mathbf{g}_K] = \mathbb{H}^{\text{H}}$ is the dual channel matrix, $X_k \in \mathbb{C}$ is the k -th TX signal, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. We let $X_k = \sqrt{P w_k} U_k$, where $U_k \sim \mathcal{CN}(0, 1)$ is the independently encoded message of TX k transmitted with power $P w_k$. Then, we consider the processed channel $\hat{U}_k = \frac{1}{\sqrt{P}} \mathbf{t}_k^{\text{H}} \mathbf{y}$, where $\mathbf{t}_k = [\mathbf{t}_{1,k}^{\text{T}} \dots \mathbf{t}_{L,k}^{\text{T}}]^{\text{T}}$ is a distributed linear combiner satisfying the information constraint $\mathbf{t}_k \in \mathcal{T}$. Let $R_k^{\text{UL}} := I(U_k; \hat{U}_k)$ be achievable ergodic rates on this channel. By standard information inequalities [7], we obtain

$$\begin{aligned} I(U_k; \hat{U}_k) &= h(U_k) - h(U_k | \hat{U}_k) \\ &\geq \log(\pi e) - h(U_k - \alpha \hat{U}_k) \\ &\geq \log(\pi e) - \log(\pi e \mathbb{E}[|U_k - \alpha \hat{U}_k|^2]). \end{aligned} \quad (\text{B.71})$$

Optimizing α according to channel statistics, i.e., choosing $\alpha = \alpha^*$ with $\alpha^* := \mathbb{E}[U_k \hat{U}_k^*] / \mathbb{E}[|\hat{U}_k|^2]$ being the solution of $\min_{\alpha} \mathbb{E}[|U_k - \alpha \hat{U}_k|^2]$, leads to the well-known UatF bound [2]

$$R_k^{\text{UL}} \geq R_k^{\text{UatF}} := \log(1 + \text{SINR}_k), \quad (\text{B.72})$$

$$\text{SINR}_k := \frac{w_k |\mathbb{E}[\mathbf{t}_k^{\text{H}} \mathbf{g}_k]|^2}{w_k \text{Var}[\mathbf{t}_k^{\text{H}} \mathbf{g}_k] + \sum_{j \neq k} w_j \mathbb{E}[|\mathbf{t}_k^{\text{H}} \mathbf{g}_j|^2] + \frac{\mathbb{E}[|\mathbf{t}_k|^2]}{P}}. \quad (\text{B.73})$$

Alternatively, we can keep $\alpha \in \mathbb{C}$ unoptimized and obtain the bound $R_k^{\text{UL}} \geq R_k^{\text{UatF}} \geq \log(\mathbb{E}[|U_k - \alpha \hat{U}_k|^2])^{-1}$, where after simple manipulations we recognize

$$\mathbb{E}[|U_k - \alpha \hat{U}_k|^2] = \mathbb{E} \left[\left\| \alpha \mathbf{W}^{\frac{1}{2}} \mathbb{H} \mathbf{t}_k - \mathbf{e}_k \right\|^2 \right] + \frac{\alpha^2}{P} \mathbb{E}[|\mathbf{t}_k|^2] = \text{MSE}_k(\alpha \mathbf{t}_k). \quad (\text{B.74})$$

The above steps also shows that solving $\min_{\alpha \in \mathbb{C}, \mathfrak{t}_k \in \mathcal{T}} \text{MSE}_k(\alpha \mathfrak{t}_k)$ is equivalent to solving

$$\underset{\mathfrak{t}_k \in \mathcal{T}}{\text{maximize}} R_k^{\text{UatF}}. \quad (\text{B.75})$$

Furthermore, we observe that $\min_{\alpha \in \mathbb{C}, \mathfrak{t}_k \in \mathcal{T}} \text{MSE}_k(\alpha \mathfrak{t}_k) = \min_{\mathfrak{t}_k \in \mathcal{T}} \text{MSE}_k(\mathfrak{t}_k)$. This is because α is a deterministic scalar, hence $\alpha \mathfrak{t}_k \in \mathcal{T}$. Therefore, Problem (B.75) and Problem (5.6) have the same optimal solution \mathfrak{t}_k^* , and the optima are related by $R_k^{\text{UatF}} = \log(\text{MSE}_k(\mathfrak{t}_k^*))^{-1}$.

Let $R_k^*(\mathbf{w})$ be the optimum of problem (B.75) for some dual UL power allocation policy $\mathbf{w} := (w_1, \dots, w_K) \in \mathbb{R}_+^K$. Let then $\mathcal{R}^{\text{UatF}}$ be the union of all rate tuples $(R_1, \dots, R_K) \in \mathbb{R}_+^K$ satisfying $R_k \leq R_k^*(\mathbf{w}) \forall k \in \mathcal{K}$, where the union is taken over all \mathbf{w} satisfying $\sum_{k=1}^K P w_k \leq P_{\text{sum}}$. By definition of $\mathcal{R}^{\text{UatF}}$, its Pareto boundary $\partial \mathcal{R}^{\text{UatF}}$ is composed by rate tuples of the type (R_1, \dots, R_K) , $R_k = R_k^*(\mathbf{w})$, achieved by some \mathbf{w} satisfying $\sum_k w_k \leq K$ and by the optimal combiners $\{\mathfrak{t}_k\}_{k=1}^K$ solving Problem (5.6) $\forall k \in \mathcal{K}$. It turns out that it is possible to fully characterize $\partial \mathcal{R}^{\text{UatF}}$ by restricting $\mathbf{w} \in \mathbb{R}_\Delta$, i.e., by using all the available power $KP = P_{\text{sum}}$. This is because R_k^{UatF} is a continuous monotonic increasing functions of P , and so is its (finite) supremum over $\mathfrak{t}_k \in \mathcal{T}$. Furthermore, it can be shown that *all* $\mathbf{w} \in \mathbb{R}_\Delta$ induce Pareto optimal rate tuples. This last statement can be proven by contradiction as follows.

Let $\mathbf{w} \in \mathbb{R}_\Delta$ and suppose that $(R_1^*(\mathbf{w}), \dots, R_K^*(\mathbf{w})) \notin \partial \mathcal{R}^{\text{UatF}}$, i.e., $\exists \mathbf{w}' \in \mathbb{R}_\Delta$, $\mathbf{w}' \neq \mathbf{w}$, s.t. $R_k^*(\mathbf{w}') > R_k^*(\mathbf{w}) \forall k \in \mathcal{K}$. We now build an iterative procedure which moves from \mathbf{w} to \mathbf{w}' and contradicts the previous supposition. Consider the following sequence of updates $\mathbf{w}^{(i)} := (w_1^{(i)}, \dots, w_K^{(i)})$ for $i = 0, \dots, K-1$, where $\mathbf{w}^{(0)} := \mathbf{w}$ and

$$w_k^{(i)} := \begin{cases} w'_k & \text{if } k \leq i \\ \frac{\sum_{j>i} w'_j}{\sum_{j>i} w_j^{(i-1)}} w_k^{(i-1)} & \text{if } k > i \end{cases}. \quad (\text{B.76})$$

The i -th step of the above procedure changes w_i into the target w'_i and scales all weights w_k with $k > i$ by a common factor s.t. the constraint $\mathbf{w}^{(i)} \in \mathbb{R}_\Delta$ is not violated. Note that this constraint also implies that $\mathbf{w}^{(K-1)} = \mathbf{w}'$ without the need of a K -th update. In the following, we use properties of R_k^* inferred by the fact that $\log\left(1 + \frac{ax}{bx+cy+d}\right)$ is continuous monotonic increasing in $x \in \mathbb{R}_+$ and continuous monotonic decreasing in $y \in \mathbb{R}_+$ for any $a \geq 0, b \geq 0, c \geq 0$, and $d > 0$, and so is its supremum over some family of parameters (a, b, c, d) . At step $i = 1$, assume w.l.o.g. that $w_1^{(1)} = w'_1 \geq w_1$, which also implies $w_k^{(1)} = \frac{\sum_{j>i} w'_j}{\sum_{j>i} w_j^{(0)}} w_k \leq w_k$ for $k > 1$. In fact, we can always reindex the RXs such that this assumption holds. When going from $x = 1$ to $x = \frac{\sum_{j>i} w'_j}{\sum_{j>i} w_j^{(0)}} \leq 1$, and then subsequently from $y = 1$

to $y = \frac{w'_1}{w_1} \geq 1$, the function

$$\log \left(1 + \frac{w_k |\mathbb{E}[\mathfrak{t}_k^H \mathfrak{g}_k]|^2 x}{(\sum_{j>1} w_j \mathbb{E}[|\mathfrak{t}_k^H \mathfrak{g}_j|^2] - w_k |\mathbb{E}[\mathfrak{t}_k^H \mathfrak{g}_k]|^2) x + w_1 \mathbb{E}[|\mathfrak{t}_k^H \mathfrak{g}_1|^2] y + \frac{\mathbb{E}[|\mathfrak{t}_k|^2]}{P}} \right) \quad (\text{B.77})$$

for $k > 1$ is continuous monotonically decreasing, and so is its supremum over $\mathfrak{t}_k \in \mathcal{T}$. Hence, we have $R_k^*(\mathbf{w}^{(1)}) \leq R_k^*(\mathbf{w}^{(0)}) \forall k > 1$. At step $i = 2$, we assume w.l.o.g. that $w'_2 \geq w_2^{(1)}$ (otherwise we can just properly reindex all RXs $k > 2$), and similarly obtain $R_k^*(\mathbf{w}^{(2)}) \leq R_k^*(\mathbf{w}^{(1)}) \forall k > 2$. By continuing until step $K - 1$, we finally obtain $R_K^*(\mathbf{w}^{(K-1)}) \leq R_K^*(\mathbf{w}^{(K-2)})$, which can be combined with the previous steps leading to the desired contradiction $R_K^*(\mathbf{w}') \leq R_K^*(\mathbf{w})$ up to a possible reindexing, i.e., at least one rate cannot be strictly increased when moving from \mathbf{w} to \mathbf{w}' .

The proof is concluded by invoking the duality principle between the UatF bound and the hardening bound [2, Theorem 4.8], which shows that $\mathcal{R}^{\text{UatF}} = \mathcal{R}^{\text{hard}}$ and that for every rate tuple $(R_1^{\text{UatF}}, \dots, R_K^{\text{UatF}})$ achieved by some $\{\mathfrak{t}_k\}_{k=1}^K$ and \mathbf{w} , there is a rate tuple $(R_1^{\text{hard}}, \dots, R_K^{\text{hard}}) = (R_1^{\text{UatF}}, \dots, R_K^{\text{UatF}})$ achievable by using the same combiners as precoders $\{\mathfrak{t}_k\}_{k=1}^K$, and by choosing $p_k = \frac{\tilde{p}_k P}{\mathbb{E}[|\mathfrak{t}_k|^2]} \forall k \in \mathcal{K}$ with $\tilde{\mathbf{p}} := [\tilde{p}_1, \dots, \tilde{p}_K]^\top$ being the solution of

$$(\mathbf{D}^{-1} - \mathbf{B})\tilde{\mathbf{p}} = (\mathbf{D}^{-1} - \mathbf{B}^\top)\mathbf{w}, \quad (\text{B.78})$$

where $\mathbf{D} := \text{diag}(d_1, \dots, d_K)$, $d_k := \text{SINR}_k \frac{\mathbb{E}[|\mathfrak{t}_k|^2]}{\mathbb{E}[|\mathfrak{t}_k^H \mathfrak{g}_k|^2]}$, and where the (k', k) -th element of $\mathbf{B} \in \mathbb{C}^{K \times K}$ is given by

$$[\mathbf{B}]_{k',k} = \begin{cases} \frac{\mathbb{E}[|\mathfrak{t}_k^H \mathfrak{g}_{k'}|^2]}{\mathbb{E}[|\mathfrak{t}_k|^2]} & \text{if } k' \neq k \\ \frac{\text{Var}[\mathfrak{t}_k^H \mathfrak{g}_k]}{\mathbb{E}[|\mathfrak{t}_k|^2]} & \text{otherwise.} \end{cases} \quad (\text{B.79})$$

The above linear system is guaranteed to have a unique solution satisfying $\sum_{k=1}^K \tilde{p}_k = \sum_{k=1}^K w_k$, which implies $\sum_{l=1}^L \mathbb{E}[|\mathfrak{x}_l|^2] = \sum_{k=1}^K w_k P = P_{\text{sum}}$.

B.3.2 Proof of Theorem 5.3

To avoid cumbersome notation, we omit the subscript k everywhere. The proof is split into three separate lemmas. We start with a minor extension of [34, Theorem 3] obtained by introducing the constraint $\mathbb{E}[|\mathfrak{t}_l(S_l)|^2] < \infty$ and specializing to the cost function considered in here.

Lemma B.3 (Existence and uniqueness). *Problem (5.6) admits a unique team optimal solution.*

Proof. Let \mathcal{H} be the space of Σ -measurable functions $\mathbf{a} : \Omega \rightarrow \mathbb{C}^{LN}$ s.t. $\mathbb{E}[\mathbf{a}^H \mathbb{O} \mathbf{a}] < \infty$. We define the inner product $\langle \mathbf{a}, \mathbf{b} \rangle := \mathbb{E}[\mathbf{b}^H \mathbb{O} \mathbf{a}]$, $\forall (\mathbf{a}, \mathbf{b}) \in \mathcal{H}^2$, and its induced norm $\|\mathbf{a}\|_{\mathcal{H}} := \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle}$, $\forall \mathbf{a} \in \mathcal{H}$. Then, the tuple $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ is a Hilbert space¹ [34]. Let us further define $\mathfrak{t}_0 := \mathbb{O}^{-1} \mathfrak{g}$, which is the unique minimizer of $c(\mathbf{H}, \mathfrak{t})$ for any realization \mathbf{H} . Firstly, we observe that $\mathfrak{t}_0 \in \mathcal{H}$, since

$$\|\mathfrak{t}_0\|_{\mathcal{H}}^2 = \mathbb{E} \left[\mathbf{e}^H \mathbb{H} (\mathbb{H}^H \mathbb{H} + P^{-1} \mathbf{I})^{-1} \mathbb{H}^H \mathbf{e} \right] \quad (\text{B.80})$$

$$\leq \mathbb{E} \left[\text{tr} \left(\mathbb{H} (\mathbb{H}^H \mathbb{H} + P^{-1} \mathbf{I})^{-1} \mathbb{H}^H \right) \right] \quad (\text{B.81})$$

$$= \mathbb{E} \left[\text{tr} \left(\mathbb{H}^H \mathbb{H} (\mathbb{H}^H \mathbb{H} + P^{-1} \mathbf{I})^{-1} \right) \right] \quad (\text{B.82})$$

$$\leq \mathbb{E} \left[\text{tr} \left((\mathbb{H}^H \mathbb{H} + P^{-1} \mathbf{I}) (\mathbb{H}^H \mathbb{H} + P^{-1} \mathbf{I})^{-1} \right) \right] = NL. \quad (\text{B.83})$$

Secondly, we observe that \mathcal{T} is a closed linear subspace of \mathcal{H} [35, Theorem 2.6.6]. Finally, simple algebraic manipulations show that the objective of Problem (5.6) can be equivalently rewritten as $\text{MSE}(\mathfrak{t}) = \|\mathfrak{t} - \mathfrak{t}_0\|_{\mathcal{H}}^2 - \|\mathfrak{t}_0\|_{\mathcal{H}}^2 + 1$. Therefore, by following [34, 35], we consider the infinite dimensional orthogonal projection problem:

$$\underset{\mathfrak{t} \in \mathcal{T}}{\text{minimize}} \|\mathfrak{t} - \mathfrak{t}_0\|_{\mathcal{H}}^2. \quad (\text{B.84})$$

The solution to Problem (5.6) corresponds to the projection of $\mathfrak{t}_0 \in \mathcal{H}$ onto the closed linear subspace $\mathcal{T} \subseteq \mathcal{H}$. By the Hilbert projection theorem, this projection is unique and always exists [70]. \square

The following result extends Lemma B.3 by following similar lines as [35, Theorem 2.6.6].

Lemma B.4 (Sufficiency of stationarity). *Suppose that $\mathbb{E}[\|\mathbb{O}\|_{\mathbb{F}}^2] < \infty$. Then, if $\mathfrak{t}^* \in \mathcal{T}$ is stationary, it is also the unique team optimal solution to Problem (5.6).*

Proof. Let us consider again the equivalent problem (B.84). Since \mathcal{T} is a closed linear subspace, the Hilbert projection theorem also states that a solution $\mathfrak{t}^* \in \mathcal{T}$ is the unique optimal solution if and only if the following orthogonality conditions

¹In fact, the positive matrix square root $\mathbb{O}^{\frac{1}{2}}$ induces an isometry between $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ and the perhaps more familiar Hilbert space of measurable functions such that $\mathbb{E}[\|\mathbf{a}\|^2] < \infty$, equipped with the standard inner product $\langle \mathbf{a}, \mathbf{b} \rangle := \mathbb{E}[\mathbf{b}^H \mathbf{a}]$.

[70] hold: $(\forall \mathfrak{t} \in \mathcal{T})$

$$\begin{aligned}
& \langle \mathfrak{t}^* - \mathfrak{t}_0, \mathfrak{t} \rangle = 0, \\
& \iff \mathbf{E} \left[\mathfrak{t}^H \mathbb{O} (\mathfrak{t}^* - \mathbb{O}^{-1} \mathfrak{g}) \right] = 0, \\
& \iff \mathbf{E} \left[\sum_{l=1}^L \mathfrak{t}_l^H \left(\mathbf{E}[\mathbb{O}_{l,l} | S_l] \mathfrak{t}_l^* + \sum_{j \neq l} \mathbf{E}[\mathbb{O}_{l,j} \mathfrak{t}_j^* | S_l] - \mathbf{E}[\mathfrak{g}_l | S_l] \right) \right] = 0,
\end{aligned} \tag{B.85}$$

where the last equality follows by the law of total expectation, provided that the inner expectations are finite. Finiteness of $\mathbf{E}[\mathfrak{g}_l | S_l]$ and $\mathbf{E}[\mathbb{O}_{l,l} | S_l]$ follows by the assumption $\mathbf{E}[\|\mathbb{H}\|_{\mathbb{F}}^2] < \infty$. Finiteness of $\mathbf{E}[\mathbb{O}_{l,j} \mathfrak{t}_j^* | S_l]$ follows by applying the Cauchy-Schwarz inequality elementwise, and by using $\mathbf{E}[\|\mathfrak{t}_j\|^2] < \infty$ and $\mathbf{E}[\|\mathbb{O}\|_{\mathbb{F}}^2] < \infty$. The proof is concluded by observing that if the stationary conditions in (5.11) are satisfied for some \mathfrak{t}^* , then the orthogonality conditions are satisfied and \mathfrak{t}^* is the unique optimal solution. \square

To conclude the proof, it remains to show the converse statement of Lemma B.4. In the following, we depart from [35, Theorem 2.6.6] and use a different argument tailored to the cost function considered in here.

Lemma B.5 (Necessity of stationarity). *Suppose that $\mathbf{E}[\|\mathbb{O}\|_{\mathbb{F}}^2] < \infty$. Then, if $\mathfrak{t}^* \in \mathcal{T}$ is the unique team optimal solution to Problem (5.6), it is also stationary.*

Proof. We start by using the notion of *person-by-person optimality* given by Definition A.2 in Appendix A, which states that a necessary condition for a tuple \mathfrak{t}^* to be globally optimal is that it must satisfy

$$\text{MSE}(\mathfrak{t}^*) = \min_{\mathfrak{t}_l \in \mathcal{T}_l} \text{MSE}(\mathfrak{t}_{-l}^*, \mathfrak{t}_l), \quad \forall l \in \mathcal{L}. \tag{B.86}$$

We relax the the above conditions by letting $\mathcal{T}_{l,\text{unc}}$ be the unconstrained version of \mathcal{T}_l , i.e., where we remove the constraint $\mathbf{E}[\|\mathfrak{t}_l\|^2] < \infty$. We then have

$$\begin{aligned}
\infty > \text{MSE}(\mathfrak{t}^*) &= \min_{\mathfrak{t}_l \in \mathcal{T}_l} \text{MSE}(\mathfrak{t}_{-l}^*, \mathfrak{t}_l) \\
&\geq \min_{\mathfrak{t}_l \in \mathcal{T}_{l,\text{unc}}} \text{MSE}(\mathfrak{t}_{-l}^*, \mathfrak{t}_l) \\
&= \min_{\mathfrak{t}_l \in \mathcal{T}_{l,\text{unc}}} \mathbf{E} \left[\mathbf{E} \left[c(\mathbb{H}, \mathfrak{t}_{-l}^*, \mathfrak{t}_l) \mid S_l \right] \right] \\
&\geq \mathbf{E} \left[\min_{\mathfrak{t}_l \in \mathcal{T}_{l,\text{unc}}} \mathbf{E} \left[c(\mathbb{H}, \mathfrak{t}_{-l}^*, \mathfrak{t}_l) \mid S_l \right] \right] \\
&= \mathbf{E} \left[\mathbf{E} \left[c(\mathbb{H}, \mathfrak{t}_{-l}^*, \mathfrak{t}_l^{**}) \mid S_l \right] \right] \\
&= \text{MSE}(\mathfrak{t}_{-l}^*, \mathfrak{t}_l^{**})
\end{aligned} \tag{B.87}$$

where \mathfrak{t}_l^{**} is given by the first-order optimality condition $\nabla_{\mathfrak{t}_l} \xi_l(S_l, \mathfrak{t}_l) = \mathbf{0}$ applied to the convex function $\xi_l(S_l, \mathfrak{t}_l) := \mathbb{E} [c(\mathbb{H}, \mathfrak{t}_{-l}^*, \mathfrak{t}_l) | S_l]$. Note that

$$\mathfrak{t}_l^{**}(S_l) := (\mathbb{E}[\mathbb{O}_{l,l} | S_l])^{-1} \left(\mathbb{E}[\mathfrak{g}_l | S_l] - \sum_{j \neq l} \mathbb{E}[\mathbb{O}_{l,j} \mathfrak{t}_j^* | S_l] \right) \in \mathcal{T}_{l,\text{unc}}, \quad (\text{B.88})$$

because it is given by sums and products of measurable functions (we recall that $\mathbb{O}_{l,l} \succ \mathbf{0}$), and that all the expectations are finite as discussed in the proof of Lemma B.4. Finally, we observe that $\mathfrak{t}_l^{**} \in \mathcal{T}_l$, i.e., $\mathbb{E}[\|\mathfrak{t}_l^{**}\|^2] < \infty$, because from the original problem formulation (5.6) we notice that $\text{MSE}(\mathfrak{t}_{-l}^*, \mathfrak{t}_l^{**})$ is given by a sum of non-negative terms, one of which is precisely $\frac{1}{P} \mathbb{E}[\|\mathfrak{t}_l^{**}\|^2]$, and $\text{MSE}(\mathfrak{t}_{-l}^*, \mathfrak{t}_l^{**}) \leq \text{MSE}(\mathfrak{t}^*) < \infty$. Therefore, the inequalities in (B.87) are in fact equalities, and the optimal solution must satisfy $\text{MSE}(\mathfrak{t}^*) = \text{MSE}(\mathfrak{t}_{-l}^*, \mathfrak{t}_l^{**})$, $\forall l \in \mathcal{L}$. This proves that an optimal solution must satisfy the stationarity conditions given by (5.11). \square

B.3.3 Proof of Lemma 5.1

We use the same notation and definitions as in the proof of Theorem 5.3 given in Appendix B.3.2. The optimality gap can be expressed as follows:

$$\text{MSE}(\mathfrak{t}) - \text{MSE}(\mathfrak{t}^*) \quad (\text{B.89})$$

$$= \|\mathfrak{t} - \mathfrak{t}_0\|_{\mathcal{H}}^2 - \|\mathfrak{t}^* - \mathfrak{t}_0\|_{\mathcal{H}}^2 \quad (\text{B.90})$$

$$= \|\mathfrak{t} - \mathfrak{t}^*\|_{\mathcal{H}}^2 \quad (\text{B.91})$$

$$= \langle \mathfrak{t} - \mathfrak{t}^*, \mathfrak{t} - \mathfrak{t}^* \rangle + \langle \mathfrak{t}^* - \mathfrak{t}_0, \mathfrak{t} - \mathfrak{t}^* \rangle \quad (\text{B.92})$$

$$= \langle \mathfrak{t} - \mathfrak{t}_0, \mathfrak{t} - \mathfrak{t}^* \rangle \quad (\text{B.93})$$

$$= \mathbb{E} \left[\sum_{l=1}^L (\mathfrak{t}_l - \mathfrak{t}_l^*)^{\text{H}} \left(\mathbb{E}[\mathbb{O}_{l,l} | S_l] \mathfrak{t}_l + \sum_{j \neq l} \mathbb{E}[\mathbb{O}_{l,j} \mathfrak{t}_j | S_l] - \mathbb{E}[\mathfrak{g}_l | S_l] \right) \right], \quad (\text{B.94})$$

where (B.91) follows from Pythagoras' theorem, (B.92) from the orthogonality condition $\langle \mathfrak{t}^* - \mathfrak{t}_0, \mathfrak{a} \rangle = 0$, $\forall \mathfrak{a} \in \mathcal{T}$ and $\mathfrak{t} - \mathfrak{t}^* \in \mathcal{T}$, and (B.94) by applying the law of total expectation as in (B.85). Then, the proof follows from

$$\begin{aligned} \|\mathfrak{t} - \mathfrak{t}^*\|_{\mathcal{H}}^2 &= \mathbb{E}[(\mathfrak{t} - \mathfrak{t}^*)^{\text{H}} \mathbb{Z}] \\ &= \langle \mathbb{O}^{-1} \mathbb{Z}, \mathfrak{t} - \mathfrak{t}^* \rangle \\ &\leq \|\mathbb{O}^{-1} \mathbb{Z}\|_{\mathcal{H}} \|\mathfrak{t} - \mathfrak{t}^*\|_{\mathcal{H}}, \end{aligned} \quad (\text{B.95})$$

where the last step is the Cauchy–Schwarz inequality, and where we use $\mathbb{O}^{-1} \preceq P\mathbf{I}$ which ensures $\|\mathbb{O}^{-1} \mathbb{Z}\|_{\mathcal{H}}^2 = \mathbb{E}[\mathbb{Z}^{\text{H}} \mathbb{O}^{-1} \mathbb{Z}] \leq P \mathbb{E}[\mathbb{Z}^{\text{H}} \mathbb{Z}] < \infty$.

B.3.4 Proof of Theorem 5.4 (additional details)

We rearrange the system at hand as $(\mathbf{D} + \mathbf{U}\mathbf{\Pi}^\top)\mathbf{C} = \mathbf{U}$, where $\mathbf{C} := [\mathbf{C}_1 \ \dots \ \mathbf{C}_L]^\top$, $\mathbf{\Pi} := [\mathbf{\Pi}_1 \ \dots \ \mathbf{\Pi}_L]^\top$, $\mathbf{U} := [\mathbf{I}_K \ \dots \ \mathbf{I}_K]^\top$, $\mathbf{D} := \text{diag}(\mathbf{I} - \mathbf{\Pi}_1, \dots, \mathbf{I} - \mathbf{\Pi}_L)$. The proof follows if $\mathbf{D} + \mathbf{U}\mathbf{\Pi}^\top$ is invertible, giving the optimal coefficients $\mathbf{C} = (\mathbf{D} + \mathbf{U}\mathbf{\Pi}^\top)^{-1}\mathbf{U}$. By Lemma B.6 given in Appendix B.3.7, $\mathbf{D} + \mathbf{U}\mathbf{\Pi}^\top$ is invertible if both \mathbf{D} and $\mathbf{D}^{-1} + \mathbf{\Pi}^\top\mathbf{U} = \mathbf{D}^{-1} + \sum_l \mathbf{\Pi}_l$ are invertible. Standard arguments show that $\mathbf{0} \preceq \mathbf{\Pi}_l \prec \mathbf{I}$. Therefore, \mathbf{D} is Hermitian positive definite (hence invertible), and so is $\mathbf{D}^{-1} + \sum_l \mathbf{\Pi}_l$, concluding the proof.

B.3.5 Proof of Theorem 5.5 (additional details)

Assume $\mathbf{0} \preceq \mathbf{\Pi}_l \prec \mathbf{I}$ for a fixed $l \in \mathcal{L}$. Then, let $\hat{\mathbf{O}}_l := \hat{\mathbf{H}}_l^\mathbf{H}\hat{\mathbf{H}}_l + \mathbf{\Sigma}_l + P^{-1}\mathbf{I}$ and observe that this implies $\hat{\mathbf{O}}_l - \hat{\mathbf{H}}_l^\mathbf{H}\mathbf{\Pi}_l\hat{\mathbf{H}}_l = \hat{\mathbf{H}}_l^\mathbf{H}(\mathbf{I} - \mathbf{\Pi}_l)\hat{\mathbf{H}}_l + \mathbf{\Sigma}_l + P^{-1}\mathbf{I} \succ \mathbf{0}$. Therefore, we obtain

$$\begin{aligned} \hat{\mathbf{H}}_l \left(\hat{\mathbf{O}}_l - \hat{\mathbf{H}}_l^\mathbf{H}\mathbf{\Pi}_l\hat{\mathbf{H}}_l \right)^{-1} \hat{\mathbf{H}}_l^\mathbf{H} &= \hat{\mathbf{H}}_l \left[\hat{\mathbf{O}}_l \left(\mathbf{I} - \mathbb{T}_l\mathbf{\Pi}_l\hat{\mathbf{H}}_l \right) \right]^{-1} \hat{\mathbf{H}}_l^\mathbf{H} \\ &= \hat{\mathbf{H}}_l \left(\mathbf{I} - \mathbb{T}_l\mathbf{\Pi}_l\hat{\mathbf{H}}_l \right)^{-1} \mathbb{T}_l \\ &= \mathbb{P}_l \left(\mathbf{I} - \mathbf{\Pi}_l\mathbb{P}_l \right)^{-1} \end{aligned} \quad (\text{B.96})$$

where the last two equalities follow from Lemma B.7 and Lemma B.8 given in Appendix B.3.7, respectively. These lemmas ensure that all the above inverses exist, and in particular $(\mathbf{I} - \mathbf{\Pi}_l\mathbb{P}_l)^{-1}$. Furthermore, the above chain of equalities also show that

$$\begin{aligned} \mathbf{\Pi}_{l-1} &= \mathbb{E}[\mathbb{P}_l\mathbb{V}_l] + \mathbf{\Pi}_l\mathbb{E}[\tilde{\mathbb{V}}_l] \\ &= \mathbf{\Pi}_l + (\mathbf{I} - \mathbf{\Pi}_l)\mathbb{E}[\mathbb{P}_l(\mathbf{I} - \mathbf{\Pi}_l\mathbb{P}_l)^{-1}](\mathbf{I} - \mathbf{\Pi}_l) \\ &= \mathbf{\Pi}_l + (\mathbf{I} - \mathbf{\Pi}_l)^{\frac{1}{2}}\mathbb{E}[\tilde{\mathbb{P}}_l](\mathbf{I} - \mathbf{\Pi}_l)^{\frac{1}{2}}, \end{aligned} \quad (\text{B.97})$$

where $\tilde{\mathbb{P}}_l := \tilde{\mathbb{H}}_l \left(\tilde{\mathbb{H}}_l^\mathbf{H}\tilde{\mathbb{H}}_l + \mathbf{\Sigma}_l + P^{-1}\mathbf{I} \right)^{-1} \tilde{\mathbb{H}}_l^\mathbf{H}$, and $\tilde{\mathbb{H}}_l := (\mathbf{I} - \mathbf{\Pi}_l)^{\frac{1}{2}}\hat{\mathbf{H}}_l$. By standard argument, it can be shown that $\mathbf{0} \preceq \tilde{\mathbb{P}}_l \prec \mathbf{I}$ holds, and hence $\mathbf{0} \preceq \mathbf{\Pi}_{l-1} \prec \mathbf{I}$. Overall, the above discussion proves that $\mathbf{0} \preceq \mathbf{\Pi}_l \prec \mathbf{I}$ implies the existence of $(\mathbf{I} - \mathbf{\Pi}_l\mathbb{P}_l)^{-1}$ and that $\mathbf{0} \preceq \mathbf{\Pi}_{l-1} \prec \mathbf{I}$. By finally observing that $\mathbf{0} \preceq \mathbb{P}_{L-1} \prec \mathbf{I}$ and hence $\mathbf{0} \preceq \mathbf{\Pi}_{L-1} = \mathbb{E}[\mathbb{P}_{L-1}] \prec \mathbf{I}$, the proof is concluded by repeating the previous argument recursively.

B.3.6 Proof of Lemma 5.3

We recall the following results from random matrix theory, provided without proof: for $\text{vec}(\mathbb{H}_l) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, we have $\mathbb{E}[\mathbb{H}_l(\mathbb{H}_l^\mathbf{H}\mathbb{H}_l)^{-1}\mathbb{H}_l^\mathbf{H}] = \frac{N}{K}\mathbf{I}$ and $\mathbb{E}[\mathbb{H}_l(\mathbb{H}_l^\mathbf{H}\mathbb{H}_l)^{-2}\mathbb{H}_l^\mathbf{H}] =$

$\frac{N}{K(K-N)}\mathbf{I}$. We define the projection matrix $\mathbb{P}_l := \mathbb{H}_l(\mathbb{H}_l^H\mathbb{H}_l)^{-1}\mathbb{H}_l^H$ onto $\text{span}(\mathbb{H}_l)$, the projection matrix $\mathbb{P}_l^\perp := \mathbf{I} - \mathbb{P}_l$ onto its orthogonal complement, and let $\mathbf{t}_k = (\mathbf{t}_{1,k}, \dots, \mathbf{t}_{L,k})$ as in (5.38). A simple recursive calculation shows the identity $\mathbf{e}_k - \sum_{j=1}^l \mathbb{H}_j \mathbf{t}_{j,k} = \mathbb{P}_l^\perp \mathbb{P}_{l-1}^\perp \dots \mathbb{P}_1^\perp \mathbf{e}_k$. The first part of the objective in (5.6) is then given by

$$\mathbb{E} \left[\left\| \mathbf{e}_k - \sum_{j=1}^L \mathbb{H}_j \mathbf{t}_{j,k} \right\|^2 \right] = \mathbb{E} [\| \mathbb{P}_L^\perp \mathbb{P}_{L-1}^\perp \dots \mathbb{P}_1^\perp \mathbf{e}_k \|^2] \quad (\text{B.98})$$

$$= \mathbf{e}^H \mathbb{E} [\mathbb{P}_1^\perp, \dots, \mathbb{P}_{L-1}^\perp \mathbb{P}_L^\perp \mathbb{P}_L^\perp \mathbb{P}_{L-1}^\perp \dots \mathbb{P}_1^\perp] \mathbf{e} \quad (\text{B.99})$$

$$= \left(1 - \frac{N}{K}\right) \mathbf{e}_k^H \mathbb{E} [\mathbb{P}_1^\perp, \dots, \mathbb{P}_{L-1}^\perp \mathbb{P}_{L-1}^\perp \dots \mathbb{P}_1^\perp] \mathbf{e}_k \quad (\text{B.100})$$

$$= \left(1 - \frac{N}{K}\right)^L, \quad (\text{B.101})$$

where we used the Hermitian symmetry and idempotency of projection matrices, and the independence between \mathbb{H}_l and $\{\mathbb{H}_j\}_{j \neq l}$. We now measure the suboptimality of \mathbf{t}_k by using Lemma 5.1, specialized to the current setting similarly to Lemma 5.2. We have:

$$\mathbf{z}_{l,k}(S_l) = (\mathbb{H}_l^H \mathbb{H}_l + P^{-1} \mathbf{I}) \mathbf{t}_{l,k}(S_l) + \mathbb{H}_l^H \left(\sum_{j \neq l} \mathbb{E} [\mathbb{H}_j \mathbf{t}_{j,k} | S_l] - \mathbf{e}_k \right) \quad (\text{B.102})$$

$$= P^{-1} \mathbf{t}_{l,k}(S_l) - \mathbb{H}_l^H \mathbb{E} [\mathbb{P}_L^\perp \mathbb{P}_{L-1}^\perp \dots \mathbb{P}_1^\perp \mathbf{e}_k | S_l] \quad (\text{B.103})$$

$$= P^{-1} \mathbf{t}_{l,k}(S_l) - \left(1 - \frac{N}{K}\right)^{L-l+1} \mathbb{H}_l^H \mathbb{P}_l^\perp \dots \mathbb{P}_1^\perp \mathbf{e}_k \quad (\text{B.104})$$

$$= P^{-1} \mathbf{t}_{l,k}(S_l), \quad (\text{B.105})$$

where the last step follows from the definition of projection matrices, which gives $\mathbb{H}_l^H \mathbb{P}_l^\perp \mathbf{u} = \mathbf{0}$ for any $\mathbf{u} \in \mathbb{C}^K$. Furthermore, we have

$$\mathbb{E} [\|\mathbf{t}_{l,k}\|^2] = \mathbb{E} \left[\left(\mathbf{e}_k - \sum_{j=1}^{l-1} \mathbb{H}_j \mathbf{t}_{j,k} \right)^H \mathbb{H}_l (\mathbb{H}_l^H \mathbb{H}_l)^{-2} \mathbb{H}_l^H \left(\mathbf{e}_k - \sum_{j=1}^{l-1} \mathbb{H}_j \mathbf{t}_{j,k} \right) \right] \quad (\text{B.106})$$

$$= \frac{N}{K(N-K)} \left(1 - \frac{N}{K}\right)^{l-1} < \infty. \quad (\text{B.107})$$

Therefore, Lemma 5.1 applies and, by using the looser bound in (5.16), we readily obtain

$$\text{MSE}_k(\mathfrak{t}_k) - \text{MSE}_k(\mathfrak{t}_k^*) \leq \frac{1}{P} \sum_{l=1}^L \mathbb{E} [\|\mathfrak{t}_{l,k}\|^2] \xrightarrow{P \rightarrow \infty} 0, \quad (\text{B.108})$$

and $\text{MSE}_k(\mathfrak{t}_k^*) \geq \text{MSE}_k(\mathfrak{t}_k) - P^{-1} \sum_{l=1}^L \mathbb{E} [\|\mathfrak{t}_{l,k}\|^2] = (1 - \frac{N}{K})^L$.

B.3.7 Linear algebra background

Lemma B.6 (Woodbury matrix identity). *Let $\mathbf{A} \in \mathbb{C}^{n \times n}$, $\mathbf{B} \in \mathbb{C}^{n \times m}$, $\mathbf{C} \in \mathbb{C}^{m \times n}$, and $\mathbf{D} \in \mathbb{C}^{m \times m}$. If \mathbf{A} , \mathbf{C} , and $\mathbf{D}^{-1} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ are invertible, then $\mathbf{A} + \mathbf{B}\mathbf{D}\mathbf{C}$ is invertible and*

$$(\mathbf{A} + \mathbf{B}\mathbf{D}\mathbf{C})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{D}^{-1} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}. \quad (\text{B.109})$$

Lemma B.7 (Inverse of product). *Let \mathbf{A} and \mathbf{B} be two square matrices of the same dimension. If $\mathbf{A}\mathbf{B}$ is invertible, then \mathbf{A} and \mathbf{B} are also invertible, and $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.*

Lemma B.8 (Push-through identity). *Let $\mathbf{A} \in \mathbb{C}^{n \times m}$ and $\mathbf{B} \in \mathbb{C}^{m \times n}$ be two matrices such that $\mathbf{I} + \mathbf{A}\mathbf{B}$ is invertible. Then, $\mathbf{I} + \mathbf{B}\mathbf{A}$ is also invertible, and $\mathbf{B}(\mathbf{I} + \mathbf{A}\mathbf{B})^{-1} = (\mathbf{I} + \mathbf{B}\mathbf{A})^{-1}\mathbf{B}$.*

Bibliography

- [1] Ericsson, “Ericsson mobility report.” ericsson.com/mobility-report, 2021.
- [2] E. Björnson, J. Hoydis, and L. Sanguinetti, “Massive MIMO networks: Spectral, energy, and hardware efficiency,” *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.
- [3] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge university press, 2008.
- [4] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.
- [5] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.
- [6] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.
- [7] A. El Gamal and Y.-H. Kim, *Network information theory*. Cambridge university press, 2011.
- [8] Ö. T. Demir, E. Björnson, and L. Sanguinetti, “Foundations of user-centric cell-free massive mimo,” *Foundations and Trends® in Signal Processing*, vol. 14, no. 3-4, pp. 162–472, 2021.
- [9] A. D. Wyner, “Shannon-theoretic approach to a Gaussian cellular multiple-access channel,” *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 1713–1727, 1994.
- [10] S. Shamai and B. M. Zaidel, “Enhancing the cellular downlink capacity via co-processing at the transmitting end,” in *IEEE VTS 53rd Vehicular Technology Conference*, vol. 3, pp. 1745–1749 vol.3, 2001.

- [11] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, “Network coordination for spectrally efficient communications in cellular systems,” *IEEE Wireless Communications*, vol. 13, pp. 56–61, Aug. 2006.
- [12] S. Venkatesan, A. Lozano, and R. A. Valenzuela, “Network MIMO: Overcoming intercell interference in indoor wireless systems,” in *41st Asilomar Conference on Signals, Systems and Computers*, pp. 83–87, 2007.
- [13] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, “Cloud RAN for mobile networks - a technology overview,” *IEEE Communication Surveys and Tutorials*, vol. 17, no. 1, pp. 405–426, 2015.
- [14] O. Simeone, A. Maeder, M. Peng, O. Sahin, and W. Yu, “Cloud radio access network: Virtualizing wireless access for dense heterogeneous systems,” *Journal of Communication and Networks*, vol. 18, pp. 135–149, Apr. 2016.
- [15] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, “Cell-free massive MIMO versus small cells,” *IEEE Transactions on Wireless Communications*, vol. 16, pp. 1834–1850, Mar. 2017.
- [16] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, “Multi-cell MIMO cooperative networks: A new look at interference,” *IEEE Journal on Selected Areas in Communications*, vol. 28, pp. 1380–1408, Sept. 2010.
- [17] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, “Prospective multiple antenna technologies for beyond 5G,” *IEEE Journal on Selected Areas in Communications*, vol. 38, pp. 1637–1660, Aug. 2020.
- [18] P. de Kerret and D. Gesbert, “Degrees of freedom of the network MIMO channel with distributed CSI,” *IEEE Transactions on Information Theory*, vol. 58, pp. 6806–6824, Nov. 2012.
- [19] D. Gesbert and P. de Kerret, “Team methods for device cooperation in wireless networks,” in *Cooperative and Graph Signal Processing*, pp. 469–487, Elsevier, 2018.
- [20] A. Bazco-Nogueras, P. de Kerret, D. Gesbert, and N. Gresset, “On the degrees-of-freedom of the K-user distributed broadcast channel,” *IEEE Transactions on Information Theory*, vol. 66, no. 9, pp. 5642–5659, 2020.
- [21] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, “Precoding and power optimization in cell-free massive MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 16, pp. 4445–4459, July 2017.

- [22] L. Du, L. Li, H. Q. Ngo, T. C. Mai, and M. Matthaiou, “Cell-free massive MIMO: Joint maximum-ratio and zero-forcing precoder with power control,” *IEEE Transactions on Communications*, vol. 69, pp. 3741–3756, June 2021.
- [23] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. J. Smith, and R. Valenzuela, “Increasing downlink cellular throughput with limited network MIMO coordination,” *IEEE Transactions on Wireless Communications*, vol. 8, pp. 2983–2989, June 2009.
- [24] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, “Networked MIMO with clustered linear precoding,” *IEEE Transactions on Wireless Communications*, vol. 8, pp. 1910–1921, Aug. 2009.
- [25] E. Björnson and E. Jorswieck, *Optimal resource allocation in coordinated multi-cell systems*. Now Publishers Inc, 2013.
- [26] S. Buzzi and C. D’Andrea, “Cell-free massive MIMO: User-centric approach,” *IEEE Commun. Lett.*, vol. 6, pp. 706–709, Dec. 2017.
- [27] G. Interdonato, E. Björnson, H. Q. Ngo, P. Frenger, and E. G. Larsson, “Ubiquitous cell-free massive MIMO communications,” *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, p. 197, 2019.
- [28] E. Björnson and L. Sanguinetti, “Scalable cell-free massive MIMO systems,” *IEEE Transactions on Communications*, vol. 68, pp. 4247–4261, July 2020.
- [29] H. Dahrouj and W. Yu, “Coordinated beamforming for the multicell multi-antenna wireless system,” *IEEE Transactions on Wireless Communications*, vol. 9, pp. 1748–1759, May 2010.
- [30] V. R. Cadambe and S. A. Jafar, “Interference alignment and degrees of freedom of the k -user interference channel,” *IEEE Transactions on Information Theory*, vol. 54, pp. 3425–3441, Aug. 2008.
- [31] C. E. Shannon, “Channels with side information at the transmitter,” *IBM Journal of Research and Development*, vol. 2, no. 4, pp. 289–293, 1958.
- [32] G. Caire and K. R. Kumar, “Information theoretic foundations of adaptive coded modulation,” *Proceedings of the IEEE*, vol. 95, pp. 2274–2298, Dec. 2007.
- [33] J. Marschak and R. Radner, *Economic Theory of Teams*. Yale University Press, 1972.

- [34] R. Radner, “Team decision problems,” *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 857–881, 1962.
- [35] S. Yüksel and T. Başar, *Stochastic networked control systems: Stabilization and optimization under information constraints*. Springer Science & Business Media, 2013.
- [36] Z. H. Shaik, E. Björnson, and E. G. Larsson, “MMSE-optimal sequential processing for cell-free massive MIMO with radio stripes,” *IEEE Transactions on Communications*, 2021. (Early Access).
- [37] J. Rodríguez Sánchez, F. Rusek, O. Edfors, M. Sarajlić, and L. Liu, “Decentralized massive MIMO processing exploring daisy-chain architecture and recursive algorithms,” *IEEE Transactions on Signal Process.*, vol. 68, pp. 687–700, Jan. 2020.
- [38] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: information-theoretic and communications aspects,” *IEEE Transactions on Information Theory*, vol. 44, pp. 2619–2692, June 1998.
- [39] G. Caire, “On the ergodic rate lower bounds with applications to massive MIMO,” *IEEE Transactions on Wireless Communications*, vol. 17, pp. 3258–3268, May 2018.
- [40] D. Gündüz and O. Simeone, “On the capacity region of a multiple access channel with common messages,” in *2010 IEEE International Symposium on Information Theory*, pp. 470–474, 2010.
- [41] G. Caire and S. Shamai (Shitz), “On the capacity of some channels with channel state information,” *IEEE Transactions on Information Theory*, vol. 45, pp. 2007–2019, June 1999.
- [42] G. Keshet, Y. Steinberg, N. Merhav, *et al.*, “Channel coding in the presence of side information,” *Foundations and Trends® in Communications and Information Theory*, vol. 4, no. 6, pp. 445–586, 2008.
- [43] R. G. Gallager, *Information theory and reliable communication*, vol. 2. Springer, 1968.
- [44] D. Slepian and J. K. Wolf, “A coding theorem for multiple access channels with correlated sources,” *Bell Syst. Tech. J.*, vol. 52, pp. 1037–1076, Sep. 1973.

- [45] A. Lapidoth and Y. Steinberg, “The multiple-access channel with causal side information: Common state,” *IEEE Transactions on Information Theory*, vol. 59, pp. 32–50, Jan. 2013.
- [46] S. Jafar, “Capacity with causal and noncausal side information: A unified view,” *IEEE Transactions on Information Theory*, vol. 52, pp. 5468–5474, Dec. 2006.
- [47] G. Como and S. Yuksel, “On the capacity of memoryless finite-state multiple-access channels with asymmetric state information at the encoders,” *IEEE Transactions on Information Theory*, vol. 57, pp. 1267–1273, March 2011.
- [48] A. Lapidoth and Y. Steinberg, “The multiple-access channel with causal side information: Double state,” *IEEE Transactions on Information Theory*, vol. 59, pp. 1379–1393, March 2013.
- [49] N. Şen, F. Alajaji, S. Yüksel, and G. Como, “Memoryless multiple access channel with asymmetric noisy state information at the encoders,” *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7052–7070, 2013.
- [50] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, “Multiuser MIMO achievable rates with downlink training and channel state feedback,” *IEEE Transactions on Information Theory*, vol. 56, pp. 2845–2866, June 2010.
- [51] Y. Steinberg, “Coding for the degraded broadcast channel with random parameters, with causal and noncausal side information,” *IEEE Transactions on Information Theory*, vol. 51, pp. 2867–2877, Aug. 2005.
- [52] H. Joudeh and B. Clerckx, “Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: A rate-splitting approach,” *IEEE Transactions on Communications*, vol. 64, pp. 4847–4861, Nov. 2016.
- [53] G. Caire and S. Shamai, “On the achievable throughput of a multiantenna gaussian broadcast channel,” *IEEE Transactions on Information Theory*, vol. 49, pp. 1691–1706, July 2003.
- [54] H. Weingarten, Y. Steinberg, and S. S. Shamai, “The capacity region of the gaussian multiple-input multiple-output broadcast channel,” *IEEE Transactions on Information Theory*, vol. 52, pp. 3936–3964, Sept. 2006.
- [55] E. J. Candes and T. Tao, “The power of convex relaxation: near-optimal matrix completion,” *IEEE Transactions on Information Theory*, vol. 56, pp. 2053–2080, May 2010.

- [56] T. Kato, *A short introduction to perturbation theory for linear operators*. Springer-Verlag New York, 1982.
- [57] V. Lau, Y. Liu, and T.-A. Chen, “On the design of MIMO block-fading channels with feedback-link capacity constraint,” *IEEE Transactions on Communications*, vol. 52, pp. 62–70, Jan. 2004.
- [58] Y. Linde, A. Buzo, and R. Gray, “An algorithm for vector quantizer design,” *IEEE Transactions on Communications*, vol. 28, pp. 84–95, Jan. 1980.
- [59] S. S. Christensen, R. Agarwal, E. de Carvalho, and J. M. Cioffi, “Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design,” *IEEE Transactions on Wireless Communications*, vol. 7, pp. 4792–4799, July 2008.
- [60] F. Negro, I. Ghauri, and D. T. M. Slock, “Sum rate maximization in the noisy MIMO interfering broadcast channel with partial CSIT via the expected weighted MSE,” in *International Symposium on Wireless Communication Systems (ISWCS)*, pp. 576–580, IEEE, 2012.
- [61] S. Shi, M. Schubert, and H. Boche, “Rate optimization for multiuser mimo systems with linear processing,” *IEEE Transactions on Signal Process.*, vol. 56, pp. 4020–4030, Aug. 2008.
- [62] G. Wahba, *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics, 1990.
- [63] E. Björnson and L. Sanguinetti, “Making cell-free massive MIMO competitive with MMSE processing and centralized implementation,” *IEEE Transactions on Wireless Communications*, vol. 19, pp. 77–90, Jan. 2020.
- [64] D. Neumann, T. Wiese, M. Joham, and W. Utschick, “A bilinear equalizer for massive MIMO systems,” *IEEE Transactions on Signal Process.*, vol. 66, no. 14, pp. 3740–3751, 2018.
- [65] 3GPP, “Further advancements for E-UTRA physical layer aspects (release 9),” 3GPP TS 36.814, 3rd Generation Partnership Project (3GPP), Mar. 2017.
- [66] W. Yu and T. Lan, “Transmitter optimization for the multi-antenna downlink with per-antenna power constraints,” *IEEE Transactions on Signal Process.*, vol. 55, pp. 2646–2660, June 2007.

- [67] J. Krainak, J. Speyer, and S. Marcus, "Static team problems—part i: Sufficient conditions and the exponential cost criterion," *IEEE Transactions on Automatic Control*, vol. 27, pp. 839–848, Apr. 1982.
- [68] W. Yu, W. Rhee, S. Boyd, and J. M. Cioffi, "Iterative water-filling for gaussian vector multiple-access channels," *IEEE Transactions on Information Theory*, vol. 50, pp. 145–152, Jan. 2004.
- [69] N. Liu and S. Ulukus, "Capacity region and optimum power control strategies for fading Gaussian multiple access channels with common data," *IEEE Transactions on Communications*, vol. 54, pp. 1815–1826, Oct. 2006.
- [70] D. G. Luenberger, *Optimization by vector space methods*. John Wiley & Sons, 1997.