Fundamental Limits and Algorithms in Decentralized and Cooperative Wireless Networks

Antonio Bazco Nogueras

Fundamental Limits and Algorithms in Decentralized and Cooperative Wireless Networks

Dissertation

submitted to

Sorbonne Université

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

Author:

Antonio BAZCO NOGUERAS

Successfully defended on the 28th November, 2019, before a committee composed of:

Reviewer	**Prof.**	**Ignacio SANTAMARÍA**	Universidad de Cantabria, Spain
Reviewer	**HDR**	**Bruno CLERCKX**	Imperial College London, U.K.
Examiner	**Prof.**	**Michèle WIGGER**	Télécom Paris, France
Examiner	**Prof.**	**Osvaldo SIMEONE**	King's College London, U.K.
Examiner	**Prof.**	**Petros ELIA**	EURECOM, France
Industrial Advisor	**Dr.**	**Nicolas GRESSET**	Mitsubishi Electric R&D Centre Europe, France
Thesis Advisor	**Prof.**	**David GESBERT**	EURECOM, France
Thesis Co-Advisor	**HDR**	**Paul de KERRET**	EURECOM, France

Limites Fondamentales et Algorithmes dans des Réseaux Sans-Fils Décentralisés et Coopératifs

Thèse

soumise à

Sorbonne Université

pour l'obtention du Grade de Docteur

présentée et soutenue publiquement par:

Antonio BAZCO NOGUERAS

Soutenance de thèse effectuée le 28 Novembre 2019 devant le jury composé de:

Rapporteur	**Prof.**	**Ignacio SANTAMARÍA**	Universidad de Cantabria, Espagne
Rapporteur	**HDR**	**Bruno CLERCKX**	Imperial College London, Royaume-Uni
Examinateur	**Prof.**	**Michèle WIGGER**	Télécom Paris, France
Examinateur	**Prof.**	**Osvaldo SIMEONE**	King's College London, Royaume-Uni
Examinateur	**Prof.**	**Petros ELIA**	EURECOM, France
Co-encadrant Industriel	**Dr.**	**Nicolas GRESSET**	Mitsubishi Electric R&D Centre Europe, France
Directeur de Thèse	**Prof.**	**David GESBERT**	EURECOM, France
Co-Directeur de Thèse	**HDR**	**Paul de KERRET**	EURECOM, France

Abstract

Cooperative transmission between different nodes is expected to be one of the main tools towards managing interference in the increasingly complex and cluttered wireless networks. Network cooperation is known to bring multiplicative gains under certain ideal assumptions, such as having perfect information about the propagation channel towards the users at every cooperating node. However, those ideal assumptions are not feasible in many of the current network settings and applications. There exist several reasons whereby the desired hypotheses do not hold. In order to achieve the promised gains, the network should be able to perfectly share all the information –in a timely manner– among the different cooperating nodes, which can be non co-located. Yet, current wireless transmissions cope with many challenging constraints, as tight delay constraints, fast-changing channels or rate-limited backhaul links –to name just a few–, that impede the perfect sharing of the locally collected information. The topic of analyzing how the non-fulfillment of the ideal hypotheses impacts the performance of cooperative settings has generated great interest in the research community. Nevertheless, the main focus has been concentrated on scenarios in which, although the channel state information is imperfectly or only partially collected, all the nodes share the same information. This last assumption is neither feasible in many current wireless networks scenarios, as stated by the previously described constraints.

This thesis aims for shedding light on the performance of decentralized cooperative settings in which the information available at each node may be different and potentially of different accuracy. In particular, we focus on the so-called distributed Network MIMO, in which a set of transmitters jointly serve a set of users. The setting considered is characterized by two main aspects: The perfect sharing of the user's information data and the imperfect sharing of the channel state information. We analyze the distributed Network MIMO setting so as to characterize its fundamental limits and provide novel algorithms.

Specifically, the analysis is carried out from two different perspectives. We start by characterizing the Degrees-of-Freedom metric of the setting. The Degrees-of-Freedom is an approximation of the capacity at high signal-to-noise ratio that allows to identify

insightful understandings of the network behavior. The contribution is twofold, as we provide both achievable schemes that increase considerably the performance with respect to the solutions available in the literature, and upper-bounds that illustrate up to which scale the distributed setting performance is harmed with respect to the perfect-sharing setting. It turns out that, in some configurations, the distributed setting is able to attain the Degrees-of-Freedom performance of the ideal setting with perfect sharing of the channel information. The second perspective consists in restricting the transmission to the conventional paradigm of Zero-Forcing and studying whether the performance losses from decentralized information can be precisely calculated. More precisely, we analyze the achievable rate at high signal-to-noise ratio with the goal of quantifying the rate loss from decentralization, i.e., we compute the difference of rate between the distributed setting and the ideal centralized setting with perfect sharing. We propose a novel zero-forcing scheme tailored to the decentralized configuration that asymptotically attains the centralized rate. On the basis of the aforementioned analysis, we also tackle related challenges such as the best channel information allocation, the impact of instantaneous power constraint and other challenges arising from the setting considered.

Abrégé

La transmission coopérative entre différents nœuds devrait être l'un des principaux outils de gestion des interférences dans des réseaux sans fil de plus en plus complexes et encombrés. La coopération en réseau est connue pour apporter des gains multiplicatifs sous certaines hypothèses idéales, telles que le fait de disposer d'informations parfaites sur le canal de propagation vers les utilisateurs à chaque nœud coopérant. Toutefois, ces hypothèses idéales ne sont pas réalisables dans bon nombre des paramètres et des applications réseau actuels. Il existe plusieurs raisons pour lesquelles les hypothèses souhaitées ne peuvent pas être considérées. Afin de atteindre les gains promis, le réseau devrait être en mesure de partager parfaitement toutes les informations –dans un délai convenable– entre les différents nœuds de coopération, qui peuvent être situés dans des lieux différents. Pourtant, les transmissions sans fil actuelles doivent faire face à de nombreuses contraintes, comme des délais serrés, des canaux en évolution rapide ou des liaisons de retour à débit limité, pour n'en nommer que quelques-unes, qui empêchent le partage parfait des informations recueillies localement. Le sujet de l'analyse de l'impact de la non-réalisation des hypothèses idéales sur la performance des scénarios coopératifs a suscité un grand intérêt dans le milieu de la recherche. Néanmoins, l'accent a surtout été mis sur des scénarios dans lesquels, bien que les informations sur l'état des canaux soient imparfaites ou seulement partiellement collectées, tous les nœuds partagent les mêmes informations. Cette dernière hypothèse n'est pas réalisable dans de nombreux scénarios actuels de réseaux sans fil, comme l'indiquent les contraintes décrites précédemment.

Cette thèse vise à mettre en lumière la performance des scénarios coopératifs décentralisés dans lesquels l'information disponible à chaque nœud peut être différente et potentiellement d'une précision différente. En particulier, nous nous concentrons sur ce qu'on appelle le réseau distribué MIMO, dans lequel un ensemble d'émetteurs servent conjointement un ensemble d'utilisateurs. Le cas considéré se caractérise par deux aspects principaux : Le partage parfait des données d'information de l'utilisateur et le partage imparfait des informations sur l'état du canal. Nous analysons le réglage distribué du réseau MIMO afin de caractériser ses limites fondamentales et de fournir de nouveaux algorithmes.

Plus précisément, l'analyse est effectuée sous deux angles différents. Nous commençons par caractériser les Degrés de Liberté du scénario. Les Degrés de Liberté est une approximation de la capacité à un rapport signal/bruit élevé qui permet d'identifier des aperçus perspicaces du comportement du réseau. La contribution est double puisque nous fournissons à la fois des schémas réalisables qui augmentent considérablement la performance par rapport aux solutions disponibles dans la littérature et des limites supérieures qui illustrent jusqu'à quelle échelle la performance du scénario distribué est affectée par rapport au réseau avec du partage parfait. Il s'avère que, dans certaines configurations, le réglage distribué est capable d'atteindre les performances de Degrés de Liberté du réglage idéal avec un partage parfait des informations du canal. La deuxième perspective consiste à limiter la transmission au paradigme conventionnel du Zéro-Forçage et à étudier si les pertes de performance à cause de la décentralisation peuvent être calculées avec précision. Particulièrement, nous analysons le débit réalisable à un rapport signal/bruit élevé dans le but de quantifier la perte de débit due à la décentralisation, c'est-à-dire que nous calculons la différence de débit entre le réglage distribué et le réglage centralisé idéal avec un partage parfait. Nous proposons un nouveau système de zéro-forçage adapté à la configuration décentralisée qui atteint asymptotiquement le débit centralisé. Sur la base de l'analyse susmentionnée, nous nous attaquons également à des défis connexes tels que la meilleure allocation de l'information de canal, l'impact de la contrainte de puissance instantanée et d'autres défis découlant du contexte considéré.

Acknowledgements

First and foremost, I would like to express my gratitude to the three people that have guided and supported me throughout the completion of my Ph.D. degree. I am deeply grateful to my thesis co-advisor and friend Paul de Kerret for the countless, yet always appealing, technical discussions. Without his constructive criticism and continuous encouragement this thesis would never have been possible. I really appreciate his profound technical knowledge, in conjunction with his human kindness and comprehension. I would like to thank my thesis advisor David Gesbert for his guidance and support. His well known academic wisdom and vision can only compete with his personal qualities, which helped to create a delightful atmosphere within the research group. I am also very grateful to my industrial advisor Nicolas Gresset for his technical advise and engaging discussions. His industrial expertise and his love of knowledge have been highly uplifting. I really appreciate the privilege of having all three of them as supervisors, and their mentorship will always be a source of inspiration.

It has been a great pleasure and an honor to have the opportunity to pursue my Ph.D. degree in EURECOM within the framework of an industrial collaboration with Mitsubishi Electric R&D Centre Europe (MERCE). It is hard to conceive a better environment for the start and development of a research career, not only for their independent excellence but also for their synergistic interplay. I would also like to thank Prof. Syed A. Jafar for hosting me at University of California, Irvine, as well as to the rest of his group for their warm hospitality and the enlightening technical discussions.

All the long journeys are highly shaped by the people that accompany and support you. For this reason, I would like to thank my office-mates, group-mates, and *coffee*-mates for all the joyful time that we have spent together. Special thanks to my friends Adrian Lahuerta, Flavio Maschietti, and Pasquale Lisena, and to Lorenzo Miretti for the enjoyable technical conversations. I am also very grateful to Loïc Brunel, David Mottier, Marie Plantard, and many others in MERCE, for their warm welcome and help.

Last but not least, I would like to express my heartfelt gratitude to my family, for their endless love and continuous far-reaching support, and to Ane, for her support, patience, and love, which has proven strong enough to make distance negligible.

Contents

List of Figures

List of Tables

Acronyms and Abbreviations

The acronyms and abbreviations used throughout the manuscript are specified in the following. They are presented here in their singular form, and their plural forms are constructed by adding and s, e.g. TX (transmitter) and TXs (transmitters). The meaning of an acronym is also indicated the first time that it is used. Plural for

AWGN	Additive White Gaussian Noise.
BC	Broadcast Channel.
CDF	Cumulative Density Function.
C-CSIT	Centralized Channel State information at the Transmitters.
CSI	Channel State Information.
CSIT	Channel State Information at the Transmitter.
CSIR	Channel State Information at the Receiver.
D-CSIT	Distributed Channel State information at the Transmitters.
DL	Downlink.
DPC	Dirty Paper Coding.
DoF	Degrees of Freedom.
eq.	Equation.
FDD	Frequency Division Duplexing.
GDoF	Generalized Degrees of Freedom.
i.i.d.	Independent and identically distributed.
IA	Interference Alignment.
IC	Interference Channel.
IoT	Internet of Things.
H-CSIT	Hierarchical Channel State information at the Transmitters.
LHS	Left hand side.
MAP	Maximum a Posteriori.
MIMO	Multiple Input Multiple Output.
MISO	Multiple Input Single Output.
PDF	Probability Density Function.
PMF	Probability Mass Function.
RHS	Right hand side.
RVQ	Random Vector Quantization.
RX	Receiver(s).

RZF	Regularized Zero-Forcing.
SIMO	Single Input Multiple Output.
SINR	Signal to Noise and Interference Ratio.
SNR	Signal-to-Noise Ratio.
SISO	Single-Input Single-Output.
TDD	Time Division Duplex.
TS	Time Slot.
TX	Transmitter(s).
UL	Uplink.
URLLC	Ultra-Reliable Low-Latency Robust Communications.
w.l.o.g.	Without loss of generality.
ZF	Zero Forcing.

Notations

The next list describes an overview on the notation used throughout this manuscript. We use boldface uppercase letters ($\mathbf{A}$) for matrices, boldface lowercase letters for vectors ($\mathbf{a}$), and regular lowercase letters for scalars (a). Events are represented by regular uppercase letters (A) and sets by calligraphic uppercase letters ($\mathcal{A}$). Unless otherwise stated, the same notation is used for a random variable and its realization. For any matrix $\mathbf{A}$, $\sqrt{\mathbf{A}}$ denotes the matrix composed of the square root of the elements of $\mathbf{A}$.

$(x_1, x_2)^+$	$\max(x_1, x_2, 0)$.
$\angle a$	The phase of a complex scalar $a = \lvert a\rvert e^{\imath \angle a}$.
argmax	Points or elements of the domain of some function at which the function values are maximized.
argmin	Points or elements of the domain of some function at which the function values are minimized.
$\mathbf{A}^{\mathrm{H}}$	Conjugate transpose of matrix $\mathbf{A}$.
$\mathbf{A}^{\mathrm{T}}$	Transpose of matrix $\mathbf{A}$.
$\mathcal{N}_{\mathbb{C}}(0, 1)$	Circularly symmetric complex Gaussian distribution.
$\mathbb{E}_1(x)$	Exponential integral defined as $\mathbb{E}_1(x) \triangleq \int_1^\infty \frac{e^{-xt}}{t}\,\mathrm{d}t$.
$\Gamma_d(k, \theta)$	Gamma Distribution with shape k and scale θ.
$\imath$	Imaginary unit, defined as $\imath \triangleq \sqrt{-1}$.
$\mathbb{1}$	Indicator function.
$\bar{i}$	For an index $i \in \mathbb{N}_I$, $\bar{i} = \mathbb{N}_I \backslash i$.
ln	Natural logarithm.
$\measuredangle(\mathbf{x}, \mathbf{y})$	Angle between two vectors $\mathbf{x}, \mathbf{y}$. $\lvert\cos(\measuredangle(\mathbf{x}, \mathbf{y}))\rvert = \lvert\mathbf{x}^{\mathrm{H}}\mathbf{y}\rvert/(\lVert\mathbf{x}\rVert\lVert\mathbf{y}\rVert)$.
$\odot$	Hadamard (element-wise) matrix product.
$\xrightarrow{a.s.}$	Converges almost surely to.
$\mathbb{R}^+$	Set of positive real numbers : $\{x \in \mathbb{R} : x > 0\}$.

$\mathrm{Re}(z)$, $\mathrm{Im}(z)$	Real and imaginary part of $z \in \mathbb{C}$.
$\triangleq$	Definition.
$\mathbf{1}_{R\times C}$	All-ones matrix of size $R \times C$.
$\mathbf{0}_{R\times C}$	Zero matrix of size $R \times C$.
$a \perp\!\!\!\perp b$	The random variables a and b are independent.
$f * g(x)$	Convolution of the functions $f(x)$ and $g(x)$.
$x \sim X$	Variable x distributed following the X distribution.
a.s.	almost surely.

Part I

Motivation and Models

Chapter 1

Introduction

Wireless communications technologies have considerably evolved in the past decade. It is expected that the imminent deployment of the fifth generation cellular network technology (5G) [1] will bring not only several order higher data rates, but also a collection of new and diversified use cases. Indeed, this diversification of services is one of the main targets of the 5G development [2]. Three main use cases are contemplated in the upcoming nascent generation: Ultra-Reliable Low-Latency Robust Communications (URLLC) [3–5], Enhanced Mobile Broadband (eMBB) [6], and Massive Machine-type Communication (mMTC) [7, 8]. The conjunction of those three aspects is intended to contribute to the blossoming of unseen functionalities such as tactile Internet [9, 10], UAV-aided networks [11, 12], vehicular networks [13, 14], or Internet-of-Things (IoT) [15]. In order to be able to provide these new applications, the network will leverage innovative technologies [16], such as millimeters-wave communications [17–19], caching [20, 21], device-to-device communications [22], or massive-MIMO [23–26]. Furthermore, the expected higher network density –in terms of cells and devices– makes interference management one of the essential problems for wireless transmissions [2].

One consequence that arises from the network portrayal previously described is the increase of the heterogeneity in the network, both between nodes communicating with each other and between different networks sharing the same resources. This heterogeneity also affects the backhaul capabilities of the different nodes. Moreover, the situations in which the communicating nodes are moving around at high speed are burgeoning. Both heterogeneity and high mobility impede the possibility of a centralized management of the wireless communications, and hence the necessity of understanding how distributed systems behave and what are their fundamental limits.

1.1 Cooperative Transmission

The notable densification of the network results in a necessity of cooperation to avoid the congestion of the wireless medium. Multi-user cooperative networks and the extend of its theoretical capabilities have been thoroughly analyzed in the literature [27–31]. Network cooperation can take shape in many different forms. Originally, cooperation was reduced to static policies that ensured a certain functioning, e.g. partial frequency reuse. With the escalation of the network complexity, cooperative methods evolved to address the soaring requirements [32]. Intuitively, the cooperation gains are subordinated to which information is shared among the cooperating nodes. We can distinguish two categories of essential information: The system –or channel– information and the user's data information.

Considering the user's data information, the cooperation mechanisms depend on whether this information is available at all the nodes or not. If the user's data are not shared, such that each cooperative node is endowed with different data information, the cooperation can be carried through coordinated beamforming [33, 34] or coordinated scheduling [35]. In the opposite scenario, with sharing of the user's data, we can apply strategies with stronger cooperation along with the previous ones. One of the main cooperative strategies is the Coordinated Multipoint (CoMP) transmission [36]. This configuration, also known as Cooperative Multiple-input multiple-output (MIMO), joint transmission, or Network MIMO, benefits from the data sharing so that the interference can be canceled or even turned into useful signal.

Multi-user joint transmission in wireless networks is known to bring multiplicative improvements in network rates [37], but only under the assumption of perfect Channel State Information (CSI). This perfect CSI scenario has been profoundly studied [37–41]. Unfortunately, perfect CSI acquisition is not possible in most of the current network applications because of its complexity and resource consumption. Consequently, the literature has striven to unravel how imperfect or quantized CSI at the transmitters (CSIT) affects the performance.

1.2 Precoding Under Non-Ideal Backhaul and CSI

1.2.1 Imperfect Channel State Information

Motivated by the infeasibility of the previous ideal assumption, settings in which the information available at the communicating nodes does not meet the perfect CSI assumption were thoroughly investigated. Thereby, the community focused on settings where the information available is timely but imperfect [31, 42–50], or where the information

is delayed [51–57]. This research topic has kept active during the last decade, and a large number of works have developed generalized schemes for the case of partially outdated [54, 58, 59], alternating [60], or evolving CSIT [61]. The attempt to comprehend the behavior of the current networks has led to the study of elaborated and complex settings. For example, the Cognitive Interference Channel, in which only some nodes have access to the other nodes' information [62, 63], or the relay network [64–68].

Significantly, even though the aforementioned works assumed an imperfect acquisition or estimation of the CSI, they consider that all the cooperating nodes share the same imperfect information. Hereinafter, we refer to the setting where the transmission is optimized on the basis of a *single* imperfect/outdated channel estimate being common at every transmit antenna as Centralized CSIT (C-CSIT) setting. Nevertheless, current and upcoming wireless networks characteristics make this assumption of *perfect sharing* impractical in many applications. This is due to, for example, the proliferation of heterogeneous networks for which some of the nodes have a wireless, fluctuating, or limited backhaul [69–71], or URLLC applications [3, 4, 72], in which the perfect sharing of the information would result in an inoperable delay. Settings in which simple devices with low capabilities aim to communicate in a dense environment –as in IoT applications– also fall into the use cases in which the sharing of information is both desirable and challenging. This evolution of different use cases boosts the interest of distributed information settings, in which the information available at the communicating nodes is not only imperfect but different from one node to another. This type of settings can be included in the so-called Team Decision problems [73, 74], in which different agents aiming at the same goal attempt to cooperate in the absence of perfect communication between them.

Recently, the increasing importance of cooperation of non-collocated transmitters –as, for example, in Unmanned Aerial Device (UAV) aided networks [11]– has led to an increasing number of works challenging this assumption of centralized CSIT. In [75, 76], methods have been developed to reduce the CSIT required to achieve MIMO Interference Alignment (IA), and the high-SNR regime with delayed and local CSIT in the Interference Channel (IC) is also studied in several works [77–79]. The assumption of centralized CSIT has also been challenged in capacity analysis for the Multiple Access Channel [80] and the Relay Channel [81], among others.

1.2.2 Precoding Under Distributed Channel State Information Setting

The C-CSIT model assumption can model a multi-antenna transmitter or a joint transmission from different non-colocated transmitters in the case where we assume a ideal Cloud Radio Access Network (C-RAN) [16]. In these cases, it is feasible to assume that the imperfect information is *perfectly shared* between the non-colocated transmitting

antennas. Nevertheless, upcoming heterogeneous networks include a huge variety of devices, such as user terminals, drone-enabled relays, pico base stations, etc., seeking to cooperate for transmission despite the lack of an ideal backhaul linking them. Other scenarios featuring existing backhaul links may favor local processing over centralized one in order to meet the tight latency constraints derived from 5G and tactile internet applications [9].

This aspect fosters analyzing what happens when this CSI is not perfectly shared among the devices, i.e., when each node may have a different CSI. This new setting is called *Distributed CSIT* (D-CSIT) setting [82]. In this network configuration, each node is endowed with an imperfect information about the system state. This information can be different from node to node, and moreover the accuracy at one node can differ from one parameter to another. Hence, the heterogeneity of the current wireless networks is correctly contained in this model. In this thesis, we focus on the distributed Network MIMO setting to study the impact of those discrepancies between cooperating nodes. While it was suggested in the past literature that Distributed CSIT scenarios can severely impact on the network performance in comparison with classical limited-yet-centralized CSIT ones [82], a crucial problem is how transmitters can cooperatively combat the lack of mutual CSI consistency in order to reduce the gap with respect to the centralized system performance.

Several works have focused on this Distributed CSIT setting [83], e.g., analyzing Interference Alignment performance [76] or studying the Regularized Zero-Forcing performance in the large system limit [84]. However, many of the issues and challenges introduced by this setting are still open problems. As a result, there is a clear interest in looking at the scenario in which each transmitter may have a different information about the channel [85]. There exists a great number of different distributed settings [83, 86–91]. Nonetheless, this thesis is aimed at the so-called Distributed Network MIMO, wherein the transmitters have access to all the information symbols of the users, yet do not share the same CSIT [82]. This model arises in scenarios in which the data can be buffered or cached [92–94], but the CSI needs to be available with very small delay, such as high-mobility scenarios, or IoT or V2X networks with fast channel varying but low data rate [15, 95, 96]. In general, any use case in which latency constraints impede efficient CSIT sharing within the channel coherence time.

1.3 Thesis Outline and Main Contributions

This thesis is divided into three parts. Prior to present the contributions of this thesis, we introduce its motivation and scope in the first part, that is composed of two chapters.

The current chapter is committed to motivating the study of the topic considered in this thesis, as well as providing an overview of the state of the art and the main contributions. The other introductory chapter, Chapter 2, comprises a comprehensive description of the model and tools considered. In particular, it describes the mathematical model for the assumption of decentralized information, the figures of merit and the notations employed. Besides that, Chapter 2 also discloses some practical scenarios that motivate and bring about the theoretical model.

In each of the two other parts we aim to shed light on the fundamental limits of cooperative and decentralized communication from a different perspective. Part II seeks to characterize the optimal Degrees-of-Freedom (DoF) of the decentralized Network MISO setting, whereas Part III confronts the problem from a different point of view. In this part, we analyze the performance of Zero-Forcing precoding schemes in the distributed cooperative setting. The choice of Zero-Forcing is motivated by its simplicity and the fact that it is prevalently employed in practical transmissions with spatial multiplexing. Let us discuss these sections in more detail.

The DoF characterization carried through Part II usually requires a twofold analysis: The achievability analysis, in which we develop schemes that can attain a certain performance, and the converse, in which we establish upper-bounds on the attainable performance. We tackle both aspects in Part II, which comprises the following chapters:

- **Chapter 3:** In this chapter, we consider the simple single-antenna setting with 2 transmitters (TXs) and 2 receivers (RXs) and we study the Generalized Degrees-of-Freedom (GDoF) metric of a joint transmission in which the two TXs are endowed with a different channel information. Recently, the GDoF of the centralized setting in which both TXs share the channel information has been obtained in [50]. The main contribution of this chapter is the characterization of the sum GDoF for the decentralized counterpart setting in which each TX may have a different channel estimation. We show that the centralized GDoF performance is attained for any path-loss topology and whichever TX has the best estimate for each channel coefficient. This interesting result is obtained thanks to a novel precoding scheme that adapts to the setting configuration, and which is built on the idea that a TX only uses its instantaneous channel information if that information is the most accurate one among the TXs.

  The work presented in this chapter has resulted in the following publications:

  > *[97] Antonio Bazco, Paul de Kerret, David Gesbert, and Nicolas Gresset, "Generalized Degrees-of-Freedom of the 2-user MISO Broadcast Channel with Distributed CSIT," in Proc. IEEE International Symposium on Information*

Theory (ISIT), June 2017, pp. 1092–1096.

[98] Antonio Bazco-Nogueras, Paul de Kerret, David Gesbert, and Nicolas Gresset, "Distributed CSIT does not reduce the Generalized DoF of the 2-user MISO Broadcast Channel," in IEEE Wireless Communications Letters, June 2019, pp. 685-688.

- **Chapter 4:** The previous chapter benefits from the structure of the setting considered inasmuch as there exists only one interfered RX, and thus a single TX can manage the interference if we design a suitable scheme. Hence, the extension of those results to an expanded setting with more nodes is not straightforward. In this chapter, we study the $K \times K$ Network MISO with distributed CSIT so as to determine to what extent the previous results are generalizable. Our main contributions are twofold: First, we derive an intuitive centralized upper-bound for the setting with distributed CSIT. This upper-bound is based on a genie-aided setting in which the TXs are allowed to share their local CSIT with the other TXs. Consequently, the genie-aided setting is a centralized scenario in which every TX obtains K different estimates. We show that this setting attains the same DoF as a centralized setting in which the TXs are only endowed with the most accurate estimate among the K available estimates. Second, we develop an achievable scheme that increases considerably the DoF performance with respect to the known approaches in the literature. In a similar manner as in the previous chapter, the transmission scheme varies accordingly to the CSI configuration. This scheme shows that, for a certain CSI accuracy regime, it is still possible to attain the DoF of the genie-aided setting for any size of the network. The key to achieve this result is to capitalize on the idea that the unavoidable interference can be exploited as side information at the receiver. This approach was aptly applied in the literature related with delayed CSIT. However, here it is employed in a different manner, as it is retransmitted before its actual generation. Moreover, the achievable scheme illustrates how important is to choose appropriately *who* transmits and to *whom* it transmits in cooperative decentralized settings, as it turns out that the maximum DoF is sometimes obtained only if a part of the TXs do not transmit any signal.

  These results have led to the following publications.

  [99] Antonio Bazco, Paul de Kerret, David Gesbert, and Nicolas Gresset, "Méthode de transmission robuste au partage imparfait de l'information de canal entre transmetteurs," in Proc. Colloque GRETSI, September 2017.

[100] Antonio Bazco-Nogueras, Paul de Kerret, David Gesbert, and Nicolas Gresset, "On the Degrees-of-Freedom of the K-user Distributed Broadcast Channel," submitted to IEEE Transactions on Information Theory, 2018.

- **Chapter 5:** The two previous chapters are more focused on the achievability analysis, as the upper-bound is obtained from a centralized genie-aided setting. In Chapter 5 we turn our attention to distributed upper-bounds. In particular, we consider the Network MIMO setting, in which M transmit antennas jointly serve 2 multi-antenna RXs. Previously, we have assumed that the TXs are endowed with different CSI which can also be of different accuracy. In order to distinguish between the impact of imperfect CSI and distributed CSI, we simplify the CSIT configuration in this chapter, such that we consider that m transmit antennas have access to perfect CSI of the whole channel matrix, whereas the $M-m$ transmit antennas have only access to finite precision CSI. We present a distributed upper-bound for this regime, which is shown to be tight a certain regime. Specifically, for the regime in which the number of informed transmit antennas is bigger or equal than the minimum number of antennas at the users. We provide also a transmission scheme attaining the said upper-bound, and we generalize it to obtain a general lower-bound applicable at any regime. This chapter is composed of partial results that are not published yet.

The analysis of Part III is motivated by the results of Part II, as one of the main questions arising from the previous chapters is if those results extend to finer metrics than DoF and GDoF. In order to answer this question, we restrict our analysis to simple zero-forcing transmission schemes. The objective is to disclose the loss of performance on account of *not sharing perfectly the CSIT*. Therefore, we study the rate gap of the decentralized scenario with respect to the centralized setting in which the channel information at the transmitters is perfectly shared. We consider in this part that the precoder satisfies an instantaneous power constraint. This is important due to the decentralized structure of the network considered, since a transmitter cannot know the power normalization applied at the other transmitter because each one computes it on the basis of its own channel information, which can be different. Part III comprises two different chapters.

- **Chapter 6:** In this chapter, we consider the simple single-antenna setting with 2 TXs and 2 RXs as in the initial chapter of Part II. However, we analyze now the rate gap of the distributed setting when the joint transmission makes use of zero-forcing schemes to attenuate the interference. The contribution of this chapter is manifold. First, we show that the rate achieved with zero-forcing transmission in the distributed CSIT setting converges at high SNR to the rate attained in the

centralized genie-aided setting where the best channel estimate is shared between the transmitters. This result implies that there is not rate gap on the asymptotic regime. Second, we develop a zero-forcing-type precoding scheme tailored for the distributed setting. This precoding scheme builds on the main insights of Chapter 3 for the DoF analysis. Third, we propose novel precoding strategies that allow to increase considerably the performance at low-to-medium SNR. Among these strategies, it is noteworthy the fact that reducing the accuracy of the channel information at one TX can improve the performance. This behavior arises from an implicit compromise between the accuracy of the decision locally taken by a certain transmitter and the consistency –or agreement– between the decisions of both transmitters.

The work presented in this chapter has produced the following publications:

> *[101] Antonio Bazco-Nogueras, Lorenzo Miretti, Paul de Kerret, David Gesbert, and Nicolas Gresset, "Achieving Vanishing Rate Loss in Decentralized Network MIMO," in Proc. IEEE International Symposium on Information Theory (ISIT), July 2019, pp. 1457-1461.*

> *[102] Antonio Bazco-Nogueras, Lorenzo Miretti, Paul de Kerret, David Gesbert, and Nicolas Gresset, "Transmission Robuste de Zéro-Forçage Asymptotiquement Optimale pour Coopération Imparfaite de Transmetteurs," in Proc. Colloque GRETSI, August 2019.*

- **Chapter 7:** We extend the analysis of the previous chapter for the general $M \times K$ Network MISO setting with multiple-antenna transmitters. In a similar vein as in Chapter 4, the goal of this chapter is to comprehend the main insights of the results of the simple setting by confronting the analysis to a more general case. The contribution of this chapter is to show that the rate of the centralized genie-aided setting is asymptotically reached for all the channel information configurations for which the distributed setting attains the centralized DoF. In other words, the decentralized setting not only achieves the same multiplexing gain as the centralized setting, but also the beamforming gain. We futher study how this result extends to the non-asymptotic SNR regime. We develop a transmission scheme to achieve these results. This scheme also makes use of the idea that reducing the accuracy of information at some nodes improves the agreement between all the transmitters. However, it differs from the scheme of the 2×2 case in the fact that the transmitters with more accurate information now attempt to correct the interference generated by the other transmitters.

These results have been submitted to the following publication.

> *[103] Antonio Bazco-Nogueras, Paul de Kerret, David Gesbert, and Nicolas Gresset, "Asymptotically Achieving Centralized Rate on the $M \times K$ Decentralized Network MISO," submitted to IEEE Transactions on Information Theory, 2019.*

To finalize, we discuss in Chapter 8 the main conclusions that emerge from the work developed during this thesis.

Chapter 2

Problem Statement and System Model

We consider a cooperative wireless network in which several transmitters aim to jointly serve several users. In order to apply this Cooperative Multi-point transmission (CoMP), all the transmitting nodes share the information data symbols destined to be decoded at the receivers. Hence, the joint transmission is intended to cancel or avoid that the interference from other users impacts the performance. The main particularity of our model is the assumption of decentralized channel state information. Consequently, each transmitter owns a particular channel estimate, possibly different with respect to the one available at the other transmitters. We motivate in the following this scenario, and we define the mathematical model that we assume.

2.1 Motivation and Practical Examples of the Distributed CSIT setting

2.1.1 Perfect User Data Sharing

In order to account for TX-dependent limited feedback in the network MIMO channel, we focus in this thesis on a wireless configuration in which the user's data symbols are available and jointly transmitted from all the TXs, whereas the channel estimates could only be imperfectly obtained at the TXs. Such assumptions, although seemingly

contradictory at first sight, are actually very relevant in current wireless networks, and even more in future 5G-and-beyond wireless networks. The main reason is that, in many scenarios of interest, the latency constraint for data delivery is significantly looser than the CSI outdating constraint, as the later is related to the coherence time and hence very short in many relevant mobility scenarios. This property has for consequence that the data sharing (or caching) between TXs can be achieved in practice while *timely* CSI acquisition and sharing becomes the main bottleneck.

Our assumption that the user's data symbols are available at all the TXs is made possible without putting in question the scenario described above because of two recent major techniques envisioned for future 5G-and-beyond wireless networks: Caching [92, 104, 105] and Cloud-Ran/Fog-Ran [106, 107].

Through caching, the user's data symbols are pre-fetched at the TX nodes before the transmission occurs [92]. Caching is an increasingly important feature that already exists in many scenarios [20, 94] and is envisioned in many more [21]. With the user's data symbols available at the TXs, even at mobile and cost-efficient ones, the accurate and timely acquisition of the multi-user channel becomes the main bottleneck for efficient interference reduction. This leads to a D-CSIT configuration wherever the cooperation links are not obtained with sufficient accuracy.

In the Cloud-RAN paradigm, the centralization of the processing of all nodes is envisioned so as to gain full benefits of cooperation. This centralization is however limited by its cost and its delay, such that partial centralization is considered a promising solution [108]. Considering decentralized precoding at the TXs allows to reduce the *delay* in CSI acquisition. In that case, the backhaul links are solely used to convey the user's data since, for many data-oriented applications, the application's latency requirements are orders of magnitude slower than the rate at which the fading channel evolves. The CSI estimates are directly exchanged between the TXs through direct links, thus reducing the delay of the complete CSI acquisition at the TXs. This CSI exchange between TXs through limited resources leads to a D-CSIT configuration.

2.1.2 Imperfect CSI Acquisition and Sharing

Although the assumption of having decentralized channel state information may seem contradictory with the assumption of perfect sharing of user data symbols, both aspects co-exist in many scenarios of interest. This model is motivated by the different timescales of latency that information data and CSI may experience in a range of emerging applications. Indeed, CSI sharing is constrained by the channel coherence time, which can be very short in mobility scenarios (ms). This, together with the fact that the communication link between TXs is usually latency-limited backhaul implies that perfect CSI sharing is hard

to achieve. On the other hand, many data applications have delivery time restrictions which are orders of magnitude weaker, such that it can be pre-fetched or cached at the various TXs and ready to be synchronously transmitted.

To convey the main idea, a non-limiting practical scenario is depicted in Fig. 2.1. In this 2×2 scenario, under the assumption of centralized configuration, both TXs are connected via an ideal –no rate-limited, instantaneous– backhaul, e.g. an optic fiber connection. Conversely, in the distributed setting, the TXs are connected by an imperfect, rate-limited or delayed link –e.g., wireless link–, such that they may share some CSI. Then, if we assume a short channel coherence time, the TXs will not be able to perfectly share its CSI, although they may transmit some noisy, quantized or compressed version of the locally available information.

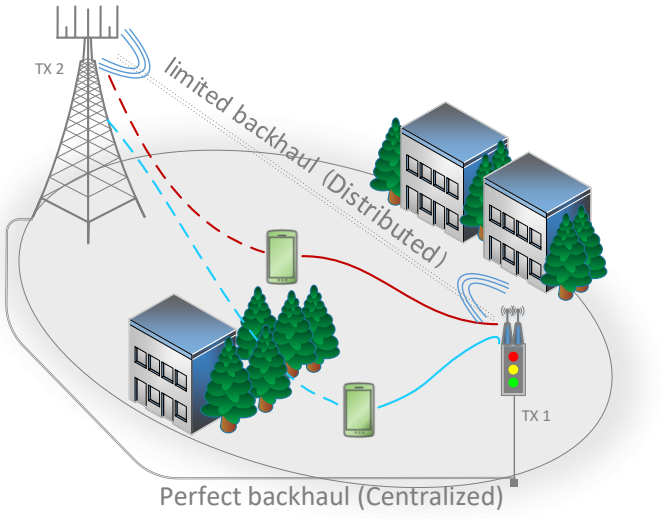
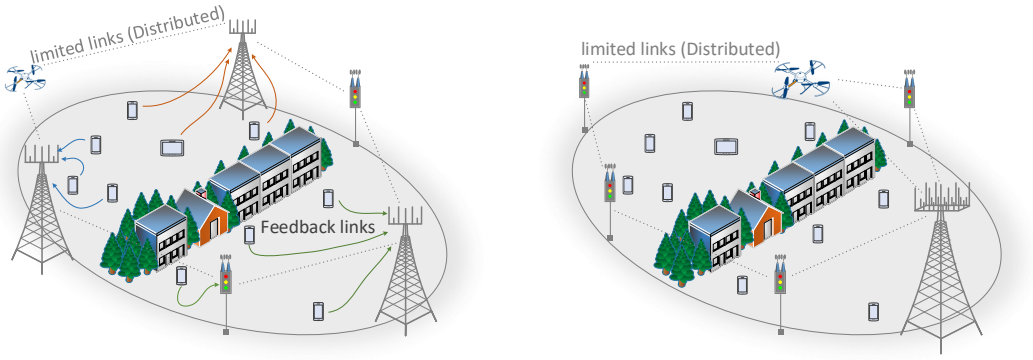


Figure 2.1 – Distributed vs Centralized CSIT example.

In the scenario described, depending on the CSI acquisition protocol, the CSI allocation will be different. For example, if we assume a Time Division Duplexing (TDD) setting in which the TX is estimating the uplink channel to make use of the reciprocity property, each TX will have a probably good estimate of its own local links towards both users. This implies that, in Fig. 2.1, TX 1 would obtain a fed-back estimate of the solid-line links and TX 2 would obtain a fed-back estimate of the dashed-line links. Then, the TXs could acquire a coarse estimate of the other TX's information via the TX-TX link. If Frequency Division Duplexing (FDD) is used, such that the RX sends to the TXs a quantized version of the channel, we can have: *a*) If the user is connected to only one TX, each TX will have good CSI of a single user –e.g., if TX 1 is connected with RX 1 it will obtain the dark red links, if it is connected to RX 2, the light blue links–; *b*) if the RX



(a) Multiple Base-Stations. Each one knows better the CSI of a subset of RXs. They obtain a less accurate estimate of the other RXs via the wireless TX-TX links.

(b) Master Base Station with remote radio-heads. It obtains an estimate of the whole channel matrix, then it transmits noisy or compressed CSI to the auxiliary TXs.

Figure 2.2 – Distributed CSIT setting use cases.

is connected to both, he may use a different feedback rate to adapt to the link capacity towards each of the TXs. Finally, TX 2 could also be a remote radio-head to enhance downlink transmission such that the RXs feed back the whole channel matrix to TX 1, and TX 1 attempts to send the most appropriate information to TX 2 via the limited wireless link. In this way, many different CSI configurations can appear as function of the system characteristics, all of them enclosed in the Distributed CSIT setting, that is rigorously defined in the following section.

We assume that a limited cooperation between TXs has taken place before the transmission phase, leading to a certain CSI accuracy configuration. Hence, we assume hereinafter that the average CSIT accuracy remains constant for a certain time. The problem of studying the best strategy of CSI sharing in a limited and constrained communication is a very interesting research problem per-se, but it is outside the scope of this work. Thus, we do not discuss the exact CSI acquisition mechanism. The generality of our model is exactly meant to adapt to any CSI sharing scenario.

The extension of the example for an arbitrary number of nodes follows easily. For instance, the model encloses a setting in which the RXs feed back their channel vectors to the TXs and 1) either a RX i sends the CSI to a TX j, and TX j shares a noisy or compressed information to the other TXs, or 2) each RX i sends a CSI of different accuracy towards each TX, depending on the link quality. The scenario is depicted in Fig. 2.2a. Furthermore, we can also model a scenario in which a main, multi or massive antenna base station serves a set of users with the help of some single or multi antenna remote radio-head or simple TXs, as depicted in Fig. 2.2b. We present a more comprehensive discussion about the possible network configurations in Chapter 4.

2.2 Downlink Transmitter Cooperation

In this thesis, we focus mainly on the downlink (DL) transmission, although most of the findings could be also applicable for the uplink (UL) transmission. The main reason to consider DL transmission is that it is a more challenging and pertinent case concerning the problem of CSI acquisition and sharing [109]. We consider the Network MISO setting in which M TXs jointly serve K single-antenna receivers (RXs), and where TX j has N_j antennas. We denote the total number of transmit antennas as

$$N_T \triangleq \sum_{j=1}^{M} N_j. \tag{2.1}$$

The TXs seek to deliver a certain message W_i to each RX i. The messages W_i are independent and identically distributed (i.i.d.), each one drawn from a circularly-symmetric complex Gaussian distribution $\mathcal{N}_{\mathbb{C}}(0,1)$. Those messages are mapped into data symbols s_i, and the vector $\mathbf{s} \triangleq [s_1, \ldots, s_K]^{\mathrm{T}}$ is assumed to be known by all the TXs.

The channel from the M TXs to the K RXs is represented by the channel matrix $\mathbf{H} \in \mathbb{C}^{K \times N_T}$. Since we will make extensive use of several sub-matrices of $\mathbf{H}$, we appropriately define them in the following. Therefore, the channel coefficient from the n-th antenna of TX k to RX i is denoted as $\mathrm{h}_{i,k,n}$, and the vector of channel coefficients from TX k to RX i is represented by $\mathbf{h}_{i,k}^{\mathrm{H}} \in \mathbb{C}^{1 \times N_k}$, where $(\cdot)^{\mathrm{H}}$ denotes the conjugate transpose. For the cases in which we assume single-antenna TXs, $\mathbf{h}_{i,k}^{\mathrm{H}}$ matches $\mathrm{h}_{i,k,n}$ and thus we will denote it as $\mathrm{h}_{i,k}$. The vector of channel coefficients from all the TXs to RX i is represented by $\mathbf{h}_i^{\mathrm{H}} \in \mathbb{C}^{1 \times N_T}$. Similarly, the sub-matrix of channel coefficients from TX k towards all the RXs is denoted by $\mathbf{H}_{*,k} \in \mathbb{C}^{K \times N_k}$; in the case of single-antenna TXs, $\mathbf{H}_{*,k}$ is a vector and hence it is denoted as $\mathbf{h}_{*,k}$. Consequently, we can write the channel matrix as

$$\mathbf{H} \triangleq \begin{bmatrix} \mathbf{h}_1^{\mathrm{H}} \\ \vdots \\ \mathbf{h}_K^{\mathrm{H}} \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{H}_{*,1} & \ldots & \mathbf{H}_{*,M} \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{h}_{1,1}^{\mathrm{H}} & \ldots & \mathbf{h}_{1,M}^{\mathrm{H}} \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{K,1}^{\mathrm{H}} & \ldots & \mathbf{h}_{K,M}^{\mathrm{H}} \end{bmatrix}. \tag{2.2}$$

The channel coefficients are assumed to be drawn from a distribution with density such that all the channel sub-matrices are full rank with probability one. We will consider sharper assumptions on the channel distribution within each chapter. We assume that all the TXs are endowed with the data symbols vector $\mathbf{s}$. They precode the vector $\mathbf{s}$ with a precoder $\mathbf{T}$. The received signal at the RXs is then given by

$$\mathbf{y} \triangleq \sqrt{P}\mathbf{H}\mathbf{T}\mathbf{s} + \mathbf{n}, \tag{2.3}$$

where P is transmit power, $\mathbf{y} \triangleq [\mathrm{y}_1, \ldots, \mathrm{y}_K]^{\mathrm{T}}$ is the received signal vector and y_i is the

received signal at RX i. The vector $\mathbf{n} \in \mathbb{C}^K$ stands for the Additive White Gaussian Noise (AWGN) distributed as $\mathcal{N}_{\mathbb{C}}(0,1)$.

In a similar manner as for the channel matrix, the precoding matrix $\mathbf{T}$ can be decomposed in several sub-matrices of interest. Hence, $\mathbf{T}_k \in \mathbb{C}^{N_k \times K}$ denotes the precoding matrix applied at TX k; in case of single-antenna TXs setting, TX k applies a precoding vector and hence we denote it as $\mathbf{t}_{\mathrm{TX}\,k} \in \mathbb{C}^{1\times K}$. The global precoding vector applied to the data symbols of RX i is denoted as $\mathbf{t}_i \in \mathbb{C}^{N_T \times 1}$. The precoding vector applied at TX k for the data symbol of RX i is represented by $\mathbf{t}_{i,k}$ and, in case of single-antenna TXs setting, $\mathbf{t}_{i,k}$ is scalar and hence we denote it as $\mathrm{t}_{i,k}$. Consequently, we can write that

$$\mathbf{T} \triangleq \begin{bmatrix} \mathbf{T}_1 \\ \vdots \\ \mathbf{T}_M \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{t}_1 & \dots & \mathbf{t}_K \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{t}_{1,1} & \dots & \mathbf{t}_{K,1} \\ \vdots & \ddots & \vdots \\ \mathbf{t}_{1,M} & \dots & \mathbf{t}_{K,M} \end{bmatrix}. \tag{2.4}$$

Throughout this manuscript, we assume different power constraints for the transmit signal. In particular, we consider a average power constraint from Chapter 3 to Chapter 5, such that there exists a constant $c \in \mathbb{R}^+$ satisfying $\mathbb{E}\left[\|\mathbf{T}_k\|^2\right] \leq c$. However, for Chapter 6 and Chapter 7 we assume that the precoding vector has a per-TX instantaneous unit-norm constraint, such that

$$\|\mathbf{T}_k\| \leq c. \tag{2.5}$$

Note that, even if we set $\|\mathbf{T}_k\| = c$, the transmit power varies over the time as the power of the data symbols s_i varies. With a huge abuse of notation, and for sake of concision, we refer hereinafter to the constraint in (2.5) as *instantaneous power constraint*, although strictly speaking it is an *instantaneous power constraint on the precoding vector*. This is done in opposition to the less restrictive average power constraint on the precoder.

2.3 System Figures of Merit

We present the main metrics used throughout this manuscript to characterize and compare the performance of the different scenarios and settings. Since this thesis is aimed at the high-SNR regime, the figures of merit considered are tailored to this regime.

2.3.1 Average Rate

The main performance metric considered is the expected value of the user rate. We assume that every user $i \in \mathbb{N}_K$ wishes to receive a message $W_i \in \mathbb{W}_i$. After n channel uses, the rate $R_i(P)$ is achievable for RX i if $R_i(P) = \frac{\log|\mathbb{W}_i|}{n}$ and the probability of

wrong decoding goes to zero as n goes to infinity. Then, the expected sum rate is defined as $R(P) \triangleq \sum_{i=1}^{K} R_i(P)$. The sum capacity $\mathcal{C}(P)$ is defined as the supremum of the sum of all achievable rates [110]. In particular, under the assumption that the data symbols are i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$, the expected rate for RX i is given by

$$R_i(P) \triangleq \mathbb{E}\left[\log_2\left(1+\frac{\frac{P}{K}|\mathbf{h}_i^{\mathrm{H}}\mathbf{t}_i|^2}{1+\sum_{\ell\neq i}\frac{P}{K}|\mathbf{h}_i^{\mathrm{H}}\mathbf{t}_\ell|^2}\right)\right]. \tag{2.6}$$

In the following, we may omit the explicit reference to the transmit power P. Importantly, due to the unit-norm AWGN assumption, the average Signal-to-Noise ratio (SNR) coincides with the transmit power. Accordingly, the nominal parameter P will be likewise referred as *nominal SNR*.

Finding fundamental limits of the rate in complex multi-user systems, such as the Network MISO setting with imperfect or distributed CSIT, has been shown to be an elusive problem. For this reason, several asymptotic metrics has been widely used in the literature. Such metrics have been proven instrumental to improve the understanding of multi-user networks. We present in the following the asymptotic metrics that we consider.

2.3.2 Degrees-of-Freedom

The DoF metric, also known as multiplexing gain or pre-log factor, is defined as [111]

$$\mathrm{DoF} \triangleq \lim_{P\to\infty}\frac{\mathcal{C}(P)}{\log_2(P)}. \tag{2.7}$$

In a similar way, the DoF for a particular RX is denoted as DoF_i. Intuitively, the DoF is the first order approximation of the capacity, and it represents the *slope* of the rate as a function of the logarithm of the SNR P when P approaches to infinity. Fig. 2.3 illustrates its meaning. The DoF of a point-to-point single-antenna transmission is equal to 1. Hence, the intuition behind this metric is that a setting with a DoF=D is equivalent in the asymptotic regime to D independent point-to-point channels.

Despite the fact that DoF presents several limitations as figure of merit –we discuss them in the following–, it has been key in the characterization of complex problems, e.g. delayed CSIT [31, 112], distributed CSIT [113, 114], mixed CSIT [59, 115], Interference Alignment (IA) [40, 116], caching [94, 117–119], etc.

2.3.3 Generalized Degrees-of-Freedom

One of the weakness of DoF is that it does not take into account the network topology –we refer to network topology as the path-loss characterization of the system–. This condition comes from its limiting nature. Indeed, as P grows, the possible impact of finite path-loss differences between links vanishes. This behavior is clear by analyzing

the limit in (2.7), since any finite path-loss can be expressed as a multiplicative constant *inside the logarithm* such that it does not affect the limit. This issue is not relevant if the channel links are in the same order of magnitude. However, it implies that DoF analysis does not ensure truthful insights when the difference between channel strengths becomes significant. The following example illustrates this behavior.

> **Example 2.1.** Suppose a setting in which 2 single-antenna TXs jointly transmit towards 2 single-antenna RXs, with no CSI at the TXs. Consider that the channels TX 1-RX 1 and TX 2-RX 2 have a unit variance, whereas the channels TX 1-RX 2 and TX 2-RX 1 have variance $\gamma \in (0, 1]$ that does not depend on P.
>
> In this scenario, the DoF analysis provides that the DoF is equal to 1 independently of the value of γ [120], i.e., it only attains the DoF of a single-RX transmission. This result comes from the fact that, for any $\gamma > 0$, all the channel strengths scale proportionally to P. Nevertheless, if the value of γ is very small, for example $\gamma = 10^{-10}$, the network topology is almost equivalent to two parallel independent channels, and hence the achievable rate is almost twice the rate of a single-RX transmission. Therefore, in this example, DoF analysis does not provide a truthful characterization of the sum-rate behavior in realistic transmissions.

The fact that DoF analysis does not take the network topology into account implies that it may not be the right metric for large networks analysis [121] or unbalanced network topologies. One of the solutions to overcome this limitation is to consider the so-called locally-connected networks [122], in which the channel links are supposed to be non-zero only for a local neighborhood of the considered node. Although this assumption allows for insightful analysis for large networks, it does not avoid the problem highlighted in Example 2.1.

The Generalized Degrees of Freedom (GDoF) concept was introduced in [41] with the purpose of overcoming this limitation and taking into account the path-loss topology. GDoF is an extension of the DoF model where the path-loss is modeled as a function of the SNR. Hence, the network structure and its topology impact the analysis. Indeed, GDoF has the same definition of DoF, i.e.,

$$\text{GDoF} \triangleq \lim_{P\to\infty} \frac{\mathcal{C}(P)}{\log_2(P)}. \tag{2.8}$$

Nevertheless, the difference lies in the channel model: Let us consider an arbitrary channel coefficient $\mathrm{h}_{i,k}$. In the GDoF channel model, it is defined as

$$\mathrm{h}_{i,k} \triangleq \sqrt{P^{\gamma_{i,k}-1}}\mathrm{g}_{i,k}, \tag{2.9}$$

where $g_{i,k}$ is drawn from a distribution that does not depend on the parameter P, i.e., as $h_{i,k}$ in the previous channel model. The parameter $\gamma_{i,k} \in [0,1]$ represents the relative channel strength for the link. In particular, $\gamma_{i,k} = 1$ can be seen to be equivalent to have negligible path-loss; moreover, if $\gamma_{i,k} = 1$ for all links, we recover the DoF model. On the other side, $\gamma_{i,k} = 0$ is equivalent to have a blocked link, in the sense that any signal received from that channel link lies under the noise floor even if $P \to \infty$.

GDoF has been proven an interesting approach since optimal achievable schemes for GDoF analysis also achieve capacity within a constant number of bits [41,123,124], and it has been extensively used in the literature for characterizing complex settings [125–129]. We consider that the parameters γ_{ik} are known by anyone, since they are assumed to be long-term coefficients that vary slowly.

Remark 2.1. In the GDoF analysis, the nominal parameter P does not represent the transmit power, as for every value of P the setting changes. Conversely, it allows to group the configurations that have the same capacity when it is normalized by $\log_2(P)$. In other words, the GDoF analysis keeps constant the ratio of capacities between the channel link: A link with $\gamma_{i,k} = \frac{1}{2}$ has half of the capacity of a link with $\gamma_{i,k} = 1$, for any value of P. □

2.3.4 Affine Approximation of the Rate

Although GDoF overcomes the lack of sensitivity of DoF regarding the path-loss topology and, in some cases, it allows to achieve capacity within a known number of bits, it still undergoes the other main limitation of DoF: The metric does not provide any bounded knowledge about the achievable rate. Indeed, the definition of DoF in (2.7) implies that the rate can be written as

$$R = \mathrm{DoF} \log_2(P) + o\big(\log_2(P)\big), \tag{2.10}$$

and the term $o\big(\log_2(P)\big)$ is not bounded. In Fig. 2.3, we show how two settings with the same DoF –which represents the *slope of the rate*– can achieve considerably different rates. Nevertheless, the metric can be refined to offer results of achievable rate with a bounded gap. In particular, we consider the affine approximation of the rate at high SNR, introduced in [130]. According to this approximation, the achievable rate can be expressed as [130]

$$R = \mathrm{DoF} \log_2(P) - \mathcal{R}_\infty + o(1), \tag{2.11}$$

where $\mathcal{R}_\infty$ denotes the *rate offset* –or vertical offset–. The approximation in (2.11) can also be written in terms of the *power offset* –horizontal offset– $\mathcal{L}_\infty$, where $\mathcal{R}_\infty = \mathrm{DoF}\, \mathcal{L}_\infty$.

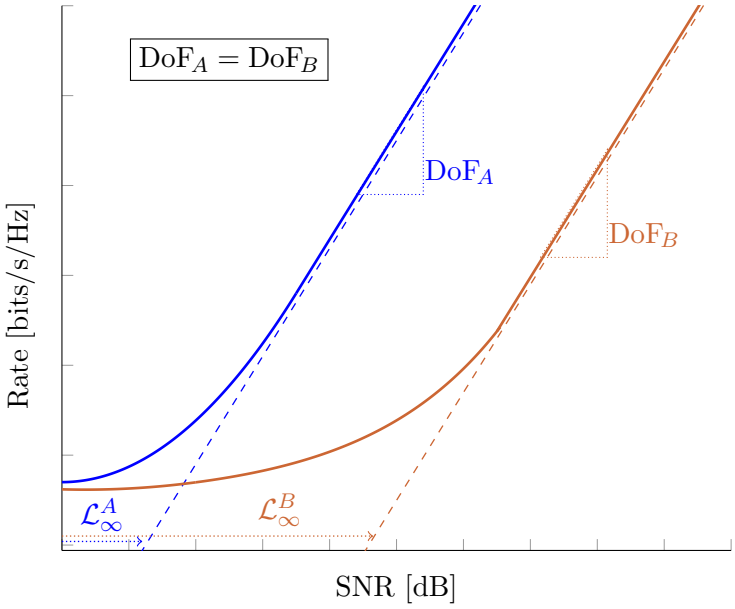


Figure 2.3 – Qualitative illustration of the affine approximation of two different setting with the same DoF (slope) but different rate offset $\mathcal{R}_\infty = \mathrm{DoF}\,\mathcal{L}_\infty$.

An illustrative visualization is shown in Fig. 2.3. The term $\mathcal{L}_\infty$ represents the zero-order term with respect to a reference setting with the same slope but whose affine approximation intersects the origin. Note that the rate offset is defined as

$$\mathcal{R}_\infty \triangleq \lim_{P\to\infty} \mathrm{DoF}\log_2(P) - R(P), \tag{2.12}$$

where $R(P)$ represents the rate as function of the SNR P. This measure has shown instrumental in several findings. In [29], Lozano *et al.* analyze the multiple-antenna point-to-point scenario, revealing that some system features which do not impact the DoF (as antenna correlation, fading...) do considerably impact the zero-order term, affecting the performance of the system at any possible SNR. In addition to expose the limitations of having only information about the DoF, [29] also reveals that the affine expansion offers appreciably tight approximations also at medium-to-low SNR. This characterization has been also established for the Broadcast Channel (BC) with perfect CSIT using Dirty-paper coding and linear precoding [131], and for the BC with imperfect CSIT [132]. In [132], the quantized feedback scenario was studied under the assumption of Zero-Forcing (ZF) schemes, showing that the channel-to-estimation-noise ratio must be proportional to SNR^α in order to attain a DoF per user of $\mathrm{DoF}_{\mathrm{RX}\,i} = \alpha$. Furthermore, having a ratio of SNR^α was shown to be equivalent to send a quantized

feedback of $\alpha \log_2(\text{SNR})$ bits, what could be attained if the feedback resources scale as the capacity of the channel. In the same vein as in [29], this approximation is expected to be adequate also to characterize the performance below the high-SNR regime for a broad set of configurations and settings.

2.4 Distributed CSIT Model

The main particularity of this thesis is the consideration that the TXs do not share perfectly their CSI. In this section, we present the general mathematical model and the assumptions that hold throughout the entire thesis. This scenario can be seen as a multi-agent cooperative decision with common goal, where each node knows the structure of the system but not the information that the others have received [73]. Further details and considerations are included in the corresponding chapters. We start by introducing the imperfect CSIT model for the centralized CSIT setting.

2.4.1 Centralized CSIT Model

In the Centralized CSIT setting (C-CSIT), there is a single estimate of the channel matrix $\mathbf{H} \in \mathbb{C}^{K \times N_T}$, shared by all the TXs. We denote the estimate of a certain channel submatrix with a hat $(\hat{\cdot})$, i.e., $\hat{\mathbf{H}}$, $\hat{\mathbf{h}}_i$, $\hat{\mathbf{H}}_{*,k}$, denote the estimate of $\mathbf{H}$, $\mathbf{h}_i$, $\mathbf{H}_{*,k}$, respectively. Let us first consider a single channel coefficient for a single-antenna TX setting. Then, the imperfect CSIT assumption is modeled such that

$$\hat{\mathrm{h}}_{i,k} \triangleq \sqrt{1 - Z_{i,k}}\, \mathrm{h}_{i,k} + \sqrt{Z_{i,k}}\, \delta_{i,k}, \tag{2.13}$$

where $\delta_{i,k}$ is the additive noise variable and Z denotes the variance scaling of that noise. Let $\mathbf{1}_{n \times m}$, $\mathbf{0}_{n \times m}$, denote respectively the all-ones matrix and the all-zeros matrix of size $n \times m$. Based on (2.13), the channel matrix estimate can be written as

$$\hat{\mathbf{H}} \triangleq \sqrt{\mathbf{1}_{K \times N_T} - \mathbf{Z}} \odot \mathbf{H} + \sqrt{\mathbf{Z}} \odot \mathbf{\Delta}, \tag{2.14}$$

where $\odot$ represents the Hadamard –element-wise– product and $\mathbf{\Delta}$ is a noise random matrix that encloses the additive estimation noise and whose covariance matrix is bounded. The variance scaling of the estimation noise is provided by the deterministic matrix $\mathbf{Z}$. Thus, $\mathbf{Z}$ encloses the average accuracy of the estimates. Intuitively, if $\mathbf{Z} = \mathbf{0}_{K \times N_T}$, the estimate is perfect, since $\hat{\mathbf{H}} = \mathbf{H}$. Conversely, if $\mathbf{Z} = \mathbf{1}_{K \times N_T}$, the estimate is composed only of random noise. We further define the i-th row of $\mathbf{\Delta}$ as $\boldsymbol{\delta}_i$, such that $\mathbf{\Delta} \triangleq [\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_K]^{\mathrm{T}}$. This setting provides a great generality, as it comprises cases where each channel coefficient is known with a different accuracy and with any possible correlation, incorporating any

heterogeneous scenario.

2.4.2 Distributed CSIT Model

The Distributed CSIT (D-CSIT) model is characterized by the consideration that each TX is endowed with a possibly different estimate. Thus, the key singularity of this setting is that, for any channel coefficient, there exist as many estimates as TXs, each one of them *locally available at a single TX*. The extension from the centralized model in (2.14) is direct. Let us denote the estimate at TX j as $\hat{\mathbf{H}}^{(j)}$[1]. Then, $\hat{\mathbf{H}}^{(j)}$ is defined as

$$\hat{\mathbf{H}}^{(j)} \triangleq \sqrt{\mathbf{1}_{K\times N_T} - \mathbf{Z}^{(j)}} \odot \mathbf{H} + \sqrt{\mathbf{Z}^{(j)}} \odot \mathbf{\Delta}^{(j)}. \tag{2.15}$$

Hence, each TX has a different estimation noise ($\mathbf{\Delta}^{(j)}$) with a a different power ($\mathbf{Z}^{(j)}$). Furthermore, we define the distributed estimate counterpart of the different sub-matrices defined in Section 2.2 consistently. For example, the estimate of the vector channel towards RX i ($\mathbf{h}_i$) is denoted as $\hat{\mathbf{h}}_i^{(j)}$.

Remark 2.2. It is critical to understand well how the *Distributed* CSIT setting differs from the many different *heterogeneous* CSIT configurations studied in the literature. Indeed, a heterogeneous CSIT configuration typically refers to a *centralized* CSIT setting (i.e., with a channel estimate common to all TXs), where each element of the channel is known with a different quality owing to specific feedback mechanisms [49, 133–135]. In contrast, the distributed setting considered here has as many different channel estimates as TXs, where each TX does not have access to the CSIT knowledge at the other TXs. □

2.4.3 Estimation Noise Scaling

The main reason why we consider the model of (2.15) with the variance of each noise coefficient explicitly indicated with the element-wise product is that our analysis is mainly dependent of the scaling of that variance. In particular, we only require from $\mathbf{\Delta}^{(j)}$ that it is a random variable with bounded covariance matrix and density (although further assumptions are considered for certain chapters). In turn, we analyze the impact of the value $\mathbf{Z}^{(j)}$ when it is modeled as a function of the SNR. It is known that, for the centralized CSIT case, the signal-to-noise ratio of the estimate should scale as P^{α}, with $\alpha > 0$, in order to avoid the collapse of the multiplexing gain [50, 132]. Then, we consider an exponential scaling with P also for the D-CSIT setting. Consequently, we assume

[1] Henceforward, and as a general rule, we consistently use the sub-index i to refer to RXs, the sub-index k to refer to channels or parameters asociated to TXs, and the super-index $^{(j)}$ to indicate *where* the term *is known*. Hence, $x_k^{(j)}$ denotes the estimate at TX j of a parameter x_k of TX k (channel coefficient, power normalization, etc.).

that the matrix $\mathbf{Z}^{(j)}$ is defined such that its (i,k) coefficient is given by

$$\mathbf{Z}_{i,k}^{(j)} = P^{-\alpha_{i,k}^{(j)}}, \tag{2.16}$$

where $0 \leq \alpha_{i,k}^{(j)} \leq 1$. The coefficient $\alpha_{i,k}^{(j)}$ is the *accuracy scaling* parameter that measures the quality of estimation of the channel matrix at TX j. The CSI accuracy at the TXs is characterized throughout this thesis with these parameters $\alpha_{i,k}^{(j)}$. We define the set of *accuracy scaling* parameters as $\boldsymbol{\alpha}$, such that

$$\boldsymbol{\alpha} \triangleq \{\alpha_{i,k}^{(j)}\}_{i\in\mathbb{N}_K,\ j,k\in\mathbb{N}_M}. \tag{2.17}$$

The parameters $\alpha_{i,k}^{(j)}$ are assumed to be long-term coefficients that vary slowly. Based on that, it is assumed that every TX knows the full set $\boldsymbol{\alpha}$, as it only requires to share few bits over a long period of time. Moreover, we will refer to the TX with the greatest parameter $\alpha_{i,k}^{(j)}$ as the "most-informed" or the "best-informed" TX. In the following, we present some intuitions and justifications of the exponential model.

Limited Feedback

This scaling model arises when we consider a setting in which the RXs quantize their perfect channel information and feed that quantization back to the TXs. Jindal demonstrated in [132] that, in the (centralized) MISO BC setting, if the channel vector is quantized with B bits, the estimation noise variance scales as $2^{-\frac{B}{M-1}}$. Hence, letting the number of quantization bits scale as $B = (M-1)\alpha\log_2(P)$, the estimation noise scaling becomes

$$2^{-\frac{B}{M-1}} = P^{-\alpha}, \tag{2.18}$$

which matches the model in (2.16). Note that the assumption that the number of bits is proportional to $\log(P)$ is a feasible assumption since it is equivalent to say that the feedback rate scales linearly with the capacity of the channel link.

Multiplexing Gain

It is known that, if the noise variance does not scale as $P^{-\alpha}$, for $\alpha > 0$, the multiplexing gain (DoF) of the setting is lost with respect to the case with no CSIT [120, 132]. Since our analysis is focused on the high-SNR regime, this CSIT accuracy model enables us to obtain a comprehensive characterization of the network behavior in the asymptotic regime. On this basis, the parameter $\alpha_{i,k}^{(j)}$ can be restricted to the interval $[0, 1]$. This is due to the fact that, with $\alpha_{i,k}^{(j)} = 1$, the estimation error lies on the noise floor –since it

has a scaling that does not grow with P– and hence it is negligible in the asymptotic regime. In particular, for the GDoF channel model, $\alpha_{i,k}^{(j)}$ can be restricted to $[0, \gamma_{i,k}]$, for the reason previously exposed.

2.4.4 Sorted CSIT Setting

There exists a particular case that is of significant relevance on its own: The CSIT configuration in which the TXs can be ordered by their CSIT accuracy. In this configuration, so-called *sorted CSIT setting*, we can write w.l.o.g. that

$$1 \geq \alpha_{i,k}^{(1)} \geq \alpha_{i,k}^{(2)} \geq \cdots \geq \alpha_{i,k}^{(M)} \geq 0. \tag{2.19}$$

For the sake of concision and readability, we consider also a simplified version of this setting in which a given TX has the same *accuracy scaling* parameter for any channel coefficient, i.e., that $\alpha_{i,k}^{(j)} = \alpha^{(j)}$, $\forall i \in \mathbb{N}_K, k \in \mathbb{N}_M$. Hence, (2.19) becomes

$$1 \geq \alpha^{(1)} \geq \alpha^{(2)} \geq \cdots \geq \alpha^{(M)} \geq 0, \tag{2.20}$$

which implies that TX 1 is the *best-informed* TX, i.e, whose CSI has the highest accuracy.

This model encloses for example a scenario in which a main, multi or massive antenna base station serves a set of users with the help of some single or multi antenna remote radio-heads or simple TXs, as depicted in Fig. 2.2b. Moreover, it can be seen as a simplification for a more decentralized setting, in which each TX obtains the channel information of its attached users. Then, a TX gathers the information from all the other TXs but, due to the tight latency constraints or the limited capacity of the backhaul links, can only send back a compressed version of the channel matrix. This configuration is represented by Fig. 2.2a.

2.4.5 Hierarchical CSIT Model

Another important case which belongs to the D-CSIT setting is the so-called Hierarchical CSIT (H-CSIT) setting: Consider the D-CSIT setting with M TXs in which each TX owns a different channel estimate of the coefficient $\mathrm{h}_{i,k}$, denoted as $\hat{\mathrm{h}}_{i,k}^{(j)}$. The setting follows a Hierarchical CSIT configuration if and only if each TX knows the estimate at the TXs whose estimate is less accurate. That is, for each channel coefficient, we can order the TXs such that TX 1 is the best-informed TX and TX M is the TX with the least accurate estimate. Then, the estimate of TX $j+1$ is *included* in the information available at TX j such that TX j knows

$$\hat{\mathrm{h}}_{i,k}^{(M)}, \ \ldots, \ \hat{\mathrm{h}}_{i,k}^{(j+1)}, \ \hat{\mathrm{h}}_{i,k}^{(j)}. \tag{2.21}$$

This CSIT structure appears in many heterogeneous networks, in which e.g. a main TX shares with the other TXs its channel estimate but has to compress it to transmit through a rate-limited link, or in cases in which there exits multi-level quantization schemes. Several works in the literature have focused on such setting [83, 136–138]. with the aftermath that the explicit structure of the CSIT enables important gains with respect to the fully distributed setting.

2.4.6 CSIR Model

In this work, we focus on the impact of the imperfect CSI on the TX side as the CSI acquisition is widely acknowledged to be more challenging on the TX side than on the RX side. For example, in FDD, due to the need to feed back the CSI that has been estimated at the RX towards the TX. Therefore, we consider that every RX has perfect knowledge of its own channel.

2.5 Genie-aided Centralized CSIT Setting

Finding purely distributed upper-bounds is a challenging subject that remains open for most of the settings. We tackle this problem in Chapter 5. However, any decentralized scenario with distributed estimates has an *ideal* centralized counterpart in which a genie provides the best estimate of each parameter to every node. Based on that, we introduce the notion of "genie-aided centralized scenario" that will be used all over this thesis. A genie-aided centralized scenario is a C-CSIT setting –in which all the TXs are endowed with the same CSI– that is obtained from a D-CSIT setting by means of providing the TXs with CSI available at other TXs. We must note that we consider two possible genie-aided settings, each one for a different part of this thesis.

1. The first genie-aided scenario, considered for the DoF analysis, is such that each TX shares its CSI with any other TX. Thus, every TX owns the set of M estimates.

2. The second genie-aided scenario, less loosened, considers that all the TXs are endowed only with the estimate of best average accuracy.

This ideal setting provides us with a benchmark for the performance on the D-CSIT setting. In this way, we are able to analyze which is the impact of having distributed information or, in other words, *the cost of not sharing the CSI*. In general, the genie-aided centralized setting is represented by a MISO BC setting with N_T transmit antennas.

Remark 2.3. It is important to observe that in the first genie-aided model every TX owns the set of M estimates of the M TXs, whereas in the second one each TX owns

only the best estimate among all the TXs, *instead of* its own estimate. The later model allows a fairer comparison between the D-CSIT and the C-CSIT scenarios. The former would benefit from the fact that the knowledge of M estimates allows to reduce the noise variance by a factor proportional to M. However, it proves useful as achieving the performance of such an ideal setting strengthens the results of the D-CSIT setting. □

2.6 Asymptotic Notation

This thesis is mainly focused on the asymptotic analysis of the rate in the high-SNR regime. In the interest of clarity, we specify in the following the notation considered to express certain asymptotic properties. In particular, we base our notation in the prevalent Bachmann–Landau notation [139]. As lucidly summarized in [140], mathematicians and researchers have applied different notations to refer to the asymptotic behavior or scaling of functions, e.g. relational notation as $\prec, \succ, \asymp$ or the dot-notation $\dot{=}, \dot{\leq}$, but also set-notations as the Bachmann–Landau notation. In this thesis, we make use of the later because it can be placed inside the mathematical derivations and offers a more flexible handling. Specifically, we follow the limit interpretation given below [140].

Definition 2.1. Let g be a real valued function. Let f be a real or complex valued function. One writes[2]

$$f(x) = \mathcal{O}(g(x)) \tag{2.22}$$

if and only if $\limsup_{x\to\infty} \frac{|f(x)|}{g(x)} < \infty$.

Definition 2.2. Let g be a real valued function. Let f be a real or complex valued function. One writes

$$f(x) = o(g(x)) \tag{2.23}$$

if and only if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.

Note that $f(x) = o(g(x)) \Rightarrow f(x) = \mathcal{O}(g(x))$, but the converse is not true. We present in the following another asymptotic notation that is disjoint with $o(g(x))$ and which was

[2] The "=" sign in $f(x) = \mathcal{O}(g(x))$ is an abuse of notation that actually means $f(x) \in \mathcal{O}(g(x))$. Indeed, it is a one-way equality [140] since $\mathcal{O}(x) = \mathcal{O}(x^2)$ but $\mathcal{O}(x^2) \neq \mathcal{O}(x)$. This "=" notation has been extensively used in the literature and we use it for convenience. However, its exact meaning must be clear.

firstly introduced by Knuth in [140].

Definition 2.3. Let g be a real valued function. Let f be a real or complex valued function. One writes

$$f(x) = \Theta(g(x)) \tag{2.24}$$

if and only if there exist positive constants $C_1 > 0$, $C_2 < \infty$, such that

$$\limsup_{x\to\infty} \frac{|f(x)|}{g(x)} < C_2 < \infty, \qquad \text{and} \qquad \liminf_{x\to\infty} \frac{f(x)}{g(x)} > c_1. \tag{2.25}$$

The previously enunciated O-notations are defined for limits of function when their arguments approaches infinity. Besides this notation, we introduce also equivalent asymptotic notations for random variables –see [141] for a detailed discussion about the different probabilistic versions of asymptotic notations–. This notation will be very useful throughout the rest of the manuscript.

Definition 2.4. Let χ be a random variable with probability density function that depends on a parameter θ and is denoted by $f_{\chi|\theta}$. Let g be a real valued function. Consider the random variable χ_∞ whose probability density function is given by

$$f_{\chi_\infty}(x) \triangleq \lim_{\theta\to\infty} f_{\chi|\theta}\left(\frac{x}{g(\theta)}\right). \tag{2.26}$$

One writes

$$\chi = \Theta_p(g(\theta)) \tag{2.27}$$

if and only if the limiting probability density function is bounded, i.e.,

$$\max_x f_{\chi_\infty}(x) = f_\chi^{\max} < \infty, \tag{2.28}$$

and there exist positive constants $C_1 > 0$, $C_2 < \infty$, such that

$$C_1 < \mathbb{E}[\|\chi_\infty\|^2] < C_2. \tag{2.29}$$

The notation $\Theta_p(g(\theta))$ is the equivalent of $\Theta(f(x))$ for random variables. Intuitively, it implies that the random variable χ can be decomposed as

$$\chi = g(\theta)\tilde{\chi}, \tag{2.30}$$

such that the random variable $\tilde{\chi}$ is a "normalized" variable for which the peak of the Probability Density Function (PDF) does not scale with θ. As a matter of example, consider a transmission of data symbols drawn from a complex Gaussian random variable $h \sim \mathcal{N}_{\mathbb{C}}(0,1)$. Suppose that we transmit those data symbols with a power P, such that the transmitted signal x is given by

$$x \triangleq \underbrace{\sqrt{P}}_{g(P)} \underbrace{h}_{\tilde{x}}. \tag{2.31}$$

Then, it follows that $x = \Theta_{\mathsf{p}}(\sqrt{P})$. Moreover, Definition 2.4 leads to the following corollary.

Corollary 2.1. *Let $x = \Theta_{\mathsf{p}}(\sqrt{P})$. Then,*

$$\mathbb{E}[\|x\|^2] = \Theta(P). \tag{2.32}$$

Corollary 2.1 sets a relation between both functional and probabilistic notations (Definition 2.3 and Definition 2.4, respectively). Therefore, a random variable in $\Theta_{\mathsf{p}}(y)$ has an expected squared norm in $\Theta(y^2)$. We similarly define the probabilistic equivalent of $\mathcal{O}(f(x))$ in Definition 2.1.

Definition 2.5. Let χ be a random variable with probability density function that depends on a parameter θ and is denoted by $f_{\chi|\theta}$. Let g be a real valued function. Consider the random variable χ_∞ whose probability density function is given by

$$f_{\chi_\infty}(x) \triangleq \lim_{\theta\to\infty} f_{\chi|\theta}\left(\frac{x}{g(\theta)}\right). \tag{2.33}$$

One writes

$$\chi = \mathcal{O}_{\mathsf{p}}(g(\theta)) \tag{2.34}$$

if and only if the limiting probability density function is bounded, i.e., if $f_\chi^{\max} \triangleq \max_x f_{\chi_\infty}(x)$ satisfies that $f_\chi^{\max} < \infty$, and there exists a positive constant $C_1 < \infty$, such that

$$\mathbb{E}[\|\chi_\infty\|^2] < C_1. \tag{2.35}$$

Together with the previous notations, it is normally assumed that $f(x) \sim g(x)$ denotes that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$. However, throughout this manuscript we use $\sim$ to denote that a

random variable has a certain distribution, i.e., we write $X \sim F$ if and only if the random variable X is distributed as F. Besides the standard O-notations and their probabilistic counterparts, we make use of another asymptotic notation that will prove useful in the DoF analysis. This last notation is a relaxation of Definition 2.3.

Definition 2.6. Let g be a real valued function. Let f be a real or complex valued function. One writes

$$f(x) = \Theta_{\log}(g(x)) \tag{2.36}$$

if and only if

$$\lim_{x\to\infty} \frac{\log(|f(x)|)}{\log(g(x))} = 1. \tag{2.37}$$

Why is this last notation interesting? As previously mentioned, it can be seen as a relaxation of $\Theta\big(g(x)\big)$, i.e.,

$$\Theta_{\log}\left(g(x)\right) \subset \Theta\big(g(x)\big). \tag{2.38}$$

It turns out that, for DoF analysis, it is enough to satisfy $\Theta_{\log}\left(g(x)\right)$. As a matter of example, let us denote the expected SINR as $f(P) = \mathbb{E}[\text{SINR}]$ and suppose that $f(P) = \Theta_{\log}(P^d)$, for any $0 \leq d \leq 1$. By upper-bounding the rate through Jensen's inequality, the DoF is given by

$$\lim_{P\to\infty} \frac{\log_2(1 + f(P))}{\log_2(P)} = d. \tag{2.39}$$

From Definition 2.6, the assignment $f(P) = \Theta_{\log}(P^d)$ is equivalent to

$$\lim_{P\to\infty} \frac{\log(f(P))}{\log(P^d)} = 1. \tag{2.40}$$

Hence, we can see that (2.39) and (2.40) are equivalent expressions. For this reason, in the remaining of the document, a scenario in which the SINR term scales as $\Theta_{\log}(P^d)$ can be understood as having a DoF $= d$.

Part II

DoF & GDoF Analysis

Chapter 3

GDoF Analysis of the 2x2 Distributed Network MISO

In this chapter we analyze the GDoF of the 2×2 setting under the D-CSIT assumption. It is known that, in the specific case where each TX has an homogeneous estimate of the whole channel matrix (such that every channel coefficient is estimated at a given TX with the same average accuracy), the DoF of the D-CSIT setting matches the DoF of the genie-aided C-CSIT setting [82]. In this chapter, we extend the analysis to the general case with arbitrary accuracy parameters and arbitrary path-loss scaling, so as to answer the question:

For the 2×2 setting, when does the D-CSIT setting achieve the GDoF of the C-CSIT?

This generalization of the initial result of [82] is far from trivial, as the outcomes could be due to several particularities of the setting assumed in [82]; namely, the master-slave-type configuration –with one TX having a better information about the whole system–, the homogeneous accuracy assumption, or the absence of path-loss differences.

3.1 Preliminaries

As previously mentioned, Etkin *et al.* introduced the GDoF metric in [41]. The GDoF is a generalization of the DoF metric that provides a finer characterization of the setting

capacity, offering in some cases capacity results within a constant number of bits [41]. Among the vast literature on GDoF analysis, one work is of particular consideration because of its complementarity with this chapter. Davoodi, Yuan, and Jafar presented in [50] the GDoF of the MISO BC with imperfect, yet centralized[1], CSIT. Thus, we can obtain from [50] the GDoF of the genie-aided C-CSIT setting described in Section 2.5 that we employ as reference setting.

3.1.1 System Model

The transmit signal is defined as in Section 2.2. We recall that $\mathbf{h}_i^{\mathrm{H}} \triangleq [\mathrm{h}_{i,1},\ \mathrm{h}_{i,2}]$ denotes the multi-TX channel to RX i and $\mathrm{h}_{i,k}$ denotes the fading channel coefficient from TX k to RX i. The transmit signal $\mathbf{x} = \bar{P}\mathbf{T}\boldsymbol{s} \in \mathbb{C}^{2\times 1}$ fulfills the average power constraint $\mathbb{E}[\|\mathbf{x}\|^2] = P$. Following the GDoF model of Section 2.3.3, the channel coefficients are defined as

$$\mathrm{h}_{i,k} \triangleq \bar{P}^{\gamma_{i,k}-1}\mathrm{g}_{i,k}, \tag{3.1}$$

where P is the nominal SNR parameter. The parameter $\gamma_{i,k} \in [0,1]$ is the relative channel strength exponent between TX k and RX i. Finally, the normalized channel parameters $\mathrm{g}_{i,k}$ are mutually independent and drawn from a generic (in the sense that any matrix formed by i.i.d. elements according to this distribution will be full rank) continuous distribution with density and whose density peak does not scale with P.

3.1.2 Distributed CSIT Model

The D-CSIT model is slightly simplified with respect to the general model described in Section 2.4. In particular, instead of defining the estimate of $\mathrm{g}_{i,k}$ at TX j as (2.13), i.e.,

$$\hat{\mathrm{g}}_{i,k}^{(j)} \triangleq \sqrt{1 - Z_{i,k}^{(j)}}\ \mathrm{g}_{i,k} + \sqrt{Z_{i,k}^{(j)}}\ \delta_{i,k}^{(j)}, \tag{3.2}$$

we omit the term $\sqrt{1-Z}$, as it does not impact the GDoF metric due to the fact that the term is $\Theta(1)$. Moreover, we model the noise variance scaling as $Z_{i,k}^{(j)} = P^{-\alpha_{i,k}^{(j)}}$. Hence, the channel estimate at TX j is written as

$$\hat{\mathrm{g}}_{i,k}^{(j)} \triangleq \mathrm{g}_{i,k} + \bar{P}^{-\alpha_{i,k}^{(j)}}\delta_{i,k}^{(j)}, \tag{3.3}$$

and $\hat{\mathrm{h}}_{i,k}^{(j)} \triangleq \bar{P}^{\gamma_{i,k}-1}\hat{\mathrm{g}}_{i,k}^{(j)}$. This simplified model is the conventional DoF model used in the literature [54, 58, 59, 120, 142] to shape the dependency of the CSIT accuracy as a function of the SNR. The estimation noise terms $\delta_{i,k}^{(j)}$ are drawn from a generic continuous

[1] Centralized refers to a *logically* centralized setting where all the TXs have access to the same, possibly imperfect, CSI.

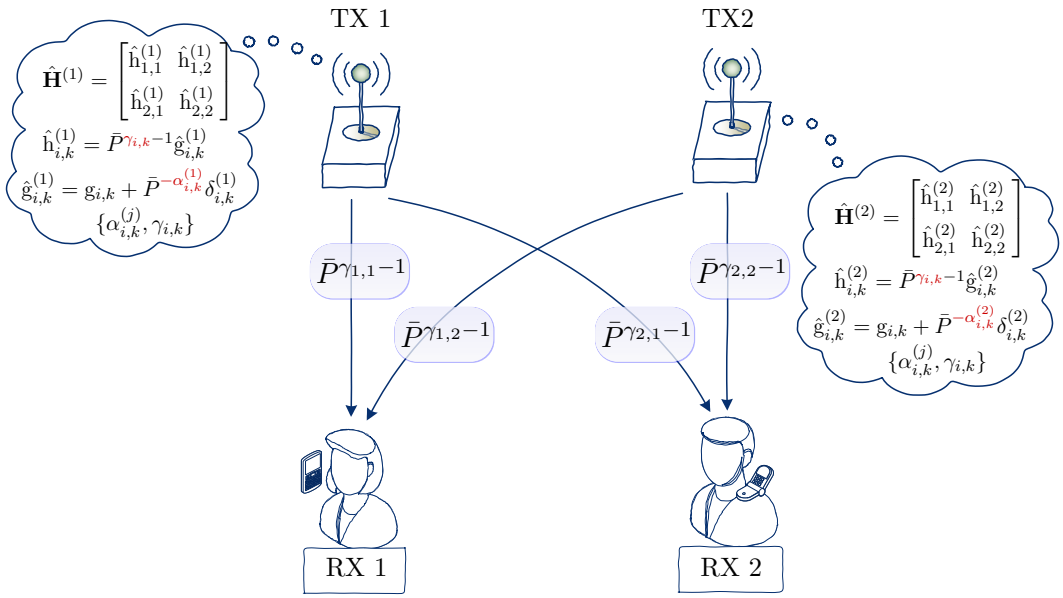


Figure 3.1 – 2×2 Network MISO with distributed CSIT. The "bubble" contains the information available at each TX. Note that the accuracy of the estimate can be different from TX to TX and from one link to another. The parameters $\gamma_{i,k}$ represent the different path-loss scaling at each link.

distribution with density, independent of P, and independent from TX to TX. We limit the values of the *CSIT quality exponent* to $\alpha^{(j)}_{i,k} \in [0, \gamma_{i,k}]$. Note that, in terms of GDoF, an estimate with *CSIT exponent* $\alpha^{(j)}_{i,k} = \gamma_{i,k}$ can be intuitively understood as being perfect [50], since the error generated by the estimation noise lies on the noise floor. Conversely, $\alpha^{(j)}_{i,k} = 0$ is intuitively understood as being useless from a GDoF perspective.

Our upper-bound analysis of the D-CSIT setting makes use of the results of Davoodi and Jafar for the C-CSIT setting in [50]. Consequently, we recall here the Bounded Density assumption usually considered in the Aligned Image Set approach [50, 120].

Definition 3.1 (Bounded Density Coefficients [120])**.** A set of random variables, $\mathcal{A}$, is said to satisfy the bounded density assumption if there exists a finite positive constant $f_{\max}$,

$$0 < f_{\max} < \infty, \tag{3.4}$$

such that for all finite cardinality disjoint subsets $\mathcal{A}_1$, $\mathcal{A}_2$ of $\mathcal{A}$, with $\mathcal{A}_1 \subset \mathcal{A}$, $\mathcal{A}_2 \subset \mathcal{A}$, $\mathcal{A}_1 \cap \mathcal{A}_2 = \emptyset$, $|\mathcal{A}_1| < \infty$, $|\mathcal{A}_2| < \infty$, the conditional probability density functions exist and are bounded as follows

$$\forall A_1, A_2, \quad f_{\mathcal{A}_1|\mathcal{A}_2}(A_1|A_2) \leq f_{\max}^{|\mathcal{A}_1|}. \tag{3.5}$$

We assume hereinafter that all channel realizations $g_{i,k}$ and estimation noise variables $\delta_{i,k}^{(j)}$ satisfy the bounded density property. This assumption excludes the cases where a channel coefficient is function of others or the CSIT is perfectly known, and it is equivalent to the restriction in [143] that the differential entropy must be greater than $-\infty$ [120].

3.1.3 Review of the Results for the Centralized CSIT Setting

As mentioned in Section 2.5, the setting in which the CSIT is perfectly shared between the TXs plays a mayor role in our analysis of the upper-bound as a reference setting with which we can compare the performance. This comparison allows us to bring out the impact of having discrepancies between TXs. In this centralized setting all the TXs share the exact same, potentially imperfect, channel estimate. Hence, there is a single channel estimate and we can remove the TX index and consider just $\hat{\mathbf{H}}$. The GDoF of the 2-user MISO BC with centralized CSIT has been derived in [50]. We state in the following their main result for sake of completeness.

Theorem 3.1. *[50] In the 2-user MISO BC with centralized CSIT, the optimal sum GDoF, denoted as* $\mathrm{GDoF}^{\mathrm{CCSIT}}(\boldsymbol{\alpha})$*, satisfies*

$$\mathrm{GDoF}^{\mathrm{CCSIT}}(\boldsymbol{\alpha}) = \min(D_1, D_2), \tag{3.6}$$

where we have defined D_1 *and* D_2 *as*

$$D_1 \triangleq \max\left(\gamma_{1,2}, \gamma_{1,1}\right) + \max\left(\left(\gamma_{2,1} - \gamma_{1,1} + \alpha_1\right)^+, \left(\gamma_{2,2} - \gamma_{1,2} + \alpha_1\right)^+\right), \tag{3.7}$$

$$D_2 \triangleq \max\left(\gamma_{2,2}, \gamma_{2,1}\right) + \max\left(\left(\gamma_{1,1} - \gamma_{2,1} + \alpha_2\right)^+, \left(\gamma_{1,2} - \gamma_{2,2} + \alpha_2\right)^+\right), \tag{3.8}$$

with the short-hand notations

$$\alpha_1 \triangleq \min\left(\alpha_{1,1}, \alpha_{1,2}\right), \tag{3.9}$$

$$\alpha_2 \triangleq \min\left(\alpha_{2,1}, \alpha_{2,2}\right). \tag{3.10}$$

Interestingly, depending on the network geometry, the path-loss can be either advantageous (since they reduce the interference power received) or detrimental (since they reduce the intended signal power received in the same level that the interference). Moreover, the GDoF performance only depends on the weakest CSIT parameter for each receiver.

Remark 3.1. This optimal sum GDoF is achieved by superposition coding, rate splitting and ZF precoding [132, 144]. A detailed discussion about the GDoF expression in (3.6)-(3.10) is provided in [50]. □

3.2 GDoF of the Distributed CSIT Setting

3.2.1 Centralized Upper-Bound

We can obtain an intuitive upper-bound for the D-CSIT setting by assuming a genie-aided centralized setting in which each TX has access to the estimates of *all* the TXs.

Lemma 3.1. *In the 2×2 D-CSIT single-antenna Network MISO, the optimal GDoF is upper-bounded by the GDoF of a C-CSIT scenario in which the channel estimates of both TXs are perfectly shared, such that each TX accesses to the set of estimates $\{\hat{\mathbf{H}}^{(1)},\hat{\mathbf{H}}^{(2)}\}$. Let us define the set of accuracy scaling parameters of the D-CSIT and C-CSIT settings as*

$$\boldsymbol{\alpha} \triangleq \{\alpha_{i,k}^{(j)}\}_{i,j,k\in\mathbb{N}_2} \qquad \textit{and} \qquad \boldsymbol{\alpha}^{\star} \triangleq \{\max_{j\in\mathbb{N}_2} \alpha_{i,k}^{(j)}\}_{i,k\in\mathbb{N}_2}, \tag{3.11}$$

respectively. Then,

$$\mathrm{GDoF}^{\mathrm{DCSIT}}(\boldsymbol{\alpha}) \leq \mathrm{GDoF}^{\mathrm{CCSIT}}(\boldsymbol{\alpha}^{\star}). \tag{3.12}$$

Proof. It is clear that the genie-aided C-CSIT setting is an upper-bound of the D-CSIT setting, as providing with extra information can not hurt the performance. It remains to prove that the GDoF of the genie-aided CSIT setting is given by $\mathrm{GDoF}^{\mathrm{CCSIT}}(\boldsymbol{\alpha}^{\star})$. Such setting corresponds to a (logically) centralized scenario with a shared CSI composed by $\{\hat{\mathbf{H}}^{(1)},\hat{\mathbf{H}}^{(2)}\}$. The particularity of having several estimates at a single node is a novel assumption that is not contemplated in the literature. However, the following proposition allows us to link this particular setting with the commonly used centralized setting.

Proposition 3.1. *Let $\hat{\mathrm{h}}^{(1)}$ and $\hat{\mathrm{h}}^{(2)}$ be two random variables defined as $\hat{\mathrm{h}}^{(j)} \triangleq \mathrm{h}+\bar{P}^{-\alpha^{(j)}}\delta^{(j)}$, where h and $\delta^{(j)}$ are independent continuous random variables satisfying the Bounded Density assumption of Definition 3.1. Then, the conditional probability density function $f_{\mathrm{h}|\hat{\mathrm{h}}^{(1)},\hat{\mathrm{h}}^{(2)}}$ satisfies that*

$$\max_{\mathrm{h}} f_{\mathrm{h}|\hat{\mathrm{h}}^{(1)},\hat{\mathrm{h}}^{(2)}} = \Theta\big(\bar{P}^{\max(\alpha^{(1)},\alpha^{(2)})}\big). \tag{3.13}$$

The proof of Proposition 3.1 is relegated to Appendix D for clarity. Actually, this proposition is a particular case of Lemma 4.1, which covers the general $K\times K$ setting

and will be presented in Chapter 4. Using Proposition 3.1, we obtain that the peak of the probability density function of this genie-aided scenario with multiple estimates has the same scaling as the centralized setting with only a single estimate $\hat{\mathbf{H}}^{\star}$, whose accuracy scaling coefficients are given by $\boldsymbol{\alpha}^{\star}$ in (3.11). It was shown in [120, Section V.8] that the DoF is characterized by the peak of the probability density function, and hence we obtain Lemma 3.1. ■

3.2.2 Distributed Lower-Bound

We have introduced in the previous section an intuitive upper-bound for the D-CSIT setting. Now, we show that this genie-aided upper-bound is achievable.

Theorem 3.2. *In the* 2×2 *single-antenna Network MISO with D-CSIT exponents* $\boldsymbol{\alpha}$*, the sum GDoF denoted by* $\mathrm{GDoF}^{\mathrm{DCSIT}}(\boldsymbol{\alpha})$ *satisfies*

$$\mathrm{GDoF}^{\mathrm{DCSIT}}(\boldsymbol{\alpha}) \geq \mathrm{GDoF}^{\mathrm{CCSIT}}(\boldsymbol{\alpha}^{\star}), \tag{3.14}$$

where $\mathrm{GDoF}^{\mathrm{CCSIT}}(\boldsymbol{\alpha}^{\star})$ *is the GDoF of the C-CSIT scenario with a single shared estimate of accuracy scaling parameters*

$$\boldsymbol{\alpha}^{\star} = \left\{ \alpha_{i,k} = \max\left(\alpha_{i,k}^{(1)},\ \alpha_{i,k}^{(2)}\right) \middle| i,k \in \mathbb{N}_2 \right\}. \tag{3.15}$$

Proof. The achievability proof relies on a proposed transmission scheme so-called Sliced ZF (S-ZF) which is presented in Section 3.3. ■

The lower-bound of Theorem 3.2 and the upper-bound of Lemma 3.1 coincide and thus

$$\mathrm{GDoF}^{\mathrm{DCSIT}}(\boldsymbol{\alpha}) = \mathrm{GDoF}^{\mathrm{CCSIT}}(\boldsymbol{\alpha}^{\star}). \tag{3.16}$$

Remarkably, the achieved GDoF is only limited by the *most accurate* estimate of each link, no matter which TX has it.

3.2.3 DoF Results

The DoF scenario models a network in which the path-loss does not scale exponentially with P. That is, where $\gamma_{i,k} = 1$ for any $i,k \in \mathbb{N}_2$. In this simplified scenario, the DoF (GDoF) result from Theorem 3.2 becomes

$$\mathrm{DoF}^{\mathrm{DCSIT}}(\boldsymbol{\alpha}) = 1 + \min_{i,k\in\mathbb{N}_2}\left(\max_{j\in\mathbb{N}_2}(\alpha_{i,k}^{(j)})\right), \tag{3.17}$$

Considering $\alpha_{i,k}^{(j)} = \alpha^{(j)}$, $\forall i,k \in \mathbb{N}_2$, we recover the results from [82].

3.2.4 An Illustrative Case

We introduce in the following a simple example to convey the main intuition behind Theorem 3.2, and to illustrate how the CSIT configuration –what CSI is known with which quality at which TX– impacts the GDoF performance. As aforementioned, each TX has its own estimate of the channel coefficient between TX k and RX i, with an error scaling as $\bar{P}^{-\alpha_{i,k}^{(j)}}$. In the following example, we consider the conventional DoF, i.e., that the channel path-loss does not scale as the SNR P ($\gamma_{i,k} = 1$ for any $i,k \in \mathbb{N}_2$). Moreover, we consider a CSIT allocation such that, for $\rho \in [0,1]$,

$$\begin{aligned}
&\text{TX } 1 \rightarrow \left\{\alpha_{1,1}^{(1)} = 0.25,\ \alpha_{1,2}^{(1)} = 0.25,\ \alpha_{2,1}^{(1)} = 0.5,\ \alpha_{2,2}^{(1)} = 0.5\right\},\\
&\text{TX } 2 \rightarrow \left\{\alpha_{1,1}^{(2)} = \rho,\ \alpha_{1,2}^{(2)} = \rho,\ \alpha_{2,1}^{(2)} = 1-\rho,\ \alpha_{2,2}^{(2)} = 1-\rho\right\}.
\end{aligned}$$

Note that as ρ increases, TX 2 becomes better informed about the links towards RX 1 and less about the links towards RX 2, while TX 1 keeps a fixed estimation quality for each user. In Fig. 3.2 we show the DoF achieved by the proposed S-ZF scheme as a function of ρ. We compare this DoF with a centralized CSIT setting with CSIT quality $\alpha_{i,k} = \max(\alpha_{i,k}^{(1)}, \alpha_{i,k}^{(2)})$, $\forall i,k \in \mathbb{N}_2$, whose DoF is computed in [50], as well as with the scheme based on conventional ZF and Time Division Multiplexing (TDM).

As stated in Theorem 3.2, the proposed scheme attains the DoF of the genie-aided centralized case, whereas the standard scheme based on conventional ZF, which is optimal in the C-CSIT setting, performs poorly when confronted with CSI discrepancies between TXs. Note that the only case where the conventional ZF scheme is performing as the proposed scheme is for $\rho = 0.25$, as both TXs have the same accuracy for the worse RX.

3.2.5 Implications on CSIT Allocation

Theorem 3.2 shows that, in terms of GDoF, the D-CSIT setting is not sensitive to *who* has the channel estimate. In the following, we illustrate this aspect from the point of view of limited-budget CSI feedback from the RXs. In settings in which the CSIT is obtained by quantized feedback from the RXs, it was shown in [132] that there exists a linear relation between the number of feedback bits per user (B) and the multiplexing gain (GDoF). Specifically, a quality exponent $\alpha_{i,k}^{(j)}$ models a feedback quantized with $B_{i,k}^{(j)} = \alpha_{i,k}^{(j)} \log_2(P)$ bits [132].

Let us consider for simplicity the DoF model, such that $\gamma_{i,k} = 1$ for any $i,k \in \mathbb{N}_2$. Suppose also that $\alpha_{1,k}^{(1)} = 1$, $\alpha_{1,k}^{(2)} = 0$, for all $k \in \mathbb{N}_2$. Thus, TX 1 knows the vector channel of RX 1 perfectly in terms of DoF. Now, let us assume that the maximum feedback rate

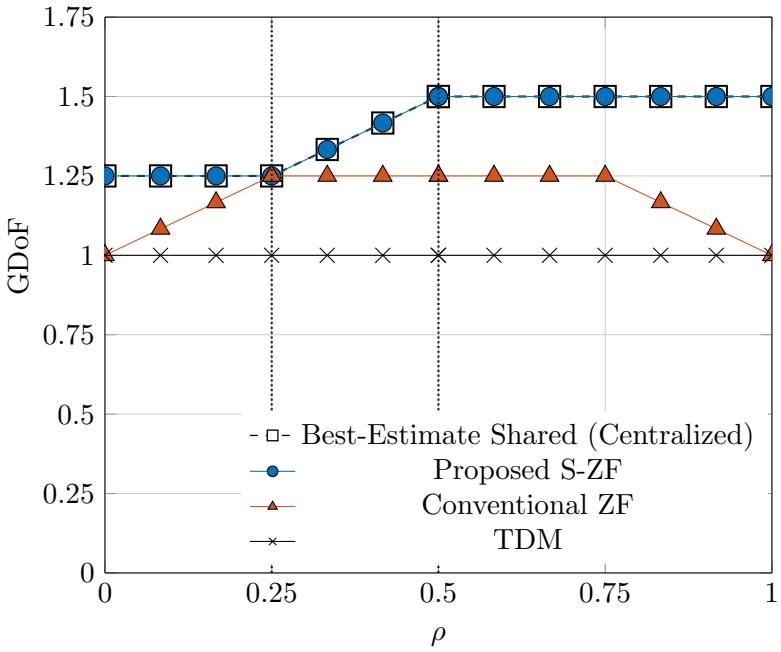


Figure 3.2 – DoF of the illustrative example setting as function of ρ.

from RX 2 towards the TXs, $B_{2,k} = B_{2,k}^{(1)} + B_{2,k}^{(2)}$, is equal to $B_{2,k} = \log_2(P)$, and therefore

$$\alpha_{2,k}^{(1)} + \alpha_{2,k}^{(2)} = 1, \qquad \forall k \in \mathbb{N}_2. \tag{3.18}$$

We vary the number of bits that RX 2 sends to each TX, and thus also $\alpha_{2,k}^{(j)}$. In particular, we suppose that RX 2 first transmits all the feedback bits to TX 1 (what implies that $\alpha_{2,k}^{(1)} = 1$ and $\alpha_{2,k}^{(2)} = 0$). Then RX 2 starts sending gradually more feedback bits to TX 2, reducing at the same time the rate towards TX 1 because of the constraint in (3.18). We can model the feedback allocation shift through a parameter $\beta \in [0, 1]$ satisfying

$$\alpha_{2,k}^{(1)} = 1 - \beta, \qquad \forall i \in \mathbb{N}_2, \tag{3.19}$$

$$\alpha_{2,k}^{(2)} = \beta, \qquad \forall i \in \mathbb{N}_2. \tag{3.20}$$

In Fig. 3.3, the sum DoF of the proposed scenario is shown as function of the feedback allocation parameter β. We can see that the sum DoF decreases as the CSI becomes more evenly distributed. This property is a direct aftermath of the DoF expression, which depends on $\min_{i,k}(\max_j(\alpha_{i,k}^{(j)}))$, what in our case is equivalent to $\max(1-\beta, \beta)$. Thus, up to $\beta = 0.5$ the DoF decreases to DoF $= 1.5$. If β keeps increasing beyond $\beta = 0.5$, the DoF raises gradually. Finally, for $\beta = 1$ we recover the maximum DoF of DoF $= 2$. Two main insights arise from Fig. 3.3. The first one is the explicit symmetry, which implies

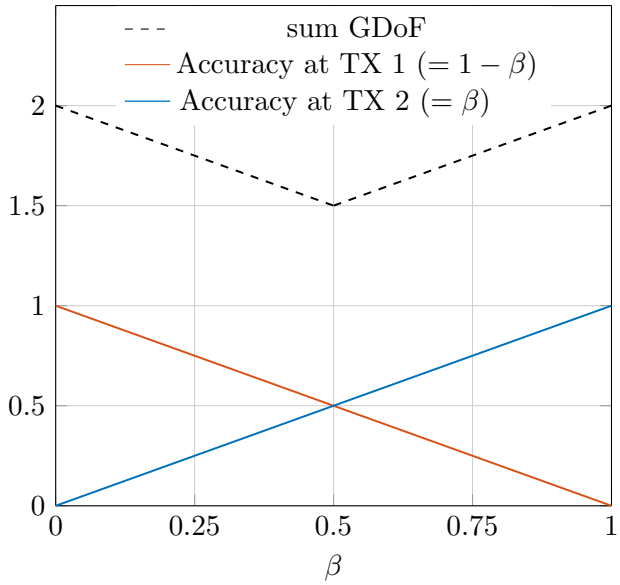


Figure 3.3 – Sum DoF as function of the CSIT allocation at the TXs.

that it does not matter which TX owns the CSIT. (For $\beta = 1$, each TX owns the CSIT of one RX.) The second insight is that in the 2-user D-CSIT setting, under a maximum feedback rate constraint for each RX, the optimal CSIT allocation decision in terms of GDoF is to transmit the CSI of a certain link only to one of the two TXs. Moreover, the CSI of different links does not need to be sent to the same TX. This insight follows from (3.15) in Theorem 3.2.

3.3 Sliced Zero-Forcing Precoding

The proof of Theorem 3.2 is presented in the next section. Prior thereto, we present a novel precoding scheme, coined as Sliced Zero-Forcing (S-ZF), which is essential in that proof. The D-CSIT setting is characterized by its flexibility, in the sense that it comprises different CSIT configurations for which the transmission scheme must be adapted. Indeed, S-ZF uses different precoder expressions –slices– depending on the CSI allocation.

As usual in interference minimizing schemes, the designs of the precoding vectors towards the different RXs can be *decoupled* [50, 137]. Consequently, we present here the precoding vector for the data symbols of RX 1, designed to cancel the interference at RX 2. The precoder for RX 2 will be obtained by a permutation of the user indexes. We decompose the precoding vector such as $\mathbf{t}_1^{\mathrm{SZF}} \triangleq \lambda_1 \mathbf{w}_1$ and

$$\mathbf{t}_1^{\mathrm{SZF}} \triangleq \begin{bmatrix} \mathrm{t}_{1,1}^{(1)} \\ \mathrm{t}_{1,2}^{(2)} \end{bmatrix} \triangleq \lambda_1 \begin{bmatrix} \mathbf{w}_{1,1}^{(1)} \\ \mathbf{w}_{1,2}^{(2)} \end{bmatrix}, \tag{3.21}$$

Table 3.1 – CSI Allocation Regimes

$\alpha_{2,1}$ \ $\alpha_{2,2}$	$\alpha_{2,2}^{(1)} \geq \alpha_{2,2}^{(2)}$	$\alpha_{2,2}^{(1)} < \alpha_{2,2}^{(2)}$
$\alpha_{2,1}^{(1)} > \alpha_{2,1}^{(2)}$	Most-informed TX (TX 1)	Locally Informed TXs
$\alpha_{2,1}^{(1)} \leq \alpha_{2,1}^{(2)}$	Non-locally Informed TXs	Most-informed TX (TX 2)

where λ_1 is a power normalization constant and $\mathrm{w}_{1,k}^{(j)}$ denotes the precoding coefficient applied at TX k and computed locally at TX j, before the power normalization. Note that the super-index (j) is used to highlight which TX owns that information, and which estimate $\hat{\mathbf{H}}^{(j)}$ has been used to compute it. Following the ZF approach [145], the vector $\mathbf{w}_1 = [\mathrm{w}_{1,1}^{(1)},\ \mathrm{w}_{1,2}^{(2)}]^{\mathrm{T}}$ is designed so as to satisfy[2]

$$\hat{\mathbf{h}}_2^{\mathrm{H}} \mathbf{w}_1 = 0, \tag{3.22}$$

although each TX computes (3.22) based on its own available CSIT. This aspect will be discussed in more detail below for each CSI configuration. Moreover, it is important to mention that one TX does not need to compute or know the coefficient applied at the other TX. The normalization constant λ_1 is chosen to fulfill the average power constraint and is then given by

$$\lambda_1 \triangleq \frac{1}{\sqrt{\mathbb{E}\left[\|\mathbf{w}_1\|^2\right]}} \tag{3.23}$$

for the constraint $\mathbb{E}[\|\mathbf{t}_1^{\mathrm{SZF}}\|^2] = 1$. The normalization constant λ_1 only depends on statistical information and can hence be computed at both TXs. Regarding the possible CSI allocation, the channel vector $\mathbf{h}_2^{\mathrm{H}}$ is composed of two coefficients, such that we can distinguish four different CSI regimes depending on which TX has better knowledge of each link. Those four regimes are shown in Table 3.1 and they can be reduced to three cases by symmetry between the TXs. For each of these regimes, we will now describe the S-ZF precoding scheme. We restrict ourselves to the precoder for the data symbols of RX 1, and then only the channel vector of RX 2 is relevant. The key intuition of this scheme is that a TX only uses its own estimate *if it is the most accurate among the two TXs*, as it will become clear in the following.

[2] If we consider Regularized ZF instead of conventional ZF, the ZF condition in (3.22) is only fulfilled asymptotically. At a certain finite transmit power, Regularized ZF does not focus only on interference cancellation ((3.22)), but it also takes into account the impact of noise.

Locally Informed TXs

In this case, each TX has the best estimate of its own channel coefficient towards RX 2. The precoding coefficient at each TX is then given by

$$\begin{aligned} \mathrm{w}_{1,1}^{(1)} &\triangleq (\hat{\mathrm{h}}_{2,1}^{(1)})^{\mathrm{H}} \left(|\hat{\mathrm{h}}_{2,1}^{(1)}|^2 + \frac{1}{\bar{P}}\right)^{-1}, \\ \mathrm{w}_{1,2}^{(2)} &\triangleq (-1)\,(\hat{\mathrm{h}}_{2,2}^{(2)})^{\mathrm{H}} \left(|\hat{\mathrm{h}}_{2,2}^{(2)}|^2 + \frac{1}{\bar{P}}\right)^{-1}. \end{aligned} \tag{3.24}$$

Non-locally Informed TXs

In this case, each TX knows more accurately the channel coefficient from the other TX towards RX 2. The precoding coefficient at each TX is given by

$$\begin{aligned} \mathrm{w}_{1,1}^{(1)} &= \hat{\mathrm{h}}_{2,2}^{(1)}, \\ \mathrm{w}_{1,2}^{(2)} &= \hat{\mathrm{h}}_{2,1}^{(2)}. \end{aligned} \tag{3.25}$$

Most-informed TX

In this last case, there exists one TX that has the best estimate of both coefficients. The S-ZF precoding is then based on the AP-ZF scheme introduced in [82]: The TX with less accurate CSIT (e.g., TX 2) transmits with a constant precoder while the *most-informed TX* (TX 1) tries to correct the interference generated by TX 2. As a matter of example, let us consider that the constant precoder at TX 2 is given by $\mathrm{w}_{1,2}^{(2)} \triangleq -1$. Thus,

$$\mathrm{w}_{1,1}^{(1)} = (\hat{\mathrm{h}}_{2,1}^{(1)})^{\mathrm{H}} \left(|\hat{\mathrm{h}}_{2,1}^{(1)}|^2 + \frac{1}{P}\right)^{-1} \hat{\mathrm{h}}_{2,2}^{(1)}. \tag{3.26}$$

In order to convey the main intuition behind the precoder expressions, Table 3.2 shows the simplified expression –with non-regularized inverses– of the precoders in (3.26)-(3.25).

Remark 3.2. The S-ZF precoding is designed so that the interfered RX receives two equal signals with opposite phase and hence the interference is canceled. The precoder allows to use only the most accurate estimate of each link. Table 3.2 shows that all the choices of $\mathbf{w}_1$ solve (3.22) with the highest possible accuracy. □

Normalization Constant

The value of the normalization constant λ_i varies accordingly to the precoder expression. The next proposition discloses the asymptotic scaling of λ_1 for each one of the possible precoders, whereas λ_2 follows from a permutation of the RX indexes. We analyze the normalization constant with Rayleigh fading in detail in Section 3.6.

Table 3.2 – Simplified precoder $\mathbf{w}_1 = [\mathrm{w}_{1,1}^{(1)},\ \mathrm{w}_{1,2}^{(2)}]^{\mathrm{T}}$ for the data symbols of RX 1 according to the CSIT configuration. "Main TX j" denotes the "Most-informed TX" (TX j) case.

	Main TX 1	Main TX 2	Local CSIT	Non-local CSIT
	$\alpha_{2,1}^{(1)} > \alpha_{2,1}^{(2)}$ $\alpha_{2,2}^{(1)} \geq \alpha_{2,2}^{(2)}$	$\alpha_{2,1}^{(1)} > \alpha_{2,1}^{(2)}$ $\alpha_{2,2}^{(1)} \geq \alpha_{2,2}^{(2)}$	$\alpha_{2,1}^{(1)} > \alpha_{2,1}^{(2)}$ $\alpha_{2,2}^{(1)} < \alpha_{2,2}^{(2)}$	$\alpha_{2,1}^{(1)} \leq \alpha_{2,1}^{(2)}$ $\alpha_{2,2}^{(1)} \geq \alpha_{2,2}^{(2)}$
$\mathbf{w}_1$	$\begin{bmatrix} (\hat{\mathrm{h}}_{2,1}^{(1)})^{-1}\hat{\mathrm{h}}_{2,2}^{(1)} \\ -1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ -(\hat{\mathrm{h}}_{2,2}^{(1)})^{-1}\hat{\mathrm{h}}_{2,1}^{(1)} \end{bmatrix}$	$\begin{bmatrix} (\hat{\mathrm{h}}_{2,1}^{(1)})^{-1} \\ -(\hat{\mathrm{h}}_{2,2}^{(2)})^{-1} \end{bmatrix}$	$\begin{bmatrix} \hat{\mathrm{h}}_{2,2}^{(1)} \\ -\hat{\mathrm{h}}_{2,1}^{(2)} \end{bmatrix}$

Proposition 3.2. *The normalization constant of the precoder for the data symbols of RX 1, for the constraint* $\mathbb{E}[\|\mathbf{t}_1^{\mathrm{SZF}}\|^2] = 1$*, satisfies that*

$$\lambda_1 = \begin{cases} \Theta_{\log}(\bar{P}^{1-\max(\gamma_{2,1},\gamma_{2,2})}) & \text{Non-locally Informed TXs} \quad (3.27a) \\ \Theta_{\log}(\bar{P}^{\min(\gamma_{2,1},\gamma_{2,2})-1}) & \text{Locally Informed TXs} \quad (3.27b) \\ \Theta_{\log}(\bar{P}^{-(\gamma_{2,2}-\gamma_{2,1})^+}) & \text{Most-informed TX (TX 1)} \quad (3.27c) \\ \Theta_{\log}(\bar{P}^{-(\gamma_{2,1}-\gamma_{2,2})^+}) & \text{Most-informed TX (TX 2)} \quad (3.27d) \end{cases}$$

Proof. The asymptotic scaling of λ_1 is directly obtained from the precoding vector definitions in (3.26), (3.24), and (3.25), and the fact that $\mathbb{E}[\|\hat{\mathrm{h}}_{i,k}^{(j)}\|] = \Theta(\bar{P}^{\gamma_{i,k}-1})$. ■

S-ZF Precoder: General Expression

The adaptive S-ZF precoder can be enclosed in a single precoding expression that encapsulates the four possible precoding vectors summarized in Table 3.2 and described in (3.24)-(3.26). Let us denote, for any index $a \subset \mathbb{N}_2$, the complementary $\bar{a} \triangleq a \pmod 2 + 1$, such that $a, \bar{a} \in \mathbb{N}_2$, and $a \neq \bar{a}$. Therefore, we can write the S-ZF precoder for RX i as

$$\mathrm{w}_{i,k}^{(j)} \triangleq (-1)^j \left(1 - \mathrm{c}_{i,j}^{(j)} + \mathrm{c}_{i,j}^{(j)} (\hat{\mathrm{h}}_{i,j}^{(j)})^{\mathrm{H}} \left(|\hat{\mathrm{h}}_{i,j}^{(j)}|^2 + \frac{1}{P}\right)^{-1}\right) \left(1 - \mathrm{c}_{i,\bar{j}}^{(j)} + \mathrm{c}_{i,\bar{j}}^{(j)} \hat{\mathrm{h}}_{i,\bar{j}}^{(j)}\right), \quad (3.28)$$

where $\mathrm{c}_{i,k}^{(j)} \in \{0, 1\}$ is defined as

$$\mathrm{c}_{i,k}^{(j)} = \begin{cases} 1 & \text{if } \alpha_{i,k}^{(j)} \geq \alpha_{i,k}^{(\bar{j})}, \\ 0 & \text{if } \alpha_{i,k}^{(j)} < \alpha_{i,k}^{(\bar{j})}. \end{cases} \quad (3.29)$$

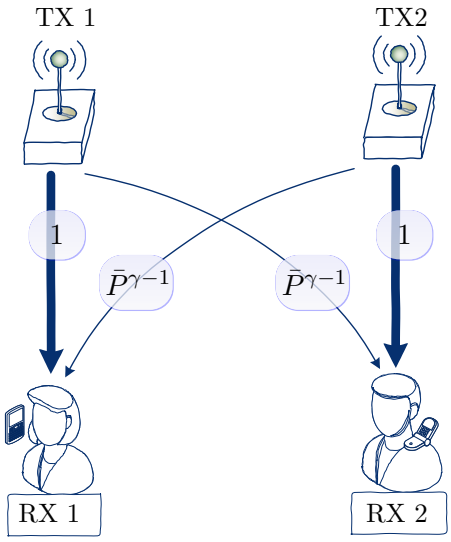


Figure 3.4 – Network topology for the *Parallel Configuration*.

The role of the parameters $c_{i,k}^{(j)}$ is to allow to swift from one precoder to another according to the CSI allocation at the TXs.

3.4 Achievability Example for a Simple Configuration

In this section, we illustrate the proposed precoding scheme for a simple CSIT and path-loss configuration, so as to convey the main intuition while avoiding cluttered and heavy notations. Specifically, we consider a particular path-loss configuration for the sake of exposition –so-called *Parallel Configuration*, and represented in Fig. 3.4–, in which $\gamma_{i,i} = 1$ and $\gamma_{i,k} = \gamma$ for $k \neq i$. Besides this, we consider that each TX has an homogeneous CSI accuracy, i.e., $\alpha_{i,k}^{(j)} = \alpha^{(j)}$, for any $i, k \in \mathbb{N}_2$, and that TX 1 is the best informed TX, i.e., $\alpha^{(1)} \geq \alpha^{(2)}$. Suppose now that we aim to send information to both users and satisfy that the interference lies on the noise floor (P^0).

Fig. 3.5 illustrates the different power levels for the transmission of the symbols of RX 2 at RX 1. (Due to the symmetry of the configuration, the received signal at RX 2 is equivalent.) We can observe several insights from Fig. 3.5: First, the transmitted power scales as $\bar{P}^{1-\gamma+\alpha^{(1)}}$. Thus, in this particular configuration, the path-loss is beneficial, as it allows to increase the received power of the intended signal while keeping fixed the interference power. Note that in other path-loss configurations this behavior is reversed and the path-loss is detrimental. Second, it shows how TX 1 reduces his transmitted power for s_2 to compensate that the channel from TX 2 is weaker, so that the interference power received at RX 1 from both TXs has the same scaling. Hence, the non-intended

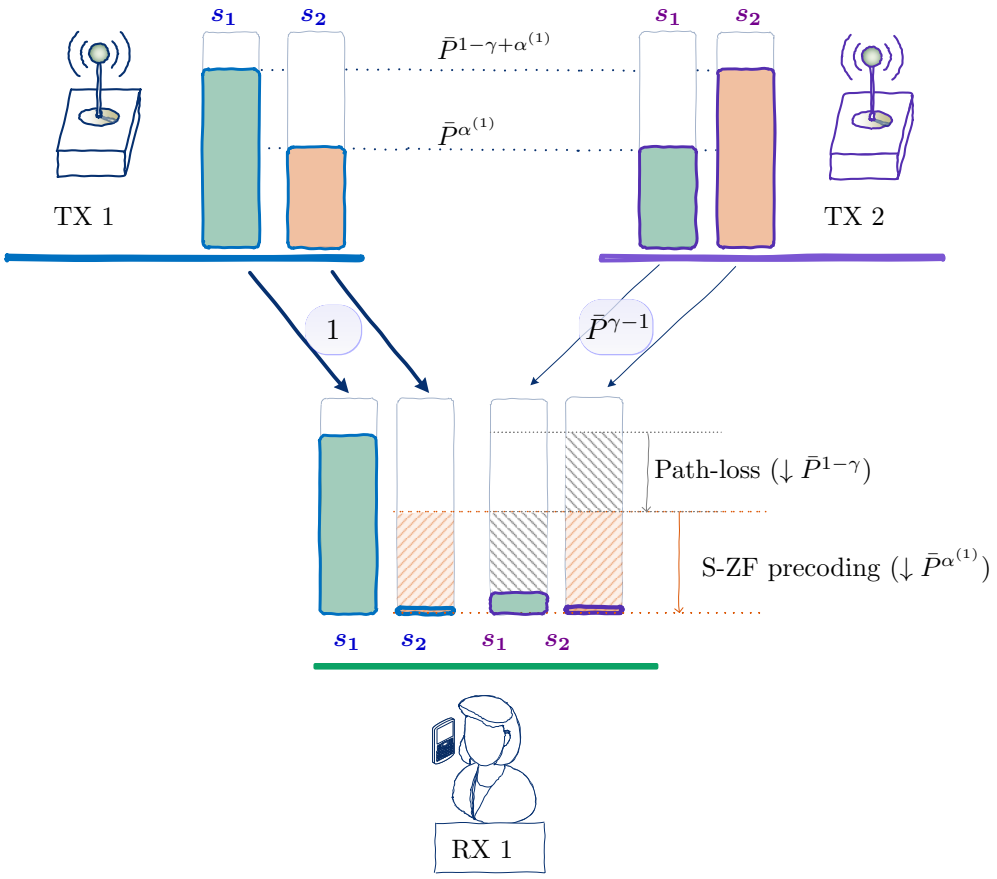


Figure 3.5 – Illustration of the different power scaling for the *Parallel Configuration*. Attenuation of the signal power due to the path-loss and the S-ZF precoding are emphasized using arrows.

symbol scales in $P^{\alpha^{(1)}}$. Then, thanks to the S-ZF precoding, it is possible to entirely cancel the interference (see Lemma 3.2).

3.5 Proof of Theorem 3.2

The achievability proof of Theorem 3.2 is based on the transmission scheme that attains the centralized GDoF bound in the centralized MISO BC of [50], but they differ on the precoding vector applied. The centralized scheme uses conventional ZF as the tool to reduce the interference. However, the use of conventional ZF in the D-CSIT setting leads to a degraded interference cancellation, such that the centralized GDoF cannot be reached. We overtake the limitation of conventional ZF with the proposed S-ZF scheme –introduced in Section 3.3– and achieve the centralized upper-bound. We present here

the main structure of the transmission scheme, emphasizing the novelty with respect to the centralized achievable scheme. A detailed description of the scheme, disclosing the appropriate transmit power of each symbol for each CSI allocation regime is relegated to Appendix A for the sake of conciseness.

3.5.1 Rate-Splitting Approach

The transmission scheme makes use of the rate-splitting approach [54, 146]. In this technique, the messages intended by the RXs (W_i for RX i) are split in several data symbols. The main idea is that each message W_i is divided in a private ($s_{i,p}$) and a common ($s_{i,c}$) messages. Then, the private part is transmitted so as to be decoded by the intended RX (RX i) and attenuated at the other RXs. Conversely, the common parts of each message are gathered in a single data symbol

$$s_c = \{s_{1,c}, \dots, s_{K,c}\}. \tag{3.30}$$

This common symbol is broadcast so as to be decoded by all the RXs –and thus its rate is limited by the worst RX channel quality–. For the decoding, each RX first decode the common symbol s_c. Then, the RX substracts the contribution of s_c to the received signal, and thus it can decode the intended private symbol $s_{i,p}$ by treating the other $s_{j,p}$, $j \neq i$, symbols as noise. The general rate-splitting approach allows for an arbitrary number of splittings, as well as common symbols for a sub-set of RXs.

3.5.2 Superposition Coding Transmission Scheme

Superposition coding schemes have been shown to achieve optimal DoF and GDoF for multiple BC settings with imperfect CSIT [48, 50, 97, 137]. In our setting, the transmission scheme fits the expression

$$\mathbf{x} = \bar{P}_{\mathrm{BC}}\mathbf{t}_{\mathrm{BC}}s_{\mathrm{BC}} + \bar{P}_{\mathrm{ZF}}(\mathbf{t}_1^{\mathrm{ZF}}s_{\mathrm{ZF}\,1} + \mathbf{t}_2^{\mathrm{ZF}}s_{\mathrm{ZF}\,2}) + \bar{P}_{\phi}\mathbf{t}_{\phi}s_{\phi}. \tag{3.31}$$

The terms s_{BC}, $s_{\mathrm{ZF}\,1}$, $s_{\mathrm{ZF}\,2}$, s_{ϕ} denote four different information symbols that are described in the following. Depending on the path-loss topology (i.e., the value of $\gamma_{i,k}$) and the CSIT allocation (i.e., the value of $\alpha_{i,k}^{(j)}$), some of those four symbols may be suppressed. In the general scheme, those symbols form a three-layer structure where each layer has a different power scaling, given by $\bar{P}_{\mathrm{BC}}$, $\bar{P}_{\mathrm{ZF}}$, and $\bar{P}_{\phi}$. Specifically:

1. *Low-power layer:* s_{ϕ} is a non-zero-forced symbol transmitted with power such that it is only received by the intended RX, if the path-loss topology allows for that.

2. *Zero-Forcing layer:* $s_{\mathrm{ZF}\,i}$, $i \in \mathbb{N}_2$, is intended to RX i and *canceled* at the other

RX using ZF-type precoding. A necessary condition for the optimality of the scheme is that the interference generated by those symbols lies below the noise floor. Therefore, they are transmitted with a power proportional to the CSI accuracy.

3. *Full-power layer:* s_{BC} is a broadcast symbol transmitted with full power, intended to be decoded at both RXs.

In order to decode every intended symbol, RX i applies successive decoding [97] to first decode s_{BC}, then its intended symbol $s_{\mathrm{ZF}\,i}$ and finally s_ϕ, if it is intended to RX i. Note that, in this case, the *Full-power layer* carries the common symbol s_c introduced in the previous section about the rate-splitting approach. Hence, this symbol $s_{\mathrm{BC}} = s_c$ may carry information for all the RXs uniformly, but we can also adapt its composition such that we balance the user rate accordingly to different considerations (e.g., max-min rate optimization). Furthermore, the private message of one of the RXs may be further split in two symbols: If we consider that s_ϕ is intended to RX i, the private message $s_{i,p}$ previously described is divided into $s_{\mathrm{ZF}\,i}$ and s_ϕ.

3.5.3 Interference Cancellation

Importantly, the precoders $\mathbf{t}_{\mathrm{BC}}$ and $\mathbf{t}_\phi$ depend only on the long-term statistical information ($\alpha_{i,k}^{(j)}$ and $\gamma_{i,k}$) and are hence not affected by the instantaneous CSI discrepancies between TXs. This implies that, in order to prove that it is possible to achieve the same GDoF as in the centralized setting with $\alpha_{i,k} = \max(\alpha_{i,k}^{(1)},\ \alpha_{i,k}^{(2)})$, we only need to show that S-ZF achieves the same level of interference attenuation as ZF in the centralized setting of reference. This is shown for the interference at RX 2 by means of the following lemma, while the same result holds for RX 1 after a permutation of indexes.

Lemma 3.2. *The S-ZF precoder achieves the same interference reduction scaling as the conventional ZF precoder computed from the best estimate of each channel coefficient, i.e.,*

$$\left|\mathbf{h}_2^{\mathrm{H}}\mathbf{t}_1^{\mathrm{SZF}}\right|^2 = \Theta_{\mathsf{p}}(P^{\min(\gamma_{2,1},\gamma_{2,2})-1-\alpha_2^{opt}}), \tag{3.32}$$

when $\mathbb{E}\left[\|\mathbf{t}_1^{\mathrm{SZF}}\|^2\right] = \Theta(1)$*, and where we have defined the short-hand notation* α_2^{opt} *as*

$$\alpha_2^{opt} \triangleq \min\left(\max_{j\,\in\,\mathbb{N}_2} \alpha_{2,1}^{(j)}, \max_{j\,\in\,\mathbb{N}_2} \alpha_{2,2}^{(j)}\right). \tag{3.33}$$

Proof. We prove separately each of the regimes of Table 3.1.

Non-locally Informed TXs: From the precoding vector expression in Table 3.2, the interference term satisfies that

$$\begin{aligned}\left|\mathbf{h}_2^{\mathrm{H}}\mathbf{t}_1^{\mathrm{SZF}}\right|^2 &= \lambda_1^2\left|-\mathrm{h}_{2,1}\hat{\mathrm{h}}_{2,2}^{(1)}+\mathrm{h}_{2,2}\hat{\mathrm{h}}_{2,1}^{(2)}\right|^2\\ &\overset{(a)}{=} \lambda_1^2\Big|-\mathrm{h}_{2,1}\mathrm{h}_{2,2}-\mathrm{h}_{2,1}\bar{P}^{-\alpha_{2,2}^{(1)}+\gamma_{2,2}-1}\delta_{2,2}^{(1)}\\ &\qquad +\mathrm{h}_{2,2}\mathrm{h}_{2,1}+\mathrm{h}_{2,2}\bar{P}^{-\alpha_{2,1}^{(2)}+\gamma_{2,1}-1}\delta_{2,1}^{(2)}\Big|^2\\ &\overset{(b)}{=} \lambda_1^2P^{\gamma_{2,1}+\gamma_{2,2}-2}\left|\bar{P}^{-\alpha_{2,1}^{(2)}}\mathrm{g}_{2,2}\delta_{2,1}^{(2)}-\bar{P}^{-\alpha_{2,2}^{(1)}}\mathrm{g}_{2,1}\delta_{2,2}^{(1)}\right|^2,\end{aligned} \tag{3.34}$$

where (a) holds because, by definition,

$$\hat{\mathrm{h}}_{2,k}^{(j)} \triangleq \mathrm{h}_{2,k}+\bar{P}^{-\alpha_{2,k}^{(j)}+\gamma_{2,k}-1}\delta_{2,k}^{(j)}, \tag{3.35}$$

and (b) comes from $\mathrm{h}_{2,k}\triangleq\bar{P}^{\gamma_{2,k}-1}\mathrm{g}_{2,k}$. Focusing on the absolute value term in (3.34) and recalling that $\alpha_2^{\mathrm{opt}}=\min(\alpha_{2,2}^{(1)},\alpha_{2,1}^{(2)})$, it holds that

$$\left|\bar{P}^{-\alpha_{2,1}^{(2)}}\mathrm{g}_{2,2}\delta_{2,1}^{(2)}-\bar{P}^{-\alpha_{2,2}^{(1)}}\mathrm{g}_{2,1}\delta_{2,2}^{(1)}\right|^2=\Theta_{\mathsf{p}}(P^{-\alpha_2^{\mathrm{opt}}}). \tag{3.36}$$

Including (3.36) in (3.34) and substituting λ_1 with (3.27a) yields

$$\left|\mathbf{h}_2^{\mathrm{H}}\mathbf{t}_1^{\mathrm{SZF}}\right|^2=\Theta_{\mathsf{p}}(P^{\min(\gamma_{2,1},\gamma_{2,2})-1-\alpha_2^{\mathrm{opt}}}), \tag{3.37}$$

what proves Lemma 3.2 for the "Non-locally Informed TXs" case. The two other regimes follow in a similar manner.

Locally Informed TXs: In this case, substituting $\mathbf{t}_1^{\mathrm{SZF}}$ by its expression[3] yields

$$\begin{aligned}\left|\mathbf{h}_2^{\mathrm{H}}\mathbf{t}_1^{\mathrm{SZF}}\right|^2 &= \lambda_1^2\left|-\mathrm{h}_{2,1}(\hat{\mathrm{h}}_{2,1}^{(1)})^{-1}+\mathrm{h}_{2,2}(\hat{\mathrm{h}}_{2,2}^{(2)})^{-1}\right|^2\\ &\overset{(a)}{=} \lambda_1^2\Big|-\hat{\mathrm{h}}_{2,1}^{(1)}\big(\hat{\mathrm{h}}_{2,1}^{(1)}\big)^{-1}+\bar{P}^{-\alpha_{2,1}^{(1)}+\gamma_{2,1}-1}\delta_{2,1}^{(1)}\big(\hat{\mathrm{h}}_{2,1}^{(1)}\big)^{-1}\\ &\qquad +\hat{\mathrm{h}}_{2,2}^{(2)}\big(\hat{\mathrm{h}}_{2,2}^{(2)}\big)^{-1}-\bar{P}^{-\alpha_{2,2}^{(2)}+\gamma_{2,2}-1}\delta_{2,2}^{(2)}\big(\hat{\mathrm{h}}_{2,2}^{(2)}\big)^{-1}\Big|^2\\ &\overset{(b)}{=} \lambda_1^2\left|\bar{P}^{-\alpha_{2,1}^{(1)}}\delta_{2,1}^{(1)}\big(\hat{\mathrm{g}}_{2,1}^{(1)}\big)^{-1}-\bar{P}^{-\alpha_{2,2}^{(2)}}\delta_{2,2}^{(2)}\big(\hat{\mathrm{g}}_{2,2}^{(2)}\big)^{-1}\right|^2\\ &\overset{(c)}{=} \Theta_{\mathsf{p}}(P^{\min(\gamma_{2,1},\gamma_{2,2})-1-\alpha_2^{\mathrm{opt}}}),\end{aligned} \tag{3.38}$$

where (a) follows from $\mathrm{h}_{2,k}=\hat{\mathrm{h}}_{2,k}^{(j)}-\bar{P}^{-\alpha_{2,k}^{(j)}+\gamma_{2,k}-1}\delta_{2,k}^{(j)}$, (b) comes from applying $\hat{\mathrm{h}}_{2,k}^{(j)}=\bar{P}^{\gamma_{2,k}-1}\hat{\mathrm{g}}_{2,k}^{(j)}$, and (c) is obtained after substituting λ_1 with its value in (3.27b).

Most-informed TX: Let us consider w.l.o.g. the case in which TX 1 is the TX with the most accurate CSI. Hence, following the same steps as in the previous cases, we obtain

$$\begin{aligned}\left|\mathbf{h}_2^{\mathrm{H}}\mathbf{t}_1^{\mathrm{SZF}}\right|^2 &= \lambda_1^2\ \left|-\mathrm{h}_{2,1}(\hat{\mathrm{h}}_{2,1}^{(1)})^{-1}\hat{\mathrm{h}}_{2,2}^{(2)}+\mathrm{h}_{2,2}\right|^2 \\ &= \lambda_1^2\ \left|\bar{P}^{-\alpha_{2,1}^{(1)}+\gamma_{2,2}-1}\delta_{2,1}^{(1)}\left(\hat{\mathrm{g}}_{2,1}^{(1)}\right)^{-1}\mathrm{g}_{2,2}-\bar{P}^{-\alpha_{2,2}^{(1)}+\gamma_{2,2}-1}\delta_{2,2}^{(1)}\right|^2 \\ &= \Theta_{\mathsf{p}}(P^{\min(\gamma_{2,1},\gamma_{2,2})-1-\alpha_2^{\mathrm{opt}}}),\end{aligned} \tag{3.39}$$

what concludes the proof of Lemma 3.2. ■

From the Corollary 2.1, Lemma 3.2 implies that

$$\mathbb{E}[\left|\mathbf{h}_2^{\mathrm{H}}\mathbf{t}_1^{\mathrm{SZF}}\right|^2] = \Theta(P^{\min(\gamma_{2,1},\gamma_{2,2})-1-\alpha_2^{\mathrm{opt}}}), \tag{3.40}$$

i.e., the average received power of the interference scales as in the centralized setting.

3.6 Normalization Constant with Rayleigh Fading

Proposition 3.2 establishes the asymptotic scaling of λ_i on the limiting high-SNR regime. Hereunder, we extend the characterization of λ_i under the assumption of Rayleigh fading by providing its exact value. Recalling (3.23), the constant λ_i is defined as

$$\lambda_i \triangleq \frac{1}{\sqrt{\mathbb{E}[\|\mathbf{w}_i\|^2]}}. \tag{3.41}$$

Let us assume that the channel is drawn from a complex gaussian distribution where all the links are independent with different variance, i.e., $\mathrm{h}_{i,k} \sim \mathcal{N}_{\mathbb{C}}(0,\sigma_{i,k}^2)$. The GDoF channel model is recovered if $\sigma_{i,k}^2 = P^{\gamma_{i,k}-1}$. Therefore, we need to obtain $\mathbb{E}[\|\mathbf{w}_i\|^2]$. We present in the following the expression of λ_i for each one of the regimes of Table 3.1. We use the GDoF model as throughout the rest of the section, and we refer to Appendix B for a general expression for any channel variance $\sigma_{i,k}^2$ and any regularization term, as well as for a full characterization of the probability density function of λ_i. We further assume that the estimate $\hat{\mathrm{h}}_{i,k}^{(j)}$ follows the same distribution as the channel coefficient $\mathrm{h}_{i,k}$.

3.6.1 Preliminaries: Statistics of the Regularized Inverse

The S-ZF precoder computes in several cases the regularized inverse of a channel coefficient $\hat{\mathrm{h}}_{i,k}^{(j)}$, which is given by $(\hat{\mathrm{h}}_{i,k}^{(j)})^{\mathrm{H}}(|\hat{\mathrm{h}}_{i,k}^{(j)}|^2+1/P)^{-1}$. We are interested in the expected

[3] We have omitted the regularization of the inverse for simplicity, as it does not have any impact in the asymptotic regime.

value of its squared absolute value.

Proposition 3.3. *The regularized inverse of a channel coefficient satisfies that*

$$\mathbb{E}\left[\left|(\hat{\mathrm{h}}_{i,k}^{(j)})^{\mathrm{H}}\big(|\hat{\mathrm{h}}_{i,k}^{(j)}|^2+\frac{1}{P}\big)^{-1}\right|^2\right] = P^{1-\gamma_{i,1}}\big(e^{P^{-\gamma_{i,1}}}E_1(P^{-\gamma_{i,1}})(1+P^{-\gamma_{i,1}})-1\big), \quad (3.42)$$

where $E_1(z)$ *is the exponential integral defined as* $E_1(z) \triangleq \int_1^\infty \frac{e^{-zt}}{t}\,\mathrm{d}t$.

Proof. The proof is relegated to Appendix B. ∎

3.6.2 Normalization Constant for the S-ZF Precoders

Non-locally Informed TXs: This case does not require inversion and λ_i is given by

$$\begin{aligned}\lambda_i &= \frac{1}{\sqrt{\mathbb{E}\left[|\mathrm{w}_{i,1}^{(1)}|^2+|\mathrm{w}_{i,2}^{(2)}|^2\right]}} \\ &= \frac{1}{\sqrt{P^{\gamma_{i,2}{}^{-1}}+P^{\gamma_{i,1}{}^{-1}}}}.\end{aligned} \quad (3.43)$$

Most-informed TX: Suppose that TX 1 is the Most-informed TX. From Proposition 3.3 and the definition of the precoder it follows that

$$\lambda_i = \frac{1}{\sqrt{P^{\gamma_{i,2}-\gamma_{i,1}}\left(e^{P^{-\gamma_{i,1}}}E_1(P^{-\gamma_{i,1}})(1+P^{-\gamma_{i,1}})-1\right)+1}}. \quad (3.44)$$

Locally Informed TXs: In this case, each TX applies the regularized inverse as precoder, and thus Proposition 3.3 yields

$$\lambda_i = \frac{1}{\sqrt{\sum_{j=1}^{2}P^{1-\gamma_{i,j}}\left(e^{P^{-\gamma_{i,j}}}E_1(P^{-\gamma_{i,j}})(1+P^{-\gamma_{i,j}})-1\right)}}. \quad (3.45)$$

3.6.3 Asymptotic scaling of the normalization constant

The asymptotic scaling of λ_i in the "Non-locally Informed TXs" case is directly obtained from (3.43) and it matches (3.27a), as expected. For the other two cases, let us consider the limit as P approaches to infinity. Consider $\gamma_{i,k} > 0$. Note that

$$\lim_{P\to\infty} 1+P^{-\gamma_{i,k}} = 1 \quad (3.46)$$

$$\lim_{P\to\infty} e^{P^{-\gamma_{i,k}}} = 1 \quad (3.47)$$

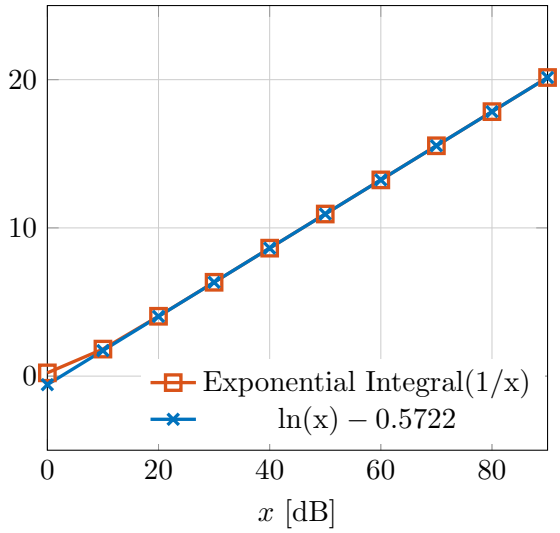


Figure 3.6 – Approximation of the $E_1(\frac{1}{P})$ when $P \to \infty$.

$$\lim_{P\to\infty} E_1(P^{-\gamma_{i,k}}) \overset{P\to\infty}{\approx} \ln(P^{\gamma_{i,k}}) - 0.57722. \tag{3.48}$$

The linear approximation of the exponential integral in (3.48) is shown in Fig. 3.6 to fit almost perfectly. The Mean Square Error (MSE) of this approximation is $\epsilon = 1.8793{*}10^{-11}$ for $P > 10$dB. Consider now the "Most-informed TX" case with TX 1 having more accurate CSI. Then,

$$\lambda_i \overset{P\to\infty}{\approx} \frac{1}{\sqrt{P^{\gamma_{i,2}-\gamma_{i,1}}\left(\ln(P^{\gamma_{i,1}}) - 1.57722\right) + 1}}. \tag{3.49}$$

Interestingly, the logarithmic term $\ln(P^{\gamma_{i,1}})$ implies that $\lambda_i = \Theta_{\log}(\bar{P}^{-(\gamma_{i,2}-\gamma_{i,1})^+})$ but conversely $\lambda_i \neq \Theta(\bar{P}^{-(\gamma_{i,2}-\gamma_{i,1})^+})$. This scaling justifies the defined notation $\Theta_{\log}(\cdot)$.

3.7 Numerical Results

We consider a parallel path-loss topology as the one introduced in Section 3.4 and depicted in Fig. 3.4, whose path-loss coefficients are given by

$$\gamma_{i,i} = 1, \qquad \gamma_{i,k} = 0.8, \qquad \forall i,k \in \mathbb{N}_2, k \neq i. \tag{3.50}$$

We further consider that the TXs have a homogeneous CSIT accuracy, such that

$$\alpha_{i,k}^{(1)} = 0.5, \qquad \alpha_{i,k}^{(2)} = 0, \qquad \forall i,k \in \mathbb{N}_2. \tag{3.51}$$

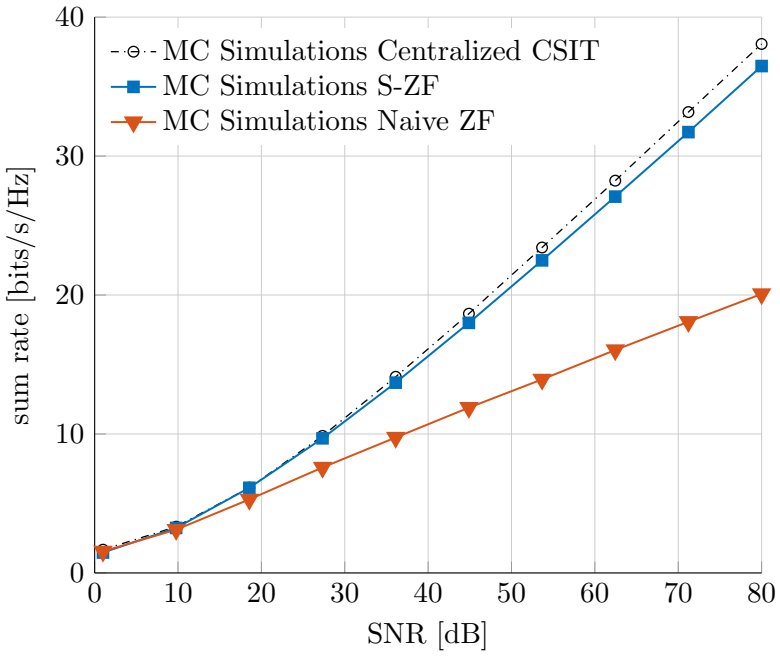


Figure 3.7 – Sum rate in terms of the SNR for the *Parallel Configuration* of Section 3.7, with $\alpha^{(1)} = 0.5$, $\alpha^{(2)} = 0$, and $\gamma = 0.8$.

In this case, the CSI accuracy at TX 2 does not scale with P and thus it is fruitless in terms of GDoF. The S-ZF scheme has been simulated and compared with two different schemes. The first one is the Centralized CSIT setting where both TXs share the CSIT information, which has been shown in Lemma 3.1 to be an upper-bound for the D-CSIT setting. The second one is the naive distributed ZF, where each TX implicitly assumes that the other TX has the same channel estimate [82], and then applies conventional ZF.

In Fig. 3.7, the GDoF is equal to the slope at high SNR of the sum-rate function over the SNR. It can be seen that S-ZF in the D-CSIT setting achieves the same GDoF as the C-CSIT case. The naive distributed Zero-Forcing is limited by the worst CSIT quality estimate, $\alpha^{(2)} = 0$, and thus the CSIT at the best TX is useless for this naive ZF and it matches the performance of the setting with no CSIT [82].

3.8 Conclusions

We have shown that, remarkably, having different CSIT at each TX does not decrease the GDoF of the 2-user Network MISO for any channel topology. Key to this surprisingly good performance is the adaptation of the role of each TX as a function of the multi-TX

multi-user CSIT configuration. The dimensionality of the setting provides the main intuition of this result: For each precoding vector, the TXs need to cancel out only one interference term. Then, having an accurate CSI at only one TX is enough to manage the interference. Besides this, we develop an adapted transmit power scheme that attains the GDoF of the ideal centralized setting with perfect CSIT sharing. Hence, this work reveals that cooperative settings are much more resilient against CSI mismatches between TXs than commonly thought in the community, what could impact the future design of feedback mechanisms. The main question that follows from this analysis is if the result can be extended to more general networks, with an arbitrary number of TXs and RXs. This subject is investigated in the following chapter.

Chapter 4

DoF Analysis of the K×K Distributed Network MISO

In the previous chapter, we have analyzed the particular setting with 2 single-antenna TXs and 2 single-antenna RXs, with the aftermath that having different CSIT scaling at each TX does not reduce the GDoF of the system. Motivated by this result, we extend the analysis to the $K \times K$ setting, aiming at finding the fundamental limitations of the distributed setting. In other words, at answering the question

To what extent can the D-CSIT setting retain the GDoF of the centralized setting?

However, the GDoF metric suffers from a curse of dimensionality, in the sense that the number of parameters increases polynomially with respect to the size of the network. In particular, in a D-CSIT $K \times K$ network there are K^2 path-loss parameters (γ) and K^3 accuracy scaling parameters (α). Hence, the tractability of the analytical characterization of the network becomes cumbersome. In the literature, many works that tackle this challenge circumvent this difficulty by assuming simplified symmetric settings [50,147–151]. In a similar vein, we focus in this chapter on the DoF analysis, such that all the channel coefficients undergo the same path-loss scaling and thus $\gamma_{i,k} = 1$ for any $i, k \in \mathbb{N}_K$. This restriction allows us to direct the analysis on the impact of decentralized CSIT and to abstract it from the path-loss topology leverage.

As a matter of example, the DoF of the 2×2 C-CSIT setting is given by

$$\mathrm{DoF}(\boldsymbol{\alpha}) = 1 + \min_{i,k \in \mathbb{N}_2} (\alpha_{i,k}). \tag{4.1}$$

If we compare the simple expression in (4.1) with the GDoF expression for the same setting in Theorem 3.1, we observe that the complexity of the expression already raises considerably for the simplest network case.

Hereunder, we analyze the D-CSIT setting with K TXs and K RXs. We first compute the DoF of the genie-aided C-CSIT setting where all TXs are given the knowledge of all the channel estimates at all TXs. We also show that this bound is tight for a range of D-CSIT configurations, coined the *Weak-CSIT regime* and defined rigorously further down. Interestingly, the optimal DoF for such D-CSIT settings only depends on the CSIT quality at the most informed TXs. Sharing the instantaneous CSIT among the TXs is hence not necessary to achieve the genie-aided centralized DoF and does not improve the optimal DoF. Building on previous fundamental principles, we also present a robust transmission scheme adapted to any CSIT configuration and any number of users and which significantly improve the achieved DoF with respect to state-of-the-art methods.

A byproduct of the content of this chapter, which completes its main contributions, is the development of new methods used as building blocks to our main algorithm and which are of interest by themselves for other applications. The first one is the idea that increasing the number of TXs with no CSIT can increase the DoF performance, as they turn out to be essential for transmitting multi-stream transmissions to a single-antenna user, and hence create an overloaded transmission. The second method is the translation to the distributed CSIT setting of the idea introduced by Maddah-Ali and Tse in [51], consisting in estimating and retransmitting the interference generated. Interestingly, and in contrast to [51], the interference terms are estimated *before* they even take place and are retransmitted in the same time slot. This principle could be applied in other wireless configurations where some nodes are more informed than others.

4.1 Preliminaries

4.1.1 Transmission Model

We focus in this chapter in a communications system where K TXs jointly serve K RXs over a Network MISO channel. We consider that each node (TX or RX) is equipped with a single-antenna. The assumption of single-antenna TX is done for ease of exposition, and the extension to multiple-antenna TX is straightforward. The transmit and received signal is defined as in Section 2.2. We assume that the transmitted multi-user signal satisfies an

average power constraint of P. The channel is assumed to be drawn from a continuous distribution with density such that all the channel matrices and their sub-matrices are full rank with probability one. Moreover, the channel coefficients are assumed to change after one channel use and to be independent from one channel use to another.

The transmitted multi-user signal $\mathbf{x}$ is obtained from the symbol vector $\boldsymbol{s} \in \mathbb{C}^{b\times 1}$ having its elements i.i.d. according to $\mathcal{N}_{\mathbb{C}}(0,1)$, where b is the number of independent data symbols delivered. We will differentiate in this chapter between the *private* data symbols, destined to be decoded at a particular user, and the *common* data symbols, broadcast and destined to be decoded at all users. Note that the term private is used only in contrast to common and does not refer to any privacy/secrecy constraint, but to the fact that only one user will decode the symbol.

4.1.2 Distributed CSIT Model

We consider the D-CSIT model described in Section 2.4, such that TX j receives the imperfect multi-user channel estimate $\hat{\mathbf{H}}^{(j)} = [\hat{\mathbf{h}}_1^{(j)}, \dots, \hat{\mathbf{h}}_K^{(j)}]^{\mathrm{H}} \in \mathbb{C}^{K\times K}$ where $(\hat{\mathbf{h}}_i^{(j)})^{\mathrm{H}}$ refers to the estimate at TX j of the channel from all TXs towards RX i. TX j then designs its transmit coefficients solely as a function of $\hat{\mathbf{H}}^{(j)}$ and the statistics of the channel. Since this chapter is focused on the DoF analysis, we use the same model as in the previous chapter, which has been described in Section 3.1.2, but with the particularity that $\bar{P}^{\gamma_{i,k}-1} = 1$ for any $i,k \in \mathbb{N}_K$. Therefore, we write the channel estimate as

$$\hat{\mathrm{h}}_{i,k}^{(j)} \triangleq \mathrm{h}_{i,k} + \bar{P}^{-\alpha_{i,k}^{(j)}} \delta_{i,k}^{(j)}. \tag{4.2}$$

We consider in this chapter the particular *Sorted CSIT* setting introduced in 2.4.4, in which we can order the TXs such that

$$1 \geq \alpha_{i,k}^{(1)} \geq \alpha_{i,k}^{(2)} \geq \cdots \geq \alpha_{i,k}^{(M)} \geq 0. \tag{4.3}$$

This assumption is made so as to avoid the unmanageable increment of possible CSIT regimes. Besides that, this regime is interesting also because it has proven to obtain important gains with respect to the lower-bound. Since the sorted structure applies to any channel coefficient in the same order, we can assume w.l.o.g. that $\alpha_{i,k}^{(j)} = \alpha^{(j)}$. Hence, $1 \geq \alpha^{(1)} \geq \cdots \geq \alpha^{(M)} \geq 0$, and the estimate matrix at TX j is given by

$$\hat{\mathbf{H}}^{(j)} = \mathbf{H} + \bar{P}^{-\alpha^{(j)}} \boldsymbol{\Delta}^{(j)}, \tag{4.4}$$

For later use, we also denote the i-th row of $\boldsymbol{\Delta}^{(j)}$ as $(\boldsymbol{\delta}_i^{(j)})^{\mathrm{H}}$, such that it holds that $\hat{\mathbf{h}}_i^{(j)} = \mathbf{h}_i + \bar{P}^{-\alpha^{(j)}} \boldsymbol{\delta}_i^{(j)}$. The multi-user distributed CSIT configuration can be hence

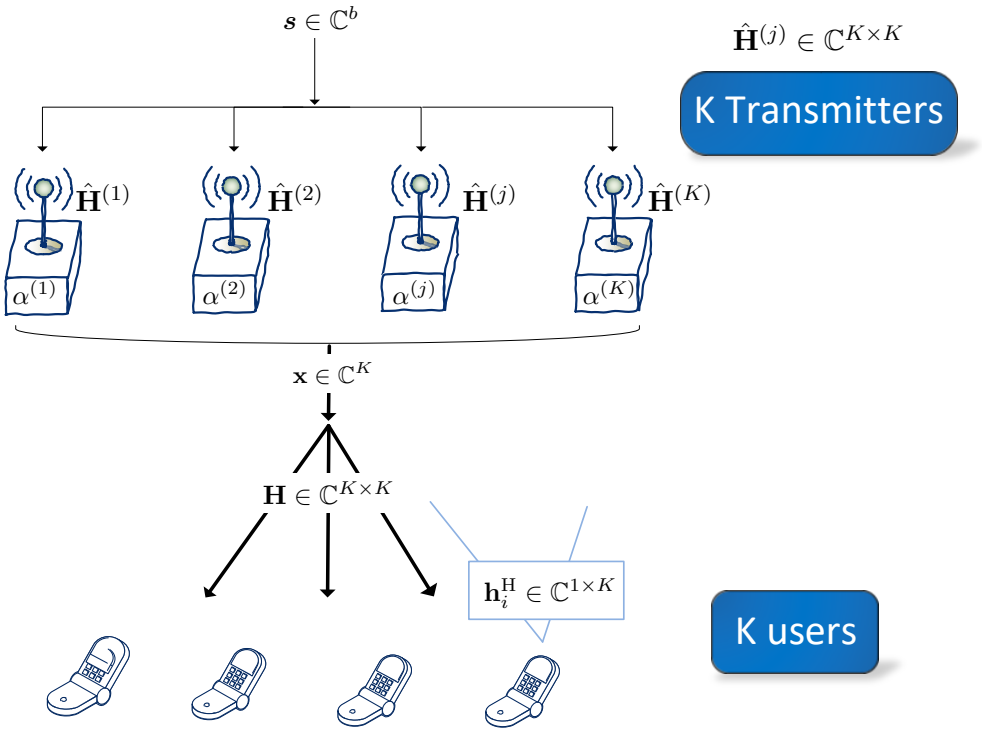


Figure 4.1 – $K \times K$ Network MISO setting with Distributed CSIT. The accuracy of the channel estimate at TX j is modeled through the CSIT scaling coefficient $\alpha^{(j)}$.

represented through the multi-TX CSIT scaling vector $\boldsymbol{\alpha} \in \mathbb{R}^K$ defined as

$$\boldsymbol{\alpha} \triangleq \{\alpha^{(1)}, \ \ldots, \ \alpha^{(K)}\}. \tag{4.5}$$

The parameters $\boldsymbol{\alpha}$ represent the average accuracy of the estimates. They are long-term coefficients that vary slowly in time and can be easily shared to all TXs. Consequently, the parameters $\boldsymbol{\alpha}$ are assumed in the following to be fixed and known at all TXs. As in the previous chapter, we assume that all channel realizations $\mathrm{h}_{i,k}$ and estimation noise variables $\delta_{i,k}^{(j)}$ satisfy the bounded density assumption of Definition 3.1. Furthermore, the channel realizations and the estimation noise are mutually independent.

4.1.3 Imperfect CSI Acquisition and Sharing

We consider in this thesis that the TX-dependent local imperfect multi-user channel estimate is obtained at the TXs from a CSI acquisition and sharing mechanism not discussed in this work. Yet, due to unavoidable delay and imperfections in the CSI sharing mechanism, this CSI sharing step leads to a setting where the TXs have received different

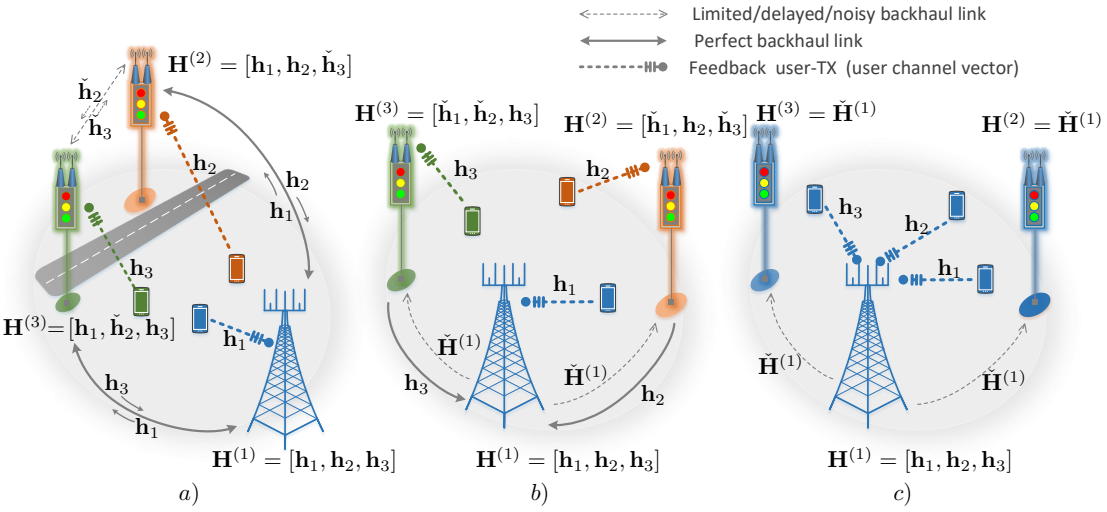


Figure 4.2 – Schematic illustration of three different example scenarios with distributed CSIT. FDD transmission is assumed. $\mathbf{h}_i$ denotes the (highly accurate) estimate of user i's channel that is fed back to the attached TX. $\mathbf{H}^{(j)}$ denotes the CSI matrix obtained at TX j after cooperation. $\check{\mathbf{H}}^{(j)}$ and $\check{\mathbf{h}}_i$ represent a quantized/coarse version of the respective estimates, transmitted through limited backhaul communications.

imperfect estimates of the true channel. We provide below several practical examples that illustrate different network configurations that lead to the D-CSIT configuration introduced in 4.1.2.

Example 4.1. In a network with several Base Stations (TXs) cooperating to jointly serve their users, each TX obtains a feedback from its attached user such that each TX will know accurately only a part of the channel state information. For example, in Time Division Duplexing (TDD) transmission with reciprocity, each TX will obtain a good estimate of its local channel; instead, in Frequency Division Duplexing (FDD) with user feedback, each TX will obtain a good estimate of the whole channel vector towards its attached user. Considering a CSI exchange step with heterogeneous links in the sense that one direction is of better quality, requires less quantization, or introduces less delay (which is a very reasonable assumption when considering heterogeneous networks where some of the TXs are UAVs [11, 12] or vehicles [152]), we obtain a setting where one of the TXs is uniformly more informed that the other. A particular example is depicted in Fig. 4.2.a, in which the link between TX 2 and TX 3 is a limited Device-to-Device (D2D) link.

Example 4.2. In addition, considering the previous setting in the case where CSI exchange is limited by a very restrictive delay, the sharing can be done by a transmission of the accurate CSI to a specific –main– TX, which then forwards a coarser version of the whole channel matrix to all TXs due to delay constraints. This retransmission from the main TX could also be broadcast, such that the resources spent on the sharing are reduced. Using layered encoding [153], every TX would obtain an estimate with a different accuracy. This setting is shown in Fig. 4.2.b.

Example 4.3. In a wireless network with one principal TX receiving feedback from all users to be served and several remote radio heads helping in the joint transmission, a distributed CSIT configuration is obtained when the CSI sharing from the main TX to the remote radio heads is done using limited and imperfect communication links, as illustrated in Fig. 4.2.c. It could also be envisioned that the remote radio heads directly acquire low precision channel state information from direct feedback from the users using layered encoding [153] or analog feedback [154]. Due to the weaker capabilities of the remote radio heads, a Distributed CSIT (D-CSIT) configuration with homogeneous quality at each TX would then be obtained.

We assume in the following that the CSI acquisition step has already occurred through limited and imperfect communication links and has led to each TX having access to its own imperfect estimate of the multi-user channel state.

4.1.4 CSIR Model

As in the important literature on delayed CSIT [51, 54, 58–61] we assume that the RX has been able to obtain perfect knowledge of the channel of the other RXs. This assumption is key to the approach used in this chapter. However, it is important to note that this assumption can be weakened as it is sufficient for the RXs to obtain the CSIT up to the best CSIT quality across the TXs (not necessarily the same estimate, but of the same quality). Furthermore, the estimate should be available at the RX for the decoding, such that its latency constraint stems from the user's data, not from the channel coherence.

4.2 DoF of the Distributed CSIT setting

As a preliminary, let us first state the optimal DoF of the centralized K-user BC setting where there exists a single estimate with a CSIT scaling coefficient α that is perfectly shared by all TXs. It was shown in [120] that the sum DoF in that configuration, denoted

by $\mathrm{DoF}^{\mathrm{CCSIT}}(\alpha)$, is equal to

$$\mathrm{DoF}^{\mathrm{CCSIT}}(\alpha) = 1 + (K-1)\alpha. \tag{4.6}$$

We can now present our main results.

4.2.1 Centralized Upper-bound

In a similar manner as for the 2×2 setting of the previous chapter, we can obtain an intuitive upper-bound for the $K \times K$ setting with distributed CSIT by assuming a genie-aided centralized setting in which each TX has access to the estimates of *all* the TXs. The following theorem is an extension of Lemma 3.1 for an arbitrary number of nodes.

Theorem 4.1. *In the $K \times K$ single-antenna Network MISO with distributed CSIT, the optimal DoF is upper-bounded by the DoF of a C-CSIT scenario in which the channel estimates of all the TXs are perfectly shared, such that each TX accesses to the set of estimates $\{\hat{\mathbf{H}}^{(j)}\}_{j\in\mathbb{N}_K}$. Concretely, let us define the set of scaling parameters of the D-CSIT and C-CSIT settings as*

$$\boldsymbol{\alpha} \triangleq \{\alpha_{i,k}^{(j)}\}_{i,j,k\in\mathbb{N}_K} \qquad \textit{and} \qquad \boldsymbol{\alpha}^{\star} \triangleq \{\max_{j\in\mathbb{N}_2} \alpha_{i,k}^{(j)}\}_{i,k\in\mathbb{N}_K}, \tag{4.7}$$

respectively. Then,

$$\mathrm{DoF}^{\mathrm{DCSIT}}(\boldsymbol{\alpha}) \leq \mathrm{DoF}^{\mathrm{CCSIT}}(\boldsymbol{\alpha}^{\star}). \tag{4.8}$$

Proof. The proof relies on the following important lemma.

Lemma 4.1. *Let $\hat{\mathcal{H}}^{(j)} \triangleq \mathcal{H} + \bar{P}^{-\alpha^{(j)}}\Delta^{(j)}$, where $\mathcal{H}$, $\Delta^{(j)}$, $\forall j \in \mathbb{N}_K$, are independent continuous random variables satisfying the Bounded Density assumption. Then, the conditional probability density function $f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\ldots,\hat{\mathcal{H}}^{(K)}}$ satisfies that*

$$\max_{\mathcal{H}} f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\ldots,\hat{\mathcal{H}}^{(K)}} = O\left(\bar{P}^{\max_{j\in\mathbb{N}_K}\alpha^{(j)}}\right). \tag{4.9}$$

Proof. The proof is relegated to Appendix D for clarity. ■

Let us now assume a genie-aided scenario where all channel estimates are exchanged between all the TXs. Such setting corresponds to a (logically) centralized scenario with a shared CSI composed by $\{\hat{\mathbf{H}}^{(1)}, \ldots, \hat{\mathbf{H}}^{(K)}\}$. Using Lemma 4.1, we obtain that the peak of the probability density function of this genie-aided scenario with multiple estimates has the same scaling as the centralized setting with only $\hat{\mathbf{H}}^{(1)}$. It then directly follows from the proof in [120, Section V.8] that the DoF of the genie-aided scenario, denoted by $\mathrm{DoF}^{\mathrm{CCSIT}}_{genie}(\boldsymbol{\alpha})$, is given by

$$\mathrm{DoF}^{\mathrm{CCSIT}}_{genie}(\boldsymbol{\alpha}) = \mathrm{DoF}^{\mathrm{CCSIT}}\left(\alpha^{(1)}\right). \tag{4.10}$$

From this equivalence, and the fact that providing with more information does not hurt, the proof is concluded. ■

Lemma 4.1 is expected to hold in a more general group of distributions, i.e., including cases where the different noise variables are partially correlated. Indeed, for the Gaussian case where the noise variables $\{\Delta^{(j)}_{i,k}\}_{\forall j \in \mathbb{N}_K}$ are drawn from partially correlated jointly Gaussian distributions, it is easy to show analytically that (4.9) is also satisfied.

Proposition 4.1. *Suppose that the estimation noise variables $\{\Delta^{(j)}_{i,k}\}_{\forall j \in \mathbb{N}_K}$ are drawn from partially correlated jointly Gaussian distributions. Then, Lemma 4.1 holds and thus it follows that*

$$\mathrm{DoF}^{\mathrm{DCSIT}}(\boldsymbol{\alpha}) \leq \mathrm{DoF}^{\mathrm{CCSIT}}(\alpha^{(1)}). \tag{4.11}$$

Proof. The proof is relegated to Appendix C.4 ■

Interestingly, the upper-bounds of both Theorem 4.1 and Lemma 4.1 also hold for the GDoF model, as the path-loss parameters do not affect the CSIT accuracy scaling, and hence the result of Lemma 4.1 is independent of the path-loss.

4.2.2 Distributed Lower-bound

In the following, we present a DoF lower-bound for the $K \times K$ Network MISO setting with distributed CSIT. We split the lower-bound in two different CSIT regimes, depending on whether it matches the centralized upper-bound of Theorem 4.1 or not. The proposed scheme achieving the lower-bound follows the same approach for both regimes.

a) Weak-CSIT Configuration

In this regime, defined rigorously in the following, the D-CSIT setting surprisingly attains the DoF of the genie-aided C-CSIT setting, as shown in the following theorem.

Theorem 4.2. *Suppose that the m first TXs have the same CSIT accuracy scaling, i.e., $\alpha^{(1)} = \cdots = \alpha^{(m)}$, $m < K$. Let us define α_m^{Weak} as*

$$\alpha_m^{\text{Weak}} \triangleq \frac{1}{1 + K(K - m - 1)}. \tag{4.12}$$

Then, if $\alpha^{(1)} \leq \alpha_m^{\text{Weak}}$, the sum DoF of the $K \times K$ Network MISO with Distributed CSIT satisfies

$$\text{DoF}^{\text{DCSIT}}(\boldsymbol{\alpha}) \geq \text{DoF}^{\text{CCSIT}}(\alpha^{(1)}). \tag{4.13}$$

Proof. The result follows directly from the analysis of the proposed scheme presented in detail in Section 4.4. ■

The regime for which $\alpha^{(1)} \leq \alpha_m^{\text{Weak}}$ is called the "m-TX Weak-CSIT" regime. For convenience, we simplify the notation and refer to it generally as the "Weak-CSIT regime". In the so-called Weak-CSIT regime, the upper-bound of Theorem 4.1 is tight. Surprisingly, for $m = 1$, the most heterogeneous case, the DoF depends only on the CSI quality at TX 1, although with the downside of reducing the range of possible CSIT configurations. For $m = K - 1$ it holds that $\alpha_{K-1}^{\text{Weak}} = 1$, and thus every CSIT configuration is included in the Weak-CSIT regime, which is consistent with the simple use of single-stream Active-Passive Zero-Forcing (AP-ZF) precoding presented in [82].

Remark 4.1. The fact that it is possible to achieve the DoF of the centralized upper-bound with badly informed TXs is a surprising result which is not expected to extend to many other CSIT configurations. Indeed, it can be intuitively seen using basic linear algebra that at least $K - 1$ dimensions are necessary to cancel $K - 1$ ZF constraints. □

b) Extension to Arbitrary CSIT Configurations

Theorem 4.2 shows the CSIT configurations for which the upper-bound of Theorem 4.1 is tight. In the following, we present a robust transmission scheme that builds on the transmission scheme attaining Theorem 4.2, which is extended to adapt to any CSIT configuration without restriction. The main challenge comes from the large number of CSIT scaling parameters, leading to an even larger (combinatorially large) number of possible CSIT configurations depending on their relative values. First, we define three terms that play an important role in the proposed transmission scheme.

Definition 4.1 (Transmitting TXs)**.** A TX is said to be a "Transmitting TX" if it is connected and sends information to the RXs. It may or may not use its instantaneous CSI for precoding.

This definition is made necessary by the distributed nature of the CSIT. Indeed, in contrast to the centralized setting where adding antennas cannot reduce the performance [49, 128], using additional antennas can decrease the achievable DoF in the distributed setting by creating additional interference. Hence, although considering a setting with K TXs, it may be beneficial in some CSI configurations to "turn off" some of the TXs and to use a smaller number of "Transmitting TXs".

> **Definition 4.2** (Active TXs)**.** A TX is said to be an "Active TX" if it is a Transmitting TX and it makes use of its instantaneous CSI.

> **Definition 4.3** (Passive TXs)**.** A TX is said to be a "Passive TX" if it is a Transmitting TX but it does not make use of its instantaneous CSI.

A more thorough explanation about these definitions is provided later on, along with the description of the proposed scheme. Interestingly, it will become clear which parameters are critical to optimize: Both the number of "Transmitting TXs" and the number of "Active TXs". In relation to these two notions, we present the following definition.

> **Definition 4.4** (Transmission Mode (n,k))**.** We define the Transmission Mode (n,k) as the transmission with k Transmitting TXs and $n \leq k$ Active TXs.

Building on these definitions, the following lower bound is exactly obtained by optimizing the performance of the proposed scheme over the different Transmission Modes.

Theorem 4.3. *The sum DoF of the $K \times K$ D-CSIT Network MISO with accuracy scaling parameters $\boldsymbol{\alpha}$ is lower-bounded by $\mathrm{DoF}^{\mathrm{APZF}}(\boldsymbol{\alpha})$, obtained by solving the following linear program:*

$$\mathrm{DoF}^{\mathrm{APZF}}(\boldsymbol{\alpha}) = \underset{\gamma_{n,k}}{\text{maximize}} \sum_{k=2}^{K}\sum_{n=1}^{k-1} \gamma_{n,k}\left(1+(k-1)\alpha^{(n)}\right) \tag{4.14}$$

$$\text{subject to} \quad \sum_{k=2}^{K}\sum_{n=1}^{k-1} \gamma_{n,k} = 1, \quad \gamma_{n,k} \geq 0, \tag{4.15}$$

$$\sum_{k=2}^{K}\sum_{n=1}^{k-1} d_{n,k}\gamma_{n,k} \geq 0, \tag{4.16}$$

where $\gamma_{n,k}$ is a time-sharing variable representing the percentage of time allocated to the Transmission Mode (n,k) and $d_{n,k} \triangleq 1 - \alpha^{(n)} - k(k-n-1)\alpha^{(n)}$.

Proof. The transmission scheme for a particular Transmission Mode is described in Section 4.4 and, building on this result, the explanation and proof of the theorem is given in Section 4.5. ■

The transmission scheme and the achieved DoF are obtained by solving a simple linear programming problem with low complexity. Interestingly, the expression $1 + (k-1)\alpha^{(n)}$ in (4.14) corresponds to the DoF achieved in the k-user centralized setting with k TXs sharing a CSIT of quality $\alpha^{(n)}$ (See (4.6)).

Remark 4.2. The linear program of Theorem 4.3 depends only on the $K-1$ best CSIT coefficients and not on $\alpha^{(K)}$. This property was already highlighted in [82] and follows from the fact that it is possible to solve $K-1$ linear equations with $K-1$ Active TXs, and thus serving K users at the same time. Consequently, it can always be assumed that $\alpha^{(K)} = 0$ without reducing the DoF. □

Besides that, it is remarkable that time sharing between only two Transmission Modes is sufficient,as presented in the following corollary.

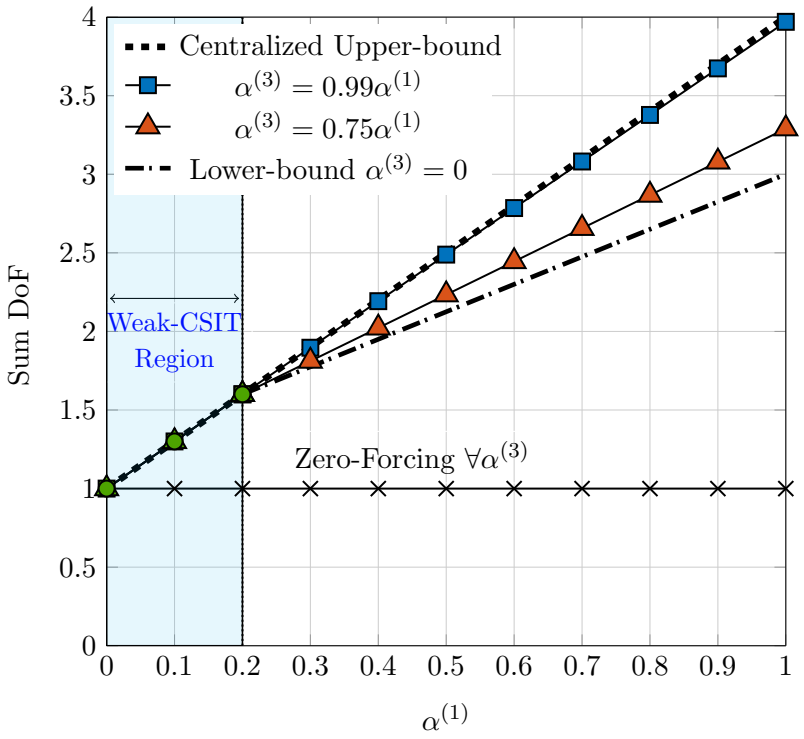
Corollary 4.1. *The linear program of Theorem 4.3 has always an optimal solution with only two Transmission Modes* (n_1, k_1) *and* (n_2, k_2)*, i.e.,*

$$\gamma_{n_1,k_1} > 0,\ \gamma_{n_2,k_2} \geq 0,\ \gamma_{n,k} = 0, \quad \forall (n,k) \notin \{(n_1,k_1),(n_2,k_2)\}. \tag{4.17}$$

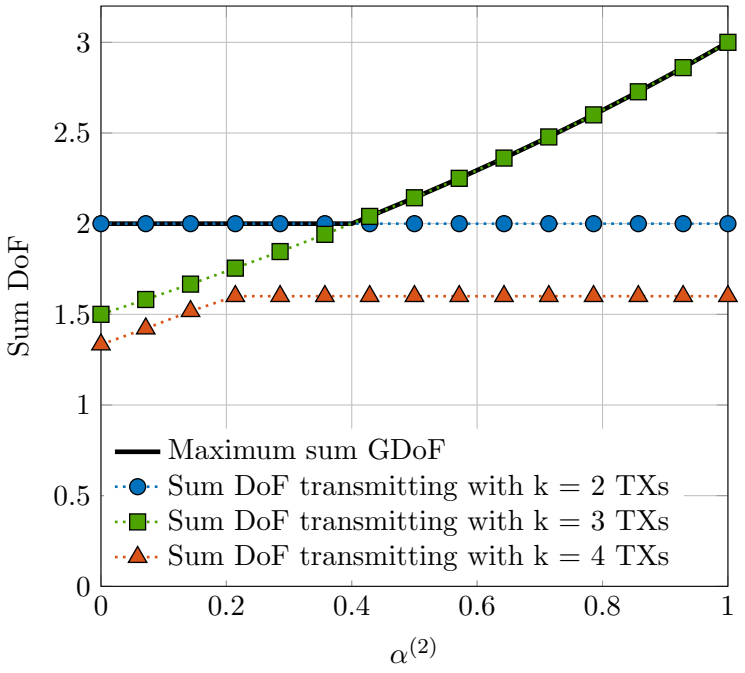
Proof. The proof is relegated to Appendix C.3. ■

Intuitively, if there are two modes of transmission, the first one is a generator of interference, i.e., it creates side-information at the RXs through the overloaded transmission, and relies on a successive second Transmission Mode to retransmit some of this interference –side information–, in order to decode the overloaded transmission. When only one mode is used, the interference is directly retransmitted through rate splitting using the common data symbol. Furthermore, the D-CSIT setting here assumed and the Delayed CSIT setting presented in [112] share an important feature: The more you overload the transmission, the better. This behavior arises as consequence of the fact that it is not possible to cancel out all the interference.

We show in Fig. 4.3a the DoF as a function of $\alpha^{(1)}$ for $K = 4$ TXs and $\alpha^{(2)} = \alpha^{(1)}$. We compare the achievable DoF with the centralized upper-bound for different values of $\alpha^{(3)}$. Up to $\alpha_2^{\text{Weak}} = 0.2$, the centralized upper-bound is achieved for any value of $\alpha^{(3)}$ –as stated in Theorem 4.2–, as well as when $\alpha^{(3)}$ becomes equal to $\alpha^{(1)}$, which is consistent with the results in [82].



(a) DoF obtained in the case with $K = 4$ TXs, for different values of $\alpha^{(3)}$ as a function of $\alpha^{(1)}$, whereas $\alpha^{(2)} = \alpha^{(1)}$, and $\alpha^{(4)} = 0$.



(b) Sum DoF of the $K = 4$-user Network MISO setting with distributed CSIT as function of $\alpha^{(2)}$, with $\alpha^{(1)} = 1$, and $\alpha^{(3)} = \alpha^{(4)} = 0$.

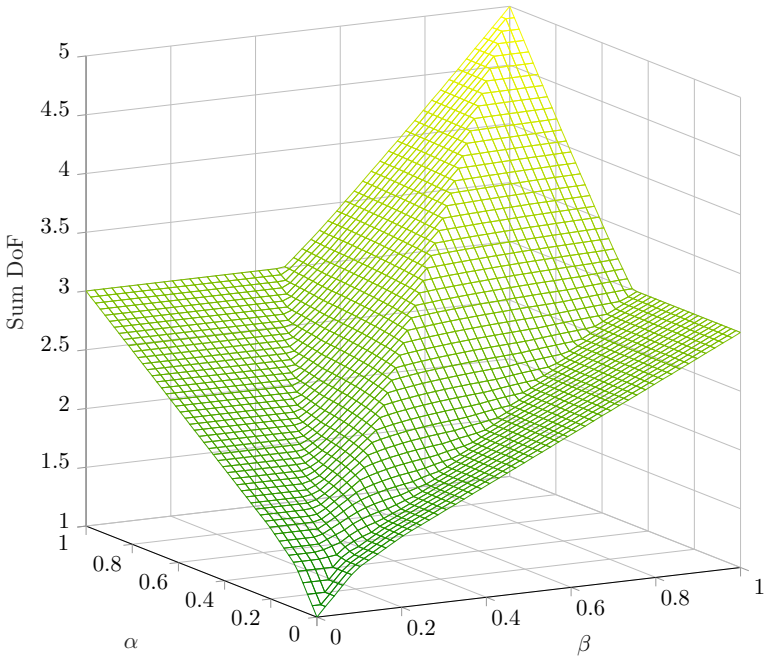


Figure 4.4 – Sum DoF for the 5×5 setting with $\alpha^{(1)} = \alpha^{(2)} = \alpha$, $\alpha^{(3)} = \alpha^{(4)} = \beta$, and $\alpha^{(5)} = 0$.

In Fig. 4.3b, we show the DoF achieved by AP-ZF for $K = 4$ TXs when we fix the number of Transmitting TXs (i.e., the value of k in Theorem 4.3) for the specific case where $\alpha^{(1)} = 1$, $\alpha^{(3)} = \alpha^{(4)} = 0$, and $\alpha^{(2)}$ varies from 0 to 1. Depending on the value of $\alpha^{(2)}$, it is optimal to use either 2 Transmitting TXs or 3 Transmitting TXs, but never $K = 4$ Transmitting TXs.

Finally, we illustrate our results for the 5×5 setting in which the CSIT scaling parameters are given by $\alpha^{(1)} = \alpha^{(2)} = \alpha$, $\alpha^{(3)} = \alpha^{(4)} = \beta$, and $\alpha^{(5)} = 0$. We depict in Fig. 4.4 the sum DoF region as function of α and β. Note that the metric is the sum DoF, not the DoF region. This figure illustrates that, even for simple and homogeneous cases, the sum DoF value does not follow a simple expression. It also shows how –in some cases– providing a TX with better CSI accuracy does not always improve the DoF: For $\alpha = 1$, we obtain the same DoF $= 3$ if $\beta < \frac{13}{30}$.

4.3 Illustrative Toy Scheme

We start by presenting a simple transmission scheme in a toy setting to exemplify the key features of our approach and convey the main intuition in a clear manner. The full scheme achieving the DoF of Section 4.2 will be described in Section 4.4.

We consider a 3-user setting with $\alpha^{(1)} = 0.1$, $\alpha^{(2)} = 0$, and $\alpha^{(3)} = 0$. The conventional regularized Zero-Forcing would achieve a DoF of 1, which is the same as for the no CSIT scenario. We will show how it is possible to achieve a DoF of $1 + 2\alpha^{(1)} = 1.2$, which is the value of the DoF that would be achieved in a centralized setting with TX 2 and TX 3 sharing the same estimate as TX 1 [120], such that there is no DoF loss from *not* sharing the CSIT between the TXs.

4.3.1 Encoding

The transmission scheme consists in a single channel use during which 3 private data symbols of $\alpha^{(1)} \log_2(P)$ bits are sent to each user –thus leading to 9 data symbols being sent in one channel use–. Additionally, a common data symbol of rate $(1 - \alpha^{(1)}) \log_2(P)$ bits is broadcast from TX 1 to all users using superposition coding [110]. Note that the information contained in this common data symbol is not only composed of "fresh" information bits destined to one user, but is also composed of side information necessary for the decoding of the private data symbols, as will be detailed below. The transmitted signal $\mathbf{x} \in \mathbb{C}^3$ is then equal to

$$\mathbf{x} = \boldsymbol{s}_1 + \boldsymbol{s}_2 + \boldsymbol{s}_3 + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} s_0 \tag{4.18}$$

where

- $\boldsymbol{s}_i \in \mathbb{C}^3$ is a vector containing the three private data symbols destined to RX i, each one with power $P^{\alpha^{(1)}}/9$ and rate $\alpha^{(1)} \log_2(P)$ bits.
- s_0 is the common data symbol transmitted *only from TX* 1 and destined to be decoded at all users, with power $P - P^{\alpha^{(1)}}$ and rate $(1 - \alpha^{(1)}) \log_2(P)$ bits.

The signal received at RX i, illustrated in Fig. 4.5, is equal to

$$y_i = \underbrace{\mathrm{h}_{i,1} s_0}_{\Theta_{\mathsf{P}}\left(\bar{P}\right)} + \underbrace{\mathbf{h}_i^{\mathrm{H}} \boldsymbol{s}_1}_{\Theta_{\mathsf{P}}\left(\bar{P}^{\alpha^{(1)}}\right)} + \underbrace{\mathbf{h}_i^{\mathrm{H}} \boldsymbol{s}_2}_{\Theta_{\mathsf{P}}\left(\bar{P}^{\alpha^{(1)}}\right)} + \underbrace{\mathbf{h}_i^{\mathrm{H}} \boldsymbol{s}_3}_{\Theta_{\mathsf{P}}\left(\bar{P}^{\alpha^{(1)}}\right)}, \tag{4.19}$$

where the power scaling is written under the bracket, and where the noise term has been neglected for clarity. Note that in this illustrative example no precoding is applied. The expression of the transmit signal in (4.18) enlightens two key features of the transmission scheme: i) the overloaded transmission in which we send *simultaneously* several data symbols to each RX, and ii) the use of superposition coding to transmit at the same time –but different power layers– the common symbol s_0 and the private symbols $\boldsymbol{s}_i$.

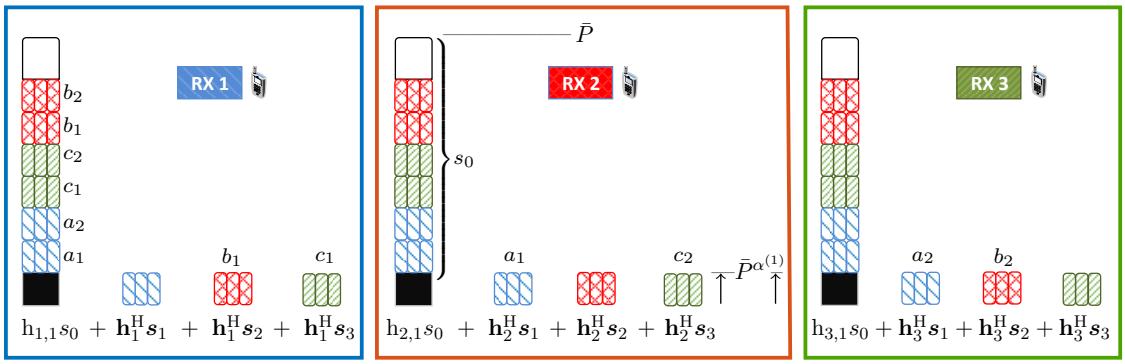


Figure 4.5 – Illustration of the received signals at every RX. Each RX receives its desired private data symbols and interference scaling both in $\bar{P}^{\alpha^{(1)}}$. Through superposition coding, it also receives the common data symbol s_0 containing a mix of fresh desired data symbols (illustrated in white), and side information to remove interference (illustrated with the color of the relevant RX).

4.3.2 Interference Estimation and Quantization at TX 1

The key element of the scheme is that the common data symbol s_0 is used to convey side information, enabling each user to decode its desired private data symbols. More specifically, TX 1 uses its local CSIT $\hat{\mathbf{H}}^{(1)}$ to *estimate* the interference terms $(\hat{\mathbf{h}}_i^{(1)})^{\mathrm{H}}\boldsymbol{s}_k$, $\forall i,k$ with $k \neq i$, which will be generated by the first layer of transmission. Then, TX 1 quantizes those $(\hat{\mathbf{h}}_i^{(1)})^{\mathrm{H}}\boldsymbol{s}_k$ terms and transmits them using the common data symbol s_0. Each interference term has a variance scaling in $\bar{P}^{\alpha^{(1)}}$ and is quantized using $\alpha^{(1)}\log_2(P)$ bits, such that the quantization noise can be made to remain at the noise floor using an appropriate uniform or Lloyd quantizer [110]. In total, the transmission of all the quantized estimated interference requires a transmission of $6\alpha^{(1)}\log_2(P)$ bits.

These $6\alpha^{(1)}\log_2(P)$ bits can be transmitted via the common data symbol s_0 if it holds that $6\alpha^{(1)}\log_2(P) \leq (1-\alpha^{(1)})\log_2(P)$, which is the case for the example considered here since $6 \times 0.1 < 1 - 0.1$. If the inequality is strict (as it is in this case), s_0 carries some additional $(1-7\alpha^{(1)})\log_2(P)$ fresh information bits —$0.3\log_2(P)$ bits here— to any particular user.

4.3.3 Decoding and DoF Analysis

It remains to verify that this transmission scheme achieves the claimed DoF. Let us consider w.l.o.g. the decoding at RX 1 as the decoding at the other users is the same up to a circular permutation of the RX's indexes. Note that signals at the noise floor are systematically omitted.

Using successive decoding [110], the common data symbol s_0 is decoded first, followed by the private data symbols $\boldsymbol{s}_1$. The data symbol s_0 of rate of $(1-\alpha^{(1)})\log_2(P)$ bits can

be decoded with a vanishing probability of error as its effective SNR can be seen in (4.19) to scale in $P^{1-\alpha^{(1)}}$. Upon decoding s_0, the quantized estimated interferences $(\hat{\mathbf{h}}_1^{(1)})^{\mathrm{H}}\boldsymbol{s}_2$ are obtained up to the quantization noise. As the quantization noise is at the noise floor, it is neglected in the following. RX 1 has then decoded:

$$\underbrace{(\hat{\mathbf{h}}_1^{(1)})^{\mathrm{H}}\boldsymbol{s}_2}_{\Theta_{\mathsf{p}}\left(P^{\alpha^{(1)}}\right)} = \mathbf{h}_1^{\mathrm{H}}\boldsymbol{s}_2 + \underbrace{\bar{P}^{-\alpha^{(1)}}(\boldsymbol{\delta}_1^{(1)})^{\mathrm{H}}\boldsymbol{s}_2}_{\Theta_{\mathsf{p}}\left(P^{0}\right)}. \tag{4.20}$$

This means that the interference terms $\mathbf{h}_1^{\mathrm{H}}\boldsymbol{s}_2$ can be suppressed up to the noise floor at RX 1. By proceeding in the same way with $(\hat{\mathbf{h}}_1^{(1)})^{\mathrm{H}}\boldsymbol{s}_3$, the remaining signal at RX 1 is

$$y_1 = \mathbf{h}_1^{\mathrm{H}}\boldsymbol{s}_1. \tag{4.21}$$

This signal, in combination with the interference terms $(\hat{\mathbf{h}}_2^{(1)})^{\mathrm{H}}\boldsymbol{s}_1$ and $(\hat{\mathbf{h}}_3^{(1)})^{\mathrm{H}}\boldsymbol{s}_1$ obtained through s_0, allows RX 1 to form a virtual received vector $\boldsymbol{y}_1^{\mathrm{v}} \in \mathbb{C}^3$ defined as

$$\boldsymbol{y}_1^{\mathrm{v}} \triangleq \begin{bmatrix} \mathbf{h}_1^{\mathrm{H}} \\ (\hat{\mathbf{h}}_2^{(1)})^{\mathrm{H}} \\ (\hat{\mathbf{h}}_3^{(1)})^{\mathrm{H}} \end{bmatrix} \boldsymbol{s}_1. \tag{4.22}$$

Each component of $\boldsymbol{y}_1^{\mathrm{v}}$ has an effective SNR scaling in $P^{\alpha^{(1)}}$ such that RX 1 can decode with a vanishing error probability its three data symbols of rate $\alpha^{(1)}\log_2(P)$ bits. Considering the three RXs, $9\alpha^{(1)}\log_2(P)$ bits have been transmitted through the private data symbols and $(1-7\alpha^{(1)})\log_2(P)$ bits through the common data symbol s_0, which yields a sum DoF of $1+2\alpha^{(1)}$.

Remark 4.3. Interestingly, the above scheme is based on interference estimation, quantization and retransmission, in a similar fashion as done in the different context of precoding with delayed CSIT (see e.g. [54, 58, 59]). Yet, we exploit in this work the distributed nature of the CSIT instead of the delayed knowledge of the CSIT such that in our scheme the interference are estimated and quantized *before even being generated*. □

4.4 Transmission Mode (n, k) with n Actives TXs and k Transmitting TXs

Leaving behind the toy example, we present in this section the Transmission Mode (n, k) with n Active TXs and k Transmitting TXs. We split the description such that we introduce the main structure of the transmission in Section 4.4.1. The precoding scheme

is described in detail in Section 4.4.2. The received signal is then studied in Section 4.4.3 and to conclude we compute the achieved DoF in Section 4.4.4 and Section 4.4.5. Note that in this section we enclose the transmit power in the data symbol vectors so as to ease the readability. Hence, a vector $\boldsymbol{s}_a \in \mathbb{C}^{1\times A}$ said to be transmitted such that each symbol is sent with power P^{α} satisfies that $\mathbb{E}[|\boldsymbol{s}_a|^2] = AP^{\alpha}$.

4.4.1 Encoding

We assume that there are only k Transmitting TXs, and only k RXs are served at a given transmission. Let $\mathcal{U}$ denote the set of RXs served and let us assume w.l.o.g. that the served RXs are the k first users, such that $\mathcal{U} = \{1, \dots, k\}$. Then, in the Transmission Mode (n, k), the transmitted signal $\mathbf{x} \in \mathbb{C}^k$ is given by

$$\mathbf{x} = \begin{bmatrix} 1 \\ \mathbf{0}_{k-1\times 1} \end{bmatrix} s_0 + \sum_{i=1}^{k} \mathbf{T}_i^{\text{APZF}} \boldsymbol{s}_i, \tag{4.23}$$

where

- $\boldsymbol{s}_i \in \mathbb{C}^{k-n}$ contains $k-n$ data symbols destined to RX i, which we hence denote as *private*, each one of rate $\alpha^{(n)} \log_2(P)$ bits and power $P^{\alpha^{(n)}}/(k(k-n))$, distributed in an i.i.d. manner. They are precoded with the AP-ZF precoder $\mathbf{T}_i^{\text{APZF}} \in \mathbb{C}^{k\times(k-n)}$ with n active TXs as described in detail in Subsection 4.4.2.
- $\boldsymbol{s}_0 \in \mathbb{C}$ is a data symbol destined to be decoded at all users and that we hence denote as *common*, of rate $(1-\alpha^{(n)}) \log_2(P)$ bits and power $P - P^{\alpha^{(n)}}$.

Therefore, $k-n$ data streams are sent to every RX but each RX has only one antenna. This overloaded transmission is necessary to take advantage of the $k-n-1$ interference terms generated by the RX's symbols at the other RXs, following the intuition from [51] that interference can be used as side information. This is detailed in Section 4.4.3.

A total of $k(k-n)\alpha^{(n)} \log_2(P)$ bits are sent in one channel use through the private data symbols. Furthermore, an additional data symbol of data rate $(1-\alpha^{(n)}) \log_2(P)$ bits is broadcast from TX 1. We will show that this common data symbol s_0 does not only contain new information bits, but also side information to enable the successful decoding of the private data symbols.

4.4.2 Precoding: AP-ZF with n Active TXs

The proposed precoder can be decoupled such that the precoder for each RX is computed independently up to a power normalization. We describe now the AP-ZF precoder serving a specific RX i with n Active TXs. This precoder allows to transmit $k-n$ streams to RX i

while reducing the interference at the n following RXs, i.e., at RXs $(i+t) \bmod [k]+1$, $\forall t \in \{1,\dots,n\}$. For ease of notation, we omit in the following the modulo operation as it is clear what an index bigger than k refers to. The precoder is obtained from distributed precoding at all TXs such that

$$\mathbf{T}_i^{\mathrm{APZF}} = \begin{bmatrix} \boldsymbol{e}_1^{\mathrm{T}} \mathbf{T}_i^{\mathrm{APZF}(1)} \\ \boldsymbol{e}_2^{\mathrm{T}} \mathbf{T}_i^{\mathrm{APZF}(2)} \\ \vdots \\ \boldsymbol{e}_k^{\mathrm{T}} \mathbf{T}_i^{\mathrm{APZF}(k)} \end{bmatrix}, \tag{4.24}$$

where $\boldsymbol{e}_\ell^{\mathrm{T}}$ refers to the ℓ-th row of the identity matrix $\mathbf{I}_{k\times k}$ and $\mathbf{T}_i^{\mathrm{APZF}(j)}$ denotes the AP-ZF precoder computed at TX j. Thus, we consider the design of $\mathbf{T}_i^{\mathrm{APZF}(j)}$ at TX j.

Remark 4.4. Note that although TX j computes the full precoder $\mathbf{T}_i^{\mathrm{APZF}(j)}$, only some coefficients are effectively used for the transmission due to the distributed precoding configuration, as made clear in (4.24). □

As a preliminary, let us define the *Active Channel* $\mathbf{H}_{\mathrm{A}} \in \mathbb{C}^{n\times n}$ as the channel coefficients from the Active TXs (TX 1 to TX n) to the RXs whose received interference is reduced (RX $i+1$ to RX $i+n$), i.e.,

$$\mathbf{H}_{\mathrm{A}} \triangleq \mathbf{H}_{i+1:i+n,1:n}. \tag{4.25}$$

Similarly, the *Passive Channel* $\mathbf{H}_{\mathrm{P}} \in \mathbb{C}^{n\times(k-n)}$ contains the channel coefficients from the Passive TXs (TX $n+1$ to TX k) to the RXs with reduced interference (RX $i+1$ to RX $i+n$), i.e.,

$$\mathbf{H}_{\mathrm{P}} \triangleq \mathbf{H}_{i+1:i+n,n+1:k}. \tag{4.26}$$

An illustration of the Active Channel and the Passive Channel is depicted in Fig. 4.6. The Passive TXs do not use their instantaneous CSIT. Hence, the *Passive Precoder* used is an arbitrarily chosen deterministic full rank matrix denoted by $\lambda_i^{\mathrm{APZF}}\mathbf{T}_i^{\mathrm{P}} \in \mathbb{C}^{(k-n)\times(k-n)}$, where $\lambda_i^{\mathrm{APZF}}$ is a constant used to satisfy an average sum power constraint and is detailed further down. On the other hand, every Active TX j, $\forall j \in \mathbb{N}_n$, computes $\mathbf{T}_i^{\mathrm{APZF}(j)} \in \mathbb{C}^{k\times k-n}$ on the basis of its own available CSIT $\hat{\mathbf{H}}^{(j)}$, such that

$$\mathbf{T}_i^{\mathrm{APZF}(j)} = \lambda_i^{\mathrm{APZF}} \begin{bmatrix} \mathbf{T}_i^{\mathrm{A}(j)} \\ \mathbf{T}_i^{\mathrm{P}} \end{bmatrix}. \tag{4.27}$$

The active precoder $\mathbf{T}_i^{\mathrm{A}(j)}$ is computed as

$$\mathbf{T}_i^{\mathrm{A}(j)} = -\left((\hat{\mathbf{H}}_{\mathrm{A}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{\mathrm{A}}^{(j)} + \frac{1}{P}\mathbf{I}_n\right)^{-1}(\hat{\mathbf{H}}_{\mathrm{A}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{\mathrm{P}}^{(j)}\mathbf{T}_i^{\mathrm{P}}. \tag{4.28}$$

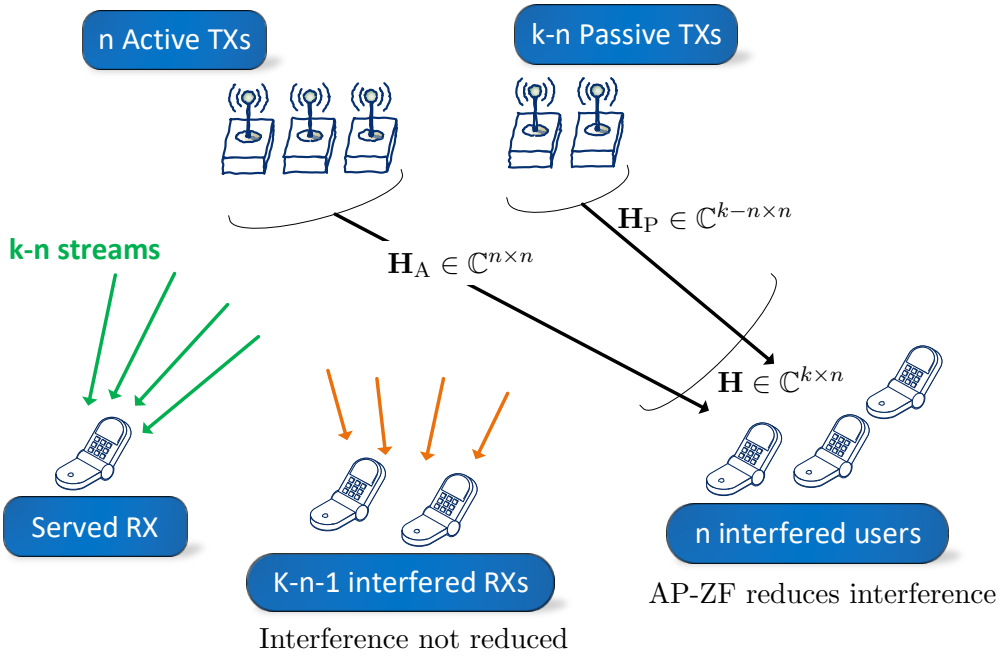


Figure 4.6 – AP-ZF illustration: The number n of Active TXs determines the number of RXs at which the interference is reduced, whereas the number of Passive TXs $(k-n)$ determines the number of independent streams that each RX can receive.

Remark 4.5. The design of the active precoder in (4.28) is an extension of the AP-ZF precoder introduced in [82]. Intuitively, the n Active TXs invert the channel to the n chosen RXs so as to cancel the interference generated by the Passive TXs. Interestingly, the number of Passive TXs limits the rank of the transmitted signal while the number of Active TXs limits the number of users whose received interference is attenuated. □

The AP-ZF precoder $\mathbf{T}_i^{\mathrm{APZF}} \in \mathbb{C}^{k \times k-n}$ actually applied in the transmission and defined in (4.24), can be written as

$$\mathbf{T}_i^{\mathrm{APZF}} \triangleq \lambda_i^{\mathrm{APZF}} \begin{bmatrix} \boldsymbol{e}_1^{\mathrm{H}} \mathbf{T}_i^{\mathrm{A}(1)} \\ \vdots \\ \boldsymbol{e}_n^{\mathrm{H}} \mathbf{T}_i^{\mathrm{A}(n)} \\ \mathbf{T}_i^{\mathrm{P}} \end{bmatrix}, \tag{4.29}$$

where the normalization coefficient $\lambda_i^{\mathrm{APZF}}$ is chosen as

$$\lambda_i^{\mathrm{APZF}} \triangleq \sqrt{\mathbb{E}\left[\left\| \begin{bmatrix} -\left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}} \mathbf{H}_{\mathrm{A}} + \frac{1}{P} \mathbf{I}_n\right)^{-1} \mathbf{H}_{\mathrm{A}}^{\mathrm{H}} \mathbf{H}_{\mathrm{P}} \mathbf{T}_i^{\mathrm{P}} \\ \mathbf{T}_i^{\mathrm{P}} \end{bmatrix} \right\|_{\mathrm{F}}^2 \right]^{-1}}. \tag{4.30}$$

This normalization constant $\lambda_i^{\mathrm{APZF}}$ requires only statistical CSI and can hence be computed at every TX, even the passive ones. It ensures that an average sum-power normalization constraint is satisfied, i.e., that

$$\mathbb{E}\left[\|\mathbf{T}_i^{\mathrm{APZF}}\|_{\mathrm{F}}^2\right] = 1. \tag{4.31}$$

The fundamental property of AP-ZF is that it effectively achieves interference reduction at the n RXs up to the *worst accuracy across the Active TXs*, as stated in the following lemma.

Lemma 4.2. *The AP-ZF precoder with n Active TXs satisfies*

$$\left\|\mathbf{h}_\ell^{\mathrm{H}}\mathbf{T}_i^{\mathrm{APZF}}\right\|_{\mathrm{F}}^2 = \mathcal{O}_{\mathsf{p}}\left(P^{-\alpha^{(n)}}\right), \quad \forall \ell \in \{i+1,\ldots,i+n\}. \tag{4.32}$$

Proof. The proof of Lemma 4.2 is given in Appendix C.2 along with the derivation of other important properties of AP-ZF precoding. ∎

4.4.3 Received Signals

Let us define the set $\mathcal{I}_i^{\mathrm{APZF}}$ as[1]

$$\mathcal{I}_i^{\mathrm{APZF}} \triangleq \{i+1,\ \ldots,\ i+k-n-1\}, \tag{4.33}$$

Intuitively, the set $\mathcal{I}_i^{\mathrm{APZF}}$ contains the RX indexes of the interfering signals that have not been attenuated towards RX i. The signal received at RX i is thus given by

$$y_i = \underbrace{\mathrm{h}_{i,1}s_0}_{\Theta_{\mathsf{p}}\left(\bar{P}\right)} + \underbrace{\mathbf{h}_i^{\mathrm{H}}\mathbf{T}_i^{\mathrm{APZF}}\boldsymbol{s}_i}_{\Theta_{\mathsf{p}}\left(\bar{P}^{\alpha^{(n)}}\right)} + \underbrace{\mathbf{h}_i^{\mathrm{H}}\sum_{\ell\in\mathcal{I}_i^{\mathrm{APZF}}}\mathbf{T}_\ell^{\mathrm{APZF}}\boldsymbol{s}_\ell}_{\Theta_{\mathsf{p}}\left(\bar{P}^{\alpha^{(n)}}\right)} + \underbrace{\mathbf{h}_i^{\mathrm{H}}\sum_{\ell\in\mathcal{U}\setminus\mathcal{I}_i^{\mathrm{APZF}}}\mathbf{T}_\ell^{\mathrm{APZF}}\boldsymbol{s}_\ell}_{\Theta_{\mathsf{p}}\left(\bar{P}^{0}\right)}, \tag{4.34}$$

where the noise term has been neglected for clarity. The last term in (4.34) scales as P^0 following the attenuation by $P^{-\alpha^{(n)}}$ due to AP-ZF precoding, as shown in Lemma 4.2. In Fig. 4.7 we illustrate the received signal at every RX for $k = 3$ Transmitting TXs and $n = 1$ Active TXs. We can see the improvement with respect to Fig. 4.5 since the number of significant interference terms is reduced by half thanks to the AP-ZF precoding. Since $n = 1$, the data symbol vector of a given RX can only be attenuated at a single RX. Note that, if we select $n = 2$, the interference could be attenuated at both interfered RXs, but

[1] We omit the modulo operation.

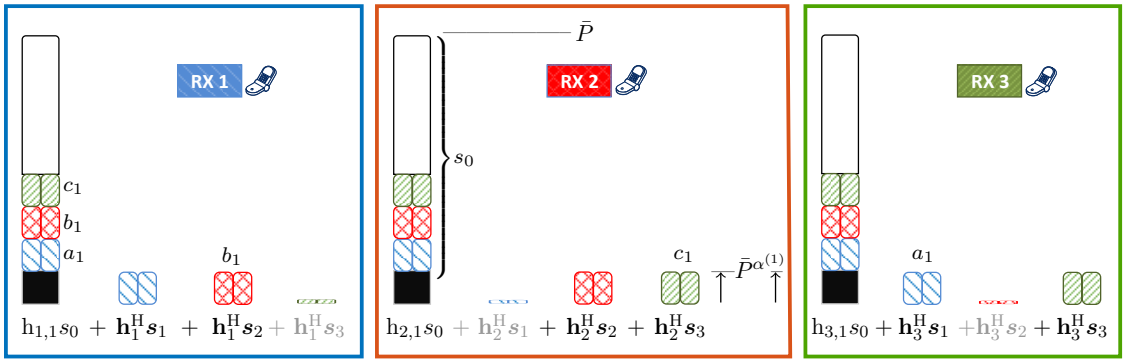


Figure 4.7 – Illustration of the received signals for the Weak-CSIT regime in the case of $k = 3$ Transmitting TXs and $n = 1$ Active TXs. Due to the AP-ZF precoding, the interference generated is reduced and thus extra new information can be sent through s_0 (white).

with a level of attenuation proportional to the CSI accuracy of the second TX, which is smaller than the accuracy at TX 1.

4.4.4 Decoding

TX 1 uses its local CSIT $\hat{\mathbf{H}}^{(1)}$ to estimate the interference terms $\mathbf{h}_i^H \mathbf{T}_\ell^{\text{APZF}} \boldsymbol{s}_\ell, \forall \ell \in \mathcal{I}_i^{\text{APZF}}$. Each interference term scales in $P^{\alpha^{(n)}}$ such that, by using $\alpha^{(n)} \log_2(P)$ bits, each term can be quantized with a quantization noise that lies at the noise floor [110]. Considering all users, this means that $k(k-n-1)\alpha^{(n)} \log_2(P)$ interference bits have to be transmitted. In order to do so, we use the broadcast data symbol s_0. If the quantity of information to be retransmitted exceeds the data rate of s_0, additional broadcast resources will need to be find to enable the successful decoding of the private data symbol. Fig. 4.8 shows an example of this problem. This is the essence of the linear optimization in Theorem 4.3 and will be discussed further in Section 4.5. We assume here that all the interference terms could be transmitted using the common data symbol s_0 and we will verify that it is indeed possible for a given RX i to decode its $(k-n)\alpha^{(n)} \log_2(P)$ intended bits.

By using successive decoding, the data symbol s_0 of rate of $(1-\alpha^{(n)}) \log_2(P)$ bits can be decoded with a vanishing probability of error as its effective SNR can be seen in (4.34) to scale as $P^{1-\alpha^{(n)}}$. Upon decoding s_0, we obtain the estimated interferences $(\hat{\mathbf{h}}_i^{(1)})^H \mathbf{T}_\ell^{\text{APZF}(1)} \boldsymbol{s}_\ell$, for $\ell \in \mathcal{I}_i^{\text{APZF}}$, up to the quantization noise at the noise floor. It then holds that

$$\begin{aligned}(\hat{\mathbf{h}}_i^{(1)})^H \mathbf{T}_\ell^{\text{APZF}(1)} \boldsymbol{s}_\ell &= \left(\mathbf{h}_i^H + \bar{P}^{-\alpha^{(1)}} (\boldsymbol{\delta}_i^{(1)})^H\right) \mathbf{T}_\ell^{\text{APZF}(1)} \boldsymbol{s}_\ell \\ &= \mathbf{h}_i^H \mathbf{T}_\ell^{\text{APZF}(1)} \boldsymbol{s}_\ell + \bar{P}^{-\alpha^{(1)}} (\boldsymbol{\delta}_i^{(1)})^H \mathbf{T}_\ell^{\text{APZF}(1)} \boldsymbol{s}_\ell \\ &= \mathbf{h}_i^H \mathbf{T}_\ell^{\text{APZF}} \boldsymbol{s}_\ell + \mathbf{h}_i^H \left(\mathbf{T}_\ell^{\text{APZF}(1)} - \mathbf{T}_\ell^{\text{APZF}}\right) \boldsymbol{s}_\ell,\end{aligned} \tag{4.35}$$

where the last equality is obtained after omitting the term $\bar{P}^{-\alpha^{(1)}}(\boldsymbol{\delta}_i^{(1)})^{\mathrm{H}}\mathbf{T}_\ell^{\mathrm{APZF}(1)}\boldsymbol{s}_\ell$, since its power scales as $P^{-\alpha^{(1)}}P^{\alpha^{(1)}} = P^0$ and thus it is negligible for the DoF analysis. Recall that $\boldsymbol{s}_\ell$ is transmitted with power scaling as $P^{\alpha^{(1)}}$. It holds that, $\forall \ell \in \mathbb{N}_k, \forall j \in \mathbb{N}_n$, the AP-ZF precoding satisfies the following property (see the proof in Appendix C.1):

$$\|\mathbf{T}_\ell^{\mathrm{APZF}(j)} - \mathbf{T}_\ell^{\mathrm{APZF}}\|_{\mathrm{F}}^2 = \mathcal{O}_{\mathsf{p}}\left(P^{-\alpha^{(j)}}\right). \tag{4.36}$$

It follows from (4.36) and the fact that $\boldsymbol{s}_\ell = \Theta_{\mathsf{p}}(\bar{P}^{\alpha^{(j)}})$ that

$$\mathbf{h}_i^{\mathrm{H}}\left(\mathbf{T}_\ell^{\mathrm{APZF}(1)} - \mathbf{T}_\ell^{\mathrm{APZF}}\right)\boldsymbol{s}_\ell = \Theta_{\mathsf{p}}(\bar{P}^0). \tag{4.37}$$

After having subtracted the quantized interference terms, the remaining signal at RX i up to the noise floor is

$$y_i = \mathbf{h}_i^{\mathrm{H}}\mathbf{T}_i^{\mathrm{APZF}}\boldsymbol{s}_i. \tag{4.38}$$

The key point of our approach is that RX i also receives through the broadcast data symbol the interference created by its own intended symbols at the *other RXs*, i.e., the estimated interference terms $(\hat{\mathbf{h}}_\ell^{(1)})^{\mathrm{H}}\mathbf{T}_i^{\mathrm{APZF}(1)}\boldsymbol{s}_i$, $\forall \ell$ such that $i \in \mathcal{I}_\ell^{\mathrm{APZF}}$ –note the swap of indexes i-ℓ with respect to previous expressions–. Each of those terms is an independent linear combination of the symbols $\boldsymbol{s}_i$, and thus RX i can form a virtual received vector $\boldsymbol{y}_i^{\mathrm{v}} \in \mathbb{C}^{k-n}$ equal to

$$\boldsymbol{y}_i^{\mathrm{v}} \triangleq \begin{bmatrix} \mathbf{h}_i^{\mathrm{H}} \\ (\hat{\mathbf{h}}_{i-1}^{(1)})^{\mathrm{H}} \\ \vdots \\ (\hat{\mathbf{h}}_{i-(k-n-1)}^{(1)})^{\mathrm{H}} \end{bmatrix} \mathbf{T}_i^{\mathrm{APZF}}\boldsymbol{s}_i. \tag{4.39}$$

Each component of $\boldsymbol{y}_y^{\mathrm{v}}$ has a SINR scaling in $P^{\alpha^{(n)}}$ and the AP-ZF precoder is of rank $k-n$ (See Lemma C.2 in Appendix C.1) such that RX i can decode its desired $k-n$ data symbols, each with the rate of $\alpha^{(n)}\log_2(P)$ bits.

Remark 4.6. The rank in (4.39) is ensured by the use of the *Passive TXs*. Hence, it is interesting to observe how uninformed TXs prove to be instrumental in the proposed scheme. □

4.4.5 DoF Analysis

We have considered the Transmission Mode (n,k) with n Active TXs and k Transmitting TXs. In this case, the transmission to each RX i creates $|\mathcal{I}_i^{\mathrm{APZF}}| = k-n-1$ interference

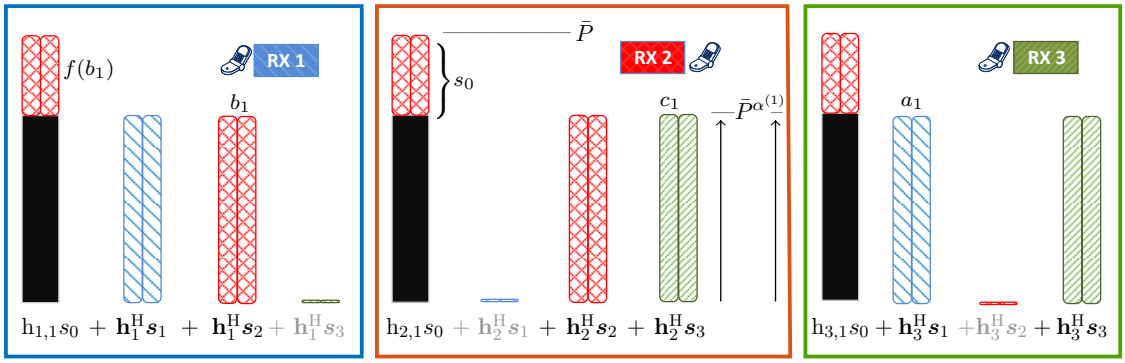


Figure 4.8 – Illustration of the signals received with $K = 3$ Transmitting TXs for the case where it is not possible to retransmit all the interference generated in the same channel use. As depicted, s_0 can not carry all a_1, b_1, c_1 as desired.

terms which gives in total $k(k-n-1)\alpha^{(n)}\log_2(P)$ bits that need to be retransmitted. Consequently, we define $\mathrm{DoF}_{n,k}^{\mathrm{Interf(-)}}$ as the DoF *consumed* in order to transmit these interference terms and which is given by

$$\mathrm{DoF}_{n,k}^{\mathrm{Interf(-)}} \triangleq k(k-n-1)\alpha^{(n)}. \tag{4.40}$$

In contrast, data symbol s_0 carries $(1-\alpha^{(n)})\log_2(P)$ bits, i.e., the DoF of the common data symbol $\mathrm{DoF}_{n,k}^{\mathrm{BC}}$ is given by

$$\mathrm{DoF}_{n,k}^{\mathrm{BC}} \triangleq 1-\alpha^{(n)}. \tag{4.41}$$

Finally, considering the $(k-n)\alpha^{(n)}\log_2(P)$ private bits for all k users leads to the private DoF denoted by $\mathrm{DoF}_{n,k}^{\mathrm{Priv}}$ and defined as

$$\mathrm{DoF}_{n,k}^{\mathrm{Priv}} \triangleq k(k-n)\alpha^{(n)}, \tag{4.42}$$

which is the DoF obtained from the private data symbols if all the interference is canceled. Putting (4.40), (4.41), and (4.42) together, the total DoF is

$$\mathrm{DoF}_{n,k} = \mathrm{DoF}_{n,k}^{\mathrm{Priv}} + \mathrm{DoF}_{n,k}^{\mathrm{BC}} - \mathrm{DoF}_{n,k}^{\mathrm{Interf(-)}} \tag{4.43}$$

at the condition that $\mathrm{DoF}_{n,k}^{\mathrm{BC}} - \mathrm{DoF}_{n,k}^{\mathrm{Interf(-)}} \geq 0$, i.e., that all interference terms could have been retransmitted. If this condition does not hold, the retransmission of the interference is managed through the time-sharing optimization of the different modes as discussed in Section 4.5. Conversely, the optimal result of Theorem 4.2 is achieved if the condition $\mathrm{DoF}_{n,k}^{\mathrm{BC}} - \mathrm{DoF}_{n,k}^{\mathrm{Interf(-)}} \geq 0$ is true for the Transmission Mode (m, K), where m is the

number of TXs with $\alpha^{(j)} = \alpha^{(1)}$. By solving the inequality $\mathrm{DoF}_{n,k}^{\mathrm{BC}} - \mathrm{DoF}_{n,k}^{\mathrm{Interf(-)}} \geq 0$, the maximum value of α_m^{Weak} is obtained from (4.40) and (4.41) as $\alpha_m^{\mathrm{Weak}} \triangleq \frac{1}{1+K(K-m-1)}$.

4.5 Proof of Theorem 4.3

Considering Transmission Mode (n, k), let us start by defining $d_{n,k}$ as the difference between the DoF available in the broadcast symbol s_0 and the DoF consumed by the interference to be retransmitted (see Section 4.4 for more details), i.e.,

$$\begin{aligned} d_{n,k} &\triangleq \mathrm{DoF}_{n,k}^{\mathrm{BC}} - \mathrm{DoF}_{n,k}^{\mathrm{Interf(-)}} \\ &= 1 - \alpha^{(n)} - k(k-n-1)\alpha^{(n)}. \end{aligned} \tag{4.44}$$

In fact, it is not required that each Transmission Mode leads to the transmission of all interference terms, it is only necessary that all interference terms were successfully transmitted at the end of the time sharing between all Transmission Modes. Mathematically, this interference retransmission constraint is written as

$$\sum_{k=2}^{K}\sum_{n=1}^{k-1} \gamma_{n,k} d_{n,k} \geq 0, \tag{4.45}$$

with $\gamma_{n,k}$ being the time sharing variable, such that $\gamma_{n,k} \geq 0$ and $\sum_{k=2}^{K}\sum_{n=1}^{k-1} \gamma_{n,k} = 1$. Interestingly, with that constraint, the sum DoF can then be rewritten as

$$\begin{aligned} \sum_{k=2}^{K}\sum_{n=1}^{k-1} \gamma_{n,k}\,\mathrm{DoF}_{n,k} &= \sum_{k=2}^{K}\sum_{n=1}^{k-1} \mathrm{DoF}_{n,k}^{\mathrm{Priv}} + \mathrm{DoF}_{n,k}^{\mathrm{BC}} - \mathrm{DoF}_{n,k}^{\mathrm{Interf(-)}} \\ &= \sum_{k=2}^{K}\sum_{n=1}^{k-1} \gamma_{n,k}\left(1 + (k-1)\alpha^{(n)}\right). \end{aligned} \tag{4.46}$$

The optimal time allocated to each Transmission Mode is obtained by maximizing the DoF over the time-sharing variables, i.e., optimizing the percentage of time in which each Transmission Mode is used. That problem leads to the following optimization problem.

$$\underset{\gamma_{n,k}}{\text{maximize}} \quad \sum_{k=2}^{K}\sum_{n=1}^{k-1} \gamma_{n,k}\left(1+(k-1)\alpha^{(n)}\right) \tag{4.47}$$

$$\text{subject to} \quad \sum_{k=2}^{K}\sum_{n=1}^{k-1} \gamma_{n,k} = 1, \quad \gamma_{n,k} \geq 0, \quad \sum_{k=2}^{K}\sum_{n=1}^{k-1} d_{n,k}\gamma_{n,k} \geq 0. \tag{4.48}$$

which concludes the proof of Theorem 4.3. ■

Interestingly, it is important to optimize over both the number of Transmitting TXs and the number of Active TXs. As shown in Fig. 4.3b, depending on the CSI allocation at each TX, using K Transmitting TXs can be detrimental. The number of Active TXs n controls how many RXs can have its received interference attenuated and up to which level that interference is reduced while the number of Transmitting TXs k controls how many users are served.

4.6 Discussion

For a given CSIT configuration, the linear program of Theorem 4.3 –recalled in (4.47)-(4.48)– has an immediate solution. Moreover, as stated in Corollary 4.1, the maximum DoF can be obtained always by means of two Transmission Modes at most. We discuss in the following several aspects of the previous results.

4.6.1 Number of Transmitting TXs

The first aspect is the fact that "turning-off" some TXs and multiplexing less RXs spatially can be beneficial. This is a consequence of the CSIT heterogeneity among the TXs, and it can be understood from the following example: Consider a setting with n TXs sharing perfect CSIT of the whole channel matrix, and an arbitrarily large number of RXs. If we start adding TXs to the joint transmission that do not have instantaneous CSIT, the DoF can be increased –as shown by the previous results–, and we can serve more RXs simultaneously. However, if the number of non-informed TXs keeps growing, at a certain point those TXs will create more interference whereas the informed TXs will be unable of canceling the interference out. Thus, after a certain total number of TXs k, it can be beneficial to schedule a sub-set of RXs (a number smaller than k) so as to avoid the collapse of DoF. This intuition shows why we consider also the Transmission Modes with $k < K$ Transmitting TXs. In order to highlight this property visually, we reuse the network setting of Fig. 4.3a for the following example.

> **Example 4.4.** Consider a network with $K = 4$ TXs and $K = 4$ RXs. Suppose that the forth TX has a useless CSI in terms of DoF ($\alpha^{(4)} = 0$), that TX 1 and TX 2 have the same CSI accuracy ($\alpha^{(1)} = \alpha^{(2)}$), and TX 3 has a CSI accuracy that is a fraction of the accuracy at TX 1, i.e., $\alpha^{(3)} = \mu\alpha^{(1)}$, with $0 \leq \mu \leq \alpha^{(1)}$. We show in Fig. 4.9 the DoF achieved by means of our proposed scheme.
>
> From the fact that $\alpha^{(4)} = 0$, applying conventional ZF in a naive manner will produce a DoF of 1, as TDM. Conversely, if $\alpha^{(3)} = \alpha^{(1)}$ ($\mu = 1$), we attain the

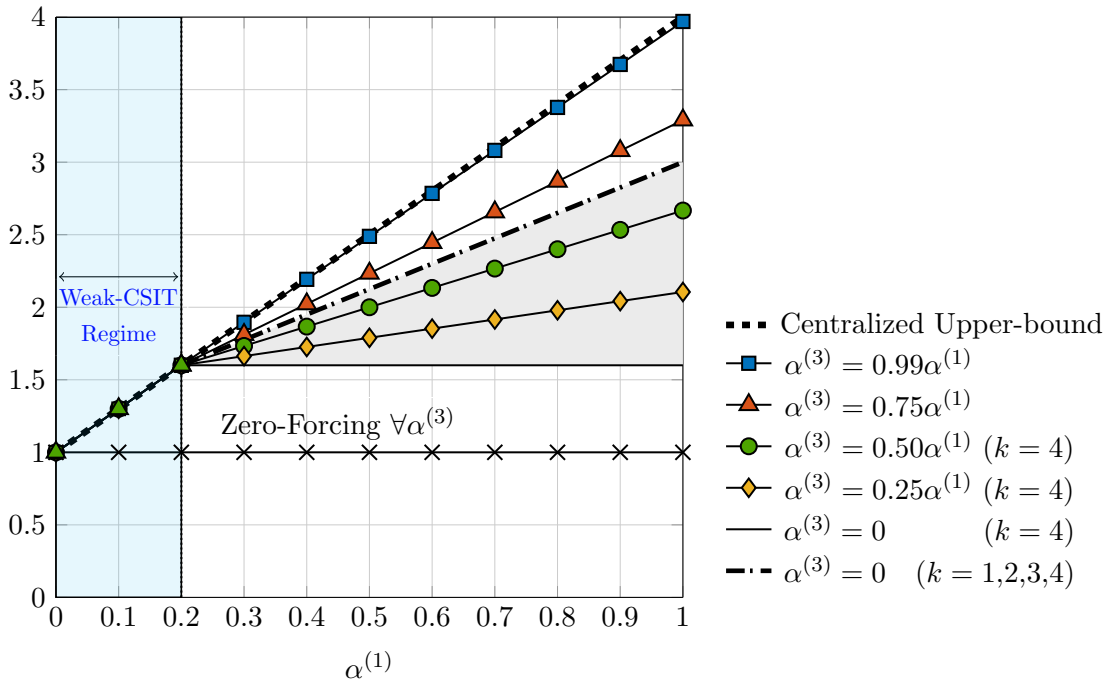


Figure 4.9 – DoF obtained in the case with $K = 4$ TXs, for different values of $\alpha^{(3)}$ as a function of $\alpha^{(1)}$, whereas $\alpha^{(2)} = \alpha^{(1)}$, and $\alpha^{(4)} = 0$. The region with gray background is the DoF gain obtained from "turning-off" some TXs and scheduling the RXs.

centralized DoF of $1 + (K - 1)\alpha^{(1)}$. In Fig. 4.3a, we presented the DoF achieved by the linear program of Theorem 4.3, in which we consider any number of Transmitting TXs. There, as well as in Fig. 4.9, it can be seen that the proposed scheme is robust against the D-CSIT setting, as the DoF increases proportionally to $\alpha^{(1)}$ even if $\alpha^{(3)} = 0$ –represented by the dash-dotted line in Fig. 4.9–.

This information was already represented in Fig. 4.3a. Besides this, we have depicted in Fig. 4.9 the maximum DoF performance if we fix the number of Transmitting TXs to $K = 4$. We observe that the performance is under the minimum performance of the general scheme for $\alpha^{(3)} \leq \frac{1}{2}$. This shows how adding an extra TX can harm the DoF performance of the joint transmission. The DoF gain that is obtained from the generalization of the scheme when $\alpha^{(3)} = 0$ is shown in gray background color. Importantly, this gap is significative, since for $\alpha^{(1)} = 1$ we double the DoF (from 1.6 to 3).

4.6.2 Number of Active TXs

The other parameter optimized in Theorem 4.3 is the number of TXs that make use of its instantaneous CSIT to compute the precoder –Active TXs–. This value is important

as it controls two dimensions of the transmission: First, the spatial dimension, in the sense that the number of RXs whose interference can be attenuated is equal to the number of Active TXs. Second, the power dimension, since the interference is attenuated up to a power that is proportional to the accuracy of the CSI at the last TX ($\alpha^{(n)}$). Specifically, during a Transmission Mode with n Active TXs, the interference received from the symbols of a certain RX i is attenuated at n RXs by a factor $\bar{P}^{\alpha^{(n)}}$.

Therefore, we can interpret this behavior as *how many bits of interference terms can we avoid to retransmit?*. From this point of view, with n Active TXs, the quantity of bits retransmitted is given by

$$\mathrm{DoF}_{n,k}^{\mathrm{Interf(-)}} = k(k-n-1)\alpha^{(n)}, \tag{4.49}$$

what follows from (4.40). Hence, the quantity of bits retransmitted is reduced by $kn\alpha^{(n)}$ bits with respect to the case with no interference cancellation through AP-ZF. Unfortunately, the value obtained if we add one Active TX more, $k(k-n)\alpha^{(n+1)}\log_2(P)$, can be either greater or smaller, and furthermore the rate obtained after correct decoding also varies. This implies that there exists a compromise on the number of Active TXs that is not straightforward. A qualitative illustration of this comments is depicted in Fig. 4.10 in the following page.

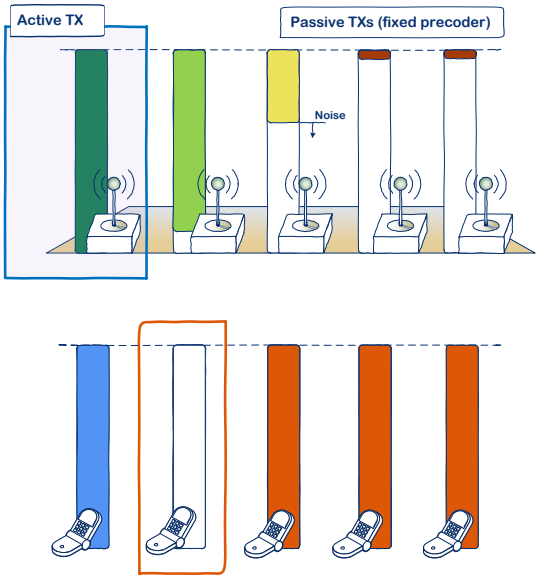
4.6.3 CSIT Allocation

We consider now a different problem with respect to the previous part of the chapter, where we have analyzed which is the best transmission scheme for a given distributed CSIT allocation. In this section, we are interested in the dual analysis: Assume that we use our proposed scheme, which is the best allocation for a given "*budget*" of CSIT?. Let us introduce a value A such that the sum of scaling coefficients α satisfies that

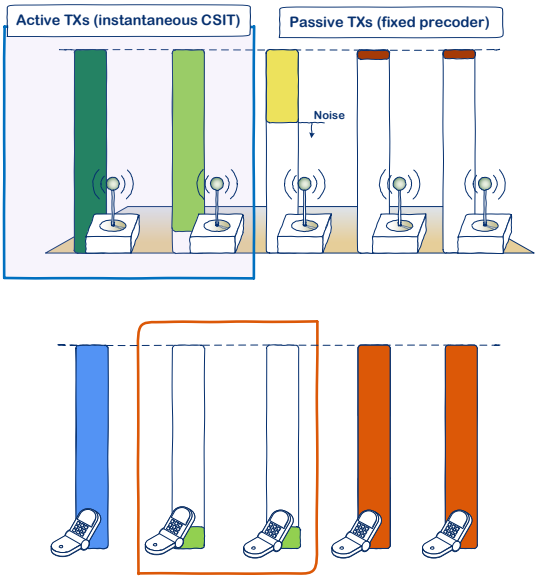
$$\sum_{j=1}^{K} \alpha^{(j)} = A. \tag{4.50}$$

In a scenario in which the RXs quantize its channel information and feed it back to the TXs, this is equivalent to say that there is a maximum number of quantization bits to be transmitted.

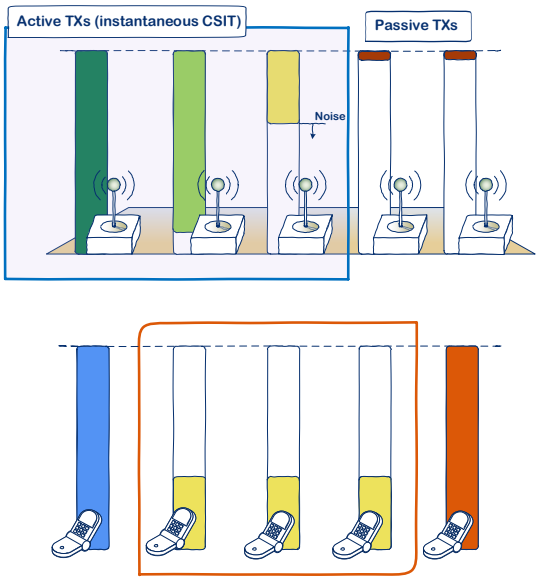
In the problem considered in Theorem 4.3, both $\alpha^{(j)}$ and $d_{n,k}$ are constant for any j, n, k, and it is a simple linear program. In the problem considered here, $\alpha^{(j)}$ becomes a variable and there is an extra constraint as $\sum_{j=1}^{K-1}\alpha^{(j)} = A$. This problem is highly complex. However, we can use Corollary 4.1 to simplify the expression.



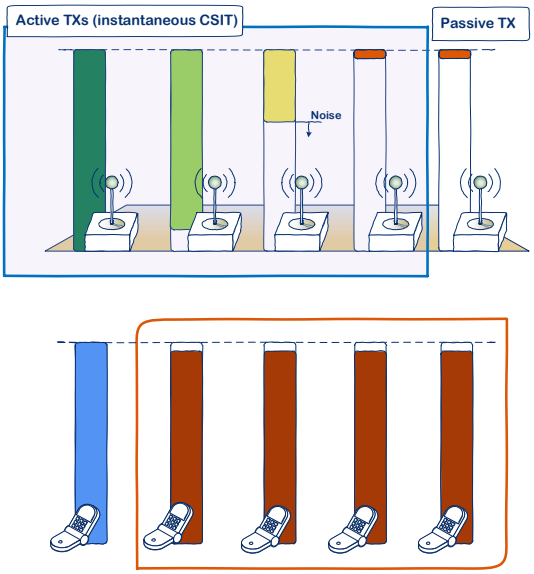
(a) Interference reduction for $n = 1$ Active TXs



(b) Interference reduction for $n = 2$ Active TXs



(c) Interference reduction for $n = 3$ Active TXs



(d) Interference reduction for $n = 4$ Active TXs

Figure 4.10 – Interference cancellation compromise as function of the number of Active TXs. The Active TXs are highlighted, as well as the RXs whose interference is attenuated through AP-ZF precoding. The first RX (blue) is the intended RX. The TX's bar represents the level of the estimation noise. The RX's bar represents the level of remaining interference. Note that increasing the number of Active TXs extend the number of RXs whose interference is attenuated, but it decreases the accuracy of such attenuation.

Lemma 4.3. *For a given total amount of CSIT A, the best CSIT allocation using the proposed scheme is given by solving the following program:*

$$\mathrm{DoF}^{\mathrm{APZF}}(\boldsymbol{\alpha}) = \underset{\substack{k_1,n_1,\\ k_2,n_2,\\ \boldsymbol{\alpha}}}{\text{maximize}}\ 1+\gamma(k_1-1)\alpha^{(n_1)}+(1-\gamma)(k_2-1)\alpha^{(n_2)} \tag{4.51}$$

$$\text{subject to}\ \ k_1,k_2 \in \{2,\dots,K\}, \tag{4.52}$$

$$n_1 \in \{1,\dots,k_1-1\}, \tag{4.53}$$

$$n_2 \in \{1,\dots,k_2-1\} \mid d_{n_2,k_2} \geq 0, \tag{4.54}$$

$$\sum_{j=1}^{K-1} \alpha^{(j)} = A. \tag{4.55}$$

where we have defined the constants $d_{n,k} \triangleq 1-\alpha^{(n)}-k(k-n-1)\alpha^{(n)}$ *and*

$$\gamma \triangleq \begin{cases} 1 & \text{if } d_{n_1,k_1} \geq 0, \\ \frac{d_{n_2,k_2}}{d_{n_2,k_2}-d_{n_1,k_1}} & \text{otherwise.} \end{cases} \tag{4.56}$$

Since the DoF and the interference generated in mode (n,k) depend only on n,k and $\alpha^{(n)}$, the DoF will be improved if we allocate the CSIT in a *homogeneous* way. For example, consider than $n_1 < n_2$. Then, the best interference cancellation accuracy is proportional to the quality at TX n_1. Hence, it is useless to have $\alpha^{(n)} > \alpha^{(n_1)}$, for any TX n such that $n < n_1$, since that extra information does not improve the DoF performance. Similarly, we have that the optimal CSIT accuracy sharing for the TXs $n_1 < n < n_2$ would be $\alpha^{(n)} = \alpha^{(n_2)}$, and intuitively $\alpha^{(n)} = 0$, for $n > n_2$. Consequently, the best CSIT allocation –when our proposed scheme is applied– follows always a layered structure. We extend the discussion on CSI allocation in Appendix E.

4.7 Conclusions

We have analyzed the DoF of the $K \times K$ Network MISO setting with distributed CSIT. We have described a novel D-CSIT robust transmission scheme that significantly improve the achieved DoF with respect to state-of-the-art precoding approaches when faced with distributed CSIT. We have first derived an upper-bound coined as the *centralized upper-bound* and consisting in a genie-aided setting where all the channel estimates are made available at all TXs. Then, we have shown how this genie-aided upper-bound was achieved by the proposed transmission scheme over a range of CSIT configurations, the

so-called "Weak-CSIT" regime. Surprisingly, this upper-bound can even be achieved with the CSI being handed at a single TX, i.e., with all other TXs having access to no CSIT. The proposed robust precoding scheme relies on new methods such as the estimation of the interference and their transmission from a single TX, as well as a modified ZF precoding allowing for an overloaded transmission. These new methods have a strong potential in other wireless configurations with TXs having access to different qualities of CSI. Converting these new innovative transmission schemes into practical transmissions schemes in realistic environments is an interesting and ongoing research direction. Such a robust precoding scheme could yield important gains in practice and make advanced precoding schemes more practical. Deriving tighter distributed upper-bounds is also an interesting and challenging research problem.

Chapter 5

DoF Analysis of the 2-user Distributed Network MIMO

In this chapter, we focus on the Network MIMO with 2 multi-antenna RXs. We consider that RX i has N_i antennas and that there are M transmit antennas. The transmit antennas can enclose an arbitrary number of physical transmitters, such that we can have M single-antenna distributed transmitters or 2 transmitters of $\frac{M}{2}$ antennas, for example. The particularity of this chapter is the assumption that only m of the M transmit antennas are endowed with accurate CSI. By analyzing this configuration, we intend to answer the question

To how many transmit antennas do we need to provide CSI?

The solution to the previous question helps also to answer the question of how much an extra informed antenna can help.

5.1 System Model

We analyze the 2-user Network MIMO where M transmit antennas aim to jointly serve 2 RXs endowed with N_1 and N_2 antennas, respectively. We assume w.l.o.g. that $N_1 \leq N_2$. Importantly, in this chapter we slightly modify the signification of TX with respect to the rest of the dissertation. Hereinafter, we refer to the i-th transmit antenna as TX_i.

This modification is motivated by the fact that in this chapter it is not relevant how many transmitters there are. The received signal at RX i is modeled as

$$\mathbf{Y}_i(t) \triangleq \bar{P}\,\mathbf{H}_i(t)\mathbf{X}(t) + \mathbf{N}_i, \tag{5.1}$$

The parameter t represents the channel use, P is the nominal SNR parameter, $\mathbf{Y}_i(t) \triangleq [Y_{i,1}(t), Y_{i,2}(t), \ldots, Y_{i,N_i}(t)]^{\mathrm{T}}$, and $\mathbf{H}_i \in \mathbb{R}^{N_i \times M}$ denotes the matrix of channel coefficients for RX i. We define the global channel matrix as $\mathbf{H} \triangleq [\mathbf{H}_1,\ \mathbf{H}_2]^{\mathrm{T}}$, $\mathbf{H} \in \mathbb{R}^{(N_1+N_2)\times M}$. The vector $\mathbf{X}(t) \triangleq [X_1(t), X_2(t), \ldots, X_M(t)]^{\mathrm{T}}$ is the transmit signal with unitary power constraint. $\mathbf{N}_i$ denotes the AWGN noise at RX i. We study here the scenario where the first m antennas (TXs) have perfect CSIT of the whole multi-user channel matrix, whilst the other $M-m$ antennas have only finite precision CSIT. We define:

- $\mathrm{TX}_\star \triangleq [\mathrm{TX}_1, \ldots, \mathrm{TX}_m]$, the m transmit antennas with perfect CSI. Consequently, we have that $\hat{\mathbf{H}}^{(j)} = \mathbf{H}$ for any $j \leq m$. The results of this chapter also hold if these TXs are endowed with a CSI with *accuracy scaling* parameter $\alpha = 1$.
- $\mathrm{TX}_\varnothing \triangleq [\mathrm{TX}_{m+1}, \ldots, \mathrm{TX}_M]$, the $M-m$ transmit antennas with finite precision CSI, i.e., $\alpha_{i,k}^{(j)} \leq 0$ for any $m < j \leq M$ and for any i, k.

Similarly, we denote as $\mathbf{H}_{i,\varnothing}$ (resp. $\mathbf{H}_{i,\star}$) the channel coefficient matrix from $\mathrm{TX}_\varnothing$ (resp. $\mathrm{TX}_\star$), and as $\mathbf{X}_\varnothing$ (resp. $\mathbf{X}_\star$) the transmit signal from $\mathrm{TX}_\varnothing$ (resp. $\mathrm{TX}_\star$). The superscript $^{[n]}$ in any variable X stands for $\{X(1), X(2), \ldots, X(n)\}$. For any set of variables $\mathcal{S}$, $H\big(\bigcap_{S_i \in \mathcal{S}} S_i\big)$ denotes the joint entropy of the elements in $\mathcal{S}$.

In this chapter, we make use of the Aligned Image Set technique introduced by Davoodi and Jafar [120] for the proof of the converse. For that, we need to consider the same assumptions as in [120]. Hence, we consider that the channel satisfies the Bounded Density assumption –presented in Definition 3.1, and that the channel coefficients are bounded away from 0 and infinity.

Definition 5.1 ([155, Definition 4])**.** For real numbers $x_1, x_2, \ldots, x_K$, define the notations $L_j^b(x_i,\ i \in \mathbb{N}_K)$, and $L_j(x_i, i \in \mathbb{N}_K)$, as

$$L_j^b(x_1, x_2, \ldots, x_k) = \sum_{i \in \mathbb{N}_K} \lfloor g_{j,i} x_i \rfloor \tag{5.2}$$

$$L_j(x_1, x_2, \ldots, x_k) = \sum_{i \in \mathbb{N}_K} \lfloor h_{j,i} x_i \rfloor \tag{5.3}$$

for distinct random variables $g_{j,i} \in \mathcal{G}$, and for some arbitrary real valued and finite constants $h_{j,i} \in \mathcal{H}$, $|h_{j,i}| \leq \delta_z < \infty$. The subscript j is used to distinguish among multiple sums.

5.2 DoF Region of the (M, N_1, N_2) Network MIMO

The main goal of this chapter is to characterize the DoF region of the Distributed Network MIMO with 2 multi-antenna RXs. For that purpose, we have considered a simple D-CSIT configuration, in which the TXs have either perfect CSI or finite precision CSI. The assumption of this particular configuration allows us to drop the impact of imperfect CSI –in the sense of noisy CSI– and highlight the impact of Decentralized CSIT. In particular, this analysis helps to understand *the benefit of providing a TX with CSI*. We present in the following theorem an outer-bound of the DoF region of the setting considered.

Theorem 5.1. *Suppose a 2-user Network MIMO in which RX i has N_i antennas and there are M transmit antennas (TXs). Suppose that $N_1 \leq N_2$ and that m antennas have perfect CSI of the whole channel matrix, whereas the other $M - m$ TXs are endowed only with finite precision CSI. Then, the DoF region is enclosed in*

$$(d_1, d_2) \in \begin{cases} d_1 \leq \min(M, N_1) \\ d_2 \leq \min(M, N_2) \\ d_1 + d_2 \leq \min(M, N_1 + N_2) \\ \frac{d_1}{\min(M, N_1+N_2)-m} + \frac{d_2-m}{N_2-m} \leq 1 \end{cases} \tag{5.4}$$

if $m < \min(N_2, M - N_1)$, and

$$(d_1, d_2) \in \begin{cases} d_1 \leq \min(M, N_1) \\ d_2 \leq \min(M, N_2) \\ d_1 + d_2 \leq \min(M, N_1 + N_2) \end{cases} \tag{5.5}$$

if $m \geq \min(N_2, M - N_1)$.

The proof is relegated to Section 5.5. As a direct consequence of the previous theorem, we can infer the following sum DoF upper-bound.

Lemma 5.1. *The sum DoF of the 2-user Network MIMO (M, N_1, N_2, m) is upper-bounded by*

$$d_1 + d_2 \leq \min\left(N_1 + N_2,\ M,\ N_2 + N_1 \frac{\min(N_1, M - N_2)}{\min(N_1 + N_2, M) - m}\right). \tag{5.6}$$

We have presented the upper-bound results for the considered setting. We introduce the achievability results hereinafter. The following theorem shows the achievable DoF for the setting with $m \geq N_1$.

Theorem 5.2. *The sum DoF of the 2-user Network MIMO (M, N_1, N_2, m) is lower-bounded by*

$$d_1 + d_2 \geq \min\left(N_1 + N_2,\ M,\ N_2 + N_1 \frac{\min(N_1, M - N_2)}{\min(N_1 + N_2, M) - m}\right) \tag{5.7}$$

if $m \geq N_1$.

Proof. The proof follows from the transmission scheme and is given in Section 5.4. ■

Thus, the sum DoF bound of Lemma 5.1 is tight for $m \geq N_1$. More importantly, this result implies that the DoF region of Theorem 5.1 can be easily shown to be exact for $m \geq N_1$. The transmission scheme achieving Theorem 5.2 is based on the Active-Passive Zero-Forcing precoding introduced in Section 4.4. For the simple case where $M = N_1 + N_2$,

$$d_1 + d_2 = \begin{cases} N_2 + N_1 & \text{if } m \geq N_2 \\ N_2 + N_1 \frac{N_1}{N_1 + N_2 - m} & \text{if } N_1 \leq m < N_2. \end{cases} \tag{5.8}$$

Unfortunately, no tight general bound is known for the case $m < N_1$, apart from particular cases. Nevertheless, we can extend the achievable scheme that attains Theorem 5.2 to obtain a general lower-bound.

Proposition 5.1. *Consider $m < N_1$. Then, the sum DoF of the 2-user Network MIMO (M, N_1, N_2, m) is lower-bounded by*

$$d_1 + d_2 \geq \max\left(\min(N_2, M),\ \min(N_2, M - m) + \frac{m^2}{\min(N_2, M - m)}\right). \tag{5.9}$$

Proof. The proof is relegated to Section 5.6, along with a specific lower bound for a particular setting. ■

5.3 Discussion

The sum DoF of the 2-user MIMO BC with perfect CSIT is $\text{DoF}^\star = \min(M, N_2 + N_1)$ [96]. Hence, it holds that we only need perfect CSIT at $m = \min(N_2, M - N_1)$ to recover

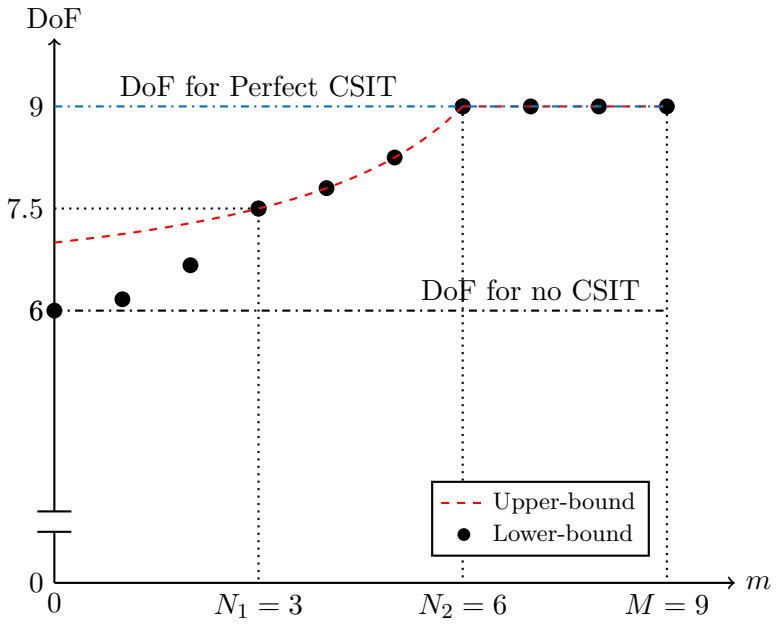


Figure 5.1 – DoF as function the number of transmit antennas with perfect CSIT (m) for the case $(M, N_1, N_2) = (9, 6, 3)$.

the maximum DoF. This aftermath extends the results of the previous chapters, in which we have shown that having the most accurate CSI at only a subset of TXs is –sometimes– enough to recover the DoF achieved with perfect CSI sharing. Moreover, we have quantized the gain from providing perfect CSI to an extra transmit antenna for the cases in which $m > N_1$. An interesting question that arises from this analysis is to determine whether adding a perfectly informed antenna can boost the DoF in more than 1. Consider the setting with $(M, N_1, N_2) = (2N, N, N)$. In this case, if $m = N$ we obtain that DoF $= 2N$, whereas if $m = 0$ we have that DoF $= N$. Thus, either the gain is exactly 1 DoF for each extra perfectly-informed antenna, or providing perfect global CSI to an extra antenna do increase the DoF in more than 1.

In Fig. 5.1, we show the sum DoF as function of m. We observe how for $m \geq N_2$ the DoF with centralized perfect CSIT is attained, and that for $N_1 \leq m \leq N_2$ the bound is tight. For the case $m < N_1$, we can see that there exists a gap between the upper and the lower bound. It is easy to infer that the upper-bound is loose from the fact that for $m = 0$, we obtain that DoF $= N_2 + 1$, but it is known that the DoF of a BC setting with finite precision CSIT is DoF $= N_2$. It is noteworthy that, the closer m is to the number of antennas of any of the RXs, the bigger increment of DoF from m to $m+1$. In Fig. 5.2, we illustrate the DoF as function of the *repartition* of antennas among the RXs, i.e., for a fixed size setting with $N_1 + N_2 = M = 20$ and m transmit antennas with perfect CSIT,

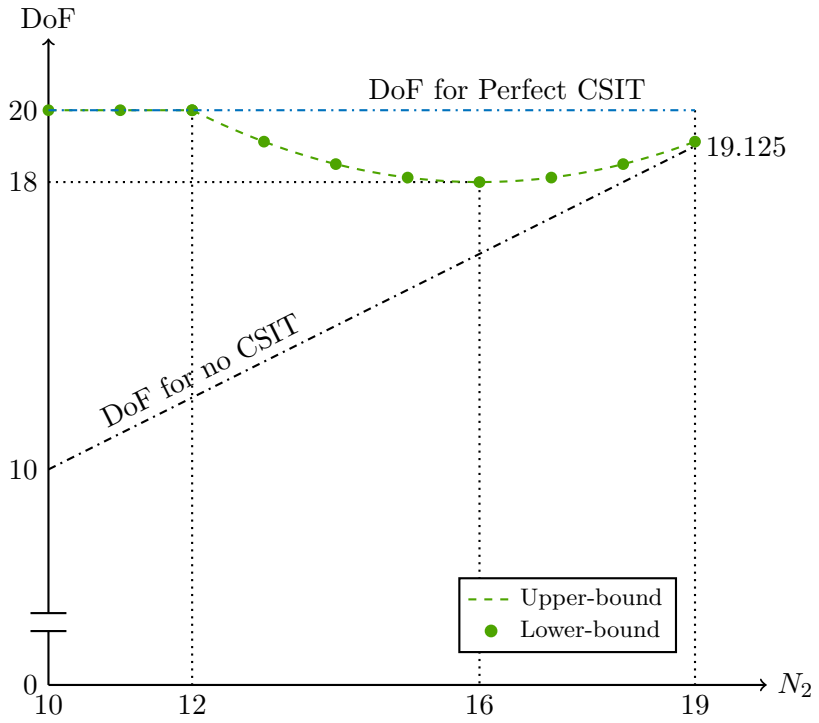


Figure 5.2 – DoF as function of N_2 for the setting $(M, m) = (20, 12)$ and $N_2 + N_1 = M$.

we plot the DoF as function of N_2. N_1 is then obtained as $20 - N_2$.

Besides this, the DoF expression obtained for the D-CSIT setting has a noteworthy similarity with the DoF expression of the C-CSIT setting with hybrid CSIT in which the TX has perfect CSI for one RX and delayed CSIT for the other RX (denoted as "PD" setting). We define

$$\begin{aligned} \mathrm{DoF}^{Distr} &= (d_1^{Distr}, d_2^{Distr}) && \text{DoF pair for the D-CSIT scenario considered.} \\ \mathrm{DoF}^{PD} &= (d_1^{PD}, d_2^{PD}) && \text{DoF pair for Hybrid C-CSIT (Perfect-Delayed) scenario.} \end{aligned}$$

By way of example, consider the scenario in which the number of antennas at the TX (M) is the same as the sum of receive antennas, i.e., $M = N_1 + N_2$. Furthermore, suppose that the number of transmit antennas with *perfect global CSIT* (m) satisfies $N_1 \leq m < N_2$. This assumption is made so as to consider the interesting bound of

$$\frac{d_1}{\min(M, N_1 + N_2) - m} + \frac{d_2 - m}{N_2 - m} \leq 1. \tag{5.10}$$

We compare the D-CSIT setting with the MIMO BC with centralized CSIT when the TX is endowed with perfect CSIT for RX 1's channel and delayed CSIT for RX 2's channel.

The upper-bound for the PD case was found by R. Tandon in [60] and, for the case with $M = N_1 + N_2$, it writes as

$$\frac{d_1^{PD}}{N_1 + N_2} + \frac{d_2^{PD}}{N_2} \leq 1. \tag{5.11}$$

This weighted expression leads to a sum DoF of

$$d_1^{PD} + d_2^{PD} = N_1 \quad + \quad N_2 - N_2 \frac{N_1}{N_1 + N_2}. \tag{5.12}$$

On the other hand, we have obtained that $d_1^{Distr} + d_2^{Distr}$ is given by

$$d_1^{Distr} + d_2^{Distr} = N_1 \quad + \quad N_2 - (N_2 - m) \frac{N_1}{N_1 + N_2 - m}. \tag{5.13}$$

We can see that there exists an analogy between both settings. In particular:

1. In the PD setting, the loss of DoF due to having delayed CSIT at RX 2 instead of perfect CSIT is $-N_2 \frac{N_1}{N_1+N_2}$.

2. In the D-CSIT case, the loss of DoF due to having perfect CSIT *only* at m antennas is $-(N_2 - m) \frac{N_1}{N_1+(N_2-m)}$.

Therefore, the D-CSIT is analogous to the PD case where only $(N_2 - m)$ have delayed CSIT instead of perfect CSIT. An intuition behind this result is that, in the D-CSIT setting, we can apply a change of basis at RX 2 so that the TXs with perfect CSIT are only listened by m antennas of RX 2. Therefore, even if some of TXs have CSI for the other $N_2 - m$ antennas, those $N_2 - m$ antennas receive only information for the finite precision CSIT TXs. Hence, it is analogous to have a delayed CSI, as the CSI cannot be used instantaneously even if it is already known.

5.4 Achievability for the Case $m \geq N_1$

We present here the DoF-optimal transmission scheme for $N_1 \leq m < N_2$, i.e., the proof of Theorem 5.2. The achievable scheme for the case $M \leq N_2$ (DoF $= M$) is trivial. For $M > N_2$, the scheme is composed of two phases. The transmission scheme makes use of the Active-Passive ZF (AP-ZF) introduced in the previous chapter. We refer to Chapter 4 for details about AP-ZF. Let us first consider the case $M = N_1 + N_2$, since the general case follows after a straightforward generalization.

5.4.1 Achievable Scheme for $M = N_1 + N_2$

The transmission scheme lasts for $N_1 + N_2 - m$ Time Slots (TS). During the first N_1 TS we transmit $M = N_1 + N_2$ symbols per TS. In particular, we transmit:

- m symbols to RX 2, which are canceled at RX 1 by using AP-ZF (since $m \geq N_1$).
- $M - m$ symbols to RX 1 canceled at m antennas of RX 2 using AP-ZF. Thus:
  - RX 2 can decode its own symbols since it has m antennas free of interference and m symbols to decode.
  - RX 2 can hence remove the contribution of its own symbols and obtain $N_2 - m$ independent linear combinations of the symbols intended by RX 1.
  - By sending those $N_2 - m$ linear combinations in a broadcast mode, we provide RX 1 with all the equations that it needs to decode its $M - m$ symbols, since it has already N_1 linear combinations.
  - Since RX 2 already knows those retransmitted linear combinations, they do not hurt its DoF because RX 2 can remove them from the received signal.

Note that RX 2 does not need to decode any message of RX 1. In the following $N_2 - m$ TS, at each TS we send N_2 symbols to RX 2 while retransmitting N_1 symbols from the interference received by RX 2 during the first phase.

- The interference retransmitted is removed perfectly at RX 2 because it is composed of the previously listened signals. Then, RX 2 can decode perfectly its own symbols.
- The symbols transmitted to RX 2 are canceled at RX 1 by AP-ZF.

Consequently, we obtain a sum DoF of

$$\frac{1}{N_1 + N_2 - m}\big(N_1(N_1 + N_2) + N_2(N_2 - m)\big) = N_2 + N_1\frac{N_1}{N_1 + N_2 - m}. \tag{5.14}$$

5.4.2 Achievable Scheme for $N_2 < M < N_1 + N_2$

We consider now the case for $N_2 < M < N_1 + N_2$, with $N_1 \leq m < N_2$. The transmission scheme follows the same structure as the $M = N_1 + N_2$ case, but with a different number of TS and transmitted symbols. Specifically, we first transmit M symbols per TS during the first N_1 TS. Thus, at each TS, we transmit:

- m symbols to RX 2, that are canceled at RX 1 using AP-ZF (since $m \geq N_1$).
- $M - m$ symbols to RX 1, that are canceled at m antennas of RX 2 using A-PZF.

- RX 2 can decode its own symbols since it has m antennas free of interference and m symbols to decode.
- RX 2 can then remove the contribution of its own symbols and obtain $N_2 - m$ independent linear combinations of the symbols intended by RX 1.
- If RX 1 obtains $M - N_1 - m$ independent linear combinations of those $N_2 - m$ interferences, RX 1 will be able to decode its $M - m$ symbols, since it has already N_1 linear combinations.
- Since RX 2 already knows those retransmitted linear combinations, they do not hurt its DoF because RX 2 can remove them from the received signal.

In the following $M - m - N_1$ TS, we send at each TS N_2 symbols to RX 2 while retransmitting N_1 symbols from the interference received by RX 2 during the first phase.

- The interference retransmitted is removed perfectly at RX 2 because it is composed of the previously listened signals. Then, RX 2 can decode perfectly its own symbols.
- The symbols transmitted to RX 2 are canceled at RX 1 by AP-ZF.

We can reconstruct and retransmit the interference created during the first phase at $m \geq N_1$ antennas. Therefore we can transmit N_1 independent linear combinations of them at each TS. Hence, we obtain a sum DoF of

$$\frac{1}{M-m}\big((m+M-m)N_1 + (N_2)(M-m-N_1)\big) = N_2 + N_1\frac{M-N_2}{M-m}, \tag{5.15}$$

what concludes the proof of Theorem 5.2. ■

5.5 Converse of Theorem 5.1

We prove Theorem 5.1 for real channels. The extension to complex variables is intuitive but cumbersome, and hence we omit it for sake of conciseness. In Theorem 5.1, the only particular bound is

$$\frac{d_1}{\min(M, N_1+N_2) - m} + \frac{d_2 - m}{N_2 - m} \leq 1, \tag{5.16}$$

since the other expressions are directly obtained by considering a genie-aided setting with perfect CSIT at every TX. This genie-aided scenario corresponds with the well-known centralized Network MIMO with perfect CSIT [96]. Since providing with additional CSI can not hurt, we obtain (5.5). Hence we consider only the case $m < \min(N_2, M - N_1)$. We split the proof in two sub-regimes: $M \leq N_1 + N_2$ and $M > N_1 + N_2$.

5.5.1 Converse for the case $M \leq N_1 + N_2$

We start by discretizing the channel, what leads to a deterministic channel model introduced in [124]. It has been shown in [120, Lemma 1] that the DoF of the deterministic channel model is an upper-bound of the DoF of the general channel model.

Deterministic Channel Model

The discretized model is such that the input signals $\bar{X}_j(t) \in \mathbb{Z}$ and output signals $\bar{Y}_i(t) \in \mathbb{Z}$ are given by

$$\bar{X}_j(t) \in \{0, 1, \ldots, \lceil \bar{P} \rceil\}, \quad \forall j \in \mathbb{N}_M, \tag{5.17}$$

$$\bar{Y}_i(t) \triangleq \sum_{j=1}^{M} \lfloor \mathbf{H}_{i,j} \bar{X}_j(t) \rfloor, \quad \forall i \in \{1, 2\}. \tag{5.18}$$

In the following, we obtain an upper-bound for this channel model. From [120], it is also an upper-bound for the general channel model that we have considered.

Weighted sum rate

We obtain (5.16) by means of bounding the weighted sum rate $n(N_2-m)R_1+n(M-m)R_2$. First of all, we present an instrumental lemma.

Lemma 5.2. *Let the number of well-informed transmit antennas (m) satisfy that $m < \min(M - N_1, N_2)$. Then,*

$$(N_2 - m)H(\bar{\mathbf{Y}}_1^{[n]}|\mathbf{H}^{[n]}, W_2) - (M - m)H(\bar{\mathbf{Y}}_2^{[n]}|\mathbf{H}^{[n]}, W_2) \leq o(\log \bar{P}). \tag{5.19}$$

Proof. The proof is relegated to Section 5.5.2. ■

We start from Fano's inequality to obtain

$$\begin{aligned} n(R_1 + R_2) &\leq I(W_1; \bar{\mathbf{Y}}_1^{[n]}|\mathbf{H}^{[n]}, W_2) + I(W_2; \bar{\mathbf{Y}}_2^{[n]} \mid \mathbf{H}^{[n]}) \\ &= H(\bar{\mathbf{Y}}_2^{[n]} \mid \mathbf{H}^{[n]}) - H(\bar{\mathbf{Y}}_2^{[n]}|\mathbf{H}^{[n]}, W_2) + H(\bar{\mathbf{Y}}_1^{[n]}|\mathbf{H}^{[n]}, W_2) \\ &\quad - \underbrace{H(\bar{\mathbf{Y}}_1^{[n]}|\mathbf{H}^{[n]}, W_2, W_1)}_{=0}. \end{aligned} \tag{5.20}$$

We recall the fact that the entropy of a random variable is bounded by its support, i.e.,

$$H(\bar{\mathbf{Y}}_2^{[n]}) \leq n\, N_2 \log(\bar{P}). \tag{5.21}$$

Lemma 5.2 and (5.21) yield

$$\begin{aligned} n(N_2-m)R_1+n(M-m)R_2 &\leq (M-m)\big(H(\bar{\mathbf{Y}}_2^{[n]}|\mathbf{H}^{[n]})-H(\bar{\mathbf{Y}}_2^{[n]}|\mathbf{H}^{[n]},W_2)\big) \\ &\quad +(N_2-m)H(\bar{\mathbf{Y}}_1^{[n]}|\mathbf{H}^{[n]},W_2) \\ &\leq n\,(M-m)N_2\log\bar{P}+o(\log\bar{P}). \end{aligned} \tag{5.22}$$

We can divide by $(M-m)(N_2-m)$ to write

$$\frac{nR_1}{M-m}+\frac{nR_2}{N_2-m}\leq\frac{nN_2}{N_2-m}\log\bar{P}+o(\log\bar{P}). \tag{5.23}$$

Since the DoF of RX i is obtained as

$$d_i=\lim_{P\to\infty}\lim_{n\to\infty}\frac{R_i}{\frac{n}{2}\log(\bar{P})}, \tag{5.24}$$

it follows that

$$\frac{d_1}{M-m}+\frac{d_2}{N_2-m}\leq\frac{N_2}{N_2-m} \tag{5.25}$$

$$\Rightarrow\quad \frac{d_1}{M-m}+\frac{d_2-m}{N_2-m}\leq 1 \tag{5.26}$$

what concludes the proof of (5.4) for $M\leq N_1+N_2$.

5.5.2 Proof of Lemma 5.2

We prove in the following Lemma 5.2, i.e, that for $m<\min(M-N_1,N_2)$,

$$(N_2-m)H(\bar{\mathbf{Y}}_1^{[n]}|\mathbf{H}^{[n]},W_2)-(M-m)H(\bar{\mathbf{Y}}_2^{[n]}|\mathbf{H}^{[n]},W_2)\leq o\big(\log\bar{P}\big). \tag{5.27}$$

We recall that $\bar{\mathbf{Y}}_i^{[n]}=[\bar{Y}_{i,1}^{[n]},\ldots,\bar{Y}_{i,N_i}^{[n]}]$, where

$$\bar{Y}_{i,j}^{[n]}=L_{i,j}(\bar{X}_1,\ldots,\bar{X}_M). \tag{5.28}$$

Note that the signals $\bar{X}_1,\ldots,\bar{X}_m$ are function of the the channel and the messages. We can apply a rotation matrix at RX 2 such that the m first transmit antennas –($\mathrm{TX}_\star$)– are only listened by m of the antennas of RX 2, that we choose w.l.o.g. to be the m first antennas. Recall that $X_\varnothing\triangleq\{X_{m+1},\ldots,X_M\}$. Hence, for any $j>m$ we can define

$$L_{\bar{Y},j}^{b[n]}(X_\varnothing)\triangleq\bar{Y}_{2,j}^{[n]}. \tag{5.29}$$

Moreover, $H(A,B) \geq H(A)$ implies that $H\big(\mathbf{Y}_2^{[n]}|\mathbf{H}^{[n]},W_2\big) \geq H\big(\bigcap_{j>m} \bar{Y}_{2,j}^{[n]}|\mathbf{H}^{[n]},W_2\big)$. From (5.29), we can write that

$$H\Big(\bigcap_{j>m} \bar{Y}_{2,j}^{[n]}|\mathbf{H}^{[n]},W_2\Big) = H\Big(\bigcap_{j>m} L_{\bar{Y},j}^{b[n]}(\bar{X}_\varnothing)|\mathbf{H}^{[n]},W_2\Big). \tag{5.30}$$

We continue by taking apart $(M-N_2)$ negative terms and applying (5.30) to write

$$\begin{aligned}
(N_2-m)H(\bar{\mathbf{Y}}_1^{[n]}|\mathbf{H}^{[n]},W_2) &- (M-m)H(\bar{\mathbf{Y}}_2^{[n]}|\mathbf{H}^{[n]},W_2) \\
&\leq (N_2-m)\Big(H(\bar{\mathbf{Y}}_1^{[n]}|\mathbf{H}^{[n]},W_2) - H\big(\bar{\mathbf{Y}}_2^{[n]}|\mathbf{H}^{[n]},W_2\big)\Big) \\
&\quad - (M-N_2)H\Big(\bigcap_{j>m} \bar{Y}_{2,j}^{[n]}|\mathbf{H}^{[n]},W_2\Big) \\
&\leq (N_2-m)\Big(H(\bar{\mathbf{Y}}_1^{[n]}|\mathbf{H}^{[n]},W_2) - H\big(\bar{\mathbf{Y}}_2^{[n]}|\mathbf{H}^{[n]},W_2\big)\Big) \\
&\quad - (M-N_2)H\Big(\bigcap_{j>m} L_{\bar{Y},j}^{b[n]}(\bar{X}_\varnothing)|\mathbf{H}^{[n]},W_2\Big).
\end{aligned} \tag{5.31}$$

Let us first describe the intuition behind the proof before deriving the result. We see in (5.31) that we have N_2-m negative entropy terms, each one of N_2 variables, and other $M-N_2$ negative entropy terms, each one of N_2-m variables. All the variables are linear combinations of the M transmit signals $(\bar{X}_i)$. Our goal is to show that all those negative terms can be reordered so as to create N_2-m terms of M linear combinations. If this statement is true, from the fact that $H(A)-H(B) \leq H(A|B)$ we can remove the contribution of the N_2-m positive terms $H(\bar{\mathbf{Y}}_1^{[n]}|\mathbf{H}^{[n]},W_2)$, since we can decode –with high probability– the M signals with M independent linear combinations. In the following we show rigorously that the previous idea is indeed applicable. Our proof is based on the following Lemma that was introduced by Davoodi and Jafar in [156] and that follows from the Aligned Image Set approach [120, 155].

Lemma 5.3 ([156, Lemma 2]). *Consider $\beta > 0$ and random variables $F_k^{[n]}, G_k^{[n]}$, $k \in \mathbb{N}_K$ that satisfy the bounded density assumption. Then, it holds that*

$$H\left(\sum_{k=1}^{K} \lceil \bar{P}^\beta F_k^{[n]} \bar{X}_k^{[n]} \rceil\right) \leq H\left(\sum_{k=1}^{K} \lceil \bar{P}^\beta G_k^{[n]} \bar{X}_k^{[n]} \rceil\right) + o(\log \bar{P}). \tag{5.32}$$

From Lemma 5.3 and the fact that $H(L(X_i)) \leq H(L^b(X_i))$ [155, 156], it follows that

$$H\big(\bigcap_{j>m} L_{\bar{Y},j}^{b[n]}(\bar{X}_\varnothing)|\mathbf{H}^{[n]},W_2\big) \geq H\big(\bigcap_{\substack{j>m \\ j\neq\ell}} L_{\bar{Y},j}^{b[n]}(\bar{X}_\varnothing),\ L^{[n]}(\bar{X}_\varnothing)|\mathbf{H}^{[n]},W_2\big) + o(\log \bar{P}), \tag{5.33}$$

where we have substituted $L^{b[n]}_{\bar{Y},\ell}(\bar{X}_\varnothing)$ by $L^{[n]}(\bar{X}_\varnothing)$. Hereinafter, we omit the $o(\log \bar{P})$ terms for ease of notation and because they are irrelevant for the DoF metric. We consider now the sum $H\big(\bar{\mathbf{Y}}_2^{[n]}|\mathbf{H}^{[n]}, W_2\big) + H\big(\bigcap_{j>m} \bar{Y}_{2,j}^{[n]}|\mathbf{H}^{[n]}, W_2\big)$. it follows that

$$
\begin{aligned}
&H\big(\bar{\mathbf{Y}}_2^{[n]}|\mathbf{H}^{[n]}, W_2\big) + H\big(\bigcap_{j>m} \bar{Y}_{2,j}^{[n]}|\mathbf{H}^{[n]}, W_2\big) \\
&= H\big(\bigcap_{k\leq m} \bar{Y}_{2,k}^{[n]}, \bigcap_{j>m} L^{b[n]}_{\bar{Y},j}(\bar{X}_\varnothing)|\mathbf{H}^{[n]}, W_2\big) + H\big(\bigcap_{j>m} L^{b[n]}_{\bar{Y},j}(\bar{X}_\varnothing)|\mathbf{H}^{[n]}, W_2\big) \\
&\overset{(a)}{\geq} H\big(\bigcap_{k\leq m} \bar{Y}_{2,k}^{[n]}, \bigcap_{j>m} L^{b[n]}_{\bar{Y},j}(\bar{X}_\varnothing)|\mathbf{H}^{[n]}, W_2\big) + H\big(\bigcap_{\substack{j>m\\ j\neq\ell}} L^{b[n]}_{\bar{Y},j}(\bar{X}_\varnothing), L^{[n]}(\bar{X}_\varnothing)|\mathbf{H}^{[n]}, W_2\big) \qquad (5.34)\\
&\overset{(b)}{\geq} H\big(\bigcap_{k\leq m} \bar{Y}_{2,k}^{[n]}, \bigcap_{j>m} L^{b[n]}_{\bar{Y},j}(\bar{X}_\varnothing), L^{[n]}(\bar{X}_\varnothing)|\mathbf{H}^{[n]}, W_2\big) + H\big(\bigcap_{\substack{j>m\\ j\neq\ell}} L^{b[n]}_{\bar{Y},j}(\bar{X}_\varnothing)|\mathbf{H}^{[n]}, W_2\big).
\end{aligned}
$$

where (a) follows from (5.33) and (b) comes from the sub-modularity property as $H(A,B)+H(B,C) \geq H(A,B,C)+H(B)$ [157, Theorem 1]. We can repeat (5.34) as many as $M-N_2$ times, such that we obtain

$$
H\big(\bigcap_{k\leq m} \bar{Y}_{2,k}^{[n]}, \bigcap_{j>m} L^{b[n]}_{\bar{Y},j}(\bar{X}_\varnothing),\ \mathbf{L}^{[n]}(\bar{X}_\varnothing)|\mathbf{H}^{[n]}, W_2\big), \qquad (5.35)
$$

where $\mathbf{L}^{[n]}(\bar{X}_\varnothing)$ is composed of $M-N_2$ independent linear combinations of $\bar{X}_\varnothing$ ($L^{[n]}(\bar{X}_\varnothing)$). Hence, in (5.35) we have gathered M independent linear combinations of the transmitted signals $\{\bar{X}_i\}_{i\in\mathbb{N}_M}$, such that it follows that

$$
\begin{aligned}
H\big(\bar{\mathbf{Y}}_1^{[n]}|\mathbf{H}^{[n]}\big) - H\big(\bigcap_{k\leq m} \bar{Y}_{2,k}^{[n]}, \bigcap_{j>m} L^{b[n]}_{\bar{Y},j}(\bar{X}_\varnothing),\ \mathbf{L}^{[n]}(\bar{X}_\varnothing)|\mathbf{H}^{[n]}, W_2\big)& \\
\leq H\big(\bar{\mathbf{Y}}_1^{[n]}\ |\ \bigcap_{k\leq m} \bar{Y}_{2,k}^{[n]}, \bigcap_{j>m} L^{b[n]}_{\bar{Y},j}(\bar{X}_\varnothing),\ \mathbf{L}^{[n]}(\bar{X}_\varnothing), \mathbf{H}^{[n]}, W_2\big)& \qquad (5.36)\\
\leq o(n).&
\end{aligned}
$$

Repeating the previously describe derivation up to N_2-m times, it holds that

$$
\begin{aligned}
&(N_2-m)H(\bar{\mathbf{Y}}_1^{[n]}|\mathbf{H}^{[n]}, W_2) - (M-m)H(\bar{\mathbf{Y}}_2^{[n]}|\mathbf{H}^{[n]}, W_2) \\
&\leq o(n),
\end{aligned} \qquad (5.37)
$$

what concludes the proof of Lemma 5.2. ■

5.5.3 Converse for the case $M > N_1 + N_2$

We define $N \triangleq N_1 + N_2$. We can split the M transmit antennas as

$$\mathrm{TX} \triangleq \big[\underbrace{\mathrm{TX}_1,\ \ldots,\ \mathrm{TX}_m}_{\mathrm{TX}_\star},\ \underbrace{\mathrm{TX}_{m+1},\ \ldots,\ \mathrm{TX}_M}_{\mathrm{TX}_\varnothing}\big] \tag{5.38}$$

The channel $\mathbf{H} \in \mathbb{C}^{N\times M}$ has $M - N$ null space dimensions. Therefore, if we could apply a rotation matrix $\mathbf{R}$ with unit determinant to make $\mathbf{HR}$'s right $M - N$ columns be zero, it would lead to an equivalent channel where the RXs do not listen to the last $(M - N)$ TXs. Defining $\mathbf{H}' \triangleq \mathbf{HR}$,

$$\mathbf{H}' = \begin{bmatrix} \mathbf{H}'_{[1:N,1:N]} & \mathbf{0}_{N\times(M-N)} \end{bmatrix}. \tag{5.39}$$

To obtain this equivalent channel, we apply an invertible linear transformation at the transmit antennas by multiplying the transmit signal $\mathbf{X}$ by $\mathbf{R}$. Hence, we transmit $\mathbf{X}' \triangleq \mathbf{RX}$ in place of $\mathbf{X}$. After this transformation, the equivalent transmitter TX' is

$$\mathrm{TX}' \triangleq \big[\underbrace{\mathrm{TX}_1,\ \ldots,\ \mathrm{TX}_m}_{\mathrm{TX}_\star},\ \underbrace{\mathrm{TX}_{m+1},\ \ldots,\ \mathrm{TX}_{N_1+N_2}}_{\mathrm{TX}'_\varnothing}\big] \tag{5.40}$$

and we can derive the upper-bound by applying the same steps as in Section 5.5.1 for $M = N_1 + N_2$, since the RXs only listen to $N_1 + N_2$ transmit antennas.

Channel Rotation with Distributed CSIT

Although the previous channel transformation is simply applied in a centralized scenario where all the transmit antennas are seen as one single entity, it is not straightforward that it can be applied in our distributed scenario, where every single transmit antenna is isolated with respect to the others and has to act *only* based on his own local information. Thereupon we show that the application of this channel transformation is indeed possible.

In the D-CSIT scenario considered, the matrix multiplication $\mathbf{RX}$ must be done *locally*. Consequently, the equivalent transmitted signal at TX_i, X_i, is obtained as

$$X'_i = \mathbf{R}_i\mathbf{X}, \tag{5.41}$$

where $\mathbf{R}_i$ is the i-th row of $\mathbf{R}$. However, the $M - m$ antennas with finite precision CSIT are not able to obtain neither $\mathbf{R}$ nor the transmit signal from the TXs with CSIT. In order to deal with this problem, we first let all the TXs in $\mathrm{TX}_\star$ cooperate among them. Similarly, we let all the non-informed TXs in $\mathrm{TX}_\varnothing$ cooperate among them. Since every TX in $\mathrm{TX}_\star$ already had perfect information of the whole channel, assuming that they are a

unique TX with m antennas does not affect the analysis. In the same way, assuming that the $M-m$ TXs with finite precision CSI form a unique TX with $M-m$ antennas does not give to them any improvement, since they still have only finite precision CSI. Furthermore, cooperation can not hurt. Therefore, we have an equivalent channel with two TXs, $\text{TX}_\star$ that transmits $\mathbf{X}_\star$, and $\text{TX}_\varnothing$ that transmits $\mathbf{X}_\varnothing$. The channel transformation is applied as

$$\mathbf{X}'_\star = \mathbf{R}_{[1:m,1:M]}\mathbf{X} \tag{5.42}$$

$$\mathbf{X}'_\varnothing = \mathbf{R}_{[m+1:M,1:M]}\mathbf{X}. \tag{5.43}$$

Composition of the Transformation Matrix

We aim to obtain a matrix $\mathbf{R} \in \mathbb{C}^{N\times M}$ such that $\mathbf{H}' \triangleq \mathbf{HR}$ satisfies

$$\mathbf{H}' = \begin{bmatrix} \mathbf{H}'_{[1:N,1:N]} & \mathbf{0}_{N\times(M-N)} \end{bmatrix}. \tag{5.44}$$

In order to obtain (5.44), we need $h'_{i,j} = 0$, for any $j \in \{N+1,\dots,M\}$, and for any i. In order to transform the j-th channel column, we solve the following linear system

$$\begin{bmatrix} h_{1,1} & h_{1,2} & \dots & h_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N,1} & h_{N,2} & \dots & h_{N,N} \end{bmatrix} \begin{bmatrix} r_{1,j} \\ \vdots \\ r_{N,j} \end{bmatrix} = \begin{bmatrix} -h_{1,j} \\ \vdots \\ -h_{N,j} \end{bmatrix}. \tag{5.45}$$

From the channel independence assumption, $\mathbf{H}_{[1:N,1:N]}$ is full rank almost surely, and therefore the system has a solution. Hence, the matrix $\mathbf{R}$ is defined as

$$\mathbf{R} \triangleq \begin{bmatrix} & r_{1,N+1} & \dots & r_{1,M} \\ \mathbf{I}_{N\times N} & \vdots & \ddots & \dots \\ & r_{N,N+1} & \dots & r_{N,M} \\ \mathbf{0}_{(M-N)\times N} & \mathbf{I}_{(M-N)\times(M-N)} & & \end{bmatrix}. \tag{5.46}$$

From (5.46), it holds that $\mathbf{H}'_{[1:N,1:N]} = \mathbf{H}_{[1:N,1:N]}$. Note that the antennas with finite precision CSIT can obtain their equivalent transmit signals as

$$\mathbf{X}'_\varnothing = \mathbf{R}_{[m+1:M,1:M]}\mathbf{X} \tag{5.47}$$

$$= \underbrace{\begin{bmatrix} \mathbf{R}_{m+1,m+1} & \dots & \mathbf{R}_{m+1,M} \\ \vdots & \ddots & \vdots \\ \mathbf{R}_{M,m+1} & \dots & \mathbf{R}_{M,M} \end{bmatrix}}_{\mathbf{R}_\varnothing} \mathbf{X}_\varnothing. \tag{5.48}$$

Therefore, the transformation at the TXs with finite precision depends only on their own transmit signals and they do not need to know $\mathbf{X}_\star$. Thus, we can assume that $\text{TX}_\varnothing$ is genie-aided and provided with the matrix $\mathbf{R}_\varnothing$. Note that (5.45) and the finite precision CSIT assumption imply that $\text{TX}_\varnothing$ can not infer any $h_{i,j}$ from the knowledge of $\mathbf{R}_\varnothing$.

5.6 On the Achievability for the Case $m < N_1$

In this section, we analyze the achievability results for the case in which $m < N_1$. First, we prove the achievable DoF presented in Proposition 5.1, which serves as lower-bound for any configuration. After that, we present a particular case that shows that the lower-bound can be improved.

5.6.1 Proof of Proposition 5.1

Let us consider the case with $M = N_1 + N_2$ for the sake of simplicity. We have seen previously that the cases $M \neq N_1 + N_2$ follow after a direct generalization. In this case, we transmit $m + N_2$ symbols per TS during the first m TS. We transmit:

- m symbols to RX 2, that are canceled at m antennas of RX 1 using AP-ZF.
- N_2 symbols to RX 1, that are canceled at m antennas of RX 2 using AP-ZF .
  - RX 2 can decode its own symbols since it has m antennas free of interference and m symbols to decode.
  - RX 2 can then remove the contribution of its own symbols and obtain $N_2 - m$ independent linear combinations of the symbols intended by RX 1.
  - If RX 1 obtains those $N_2 - m$ independent linear combinations of its own symbols, RX 1 can decode all the N_2 symbols, since it has already m linear combinations free of interference.
  - Since RX 2 already knows those retransmitted symbols, they do not hurt its DoF because RX 2 can remove them from the received signal.

In the following $N_2 - m$ TS, at each TS we send N_2 symbols to RX 2 while retransmitting m symbols from the interference received by RX 2 during the first m TS.

- The interference retransmitted can be removed perfectly at RX 2, then RX 2 can decode perfectly its own N_2 symbols.
- The symbols intended by RX 2 are canceled at m antennas of RX 1 thanks to AP-ZF.
- RX 1 has m antennas free of interference and thus it can decote its own m symbols.

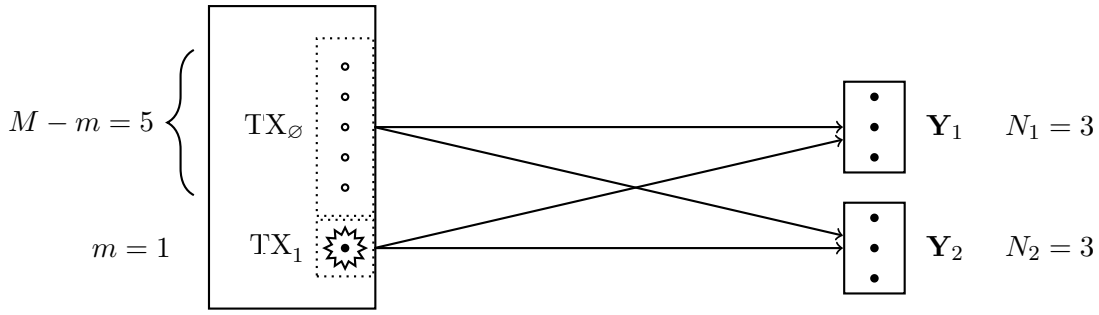


Figure 5.3 – Equivalent channel for the case $(M, N_1, N_2, m) = (6, 3, 3, 1)$.

Consequently, we obtain a DoF of

$$\frac{1}{m + N_2 - m}\big(m(m + N_2) + (N_2 - m)N_2\big) = N_2 + \frac{m^2}{N_2}. \tag{5.49}$$

5.6.2 Achievability for the Case $(M, N_1, N_2) = (6, 3, 3)$

In this section, we consider a particular setting so as to illustrate some achievability results in the regime with $m < N_1$ transmit antennas with perfect CSIT, for which no tight upper-bound is known. Let us consider a setting with $M = 6$ transmit antennas and two RXs, with $N_1 = N_2 = 3$ antennas at each RX. Suppose that only one transmit antenna has perfect CSI for the whole channel matrix, while the other 5 transmit antennas have only finite precision CSI. Thus, $m = 1$. The setting, denoted as $(M, N_1, N_2, m) = (6, 3, 3, 1)$, is illustrated in Fig. 5.3. We present here a scheme that achieves a sum DoF of 4.
The scheme consists in two phases, each one of 2 Time Slots (TS), and it is presented in Table 5.1, in which every row represents one antenna. The table is divided in three horizontal parts: The top part represents the symbols transmitted from each antenna. The middle part represents the received signal at RX 1 and the bottom part shows the received signal at TX 2.

Let us disclose the previous table by describing the transmission scheme. We send 18 symbols (a_{1-8}, b_{1-8}, c, d). Symbols a_i are intended to RX 1 and symbols b_i are intended to RX 2. The functions f, f', f'', f''', are such that the corresponding symbols a_i or b_i are canceled at the third antenna of the non-intended RX. The function f_j^i, g_j^i, are defined such that they represent the received signal at RX i. The sub-index j is used to order and identify the different received signals. f_j^i denotes the received signal during the first two TS whereas g_j^i denotes the received signal for the last two TS.

First, the messages a_{6-8} and b_{6-8} are easily obtained at the intended RX from the received signal of $t = 3$ and $t = 4$ after decoding d. Furthermore, if RX 1 obtained equations f_1^2 and f_2^2, it would be able to decode all the a_{1-5}. Similarly, if RX 2 obtained

Table 5.1 – Description of a transmission scheme achieving the optimal DoF = 4 for the setting $(M, N_1, N_2, m) = (6, 3, 3, 1)$.

	$t=1$	$t=2$	$t=3$	$t=4$
TX_1	$c + f(a_1, a_2, a_3, a_4, a_5)$	$c + f'(b_1, b_2, b_3, b_4, b_5)$	$d + f''(a_6, a_7, a_8)$	$d + f'''(b_6, b_7, b_8)$
$\text{TX}_{\varnothing,1}$	a_1, a_2, a_3, a_4, a_5	b_1, b_2, b_3, b_4, b_5	a_6, a_7, a_8	b_6, b_7, b_8
$\text{TX}_{\varnothing,2}$	a_1, a_2, a_3, a_4, a_5	b_1, b_2, b_3, b_4, b_5	a_6, a_7, a_8	b_6, b_7, b_8
$\text{TX}_{\varnothing,3}$	a_1, a_2, a_3, a_4, a_5	b_1, b_2, b_3, b_4, b_5	a_6, a_7, a_8	b_6, b_7, b_8
$\text{TX}_{\varnothing,4}$	a_1, a_2, a_3, a_4, a_5	b_1, b_2, b_3, b_4, b_5	a_6, a_7, a_8	b_6, b_7, b_8
$\text{TX}_{\varnothing,5}$	a_1, a_2, a_3, a_4, a_5	b_1, b_2, b_3, b_4, b_5	a_6, a_7, a_8	b_6, b_7, b_8
$Y_{1,1}$	$f_1^1(a_1, a_2, a_3, a_4, a_5, c)$	$f_4^1(b_1, b_2, b_3, b_4, b_5, c)$	$g_1^1(a_6, a_7, a_8, d)$	$g_4^1(b_6, b_7, b_8, d)$
$Y_{1,2}$	$f_2^1(a_1, a_2, a_3, a_4, a_5, c)$	$f_5^1(b_1, b_2, b_3, b_4, b_5, c)$	$g_2^1(a_6, a_7, a_8, d)$	$g_5^1(b_6, b_7, b_8, d)$
$Y_{1,3}$	$f_3^1(a_1, a_2, a_3, a_4, a_5, c)$	c	$g_3^1(a_6, a_7, a_8, d)$	d
$Y_{2,1}$	$f_1^2(a_1, a_2, a_3, a_4, a_5, c)$	$f_3^2(b_1, b_2, b_3, b_4, b_5, c)$	$g_1^2(a_6, a_7, a_8, d)$	$g_3^2(b_6, b_7, b_8, d)$
$Y_{2,2}$	$f_2^2(a_1, a_2, a_3, a_4, a_5, c)$	$f_4^2(b_1, b_2, b_3, b_4, b_5, c)$	$g_2^2(a_6, a_7, a_8, d)$	$g_4^2(b_6, b_7, b_8, d)$
$Y_{2,3}$	c	$f_5^2(b_1, b_2, b_3, b_4, b_5, c)$	d	$g_5^2(b_6, b_7, b_8, d)$

equations f_4^1 and f_5^1, it would be able to decode all the b_{1-5}. Hence, we select c and d as

$$c = f_1^2 \oplus f_4^1, \tag{5.50}$$

$$d = f_2^2 \oplus f_5^1. \tag{5.51}$$

Therefore, RX 1 can subtract f_4^1 from c and f_5^1 from d and obtain the necessary equations. On the other hand, RX 2 can subtract f_1^2 from c and f_2^2 from d and obtain also the necessary equations. Since we have causal CSIT, we can not encode the f_4^1 of $t = 2$ in c, but we can accept a one-block delay and transmit the received signal of the previous transmission block. The DoF loss will be negligible if the time considered is long enough. For $t = 3$ and $t = 4$, RX 1 obtains d in $t = 4$ and, after that, it can decode a_6, a_7, a_8 from $t = 3$. In the same way, RX 2 obtains d in $t = 3$ and thus it can decode b_6, b_7, b_8 at $t = 4$. Consequently, we transmit 16 information symbols in 4 TS, and thus DoF = 4. The general achievable scheme presented in Section 5.6.1 only attains a DoF of $\frac{10}{3}$, whereas the upper-bound of Lemma 5.1 yields DoF $\leq 4 + \frac{4}{5}$. Interestingly, the DoF of the the $(M, N_1, N_2) = (6, 3, 3)$ setting is equal to

$\text{DoF} = 3$	if $m = 0$,	$4 + \frac{1}{3} \leq \text{DoF} \leq 5 + \frac{1}{4}$	if $m = 2$,
$4 \leq \text{DoF} \leq 4 + \frac{4}{5}$	if $m = 1$,	$\text{DoF} = 6$	if $m \geq 3$.

From the previous results for other settings and the intuition that one extra informed antenna brings out one DoF, we could conjecture that DoF $= 4$ if $m = 1$ and DoF $= 5$ if $m = 2$. However, this characterization remains an open and interesting problem.

5.7 Conclusion

In this chapter, we have analyzed the Network MIMO in which M decentralized transmitted antennas jointly serve two multi-antenna users. We have considered the setting in which m transmit antennas are endowed with perfect CSI while the other $M - m$ antennas only have access to finite precision CSI. We have studied the DoF performance of this setting by deriving upper and lower bounds. Interestingly, we have proven that it is not necessary to have perfect CSI at every transmit antenna, but only at the number of antennas of the bigger user. We have derived a tight distributed CSIT upper-bound for the case of $N_1 \leq m$, characterizing the loss of DoF obtained from reducing the number of informed antennas. Nevertheless, there exist many open problems regarding the MIMO setting with distributed CSIT. Indeed, the upper-bound presented here is one of the first for the setting considered. We have also shown achievable transmission schemes that achieve the upper-bound for a certain regime. However, the gap between lower and upper bound in the regime with $m < N_1$ is not closed.

Part III

Performance Analysis of Distributed Zero-Forcing

Chapter 6

Rate Gap of the Distributed CSIT Setting with Random Vector Quantization

In the previous part, we have analyzed the D-CSIT setting from the perspective of the DoF and GDoF metrics. The objective was to uncover meaningful insights concerning the fundamental limits of the D-CSIT setting and to derive optimal schemes. In the following, we analyze the setting from a different angle. In particular, there are two main differences with respect to the previous analysis: First, we focus on simple ZF-type schemes; second, we go beyond the DoF interpretation and study the loss in achievable rate on account of the distributed structure of the CSI. The emphasis on ZF-type schemes is due to several considerations:

- ZF schemes represent an important and practical group of schemes that is known to provide a good compromise between performance and complexity, specially at the high-SNR regime.
- They allow for analytical tractability.
- As shown in the previous part, ZF precoding is one of the essential components of the DoF-optimal transmission schemes for the D-CSIT setting. Hence, it is interesting to analyze which is the aftermath of confronting those schemes to finer metrics and more practical analysis.

Due to the aforementioned points, we consider hereinafter a simple linear transmission scheme without superposition coding or successive interference cancellation. The TXs transmit a linear combination of the data symbols for all the RXs and aim to cancel the interference out. Further details are provided in this chapter.

As for the rate loss analysis, we seek to overcome the inherent limitations of the DoF metric –which were detailed in Section 2.3–. For this reason, we analyze hereinafter the rate difference between the C-CSIT scenario and the D-CSIT setting. This part is divided in two chapters, which differ in the size of the network considered and the CSIT acquisition model:

- The current chapter considers the 2×2 single-antenna setting. It also considers that the CSIT is obtained by means of quantized feedback sent from the RXs. The quantization at the RXs is carried out by applying Grassmanian Random Vector Quantization, that will be later explained.

- The next chapter considers the $M \times K$ setting with multi-antenna TXs. The CSIT is modeled such that the estimation noise is distributed as a Gaussian random variable.

This double distinction is made so as to convey in a compact manner the different implications that each one of the cases entails. Indeed, assuming two different CSIT models enables to provide two different approaches to obtain similar high-SNR regime results, as well as to show that the results hold for a general set of models. The division between the simple 2×2 case and the general case is interesting because the interpretation and analysis for the 2×2 setting are specific.

6.1 Preliminaries

6.1.1 Affine Approximation of Rate at High-SNR

We recall that the affine approximation of the rate, introduced in Section 2.3, allows us to write the rate as

$$R = \mathrm{DoF} \log_2(P) - \mathcal{R}_\infty + o(1). \tag{6.1}$$

We can observe that in the previous part we had focused on the characterization of the DoF term. Hereinafter, we wish to characterize completely the expression in (6.1). The term $\mathcal{R}_\infty$ represents the *rate offset*, i.e., the rate gap with respect to capacity when $P \to \infty$. We recall in Fig. 6.1 the qualitative meaning of the approximation in (6.1).

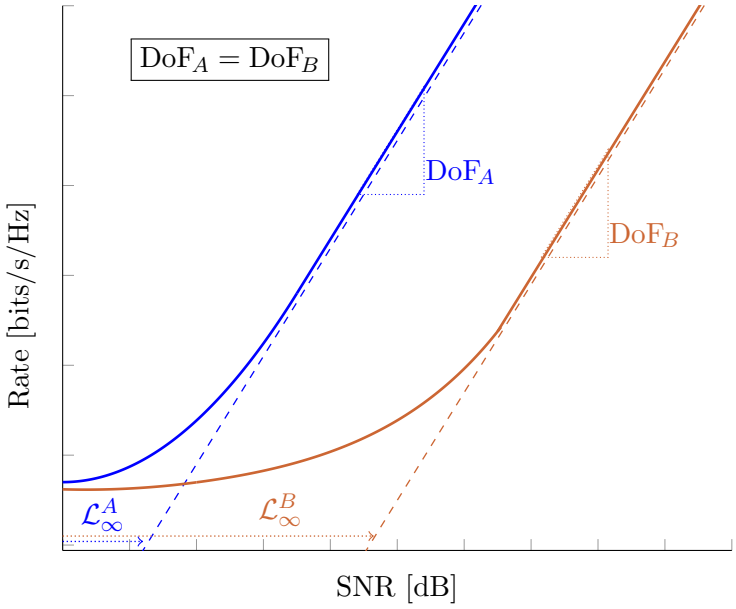


Figure 6.1 – Qualitative illustration of the affine approximation of two different setting with the same DoF (slope) but different rate offset $\mathcal{R}_\infty$.

Note that the rate offset is defined as

$$\mathcal{R}_\infty \triangleq \lim_{P\to\infty} \mathrm{DoF}\log_2(P) - R(P). \tag{6.2}$$

As stated in Section 2.3.4, several works in the literature have focused on characterizing (6.1) for some settings. Namely, the multiple-antenna point-to-point scenario [29], the BC with perfect CSIT using Dirty-paper coding and linear precoding [131], and the BC with imperfect CSIT [132]. Those works are yet focused on the centralized scenario. For the best of our knowledge, the analysis presented in the next two chapters is the first attempt to characterize the affine approximation the a distributed setting.

6.1.2 Transmission Model

As in Chapter 3, we consider a single-antenna setting with 2 TXs jointly serving 2 RXs. The signal received at RX i follows the notation of Section 2.2 such that

$$y_i = \mathbf{h}_i^{\mathrm{H}}\mathbf{x} + z_i, \tag{6.3}$$

where $\mathbf{h}_i^{\mathrm{H}} \in \mathbb{C}^{1\times 2}$ is the channel vector for RX i, $\mathbf{x} \in \mathbb{C}^{2\times 1}$ is the transmitted multi-user signal, and $z_i \in \mathbb{C}$ is the Additive White Gaussian Noise (AWGN) at RX i, being

independent of the channel and the transmitted signal, and drawn as $\mathcal{N}_{\mathbb{C}}(0,1)$. We further define the channel matrix $\mathbf{H} \in \mathbb{C}^{2\times 2}$ as $\mathbf{H} \triangleq [\mathbf{h}_1,\ \mathbf{h}_2]^{\mathrm{H}}$, with its (i,k)-th element representing the channel coefficient from TX k to RX i and being denoted as h_{ik}. The channel coefficients are assumed to be i.i.d. as $\mathcal{N}_{\mathbb{C}}(0,1)$ such that all the channel sub-matrices are full rank with probability one.

The transmitted multi-user signal $\mathbf{x} \in \mathbb{C}^{2\times 1}$ is obtained from the precoding of the symbol vector $\boldsymbol{s} \triangleq [s_1,\ s_2]^{\mathrm{T}}$. The symbols s_i are i.i.d. as $\mathcal{N}_{\mathbb{C}}(0,1)$ and s_i denotes the symbol intended by RX i. Furthermore, the transmit signal can be written as

$$\mathbf{x} \triangleq \frac{\bar{P}}{\sqrt{2}} \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}. \tag{6.4}$$

The vector $\mathbf{t}_i \in \mathbb{C}^{2\times 1}$ denotes the normalized precoding vector towards RX i, and thus $\mathbf{T} \triangleq [\mathbf{t}_1,\ \mathbf{t}_2] \in \mathbb{C}^{2\times 2}$. As introduced in Section 2.2, the precoder of TX j is denoted as $\mathbf{t}_{\mathrm{TX}\,j} \triangleq [\{\mathbf{t}_1\}_j,\ \{\mathbf{t}_2\}_j]$. Importantly, we assume a *per-TX instantaneous power constraint for the precoder*, i.e.,

$$\|\mathbf{t}_{\mathrm{TX}\,j}\| \leq 1, \quad \forall j \in \mathbb{N}_2, \tag{6.5}$$

what also implies that $\mathbb{E}\big[\|\mathbf{x}\|^2\big] \leq P$.

Remark 6.1. The assumption of a per-TX instantaneous power constraint for the precoder is an important change with respect to Part II, since we had previously considered an average power constraint. The reason why this modification is meaningful lies on the distributed nature of our setting. Indeed, with an average power constraint, the power normalization is based only on statistical information and thus it can be applied at all the TXs in a coordinated manner. This fact allowed us to develop robust schemes, as the interference cancellation can be carried out only by a subset of the TXs. However, with an instantaneous power constraint, *every TX* has to compute the power normalization values based on its own instantaneous CSI. Hence, the previous solutions are not sufficient under the new assumption, and we need to develop more elaborated strategies to prevent the performance from sinking. □

6.1.3 Grassmanian Random Vector Quantization

We are interested in analyzing the performance at the high-SNR regime. This regime has been studied extensively for the centralized setting, for which the rate offset between Dirty-Paper Coding (DPC), ZF with perfect CSIT, and ZF with imperfect CSIT has been obtained [131, 132]. Indeed, as aforementioned, there is a logical path from theoretical models to more practical ones in which 1) first the perfect-CSIT setting is analyzed.

Then, 2) the impact of having imperfect –yet centralized– CSIT is studied, and 3) the subsequent next setting to consider is the D-CSIT setting.

The first step, i.e., obtaining the rate offset on the BC with perfect CSIT using linear precoding with respect to the capacity-achieving DPC, was studied in [131]. Jindal tackled the the second step in [132]. In particular, the author investigated in [132] the performance of ZF for the case in which the CSIT is imperfect, assuming that the CSIT is obtained from a quantized version of the channel vector sent from the RXs, and the quantization is done through Grassmanian Random Vector Quantization (RVQ).

In this chapter we address the 3) step, i.e., we analyze the rate gap between the BC with imperfect centralized CSIT and with distributed CSIT. Therefore, we follow the same approach of the reference work [132] and study the performance when RVQ is applied. For the sake of completeness, we will recall in the following some properties that will be needed for the proof of our main results. For more details about RVQ, see [43,132].

Let M denote the number of transmit antennas. In RVQ, a unit-norm channel vector $\tilde{\mathbf{h}} \in \mathbb{C}^M$ is quantized using B bits to a codebook $\mathcal{C}$ containing 2^B unit-norm vectors isotropically distributed on the M-dimensional unit sphere. We consider a Grassmanian quantization scheme such that the quantized estimate –which is denoted by $\hat{\mathbf{h}} \in \mathbb{C}^M$– is obtained to minimize the angle with the true channel, i.e.,

$$\begin{aligned} \hat{\mathbf{h}} &= \underset{\boldsymbol{w}\in\mathcal{C}}{\operatorname{argmax}} \, |\tilde{\mathbf{h}}^{\mathrm{H}}\boldsymbol{w}|^2 \\ &= \underset{\boldsymbol{w}\in\mathcal{C}}{\operatorname{argmin}} \sin^2(\measuredangle(\tilde{\mathbf{h}}, \boldsymbol{w})), \end{aligned} \tag{6.6}$$

where we have introduced the angle for unit-norm vectors in $\mathbb{C}^M$ from $\measuredangle(\boldsymbol{x}, \boldsymbol{y}) \triangleq \arccos |\boldsymbol{x}^{\mathrm{H}}\boldsymbol{y}|$. We define the quantization error as

$$Z \triangleq \sin^2(\tilde{\mathbf{h}}, \hat{\mathbf{h}}). \tag{6.7}$$

Since the elements of the codebook $\mathcal{C}$ are independent of $\tilde{\mathbf{h}}$ and isotropically distributed, the quantization error Z is obtained as the minimum of 2^B independent beta $(M-1, 1)$ random variables. Upon defining $z = \sqrt{Z}$, and $\breve{z} \triangleq \sqrt{1-Z}$, we can write the true channel as a function of its quantized version as

$$\tilde{\mathbf{h}} = \breve{z}\hat{\mathbf{h}} + z\boldsymbol{\delta}, \tag{6.8}$$

where $\boldsymbol{\delta}$ is a unit-norm vector isotropically distributed in the null space of $\hat{\mathbf{h}}$, and $\boldsymbol{\delta}$ and Z are mutually independent. In this chapter, we consider that the vectors have $M = 2$ elements, and thus the quantization error Z is distributed as the minimum of 2^B standard uniform random variables [132]. We can see that the estimation model in (6.8) is included

in the general estimation model introduced in Section 2.4 and stated in (2.13).

6.1.4 Distributed CSIT Model

In this chapter, we consider that the TXs acquire its CSIT from the RXs. Specifically, we assume that the RXs have perfect and instantaneous knowledge of its own channel vector. Each RX quantizes its normalized channel with a certain number of bits using RVQ. In the D-CSIT setting, in contrast with the centralized setting of [132], each TX receives a quantized version with different number of bits.

In particular, we consider that RX i feeds back to TX j a quantized version of the normalized vector $\tilde{\mathbf{h}}_i \triangleq \frac{\mathbf{h}_i}{\|\mathbf{h}_i\|} \in \mathbb{C}^2$ using $B_i^{(j)}$ bits, denoted as $\hat{\mathbf{h}}_i^{(j)}$. We assume that RX i uses random vector quantization codebooks of $2^{B_i^{(j)}}$ codewords [132], such that the codewords are unit-norm vectors uniformly distributed on the 2-dimensional complex unit sphere. After receiving the feedback from both RXs, TX j obtains a multi-user channel estimate $\hat{\mathbf{H}}^{(j)} = [\hat{\mathbf{h}}_1^{(j)}, \hat{\mathbf{h}}_2^{(j)}]^{\mathrm{H}} \in \mathbb{C}^{2\times 2}$. In order to avoid degenerated conditions, we assume that the codebooks of different RXs do not share any codeword.

This D-CSIT model, for which the CSIT at TX j is composed of two vectors generated from codebooks with different size ($2^{B_1^{(j)}}$ and $2^{B_2^{(j)}}$), could also model a FDD transmission in which, for example, one of the following feedback mechanisms is applied:

1. The RXs send a rate-adapted feedback to each TX through different feedback messages.

2. The RXs send a single (broadcast) feedback message using layered encoding [153], such that each TX decode the message up to a different number of bits.

3. The RXs send a single feedback message to one of the TXs –this is the current standard mechanism– and each TX sends a compressed version of its CSIT to the other TX.

Moreover, it is known that, in order to avoid the collapse of DoF in the C-CSIT setting, the number of feedback bits must scale linearly with $\log_2(P)$ [120, 132]. By extension, we suppose the same scaling and let the number of bits grow linearly with $\log_2(P)$ as

$$B_i^{(j)} = \alpha_i^{(j)} \log_2(P). \tag{6.9}$$

This means that the estimation noise variance scales with $2^{-B_i^{(j)}} = P^{-\alpha_i^{(j)}}$. From [132, Lemma 1], it follows that the CSI error variance at TX j scales as $P^{-\alpha_i^{(j)}}$. Under such feedback condition, the multiplexing gain (DoF) of our setting is equal to

$$\mathrm{DoF} = 1 + \min\left(\max_j\left(\alpha_1^{(j)}\right),\ \max_j\left(\alpha_2^{(j)}\right),\ 1\right)^+, \tag{6.10}$$

as it can be inferred from Chapter 3 and [82]. The DoF collapses if the number of bits does not scale linearly with $\log_2(P)$ [120, 132]. Therefore, we assume that all $\alpha_i^{(j)}$ are strictly positive. From (6.10), $\alpha_i^{(j)}$ can be restricted to be $0 < \alpha_i^{(j)} \leq 1$. The multi-user D-CSIT configuration is represented through the multi-TX CSIT scaling matrix $\boldsymbol{\alpha}$ defined as

$$\boldsymbol{\alpha} \triangleq \begin{bmatrix} \alpha_1^{(1)} & \alpha_2^{(1)} \\ \alpha_1^{(2)} & \alpha_2^{(2)} \end{bmatrix} \in \mathbb{R}^{2\times 2}. \tag{6.11}$$

6.1.5 Genie-Aided Centralized Setting

In this chapter we consider the second genie-aided scenario introduced in Section 2.5, whose rigorous definition is presented below.

> **Definition 6.1** (Genie-aided Centralized Setting)**.** Let us assume a distributed setting in which each node has a different estimate with different average accuracy. The *Genie-aided Centralized Setting* is defined as the setting in which all the TXs are endowed with the estimate of best average accuracy.

We compare the rate achieved in the distributed scenario described in Section 6.1.4 with the respective genie-aided counterpart. This provides us with a benchmark for the performance of ZF schemes on the D-CSIT setting.

Remark 6.2. It is important to observe that the genie-aided scenario is such that every TX owns the best among the available estimates at any TX, *instead of* its own estimate –which by definition would have less accuracy–. This is in opposition to the other genie-aided scenario considered in Part II, in which each TX shares its CSI with any other TX, such that every TX owns the set of K estimates of the K TXs. We must assume this new genie-aided setting because the previous one, although instrumental for the DoF analysis, incurs in an excessive aid from the fact that each TX obtains K different estimates. □

6.2 Centralized Zero-Forcing Precoding

We restrict this chapter to ZF precoding schemes, which are one of the essential components to achieve the optimal DoF in the C-CSIT setting [120, 132] and that allow for analytical tractability. Since in the C-CSIT setting all the TXs share the same CSIT, the super-index $(\cdot)^{(j)}$ is not needed. Let us denote the shared channel estimates as $\hat{\mathbf{h}}_i$, $\hat{\mathbf{H}}$, where $\hat{\mathbf{h}}_i$ is obtained with a feedback rate of $B_i = \alpha_i \log_2(P)$ bits. Note that the centralized case is equivalent to the distributed setting in which $\hat{\mathbf{h}}_i^{(1)} = \hat{\mathbf{h}}_i^{(2)}$ for all $i \in \mathbb{N}_2$.

Let $\mathbf{v}_i^\star$ denote a unit-norm ZF precoder for RX i, computed on the basis of the estimate $\hat{\mathbf{H}}$. We can write the centralized ZF precoding matrix as $\mathbf{T}^{\mathrm{ZF}} \triangleq [\mu_1 \mathbf{v}_1^\star,\ \mu_2 \mathbf{v}_2^\star]$,

where $\mu_i \in \mathbb{R}$ is a parameter that ensures that the instantaneous power constraint $\|\mathbf{t}_{\mathrm{TX}\,j}\| \leq 1$ is fulfilled, and which will be detailed later in Section 6.3.2. Consider $\bar{i}$ as the complementary index of i, such that $\bar{i} = [i \mod 2] + 1$. From the ZF precoding definition, $\mathbf{v}_i^\star$ is a vector satisfying

$$\hat{\mathbf{h}}_{\bar{i}}^{\mathrm{H}} \mathbf{v}_i^\star = 0. \tag{6.12}$$

We assume hereinafter that $\|\mathbf{v}_i^\star\| = 1$, as the norm of the precoder can be incorporated to μ_i. In this case, since the estimation vector has a unitary norm from the RVQ properties, we can write w.l.o.g. that $\mathbf{v}_i = e^{-\imath\phi_i}[\hat{\mathrm{h}}_{\bar{i}2},\ -\hat{\mathrm{h}}_{\bar{i}1}]^{\mathrm{T}}$, where $e^{-\imath\phi_i}$ is an arbitrary phase term. The precoding matrix can be expressed as

$$\mathbf{T}^{\mathrm{ZF}} = \underbrace{\begin{bmatrix} \hat{\mathrm{h}}_{2,2} & \hat{\mathrm{h}}_{1,2} \\ -\hat{\mathrm{h}}_{2,1} & -\hat{\mathrm{h}}_{1,1} \end{bmatrix}}_{\triangleq \mathbf{V}} \underbrace{\begin{bmatrix} \mu_1 e^{-\imath\phi_1} & 0 \\ 0 & \mu_2 e^{-\imath\phi_2} \end{bmatrix}}_{\triangleq \mathbf{M}}. \tag{6.13}$$

In essence, $\mathbf{M}$ encloses the power normalization and a possible phase shifting whereas $\mathbf{V}$ encloses the correct interference cancellation introduced in (6.12). Note that the rate is invariant to any phase-shift $e^{\imath\phi_i}$ [145].

6.3 Distributed ZF: Hybrid Active-Passive ZF Precoding

Although ZF precoding schemes as the one described in Section 6.2 perform properly with centralized CSIT, their performance shrinks considerably on the D-CSIT setting. This comes from the fact that the zero-forcing accuracy is proportional to the worst quality among the TXs. Thus, conventional ZF does not achieve the centralized DoF, as it has been shown in Chapter 3. Indeed, if the precoder defined for the centralized setting in (6.13) were applied naively, we would obtain

$$\mathbf{T} = \begin{bmatrix} \hat{\mathrm{h}}_{2,2}^{(1)} & \hat{\mathrm{h}}_{1,2}^{(2)} \\ -\hat{\mathrm{h}}_{2,1}^{(1)} & -\hat{\mathrm{h}}_{1,1}^{(2)} \end{bmatrix} \odot \begin{bmatrix} \mu_1^{(1)} e^{-\imath\phi_1} & \mu_2^{(1)} e^{-\imath\phi_2} \\ \mu_1^{(2)} e^{-\imath\phi_1} & \mu_2^{(2)} e^{-\imath\phi_2} \end{bmatrix}, \tag{6.14}$$

where $\odot$ denotes the Hadamard (element-wise) product. Note that, since the precoder is computed locally, the estimated value for μ_i may be different at each TX, and thence the change from matrix product to Hadamard product.

The solution proposed in DoF-achieving schemes [82,98,100], as the ones in Chapter 3 and Chapter 4, was that the TX with worse accuracy for a certain channel coefficient does not use its estimate for computing its precoding vector. This strategy succumbs to

the assumption of instantaneous power constraint for the precoding vector ($\|\mathbf{t}_{\mathrm{TX}\,j}\| \leq 1$), since a less restricting average power constraint was considered. The only known scheme achieving the optimal DoF is obtained from [82] where the transmit power scales in $P/\log(P)$. This leads to a very inefficient power normalization, and hence to a very poor rate offset ($\mathcal{R}_\infty$).

We present a distributed precoding scheme, coined *Hybrid Active-Passive ZF Precoding* (HAP-ZF), that precludes the worst TX from harming the performance. The key for attaining such result is an asymmetric ZF scheme and the quantization of the power control, that allows the TXs to be *consistent*. We divide the HAP-ZF definition in several logical steps that help to better comprehend the benefits of the precoder.

6.3.1 Adapting Phase to CSIT Topology

The first step in the building process of the precoding scheme is to adapt the phase of the precoder to the CSIT configuration. Specifically, the idea is that, since multiplying the precoding vector by a phase-shift $e^{\imath\phi_i}$ does not impact the rate [145], we can adapt this phase shifting to the CSIT configuration such that the TX with least accurate estimate for a certain coefficient uses only the absolute value of that coefficient.

Example 6.1. Let TX 2 be the TX with worst accuracy for the channel coefficient between TX 1 and RX 1 ($\mathrm{h}_{1,1}$). From the centralized ZF precoder definition of (6.13), the precoder of TX 2 for the data symbols of RX 2 is

$$\mathbf{t}_{2,2} = -\mu_2 \hat{\mathrm{h}}_{1,1}^{(2)} e^{\imath\phi_i}. \tag{6.15}$$

Then, $e^{\imath\phi_i}$ is selected such that $\mu_2 \hat{\mathrm{h}}_{1,1}^{(2)} e^{\imath\phi_i}$ is a real number, i.e., $\phi_i = -\angle \hat{\mathrm{h}}_{1,1}^{(2)}$, where $\angle x$ represents the phase of $x \in \mathbb{C}$. In that case, the precoder applied at TX 2 is

$$\mathbf{t}_{2,2} = -\mu_2 |\hat{\mathrm{h}}_{1,1}^{(2)}|. \tag{6.16}$$

By doing so, the performance of the precoder would be intuitively improved, as the limiting node eliminates its sensitivity with respect to the error on the phase of the coefficient, and it is only affected by the error on the absolute value of the coefficient.

Therefore, the expression of the precoder changes accordingly to the D-CSIT configuration. In order to ease the mathematical derivations for the remaining of the chapter, we make the precoder expression chime with the one in (6.13) for the centralized case, such that it is decomposed in two matrices: A first matrix, denoted as $\mathbf{W} = [\mathbf{w}_1,\ \mathbf{w}_2]$, that encloses the interference nulling task of (6.12); the second matrix, denoted as $\mathbf{\Lambda}$, carries out the power normalization and is only composed by real values. Then we can

write that

$$\mathbf{T} = \underbrace{[\mathbf{w}_1 \quad \mathbf{w}_2]}_{\mathbf{W}} \odot \underbrace{\begin{bmatrix} \lambda_1^{(1)} & \lambda_2^{(1)} \\ \lambda_1^{(2)} & \lambda_2^{(2)} \end{bmatrix}}_{\mathbf{\Lambda}}. \tag{6.17}$$

We have denoted the (j,i)-th coefficient of the matrix $\mathbf{\Lambda}$ by $\lambda_i^{(j)}$. Then, the precoding vector has a different structure depending on which TX has the most accurate estimate for each channel coefficient. All the possible configurations for the precoder of RX 1 are presented in Table 6.1, whereas the ones for RX 2's precoder are omitted as they are obtained by swapping the user indexes in Table 6.1.

We further define the corresponding precoder obtained in the genie-aided C-CSIT setting where the most accurate estimate is shared as $\mathbf{T}^\star = \mathbf{W}^\star\mathbf{\Lambda}^\star$. Note that, in the centralized precoder, $\lambda_i^{(1)} = \lambda_i^{(2)}$ and thus the element-wise product can be substituted by a matrix product such that

$$\mathbf{T}^\star = \mathbf{W}^\star \underbrace{\begin{bmatrix} \lambda_1^\star & 0 \\ 0 & \lambda_2^\star \end{bmatrix}}_{\mathbf{\Lambda}^\star}. \tag{6.18}$$

Example 6.2. Let TX 2 be the TX with worst accuracy for the whole channel matrix. In that case, the precoder matrix can be expressed as

$$\begin{aligned} \mathbf{T} &= \begin{bmatrix} \hat{\mathrm{h}}_{2,2}^{(1)} & \hat{\mathrm{h}}_{1,2}^{(1)} \\ -\hat{\mathrm{h}}_{2,1}^{(2)} & -\hat{\mathrm{h}}_{1,1}^{(2)} \end{bmatrix} \odot \begin{bmatrix} \mu_1^{(1)} e^{-\imath\angle\hat{\mathrm{h}}_{2,1}^{(1)}} & \mu_2^{(1)} e^{-\imath\angle\hat{\mathrm{h}}_{1,1}^{(1)}} \\ \mu_1^{(2)} e^{-\imath\angle\hat{\mathrm{h}}_{2,1}^{(2)}} & \mu_2^{(2)} e^{-\imath\angle\hat{\mathrm{h}}_{1,1}^{(2)}} \end{bmatrix} \\ &= \begin{bmatrix} \left(\hat{\mathrm{h}}_{2,1}^{(1)}\right)^{-1}\hat{\mathrm{h}}_{2,2}^{(1)} & \left(\hat{\mathrm{h}}_{1,1}^{(1)}\right)^{-1}\hat{\mathrm{h}}_{1,2}^{(1)} \\ -1 & -1 \end{bmatrix} \odot \begin{bmatrix} \mu_1^{(1)}|\hat{\mathrm{h}}_{2,1}^{(1)}| & \mu_2^{(1)}|\hat{\mathrm{h}}_{1,1}^{(1)}| \\ \mu_1^{(2)}|\hat{\mathrm{h}}_{2,1}^{(2)}| & \mu_2^{(2)}|\hat{\mathrm{h}}_{1,1}^{(2)}| \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} \left(\hat{\mathrm{h}}_{2,1}^{(1)}\right)^{-1}\hat{\mathrm{h}}_{2,2}^{(1)} & \left(\hat{\mathrm{h}}_{1,1}^{(1)}\right)^{-1}\hat{\mathrm{h}}_{1,2}^{(1)} \\ -1 & -1 \end{bmatrix}}_{\mathbf{W}} \odot \underbrace{\begin{bmatrix} \lambda_1^{(1)} & \lambda_2^{(1)} \\ \lambda_1^{(2)} & \lambda_2^{(2)} \end{bmatrix}}_{\mathbf{\Lambda}}. \end{aligned} \tag{6.19}$$

In this specific case, $\lambda_i^{(j)}$ is given by $\lambda_i^{(j)} = \mu_i^{(1)}|\hat{\mathrm{h}}_{\bar{i}i}^{(1)}|$. Furthermore, the equivalent precoding matrix obtained in the centralized scenario is given by

$$\mathbf{T}^\star = \underbrace{\begin{bmatrix} \hat{\mathrm{h}}_{2,1}^{-1}\hat{\mathrm{h}}_{2,2} & \hat{\mathrm{h}}_{1,1}^{-1}\hat{\mathrm{h}}_{1,2} \\ -1 & -1 \end{bmatrix}}_{\mathbf{W}^\star} \underbrace{\begin{bmatrix} \lambda_1^\star & 0 \\ 0 & \lambda_2^\star \end{bmatrix}}_{\mathbf{\Lambda}^\star}. \tag{6.20}$$

Table 6.1 – Precoder $\mathbf{t}_1 = [\lambda_1^{(1)}\mathbf{w}_{1,1}^{(1)},\ \lambda_1^{(2)}\mathbf{w}_{1,2}^{(2)}]^{\mathrm{T}}$ for the data symbols of RX 1 according to the CSIT configuration.

	Main TX 1	Main TX 2	Local CSIT	Non-local CSIT
	$\alpha_{2,1}^{(1)} > \alpha_{2,1}^{(2)}$ $\alpha_{2,2}^{(1)} \geq \alpha_{2,2}^{(2)}$	$\alpha_{2,1}^{(1)} > \alpha_{2,1}^{(2)}$ $\alpha_{2,2}^{(1)} \geq \alpha_{2,2}^{(2)}$	$\alpha_{2,1}^{(1)} > \alpha_{2,1}^{(2)}$ $\alpha_{2,2}^{(1)} < \alpha_{2,2}^{(2)}$	$\alpha_{2,1}^{(1)} \leq \alpha_{2,1}^{(2)}$ $\alpha_{2,2}^{(1)} \geq \alpha_{2,2}^{(2)}$
$\mathbf{w}_1$	$\begin{bmatrix} (\hat{\mathbf{h}}_{2,1}^{(1)})^{-1}\hat{\mathbf{h}}_{2,2}^{(1)} \\ -1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ -(\hat{\mathbf{h}}_{2,2}^{(1)})^{-1}\hat{\mathbf{h}}_{2,1}^{(1)} \end{bmatrix}$	$\begin{bmatrix} (\hat{\mathbf{h}}_{2,1}^{(1)})^{-1} \\ -(\hat{\mathbf{h}}_{2,2}^{(2)})^{-1} \end{bmatrix}$	$\begin{bmatrix} \hat{\mathbf{h}}_{2,2}^{(1)} \\ -\hat{\mathbf{h}}_{2,1}^{(2)} \end{bmatrix}$
$\begin{bmatrix} \lambda_1^{(1)} \\ \lambda_1^{(2)} \end{bmatrix}$	$\begin{bmatrix} \mu_1^{(1)}\lvert\hat{\mathbf{h}}_{2,1}^{(1)}\rvert \\ \mu_1^{(2)}\lvert\hat{\mathbf{h}}_{2,1}^{(2)}\rvert \end{bmatrix}$	$\begin{bmatrix} \mu_1^{(1)}\lvert\hat{\mathbf{h}}_{2,2}^{(1)}\rvert \\ \mu_1^{(2)}\lvert\hat{\mathbf{h}}_{2,2}^{(2)}\rvert \end{bmatrix}$	$\begin{bmatrix} \mu_1^{(1)}\lvert\hat{\mathbf{h}}_{2,1}^{(1)}\rvert\lvert\hat{\mathbf{h}}_{2,2}^{(1)}\rvert \\ \mu_1^{(2)}\lvert\hat{\mathbf{h}}_{2,1}^{(2)}\rvert\lvert\hat{\mathbf{h}}_{2,2}^{(2)}\rvert \end{bmatrix}$	$\begin{bmatrix} \mu_1^{(1)} \\ \mu_1^{(2)} \end{bmatrix}$

It is important to remark that, in the centralized scenario, all the different precoder expressions obtained from Table 6.1 are equivalent, since the only difference is the phase shift and it does not affect the rate performance. Note also that for any channel coefficient there exist two estimates, $\hat{\mathrm{h}}_{i,j}^{(1)}$ and $\hat{\mathrm{h}}_{i,j}^{(2)}$, and $\mathbf{W}$ only depends on the one with higher accuracy. Indeed, the matrix $\mathbf{W}$ matches the precoding matrix for the genie-aided centralized setting $\mathbf{W}^\star$ and, as consequence, the possible performance loss comes only from the divergence between $\mathbf{\Lambda}$ and $\mathbf{\Lambda}^\star$. The idea behind this separation is that the interference nulling of 6.12 has to be extremely accurate but it can be performed by a single TX, whereas the power normalization has to be done by both TXs but it can be computed with a reduced precision, allowing the TXs to be consistent.

Remark 6.3. Table 6.1 illustrates all the possible relations between $\alpha_{2k}^{(j)}$, for all $j, k \in \mathbb{N}_2$. The precoder for RX 2 has the same number of possible configurations. Nevertheless, from the CSIT model assumed, not all of them are possible. Indeed, we have assumed that the RX feeds back its channel *vector* quantized with RVQ. Then, the CSIT accuracy at one TX for the whole channel vector is always the same, i.e., $\alpha_{2,1}^{(j)} = \alpha_{2,2}^{(j)}$. This implies that only the two first cases –in which a TX knows with a better accuracy the full channel vector of a certain RX– are the only possible ones. However, we have included all the possible cases for two reasons. The first one is to show the analogy with respect to the scheme of Chapter 3 which achieves the optimal GDoF. The second reason is because, if we abstract the mathematical model from the feedback mechanism, it is possible to assume that each coefficient may have a different accuracy but following the random distribution arising from RVQ feedback. Then, the results of this chapter still hold. □

6.3.2 Instantaneous Power Control

We have postponed the explanation about the power parameter for sake of readability, as it depends on the $\mathbf{\Lambda}$ matrix and its λ_i coefficients, which have just been introduced. The instantaneous power normalization algorithm is the same for the distributed and the centralized setting. The only difference is the information that each TX owns to compute it. Consequently, we omit any reference to the super-indexes indicating who is computing the coefficient $-(\cdot)^{(j)}$ or $(\cdot)^{\star}-$.

The power normalization strategy is performed by the parameter μ_i and follows any algorithm that belongs to a broad family of functions satisfying the per-TX instantaneous power constraint $\|\mathbf{t}_{\mathrm{TX}\,j}\| \leq 1$, $\forall j \in \mathbb{N}_2$. We recall that the term *instantaneous* refers only to the precoding vector power. The power of the transmit signal depends on the power of the data symbols, and thus it satisfies an *average* power constraint. We model the power control as a function Λ_i such that $\forall i \in \mathbb{N}_2$,

$$\lambda_i = \Lambda_i\Big(\hat{\mathbf{H}}, \boldsymbol{\alpha}, P\Big), \tag{6.21}$$

where $\lambda_i \in \mathbb{R}$. We assume that Λ_i is C^1, i.e., all its partial derivatives exist and are continuous, and that its Jacobian Matrix $\mathbf{J}_\Lambda$ satisfies $\|\mathbf{J}_\Lambda\| \leq M_{\mathbf{J}} < \infty$. Moreover, the probability density function of Λ_i, denoted as f_{Λ_i} is bounded away from infinity such that

$$\max_x f_{\Lambda_i}(x) \leq f_{\Lambda_i}^{\max} < \infty. \tag{6.22}$$

From the RVQ feedback assumption, $\hat{\mathbf{H}}$ is distributed as $\tilde{\mathbf{H}}$ and hence the marginal PDF $f_{\Lambda_i}(x)$ is the same for perfect, imperfect centralized and distributed CSIT. To conclude, since the power control acts on the normalized precoder, the instantaneous power constraint per TX implies that

$$0 \leq \lambda_i \leq 1. \tag{6.23}$$

6.3.3 Discretization of Power Normalization Parameters

Although the previous CSIT-adapted precoder attains better performance that the naive ZF scheme, it is still governed by the worst CSIT accuracy among the TXs. Indeed, as we will show in the numerical results section, it does not achieve the DoF of the centralized CSIT setting. A possible refinement would be to let the best informed TX predict what is the channel coefficient estimate at the other TX. However, it is easy to see that the error on such prediction is proportional to the worst accuracy among the two TXs. In this way, the performance is not improved.

We are interested in any case in enabling certain consistency between the TXs, such

that the power control applied at each TX is consistent. We implement this idea by *discretizing* the power normalization matrix. In particular, let $\mathcal{Q}(\cdot)$ represent the output of an arbitrary quantizer $\mathcal{Q}$ satisfying that $\mathcal{Q}(x) \leq x$. Hence, after computing the power parameter $\lambda_i^{(j)}$, TX j quantizes it to obtain $\mathcal{Q}(\lambda_i^{(j)})$. TX j applies then the quantized version to the precoding matrix. Consequently, the Hybrid Active-Passive Zero-Forcing precoder, denoted by $\mathbf{T}^{\mathrm{HAP}} \in \mathbb{C}^{2\times 2}$, is given by

$$\mathbf{T}^{\mathrm{HAP}} \triangleq \mathbf{W} \odot \begin{bmatrix} \mathcal{Q}(\lambda_1^{(1)}) & \mathcal{Q}(\lambda_2^{(1)}) \\ \mathcal{Q}(\lambda_1^{(2)}) & \mathcal{Q}(\lambda_2^{(2)}) \end{bmatrix}. \tag{6.24}$$

where $\mathbf{W}$ varies according to the CSIT configuration as illustrated in Table 6.1 and, by definition, only depends on the most accurate estimate for each channel coefficient.

At first sight, the performance obtained by using (6.24) would be degraded in comparison with the previous non-quantized precoder, as the quantizer decreases the accuracy of the parameters. This is not case, as we explain in the following. From (6.24), two possible cases arise. Namely, either $\mathcal{Q}(\lambda_1^{(1)}) = \mathcal{Q}(\lambda_1^{(2)})$ or not. If the two quantized estimations do not match, the performance is degraded. However, if they match, the precoding matrix becomes

$$\mathbf{T}^{\mathrm{HAP}} \triangleq \mathbf{W} \begin{bmatrix} \mathcal{Q}(\lambda_1^{(1)}) & 0 \\ 0 & \mathcal{Q}(\lambda_2^{(2)}) \end{bmatrix}. \tag{6.25}$$

Thus, the precoder recovers the original centralized shape of (6.18). This implies that the interference cancellation is achieved up to the centralized level, as $\mathbf{W} = \mathbf{W}^{\star}$ and the matrix of power normalization does not break the orthogonality. The only impairment comes from the reduction of transmitted power, as $\mathcal{Q}(x) \leq x$.

6.3.4 Properties of the Quantizer

In the following, we summarize some inherent properties of the quantizer $\mathcal{Q}$, as well as some desired behavior. Note that the condition $\mathcal{Q}(x) \leq x$ is mandatory so as to not infringe the instantaneous unitary power constraint. Furthermore, $\lambda_i^{(j)} \in [0, 1]$ from (6.23) and therefore $\mathcal{Q}(\lambda_i^{(j)}) \in [0, 1]$. We assume that it exists $M_{\mathcal{Q}} < \infty$ such that

$$\left| \mathbb{E}_{|\mathcal{Q}(x)>0} \left[\log_2 \left(\mathcal{Q}(x) \right) \right] \right| \leq M_{\mathcal{Q}}, \tag{P0}$$

which is a technical assumption that is satisfied by any non-degenerate quantizer. The role of $\mathcal{Q}$ is to trade-off the accuracy of the power control with the consistency of the decision at the TXs, since the ZF orthogonality of (6.12) is preserved only if both TXs obtain the same quantization value ($\mathcal{Q}(\lambda_i^{(1)}) = \mathcal{Q}(\lambda_i^{(2)})$). In order to emphasize the

relevance of the quantizer, we define Ω as the set of estimates $(\hat{\mathbf{H}}^{(1)}, \hat{\mathbf{H}}^{(2)})$ that ensure that the ZF orthogonality is not violated, excluding degenerate cases, i.e.,

$$\Omega \triangleq \left\{ (\hat{\mathbf{H}}^{(1)}, \hat{\mathbf{H}}^{(2)}) \middle| \; \forall i \in \mathbb{N}_2 \quad \mathcal{Q}\big(\lambda_i^{(1)}\big) = \mathcal{Q}\big(\lambda_i^{(2)}\big) \in \mathbb{R}^+ \right\}. \tag{6.26}$$

In simple words, Ω encloses the cases in which the TXs agree on the power normalization coefficients for both RXs and they are strictly positive. We further denote the complementary event of Ω as Ω^{c} (the *inconsistent* cases). We proceed by introducing two important properties for the quantizers.

Definition 6.2 (Asymptotically Accurate Quantizers)**.** A quantizer $\mathcal{Q}$ is said to be *asymptotically accurate* if

$$\lim_{P\to\infty} \mathcal{Q}(\lambda_i^{(j)}) = \lambda_i^{\star} \quad \text{a.s.} \quad \forall i, j \in \mathbb{N}_2, \tag{P1}$$

where a.s. stands for *almost surely*.

Definition 6.3 (Asymptotically Consistent Quantizers)**.** A quantizer $\mathcal{Q}$ is said to be *asymptotically consistent* if

$$\Pr\left(\Omega^{\mathsf{c}}\right) = o\left(\frac{1}{\log_2(P)}\right). \tag{P2}$$

Property (P2) implies that *inconsistent precoding* events are negligible in terms of asymptotic rate. We exhibit in the following lemma one particular quantizer satisfying properties (P1)-(P2). Optimizing further this quantizer is crucial to good performance at finite SNR and its optimization its an interesting topic for future research.

Lemma 6.1. *Let* $\alpha_{\min} = \min_{i,j\in\mathbb{N}_2}(\alpha_i^{(j)})$. *Let* $\mathcal{Q}_u$ *be a uniform quantizer in the interval* $[0, 1]$ *with a step size of* $\bar{P}^{\frac{-\alpha_{\min}}{c_q}}$, *with* $c_q > 1$, *such that*

$$\mathcal{Q}_u(x) \triangleq \bar{P}^{\frac{-\alpha_{\min}}{c_q}} \left\lfloor \bar{P}^{\frac{\alpha_{\min}}{c_q}} x \right\rfloor. \tag{6.27}$$

Then, $\mathcal{Q}_u$ *satisfies properties* (P0), (P1) *and* (P2).

Proof. The proof is relegated to Appendix F. ∎

6.4 Rate Gap of ZF in the Distributed CSIT Setting

Let us denote the expected sum rate achieved using HAP-ZF precoding in the D-CSIT setting by $R^{\mathrm{HAP}}(\boldsymbol{\alpha})$. Similarly, the expected sum rate attained by the centralized ZF precoder of Section 6.2 on the basis of the best estimates is denoted as $R^{\mathrm{ZF}}(\boldsymbol{\alpha}^{\star})$. Note that

$$\boldsymbol{\alpha}^{\star} = [\max_{j\in\mathbb{N}_2} \alpha_1^{(j)}, \quad \max_{j\in\mathbb{N}_2} \alpha_2^{(j)}]. \tag{6.28}$$

Accordingly, the rate gap between those settings is defined as

$$\Delta\mathrm{R} \triangleq R^{\mathrm{ZF}}(\boldsymbol{\alpha}^{\star}) - R^{\mathrm{HAP}}(\boldsymbol{\alpha}). \tag{6.29}$$

For the cases in which $\mathcal{Q}(\lambda_i^{(1)}) = \mathcal{Q}(\lambda_i^{(2)})$, we refer to both quantized parameters as $\lambda_i^{\mathcal{Q}}$, i.e.,

$$\text{given } \Omega, \quad \lambda_i^{\mathcal{Q}} = \mathcal{Q}(\lambda_i^{(j)}), \quad \forall j \in \mathbb{N}_2. \tag{6.30}$$

We can now state our main results.

Theorem 6.1. *Consider*

$$\Gamma_i \triangleq \left|\frac{\lambda_i^{\star}}{\lambda_i^{\mathcal{Q}}}\right|^2. \tag{6.31}$$

and $\Gamma_i^{\mathrm{AV}} \triangleq \mathbb{E}_{|\Omega}[\log_2(\Gamma_i)]$. *Then, the rate gap of ZF precoding with distributed CSIT is upper bounded by*

$$\Delta\mathrm{R} \leq \Gamma_1^{\mathrm{AV}} + \Gamma_2^{\mathrm{AV}} + \Pr(\Omega^{\mathsf{c}})\, R_{|\Omega^{\mathsf{c}}}^{\mathrm{ZF}}(\boldsymbol{\alpha}^{\star}), \tag{6.32}$$

where Ω *is defined in* (6.26), *and it holds that* $R_{|\Omega^{\mathsf{c}}}^{\mathrm{ZF}}(\alpha^{(1)}) \leq 2\log_2(1+P)$.

The proof is detailed in Section 6.5. This bound depends on the set Ω and thus on the quantizer selected. Intuitively, a *good* quantizer has to ensure a high probability of agreement, so as to make $\Pr(\Omega^{\mathsf{c}})$ small. This can be done by enlarging the quantization step, what will make the first term bigger, as $\mathcal{Q}(\lambda_i^{(j)})$ needs to be as close to $\lambda_i^{\star}$ as possible. This shows why finding the optimal quantizer is a challenging research topic. Nevertheless, there exists a family of quantizers that behave asymptotically optimal, as stated in the following theorem.

Theorem 6.2. *Let $\mathcal{Q}$ be an arbitrary quantizer satisfying (P0), (P1) and (P2). Then, taking the limit in Theorem 6.1 as P approaches infinity yields*

$$\lim_{P\to\infty} \Delta \mathrm{R} \leq 0. \tag{6.33}$$

Proof. The proof follows from Theorem 6.1. First, note that the sum rate $R^{\mathrm{ZF}}_{|\Omega^{\mathsf{c}}}(\alpha^{\star})$ is trivially bounded by twice the interference-free single-user rate to obtain

$$\begin{aligned} R^{\mathrm{ZF}}_{|\Omega^{\mathsf{c}}}(\boldsymbol{\alpha}^{\star}) &\leq \sum_{i=1}^{2} \log_2\left(1+\frac{P}{2}\,\mathbb{E}\left[\|\mathbf{h}_i\|^2\right]\right) \\ &= 2\log_2\left(1+P\right), \end{aligned} \tag{6.34}$$

what together with property (P2) implies that

$$\Pr\left(\Omega^{\mathsf{c}}\right) R^{\mathrm{ZF}}_{|\Omega^{\mathsf{c}}}(\alpha^{(1)}) = o(1). \tag{6.35}$$

Consequently, it only remains to show that $\lim_{P\to\infty} \mathbb{E}_{|\Omega}[\log_2(\Gamma_i)] = 0$ to conclude the proof . From the definition of Γ_i, it holds that

$$\mathbb{E}_{|\Omega}[\log_2\left(\Gamma_i\right)] = \mathbb{E}_{|\Omega}\left[\log_2\left(\lambda_i^{\star}\right)\right] - \mathbb{E}_{|\Omega}\left[\log_2\left(\mathcal{Q}(\lambda_i^{(1)})\right)\right]. \tag{6.36}$$

We recall a simple property on conditional probability. The law of total probability stays that, for any two events A, B,

$$\mathbb{E}_{|A}[\log_2(x)] = \Pr(B \mid A)\mathbb{E}_{|A\cap B}[\log_2(x)] + \Pr(B^{\mathsf{c}} \mid A)\mathbb{E}_{|A\cap B^{\mathsf{c}}}[\log_2(x)]. \tag{6.37}$$

Consider that $0 \leq x \leq 1$ and that $\Pr(B \mid A) > 0$. Then, $\mathbb{E}_{|A\cap B^{\mathsf{c}}}[\log_2(x)] \leq 0$ and

$$\mathbb{E}_{|A\cap B}[\log_2(x)] \geq \frac{1}{\Pr(B \mid A)}\mathbb{E}_{|A}[\log_2(x)]. \tag{6.38}$$

Therefore, if $\mathbb{E}_{|A}[\log_2(x)]$ exists, also $\mathbb{E}_{|A\cap B}[\log_2(x)]$ exists and it is bounded below by (6.38) and above by 0.

Now, suppose that A and B are given by $A = \{\mathcal{Q}(\lambda_i^{(1)}) > 0, \forall i\}$ and $B = \left\{\mathcal{Q}\left(\lambda_i^{(1)}\right) = \mathcal{Q}\left(\lambda_i^{(2)}\right), \forall i\right\}$. Thus, $\Omega = A \cap B$. It follows from (6.38) and (P0) that

$$\mathbb{E}_{|\Omega}[\log_2\left(\mathcal{Q}(\lambda_i^{(1)})\right)] \geq -\frac{\Pr(\mathcal{Q}(\lambda_i^{(1)}) > 0, \forall i)}{\Pr(\Omega)} M_{\mathcal{Q}}, \tag{6.39}$$

where we have applied the fact that $\Pr(B|A) = \frac{\Pr(A\cap B)}{\Pr(A)}$. Hence, $\mathbb{E}_{|\Omega}[\log_2(\mathcal{Q}(\lambda_i^{(1)}))]$ is bounded. The same result follows for $\mathbb{E}_{|\Omega}[\log_2\left(\lambda_i^{(1)}\right)]$ from the bounded density assumption of (6.22). Moreover, from the continuity of the log function and (P1), $\log_2(\mathcal{Q}(\lambda_i^{(1)}))$ converges a.s. to $\log_2(\lambda_i^{(1)})$. From all these facts, we can apply Lebesgue's Dominated Convergence Theorem [158, Theorem 16.4] to interchange expectation and limit and show that

$$\lim_{P\to\infty} \mathbb{E}_{|\Omega}\left[\log_2\left(\mathcal{Q}(\lambda_i^{(1)})\right)\right] = \mathbb{E}_{|\Omega}\left[\log_2\left(\lambda_i^{(1)}\right)\right], \tag{6.40}$$

and thus $\lim_{P\to\infty} \mathbb{E}_{|\Omega}[\log_2(\Gamma_1)] = 0$, which concludes the proof. ■

Corollary 6.1 (Rate Offset with HAP precoder)**.** *It holds from Theorem 6.2 that the rate offset $\mathcal{R}_\infty$ –defined in (6.2)– of ZF with D-CSIT is the same as for the genie-aided centralized setting, whose rate offset was shown to be constant with respect to Perfect CSIT ZF [132] (and thus with respect to the capacity-achieving Dirty Paper Coding) for $\alpha = 1$. Specifically, for a constant b, if the number of bits is $B = \log_2(P) - \log_2(b)$, the rate offset with respect to Perfect CSIT ZF is $\log_2(b)$.*

The key for attaining such performance is the trade-off between *consistency* and *accuracy* that is ruled by the quantizer. The result also implies that the logical separation between interference cancellation and power adjustment is instrumental: The power adjustment can be implemented with a low accuracy, as it is more important that the TXs agree. Conversely, the interference cancellation has to be applied with high accuracy. Interestingly, Lemma 6.1 illustrates that simple quantizers –as the uniform one– satisfy the sufficient conditions of convergence if we select the correct number of quantization levels. Moreover, since this quantizer is applied locally and no information exchange is done, the granularity of the quantizer does not increase the complexity of the scheme. Note that, if $\mathcal{Q}$ has a single quantization point, it leads to a statistical power control. In turn, if $\mathcal{Q}$ has infinite quantization points the scheme corresponds to the unquantized scheme. In both cases, part of the DoF is lost.

6.5 Proof of Theorem 6.1

We consider w.l.o.g. the rate difference at RX 1, denoted as ΔR_1, since the proof for RX 2 is obtained after switching the indexes of the RXs. ΔR_1 can be split as

$$\Delta R_1 = \Pr(\Omega)\,\Delta R_{1|\Omega} + \Pr(\Omega^c)\,\Delta R_{1|\Omega^c}. \tag{6.41}$$

First, we focus on $\Delta \mathrm{R}_{1|\Omega}$, which encloses the *consistent precoding* cases. Thus, we condition on Ω, albeit in the following we may omit to mention it explicitly for sake of concision. Conditioned on Ω it holds that $\mathcal{Q}\big(\lambda_i^{(1)}\big) = \mathcal{Q}\big(\lambda_i^{(2)}\big)$, $\forall i \in \mathbb{N}_2$, and hence we can use the notation $\lambda_i^{\mathcal{Q}}$ introduced in (6.30). Moreover, it can be observed from (6.25) that, conditioned on Ω, the HAP-ZF precoder satisfies

$$\mathbf{t}_i^{\mathrm{HAP}} = \frac{\lambda_i^{\mathcal{Q}}}{\lambda_i^{\star}} \mathbf{t}_i^{\mathrm{ZF}}, \qquad \forall i \in \mathbb{N}_2. \tag{6.42}$$

From the definition of the centralized scheme, $\lambda_i^{\star} = \lambda_i^{(\mathrm{argmax}_j \, \alpha_i^{(j)})}$. Given that $\mathcal{Q}(x) \leq x$, it follows that $\lambda_i^{\mathcal{Q}}/\lambda_i^{\star} \leq 1$, $\forall i \in \mathbb{N}_2$. Let us recall that Γ_i is defined as

$$\Gamma_i \triangleq \left| \frac{\lambda_i^{\star}}{\lambda_i^{\mathcal{Q}}} \right|^2, \tag{6.43}$$

Then, Γ_i satisfies $\Gamma_i \geq 1$ $\forall i \in \mathbb{N}_2$. Conditioned on Ω we can write that the SINR obtained through HAP-ZF precoding satisfies

$$\begin{aligned} 1 + \frac{\frac{P}{2} \left| \mathbf{h}_1^{\mathrm{H}} \mathbf{t}_1^{\mathrm{HAP}} \right|^2}{1 + \frac{P}{2} \left| \mathbf{h}_1^{\mathrm{H}} \mathbf{t}_2^{\mathrm{HAP}} \right|^2} &= 1 + \frac{\frac{1}{\Gamma_1} \frac{P}{2} \left| \mathbf{h}_1^{\mathrm{H}} \mathbf{t}_1^{\mathrm{ZF}} \right|^2}{1 + \frac{1}{\Gamma_2} \frac{P}{2} \left| \mathbf{h}_1^{\mathrm{H}} \mathbf{t}_2^{\mathrm{ZF}} \right|^2} \\ &\geq \frac{1}{\Gamma_1} \left(1 + \frac{\frac{P}{2} \left| \mathbf{h}_1^{\mathrm{H}} \mathbf{t}_1^{\mathrm{ZF}} \right|^2}{1 + \frac{P}{2} \left| \mathbf{h}_1^{\mathrm{H}} \mathbf{t}_2^{\mathrm{ZF}} \right|^2} \right), \end{aligned} \tag{6.44}$$

where the first equality follows from (6.42)-(6.43) whereas the last inequality comes from the fact that $1/\Gamma_i \leq 1$ $\forall i$. We can recognize in (6.44) the SINR at RX 1 for the centralized ZF scheme such that it holds:

$$\begin{aligned} R_{1|\Omega}^{\mathrm{HAP}}(\boldsymbol{\alpha}) &= \mathbb{E}_{|\Omega} \left[\log_2 \left(1 + \frac{\frac{P}{2} |\mathbf{h}_1^{\mathrm{H}} \mathbf{t}_1^{\mathrm{HAP}}|^2}{1 + \frac{P}{2} |\mathbf{h}_1^{\mathrm{H}} \mathbf{t}_2^{\mathrm{HAP}}|^2} \right) \right] \\ &\geq -\mathbb{E}_{|\Omega}[\log_2(\Gamma_1)] + R_{1|\Omega}^{\mathrm{ZF}}(\boldsymbol{\alpha}^{\star}). \end{aligned} \tag{6.45}$$

Since $\Delta \mathrm{R}_{1|\Omega} = R_{1|\Omega}^{\mathrm{ZF}}(\boldsymbol{\alpha}^{\star}) - R_{1|\Omega}^{\mathrm{HAP}}(\boldsymbol{\alpha})$, it follows that

$$\Delta \mathrm{R}_{1|\Omega} \leq \mathbb{E}_{|\Omega}[\log_2(\Gamma_1)]. \tag{6.46}$$

Let us now consider the *inconsistent precoding* cases, i.e., Ω^{c}. Since $R_{1|\Omega^{\mathsf{c}}}^{\mathrm{HAP}}(\boldsymbol{\alpha}) \geq 0$, the rate gap can be bounded by the centralized rate as $\Delta \mathrm{R}_{1|\Omega^{\mathsf{c}}} \leq R_{1|\Omega^{\mathsf{c}}}^{\mathrm{ZF}}(\boldsymbol{\alpha}^{\star})$. Putting these

results together in (6.41) yields

$$\Delta \mathrm{R}_1 \ \leq \mathbb{E}_{|\Omega}[\log_2(\Gamma_i)] + \Pr(\Omega^{\mathrm{c}})\, R^{\mathrm{ZF}}_{1|\Omega^{\mathrm{c}}}(\alpha^{(1)}), \tag{6.47}$$

where we have applied the fact that $\Pr(\Omega) \leq 1$. The sum rate gap $\Delta\mathrm{R}$ is obtained as $\Delta\mathrm{R} = \Delta\mathrm{R}_1 + \Delta\mathrm{R}_2$, what concludes the proof. ■

6.6 Analysis of the Power Normalization Parameters λ_i

Thus far, we have not focused on the statistics of the power normalization parameters λ_i, and we have abstracted them up to the sufficient conditions that they have to fulfill to prove the assymptotic results. In this section, we discuss them a bit more in detail. This discussion is motivated because the bound of (6.32) Theorem 6.1 strongly depends on those statistics. Note that

$$\Gamma_i^{\mathrm{AV}} = 2\mathbb{E}_{|\Omega}[\log_2(\lambda_i^{\star})] - 2\mathbb{E}_{|\Omega}\left[\log_2\left(\lambda_i^{\mathcal{Q}}\right)\right]. \tag{6.48}$$

In the following, we present the main conclusions. The derivation and further information is relegated to Appendix G. We recall that the the CSIT is composed of a quantized version of the normalized vector $\tilde{\mathbf{h}}_i \triangleq \frac{\mathbf{h}_i}{\|\mathbf{h}_i\|} \in \mathbb{C}^2$ using RVQ with $B_i^{(j)}$ bits. The assumption of Rayleigh fading ($\tilde{\mathrm{h}}_{i,k} \sim \mathcal{N}_{\mathbb{C}}(0,1)$) implies that

$$|\tilde{\mathrm{h}}_{i,k}|^2 \sim \mathrm{Uniform}(0,1), \tag{6.49}$$

$$|\tilde{\mathrm{h}}_{i,k}| \ \sim \mathrm{Triangular}(0,1), \tag{6.50}$$

$$|\tilde{\mathrm{h}}_{i,\bar{k}}|^2 = 1 - |h_{i,k}|^2. \tag{6.51}$$

From RVQ properties, the estimates $\hat{\mathrm{h}}_{i,k}^{(j)}$ follow the same distributions. We denote the precoding vector of TX j before normalization as $\mathbf{t}^o_{\mathrm{TX}\,j} \triangleq [\mathbf{t}^o_{\mathrm{TX}\,j,1},\ \mathbf{t}^o_{\mathrm{TX}\,j,2}]$. As indicated in (6.14), $\mathbf{t}^o_{\mathrm{TX}\,j,i} = \tilde{\mathrm{h}}^{(j)}_{i,\bar{j}}$. The final precoder of TX j is then $\mathbf{t}_{\mathrm{TX}\,j} = [\mu_1 \mathbf{t}^o_{\mathrm{TX}\,j,1},\ \mu_2 \mathbf{t}^o_{\mathrm{TX}\,j,2}]$. We consider two different power normalizations.

1. The coefficient μ_i is chosen such that $\mu_i \triangleq \frac{1}{\max(\|\mathbf{t}^o_{\mathrm{TX}\,1}\|, \|\mathbf{t}^o_{\mathrm{TX}\,2}\|)}$, , for any $i \in \mathbb{N}_2$.
2. We transmit $P/2$ power for each RX stream (symbol), and the precoder of each symbol is unit-norm. Hence, for any $i \in \mathbb{N}_2$, $\mu_i = 1$.

In order to ease the notation, consider that $\alpha_{i,k}^{(j)} = \alpha^{(j)}$ for all $i,k \in \mathbb{N}_2$. We assume w.l.o.g. that TX 1 is the TX with the most accurate CSIT, i.e., $\alpha^{(1)} \geq \alpha^{(2)}$. We analyze separately each one of the power normalization considered. Let f_x denote the PDF of the variable x and F_x denote the cumulative distribution function (CDF).

6.6.1 Maximum TX Norm Normalization

We consider here the first power normalization of the two cases mentioned above. Note that in this case $\mu_1 = \mu_2$ but $\lambda_1 \neq \lambda_2$, since $\lambda_i \triangleq \mu_i |\hat{h}_{\bar{i}1}^{(j)}|$, as shown in Table 6.1. Since the estimates $\hat{\mathrm{h}}_{i,k}^{(j)}$ have the same distributions as the true channel elements $\tilde{\mathrm{h}}_{i,k}$, we omit the super-index notation. Consequently, λ_i is given by

$$\lambda_i \triangleq \frac{|\tilde{\mathrm{h}}_{\bar{i},1}|}{\max(\|\mathbf{t}_{\mathrm{TX}\,1}^{o}\|, \|\mathbf{t}_{\mathrm{TX}\,2}^{o}\|)}. \tag{6.52}$$

Lemma 6.2. *For the power normalization of (6.52), it follows that*

$$f_{\lambda_i} = \frac{4x}{(1+x^2)^2} - 2x + \min\left(\frac{x}{(1-x^2)^2}, \frac{1}{x^3}\right), \tag{6.53}$$

and

$$\mathbb{E}[\log_2(\lambda_i)] = \frac{3 - 4\ln(4)}{4\ln(2)} \approx -0.917980. \tag{6.54}$$

Proof. The proof is relegated to Appendix G.1. In Fig. 6.2, the PDF of (6.53) is shown, together with the histogram of a Monte-Carlo based simulation, in order to verify the results. ■

Lemma 6.3. *Suppose that the quantization of the $\lambda_i^{(j)}$ parameters is done by the quantizer of Lemma 6.1, with step size q. Let p_i^0 be*

$$p_i^0 \triangleq \frac{1}{1 - \Pr\left(\lambda_i^{\mathcal{Q}} = 0 | \mathcal{Q}(\lambda_i^{(1)}) = \mathcal{Q}(\lambda_i^{(2)})\right)}. \tag{6.55}$$

Then,

$$\mathbb{E}_{|\Omega}\left[\log_2\left(\lambda_i^{\mathcal{Q}}\right)\right] = p_i^0\Big(\sum_{n=2}^{N-1} F_\lambda(nq)\log_2\left(\frac{n-1}{n}\right) + \log_2(N-1) + \log_2(q)\left(1 - F_\lambda(q)\right)\Big).$$

Note that, if $q^{-1} \in \mathbb{Z}$, the expression is simplified as $\mathbb{E}_{|\Omega}\left[\log_2\left(\lambda_i^{\mathcal{Q}}\right)\right] = p_i^0 \sum_{n=2}^{N} F_\lambda(nq)$.

Proof. The proof is relegated to Appendix G.1. ■

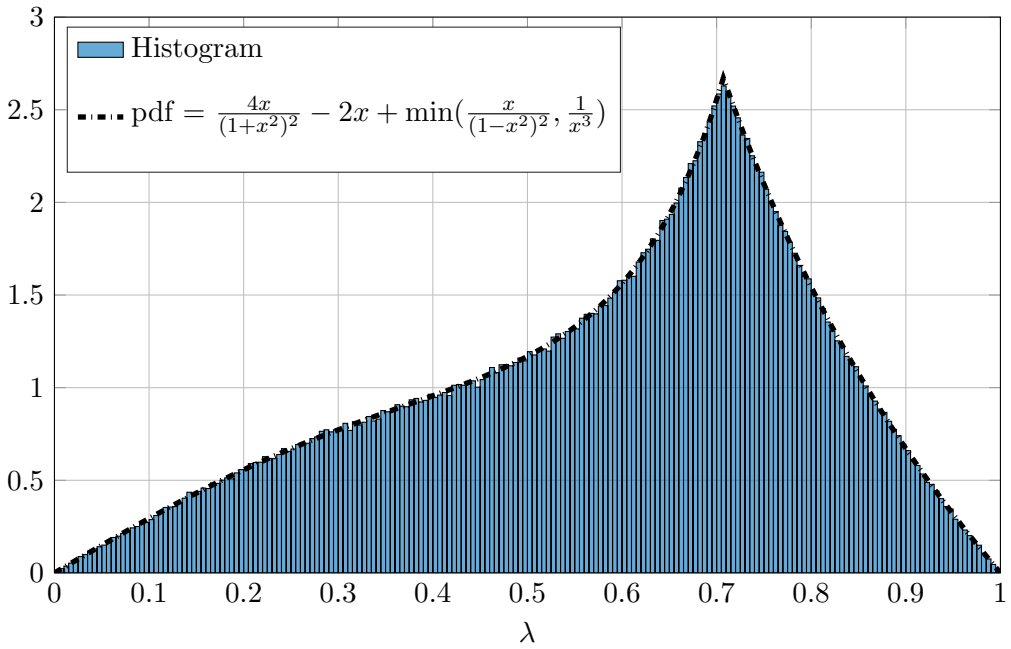


Figure 6.2 – Probability density function λ_i assuming $\mu = \max(P_{\mathrm{TX}\,1}, P_{\mathrm{TX}\,2})^{-1/2}$.

6.6.2 Unit-Norm per RX Normalization

We consider now the second normalization, i.e., $\mu_i = 1$. Therefore, $\lambda_i \triangleq \left|h_{\bar{i},1}\right|$, and thus

$$\lambda_i^2 \sim \mathrm{Uniform}(0,1), \tag{6.56}$$

$$\lambda_i \sim \mathrm{Triangular}(0,1). \tag{6.57}$$

Lemma 6.4. *For the power normalization of* (6.6.2)*, it follows that*

$$\mathbb{E}[\log_2(\lambda_i)] = \frac{-1}{2\ln(2)} \tag{6.58}$$

and

$$\mathbb{E}_{|\Omega}\left[\log_2\left(\lambda^2\right)\right] = \frac{-1}{\ln(2)} - \frac{q^2\log_2(q^2)}{1-q^2}. \tag{6.59}$$

Proof. The proof is relegated to Appendix G.2. ■

Lemma 6.5. *Suppose that the quantization of the $\lambda_i^{(j)}$ parameters is done by the quantizer of Lemma 6.1, with step size q. Let p_i^0 be*

$$p_i^0 \triangleq \frac{1}{1 - \Pr\left(\lambda_i^{\mathcal{Q}} = 0 | \mathcal{Q}(\lambda_i^{(1)}) = \mathcal{Q}(\lambda_i^{(2)})\right)}. \tag{6.60}$$

Then,

$$\mathbb{E}_{|\Omega}\left[\log_2\left(\lambda_i^{\mathcal{Q}}\right)\right] = \frac{q^2}{1-q^2}\sum_{n=1}^{N-2}\log_2(nq)(2n+1) + \log_2((N-1)q)\frac{1-(N-1)^2q^2}{1-q^2}.$$

Note that, if $q^{-1} \in \mathbb{Z}$, the expression is simplified as

$$\mathbb{E}_{|\Omega}\left[\log_2\left(\lambda_i^{\mathcal{Q}}\right)\right] = \frac{q^2}{1-q^2}\sum_{n=1}^{N-1}\log_2(nq)(2n+1). \tag{6.61}$$

Proof. The proof is relegated to Appendix G.2. ■

6.7 Numerical Results

We illustrate in the following some numerical results that corroborate the previous insights. Suppose that the quantization of the $\lambda_i^{(j)}$ parameters is computed with a uniform quantizer. For sake of simplification, we assume a basic power normalization that ensures the per-TX power constraint: Let $\mathbf{t}^o_{\mathrm{TX}\,j} \triangleq [\mathbf{t}^o_{\mathrm{TX}\,j,1},\ \mathbf{t}^o_{\mathrm{TX}\,j,2}]$ denote the precoding vector of TX j before normalization. The final precoder of TX j is $\mathbf{t}_{\mathrm{TX}\,j} = [\mu_1\mathbf{t}^o_{\mathrm{TX}\,j,1},\ \mu_2\mathbf{t}^o_{\mathrm{TX}\,j,2}]$. Then, μ_i is chosen as

$$\mu_i \triangleq \frac{1}{\max(\|\mathbf{t}^o_{\mathrm{TX}\,1}\|, \|\mathbf{t}^o_{\mathrm{TX}\,2}\|)}, \qquad \forall i \in \mathbb{N}_2. \tag{6.62}$$

Let us start with a simple CSIT configuration. Suppose the per-TX homogeneous CSIT configuration in which a TX j has the same accuracy for the whole channel matrix, i.e.,

$$\begin{aligned} \alpha_1^{(1)} &= \alpha_2^{(1)} = \alpha^{(1)}, \\ \alpha_1^{(2)} &= \alpha_2^{(2)} = \alpha^{(2)}. \end{aligned} \tag{6.63}$$

In particular, suppose that $\alpha^{(1)} = 1$ and $\alpha^{(2)} = 0.6$. Fig. 6.3 shows the simulated sum rate for this specific configuration. The simulation is computed by using Monte-Carlo

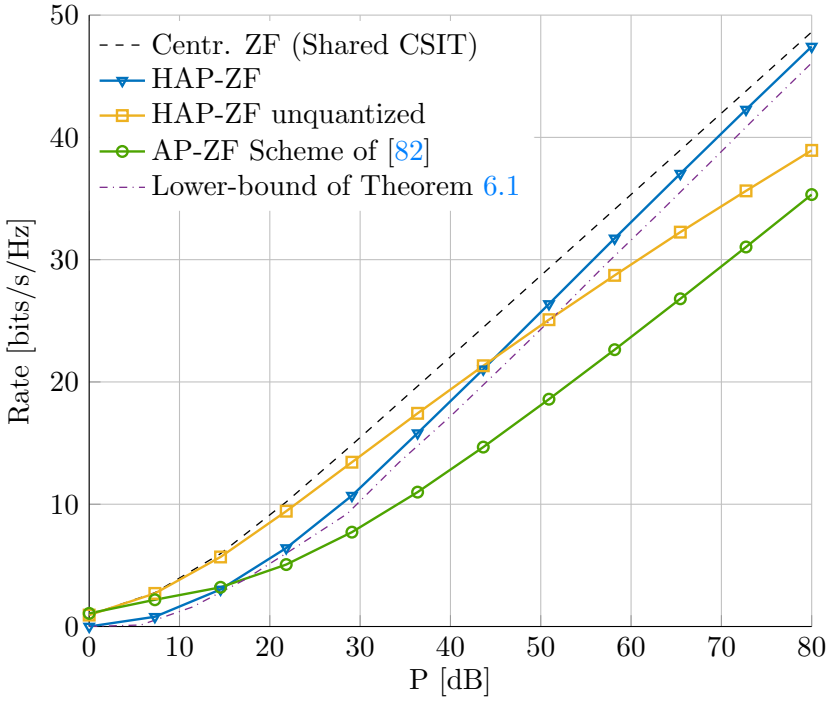


Figure 6.3 – Expected sum rate of the proposed scheme for the setting with CSIT scaling parameters $\alpha^{(1)} = 1$, $\alpha^{(2)} = 0.6$, using the uniform quantizer of Lemma 6.1.

runs and averaging over 1000 random codebooks and 1000 channel realizations. The quantization step q is selected as

$$q = \bar{P}^{-\alpha^{(2)}/2}. \tag{6.64}$$

We can see that the proposed HAP-ZF scheme leads to a vanishing rate loss with respect to the centralized case (where both TXs are provided with the best CSIT, $\hat{\mathbf{H}}^{(1)}$ in the simulated case). The lower-bound of Theorem 6.1 is considerably close to the actual rate. We recall that the only scheme known previously to achieve the centralized DoF in the D-CSIT setting was presented in [82] and it applies a scaled power normalization of $P/\log_2(P)$. Fig. 6.3 shows that, although this scheme achieves the optimal DoF –slope–, it achieves that property at the cost of a strong loss in rate offset.

Finally, the performance obtained if we do not quantize the $\lambda_i^{(j)}$ parameters is also presented (the so-called HAP-ZF unquantized in Fig. 6.3). It is clear that the unquantized scheme experiences a loss in terms of DoF. This occurs because, as aforementioned, the mismatches between the precoding coefficients of each TX break the orthogonality needed for the interference nulling. Thus, this scheme only achieves a DoF proportional to $\alpha^{(2)}$ instead of $\alpha^{(1)}$. At intermediate SNR, this unquantized scheme outperforms the proposed

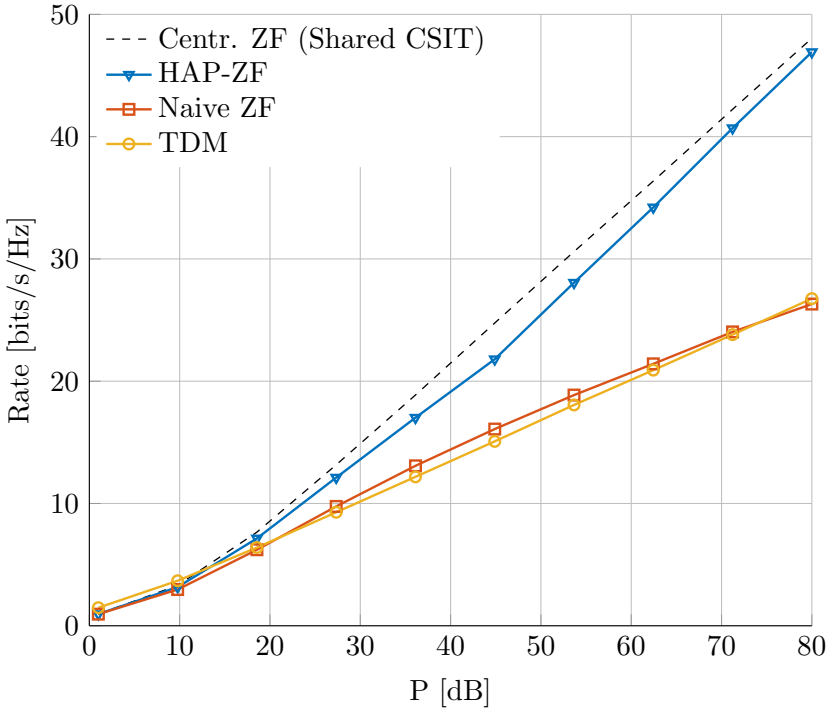


Figure 6.4 – Average sum rate comparison for the CSIT allocation in which the CSIT about RX 1 is known locally and the CSIT about RX 2 is known non-locally.

HAP-ZF precoding scheme. Yet, this is a consequence of our focus towards analytical tractability and asymptotic analysis, as the quantization step q has been selected in (6.64) only so as to satisfy the convergence conditions. Optimizing the precoder for finite SNR performance will allow to bridge the gap between the two schemes to obtain a scheme outperforming both of them.

Let us now present a different configuration. Importantly, in this case the quantization step q is optimized by exhaustive search. We still consider two levels of accuracy, this time allocated as

$$\text{At TX 1} \longrightarrow \quad \begin{matrix} \alpha_{1,1}^{(1)} = 1 & \alpha_{1,2}^{(1)} = 0.4 \\ \alpha_{2,1}^{(1)} = 0.4 & \alpha_{2,2}^{(1)} = 1 \end{matrix} \quad \begin{matrix} \alpha_{1,1}^{(2)} = 0.4 & \alpha_{1,2}^{(2)} = 1 \\ \alpha_{2,1}^{(2)} = 0.4 & \alpha_{2,2}^{(2)} = 1 \end{matrix} \quad \longleftarrow \text{At TX 2}$$

Note that this configuration corresponds to the case in which the CSIT about RX 1's channel is distributed *locally* –i.e., each TX has the most accurate estimate for its own coefficient–, and the CSIT about RX 1's channel is allocated in a non-locally way –i.e., each TX has the most accurate estimate for the other TX's coefficient–. Although this

CSIT allocation may not be found in practical use cases, we select it so as to illustrate the flexibility of the proposed scheme and the validity of the results. We show in Fig. 6.4 the rate achieved by the proposed HAP-ZF scheme together with three different reference schemes:

1. The centralized scheme for case in which the best estimate is shared by both TXs.
2. The Naive ZF, where the TXs compute the conventional ZF assuming that the other TX shares the same information. This scheme is not aware of the distributed allocation of the CSIT.
3. Simple Time Division Multiplexing (TDM). Only one RX is served at a given time.

We observe that the proposed scheme offers a considerably better performance with respect to the Naive ZF and TDM. Besides this, it can be seen that the HAP-ZF rate converges to the upper-bound of the centralized setting.

In the interest of better illustrating the convergence behavior of Theorem 6.1, we present in Fig. 6.5 the percentage of the centralized upper-bound attained by the proposed HAP-ZF scheme, for all the possible CSIT configurations. We show side-by-side the percentage attained at $P = 30,\ 50$ and 80dB. It is important to bear in mind that the values are normalized. The rate achieved by the centralized setting is 15 bits/Hz/s at 30 dB, 28 bits/Hz/s at 50 dB, and 48 bits/Hz/s at 80 dB, what can be seen from Fig. 6.4. Note that the more the SNR increases, the less percentage attains the Naive ZF scheme. Furthermore, Fig. 6.5 shows that the HAP-ZF scheme converges at different pace for each of the possible CSIT configurations. This shows how sensitive the scheme is to the probability distribution of the power parameters $\lambda_i^{(j)}$, since that distribution differs for different CSIT configurations.

6.8 Conclusions

Considering a decentralized scenario where each TX has a CSI with different SNR scaling accuracy, we have shown that there exists a linear precoding scheme that asymptotically recovers the rate of ZF precoding in the ideal centralized setting in which the best estimate is shared. Going beyond the setting considered, we have shown how using a low rate quantization of some parameters (here the power normalization) in combination with a higher-accuracy distributed decision allows to reach coordination without loosing precision. The extension of the results to more antennas and more users is considered in the following chapter. The optimization at finite SNR, as well as the extension of the main unveiled intuitions concerning distributed settings to diverse cooperative problems, are interesting and challenging open research problems.

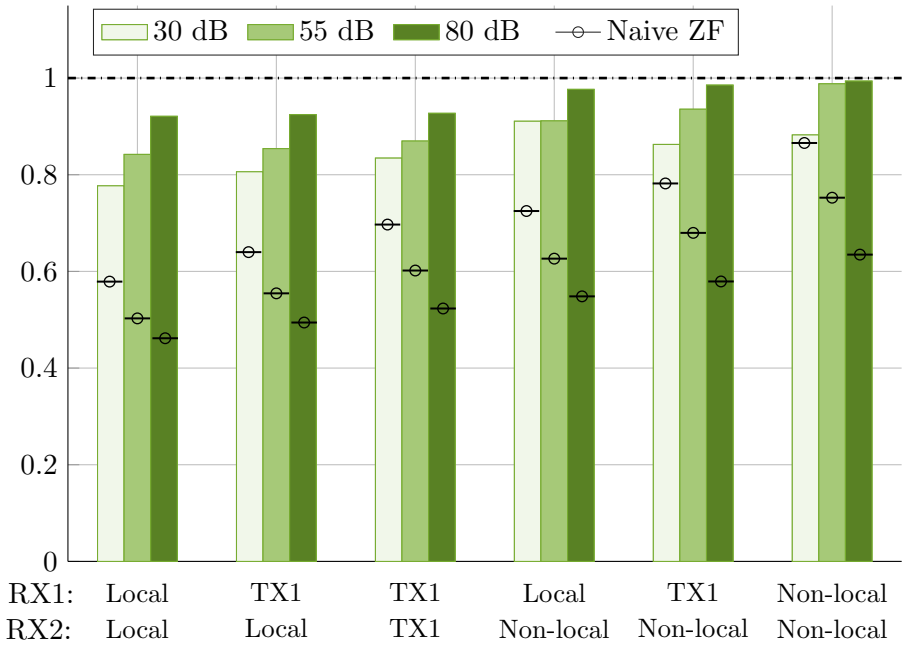


Figure 6.5 – Percentage of the C-CSIT setting rate attained by the HAP-ZF (green) and the Naive ZF (-o-) schemes, for any possible D-CSIT configuration. The configurations are described in Table 6.1.

Chapter 7

Rate Gap of the M×K Network MISO with Distributed CSIT

In the previous chapter, we have analyzed the performance of distributed ZF schemes for the 2×2 Network MISO setting, with the aftermath that it is possible to recover asymptotically the rate of the genie-aided centralized setting. Motivated by this result, we proceed as in Part II and we broaden the rate gap analysis to a more general setting. In this case, we extend it to the setting in which M multi-antenna TXs jointly serve K different RXs. In particular, we focus on the cases in which distributed ZF-type schemes can attain the DoF of the C-CSIT setting. This restriction is due to the fact that we study the asymptotic regime, and the rate gap for the cases in which the DoF is not attained grows unboundedly as the SNR increases.

As in the previous part, the schemes that were developed for the simple 2×2 case are not applicable to the general setting, as they rely on the fact that there exists only an interference constraint to resolve and therefore a single TX is enough to recover the DoF. Besides this, we consider now a CSIT model with Gaussian estimation noise. This modification of the estimation model with respect to the previous chapter, in which we assumed RVQ feedback, is applied so as to extend the scope of the results. Indeed, from the mathematical demonstrations in both chapters, the results are expected to hold for a broad set of distributions, since the analysis shows that the parameter that characterizes the asymptotic performance is the scaling of the noise variance, and not its distribution.

7.1 Preliminaries

7.1.1 Transmission and System Model

We consider the Network MISO setting in which M TXs jointly serve K single-antenna RXs. TX j has N_j antennas. We denote the total number of transmit antennas as $N_T = \sum_{j=1}^{M} N_j$. The received signal is defined as in Section 2.2. Every RX wishes to receive a data symbol s_i. The data symbols s_i are i.i.d. as $\mathcal{N}_{\mathbb{C}}(0,1)$. We define the vector of data symbols as $\mathbf{s} = [s_1, \ldots, s_K]^{\mathrm{T}}$, such that $\mathbb{E}\left[\|\mathbf{s}\|^2\right] = K$. We apply a minor modification to the notation of the channel matrix. The channel matrix is now written as

$$\mathbf{H} \triangleq \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_K \end{bmatrix} = \begin{bmatrix} \mathbf{h}_{1,1} & \ldots & \mathbf{h}_{1,M} \\ \vdots & \ddots & \vdots \\ \mathbf{h}_{K,1} & \ldots & \mathbf{h}_{K,M} \end{bmatrix} \in \mathbb{C}^{K \times N_T}. \tag{7.1}$$

Hence, $\mathbf{h}_i \in \mathbb{C}^{1 \times N_T}$ denotes the global channel vector towards RX i, and $\mathbf{h}_{i,j} \in \mathbb{C}^{1 \times N_j}$ is the channel vector from TX j to RX i. Note that we have defined the row vectors as $\mathbf{h}_i$ and $\mathbf{h}_{i,j}$ in place of the usual Hermitian notation $\mathbf{h}_i^{\mathrm{H}}$ and $\mathbf{h}_{i,j}^{\mathrm{H}}$. This is done so as to ease the notation for the remaining of the chapter. The channel coefficients are assumed to be i.i.d. as $\mathcal{N}_{\mathbb{C}}(0,1)$ such that all the channel sub-matrices are full rank with probability one. The precoding matrix is given by

$$\mathbf{T} \triangleq \begin{bmatrix} \mathbf{T}_1 \\ \vdots \\ \mathbf{T}_M \end{bmatrix} = \mu \begin{bmatrix} \mathbf{w}_1 & \ldots & \mathbf{w}_K \end{bmatrix} = \mu \begin{bmatrix} \mathbf{w}_{1,1} & \ldots & \mathbf{w}_{K,1} \\ \vdots & \ddots & \vdots \\ \mathbf{w}_{1,M} & \ldots & \mathbf{w}_{K,M} \end{bmatrix} \in \mathbb{C}^{N_T \times K}, \tag{7.2}$$

hence $\mathbf{T}_j \in \mathbb{C}^{N_j \times K}$ is the precoding matrix applied at TX j, $\mathbf{w}_i \in \mathbb{C}^{N_T \times 1}$ is the global precoding vector for the information symbols of RX i (s_i), and $\mathbf{w}_{i,j} \in \mathbb{C}^{N_j \times 1}$ is the precoding vector applied at TX j for s_i. Finally, we denote the coefficient at the n-th antenna of TX j as $\mathbf{w}_{i,j,n}$. The parameter $0 < \mu \leq 1$ is a power correction value that will be detailed later. We further define $\mathbf{T}_{j,n}$ as the precoding vector applied at the n-th antenna of TX j, with $n \in \mathbb{N}_{N_j}$. We assume that the precoder has a per-antenna instantaneous unit-norm constraint, such that

$$\|\mathbf{T}_{j,n}\| \leq 1. \tag{7.3}$$

The results presented here also hold under the assumption of per-TX instantaneous constraint ($\|\mathbf{T}_j\| \leq 1$). Note that, even if we set $\|\mathbf{T}_j\| = 1$, the transmit power varies over the time as the power of the information symbols s_i varies. With a huge abuse of notation,

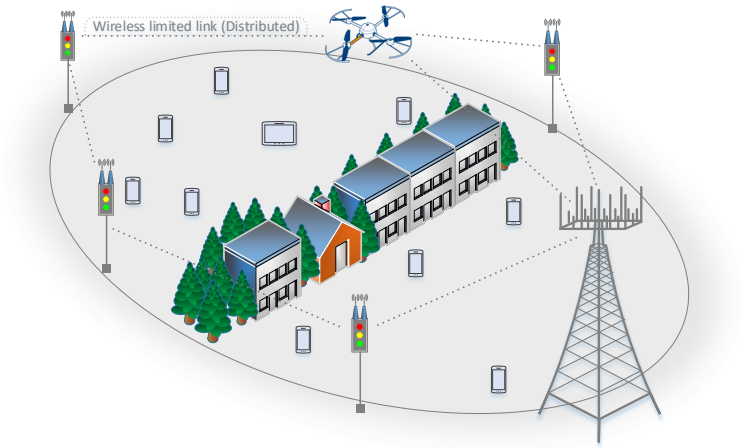


Figure 7.1 – Master Base Station with remote radio-heads. It obtains an estimate of the whole channel matrix, then it transmits noisy or compressed CSI to the auxiliary TXs.

and for sake of concision, we refer hereinafter to the unit-norm constraint of (7.3) as *instantaneous power constraint*, although strictly speaking it is an *instantaneous power constraint on the precoding vector*. This is done in opposition to the less restrictive average power constraint on the precoder ($\mathbb{E}[\|\mathbf{T}_j\|] \leq 1$) which has been assumed in Part II.

7.1.2 Distributed CSIT Model

We recall the expression of the channel estimate at TX j, presented in (2.15):

$$\hat{\mathbf{H}}^{(j)} \triangleq \sqrt{\mathbf{1}_{K\times N_T} - \mathbf{Z}^{(j)}} \odot \mathbf{H} + \sqrt{\mathbf{Z}^{(j)}} \odot \mathbf{\Delta}^{(j)}. \tag{7.4}$$

This chapter is characterized by the assumption of Gaussian estimation noise. In particular, besides the assumption of Rayleigh fading, we consider that the elements of $\mathbf{\Delta}^{(j)}$ are i.i.d. as $\mathcal{N}_{\mathbb{C}}(0,1)$ and they are independent of $\mathbf{H}$. Furthermore, we consider the Sorted CSIT configuration introduced in Section 2.4.4, as in Chapter 4. We assume for simplicity the homogeneous accuracy case, such that we can write $\mathbf{Z}^{(j)} = Z^{(j)} \in \mathbb{R}$. Consequently, (7.4) becomes

$$\hat{\mathbf{H}}^{(j)} \triangleq \sqrt{1 - Z^{(j)}}\, \mathbf{H} + \sqrt{Z^{(j)}} \mathbf{\Delta}^{(j)}. \tag{7.5}$$

Hence, a TX knows the full channel matrix with the same average accuracy. This model encloses e.g. a scenario in which a main, multi or massive antenna base station serves a set of users with the help of some single or multi antenna remote radio-head or simple TXs, as depicted in Fig. 7.1.

CSIT accuracy

Following the same reasoning of the previous chapters, our interest in the high-SNR regime motivates the assumption that the estimation error scales as

$$Z^{(j)} = P^{-\alpha^{(j)}}, \tag{7.6}$$

where $0 < \alpha^{(j)} \leq 1$. $\alpha^{(j)}$ is the *accuracy scaling* parameter that measures the quality of estimation of the channel matrix at TX j. Hence, we can order the TXs w.l.o.g. as

$$1 \geq \alpha^{(1)} \geq \alpha^{(2)} \geq \cdots \geq \alpha^{(M)} \geq 0, \tag{7.7}$$

what implies that TX 1 is the *best-informed* TX, whose CSIT has the highest accuracy. We define the set of accuracy parameters of the D-CSIT setting as

$$\boldsymbol{\alpha}_M = \{\alpha^{(j)}\}_{j \in \mathbb{N}_M}. \tag{7.8}$$

For further use, we define the estimate for the channel of RX i such that

$$\hat{\mathbf{h}}_i^{(j)} \triangleq \breve{z}^{(j)} \mathbf{h}_i + z^{(j)} \, \boldsymbol{\delta}_i^{(j)}, \tag{7.9}$$

where $z^{(j)} \triangleq \bar{P}^{-\alpha^{(j)}}$, $\breve{z}^{(j)} \triangleq \sqrt{1 - (z^{(j)})^2}$, and $\hat{\mathbf{h}}_i^{(j)}$, $\boldsymbol{\delta}_i^{(j)}$, are the i-th row of the matrices $\hat{\mathbf{H}}^{(j)}$, $\boldsymbol{\Delta}^{(j)}$, respectively. The accuracy parameters $\boldsymbol{\alpha}_M$ are assumed to be long-term coefficients that vary slowly. Based on that, it is assumed that every TX knows the full set $\boldsymbol{\alpha}_M$, as it only requires a sharing of few bits over a long period of time.

7.1.3 Genie-Aided Centralized Setting

We consider the same genie-aided centralized setting as in the previous chapter, defined in Definition 6.1. Hence, all the TXs are endowed with the estimate of best average accuracy. Note that in the sorted setting, where $\alpha^{(1)} \geq \cdots \geq \alpha^{(M)}$, the centralized setting consists on a BC setting with N_T transmit antennas and CSIT $\hat{\mathbf{H}}$ equal to $\hat{\mathbf{H}}^{(1)}$.

7.1.4 Affine Approximation of the Achievable Rate

As in the previous chapter, we want to characterize the affine approximation of the rate presented in (2.11) in the D-CSIT setting. To wit, we aim to find the values DoF_d, $\mathcal{R}_\infty^d$, such that

$$R(\boldsymbol{\alpha}) = \mathrm{DoF}_d \log_2(P) - \mathcal{R}_\infty^d + o(1). \tag{7.10}$$

7.2 Centralized Zero-Forcing Precoding

7.2.1 Centralized Zero-Forcing Schemes (Under ideal CSIT sharing)

Since we base our analysis on the comparison with respect to the well-known centralized ZF precoding schemes, we first present the set of centralized precoders to which we restrict this work. In point of fact, the presented results hold for a general type of ZF precoders, such that we just introduce the requirements that these schemes have to satisfy. We denote the CSIT accuracy of the centralized case as $\alpha^{\star}$. Hence, the genie-aided centralized setting for the *sorted distributed setting* of (7.7) is represented by $\alpha^{\star} = \alpha^{(1)}$.

First, in order to distinguish when the precoding vectors refer to the genie-aided C-CSIT setting or to the D-CSIT setting, we denote the centralized coefficients as $\mathbf{v}_{i,k}$. This is in opposition to the $\mathbf{w}_{i,k}$ notation applied in (7.2) for the D-CSIT setting. In addition, $\mathbf{V}$, $\mathbf{v}_i$ and $\mathbf{T}_j^{\text{centr}}$ are defined as the centralized counterpart of $\mathbf{W}$, $\mathbf{w}_i$ and $\mathbf{T}_j$. Hence, the vectors $\mathbf{v}_i$ are computed from any ZF precoding algorithm satisfying

$$\begin{aligned}
&1) \quad \hat{\mathbf{h}}_i \mathbf{v}_\ell \;=\; 0, \quad \forall \ell \neq i \qquad \text{(Zero-Forcing condition)} && \text{(ZF1)}\\
&2) \quad \mathbb{E}\left[\|\mathbf{v}_{i,j,n}\|^{-1}\right] \;= \Theta(1) && \text{(ZF2)}\\
&3) \quad f_{\|\mathbf{v}_i\|} \leq f_{\|\mathbf{v}_i\|}^{\max} < \infty && \text{(ZF3)}
\end{aligned}$$

where $\hat{\mathbf{h}}_i$ is the centralized estimate of the channel vector of RX i. Note that (ZF1) is nothing but the condition that defines ZF schemes, (ZF2) implies that the probability of precoding with a vanishing power is negligible, and (ZF3) that the precoding vector has a bounded probability density function, i.e., that it is neither predetermined nor constant. Hereinafter, we assume that the centralized precoding scheme satisfies (ZF1), (ZF2), (ZF3). Furthermore, we assume that the precoding vectors and matrices can be expressed as a combination of summations, products, and generalized inverses of the channel estimate. As an example, we can use the typical choice of the projection of the matched filters onto the null spaces of the interfered users, i.e.,

$$\mathbf{v}_i = \lambda_i \frac{\mathbf{P}_{\mathbf{h}_{\bar{i}}^{\perp}} \hat{\mathbf{h}}_i^{\mathrm{H}}}{\|\mathbf{P}_{\mathbf{h}_{\bar{i}}^{\perp}} \hat{\mathbf{h}}_i^{\mathrm{H}}\|}, \qquad \mathbf{P}_{\mathbf{h}_{\bar{i}}^{\perp}} \triangleq \left(\mathbf{I} - \hat{\mathbf{H}}_{\bar{i}}^{\mathrm{H}} (\hat{\mathbf{H}}_{\bar{i}} \hat{\mathbf{H}}_{\bar{i}}^{\mathrm{H}})^{-1} \hat{\mathbf{H}}_{\bar{i}}\right), \tag{7.11}$$

where the matrix $\hat{\mathbf{H}}_{\bar{i}}$ stands for the full channel matrix with the i-th row removed, and λ_i is a parameter to satisfy the power constraint of (7.3). Note that, in order to avoid degenerate cases and increase the performance at low SNR, the inversion in (7.11) can be regularized. However, as conventional regularized schemes converge to their non-regularized counterpart at high-SNR, we omit any reference to regularized inverses.

We further model the precoding scheme as a function of the CSIT, such that $\mathcal{V}$ denotes the function applied to the channel estimate:

$$\mathcal{V} : \mathbb{C}^{K\times N_T} \to \mathbb{C}^{N_T\times K} \qquad \text{and} \qquad \mathbf{V} = \mathcal{V}\big(\hat{\mathbf{H}}\big). \tag{7.12}$$

7.2.2 ZF on Distributed CSIT Settings

It is known that centralized ZF schemes performance collapses under D-CSIT assumption [82, 100]. The main reason is that the interference cancellation achieved through (ZF1) is proportional to the worst accuracy among the TXs, $\alpha^{(M)}$ in the sorted case. Thus, the question is how to prevent the least accurate TXs from harming the transmission.

The first intuitive idea is to apply the strategy that we have developed in Part II. In that case, those inaccurate TXs do not make use of its instantaneous CSI to precode, transmitting with a fixed or known precoder based on statistical information. This solution achieves the centralized DoF under the less restrictive average power constraint $\mathbb{E}\big[\|\mathbf{T}_j\|^2\big] \leq 1$. However, under the –here assumed– instantaneous power constraint $\|\mathbf{T}_{j,n}\|^2 \leq 1$, the best known performance requires a power back-off such that the best informed TXs have enough power to realign the interference generated by the fixed-precoder TXs, with the flaw that this power back-off does not vanish at high SNR.

Another possible strategy is that the best informed TX attempts to estimate what is the CSIT at any TX ℓ ($\hat{\mathbf{H}}^{(\ell)}$) based on its own estimate $\hat{\mathbf{H}}^{(1)}$. However, the error variance of $\hat{\mathbf{H}}^{(\ell)}$ as a function of $\hat{\mathbf{H}}^{(1)}$ scales proportionally to $\alpha^{(\ell)}$. Thus, trying to correct the misalignment created by TX ℓ will not succeed.

The two problems described above are the two main barriers that limit the performance on D-CSIT settings. We present in the following several definitions that help to emphasize those two limitations.

Definition 7.1 (Consistency)**.** Consider two TXs, each one endowed with a different CSI, such that $\hat{\mathbf{H}}^{(j)}$ denotes the CSI at TX j. Suppose that they aim to compute the same function $f(x)$, each one on the basis of its own CSI. The computation is said to be *Consistent* if and only if $f(\hat{\mathbf{H}}^{(1)}) = f(\hat{\mathbf{H}}^{(2)})$. Otherwise, it is said to be *Inconsistent*.

Definition 7.2 (Power Outage)**.** Let TX j compute its precoding matrix $\mathbf{T}_j$ such that, for a given function f and value A, $\mathbf{T}_j$ fulfills that $f(\mathbf{T}_j) = A$. Then, TX j is said to be in *Power Outage* if and only if the computed precoder exceeds the instantaneous power constraint.

Definition 7.3 (Feasible Consistency)**.** Two TXs apply a *Feasible Consistent* precoder if the precoding coefficients are *Consistent* and there is not *Power Outage*.

One of our main contributions is to show that these limitations can be overcome by encouraging *consistency* among the different TXs, at the cost of reducing the accuracy of precoding at some TXs.

7.3 Rate Gap of the $M \times K$ D-CSIT Network MISO Setting

Our main contributions rely on a novel ZF-type precoding scheme coined *Consistent Decentralized ZF* (CD-ZF), which is presented in detail in Section 7.3.2. Briefly, this scheme is an adaptation to distributed scenarios of the aforementioned centralized ZF precoding, such that the precoding applied at each TX is different if the TX is the best informed one or not. Let $R(\boldsymbol{\alpha}_M)$ be the expected sum rate for our D-CSIT setting. Similarly, let $R^{\star}(\alpha^{(1)})$ be the expected sum rate achieved by a ZF scheme on the genie-aided C-CSIT setting as described in Section 7.2.1. Accordingly, the rate gap between those settings is defined as $\Delta \mathrm{R} \triangleq R^{\star}(\alpha^{(1)}) - R(\boldsymbol{\alpha}_M)$. We can now state our main result.

Theorem 7.1. *In the Network MISO setting with distributed CSIT, with $N_1 \geq K-1$ and $\alpha^{(M)} > 0$, the expected sum rate achieved by ZF-type schemes in the genie-aided Centralized CSIT setting is asymptotically achieved, i.e.,*

$$\lim_{P \to \infty} R^{\star}(\alpha^{(1)}) - R(\boldsymbol{\alpha}_M) = 0. \tag{7.13}$$

Proof. The proof builds on the proposed CD-ZF precoding scheme, which is presented in Section 7.3.2, and it is relegated to Section 7.4. ■

Corollary 7.1 (Rate-Offset under Distributed CSIT)**.** *It holds from Theorem 7.1 that the rate offset $\mathcal{R}_{\infty}^{d}$ –defined in (7.10)– of ZF with distributed CSIT is the same as for the genie-aided centralized setting, whose rate offset was shown in [132] to be constant with respect to Perfect CSIT ZF –and thus with respect to the capacity-achieving Dirty Paper Coding (DPC)– for the case of $\alpha^{\star} = 1$.*

Remarkably, Theorem 7.1 implies that it is possible to achieve not only the multiplexing gain but also the beamforming gain achieved by the centralized case with N_T antennas,

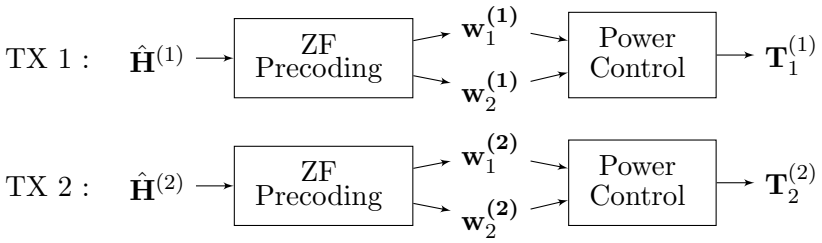
even if only N_1 antennas are endowed with the maximum accuracy. The constraint $N_1 \geq K-1$, i.e., that the TX with the most accurate CSI has a number of antennas at least equal to the number of interferable RXs, comes from the fact that if $N_1 < K-1$ the use of only ZF is not enough to achieve the DoF of the centralized setting [100], and thus $\lim_{P\to\infty} R^{\star}(\alpha^{(1)}) - R(\boldsymbol{\alpha}_M) = \infty$. We have shown in Chapter 4 that, even for the case with $N_1 = 1$, it is possible to reach the genie-aided DoF in some regimes. However, this is accomplished by means of an elaborated transmission scheme which comprises interference quantization and retransmission, superposition coding at the TXs and successive decoding at the RXs. Since in this work we focus in a simple ZF transmission, we restrict to the DoF-achieving regime $N_1 \geq K-1$.

It is known that the optimal DoF of the C-CSIT setting with accuracy $\alpha^{(1)}$ is equal to $1+(K-1)\alpha^{(1)}$, by means of superposition coding where a common message is broadcast and intended to be the decoded by all the RXs. It is noteworthy that, in the regime of interest, $N_1 \geq K-1$, the D-CSIT setting performance still converges asymptotically to the centralized performance even if superposition coding is applied. This comes from the fact that the instantaneous power applied converges to the one used in the C-CSIT setting, such that the broadcast common symbol can be sent with the same rate.

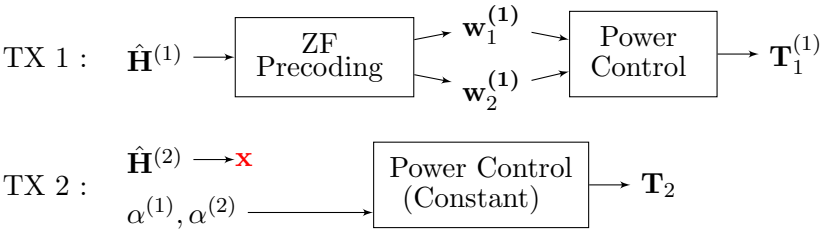
7.3.1 Achievability: A Broad View

Theorem 7.1 evidences that the issues associated which *feasible consistency* between the TXs –enunciated in Section 7.2.2– can be overtaken. Intuitively, the strategies mentioned in Section 7.2.2 are the extremes cases of consistency. Particularly, the naive ZF –whose block diagram is shown in Fig. 7.2a– represents the extreme in which consistency is not considered, whereas the AP-ZF –shown in Fig. 7.2b– embodies the extreme with perfect consistency but limited accuracy and possible Power Outage. The main question is if enforcing partial consistency might help and enhance the performance. Let the TX with best CSI attempt to estimate the decision of the other TXs. As previously mentioned, this incurs in an error proportional to the CSI accuracy of the worst TX. Moreover, as the variables are continuous, the probability of being consistent –estimating exactly what the other TX knows– is 0.

Nevertheless, we can build on the idea introduced in the previous chapter that discretizing the decision space of the TXs helps to enforce consistency as, at least, the probability of discerning what the other TX knows is strictly positive. The application of this idea is however not straightforward, since the only source of inconsistency in Chapter 6 was a single scalar power parameter, and no beamforming was possible. However, the main insight is still valid: By means of discretizing the decision space of the TXs that do not have the best CSI, we construct a probabilistic hierarchical setting,



(a) Block diagram of ZF applied naively in the 2x2 D-CSIT scenario (No Consistency).



(b) Block diagram of AP-ZF applied in the 2x2 D-CSIT scenario (Full Consistency).

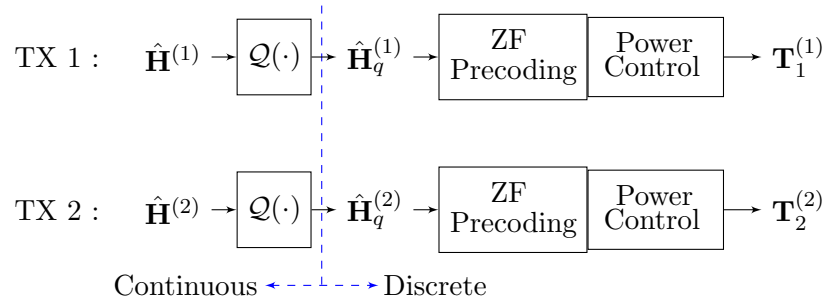
Figure 7.2 – Simple strategies for distributed precoding.

in which the best informed TX is able to estimate correctly the action taken by the other TXs with a certain probability. Interestingly, this discretization –or quantization– can be applied either to the available information –the channel matrix– or the output parameters –the precoder–. Both cases are illustrated in Fig. 7.3. This is due to the properties of linear systems and the asymptotic nature of our analysis. It is however clear that the performance at low-to-medium SNR can importantly differ for each of the cases.

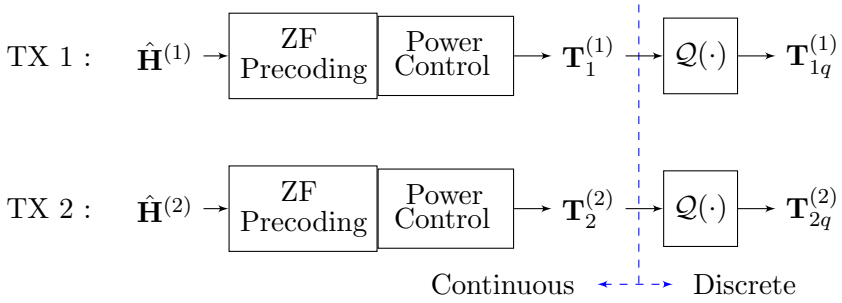
The key for attaining the surprising result of Theorem 7.1 is the proposed precoding scheme, whose rigorous description is presented in the following section. Yet, the benefit from achieving partial consistency is still not explicit and, indeed, it turns out that there exists a non-trivial compromise between consistency and accuracy that allows us to asymptotically close the rate gap. The proof of Theorem 7.1 relies on a simple idea: Let A be a set enclosing the *feasible consistent* cases in which the precoders transmit coordinately, and let A^c be its complementary event. Hence, the rate gap $\Delta \mathrm{R} \triangleq R^\star(\alpha^{(1)}) - R(\boldsymbol{\alpha}_M)$ can be expressed as

$$\Delta \mathrm{R} = \Delta \mathrm{R}_{|A} \Pr(A) + \Delta \mathrm{R}_{|A^c} \Pr(A^c). \tag{7.14}$$

If $\Delta \mathrm{R}_{|A} \to 0$ and $\Delta \mathrm{R}_{|A^c} \Pr(A^c) \to 0$, Theorem 7.1 is proven. It turns out that the transmission scheme has to be both consistent ($\Pr(A) \to 1$) and, for the consistent cases, it has to be accurate ($\Delta \mathrm{R}_{|A} \to 0$). This is rigorously shown in Section 7.4.



(a) Quantizing input (CSIT).



(b) Quantizing output (precoder).

Figure 7.3 – Two manners of discretizing decision space: At the input (information) or at the output (action).

7.3.2 Proposed Transmission Scheme: Consistent Distributed ZF

We first explain the Consistent Distributed ZF (CD-ZF) precoding scheme for the case in which the channel matrix (input) is quantized. The case when the precoding vector (output) is quantized is presented later. The proposed scheme presents an uneven structure, such that each TX applies a different strategy depending on who has higher accuracy. Furthermore, the proposed scheme computes independently the precoder for the symbols of different RXs, except for the final power normalization.

a) Quantizing the CSI ($\mathcal{Q}(\hat{\mathbf{H}}^{(j)})$)

The block diagram of this precoding scheme is depicted in Fig. 7.4. We split the description such that in the first place we explain the precoder at any TX not being the best informed one (TX 2 to TX M). Later, we present the precoder at the best-informed TX (TX 1).

The main limitation of the distributed precoding is not the error variance at the restricting TXs, but the impossibility at TX 1 (or the set of TXs with accuracy $\alpha^{(1)}$) of knowing what the other TXs are going to transmit. In order to overtake this problem, all the TXs but TX 1 quantize their estimation matrix with a known quantizer $\mathcal{Q}$. Hence,

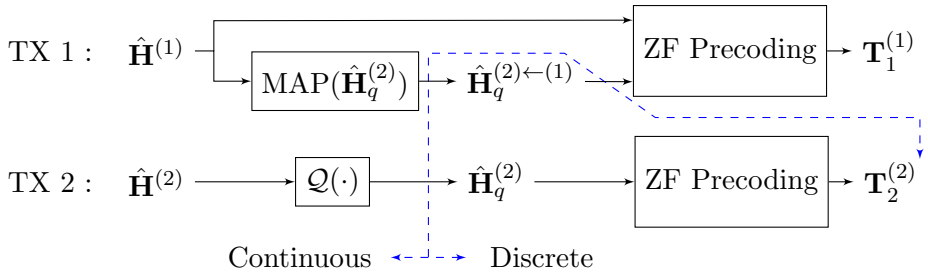


Figure 7.4 – Block diagram of CD-ZF applied in the 2x2 D-CSIT scenario.

for any $j > 1$, TX j does not use its CSIT $\hat{\mathbf{H}}^{(j)}$ to precode, but first pre-processes it. In other words, TX j applies

$$\hat{\mathbf{H}}_q^{(j)} = \mathcal{Q}(\hat{\mathbf{H}}^{(j)}). \tag{7.15}$$

The characteristics of the quantizer $\mathcal{Q}$ will be detailed later. Then, it applies naively a centralized ZF scheme as described in Section 7.2.1, based on $\hat{\mathbf{H}}_q^{(j)}$. Since the quantization transforms the continuous variable $\hat{\mathbf{H}}^{(j)}$ into a discrete one, it facilitates that the setting becomes a hierarchical setting, in which the information available at other TXs is estimated without explicit communication.

We focus now on the precoder at the most accurate TX, which attempts to correct the error of the previous TXs. Let TX 1 estimate $\hat{\mathbf{H}}_q^{(j)}$ based on its own information $\hat{\mathbf{H}}^{(1)}$, e.g. by computing the Maximum A Posteriori estimator (MAP) of $\hat{\mathbf{H}}_q^{(j)}$:

$$\hat{\mathbf{H}}_q^{(j)\leftarrow(1)} = \underset{\hat{\mathbf{H}}_q^{(j)} \in \mathcal{Q}(\mathbb{C}^{K \times N_T})}{\operatorname{argmax}} \Pr\left(\hat{\mathbf{H}}_q^{(j)} \mid \hat{\mathbf{H}}^{(1)}\right). \tag{7.16}$$

It is important to notice that the quantized value $\hat{\mathbf{H}}_q^{(j)}$ is not intended to be transmitted, but it is aimed at helping TX 1 to estimate the CSIT used at TX j, *without any explicit communication between them*. For sake of exposition, let us assume that TX 1 correctly estimates the CSIT at TX j, $\forall j \in \mathbb{N}_M$, such that

$$\hat{\mathbf{H}}_q^{(j)\leftarrow(1)} = \hat{\mathbf{H}}_q^{(j)}. \tag{7.17}$$

We will discuss about the probability that (7.17) happens in the following section. Hence, we obtain a *consistent* D-CSIT setting. The goal of TX 1 is to imitate the interference cancellation performance that the centralized ZF scheme would achieve if every other TX also owned $\hat{\mathbf{H}}^{(1)}$. In order to provide some insight, we first describe the 2×2 case.

2×2 case: Let $\bar{i} \triangleq i \pmod 2 + 1$. Mathematically, the goal is to have $|\hat{\mathbf{h}}_i^{(1)}\mathbf{w}_{\bar{i}}| = |\hat{\mathbf{h}}_i^{(1)}\mathbf{v}_{\bar{i}}|$, what can be rewritten as

$$|\hat{\mathbf{h}}_{i,1}^{(1)}\mathbf{w}_{\bar{i},1} + \hat{\mathbf{h}}_{i,2}^{(1)}\mathbf{w}_{\bar{i},2}| = |\hat{\mathbf{h}}_{i,1}^{(1)}\mathbf{v}_{\bar{i},1} + \hat{\mathbf{h}}_{i,2}^{(1)}\mathbf{v}_{\bar{i},2}|. \tag{7.18}$$

We remind that $\mathbf{w}$ stands for the distributed precoder whereas $\mathbf{v}$ stands for the centralized precoder. Under the assumption that TX 1 correctly estimates $\hat{\mathbf{H}}_q^{(2)}$, it knows $\mathbf{w}_{\bar{i},2}$. Then, TX 1 computes its precoder such that

$$\mathbf{w}_{\bar{i},1} = \mathbf{v}_{\bar{i},1} + \underbrace{(\hat{\mathbf{h}}_{i,1}^{(1)})^{\dagger}\hat{\mathbf{h}}_{i,2}^{(1)}(\mathbf{v}_{\bar{i},2} - \mathbf{w}_{\bar{i},2})}_{\boldsymbol{\phi}_{\bar{i}}}, \tag{7.19}$$

where $(\mathbf{x})^{\dagger}$ denotes the pseudo-inverse[1] of $\mathbf{x}$, which is known to have minimal Frobenius norm among all the generalized inverses [145]. The term $\boldsymbol{\phi}_i$ represents the correction term that TX 1 has to apply in order to compensate the error introduced by TX 2; note that (7.19) satisfies (7.18).

$M \times K$ case: The generalization follows directly but it needs one more step. Let the TX 1 have $N_1 \geq K - 1$ antennas. The goal is again to obtain the same interference cancellation as for the centralized precoder, such that, $\forall i \in \mathbb{N}_K$,

$$\sum_{\ell \in \mathbb{N}_K \setminus i} |\hat{\mathbf{h}}_i^{(1)}\mathbf{w}_\ell|^2 = \sum_{\ell \in \mathbb{N}_K \setminus i} |\hat{\mathbf{h}}_i^{(1)}\mathbf{v}_\ell|^2. \tag{7.20}$$

This can be attained if $\hat{\mathbf{h}}_i^{(1)}\mathbf{w}_\ell = \hat{\mathbf{h}}_i^{(1)}\mathbf{v}_\ell$, $\forall i, \ell \in \mathbb{N}_K, i \neq \ell$. Let us split the precoding and channel vectors in two parts: $\mathbf{v}_{\ell,1}$, $\mathbf{w}_{\ell,1}$ and $\hat{\mathbf{h}}_{i,1}^{(1)}$ denote the sub-vector corresponding to the antennas of TX 1, and $\mathbf{v}_{\ell,\bar{1}}$, $\mathbf{w}_{\ell,\bar{1}}$ and $\hat{\mathbf{h}}_{i,\bar{1}}^{(1)}$ represent the sub-vector corresponding to the antennas of TX 2 to TX M. The sub-matrices $\hat{\mathbf{H}}_{\bar{\ell},1}$ and $\hat{\mathbf{H}}_{\bar{\ell},\bar{1}}$ are defined in the same manner. Both sub-matrices are illustrated in Fig. (7.5) for ease of comprehension. We can expand the condition $\hat{\mathbf{h}}_i^{(1)}\mathbf{w}_\ell = \hat{\mathbf{h}}_i^{(1)}\mathbf{v}_\ell$ as a matrix equation in which $\mathbf{w}_{\ell,1}$ has to satisfy

$$\hat{\mathbf{H}}_{\bar{\ell},1}\mathbf{w}_{\ell,1} = \hat{\mathbf{H}}_{\bar{\ell},1}\mathbf{v}_{\ell,1} + \hat{\mathbf{H}}_{\bar{\ell},\bar{1}}(\mathbf{v}_{\ell,\bar{1}} - \mathbf{w}_{\ell,\bar{1}}), \tag{7.21}$$

where $N_{\bar{1}} = N_T - N_1$. The precoding vector at TX 1 is selected as

$$\mathbf{w}_{\ell,1} = \mathbf{v}_{\ell,1} + \underbrace{\hat{\mathbf{H}}_{\bar{\ell},1}^{\dagger}\hat{\mathbf{H}}_{\bar{\ell},\bar{1}}(\mathbf{v}_{\ell,\bar{1}} - \mathbf{w}_{\ell,\bar{1}})}_{\boldsymbol{\phi}_\ell}. \tag{7.22}$$

[1] $(\mathbf{x})^{\dagger}$ could also represent the regularized pseudo-inverse.

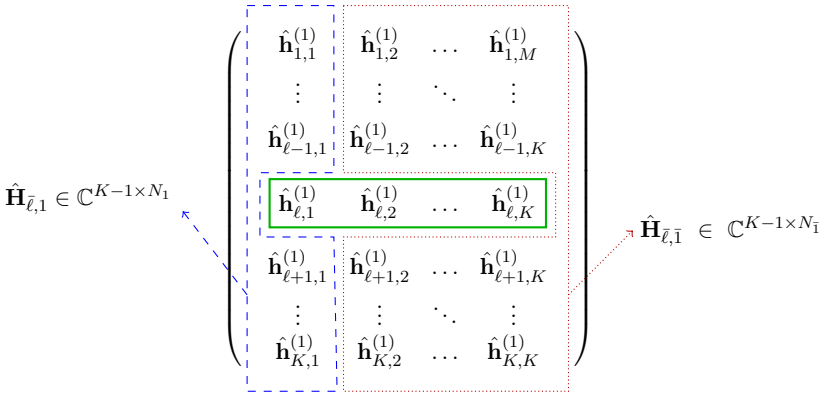


Figure 7.5 – Definition of the channel sub-matrices $\hat{\mathbf{H}}_{\bar{\ell},1}$ and $\hat{\mathbf{H}}_{\bar{\ell},\bar{1}}$.

The dimensionality of the linear system in (7.21) explains the limitation of having $N_1 \geq K-1$. In (7.21), it is only ensured that the interference received is the same as for the centralized setting. It is possible also to ensure that the receive signal $\hat{\mathbf{h}}_i^{(1)}\mathbf{w}_i$ is equal to the one of the centralized setting. However, it would require an extra antenna at TX 1 –since there is an extra equation in the equations system–, and it is not necessary as the received intended signal turns out to be statistically equivalent without enforcing $\hat{\mathbf{h}}_i^{(1)}\mathbf{w}_i$.

b) Quantizing the precoding coefficients

The scheme proposed above makes use of the quantization of the channel matrix at the TXs from TX 2 to TX M so as to allow TX 1 to know with a certain probability the CSIT at those TXs. Then, TX 1 can compute the precoder coefficients obtained at the other TXs. Another possibility is to apply the quantization at the end of the computation. In this case, TX j, $j \geq 2$, computes its precoder $\mathbf{T}_j^{(j)}$ based on its own non-quantized CSIT $\hat{\mathbf{H}}^{(j)}$. Then, it quantizes[2] its precoder to obtain

$$\mathbf{Q}_j^{(j)} \triangleq \mathcal{Q}(\mathbf{T}_j^{(j)}), \tag{7.23}$$

which will be the effective precoding vector applied. Then, TX 1 will perform the same algorithm as in the previous case where the channel matrix was quantized. The difference is that in this case it directly computes

$$\mathbf{Q}_j^{(j)\leftarrow(1)} = \underset{\mathbf{Q}_j^{(j)}\in\mathcal{Q}(\mathbb{C}^{N_j\times K})}{\operatorname{argmax}} \Pr\left(\mathbf{Q}_j^{(j)} \mid \hat{\mathbf{H}}^{(1)}\right), \tag{7.24}$$

[2]Note that we assume an instantaneous power constraint on $\mathbf{T}_j$. Hence, in order to prevent possible infringements of the constraint due to the quantization, if the quantization is done on the precoding vector it has to be applied by truncating (closer value towards 0).

such that TX 1 will know –with a certain probability– the precoding vector $\mathbf{w}_{i,j}$ applied at the TX j for RX i. The asymptotic results of this work hold for both cases (either quantizing the channel matrix or the precoding vectors). We focus hereinafter in the scheme that quantizes the channel matrix for sake of a better understanding, as the proof is less devious, and because the proof for the other case follows the same approach.

Feasibility and Consistency

In the previous description of the scheme, it has been assumed that TX 1 obtains a feasible consistent precoder, such that it correctly estimates the CSI at the other TXs and that the obtained precoding vector can be used for transmission. However, the transmission scheme will suffer from the two main issues described in Section 7.2.2: Power outage –since it has to satisfy that $\|\mathbf{T}_{1,n}\| \leq 1$, $\forall n \in \mathbb{N}_{N_1}$–, and *Consistency* –as the quantization of the CSIT at TX j allows TX 1 to obtain that CSIT only with a certain probability–. In the following we present some properties that will be instrumental to deal with those limitations. Let us focus first on the *consistency* problem. We introduce a set of quantizers that are essential in the proof of Theorem 7.1.

Definition 7.4 (Asymptotically Consistent Quantizers)**.** A quantizer $\mathcal{Q}$ is said to be *Asymptotically Consistent* if the probability of correct estimation of the MAP estimator at TX 1 satisfies

$$\Pr\left(\hat{\mathbf{H}}_q^{(j)\leftarrow(1)} = \hat{\mathbf{H}}_q^{(j)}\right) = o\left(\frac{1}{\log_2(P)}\right), \quad \forall j \in \mathbb{N}_M. \tag{P1}$$

Property (P1) implies that it is possible to induce that the probability of having *inconsistent* precoding among TXs vanishes faster than $1/\log_2(P)$. This fact implies that the rate impact of inconsistent precoding events vanishes asymptotically –as we will detail later–. Clearly, it remains to prove that there exists some quantizer $\mathcal{Q}_c$ satisfying (P1). Surprisingly, very simple quantizers as the one presented below satisfy it.

Lemma 7.1. *Let $\mathcal{Q}_u(\mathbf{X})$ be a scalar uniform quantizer with quantization step $q = \bar{P}^{-\alpha_q}$, where α_q is such that $\alpha^{(j)} > \alpha_q > 0$, $\forall j \in \mathbb{N}_M$. Then, $\mathcal{Q}_u$ is an* Asymptotically Consistent Quantizer *and* $\Pr\left(\hat{\mathbf{H}}_q^{(j)\leftarrow(1)} \neq \hat{\mathbf{H}}_q^{(j)}\right) = o\left(\frac{1}{\log_2(P)}\right)$*, for any $j \in \mathbb{N}_M$.*

Proof. The proof is relegated to Appendix H.5. ■

Note that $\mathcal{Q}_u$ is a scalar quantizer. Thus, the notation $\mathbf{A}_q = \mathcal{Q}_u(\mathbf{A})$, where $\mathbf{A}$ is a matrix, denotes –with an abuse of notation– that $\mathbf{A}_q$ is composed of the independent

scalar quantization of the real and imaginary part of each element in $\mathbf{A}$. Obviously, a better results would be obtained by applying vector quantization. However, as any quantizer satisfying (P1) is adequate for proving Theorem 7.1, we present $\mathcal{Q}_u$ for the sake of simplicity. The analysis at medium and low SNR –that requires an optimization on the quantizer used– is an interesting research topic that is delegated to future works.

Let us assume that the uniform quantizer of Lemma 7.1 is applied. Then, the naive precoder of TX 2 incurs in an error with respect to the centralized precoder that is proportional to the quantization step, as stated in the following lemma.

Lemma 7.2. *Let TX j, $2 \leq j \leq M$, quantize its CSIT with a scalar uniform quantizer with quantization step $q = \bar{P}^{-\alpha_q}$, $\alpha^{(M)} > \alpha_q > 0$. The naive precoder at TX j satisfies that*

$$\mathbb{E}\left[\|\mathbf{v}_{i,j} - \mathbf{w}_{i,j}\|\ \right] = \mathcal{O}\left(\bar{P}^{-\alpha_q}\right), \tag{7.25}$$

$$\mathbb{E}\left[\|\mathbf{v}_{i,j} - \mathbf{w}_{i,j}\|^2\right] = \mathcal{O}\left(P^{-\alpha_q}\right). \tag{7.26}$$

Proof. The proof is provided in Appendix H.6. ■

Lemma 7.2 is based on error propagation properties of linear systems. Thus, it is expected to hold for a broad set of noisy estimation models whose error variance scales as P^{-a} for any $a > 0$. For example, it holds for the quantized feedback model of [132], in which random vector quantization is assumed and the number of quantization bits scales with P, as shown in Chapter 6. Furthermore, Lemma (7.2) leads to the following corollary. The proof of this corollary is provided in Appendix H.2.

Corollary 7.2. *Let TX j, $2 \leq j \leq M$, quantize its CSIT with a scalar uniform quantizer with quantization step $q = \bar{P}^{-\alpha_q}$, $\alpha^{(M)} > \alpha_q > 0$. The global precoder satisfies that*

$$\mathbb{E}\left[\|\mathbf{v}_i - \mathbf{w}_i\|^2\right] = \mathcal{O}\left(P^{-\alpha_q}\right), \tag{7.27}$$

and hence $\mathbb{E}\left[\|\mathbf{v}_i - \mathbf{w}_i\|\right] = \mathcal{O}\left(\bar{P}^{-\alpha_q}\right)$.

Let us focus now on the probability of *power outage*. Let $\mathcal{P}_o$ denote the event of power outage. The precoder at the n-th antenna of TX j is given by $\mathbf{T}_{j,n}$. In that case,

$$\mathcal{P}_o \triangleq \left\{ \bigcup_{\substack{n \in \mathbb{N}_{N_j} \\ j \in \mathbb{N}_M}} \|\mathbf{T}_{j,n}\| > 1 \right\} \tag{7.28}$$

and the following lemma holds.

Lemma 7.3. *Let $\mu = 1 - \varepsilon$, where $\varepsilon > 0$, $\varepsilon \in \Theta(\bar{P}^{-\alpha_\mu})$ and $\alpha_\mu < \alpha_q$. Then,*

$$\Pr\left(\mathcal{P}_o\right) = o\left(\frac{1}{\log_2(P)}\right). \tag{7.29}$$

Proof. See Appendix H.1 ■

Similarly to property (P1), Lemma 7.3 implies that *power outage* events are negligible in terms of asymptotic rate. The only TX that may incur in power outage is TX 1, as the other TXs apply the naive centralized precoder and hence they will always satisfy the power constraint.

7.3.3 Hierarchical CSIT Setting

Theorem 7.1 shows that it is possible to attain asymptotically the rate of the centralized setting. Its performance at low-to-medium SNR is however limited by the probability of obtaining a *feasible consistent* precoder. This probability depends on the quantizer applied, the power back-off considered and the values of $\alpha^{(j)}$, and hence it is challenging to obtain. As shown in Section 7.3.2, the precoder is computed assuming a correct estimation of the CSI at the other TXs. Consequently, if the probability of *consistency* is low, the scheme does not perform properly and this probability decreases as the network size increases, since TX 1 needs to estimate correctly more parameters.

This limitation is inherent to the D-CSIT setting here assumed, in which each TX only knows its own CSI. However, there exists another practical setting with distributed CSI but in which there is more structure in the network CSI: The Hierarchical CSIT setting (H-CSIT). In this setting, introduced in Section 2.4.5, each TX is endowed with its own multi-user CSI $\mathbf{H}^{(j)}$, as in the D-CSIT setting, but it is also endowed with the CSI of the TXs having less accuracy than itself. Namely, in the sorted CSI scenario with $\alpha^{(1)} \geq \cdots \geq \alpha^{(M)}$, TX j knows $\{\hat{\mathbf{H}}^{(j)}, \hat{\mathbf{H}}^{(j+1)}, \ldots, \hat{\mathbf{H}}^{(M)}\}$.

This scenario, although it may seem less practical, arises in many heterogeneous networks. Fig. 7.1 depicts an example: Suppose that the RXs are all connected to the same main TX (e.g. TX 1), and the other TXs are remote radio-heads that receive a coarse version of the CSI by means of a wireless link from TX 1. In this use case, TX 1 will know the CSI available at each other TX. If the CSI sharing is done through dedicated links for each TX, each TX would receive CSI with accuracy proportional to its own link. If the CSI is broadcast, they may obtain an estimate with different accuracy if layered encoding [153] or analog feedback [154] is used.

Corollary 7.3. *Theorem 7.1 also holds in the Hierarchical CSIT setting and hence* $\lim_{P\to\infty} R^{\star}(\alpha^{(1)}) - R(\boldsymbol{\alpha}_M) = 0$.

Proof. The proof follows directly from the proof of Theorem 7.1 in Section 7.4. ■

In this setting, TX 1 already knows $\hat{\mathbf{H}}^{(j)}$, $\forall j \in \mathbb{N}_M$. Hence, the discretization of the variables at the other TXs is not needed, and the precoders are *consistent* with probability 1. Therefore, the only effect that may restrain TX 1 to achieve the centralized performance is the power outage. The performance at medium SNR will improve with respect to the general D-CSIT case, and moreover, it is not affected by the size of the network, as we will see in the numerical examples of Section 7.6.

7.3.4 Finite Precision CSIT Setting

Previously, it has been assumed that $\alpha^{(j)} > 0$ $\forall j \in \mathbb{N}_M$. Let us now assume that some TXs may have a finite precision CSIT, i.e., that $\exists j \in \mathbb{N}_M$ such that $\alpha^{(j)} = 0$. Since the TXs are sorted is descending order of the CSIT accuracy, consider w.l.o.g. that $\alpha^{(M)} = 0$.

In this case, the AP-ZF introduced in Chapter 4 is a more suitable scheme, as the TXs from TX 2 to TX M do not use its own CSIT for precoding. This scheme can be seen as a special case of CD-ZF, in which the TXs 2 to M use a known, pre-defined precoder independent of its CSI. For example, $\mathbf{w}_{i,j} = \frac{1}{\sqrt{K}}\mathbf{1}_{N_j \times 1}$ can be chosen for sake of simplicity, but $\mathbf{w}_{i,j}$ could also be computed e.g. from a pseudo-random sequence known at every TX following a Gaussian distribution or the distribution of the centralized precoder $\mathbf{v}_{i,j}$. Hence, TX 1 knows which is the precoder at the other TXs –similar as in the Hierarchical CSIT setting– and it can cancel out the interference generated by the other TXs. Particularly, the AP-ZF precoder for RX ℓ at TX 1 [82, 100] is given by

$$\mathbf{w}_{\ell,1} = -\hat{\mathbf{H}}_{\bar{\ell},1}^{\dagger}\hat{\mathbf{H}}_{\bar{\ell},\bar{1}}\mathbf{w}_{\ell,\bar{1}}. \tag{7.30}$$

Under this assumption of finite precision CSIT, Theorem 7.1 does not hold and the following corollary is obtained. The proof is relegated to Appendix H.7.

Corollary 7.4. *Let* $N_1 \geq K - 1$, *and* $\alpha^{(M)} = 0$. *In the Network MISO setting with distributed CSIT or hierarchical CSIT, with instantaneous power constraint for the precoder, the use of CD-ZF or AP-ZF leads to*

$$\lim_{P\to\infty} R^{\star}(\alpha^{(1)}) - R(\boldsymbol{\alpha}_M) = \infty. \tag{7.31}$$

In the D-CSIT setting, it is clear that CD-ZF requires $\alpha^{(j)} > 0$ for all $j \in \mathbb{N}_M$ because, if $\alpha^{(j)} = 0$, the probability of correct estimation at TX 1 does not increase as the SNR approaches infinity, i.e., $\Pr\left(\hat{\mathbf{H}}_q^{(j)\leftarrow(1)} = \hat{\mathbf{H}}_q^{(j)}\right) = \Theta(1)$. For that reason, the rate achieved with CD-ZF does not converge to the centralized performance. Moreover, for the Hierarchical CSIT scenario or for the use of AP-ZF precoding the main limitation consists in the fact that the probability of being in power outage at TX 1 does not vanish, and hence $\Pr(\mathcal{P}_o) = \Theta(1)$ because the distribution of the precoder at the other TXs remains the same for any P. This is explained more in detail in Appendix H.7.

Let us now relax the power normalization constraint from instantaneous precoder norm to the common average power constraint, i.e.,

$$\mathbb{E}\left[\|\mathbf{T}_{j,n}\|^2\right] \leq 1, \qquad \forall j \in \mathbb{N}_M,\ n \in \mathbb{N}_{N_j}. \tag{7.32}$$

Theorem 7.2. *Let $N_1 > K-1$, and $\alpha^{(j)} = 0$, $\forall j > 1$. In the Network MISO setting with distributed CSIT, under average power constraint, the rate gap is bounded, i.e., $\lim_{P\to\infty} R^\star(\alpha^{(1)}) - R(\boldsymbol{\alpha}_M) \leq c$, with $c < \infty$. Furthermore, it holds that*

$$\lim_{P\to\infty} R^\star(\alpha^{(1)}) - R(\boldsymbol{\alpha}_M) \leq K\Big(\log_2\left(\mathbb{E}\left[\|\hat{\mathbf{H}}_{K,1}^\dagger\|^2\right]\right) + \log_2\left(4K^2(K-1)\right)\Big), \tag{7.33}$$

where $\hat{\mathbf{H}}_{K,1}$ represents the first N_1 columns of the channel matrix estimate at TX 1.

Proof. The proof is relegated to Appendix H.8. ■

The bound in (7.33) is not tight, but it is useful at it shows that the rate gap with respect to the centralized scenario is bounded when some TXs are endowed with finite precision CSIT (or no CSIT at all). Note that the gap in (7.33) scales in K as $K\log_2(K)$. Interestingly, the gap between capacity-achieving Dirty-Paper Coding (DPC) and centralized ZF with perfect CSIT was shown in [131] to scale also as $K\log_2(K)$ when $N_T = K$.

7.4 Proof of Theorem 7.1

In order to prove Theorem 7.1, we need to demonstrate that the user rate gap $\Delta\mathrm{R}_i = R_i^\star(\alpha^{(1)}) - R_i(\boldsymbol{\alpha}_M)$ vanishes. Then, by symmetry, $\Delta\mathrm{R} = \sum_{i\in\mathbb{N}_K} \Delta\mathrm{R}_i$ will also vanish. The proof is divided in several steps: First, we show that both main issues previously exposed, *power outage* and *inconsistent* precoding, can be made negligible in terms of rate loss. Then, we prove that the rate gap vanishes by showing that in the distributed setting

both the interference received and total power received converge to their counterparts of the centralized setting.

7.4.1 Neglecting Non-Consistent Events

The proof of Theorem 7.1 builds on Lemma 7.1 and Lemma 7.3. Indeed, the proposed scheme will perform poorly if the precoder is not feasible consistent, as it assumes that it is. Howbeit, both Lemma 7.1 and Lemma 7.3 illustrate that those events can be made very unlikely. Let $\mathcal{H}_{\neq}$ denote the set *inconsistent* events, i.e., $\mathcal{H}_{\neq} = \left\{ \bigcup_{2\leq j\leq M} \hat{\mathbf{H}}_q^{(j)\leftarrow(1)} \neq \hat{\mathbf{H}}_q^{(j)} \right\}$. Hence, the probability of having *feasible consistent* precoding is $\Pr\left(\mathcal{P}_o^{\mathsf{c}} \cap \mathcal{H}_{\neq}^{\mathsf{c}}\right)$. By means of the law of total probability, we can split the expected sum rate R as

$$R = \Pr\left(\mathcal{P}_o \cup \mathcal{H}_{\neq}\right) R_{|\mathcal{P}_o \cup \mathcal{H}_{\neq}} + \Pr\left(\mathcal{P}_o^{\mathsf{c}} \cap \mathcal{H}_{\neq}^{\mathsf{c}}\right) R_{|\mathcal{P}_o^{\mathsf{c}} \cap \mathcal{H}_{\neq}^{\mathsf{c}}}. \tag{7.34}$$

Note that the expected sum rate achieved for any event is $\mathcal{O}(\log_2(P))$. Hence, it follows from Lemma 7.1 and Lemma 7.3 that

$$\Pr\left(\mathcal{P}_o \cup \mathcal{H}_{\neq}\right) R_{|\mathcal{P}_o \cup \mathcal{H}_{\neq}} = o\left(\frac{1}{\log_2(P)}\right) \mathcal{O}(\log(P)), \tag{7.35}$$

and consequently

$$R = R_{|\mathcal{P}_o^{\mathsf{c}} \cap \mathcal{H}_{\neq}^{\mathsf{c}}} + o(1). \tag{7.36}$$

Thus, in the remaining of the proof we assume w.l.o.g. that TX 1 knows $\hat{\mathbf{H}}_q^{(j)}$, $\forall j \in \mathbb{N}_M$, and that there is not power outage, as both cases become negligible at high SNR. This implies that the setting becomes hierarchical, as TX 1 correctly estimates the quantized CSIT of the other TXs. It is important to remark that this simplification is only possible because of the proposed scheme, in which we apply a correct power back-off and quantization step. Indeed, the surprising outcome is not (7.36) but the fact that $R_{|\mathcal{P}_o^{\mathsf{c}} \cap \mathcal{H}_{\neq}^{\mathsf{c}}}$ converges to the centralized setting rate. Furthermore, the achievable rate at low or medium SNR regimes might be increased by means of more complex schemes. For example, allowing several layers of quantization, such that TX $(j-1)$ tries to correct the interference generated by TX j, in a similar manner to the algorithm presented in [137].

7.4.2 Reformulating the Rate Gap

Note that we can rewrite the rate gap for RX i as

$$\Delta \mathrm{R}_i = \mathbb{E}\left[\log_2\left(1 + \frac{\frac{P}{K}\left|\mathbf{h}_i\mathbf{v}_i\right|^2}{1 + \frac{P}{K}\sum_{\ell\neq i}\left|\mathbf{h}_i\mathbf{v}_\ell\right|^2}\right)\right] - \mathbb{E}\left[\log_2\left(1 + \frac{\frac{P}{K}\left|\mu\mathbf{h}_i\mathbf{w}_i\right|^2}{1 + \frac{P}{K}\sum_{\ell\neq i}\left|\mu\mathbf{h}_i\mathbf{w}_\ell\right|^2}\right)\right]$$

$$= \mathbb{E}\left[\log_2\Big(\underbrace{\frac{1+\frac{P}{K}\sum_{\ell\in\mathbb{N}_K}|\mathbf{h}_i\mathbf{v}_\ell|^2}{1+\frac{P}{K}\sum_{\ell\in\mathbb{N}_K}|\mu\mathbf{h}_i\mathbf{w}_\ell|^2}}_{\mathcal{F}_\mathcal{D}}\Big)\right] + \mathbb{E}\left[\log_2\Big(\underbrace{\frac{1+\frac{P}{K}\sum_{\ell\neq i}|\mu\mathbf{h}_i\mathbf{w}_\ell|^2}{1+\frac{P}{K}\sum_{\ell\neq i}|\mathbf{h}_i\mathbf{v}_\ell|^2}}_{\mathcal{F}_\mathcal{I}}\Big)\right]. \quad (7.37)$$

This rewriting of $\Delta\mathrm{R}_i$ allows us to separate the ratio of received interference power ($\mathcal{F}_\mathcal{I}$) and the ratio of total received power ($\mathcal{F}_\mathcal{D}$). In the following, we will prove that $\lim_{P\to\infty}\Delta\mathrm{R}_i = 0$ by showing that $\lim_{P\to\infty}\mathbb{E}[\log_2(\mathcal{F}_i)] = 0$ for both $\mathcal{F}_\mathcal{D}$ and $\mathcal{F}_\mathcal{I}$. We start with $\mathcal{F}_\mathcal{I}$ for simplicity, and later we apply a similar argument to $\mathcal{F}_\mathcal{D}$.

7.4.3 Analysis of the Interference Ratio ($\mathcal{F}_\mathcal{I}$)

We prove the convergence by upper and lower bounding $\mathcal{F}_\mathcal{I}$ and then showing that both bounds converge to 0. We recall that we assume that TX 1 is able to transmit the desired precoding vector of (7.19) since the opposite case only yields an $o(1)$ rate contribution. Let us start with the upper-bound. Note that, since $\mu \leq 1$,

$$\mathbb{E}\left[\log_2\left(\frac{1+\frac{P}{K}\sum_{\ell\neq i}|\mu\mathbf{h}_i\mathbf{w}_\ell|^2}{1+\frac{P}{K}\sum_{\ell\neq i}|\mathbf{h}_i\mathbf{v}_\ell|^2}\right)\right] \leq \mathbb{E}\left[\log_2\Big(\underbrace{\frac{1+\frac{P}{K}\sum_{\ell\neq i}|\mathbf{h}_i\mathbf{w}_\ell|^2}{1+\frac{P}{K}\sum_{\ell\neq i}|\mathbf{h}_i\mathbf{v}_\ell|^2}}_{\mathcal{F}'_\mathcal{I}}\Big)\right], \quad (7.38)$$

where we have introduced the notation $\mathcal{F}'_\mathcal{I}$ for the sake of readability. Let η be a scalar $0 \leq \eta \leq 1$. We can split the expectation under the condition that the term $\mathcal{F}'_\mathcal{I}$ is smaller than $1+\eta$ or not. Therefore,

$$\begin{aligned}\mathbb{E}\left[\log_2\left(\mathcal{F}'_\mathcal{I}\right)\right] &= \Pr\left(\mathcal{F}'_\mathcal{I} < 1+\eta\right)\mathbb{E}_{\mathcal{F}'_\mathcal{I}<1+\eta}\left[\log_2\left(\mathcal{F}'_\mathcal{I}\right)\right] \\ &\quad + \Pr\left(\mathcal{F}'_\mathcal{I} \geq 1+\eta\right)\mathbb{E}_{\mathcal{F}'_\mathcal{I}\geq 1+\eta}\left[\log_2\left(\mathcal{F}'_\mathcal{I}\right)\right].\end{aligned} \quad (7.39)$$

Now we present the a useful lemma.

Lemma 7.4. *Let $\eta = \bar{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$ and ε arbitrarily small. Then,*

$$\Pr\left(\mathcal{F}'_\mathcal{I} \geq 1+\eta\right) = o\left(\frac{1}{\log_2(P)}\right) \quad \textit{and} \quad \Pr\left(\frac{1}{\mathcal{F}'_\mathcal{I}} \geq 1+\eta\right) = o\left(\frac{1}{\log_2(P)}\right). \quad (7.40)$$

Proof. The proof is relegated to Section 7.5.1. ■

Let $\eta = P^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$ and ε arbitrarily small. Then, (7.39) becomes

$$\begin{aligned}\mathbb{E}\left[\log_2\left(\mathcal{F}'_\mathcal{I}\right)\right] &\leq \mathbb{E}_{\mathcal{F}'_\mathcal{I}<1+\eta}\left[\log_2\left(\mathcal{F}'_\mathcal{I}\right)\right] + o\left(\frac{1}{\log_2(P)}\right)\mathbb{E}_{\mathcal{F}'_\mathcal{I}\geq 1+\eta}\left[\log_2\left(\mathcal{F}'_\mathcal{I}\right)\right] \\ &\leq \log_2(1+\eta) \;+\; o\left(1\right)\end{aligned} \quad (7.41)$$

since $\mathbb{E}_{\mathcal{F}'_{\mathcal{I}} \geq 1+\eta}\left[\log_2\left(\mathcal{F}'_{\mathcal{I}}\right)\right] = \mathcal{O}(\log_2(P))$. We now lower-bound the expectation. Note that

$$\mathbb{E}\left[\log_2\left(\mathcal{F}_{\mathcal{I}}\right)\right] \geq \log_2(\mu^2) + \mathbb{E}\left[\log_2\left(\mathcal{F}'_{\mathcal{I}}\right)\right]. \tag{7.42}$$

Furthermore, lower-bounding (7.42) is equivalent to upper-bound $\mathbb{E}\left[\log_2\left(\frac{1}{\mathcal{F}'_{\mathcal{I}}}\right)\right]$. By applying Lemma 7.4 and similar as in (7.41), we obtain that $\mathbb{E}\left[\log_2\left(\frac{1}{\mathcal{F}'_{\mathcal{I}}}\right)\right] \leq \log_2(1+\eta) \; + \; o\left(1\right)$ and hence

$$\mathbb{E}\left[\log_2\left(\mathcal{F}_{\mathcal{I}}\right)\right] \geq \log_2(\mu^2) - \log_2(1+\eta) \; + \; o\left(1\right). \tag{7.43}$$

Consequently, the term $\mathbb{E}\left[\log_2\left(\mathcal{F}_{\mathcal{I}}\right)\right]$ can be bounded as

$$\log_2(\mu^2) - \log_2(1+\eta) \; + \; o\left(1\right) \leq \mathbb{E}\left[\log_2\left(\mathcal{F}_{\mathcal{I}}\right)\right] \leq \log_2(1+\eta) \; + \; o\left(1\right). \tag{7.44}$$

Since $\lim_{P\to\infty} \mu = 1$, $\lim_{P\to\infty} \eta = 0$, and $\lim_{P\to\infty} o(1) = 0$, it follows that

$$\lim_{P\to\infty} \mathbb{E}\left[\log_2\left(\mathcal{F}_{\mathcal{I}}\right)\right] = 0. \tag{7.45}$$

7.4.4 Analysis of the Received Signal Ratio ($\mathcal{F}_{\mathcal{D}}$)

It remains to prove that the first expectation of (7.37) also converges to zero. As for $\mathcal{F}_{\mathcal{I}}$, we can write

$$\mathbb{E}[\log_2(\mathcal{F}_{\mathcal{D}})] \leq \log_2\left(\frac{1}{\mu^2}\right) + \mathbb{E}\left[\log_2\Big(\underbrace{\frac{1+\frac{P}{K}\sum_{\ell\in\mathbb{N}_K}|\mathbf{h}_i\mathbf{v}_\ell|^2}{1+\frac{P}{K}\sum_{\ell\in\mathbb{N}_K}|\mathbf{h}_i\mathbf{w}_\ell|^2}}_{\mathcal{F}'_{\mathcal{D}}}\Big)\right]. \tag{7.46}$$

Moreover, the equivalent to Lemma 7.4 also holds for $\mathcal{F}'_{\mathcal{D}}$.

Lemma 7.5. *Let $\eta = \bar{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$ and ε arbitrarily small. Then,*

$$\Pr\left(\mathcal{F}'_{\mathcal{D}} \geq 1+\eta\right) = o\left(\frac{1}{\log_2(P)}\right) \quad \textit{and} \quad \Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{D}}} \geq 1+\eta\right) = o\left(\frac{1}{\log_2(P)}\right). \tag{7.47}$$

Proof. The proof is relegated to Section 7.5.2. ■

Thus, applying the same step as in (7.41) we obtain that

$$\mathbb{E}\left[\log_2\left(\mathcal{F}_{\mathcal{D}'}\right)\right] \leq \log_2(1+\eta) + o(1). \tag{7.48}$$

We can lower-bound $\mathbb{E}[\log_2(\mathcal{F}_\mathcal{D})]$ similarly to obtain that

$$-\log_2(1+\eta) \;+\; o(1) \leq \mathbb{E}[\log_2(\mathcal{F}_\mathcal{D})] \leq \log_2\left(1/\mu^2\right) + \log_2(1+\eta) \;+\; o(1). \tag{7.49}$$

Since $\lim_{P\to\infty}\mu = 1$, $\lim_{P\to\infty}\eta = 0$, $\lim_{P\to\infty} o(1) = 0$, it follows that

$$\lim_{P\to\infty} \mathbb{E}[\log_2(\mathcal{F}_\mathcal{D})] = 0. \tag{7.50}$$

7.4.5 Merging Previous Sections

It follows that, since $\lim_{P\to\infty}\Delta\mathrm{R} = \lim_{P\to\infty}\sum_{i=1}^{K}\Delta\mathrm{R}_i$,

$$\begin{aligned}\lim_{P\to\infty}\Delta\mathrm{R} &= \lim_{P\to\infty} K(\mathbb{E}[\log_2(\mathcal{F}_\mathcal{D})] + \mathbb{E}[\log_2(\mathcal{F}_\mathcal{I})]) \\ &= 0,\end{aligned} \tag{7.51}$$

what concludes the proof of Theorem 7.1.

7.5 Proof of Lemma 7.4 and Lemma 7.5

In this section we prove Lemma 7.4 and Lemma 7.5, which are instrumental for the proof of the main results.

7.5.1 Proof of Lemma 7.4

We aim to prove that, for $\eta = \bar{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$ and ε arbitrarily small, it holds that

$$\Pr\left(\mathcal{F}'_\mathcal{I} \geq 1+\eta\right) = o\left(\frac{1}{\log_2(P)}\right), \tag{7.52}$$

$$\Pr\left(\frac{1}{\mathcal{F}'_\mathcal{I}} \geq 1+\eta\right) = o\left(\frac{1}{\log_2(P)}\right). \tag{7.53}$$

We start by noting that $\mathbf{h}_i$ can be written as $\mathbf{h}_i = \frac{1}{\breve{z}^{(j)}}(\hat{\mathbf{h}}_i^{(j)} - z^{(j)}\,\boldsymbol{\delta}_i^{(j)})$ from the definition of the estimate in (7.9). Let us introduce the notations $\breve{z}_{\text{inv}}^{(j)} = \frac{1}{\breve{z}^{(j)}}$ and $z_n^{(j)} = \frac{z^{(j)}}{\breve{z}^{(j)}}$. Hence, it follows that

$$\begin{aligned}|\mathbf{h}_i\mathbf{w}_\ell| &\overset{(a)}{=} |\breve{z}_{\text{inv}}^{(1)}\hat{\mathbf{h}}_i^{(1)}\mathbf{w}_\ell - z_n^{(1)}\boldsymbol{\delta}_i^{(1)}\mathbf{w}_\ell| \\ &\overset{(b)}{=} |\breve{z}_{\text{inv}}^{(1)}\hat{\mathbf{h}}_i^{(1)}\mathbf{v}_\ell - z_n^{(1)}\boldsymbol{\delta}_i^{(1)}\mathbf{w}_\ell + z_n^{(1)}\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell - z_n^{(1)}\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell| \\ &\overset{(c)}{=} |\mathbf{h}_i\mathbf{v}_\ell - z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell - \mathbf{v}_\ell)|,\end{aligned} \tag{7.54}$$

where (a) and (c) come from the D-CSIT model of (7.9) and (b) from the precoder definition in (7.18) since $\hat{\mathbf{h}}_i^{(1)}\mathbf{v}_\ell = \hat{\mathbf{h}}_i^{(1)}\mathbf{w}_\ell$. Hence, from the triangular inequality it follows

$$\frac{1+\frac{P}{K}\sum_{\ell\neq i}|\mathbf{h}_i\mathbf{w}_\ell|^2}{1+\frac{P}{K}\sum_{\ell\neq i}|\mathbf{h}_i\mathbf{v}_\ell|^2} \leq 1+\frac{\frac{P}{K}\sum_{\ell\neq i}\left(|z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|^2+2\frac{P}{K}|\mathbf{h}_i\mathbf{v}_\ell||z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|\right)}{1+\frac{P}{K}\sum_{\ell\neq i}|\mathbf{h}_i\mathbf{v}_\ell|^2}. \tag{7.55}$$

Let us recall that

$$\Pr\left(\sum_{k=1}^{K}A_k\geq c\right)\leq\sum_{k=1}^{K}\Pr\left(A_k\geq\frac{c}{K}\right). \tag{7.56}$$

From (7.55) and (7.56), it follows that

$$\begin{aligned}
&\Pr\left(\mathcal{F}_{\mathcal{I}}'\geq 1+\eta\right)\\
&\leq\sum_{\ell\neq i}\Pr\left(\frac{\frac{P}{K}\left(|z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|^2+2\frac{P}{K}|\mathbf{h}_i\mathbf{v}_\ell||z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|\right)}{1+\frac{P}{K}\sum_{\ell\neq i}|\mathbf{h}_i\mathbf{v}_\ell|^2}\geq\frac{\eta}{K}\right)\\
&\leq\sum_{\ell\neq i}\Pr\left(\frac{|z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|^2+2|\mathbf{h}_i\mathbf{v}_\ell||z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|}{|\mathbf{h}_i\mathbf{v}_\ell|^2}\geq\frac{\eta}{K}\right)\\
&\overset{(a)}{=}(K-1)\Pr\left(\frac{|z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|^2+2|\mathbf{h}_i\mathbf{v}_\ell||z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|}{|\mathbf{h}_i\mathbf{v}_\ell|^2}\geq\frac{\eta}{K}\right)\\
&\overset{(b)}{\leq}(K-1)\left(\Pr\left(\frac{|z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|^2}{|\mathbf{h}_i\mathbf{v}_\ell|^2}\geq\frac{\eta}{2K}\right)+\Pr\left(\frac{2|\mathbf{h}_i\mathbf{v}_\ell||z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|}{|\mathbf{h}_i\mathbf{v}_\ell|^2}\geq\frac{\eta}{2K}\right)\right)\\
&\overset{(c)}{\leq}2(K-1)\Pr\left(\frac{|z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|}{|\mathbf{h}_i\mathbf{v}_\ell|}\geq\frac{\eta}{4K}\right),
\end{aligned} \tag{7.57}$$

where (a) comes from symmetry, (b) from (7.56), and (c) because $\eta<1$. Let us introduce now a parameter $\gamma\in\mathbb{R}$. We can continue as

$$\begin{aligned}
\Pr\left(\frac{|z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|}{|\mathbf{h}_i\mathbf{v}_\ell|}\geq\frac{\eta}{4K}\right)&=\Pr\left(|\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|\geq\frac{\eta}{4K}|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell|\right)\\
&\leq\Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell|<\bar{P}^{-\gamma}\right)+\int_{|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell|\geq\bar{P}^{-\gamma}}\frac{\mathbb{E}\left[|\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|\right]}{\frac{\eta}{4K}y}f_{|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell|}(y)\,\mathrm{d}y
\end{aligned} \tag{7.58}$$

where the first equality comes from the fact that $|\mathbf{h}_i\mathbf{v}_\ell| = z_n^{(1)}|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell|$ and the last inequality from the Law of Total Probability and Markov's Inequality. $f_{|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell|}$ stands for the probability density function of $|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell|$. Let us focus first on the first term of (7.58), $\Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell|<\bar{P}^{-\gamma}\right)$, which satisfies the following proposition, whose proof is relegated to Appendix H.3.

Proposition 7.1. *Let $\gamma > 0$. Then,*

$$\Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell| < \bar{P}^{-\gamma}\right) = o\left(\frac{1}{\log_2(P)}\right). \tag{7.59}$$

On the other hand, the integral term of (7.58) can be rewritten as

$$\begin{aligned}\int_{|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell|\geq\bar{P}^{-\gamma}} &\mathbb{E}\Big[|\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|\Big]\,\frac{f_{|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell|}(y)}{\frac{\eta}{4K}y}\,\mathrm{d}y \\ &= \frac{4K}{\eta}\mathbb{E}\Big[|\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|\Big]\,\mathbb{E}_{||\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell|\geq\bar{P}^{-\gamma}}\left[\frac{1}{|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell|}\right] \\ &\leq \frac{4K}{\eta}\mathbb{E}\Big[|\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|\Big]\,\bar{P}^{\gamma}.\end{aligned} \tag{7.60}$$

Now we introduce a useful result, whose proof is relegated to Appendix H.4.

Proposition 7.2. *It follows that*

$$\mathbb{E}\Big[|\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|\Big] = \mathcal{O}(\bar{P}^{-\alpha_q}). \tag{7.61}$$

By introducing Proposition 7.1 and Proposition 7.2 in (7.58), it is straight to see that

$$\Pr\left(\mathcal{F}_{\mathcal{I}}' \geq 1+\eta\right) \leq o\left(\frac{1}{\log_2(P)}\right) + \frac{8K(K-1)}{\eta}\mathcal{O}(\bar{P}^{-\alpha_q})\bar{P}^{\gamma}. \tag{7.62}$$

Since $\eta = \bar{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$,

$$\Pr\left(\mathcal{F}_{\mathcal{I}}' \geq 1+\eta\right) \leq o\left(\frac{1}{\log_2(P)}\right) + \bar{P}^{\varepsilon}\mathcal{O}(\bar{P}^{-\alpha^{(2)}})\bar{P}^{\gamma}. \tag{7.63}$$

Let us select γ such that $\gamma > 0$ and $\varepsilon + \gamma - \alpha_q < 0$. Then, it follows that

$$\Pr\left(\mathcal{F}_{\mathcal{I}}' \geq 1+\eta\right) = o\left(\frac{1}{\log_2(P)}\right), \tag{7.64}$$

what concludes the proof of the first statement of Lemma 7.4. We prove in the following the second statement, i.e.,

$$\Pr\left(\frac{1}{\mathcal{F}_{\mathcal{I}}'} \geq 1+\eta\right) = o\left(\frac{1}{\log_2(P)}\right). \tag{7.65}$$

This is obtained by switching the vectors $\mathbf{v}_\ell$ and $\mathbf{w}_\ell$ and applying the same steps as in the proof of the first statement. To begin with, by following the steps in (7.57) we can easily obtain that

$$\begin{aligned}\Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{I}}} \geq 1+\eta\right) &= \Pr\left(\frac{1+\frac{P}{K}\sum_{\ell\neq i}|\mathbf{h}_i\mathbf{v}_\ell|^2}{1+\frac{P}{K}\sum_{\ell\neq i}|\mathbf{h}_i\mathbf{w}_\ell|^2} \geq 1+\eta\right) \\ &\leq 2(K-1)\Pr\left(\frac{|z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|}{|\mathbf{h}_i\mathbf{w}_\ell|} \geq \frac{\eta}{4K}\right).\end{aligned} \tag{7.66}$$

Furthermore, the final expression in (7.66) is equal to the one in (7.57) except from the fact that the denominator is $|\mathbf{h}_i\mathbf{w}_\ell|$ instead of $|\mathbf{h}_i\mathbf{v}_\ell|$. Hence, continuing as in (7.58)-(7.63), we obtain that

$$\Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{I}}} \geq 1+\eta\right) = o\left(\frac{1}{\log_2(P)}\right), \tag{7.67}$$

what concludes the proof of Lemma 7.4.

7.5.2 Proof of Lemma 7.5

We aim to prove that, $\forall\eta = \bar{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$ and ε arbitrarily small, it holds that

$$\Pr\left(\mathcal{F}'_{\mathcal{D}} \geq 1+\eta\right) = o\left(\frac{1}{\log_2(P)}\right), \tag{7.68}$$

$$\Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{D}}} \geq 1+\eta\right) = o\left(\frac{1}{\log_2(P)}\right). \tag{7.69}$$

Firstly, we focus on the first statement. Note that, applying similar steps to (7.54), it holds that

$$|\mathbf{h}_i\mathbf{v}_\ell|^2 \leq |\mathbf{h}_i\mathbf{w}_\ell|^2 + |z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|^2 + 2|\mathbf{h}_i\mathbf{w}_\ell||z_n^{(1)}\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell-\mathbf{v}_\ell)|, \tag{7.70}$$

$$|\mathbf{h}_i\mathbf{v}_i|^2 \leq |\mathbf{h}_i\mathbf{w}_i|^2 + |\mathbf{h}_i(\mathbf{w}_i-\mathbf{v}_i)|^2 + 2\,|\mathbf{h}_i\mathbf{w}_i|\,|\mathbf{h}_i(\mathbf{w}_i-\mathbf{v}_i)|\,. \tag{7.71}$$

Hence, following the steps applied in (7.55)-(7.57), we can write that

$$\begin{aligned}\Pr\left(\frac{1+\frac{P}{K}\sum_{\ell\in\mathbb{N}_K}|\mathbf{h}_i\mathbf{v}_\ell|^2}{1+\frac{P}{K}\sum_{\ell\in\mathbb{N}_K}|\mathbf{h}_i\mathbf{w}_\ell|^2} \geq 1+\eta\right) &\leq \Pr\left(\mathcal{D}_1+\mathcal{D}_2+\mathcal{D}_3+\mathcal{D}_4 \geq \eta\right) \\ &\leq \sum_{i=1}^{4}\Pr\left(\mathcal{D}_i \geq \frac{\eta}{4}\right),\end{aligned} \tag{7.72}$$

where we have introduced the notations

$$\mathcal{D}_1 \triangleq \sum_{\ell \neq i} \frac{|z_n^{(1)} \boldsymbol{\delta}_i^{(1)} (\mathbf{w}_\ell - \mathbf{v}_\ell)|^2}{|\mathbf{h}_i \mathbf{w}_i|^2} \tag{7.73}$$

$$\mathcal{D}_2 \triangleq \sum_{\ell \neq i} \frac{2\,|\mathbf{h}_i \mathbf{w}_\ell|\,|z_n^{(1)} \boldsymbol{\delta}_i^{(1)} (\mathbf{w}_i - \mathbf{v}_\ell)|}{|\mathbf{h}_i \mathbf{w}_i|^2} \tag{7.74}$$

$$\mathcal{D}_3 \triangleq \frac{|\mathbf{h}_i (\mathbf{w}_i - \mathbf{v}_i)|^2}{|\mathbf{h}_i \mathbf{w}_i|^2} \tag{7.75}$$

$$\mathcal{D}_4 \triangleq \frac{2\,|\mathbf{h}_i \mathbf{w}_i|\,|\mathbf{h}_i (\mathbf{w}_i - \mathbf{v}_i)|}{|\mathbf{h}_i \mathbf{w}_i|^2}. \tag{7.76}$$

The first inequality in (7.72) is obtained by applying (7.70)-(7.71) and eliminating the term $1 + \frac{P}{K} \sum_{\ell \neq i} |\mathbf{h}_i \mathbf{w}_\ell|^2$ from the denominator. From the analysis of $\mathcal{F}_{\mathcal{I}}$ in the previous section –see (7.58)– it follows easily that, if $\eta = \bar{P}^{-\varepsilon}$, with $\alpha_q > \varepsilon > 0$,

$$\begin{aligned} \Pr\left(\mathcal{D}_1 \geq \frac{\eta}{4}\right) &\leq \sum_{\ell \neq i} \Pr\left(z_n^{(1)} \frac{|\boldsymbol{\delta}_i^{(1)} (\mathbf{w}_\ell - \mathbf{v}_\ell)|^2}{|\mathbf{h}_i \mathbf{w}_i|^2} \geq \frac{\eta}{4K}\right) \\ &= o\left(\frac{1}{\log_2(P)}\right). \end{aligned} \tag{7.77}$$

Similarly,

$$\begin{aligned} \Pr\left(\mathcal{D}_2 \geq \frac{\eta}{4}\right) &= \sum_{\ell \neq i} \Pr\left(z_n^{(1)} \frac{2\,|\mathbf{h}_i \mathbf{w}_\ell|\,|\boldsymbol{\delta}_i^{(1)} (\mathbf{w}_i - \mathbf{v}_\ell)|}{|\mathbf{h}_i \mathbf{w}_i|^2} \geq \frac{\eta}{4K}\right) \\ &= o\left(\frac{1}{\log_2(P)}\right). \end{aligned} \tag{7.78}$$

For the two remaining terms, $\mathcal{D}_3$ and $\mathcal{D}_4$, note that

$$\begin{aligned} \Pr\left(\mathcal{D}_3 \geq \frac{\eta}{4}\right) + \Pr\left(\mathcal{D}_4 \geq \frac{\eta}{4}\right) &= \Pr\left(\frac{|\mathbf{h}_i (\mathbf{w}_i - \mathbf{v}_i)|^2}{|\mathbf{h}_i \mathbf{w}_i|^2} \geq \frac{\eta}{4}\right) \\ &\quad + \Pr\left(\frac{2\,|\mathbf{h}_i (\mathbf{w}_i - \mathbf{v}_i)|}{|\mathbf{h}_i \mathbf{w}_i|} \geq \frac{\eta}{4}\right) \\ &\leq 2 \Pr\left(\frac{|\mathbf{h}_i (\mathbf{w}_i - \mathbf{v}_i)|}{|\mathbf{h}_i \mathbf{w}_i|} \geq \frac{\eta}{16}\right) \\ &= 2 \Pr\left(\frac{|\tilde{\mathbf{h}}_i (\mathbf{w}_i - \mathbf{v}_i)|}{|\tilde{\mathbf{h}}_i \mathbf{w}_i|} \geq \frac{\eta}{16}\right). \end{aligned} \tag{7.79}$$

where $\tilde{\mathbf{h}} = \frac{\mathbf{h}}{\|\mathbf{h}\|}$ is unit-norm and it is distributed isotropically on the N_T-dimensional unit-sphere [132]. We can continue as in (7.58) to write

$$\Pr\left(\frac{|\tilde{\mathbf{h}}_i(\mathbf{w}_i-\mathbf{v}_i)|}{|\tilde{\mathbf{h}}_i\mathbf{w}_i|}\geq\frac{\eta}{16}\right)\leq\Pr\left(|\tilde{\mathbf{h}}_i\mathbf{w}_i|<\bar{P}^{-\gamma}\right)+\int_{|\tilde{\mathbf{h}}_i\mathbf{w}_i|\geq\bar{P}^{-\gamma}}\frac{\mathbb{E}[|\tilde{\mathbf{h}}_i(\mathbf{w}_i-\mathbf{v}_i)|]}{\frac{\eta}{16}y}f_{|\tilde{\mathbf{h}}_i\mathbf{w}_i|}(y)\,\mathrm{d}y$$
$$\leq\mathcal{O}(\bar{P}^{-\gamma})+16\bar{P}^{\varepsilon}\mathbb{E}[|\tilde{\mathbf{h}}_i(\mathbf{w}_i-\mathbf{v}_i)|]\,\bar{P}^{\gamma}. \tag{7.80}$$

The fact that $\|\tilde{\mathbf{h}}_i\|=1$ implies that $\mathbb{E}[|\tilde{\mathbf{h}}_i(\mathbf{w}_i-\mathbf{v}_i)|]\leq\mathbb{E}[\|\mathbf{w}_i-\mathbf{v}_i\|]$. Moreover, Corollary 7.2 states that $\mathbb{E}[\|\mathbf{w}_i-\mathbf{v}_i\|]=\mathcal{O}(\bar{P}^{-\alpha_q})$. Consequently, by selecting γ such that $\gamma>0$ and $\varepsilon+\gamma-\alpha_q<0$, it follows from (7.80) that

$$\begin{aligned}\Pr\left(\frac{|\tilde{\mathbf{h}}_i(\mathbf{w}_i-\mathbf{v}_i)|}{|\tilde{\mathbf{h}}_i\mathbf{w}_i|}\geq\frac{\eta}{16}\right)&\leq\mathcal{O}(\bar{P}^{-\gamma})+\bar{P}^{\varepsilon}\mathcal{O}(\bar{P}^{-\alpha_q})\bar{P}^{\gamma}\\&=o\left(\frac{1}{\log_2(P)}\right).\end{aligned} \tag{7.81}$$

We can introduce the result of (7.81) into (7.79) to obtain from (7.72) that

$$\Pr\left(\mathcal{F}'_{\mathcal{D}}\geq1+\eta\right)=o\left(\frac{1}{\log_2(P)}\right). \tag{7.82}$$

It would remain to prove that $\Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{D}}}\geq1+\eta\right)=o\left(\frac{1}{\log_2(P)}\right)$. To do so, we just need to apply the same previous steps, in which $\mathbf{w}$ and $\mathbf{v}$ are interchanged. Following those steps and following similar argument as in the proof for $\Pr\left(\frac{1}{\mathcal{F}'_{\mathcal{I}}}\right)$, it is obtained directly. For this reason, and for sake of concision, we omit the derivation.

7.6 Numerical Results

In this section we provide some performance analysis for the previous asymptotic results. First, we consider a scenario in which the most-informed TX has a CSI accuracy scaling parameter $\alpha^{(1)}=1$ for the whole channel matrix, and the rest of TXs have a CSI accuracy scaling parameter $\alpha^{(j)}=0.6$, for any $j>1$. Intuitively, this configuration can model a setting in which a main TX receives a quantized CSI feedback from all the RXs, and then it shares a compressed version of the CSI to the other auxiliary transmit antennas. We present the performance of several schemes:

- The ideal C-CSIT setting, in which all the TXs are endowed with the CSI of TX 1.
- The CD-ZF scheme with Hierarchical CSIT –TX 1 knows the other TXs' CSI–.
- The CD-ZF scheme when the CSIT is non hierarchical –general D-CSIT setting–.

- The AP-ZF scheme when the CSIT is non hierarchical –general D-CSIT setting–.
- The Naive ZF scheme when the CSIT is non hierarchical –general D-CSIT setting–.
- The performance of transmitting only from TX 1 and turning off the other TXs.

We analyze different network configurations. In Fig. 7.6, we show the rate performance for a setting with 2 single-antenna TXs and 2 RXs with the assumption of instantaneous power constraint for the precoder, whereas Fig. 7.7 illustrates a setting with $M = 4$ TXs, $N_1 = 3$ and $N_2 = N_3 = N_4 = 1$ transmit antennas, and $K = 3$ RXs with average power constraint. The unit of the vertical axis is bits/Hz/s at all the figures. Several insights emerge from the figures.

First, we observe how the proposed CD-ZF scheme performs almost as good as the ideal perfect-sharing C-CSIT setting for the H-CSIT configuration. This fact holds for any network size, yet considering that $N_1 \geq K - 1$. Besides this, the CD-ZF scheme is shown to tend towards the centralized rate also for the general D-CSIT setting, where the CSI at other TXs is not available at TX 1. However, we can see how the convergence is slow and at practical SNR regimes the CD-ZF scheme outperforms the single-TX transmission or the Naive ZF only in a slight manner. This is an aftermath of the scheme definition. Indeed, the CD-ZF scheme performs in an almost optimal manner if TX 1 correctly estimate the CSI at the other TXs; however, the probability of correct estimation increases slowly. Thus, the performance at medium SNR is limited. It is important to note that the CD-ZF scheme here presented is not optimized, as our objective was to show the asymptotic behavior. For example, we assume a scalar quantizer that independently quantizes every real and imaginary part of each channel coefficient. Considerably higher probabilities of consistency would be obtained if the quantization phase is optimized, e.g. by using vector quantization. Nevertheless, the aforementioned points show how important is to provide the CSI with structure (or hierarchy), as it has been proven indispensable to boost the performance. Moreover, this CSI structure is sometimes given by the network configuration, such that it does not imply an extra aspect to develop. Another point to be considered is that CD-ZF allows to obtain centralized performance with one informed antenna less than the single-TX transmission. This consideration can be seen in Fig. 7.6, as the single-TX transmission does not achieve even the centralized DoF.

In Fig. 7.8, we consider a different CSI configuration. In this case, $\alpha^{(2)} = 0$, i.e., the CSI accuracy at TX 2 does not scale with the SNR. We consider a setting with $M = 2$ TXs, $N_1 = 3$ and $N_2 = 1$ transmit antennas, and $K = 4$ RXs, with average power constraint. We do not plot the CD-ZF scheme in this configuration because the performance collapses, as discussed in the previous sections. This saturation is due to the fact that the probability of correct estimation of the other TXs' CSI does

not increase proportionally to the SNR. As previously mentioned, in this setting the single-TX transmission does not achieve the centralized DoF because it can only use a single antenna. We observe how the AP-ZF scheme achieves the centralized DoF, and moreover it attains a rate within a constant gap with respect to the centralized rate, what proves Theorem 7.2. Naive ZF, for its part, suffers from its dependency to the worst-TX accuracy.

7.7 Conclusions

We have presented an achievable scheme for the D-CSIT setting that attains the same asymptotic rate as the Zero-Forcing-type schemes in the centralized setting where every TX is endowed with the best estimate among all the TXs. This interesting result reveals that the performance degradation generated from the CSI mismatches between TXs can be asymptotically overcome by a properly designed precoding scheme which is aware of the distributed nature of the setting. Furthermore, it has been shown that the quantization of the information available at certain nodes is helpful as it facilitates the consistency of the decision at all the transmitters. This last result could be applied to a broad set of distributed problems, in which the trade-off between global consistency and local accuracy has not been deeply analyzed yet.

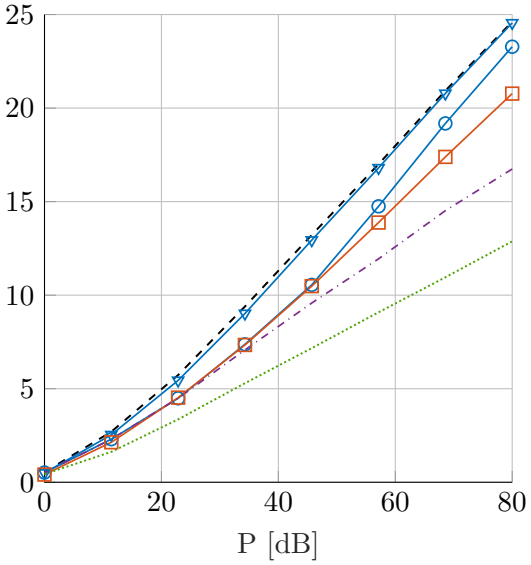


Figure 7.6 – Setting with 2 single-antenna TXs and 2 RXs with instantaneous power constraint.

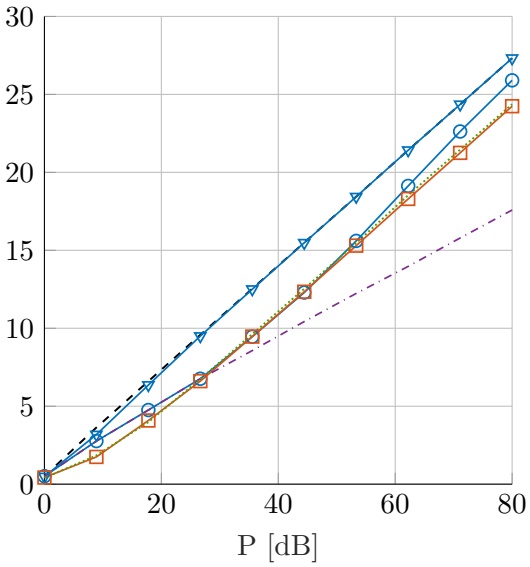
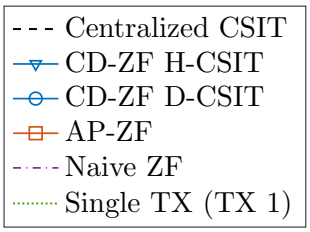


Figure 7.7 – Setting with $M = 4$ TXs, $N_1 = 3$ and $N_2 = N_3 = N_4 = 1$ transmit antennas, and $K = 3$ RXs with average power constraint.



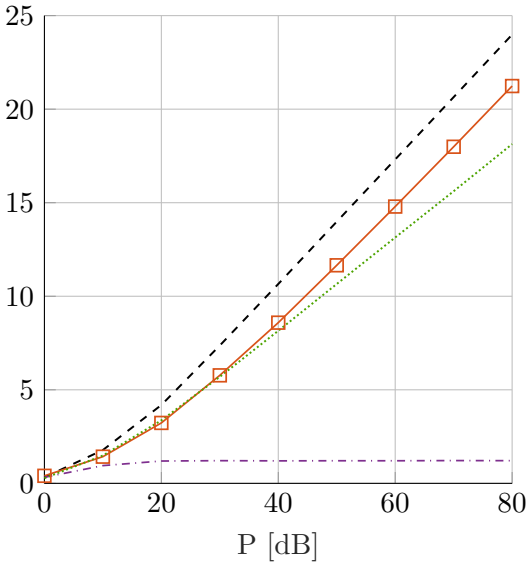


Figure 7.8 – Setting with $M = 2$ TXs, $N_1 = 3$ and $N_2 = 1$ transmit antennas, and $K = 4$ RXs with average power constraint and $\alpha^{(2)} = 0$.

Chapter 8

Conclusions and Perspectives

In this dissertation, we have focused on characterizing the high SNR regime of the Distributed Network MIMO, in which a set of TXs jointly serve a set of users with the particularity that the information available at each TX may be different or it can have a different accuracy. We have looked at this problem from different perspectives. First, we have analyzed the Generalized Degrees of Freedom metric. We have shown that the distributed setting attains the GDoF performance of the ideal centralized setting in which the TXs can perfectly share their CSIT. The case with 2 TXs and 2 RXs is of particular interest, as it turns out that it does not matter neither which TX has the best estimation for each link nor the path-loss configuration of the setting. Thus, in this particular configuration, the performance is robust to decentralization. For the general $K \times K$ setting, we have extended the previous result to a certain regime of CSI accuracy. It is intuitive that the distributed setting can not attain the ideal centralized performance for any possible case. However, we show that for any value of K, there exists a certain regime of CSIT accuracies for which the distributed setting achieves the ideal DoF performance.

Besides the GDoF analysis, we were interested in understanding to what extend the previous results can be translated to achievable rate, since the GDoF metric only provides information about the scaling. For that purpose, we have assumed a simple zero-forcing transmission and analyzed the rate gap between a transmission in our distributed setting and a transmission in an ideal centralized setting in which the best channel estimate is perfectly shared among the TXs. Following the same approach as in the GDoF analysis, we have first considered the 2×2 setting, and later we have extended the analysis to the $M \times K$ setting with multi-antenna TXs. Surprisingly, a similar conclusion as for the GDoF

analysis is obtained in the rate analysis of the 2×2 case: For any CSIT configuration, the distributed setting *asymptotically* attains the ideal centralized rate. We have further extended this result to the general multi-node setting, showing that for the cases in which the distributed setting attains the centralized DoF, it also attains –asymptotically– the centralized rate for a zero-forcing transmission.

Nonetheless, the manner of achieving this asymptotic result is interesting on its own. It is noteworthy that such optimal performance is obtained by reducing the accuracy of the precoder. In particular, the main insight derived from the proposed achievable scheme is that there exists an implicit compromise between the local accuracy at each TX and the consistency between TXs. Thus, reducing the precision at certain nodes –normally at those whose accuracy is not the best– allows us to improve the consistency of the decision taken by all the TXs, since the best informed TXs are able to estimate those reduced-precision parameters, and thereby correct the interference generated. The asymptotic convergence implies that the distributed setting attains not only the same multiplexing gain but also the same beamforming gain. In other words, for a setting with K single-antenna RXs and N_T transmit antennas, we can achieve the beamforming gain of the N_T antennas with only $K-1$ well-informed antennas.

In conclusion, we have shown how cooperation gains are less sensitive to CSIT impairments that what it was usually assumed. The key insight is that we have to develop schemes that are aware and reactive to the CSIT allocation, since common schemes induce a significant shrinking of performance. Although we are able to compensate the decentralization of the information with suitable algorithms in some cases, reaching the centralized performance is impossible for many distributed settings. A valuable conclusion of this dissertation is the idea that adding structure to the CSIT configuration boosts the achievable performance. We have observed this perception in the Hierarchical CSIT setting. Hence, providing some TXs with the limited CSI available at other TXs can boost the performance, specially at low-to-medium SNR. Motivated by this behavior, an interesting analysis to be done is to study how this reduction of performance scales when confronting it with the reduction of overhead that is implicit in the distributed settings. Hence, a reduced performance can be advantageous with respect to the ideal centralized one if the later implies an unfeasible quantity of information to be shared. The approaches developed to attain the aforementioned results are thought to be useful in many diverse decentralized settings or team decision problems. Indeed, the trade-off between consistency and accuracy is an inherent compromise in team decision problems.

The outcomes of this thesis illustrate how coarse limiting metrics as DoF are instrumental in the understanding of complex networks, since the main intuitions provided by them help considerable to develop new approaches applicable in practical scenarios.

The aftermath of the carried out analysis is constrained by the considered assumptions. Indeed, the analysis is asymptotic, although the results has proven extensible to the low-to-medium SNR regime. Even if this aspect can limit the conclusions, it also provides interesting future research paths. In particular, it is necessary to understand how dependent are the exposed results to the assumption of perfect sharing of the user's information data. The scenario in which the user data is also distributed across the network is an appealing and challenging problem. Another interesting topic is to shift the perspective here considered –that of developing the best transmission for a given CSI allocation– towards the reverse analysis of optimizing the CSI allocation subject to a certain transmission strategy and feedback protocol. Furthermore, considering the CSIT sharing load as part of the performance metric is a subsequent step in the analysis of decentralized networks. In fact, in order to compare distributed settings with its centralized counterpart in a fair manner, it is necessary to consider two facets of the CSI sharing overhead: That it impacts the performance, but also that it is a constraint of infeasibility, as the centralized setting may be unachievable due to delay constraints. The analysis at low SNR of the scenario considered is also an interesting extension of the presented work.

To conclude, the future 5G-and-beyond wireless networks are expected to cope with very heterogeneous scenarios and previously unseen specifications, as a massive number of devices communicating at the same time, sometimes in a sporadic manner (IoT), networks with mobile devices moving at very high speed (V2X networks), or exceedingly demanding delay constraints (URLLC). These scenarios arise as consequence of the broadening of use cases, as novel paramount applications are envisioned, for example haptic communications, remote medical services, or industry automation. Thus, the analysis of other cooperative scenarios, as well as non-cooperative ones where the inclusion of partial cooperation is possible, is a very appealing topic that can be analyzed in the future research.

Appendices

Appendix A

GDoF of the 2x2 setting: Achievability of Theorem 3.2

Let us start by denoting best CSIT accuracy across TXs as

$$\alpha_{i,k}^{\max} \triangleq \max\left(\alpha_{i,k}^{(1)}, \alpha_{i,k}^{(2)}\right). \tag{A.1}$$

We hence define α_i' as

$$\alpha_1' \triangleq \min(\alpha_{1,1}^{\max}, \alpha_{1,2}^{\max}) \qquad \text{and} \qquad \alpha_2' \triangleq \min(\alpha_{2,1}^{\max}, \alpha_{2,2}^{\max}). \tag{A.2}$$

Note that α_1' and α_2' are the only α-parameters that impact the GDoF expression of Theorem 3.2. We can assume w.l.o.g. that $\gamma_{1,1}$ is the strongest channel, i.e.,

$$\gamma_{1,1} \geq \max(\gamma_{1,2}, \gamma_{2,1}, \gamma_{2,2}). \tag{A.3}$$

Hereinafter, we prove that the GDoF expression of Theorem 3.2 is achievable by means of the proposed S-ZF. The proof is akin to the one for the MISO BC with centralized CSIT in [50], with the particularity that the interference cancellation is carried out by the S-ZF scheme. We split the demonstration in two different path-loss regimes in order to ease the readability. Those two path-loss regimes are

1. Regime $\gamma_{2,1} \leq \gamma_{2,2}$: The strongest link of each TX is geared to different RXs.
2. Regime $\gamma_{2,1} > \gamma_{2,2}$: In this regime TX 1 owns the strongest link towards both RXs.

We recall that, as explained in Section 3.5.2, the transmitted signal follows the structure

$$\mathbf{x} = \bar{P}_{\text{BC}}\mathbf{t}_{\text{BC}}s_{\text{BC}} + \bar{P}_{\text{ZF}}(\mathbf{t}_1^{\text{SZF}}s_{\text{ZF}\,1} + \mathbf{t}_2^{\text{SZF}}s_{\text{ZF}\,2}) + \bar{P}_{\phi}\mathbf{t}_{\phi}s_{\phi}, \tag{A.4}$$

where s_ϕ is a non-zero-forced symbol transmitted with power such that it is only received by the intended RX, if the path-loss topology allows for that, $s_{\mathrm{ZF}\,i}$ is intended to RX i and *canceled* at the other RX using S-ZF precoding, and s_{BC} is a broadcast symbol transmitted with full power, intended to be decoded at both RXs. In the following, we refer to the $s_{\mathrm{ZF}\,i}$ symbols as the S-ZF symbols.

Before presenting the transmission scheme, we present a lemma that will be useful throughout this appendix. Importantly, we can observe from Table 3.2 that the S-ZF precoding vector for RX i always satisfies that the difference of transmit power at each TX differs in $P^{|\gamma_{\bar{i},2}-\gamma_{\bar{i},1}|}$. This comes from the fact that, in order to implement the zero-forcing of the interference, the received signal at RX i from both TXs has to have the same power level; hence, each TX needs to transmit with a different power so as to compensate the different path-loss that they endure.

Lemma A.1. *Consider the 2-user D-CSIT Network MISO setting. Suppose a transmission with S-ZF precoding such that* $\mathbb{E}[\|\mathbf{t}_i^{\mathrm{SZF}}\|^2] = 1$. *Then, the intended signal received at RX i, satisfies*

$$\left|\mathbf{h}_i^{\mathrm{H}}\mathbf{t}_i^{\mathrm{SZF}}\right|^2 = \Theta_{\mathsf{p}}\left(P^{-1}\max(P^{\gamma_{i,1}-(\gamma_{\bar{i},1}-\gamma_{\bar{i},2})^+}, P^{\gamma_{i,2}-(\gamma_{\bar{i},2}-\gamma_{\bar{i},1})^+})\right), \tag{A.5}$$

whereas the interference at the same RX i satisfies

$$\left|\mathbf{h}_i^{\mathrm{H}}\mathbf{t}_{\bar{i}}^{\mathrm{SZF}}\right|^2 = \mathcal{O}_{\mathsf{p}}\left(P^{\min(\gamma_{i,1},\gamma_{i,2})-1-\alpha'_i}\right). \tag{A.6}$$

Proof. The scaling of the interference term in (A.6) follows directly from Lemma 3.2. We focus on the proof of (A.5). Note that $\mathbf{h}_i^{\mathrm{H}}$ and $\mathbf{t}_i^{\mathrm{SZF}}$ are independent. It follows that

$$|\mathrm{h}_{i,1}^{\mathrm{H}}\mathrm{t}_{i,1}^{(1)}|^2 = \Theta_{\mathsf{p}}(P^{\gamma_{i,1}-1}P^{-(\gamma_{\bar{i},1}-\gamma_{\bar{i},2})^+}). \tag{A.7}$$

Similarly,

$$|\mathrm{h}_{i,2}^{\mathrm{H}}\mathrm{t}_{i,2}^{(2)}|^2 = \Theta_{\mathsf{p}}(P^{\gamma_{i,2}-1}P^{-(\gamma_{\bar{i},2}-\gamma_{\bar{i},1})^+}). \tag{A.8}$$

From the fact that $\Theta_{\mathsf{p}}(A) + \Theta_{\mathsf{p}}(B) = \Theta_{\mathsf{p}}(\max(A, B))$, we obtain (A.5). ■

As main insight, the terms $\pm(\gamma_{\bar{i},2} - \gamma_{\bar{i},1})$ in (A.5), as well as the term $\min(\gamma_{i,1}, \gamma_{i,2})$ in (A.6), come from the fact that the TX with greater channel strength towards the interfered RX reduces his power to match the power received from the other TX so as to be able to cancel the interference.

A.1 Case $\gamma_{2,1} \leq \gamma_{2,2}$

In that case, the sum GDoF expression of Theorem 3.2 reads as

$$\mathrm{GDoF}^{\mathrm{DCSIT}}(\boldsymbol{\alpha}) = \min\left(\gamma_{1,1} + (\gamma_{2,2} - \gamma_{1,2} + \alpha_1')^+,\ \gamma_{2,2} + \gamma_{1,1} - \gamma_{2,1} + \alpha_2'\right). \tag{A.9}$$

Note that the stronger link for RX 2 has a path-loss exponent of $\max(\gamma_{2,1},\ \gamma_{2,2}) = \gamma_{2,2}$. Therefore, any signal transmitted with less power than $P^{1-\gamma_{2,2}}$ lies on the noise floor at RX 2. We can transmit information to RX 1 with power $P^{1-\gamma_{2,2}}$ and rate $(\gamma_{1,1} - \gamma_{2,2})\log_2(P)$ bits without generating interference at RX 2.

Transmitted signal

Let us define $\rho \in [0, 1]$ such that the rate of the S-ZF symbols is $\rho \log_2(P)$ bits per transmission. Omitting the time indexes, the transmitted signal is given by (A.4), where

- $s_{\mathrm{BC}} \in \mathbb{C}$ is a common symbol of rate $(\gamma_{2,2} - \rho)\log_2(P)$ bits that is decoded at both RXs. The precoder $\mathbf{t}_{\mathrm{BC}}$ is the uniform multicast precoder $\mathbf{t}_{\mathrm{BC}} = [1,\ 1]^{\mathrm{T}}$ and $\bar{P}_{\mathrm{BC}}$ is given by

$$\bar{P}_{\mathrm{BC}} = \frac{\sqrt{P - 2P^{\rho+1-\gamma_{2,2}} - P^{1-\gamma_{2,2}}}}{\sqrt{2}}. \tag{A.10}$$

- $s_{\mathrm{ZF}\,i} \in \mathbb{C}$ is a S-ZF symbol intended to RX i of rate $\rho \log_2(P)$ bits, where

$$\rho = \min\left((\gamma_{2,2} - \gamma_{1,2} + \alpha_1')^+,\ \gamma_{2,2} - \gamma_{2,1} + \alpha_2'\right). \tag{A.11}$$

  The term $\mathbf{t}_i^{\mathrm{SZF}}$ is the normalized S-ZF precoder for RX i introduced in Section 3.3, such that $\mathbb{E}[\|\mathbf{t}_i^{\mathrm{SZF}}\|^2] = 1$. The transmission power is

$$\bar{P}_{\mathrm{ZF}} = \bar{P}^{\rho+1-\gamma_{2,2}}. \tag{A.12}$$

- $s_\phi \in \mathbb{C}$ is a symbol of rate $(\gamma_{1,1} - \gamma_{2,2})\log_2(P)$ bits that carries information intended by RX 1 and it does not generate interference at the other RX. $\mathbf{t}_\phi$ is the matched precoder with power transmission $\bar{P}_\phi = \bar{P}^{1-\gamma_{2,2}}$.

Received signal

The received signal at RX 1 is

$$\mathrm{y}_1 = \underbrace{\bar{P}_{\mathrm{BC}}\mathbf{h}_1^{\mathrm{H}}\mathbf{t}_{\mathrm{BC}}s_{\mathrm{BC}}}_{\Theta_{\mathsf{p}}(\bar{P}^{\gamma_{1,1}})} + \underbrace{\bar{P}_{\mathrm{ZF}}\mathbf{h}_1^{\mathrm{H}}\mathbf{t}_1^{\mathrm{SZF}}s_{\mathrm{ZF}\,1}}_{\Theta_{\mathsf{p}}(\bar{P}^{\rho+\gamma_{1,1}-\gamma_{2,2}})} + \underbrace{\bar{P}_\phi\mathbf{h}_1^{\mathrm{H}}\mathbf{t}_\phi s_\phi}_{\Theta_{\mathsf{p}}(\bar{P}^{\gamma_{1,1}-\gamma_{2,2}})} + \underbrace{\bar{P}_{\mathrm{ZF}}\mathbf{h}_1^{\mathrm{H}}\mathbf{t}_2^{\mathrm{SZF}}s_{\mathrm{ZF}\,2}}_{\mathcal{O}_{\mathsf{p}}(\bar{P}^0)}. \tag{A.13}$$

The different power scaling of each symbol can be obtained from Lemma A.1 and the path-loss topology. In particular, starting from Lemma A.1 we can write that

$$\begin{aligned} P_{\mathrm{ZF}} \left|\mathbf{h}_1^{\mathrm{H}} \mathbf{t}_1^{\mathrm{SZF}}\right|^2 &= \Theta_{\mathsf{p}} \left(P_{\mathrm{ZF}} P^{-1} \max(P^{\gamma_{1,1}-(\gamma_{2,1}-\gamma_{2,2})^+}, P^{\gamma_{1,2}-(\gamma_{2,2}-\gamma_{2,1})^+})\right) \\ &\overset{(a)}{=} \Theta_{\mathsf{p}} \left(P^{\rho+1-\gamma_{2,2}-1} \max(P^{\gamma_{1,1}}, P^{\gamma_{1,2}-(\gamma_{2,2}-\gamma_{2,1})})\right) \\ &= \Theta_{\mathsf{p}} \left(P^{\gamma_{1,1}-\gamma_{2,2}+\rho}\right), \end{aligned} \tag{A.14}$$

where (a) comes from the fact that $(\gamma_{2,1}-\gamma_{2,2})^+ = 0$ and $(\gamma_{2,2}-\gamma_{2,1})^+ = \gamma_{2,2}-\gamma_{2,1}$. Also due to Lemma A.1, the contribution of the interfering symbol $s_{\mathrm{ZF}\,2}$ lies on the noise floor thanks to the S-ZF precoder:

$$\begin{aligned} P_{\mathrm{ZF}} \left|\mathbf{h}_1^{\mathrm{H}} \mathbf{t}_2^{\mathrm{SZF}}\right|^2 &\overset{(a)}{=} \Theta_{\mathsf{p}} \left(P^{\rho+1-\gamma_{2,2}} P^{\gamma_{1,2}-1-\alpha_1'}\right) \\ &\overset{(b)}{=} \mathcal{O}_{\mathsf{p}}(P^0), \end{aligned} \tag{A.15}$$

where (a) comes from $P_{\mathrm{ZF}} = P^{\rho+1-\gamma_{2,2}}$ and $\gamma_{1,1} \geq \gamma_{1,2}$, whereas (b) comes from the definition of ρ in (A.11) since it holds that $\rho \leq \gamma_{2,2}-\gamma_{1,2}+\alpha_1'$. The received signal at RX 2 is analyzed in the same way. Hence

$$\mathrm{y}_2 = \underbrace{\bar{P}_{\mathrm{BC}} \mathbf{h}_2^{\mathrm{H}} \mathbf{t}_{\mathrm{BC}} s_{\mathrm{BC}}}_{\Theta_{\mathsf{p}}(\bar{P}^{\gamma_{2,2}})} + \underbrace{\bar{P}_{\mathrm{ZF}} \mathbf{h}_2^{\mathrm{H}} \mathbf{t}_2^{\mathrm{SZF}} s_{\mathrm{ZF}\,2}}_{\Theta_{\mathsf{p}}(\bar{P}^{\rho})} + \underbrace{\bar{P}_{\mathrm{ZF}} \mathbf{h}_2^{\mathrm{H}} \mathbf{t}_1^{\mathrm{SZF}} s_{\mathrm{ZF}\,1}}_{\mathcal{O}_{\mathsf{p}}(\bar{P}^0)} + \underbrace{\bar{P}_{\phi} \mathbf{h}_2^{\mathrm{H}} \mathbf{t}_{\phi} s_{\phi}}_{\Theta_{\mathsf{p}}(\bar{P}^0)}. \tag{A.16}$$

The power scaling can be derived following the same steps as in (A.14) and (A.15). Then,

$$\begin{aligned} P_{\mathrm{ZF}} \left|\mathbf{h}_2^{\mathrm{H}} \mathbf{t}_2^{\mathrm{SZF}}\right|^2 &= \Theta_{\mathsf{p}} \left(P^{\rho+1-\gamma_{2,2}-1} \max(P^{\gamma_{2,1}-(\gamma_{1,1}-\gamma_{1,2})}, P^{\gamma_{2,2}})\right) \\ &= \Theta_{\mathsf{p}} \left(P^{\rho}\right). \end{aligned} \tag{A.17}$$

The scaling for the interfering signal $s_{\mathrm{ZF}\,1}$ yields

$$\begin{aligned} \left|\mathbf{h}_2^{\mathrm{H}} \mathbf{t}_1^{\mathrm{SZF}}\right|^2 &= \Theta_{\mathsf{p}} \left(P^{\rho+1-\gamma_{2,2}-1} P^{\gamma_{2,1}-\alpha_2'}\right) \\ &= \mathcal{O}_{\mathsf{p}} \left(P^0\right), \end{aligned} \tag{A.18}$$

where (A.18) comes from the definition of ρ in (A.11) since $\rho \leq \gamma_{2,2}-\gamma_{2,1}+\alpha_2'$.

Decoding and Achievable GDoF

We can see in (A.13) that RX 1 receives the common symbol s_{BC} with a SNR scaling as $P^{\gamma_{2,2}-\rho}$, and therefore it can decode s_{BC} by treating $s_{\mathrm{ZF}\,1}$ and s_{ϕ} as noise. After decoding the common symbol and removing its contribution to the received signal, $s_{\mathrm{ZF}\,1}$ can be decoded by treating s_{ϕ} as noise, since the SNR scales as P^{ρ}. And finally, s_{ϕ} is decoded

after removing the symbol $s_{\mathrm{ZF}\,1}$ from the received signal. Likewise, for the received sigmal at RX 2, (A.16) shows that the SNR for s_{BC} scales with $P^{\gamma_{2,2}-\rho}$ if $s_{\mathrm{ZF}\,2}$ is treated as noise. After decoding s_{BC} and removing its contribution from the received signal, $s_{\mathrm{ZF}\,2}$ can be decoded as its SNR scales with P^{ρ}.

The symbols are sent with a rate that is proportional to the SNR scaling, hence they can be decoded with a vanishing error probability. We decode the common symbol s_{BC} with rate $(\gamma_{2,2}-\rho)\log_2(P)$ bits, $s_{\mathrm{ZF}\,1}$ and $s_{\mathrm{ZF}\,2}$ with rate $\rho\log_2(P)$ bits and s_{ϕ} with rate $(\gamma_{1,1}-\gamma_{2,2})\log_2(P)$ bits. That allows us to achieve a GDoF of

$$\begin{aligned}\mathrm{GDoF}^{\mathrm{DCSIT}}(\boldsymbol{\alpha}) &= (\gamma_{2,2}-\rho)+(\gamma_{1,1}-\gamma_{2,2})+2\rho \\ &= \rho+\gamma_{1,1} \\ &= \min(\gamma_{1,1}+(\gamma_{2,2}-\gamma_{1,2}+\alpha_1')^+, \gamma_{2,2}+\gamma_{1,1}-\gamma_{2,1}+\alpha_2').\end{aligned} \tag{A.19}$$

This corresponds to the GDoF of the C-CSIT setting –see Theorem 3.1– and hence we attain the upper-bound.

A.2 Case $\gamma_{2,2} < \gamma_{2,1}$

For the second regime, the sum GDoF expression given in Theorem 3.1 is

$$\begin{aligned}\mathrm{GDoF}^{\mathrm{DCSIT}}(\boldsymbol{\alpha}) = \min\Big(&\gamma_{1,1}+\max\left((\gamma_{2,2}-\gamma_{1,2}+\alpha_1')^+, (\gamma_{2,1}-\gamma_{1,1}+\alpha_1')^+\right), \\ &\gamma_{1,1}+(\gamma_{2,1}-\gamma_{1,1}+\gamma_{1,2}-\gamma_{2,2})^+ + \alpha_2'\Big).\end{aligned} \tag{A.20}$$

In a similar way as in the previous case, any signal sent with power below $P^{1-\gamma_{2,1}}$ lies on the noise floor for RX 2. Thus, we will transmit to RX 1 a non-interfering symbol with power $P^{1-\gamma_{2,1}}$ and rate $(\gamma_{1,1}-\gamma_{2,1})\log_2(P)$ bits.

Transmitted signal

Let $\rho \in [0,1]$ be defined as the rate-parameter for the S-ZF symbols, such that the rate is $\rho\log_2(P)$ bits. By omitting the time indexes, the transmitted signal is given by (A.4), with

- s_{BC} is a common symbol of rate $(\gamma_{2,2}-\rho)\log_2(P)$ bits that is decoded at both RXs. The precoder $\mathbf{t}_{\mathrm{BC}}$ is obtained as $\mathbf{t}_{\mathrm{BC}} = [1,\ 1]^{\mathrm{T}}$ and $\bar{P}_{\mathrm{BC}}$ is given by

$$\bar{P}_{\mathrm{BC}} = \frac{\sqrt{P - 2P^{\rho+1-\gamma_{2,1}+\min(\gamma_{1,1}-\gamma_{1,2},\ \gamma_{2,1}-\gamma_{2,2})} - P^{1-\gamma_{2,1}}}}{\sqrt{2}} \tag{A.21}$$

- $s_{\mathrm{ZF}\,i}$ is a S-ZF symbol intended to RX i of rate $\rho\log_2(P)$ bits, where ρ is given by

$$\rho = \min\Bigg(\max\Big((\gamma_{2,2} - \gamma_{1,2} + \alpha_1')^+,\ (\gamma_{2,1} - \gamma_{1,1} + \alpha_1')^+\Big), \\ \alpha_2' + (\gamma_{2,1} - \gamma_{1,1} + \gamma_{1,2} - \gamma_{2,2})^+\Bigg). \tag{A.22}$$

The term $\mathbf{t}_i^{\mathrm{SZF}}$ is the normalized S-ZF precoder for RX i introduced in Section 3.3, such that $\mathbb{E}[\|\mathbf{t}_i^{\mathrm{SZF}}\|^2] = 1$. The transmission power is

$$\bar{P}_{\mathrm{ZF}} = \bar{P}^{\rho+1-\gamma_{2,1}+\min(\gamma_{1,1}-\gamma_{1,2},\ \gamma_{2,1}-\gamma_{2,2})}. \tag{A.23}$$

- s_ϕ is a symbol of rate $(\gamma_{1,1} - \gamma_{2,2})\log_2(P)$ bits that carries information intended by RX 1, and it does not generate interference at the other RX. $\mathbf{t}_\phi$ is the matched precoder with transmission power $\bar{P}_\phi = \bar{P}^{1-\gamma_{2,1}}$.

Received signal

The received signal at RX 1 is

$$\mathrm{y}_1 = \underbrace{\bar{P}_{\mathrm{BC}}\mathbf{h}_1^{\mathrm{H}}\mathbf{t}_{\mathrm{BC}}s_{\mathrm{BC}}}_{\Theta_{\mathsf{p}}(\bar{P}^{\gamma_{1,1}})} + \underbrace{\bar{P}_{\mathrm{ZF}}\mathbf{h}_1^{\mathrm{H}}\mathbf{t}_1^{\mathrm{SZF}}s_{\mathrm{ZF}\,1}}_{\Theta_{\mathsf{p}}(\bar{P}^{\rho+\gamma_{1,1}-\gamma_{2,1}})} + \underbrace{\bar{P}_{\phi}\mathbf{h}_1^{\mathrm{H}}\mathbf{t}_{\phi}s_{\phi}}_{\Theta_{\mathsf{p}}(\bar{P}^{\gamma_{1,1}-\gamma_{2,1}})} + \underbrace{\bar{P}_{\mathrm{ZF}}\mathbf{h}_1^{\mathrm{H}}\mathbf{t}_2^{\mathrm{SZF}}s_{\mathrm{ZF}\,2}}_{\mathcal{O}_{\mathsf{p}}(\bar{P}^0)}. \tag{A.24}$$

The different power scaling of each symbol can be obtained from Lemma A.1 and the path-loss topology. In particular, starting from Lemma A.1 we can write that

$$\begin{aligned} P_{\mathrm{ZF}}\left|\mathbf{h}_1^{\mathrm{H}}\mathbf{t}_1^{\mathrm{SZF}}\right|^2 &= \Theta_{\mathsf{p}}\left(P_{\mathrm{ZF}}P^{-1}\max(P^{\gamma_{1,1}-(\gamma_{2,1}-\gamma_{2,2})^+}, P^{\gamma_{1,2}-(\gamma_{2,2}-\gamma_{2,1})^+})\right) \\ &\overset{(a)}{=} \Theta_{\mathsf{p}}\left(P^{\rho+1-\gamma_{2,1}+\min(\gamma_{1,1}-\gamma_{1,2},\gamma_{2,1}-\gamma_{2,2})-1}\max(P^{\gamma_{1,1}-(\gamma_{2,1}-\gamma_{2,2})}, P^{\gamma_{1,2}})\right) \\ &= \Theta_{\mathsf{p}}\left(P^{\rho-\gamma_{2,1}+\min(\gamma_{1,1}-\gamma_{1,2},\ \gamma_{2,1}-\gamma_{2,2})+\max(\gamma_{1,1}-\gamma_{2,1}+\gamma_{2,2},\gamma_{1,2})}\right) \\ &= \Theta_{\mathsf{p}}\left(P^{\gamma_{1,1}-\gamma_{2,1}+\rho}\right), \end{aligned} \tag{A.25}$$

where (a) comes from $(\gamma_{2,1} - \gamma_{2,2})^+ = \gamma_{2,1} - \gamma_{2,2}$ and $(\gamma_{2,2} - \gamma_{2,1})^+ = 0$. The contribution of the interfering symbol $s_{\mathrm{ZF}\,2}$ is obtained from the definition of ρ in (A.22) as

$$\begin{aligned} P_{\mathrm{ZF}}\left|\mathbf{h}_1^{\mathrm{H}}\mathbf{t}_2^{\mathrm{SZF}}\right|^2 &= \Theta_{\mathsf{p}}\left(P^{\rho+1-\gamma_{2,1}+\min(\gamma_{1,1}-\gamma_{1,2},\gamma_{2,1}-\gamma_{2,2})-1}P^{\gamma_{1,2}-\alpha_1'}\right) \\ &= \mathcal{O}_{\mathsf{p}}(P^0). \end{aligned} \tag{A.26}$$

Similarly, the received signal at RX 2 can be written as

$$\mathrm{y}_2 = \underbrace{\bar{P}_{\mathrm{BC}}\mathbf{h}_2^{\mathrm{H}}\mathbf{t}_{\mathrm{BC}}s_{\mathrm{BC}}}_{\Theta_{\mathsf{p}}(\bar{P}^{\gamma_{2,1}})} + \underbrace{\bar{P}_{\mathrm{ZF}}\mathbf{h}_2^{\mathrm{H}}\mathbf{t}_2^{\mathrm{SZF}}s_{\mathrm{ZF}\,2}}_{\Theta_{\mathsf{p}}(\bar{P}^{\rho})} + \underbrace{\bar{P}_{\mathrm{ZF}}\mathbf{h}_2^{\mathrm{H}}\mathbf{t}_1^{\mathrm{SZF}}s_{\mathrm{ZF}\,1}}_{\mathcal{O}_{\mathsf{p}}(\bar{P}^0)} + \underbrace{\bar{P}_{\phi}\mathbf{h}_2^{\mathrm{H}}\mathbf{t}_{\phi}s_{\phi}}_{\Theta_{\mathsf{p}}(\bar{P}^0)}. \tag{A.27}$$

The power scaling of the intended signal $s_{\mathrm{ZF}\,2}$ satisfies that

$$\begin{aligned} P_{\mathrm{ZF}} \left|\mathbf{h}_2^{\mathrm{H}} \mathbf{t}_2^{\mathrm{SZF}}\right|^2 &= \Theta_{\mathsf{p}} \left(P^{\rho+1-\gamma_{2,1}+\min(\gamma_{1,1}-\gamma_{1,2},\gamma_{2,1}-\gamma_{2,2})-1} \max(P^{\gamma_{2,1}-(\gamma_{1,1}-\gamma_{1,2})}, P^{\gamma_{2,2}})\right) \\ &= \Theta_{\mathsf{p}} \left(P^{\rho}\right). \end{aligned} \tag{A.28}$$

The power scaling of the interfering signal $s_{\mathrm{ZF}\,1}$ satisfies that

$$\begin{aligned} \left|\mathbf{h}_2^{\mathrm{H}} \mathbf{t}_1^{\mathrm{SZF}}\right|^2 &= \Theta_{\mathsf{p}} \left(P^{\rho+1-\gamma_{2,1}+\min(\gamma_{1,1}-\gamma_{1,2},\gamma_{2,1}-\gamma_{2,2})-1} P^{\gamma_{2,1}-\alpha_2'}\right) \\ &= \mathcal{O}_{\mathsf{p}} \left(P^{0}\right), \end{aligned} \tag{A.29}$$

where (A.29) follows from the definition of ρ in (A.22) since it holds that $\rho \leq \gamma_{2,1} - \min(\gamma_{1,1} - \gamma_{1,2}, \gamma_{2,1} - \gamma_{2,2}) - \gamma_{2,1} + \alpha_2'$.

Decoding and Achievable GDoF

The decoding is applied as for the first regime, with the only difference that the rate of each symbol is adapted to the topology here considered. Hence, we can decode each symbol with a rate proportional to the SNR scaling: The common symbol s_{BC} can be decoded with rate $(\gamma_{2,1} - \rho)\log_2(P)$ bits, $s_{\mathrm{ZF}\,1}$ and $s_{\mathrm{ZF}\,2}$ with rate $\rho \log_2(P)$ bits and s_ϕ with rate $(\gamma_{1,1} - \gamma_{2,1})\log_2(P)$ bits. That allows us to achieve a GDoF of

$$\begin{aligned} \mathrm{GDoF}^{\mathrm{DCSIT}}(\boldsymbol{\alpha}) &= (\gamma_{2,1} - \rho) + (\gamma_{1,1} - \gamma_{2,1}) + 2\rho \\ &= \gamma_{1,1} + \rho \\ &= \min\Big(\gamma_{1,1} + \max\left((\gamma_{2,2} - \gamma_{1,2} + \alpha_1')^+, (\gamma_{2,1} - \gamma_{1,1} + \alpha_1')^+\right), \\ &\qquad \gamma_{1,1} + (\gamma_{2,1} - \gamma_{1,1} + \gamma_{1,2} - \gamma_{2,2})^+ + \alpha_2'\Big). \end{aligned} \tag{A.30}$$

The expression in (A.30) corresponds to the GDoF of the Centralized CSIT setting –see Theorem 3.1– which concludes the proof. ■

Appendix B

Statistics of the average power constraint for the 2x2 setting

In this appendix, we analyze the normalization constant λ of the Sliced ZF scheme introduced in Chapter 3. We recall that λ_i is defined as

$$\lambda_i \triangleq \frac{1}{\sqrt{\mathbb{E}\left[\|\mathbf{w}_{i,1}^{(1)}\|^2 + \|\mathbf{w}_{i,2}^{(2)}\|^2\right]}}. \tag{B.1}$$

First of all, we characterize the probability density function and the expected value of the regularized inverse. Then, based on those results, we prove Proposition (3.3).

B.1 Generic Regularized Inverse Term

For sake of completeness, we consider a generic regularized inverse for a Gaussian variable with arbitrary variance, and then we particularize for the case of interest in the GDoF setting. Let us consider a random variable $\mathrm{h} \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$, and an arbitrary regularization constant $\eta > 0$.

B.1.1 Distribution of the Regularized Inverse Term

We wish to obtain the probability density function (PDF) of $\left|\mathrm{h}^{\mathrm{H}}\left(|\mathrm{h}|^2 + \eta\right)^{-1}\right|^2$, as well as its expected value. Note that the square of the absolute value of a complex Gaussian distribution with variance σ^2 follows an exponential distribution of parameter $\mu = \frac{1}{\sigma^2}$. Therefore, $|\mathrm{h}|^2 \sim \mathrm{Exp}(\frac{1}{\sigma^2})$. Let us introduce the notation

$$\chi \triangleq |\mathrm{h}|^2 \sim \mathrm{Exp}\Big(\frac{1}{\sigma^2}\Big). \tag{B.2}$$

Furthermore, we denote the CDF of χ as F_χ. We obtain in the following the PDF and expected value of $\chi|(\chi+\eta)^{-1}|^2$. For ease of notation we denote

$$\Psi = \frac{\chi}{(\chi+\eta)^2}. \tag{B.3}$$

The CDF of Ψ, is given by $F_\Psi(\psi) = \text{Prob}\left(\frac{\chi}{(\chi+\eta)^2} < \psi\right)$. Isolating χ yields

$$F_\Psi(\psi) = \begin{cases} 0 & \text{if } \psi < 0 \\ \text{Prob}\left(\chi < \frac{1-2\eta\psi-\sqrt{1-4\eta\psi}}{2\psi}\right) + \text{Prob}\left(\chi > \frac{1-2\eta\psi+\sqrt{1-4\eta\psi}}{2\psi}\right) & \text{if } 0 \le \psi < \frac{1}{4\eta} \\ 1 & \text{if } \psi \ge \frac{1}{4\eta} \end{cases}$$

Then, we can obtain the PDF of Ψ thanks to the fact that $f_\Psi(\psi) = \frac{\mathrm{d}}{\mathrm{d}\chi}F_\Psi(\psi)$. Let us introduce the notation

$$\bar{\psi}^- \triangleq \frac{1-2\eta\psi-\sqrt{1-4\eta\psi}}{2\psi} \qquad \text{and} \qquad \bar{\psi}^+ \triangleq \frac{1-2\eta\psi+\sqrt{1-4\eta\psi}}{2\psi}.$$

Thus,

$$\begin{aligned} f_\Psi(\psi) &= \frac{\mathrm{d}}{\mathrm{d}\chi}\left(F_\chi\left(\bar{\psi}^-\right) + 1 - F_\chi\left(\bar{\psi}^+\right)\right) \\ &= \frac{\bar{\psi}^-}{\psi\sqrt{1-4\eta\psi}} f_\chi\left(\bar{\psi}^-\right) + \frac{\bar{\psi}^+}{\psi\sqrt{1-4\eta\psi}} f_\chi\left(\bar{\psi}^+\right). \end{aligned}$$

For sake of example, consider the case in which $\eta = 1$ and $\sigma^2 = 10$ dB and 20 dB. (We have enclosed the transmit power in the channel coefficient; this example is equivalent to having a unit-variance channel and a transmit power of 10 dB and 20 dB.) Fig. (B.1) shows the PDF (Fig. B.1a-B.1c) and the CDF (Fig. B.1b-B.1d) of these two examples.

B.1.2 Expected value

We continue by calculating the expected value of Ψ. For that, we use the *Law of the unconscious statistician*. Hence, the expected value of the regularized inverse is

$$\mathbb{E}\left[\Psi\right] = \int_0^\infty \frac{x}{(x+\eta)^2} f_\chi(x)\,\mathrm{d}x. \tag{B.4}$$

The PDF of χ is given by $f_\chi(x) = \mu e^{-\mu x}$, with $\mu = \frac{1}{\sigma^2}$. Considering the change of variables $x' = x + \eta$ yields

$$\mathbb{E}\left[\Psi\right] = \int_\eta^\infty \frac{1}{x'}\mu e^{-\mu(x'-\eta)}\,\mathrm{d}x' - \int_\eta^\infty \frac{\eta}{(x')^2}\mu e^{-\mu(x'-\eta)}\,\mathrm{d}x'. \tag{B.5}$$

We follow by integrating the RHS and obtaining

$$\begin{aligned}\mathbb{E}\left[\Psi\right] &= \mu e^{\eta\mu} E_1(\eta\mu) - \mu(1 - e^{\eta\mu}\eta\mu E_1(\eta\mu)) \\ &= \mu\big(e^{\eta\mu} E_1(\eta\mu)(1+\eta\mu) - 1\big),\end{aligned} \tag{B.6}$$

where $E_1(z)$ is the exponential integral defined as

$$E_1(z) \triangleq \int_1^\infty \frac{e^{-zt}}{t}\,\mathrm{d}t. \tag{B.7}$$

B.1.3 GDoF Path-Loss Model: Proof of Proposition (3.3)

Consider now that the variance of the channel coefficient h is $\sigma^2 = P^{\gamma-1}$, and that the regularization parameter is $\eta = 1/P$. In that case, $\mu = P^{1-\gamma}$ and (B.6) becomes

$$\mathbb{E}\left[\left|\mathrm{h}^{\mathrm{H}}\big(\,|\mathrm{h}|^2+\eta\big)^{-1}\right|^2\right] = P^{1-\gamma}\big(e^{P^{-\gamma}} E_1(P^{-\gamma})(1+P^{1-\gamma}) - 1\big), \tag{B.8}$$

what gives Proposition (3.3). ■

B.2 General Expression of the Normalization Parameter

We focus now on the general expression of the normalization constant for every regime, which we obtain from (B.6) and the precoder definition. Let $\sigma^2_{i,k}$ denote the variance of the channel from TX j to RX i ($\mathrm{h}_{i,k}$), and η the regularization constant. We consider the precoder for RX i's data symbols. Recall that

$$\lambda_i = \frac{1}{\sqrt{\mathbb{E}\left[|\mathrm{w}_{i,1}^{(1)}|^2 + |\mathrm{w}_{i,2}^{(2)}|^2\right]}}.$$

Non-locally Informed TXs: Thi precoding vector does not require inversion and λ is given by

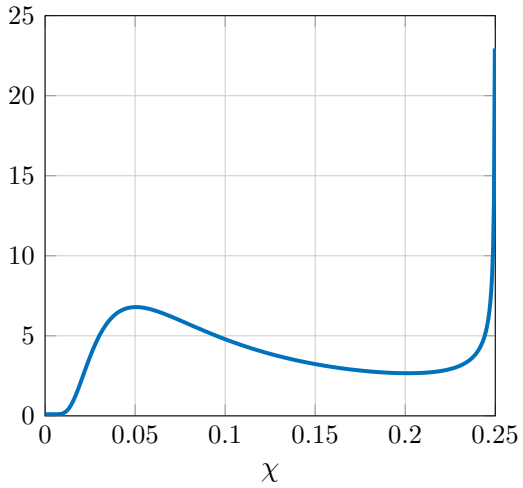
$$\lambda_i = \frac{1}{\sqrt{\sigma^2_{\bar{i},2} + \sigma^2_{\bar{i},1}}}.$$

Most-informed TX: Suppose that TX 1 is the Most-informed TX. Let σ^{-2} be defined as $\sigma^{-2} \triangleq (\sigma^2)^{-1}$. Then precoding vector writes as

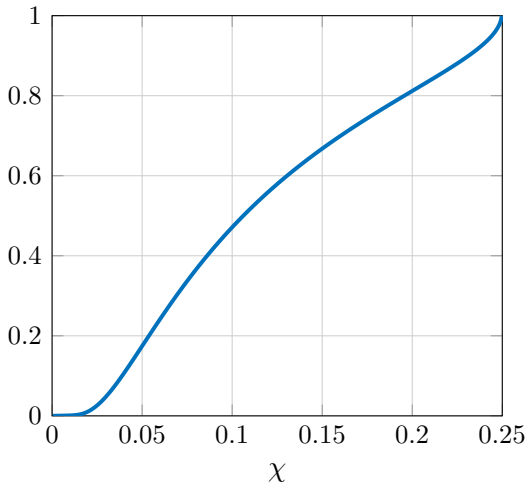
$$\lambda_i = \frac{1}{\sqrt{\sigma^{-2}_{\bar{i},1}\left(e^{\eta\sigma^{-2}_{\bar{i},1}} E_1(\eta\sigma^{-2}_{\bar{i},1})(1+\eta\sigma^{-2}_{\bar{i},1}) - 1\right)\sigma^2_{\bar{i},2} + 1}}.$$

Locally Informed TXs: In this case, each TX applies the regularized inverse as precoder, and thus

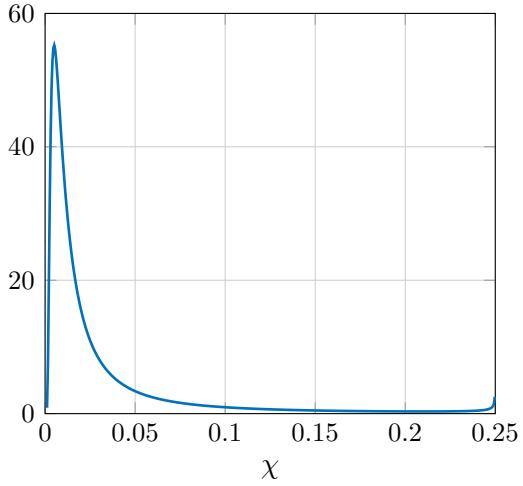
$$\lambda_i = \frac{1}{\sqrt{\sigma_{\bar{i},1}^{-2}\left(e^{\eta\sigma_{\bar{i},1}^{-2}}E_1(\eta\sigma_{\bar{i},1}^{-2})(1+\eta\sigma_{\bar{i},1}^{-2})-1\right)+\sigma_{\bar{i},2}^{-2}\left(e^{\eta\sigma_{\bar{i},2}^{-2}}E_1(\eta\sigma_{\bar{i},2}^{-2})(1+\eta\sigma_{\bar{i},2}^{-2})-1\right)}}.$$



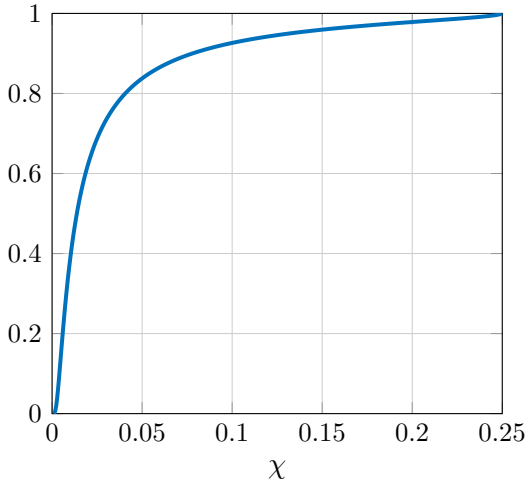
(a) PDF Reg. Inv. with P = 10 dB



(b) CDF Reg. Inv. with P = 10 dB



(c) PDF Reg. Inv. with P = 20 dB



(d) CDF Reg. Inv. with P = 20 dB

Figure B.1 – Probability Density Function (PDF) and Cumulative Density Function (CDF) for $\sigma^2 = 10$ dB and $\sigma^2 = 20$ dB, with $\eta = 1$.

Appendix C

Proofs of Chapter 4 and Properties of AP-ZF

C.1 Properties of AP-ZF

In this appendix, we start by showing some simple but important properties of the AP-ZF precoder. We consider the precoder for RX i's data symbols. From symmetry, the precoder satisfies the same properties for any RX, such that we omit hereinafter the RX's sub-index i for clarity.

Lemma C.1. *Let $\mathbf{H} \in \mathcal{C}^{n \times K}$ denote the channel matrix towards the n RXs whose received interference is canceled. With perfect channel knowledge at all Active TXs, the AP-ZF precoder with n Active TXs and $K - n$ Passive TXs satisfies*

$$\mathbf{H}\mathbf{T}^{\mathrm{APZF}\star} \xrightarrow[P\to\infty]{} \mathbf{0}_{n\times(K-n)}, \tag{C.1}$$

where $\mathbf{T}^{\mathrm{APZF}\star}$ denotes the AP-ZF precoder according to the description in Section 4.4.2 but based on perfect CSIT, and it is given as

$$\mathbf{T}^{\mathrm{APZF}\star} \triangleq \lambda^{\mathrm{APZF}} \begin{bmatrix} \mathbf{T}^{\mathrm{A}\star} \\ \mathbf{T}^{\mathrm{P}} \end{bmatrix}. \tag{C.2}$$

Proof. Using the well known Resolvent identity [159, Lemma 6.1], we can write that

$$\left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{A}} + \frac{1}{P}\mathbf{I}_n\right)^{-1} - \left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{A}}\right)^{-1} = -\left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{A}}\right)^{-1}\frac{1}{P}\mathbf{I}_n\left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{A}} + \frac{1}{P}\mathbf{I}_n\right)^{-1}. \tag{C.3}$$

We can then compute the leaked interference as

$$\begin{aligned}\mathbf{H}\mathbf{T}^{\mathrm{APZF}\star} &= \lambda^{\mathrm{APZF}}\mathbf{H}_{\mathrm{A}}\mathbf{T}^{\mathrm{A}\star} + \lambda^{\mathrm{APZF}}\mathbf{H}_{\mathrm{P}}\mathbf{T}^{\mathrm{P}} \\ &\overset{(a)}{=} \lambda^{\mathrm{APZF}}\mathbf{H}_{\mathrm{A}}\left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{A}}\right)^{-1}\frac{1}{P}\mathbf{I}_n\left(\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{A}} + \frac{1}{P}\mathbf{I}_n\right)^{-1}\mathbf{H}_{\mathrm{A}}^{\mathrm{H}}\mathbf{H}_{\mathrm{P}}\mathbf{T}^{\mathrm{P}},\end{aligned} \tag{C.4}$$

where equality (a) follows from inserting (C.3) inside the AP-ZF precoder and simplifying. It follows that the leaked interference tends to zero as the available power P tend to infinity. ■

Lemma C.2. *The AP-ZF precoder with n Active TXs and $K-n$ Passive TXs is of rank $K-n$.*

Proof. The passive precoder was chosen such that $\mathbf{T}^{\mathrm{P}}$ is full rank (rank $= K-n$). The precoder $\mathbf{T}^{\mathrm{A}(j)}$ is a linear combination of $\mathbf{T}^{\mathrm{P}}$ for each j, such that the effective AP-ZF precoder $\mathbf{T}^{\mathrm{APZF}}$ resulting from distributed precoding is exactly of rank $K-n$. ■

C.2 Proof of Lemma 4.2

Following a similar approach as in [82], we can use once more the resolvent identity [159, Lemma 6.1] to approximate the matrix inverse and show that, $\forall j \leq n$,

$$\left\|\mathbf{T}^{\mathrm{APZF}(j)} - \mathbf{T}^{\mathrm{APZF}\star}\right\|_{\mathrm{F}}^2 = \mathcal{O}_{\mathsf{p}}(P^{-\alpha^{(j)}}). \tag{C.5}$$

It then follows that

$$\begin{aligned}\left\|\mathbf{H}\mathbf{T}^{\mathrm{APZF}}\right\|_{\mathrm{F}}^2 &\overset{(a)}{=} \Theta\Big(\left\|\mathbf{H}\left(\mathbf{T}^{\mathrm{APZF}} - \mathbf{T}^{\mathrm{APZF}\star}\right)\right\|_{\mathrm{F}}^2\Big) \\ &\leq \Theta\Big(\|\mathbf{H}\|_{\mathrm{F}}^2\left\|\mathbf{T}^{\mathrm{APZF}} - \mathbf{T}^{\mathrm{APZF}\star}\right\|_{\mathrm{F}}^2\Big) \\ &\leq \Theta\Big(\|\mathbf{H}\|_{\mathrm{F}}^2\sum_{j=1}^{n}\left\|\mathbf{T}^{\mathrm{APZF}(j)} - \mathbf{T}^{\mathrm{APZF}\star}\right\|_{\mathrm{F}}^2\Big) \\ &\overset{(b)}{=} \mathcal{O}_{\mathsf{p}}\left(P^{-\min_{j\in\{1,\ldots,n\}}\alpha^{(j)}}\right),\end{aligned} \tag{C.6}$$

where (a) comes from Lemma C.1 and (b) follows from (C.5).

Remark C.1. The interference attenuation of AP-ZF precoding is only limited by the worst CSIT accuracy at the Active TXs, and does not depend on the CSI accuracy at the Passive TXs. □

C.3 Proof of Corollary 4.1

In this appendix we prove Corollary 4.1, i.e., that the solution of Theorem 4.3 is composed of two phases (Transmission Modes) at most. This is equivalent to prove that $\exists k_1, k_2 \in \mathcal{K},\ n_1 < k_1,\ n_2 < k_2$ such that

$$\gamma_{n_1,k_1} > 0, \qquad \gamma_{n_2,k_2} \geq 0, \qquad \gamma_{n,k} = 0, \qquad \forall (n,k) \neq (n_1,k_1),(n_2,k_2), \tag{C.7}$$

is always an optimal solution of the maximization problem stated in Theorem 4.3. We know from the problem definition that the number of variables $\gamma_{n,k}$ (denoted as U) of the problem is $\sum_{i=2}^{K}(i-1) = \frac{K(K-1)}{2}$. To simplify the notation, we apply a unique sub-index to the variables such that our variables become $\{\gamma_u \mid u \in \{1,\dots,U\}\}$. The index u is defined as $u \triangleq f_b(n,k)$, where f_b is a bijective function, e.g. $u = \frac{(k-1)(k-2)}{2} + n$. We recall the optimization problem to obtain $\text{DoF}^{\text{APZF}}(\boldsymbol{\alpha})$ but, for sake of clarity, we present it in vector notation. For that, let $\boldsymbol{\gamma}$ be the vector containing the time-sharing variables γ_u, i.e., $\boldsymbol{\gamma} \triangleq [\gamma_1, \gamma_2, \dots, \gamma_U]$. Similarly, we define the vector $\mathbf{F}^{\alpha}$ as the concatenation of the effective DoF of each mode in (4.14), such that $\mathbf{F}^{\alpha}_u = 1 + (k-1)\alpha^{(n)}$, with k, n given by $(n,k) = f_b^{-1}(u)$. Finally, the vector of terms $d_{n,k} \triangleq 1 - \alpha^{(n)} - k(k-n-1)\alpha^{(n)}$ for the constraint (4.16) is denoted as $\boldsymbol{d}$. Hence, the problem of Theorem 4.3 can be expressed as

$$\begin{aligned}
\text{DoF}^{\text{APZF}}(\boldsymbol{\alpha}) = \underset{\boldsymbol{\gamma}}{\text{maximize}} \quad & \mathbf{F}^{\alpha}\boldsymbol{\gamma} && \text{(C.8)}\\
\text{subject to} \quad & \|\boldsymbol{\gamma}\|_1 = 1, && \text{(C.9)}\\
& \boldsymbol{\gamma} \succeq 0, && \text{(C.10)}\\
& \boldsymbol{d}\boldsymbol{\gamma} \geq 0, && \text{(C.11)}
\end{aligned}$$

where $\mathbf{F}^{\alpha}$, $\boldsymbol{d}$, are constant vectors. Let us remind that if a linear programming problem has an optimal solution then it is an extreme point of the feasible set [160].

The feasible set given by conditions (C.9)-(C.10), which is denoted by $\mathcal{C}$, is the *probability simplex* [161] determined by the unit vectors $e_1, \dots, e_U \in \mathbb{R}^U$, hence it is a $(U-1)$-dimensional simplex. On the other hand, condition in (C.11) represents a half-space determined by the vector hyperplane [161] denoted as $\mathcal{V}$ and given by $\boldsymbol{d}\boldsymbol{\gamma} = 0$. We can have different cases depending on how the *probability simplex* $\mathcal{C}$ and the half-space determined by the hyperplane $\mathcal{V}$ intersect:

1. If $\mathcal{C} \cap \{\boldsymbol{\gamma} \mid \boldsymbol{d}\boldsymbol{\gamma} \geq 0\} = \mathcal{C}$ ($\mathcal{C}$ is a subset of the half-space), the feasible region is $\mathcal{C}$ and the extreme points are the unit vectors e_u. Then, the solution of the problem uses only a single mode –because in e_u the only non-zero variable is the u-th variable–.

2. If $\mathcal{C} \cap \{\boldsymbol{\gamma} \mid \boldsymbol{d}\boldsymbol{\gamma} \geq 0\} = \emptyset$, there is not feasible solution. However, this is not possible

since we have shown that this linear program is always feasible, just choosing $\gamma_{k-1,k} = 1$, with $k \in \{2, \ldots, K\}$.

3. If $\mathcal{C} \cap \{\boldsymbol{\gamma} \mid \boldsymbol{d\gamma} \geq 0\} \subset \mathcal{C}$, we need to prove that all the extreme points of the resulting set satisfy (C.7). Those extreme points either they will be the extreme points of $\mathcal{C}$ or they will belong to the intersection between $\mathcal{C}$ and $\mathcal{V}$.

From linear algebra, we know that the intersection of an l-dimensional and an m-dimensional sub-space in the n-dimensional space $\mathbb{R}^n$ has dimension p such that

$$p \geq l + m - n. \tag{C.12}$$

Thus, in order to obtain the extreme points ($p = 0$) of the feasible set, we must obtain the intersection between $\mathcal{V}$ (m $= U - 1$), and the *edges* of $\mathcal{C}$, i.e., the 1-faces (segments) that define $\mathcal{C}$, in the space $\mathbb{R}^U$. The edges of $\mathcal{C}$ are segments that connect two points with a single non-zero variable –the unit vectors–, and therefore they belong to a line of only two non-zero variables. Given that the intersection of $\mathcal{V}$ with one edge must be a point of the edge, it holds that all the extreme points satisfy (C.7), and therefore Corollary 4.1 is proven. From the previous analysis, it follows that the feasibility set is convex.

Moreover, as Theorem 4.3 is always composed of at most two Transmission Modes, it can be expressed as the following integer linear program:

$$\mathrm{DoF}^{\mathrm{APZF}}(\boldsymbol{\alpha}) = \underset{k_1,n_1,k_2,n_2}{\text{maximize}}\ 1 + \gamma(k_1 - 1)\alpha^{(n_1)} + (1-\gamma)(k_2 - 1)\alpha^{(n_2)} \tag{C.13}$$

$$\text{subject to}\ \ k_1, k_2 \in \{2, \ldots, K\}, \tag{C.14}$$

$$n_1 \in \{1, \ldots, k_1 - 1\}, \tag{C.15}$$

$$n_2 \in \{1, \ldots, k_2 - 1\} \mid d_{n_2,k_2} \geq 0, \tag{C.16}$$

where γ is given by

$$\gamma \triangleq \begin{cases} 1 & \text{if } d_{n_1,k_1} \geq 0 \\ \frac{d_{n_2,k_2}}{d_{n_2,k_2} - d_{n_1,k_1}} & \text{otherwise} \end{cases}. \tag{C.17}$$

■

C.4 Proof of Proposition 4.1

In this section we prove Proposition 4.1, i.e., that Lemma 4.1 holds for the case where the different estimates of a certain link are correlated, and both noises and channel are

drawn from Gaussian distributions. Specifically, we assume that the setting satisfy the following assumptions.

Assumption.

(C.H1) $\mathrm{h}_{i,k} \sim \mathcal{N}(0,1), \quad \forall i,k \in \mathbb{N}_K.$

(C.H2) $\delta_{i,k}^{(j)} \sim \mathcal{N}(0,1), \quad \forall i,k,j \in \mathbb{N}_K.$

(C.H3) $\hat{\mathrm{h}}_{i,k}^{(j)} = \mathrm{h}_{i,k} + \bar{P}^{-\alpha_{i,k}^{(j)}} \delta_{i,k}^{(j)}.$

(C.H4) $\mathrm{h}_{i,k} \perp\!\!\!\perp \mathrm{h}_{i',k'}, \qquad \forall (i,k) \neq (i',k')$

(C.H5) $\delta_{i,k}^{(j)} \perp\!\!\!\perp \delta_{i',k'}^{(\ell)}, \qquad \forall j,\ell \in \mathbb{N}_K, \quad \forall (i,k) \neq (i',k').$

(C.H6) $\mathrm{Prob}\left(\sum_{j=1}^{K} \boldsymbol{a}_j \delta_{i,k}^{(j)} = 0 \right) < 1.$

Assumption (C.H6) implies that none of the noise variables $\{\delta_{i,k}^{(j)}\}_{j \in \mathbb{N}_K}$ is a deterministic linear combination of the others. From (C.H4)-(C.H5), we can restrict ourselves to a single arbitrary link. Consequently, let h be the channel coefficient of a link between an arbitrary TX and an arbitrary RX, i.e., $\mathrm{h} = \mathrm{h}_{i,k}$, with $i \in \mathbb{N}_K$, $k \in \mathbb{N}_K$. Henceforth, we omit the sub-indexes (i,k) for sake of readability.

Besides the previous assumptions, we consider that the estimation noise at TX j is correlated with the estimation noise at TX ℓ with a correlation factor $\rho_{j,\ell}$, such that

$$\rho_{j,\ell} \triangleq \mathrm{cov}(\delta^{(j)}, \delta^{(\ell)}). \tag{C.18}$$

Note that (C.H6) precludes the cases with $|\rho_{j,\ell}| = 1$. For the sake of completeness, we recall the following result on multivariate Gaussian distribution.

Theorem C.1. *[162, Theorem 23.7.4. p. 484] Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be centered and jointly Gaussian with covariance matrix* $\mathbf{K_{XX}}$ *and* $\mathbf{K_{YY}}$*. Assume that* $\mathbf{K_{YY}} \succ 0$*. Then the conditional distribution of* $\mathbf{X}$ *conditioned on* $\mathbf{Y} = \boldsymbol{y}$ *is a multivariate Gaussian of mean* $\mathbb{E}\left[\mathbf{X}\mathbf{Y}^{\mathrm{H}}\right] \mathbf{K}_{\mathbf{YY}}^{-1} \boldsymbol{y}$ *and covariance matrix*

$$\mathbf{K_{XX}} - \mathbb{E}\left[\mathbf{X}\mathbf{Y}^{\mathrm{H}}\right] \mathbf{K}_{\mathbf{YY}}^{-1} \mathbb{E}\left[\mathbf{Y}\mathbf{X}^{\mathrm{H}}\right] \tag{C.19}$$

In our case, $\mathbf{X}$ represents the channel (h) and $\mathbf{Y}$ denotes the set of estimates $\{\hat{\mathrm{h}}^{(j)}\}_{j \in \mathbb{N}_K}$. Applying Theorem C.1, the conditional distribution $p_{\mathrm{h}|\hat{\mathrm{h}}^{(1)},\ldots,\hat{\mathrm{h}}^{(K)}}$ is multivariate Gaussian. Our goal is to compute the covariance matrix of this conditional distribution, denoted

by $\mathbf{K}_{\mathbf{X}|\mathbf{Y}}$. Let us denote the covariance matrix between the noise random variables as $\boldsymbol{\Delta} \in \mathbb{C}^{K\times K}$, such that

$$\boldsymbol{\Delta} = \begin{bmatrix} P^{-\alpha^{(1)}} & \dots & \rho_{1,K}\bar{P}^{-\alpha^{(1)}-\alpha^{(K)}} \\ \rho_{1,2}\bar{P}^{-\alpha^{(2)}-\alpha^{(1)}} & \dots & \rho_{2,K}\bar{P}^{-\alpha^{(2)}-\alpha^{(K)}} \\ \vdots & \ddots & \vdots \\ \rho_{1,K}\bar{P}^{-\alpha^{(K)}-\alpha^{(1)}} & \dots & P^{-\alpha^{(K)}} \end{bmatrix}. \tag{C.20}$$

Let $\mathbf{K}_{\mathbf{YY}} \in \mathbb{C}^{K\times K}$ denote the covariance matrix of $\mathbf{Y}$. It is given by

$$\mathbf{K}_{\mathbf{YY}} = (\mathbf{1}_K + \boldsymbol{\Delta}), \tag{C.21}$$

where $\mathbf{1}_K$ is the all-ones matrix of size $K \times K$.

Remark C.2. From (C.H6), $\boldsymbol{\Delta}$ is non-singular. This follows because, as explained in [162, Section 23.4.3. p. 466], a singular covariance matrix implies that there is at least one component of the random vector such that it is determined with probability one by an affine function of other components. □

Let $\mathbf{D}_{\boldsymbol{\alpha}} \in \mathbb{R}^{K\times K}$ be the diagonal matrix obtained from the vector of error variances, i.e.,

$$\mathbf{D}_{\boldsymbol{\alpha}} \triangleq \operatorname{diag}\left(\left[\bar{P}^{-\alpha^{(1)}}, \dots, \bar{P}^{-\alpha^{(K)}}\right]\right). \tag{C.22}$$

From its structure in (C.20), we can write the matrix $\boldsymbol{\Delta}$ as $\boldsymbol{\Delta} = \mathbf{D}_{\boldsymbol{\alpha}}\mathbf{P}\mathbf{D}_{\boldsymbol{\alpha}}$, where $\mathbf{P} \in \mathbb{C}^{K\times K}$ is a symmetric matrix that only depends on the correlation coefficients $\rho_{j,\ell}$. Therefore,

$$\boldsymbol{\Delta}^{-1} = \mathbf{D}_{\boldsymbol{\alpha}}^{-1}\mathbf{P}^{-1}\mathbf{D}_{\boldsymbol{\alpha}}^{-1}. \tag{C.23}$$

It follows from Remark C.2 that the inverse $\mathbf{P}^{-1}$ exists. Thus, the j-th row, ℓ-th column coefficient of the matrix $\boldsymbol{\Delta}^{-1}$ can be written as

$$\boldsymbol{\Delta}^{-1}_{j,\ell} = \bar{P}^{\alpha^{(j)}+\alpha^{(\ell)}}\mathbf{P}^{-1}_{j,\ell}, \tag{C.24}$$

and thus $\boldsymbol{\Delta}^{-1}_{j,\ell} = O(\bar{P}^{\alpha^{(j)}+\alpha^{(\ell)}})$. Moreover, the all-ones matrix $\mathbf{1}_K$ can be expressed as $\mathbf{1}_K = \mathbf{1}_{K,1}\mathbf{1}_{K,1}^{\mathrm{T}}$. With this notation and by using Theorem C.1, the covariance matrix $\mathbf{K}_{\mathbf{X}|\mathbf{Y}}$ is then expressed as

$$\begin{aligned} \mathbf{K}_{\mathbf{X}|\mathbf{Y}} &= 1 - \mathbf{1}_{K,1}^{\mathrm{T}}\left(\mathbf{1}_{K,1}\mathbf{1}_{K,1}^{\mathrm{T}} + \boldsymbol{\Delta}\right)^{-1}\mathbf{1}_{K,1} \\ &= 1 - \mathbf{1}_{K,1}^{\mathrm{T}}\boldsymbol{\Delta}^{-1}\mathbf{1}_{K,1}\left(1 + \mathbf{1}_{K,1}^{\mathrm{T}}\boldsymbol{\Delta}^{-1}\mathbf{1}_{K,1}\right)^{-1}, \end{aligned} \tag{C.25}$$

where the last equality follows from the Matrix Inversion Lemma [163, Chapter 3.1.1], specifically because $(\mathbf{BCD}+\mathbf{A})^{-1}\mathbf{BC}=\mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1}+\mathbf{DA}^{-1}\mathbf{B})^{-1}$. Note that

$$\mathbf{1}_{K,1}^{\mathrm{T}}\boldsymbol{\Delta}^{-1}\mathbf{1}_{K,1}=\sum_{j=1}^{K}\sum_{\ell=1}^{K}\boldsymbol{\Delta}_{j,\ell}^{-1}. \tag{C.26}$$

Applying (C.26) to (C.25) we have

$$\begin{aligned}\mathbf{K}_{\mathbf{X}|\mathbf{Y}}&=1-\mathbf{1}_{K,1}^{\mathrm{T}}\boldsymbol{\Delta}^{-1}\mathbf{1}_{K,1}\left[1+\mathbf{1}_{K,1}^{\mathrm{T}}\boldsymbol{\Delta}^{-1}\mathbf{1}_{K,1}\right]^{-1}\\&=\frac{1}{1+\sum_{j=1}^{K}\sum_{\ell=1}^{K}\boldsymbol{\Delta}_{j,\ell}^{-1}}.\end{aligned} \tag{C.27}$$

Hence, from (C.27) and (C.24), the conditional probability density function is Gaussian with the variance of its elements scaling in $P^{\max_j \alpha^{(j)}}$ such that it satisfies that

$$\max_{\mathrm{h}} f_{\mathrm{h}|\hat{\mathrm{h}}^{(1)},\ldots,\hat{\mathrm{h}}^{(K)}}=O\Big(\sqrt{P^{\max_{j\in\mathbb{N}_K}\alpha^{(j)}}}\Big), \tag{C.28}$$

what concludes the proof of Proposition 4.1. ■

Appendix D

Proof of Lemma 4.1

In this appendix, we prove Lemma 4.1 –and so Theorem 4.1– for a broad general case, where the estimation noise random variables are mutually independent and they are drawn from continuous distributions with density. We first enunciate some definitions and hypothesis that are taken on the random variables and their PDFs. Later, we prove the lemma for the case with $K = 2$ and, to conclude, we prove the general case with $K > 2$ by induction. From the independence between different channel coefficients, we restrict ourselves to an arbitrary link such that we omit the sub-indexes i, k.

D.1 Preliminaries

As opposed to the general notation used throughout this manuscript, in this appendix we use a different notation for expressing a random variable and its realization. Henceforth, random variables are denoted by calligraphic upper-case letters ($\mathcal{X}$), and the realization of the variable $\mathcal{X}$ is denoted by regular lower-case letters (x). We recall that the PDF of a variable $\mathcal{X}$ is denoted as $f_{\mathcal{X}}$. Let us first introduce several important definitions.

Definition D.1. For $\varepsilon > 0$, The ε-support of a random variable $\mathcal{X}$ is defined as

$$\mathcal{S}_{\mathcal{X}}^{\varepsilon} = \{x \mid f_{\mathcal{X}}(x) > \varepsilon\}. \tag{D.1}$$

Based on the Definition D.1, we present the notion of *Bounded Support* as follows.

Definition D.2 (Bounded Support)**.** A random variable $\mathcal{X}$ has bounded support if there exists a constant $M_{\mathcal{X}} < \infty$ such that $x \leq |M_{\mathcal{X}}| \ \forall x \in \mathcal{S}_{\mathcal{X}}^{\varepsilon}, \forall \varepsilon$.

Definition D.3 (Bounded Probability Density function)**.** A random variable $\mathcal{X}$ has bounded probability density function if there exists a constant $f_{\mathcal{X}}^{\max} < \infty$ such that $f_{\mathcal{X}}(x) \leq f_{\mathcal{X}}^{\max}$ for any x.

Let $\mathcal{X} \perp\!\!\!\perp \mathcal{Y}$ denote that the random variables $\mathcal{X}$ and $\mathcal{Y}$ are independent. We consider the following assumptions on the random variables.

Assumption.

(D.H1) $\mathcal{H},\ \Delta^{(j)},\ \forall j \in \mathbb{N}_K$, are continuous random variables with bounded support and bounded probability density function.

(D.H2) $\hat{\mathcal{H}}^{(j)} \triangleq \mathcal{H} + \bar{P}^{-\alpha^{(j)}}\Delta^{(j)}$.

(D.H3) $\mathcal{H},\ \Delta^{(j)} \perp\!\!\!\perp P,\ \alpha^{(j)}$.

(D.H4) $\mathcal{H}_{i,k} \perp\!\!\!\perp \mathcal{H}_{i',k'}, \quad \forall (i,k) \neq (i',k')$.

(D.H5) $\Delta_{i,k}^{(j)} \perp\!\!\!\perp \Delta_{i',k'}^{(j')}, \quad \forall (i,k,j) \neq (i',k',j')$.

We denote the observed values of the aforementioned variables as $\mathrm{h} \sim \mathcal{H}$, $\delta^{(j)} \sim \Delta^{(j)}$, $\hat{\mathrm{h}}^{(j)} \sim \hat{\mathcal{H}}^{(j)}$, and consequently $\hat{\mathrm{h}}^{(j)} \triangleq \mathrm{h} + \bar{P}^{-\alpha^{(j)}}\delta^{(j)}$. Furthermore, let us consider that the realizations are in the ε-support of their respective variables. As a refresher, and because we will make extensive use of it, we recall the well-known formula for the PDF of a random variable multiplied by a positive constant.

Proposition D.1. *Let $\mathcal{X}$ be a continuous random variable with PDF $f_{\mathcal{X}}(x)$, and let $c \in \mathbb{R}$ be a constant satisfying $c > 0$. Then, the random variable $c \cdot \mathcal{X}$ is also a continuous random variable whose probability density function is given by*

$$f_{c\mathcal{X}}(x) = \frac{1}{c} f_{\mathcal{X}}\left(\frac{x}{c}\right). \tag{D.2}$$

Furthermore, we present a useful lemma on the convergence of the estimate variables $\hat{\mathcal{H}}^{(j)}$ that will be useful for the proof.

Lemma D.1. *Let $\hat{\mathcal{H}}^{(j)}$, with $j \in \mathbb{N}_K$, be defined from assumptions (D.H1)-(D.H5), such that $\alpha^{(j)} > 0$. Then, $f_{\hat{\mathcal{H}}^{(j)}}$ converges almost surely to $f_{\mathcal{H}}$, i.e.,*

$$\lim_{P \to \infty} f_{\hat{\mathcal{H}}^{(j)}}(x) = f_{\mathcal{H}}(x). \tag{D.3}$$

The previous lemma leads to the next corollary for the conditional PDF of the estimation noise.

Corollary D.1. *Assume that* $\alpha^{(1)} > 0$. *Then,*

$$\lim_{P\to\infty} f_{\Delta^{(1)}|\hat{\mathcal{H}}^{(1)},\dots,\hat{\mathcal{H}}^{(K)}}(y|\hat{\mathrm{h}}^{(1)},\ \dots,\ \hat{\mathrm{h}}^{(K)}) = f_{\Delta^{(1)}}(y). \tag{D.4}$$

Proof. The proof of both lemma and corollary is relegated to Section D.4. ■

Finally, we recall here the Lebesgue's Dominated Convergence Theorem [158].

Theorem D.1 ([158, Theorem 16.4]). *Let* $\{f_n\}$ *be a sequence of functions on the measure space* (Ω, Σ, μ), *where* Ω *is a non-empty sample space,* Σ *is a* σ*-algebra on the space* Ω, *and* μ *a measure on* (Ω, Σ). *Suppose that*

$$\lim_{n\to\infty} f_n(x) = f(x) \tag{D.5}$$

almost surely. Further suppose that exists an integrable non-negative function G *such that*

$$|f_n(x)| \leq G(x),\ \forall n, \tag{D.6}$$

almost surely. Then $\{f_n\}$ *and* f *are integrable and*

$$\lim_{n\to\infty} \int_\Omega f_n(x) d\mu(x) = \int_\Omega f(x) d\mu(x). \tag{D.7}$$

D.2 Proof for the K=2 estimates Case

Before analyzing the conditional PDF $f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\hat{\mathcal{H}}^{(2)}}$, let us introduce two claims on the PDF of $f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)}}$ and $f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}$. The proof of both claims is relegated at the end of the section.

Claim D.1. *Let* $\mathcal{H}$ *and* $\hat{\mathcal{H}}^{(1)}$ *be defined as in* (D.H1)-(D.H5). *Then,*

$$\max f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)}}(\mathrm{h} \mid \hat{\mathrm{h}}^{(1)}) = \Theta\big(\bar{P}^{\alpha^{(1)}}\big). \tag{D.8}$$

Claim D.2. *Let $\mathcal{H}$, $\hat{\mathcal{H}}^{(1)}$, and $\hat{\mathcal{H}}^{(2)}$ be defined as in (D.H1)-(D.H5). Let us define $\upsilon_P(y)$ as*

$$\upsilon_P(y) \triangleq f_{\Delta^{(2)}}\big(\delta^{(2)} - \bar{P}^{\alpha^{(2)}-\alpha^{(1)}}(\delta^{(1)} + y)\big) f_{-\Delta^{(1)}|\hat{\mathcal{H}}^{(1)}}(y|\hat{\mathrm{h}}^{(1)}). \tag{D.9}$$

Then,

$$f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(2)} \mid \hat{\mathrm{h}}^{(1)}) = \bar{P}^{\alpha^{(2)}} \int_{-\infty}^{\infty} \upsilon_P(y)\, \mathrm{d}y. \tag{D.10}$$

Furthermore, it holds that

$$\lim_{P\to\infty} \upsilon_P(y) = \begin{cases} f_{\Delta^{(2)}}(\delta^{(2)} - \delta^{(1)} - y) f_{-\Delta^{(1)}}(y) & \text{if } \alpha^{(1)} = \alpha^{(2)} \quad \text{(D.11a)} \\ f_{\Delta^{(2)}}(\delta^{(2)}) f_{-\Delta^{(1)}}(y) & \text{if } \alpha^{(1)} > \alpha^{(2)} \quad \text{(D.11b)} \end{cases}$$

For the $K = 2$ case, Lemma 4.1 states that $\max f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\hat{\mathcal{H}}^{(2)}} = \Theta\big(\bar{P}^{\alpha^{(1)}}\big)$. From the statement of Claim D.1, in order to prove Lemma 4.1 we need to demonstrate that

$$\lim_{P\to\infty} \frac{f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\hat{\mathcal{H}}^{(2)}}(\mathrm{h}|\hat{\mathrm{h}}^{(1)}, \hat{\mathrm{h}}^{(2)})}{f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)}}(\mathrm{h}|\hat{\mathrm{h}}^{(1)})} = \Theta(1), \tag{D.12}$$

i.e., that the limit exists and it is bounded away from 0 and ∞. Let us start by noting that

$$\begin{aligned} f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\hat{\mathcal{H}}^{(2)}}(h|\hat{\mathrm{h}}^{(1)}, \hat{\mathrm{h}}^{(2)}) &= \frac{f_{\mathcal{H},\hat{\mathcal{H}}^{(1)},\hat{\mathcal{H}}^{(2)}}(\mathrm{h}, \hat{\mathrm{h}}^{(1)}, \hat{\mathrm{h}}^{(2)})}{f_{\hat{\mathcal{H}}^{(1)},\hat{\mathcal{H}}^{(2)}}(\hat{\mathrm{h}}^{(1)}, \hat{\mathrm{h}}^{(2)})} \\ &\overset{(a)}{=} \frac{f_{\mathcal{H},\hat{\mathcal{H}}^{(1)}}(\mathrm{h}, \hat{\mathrm{h}}^{(1)}) f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}}(\hat{\mathrm{h}}^{(2)} - \mathrm{h})}{f_{\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(1)}) f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(2)} \mid \hat{\mathrm{h}}^{(1)})} \\ &= f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)}}(\mathrm{h}|\hat{\mathrm{h}}^{(1)}) \frac{f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}}(\hat{\mathrm{h}}^{(2)} - \mathrm{h})}{f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(2)} \mid \hat{\mathrm{h}}^{(1)})}, \end{aligned} \tag{D.13}$$

where (a) comes from the independence between $\mathcal{H}$, $\Delta^{(1)}$, $\Delta^{(2)}$. Equation (D.13) yields

$$\frac{f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\hat{\mathcal{H}}^{(2)}}(\mathrm{h}|\hat{\mathrm{h}}^{(1)}, \hat{\mathrm{h}}^{(2)})}{f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)}}(\mathrm{h}|\hat{\mathrm{h}}^{(1)})} = \frac{f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}}(\hat{\mathrm{h}}^{(2)} - \mathrm{h})}{f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(2)} \mid \hat{\mathrm{h}}^{(1)})}. \tag{D.14}$$

Hereinafter we focus on the RHS of (D.14). Note that we can write $f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}}(\hat{\mathrm{h}}^{(2)} - \mathrm{h}) = \bar{P}^{-\alpha^{(2)}} f_{\Delta^{(2)}}(\delta^{(2)})$, what follows from Proposition D.1. From (D.10) in Claim D.2, the

RHS of (D.14) can be expressed as

$$\frac{f_{\bar{P}-\alpha^{(2)}\Delta^{(2)}}(\hat{\mathrm{h}}^{(2)}-\mathrm{h})}{f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(2)}\mid\hat{\mathrm{h}}^{(1)})}=\frac{f_{\Delta^{(2)}}(\delta^{(2)})}{\int_{-\infty}^{\infty}\upsilon_P(y)\,\mathrm{d}y}. \tag{D.15}$$

We consider separately the two possible subcases $\alpha^{(1)}=\alpha^{(2)}$ and $\alpha^{(1)}>\alpha^{(2)}$:

a) Case $\alpha^{(1)}=\alpha^{(2)}$:

From the Lebesgue's Dominated Convergence Theorem (Theorem D.1), the bounded probability density assumption, and (D.11a), the limit exists and it holds that

$$\lim_{P\to\infty}\int_{-\infty}^{\infty}\upsilon_P(y)\,\mathrm{d}y=f_{-\Delta^{(1)}}*f_{\Delta^{(2)}}(\delta^{(2)}-\delta^{(1)}). \tag{D.16}$$

where $f*g(x)$ stands for the convolution $(f*g)(x)$ between $f(x)$ and $g(x)$. From (D.15) and (D.16), it holds that

$$\lim_{P\to\infty}\frac{f_{\bar{P}-\alpha^{(2)}\Delta^{(2)}}(\hat{\mathrm{h}}^{(2)}-\mathrm{h})}{f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(2)}\mid\hat{\mathrm{h}}^{(1)})}=\frac{f_{\Delta^{(2)}}(\delta^{(2)})}{f_{-\Delta^{(1)}}*f_{\Delta^{(2)}}(\delta^{(2)}-\delta^{(1)})}. \tag{D.17}$$

From the bounded density assumption, it exists a $f^{\max}_{\Delta^{(i)}}<\infty$ such that, for all x, $f_{\Delta^{(i)}}(x)\leq f^{\max}_{\Delta^{(i)}}$. Then, it holds that

$$f_{-\Delta^{(1)}}*f_{\Delta^{(2)}}(x)\;\leq\;\max(f^{\max}_{\Delta^{(1)}},f^{\max}_{\Delta^{(2)}}). \tag{D.18}$$

Let $\mathbb{1}$ be the indicator function and let then τ be

$$\tau\triangleq\int_{-\infty}^{\infty}\mathbb{1}_{x\in\mathcal{S}^{\varepsilon}_{-\Delta^{(1)}}}\times\mathbb{1}_{(\delta^{(2)}-\delta^{(1)}-x)\in\mathcal{S}^{\varepsilon}_{\Delta^{(2)}}}\,\mathrm{d}x. \tag{D.19}$$

Then, it follows that

$$f_{-\Delta^{(1)}}*f_{\Delta^{(2)}}(\delta^{(2)}-\delta^{(1)})\;>\;\epsilon^2\tau \tag{D.20}$$

and $\tau>0$ if $\delta^{(1)}\in\mathcal{S}^{\varepsilon}_{\Delta^{(1)}}$ and $\delta^{(2)}\in\mathcal{S}^{\varepsilon}_{\Delta^{(2)}}$. From (D.18) and (D.20), (D.17) satisfies

$$\frac{\varepsilon}{\max(f^{\max}_{\Delta^{(1)}},f^{\max}_{\Delta^{(2)}})}\;<\;\frac{f_{\Delta^{(2)}}(\delta^{(2)})}{f_{-\Delta^{(1)}}*f_{\Delta^{(2)}}(\delta^{(2)}-\delta^{(1)})}\;<\;\frac{f^{\max}_{\Delta^{(2)}}}{\varepsilon^2\tau}.$$

This implies (D.12) and thus the proof is concluded for the $\alpha^{(2)}=\alpha^{(1)}$ case.

a) Case $\alpha^{(1)} > \alpha^{(2)}$

From the Lebesgue's Dominated Convergence Theorem, the bounded probability density assumption, and (D.11b), the limit exists and it holds that

$$\begin{aligned}\lim_{P\to\infty} v_P(y) &= \int_{-\infty}^{\infty} f_{\Delta^{(2)}}(\delta^{(2)}) f_{-\Delta^{(1)}}(y)\,\mathrm{d}y \\ &= f_{\Delta^{(2)}}(\delta^{(2)}).\end{aligned} \tag{D.21}$$

By applying (D.21) in (D.15) we obtain that

$$\lim_{P\to\infty} \frac{f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)},\hat{\mathcal{H}}^{(2)}}(\mathrm{h}|\hat{\mathrm{h}}^{(1)},\hat{\mathrm{h}}^{(2)})}{f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)}}(\mathrm{h}|\hat{\mathrm{h}}^{(1)})} = 1. \tag{D.22}$$

This concludes the proof of Lemma 4.1 for the 2-estimate case. ■

D.2.1 Proof of Claim D.1

Using Bayes' formula we obtain that

$$\begin{aligned}f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)}}(\mathrm{h} \mid \hat{\mathrm{h}}^{(1)}) &= \frac{f_{\bar{P}^{-\alpha^{(1)}}\Delta^{(1)}}(\bar{P}^{-\alpha^{(1)}}\delta^{(1)}) f_{\mathcal{H}}(\mathrm{h})}{f_{\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(1)})} \\ &= \bar{P}^{\alpha^{(1)}} \frac{f_{\Delta^{(1)}}(\delta^{(1)}) f_{\mathcal{H}}(\mathrm{h})}{f_{\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(1)})},\end{aligned} \tag{D.23}$$

where the last equality comes from Proposition D.1. Let us consider separately the cases where $\alpha^{(1)} = 0$ and where $\alpha^{(1)} > 0$.

a) $\alpha^{(1)} = 0$**:** In this case, (D.23) does not depend on P, since $P^0 = 1,\ \forall P > 0$. From the bounded probability density assumption, $f_{\mathcal{H}}$ and $f_{\Delta^{(1)}}$ are bounded away from ∞. Moreover, if $\hat{\mathrm{h}}^{(1)} \in \mathcal{S}^{\varepsilon}_{\hat{\mathcal{H}}^{(1)}}$, then $f_{\hat{\mathcal{H}}^{(1)}}$ is also lower-bounded by ε. Thus,

$$\max f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)}}(\mathrm{h} \mid \hat{\mathrm{h}}^{(1)}) = \Theta\left(\bar{P}^0\right). \tag{D.24}$$

b) $\alpha^{(1)} > 0$**:** From Lemma D.1, we have that $f_{\hat{\mathcal{H}}^{(1)}}$ converges almost surely (a.s.) to $f_{\mathcal{H}}$, and from the bounded probability density assumption that $\max f_{\Delta^{(1)}} < \infty$. Thus, from (D.23) it holds that

$$\max f_{\mathcal{H}|\hat{\mathcal{H}}^{(1)}}(\mathrm{h} \mid \hat{\mathrm{h}}^{(1)}) = \Theta\big(\bar{P}^{\alpha^{(1)}}\big). \tag{D.25}$$

■

D.2.2 Proof of Claim D.2

Since $\Delta^{(j)}$ is independent of $\mathcal{H}$, and $\hat{\mathcal{H}}^{(2)} = \hat{\mathcal{H}}^{(1)} - \bar{P}^{-\alpha^{(1)}}\Delta^{(1)} + \bar{P}^{-\alpha^{(2)}}\Delta^{(2)}$, it follows that

$$\begin{aligned} f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(2)} \mid \hat{\mathrm{h}}^{(1)}) &= f_{\hat{\mathcal{H}}^{(1)} - \bar{P}^{-\alpha^{(1)}}\Delta^{(1)} + \bar{P}^{-\alpha^{(2)}}\Delta^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(1)} + \hat{\mathrm{h}}^{(2)} - \hat{\mathrm{h}}^{(1)} \mid \hat{\mathrm{h}}^{(1)}) \\ &= f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)} - \bar{P}^{-\alpha^{(1)}}\Delta^{(1)}|\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(2)} - \hat{\mathrm{h}}^{(1)} \mid \hat{\mathrm{h}}^{(1)}). \end{aligned} \tag{D.26}$$

Note that $\hat{\mathrm{h}}^{(2)} - \hat{\mathrm{h}}^{(1)} = \bar{P}^{-\alpha^{(2)}}\delta^{(2)} - \bar{P}^{-\alpha^{(1)}}\delta^{(1)}$. From the independence of $\Delta^{(1)}$ and $\Delta^{(2)}$, we can rewrite (D.26) in terms of the convolution as

$$\begin{aligned} f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(2)} \mid \hat{\mathrm{h}}^{(1)}) &= f_{-\bar{P}^{-\alpha^{(1)}}\Delta^{(1)}|\hat{\mathcal{H}}^{(1)}} * f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}}(\hat{\mathrm{h}}^{(2)} - \hat{\mathrm{h}}^{(1)}|\hat{\mathrm{h}}^{(1)}) \\ &= \int_{-\infty}^{\infty} f_{\bar{P}^{-\alpha^{(2)}}\Delta^{(2)}}(\hat{\mathrm{h}}^{(2)} - \hat{\mathrm{h}}^{(1)} - x) f_{-\bar{P}^{-\alpha^{(1)}}\Delta^{(1)}|\hat{\mathcal{H}}^{(1)}}(x|\hat{\mathrm{h}}^{(1)})\,\mathrm{d}x. \end{aligned} \tag{D.27}$$

Consider the change of PDF of Proposition D.1. If we apply it to pass from $f_{-\bar{P}^{-\alpha^{(1)}}\Delta^{(1)}}$ to $f_{-\Delta^{(1)}}$, we can express $f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(2)} \mid \hat{\mathrm{h}}^{(1)})$ as

$$\int_{-\infty}^{\infty} \bar{P}^{\alpha^{(2)}} f_{\Delta^{(2)}}\big(\bar{P}^{\alpha^{(2)}}(\hat{\mathrm{h}}^{(2)} - \hat{\mathrm{h}}^{(1)} - x)\big) \bar{P}^{\alpha^{(1)}} f_{-\Delta^{(1)}|\hat{\mathcal{H}}^{(1)}}(\bar{P}^{\alpha^{(1)}}x \mid \hat{\mathrm{h}}^{(1)})\,\mathrm{d}x. \tag{D.28}$$

Changing the integration variable to $y = \bar{P}^{\alpha^{(1)}}x$ (and thus $\mathrm{d}x = \bar{P}^{-\alpha^{(1)}}\,\mathrm{d}y$) yields

$$f_{\hat{\mathcal{H}}^{(2)}|\hat{\mathcal{H}}^{(1)}}(\hat{\mathrm{h}}^{(2)} \mid \hat{\mathrm{h}}^{(1)}) = \bar{P}^{\alpha^{(2)}} \int_{-\infty}^{\infty} \upsilon_P(y)\,\mathrm{d}y \tag{D.29}$$

where

$$\upsilon_P(y) \triangleq f_{\Delta^{(2)}}\big(\delta^{(2)} - \bar{P}^{\alpha^{(2)}-\alpha^{(1)}}(\delta^{(1)} + y)\big) f_{-\Delta^{(1)}|\hat{\mathcal{H}}^{(1)}}(y|\hat{\mathrm{h}}^{(1)}) \tag{D.30}$$

comes from applying $\hat{\mathrm{h}}^{(i)} = \mathrm{h} + \bar{P}^{-\alpha^{(i)}}\delta^{(i)}$. We have obtained (D.10) of Claim D.2. Let us obtain the limit of $\upsilon_P(y)$ as $P \to \infty$. This limit is directly obtained from continuity of $f_{\Delta^{(1)}}$ and $f_{\Delta^{(2)}}$ and Corollary D.1, and it has two possible expressions depending on the relation between $\alpha^{(1)}$ and $\alpha^{(2)}$. Specifically, it holds that

$$\lim_{P\to\infty} \upsilon_P(y) = f_{\Delta^{(2)}}(\delta^{(2)} - \delta^{(1)} - y) f_{-\Delta^{(1)}}(y) \tag{D.31}$$

if $\alpha^{(1)} = \alpha^{(2)}$, and that

$$\lim_{P\to\infty} \upsilon_P(y) = f_{\Delta^{(2)}}(\delta^{(2)}) f_{-\Delta^{(1)}}(y) \tag{D.32}$$

if $\alpha^{(1)} > \alpha^{(2)}$, what concludes the proof of Claim D.2. ■

D.3 Proof for $K > 2$ estimates

In this section we prove by induction that Lemma 4.1 also holds for any number K of estimates. We have proved that it is true for the base cases $K = 1$ –trivial– and $K = 2$. In the following, we prove the induction step. We denote the set of estimates as $\overline{\mathcal{G}}_{\mathcal{K}} \triangleq \{\hat{\mathcal{H}}^{(1)}, \dots, \hat{\mathcal{H}}^{(K)}\}$ and, consistently, the set of given values as $\overline{g}_{\mathcal{K}} \triangleq \{\hat{\mathrm{h}}^{(1)}, \dots, \hat{\mathrm{h}}^{(K)}\}$. Let us assume that Lemma 4.1 is verified for a given K. We consider $K+1$ estimates. Then, from the mutual independence of the estimation noise variables $\Delta^{(j)}$ and Bayes' formula we obtain that

$$f_{\mathcal{H}|\overline{\mathcal{G}}_{\mathcal{K}},\hat{\mathcal{H}}^{(K+1)}}\big(\mathrm{h} \mid \overline{g}_{\mathcal{K}}, \hat{\mathrm{h}}^{(K+1)}\big) = \underbrace{\frac{f_{\mathcal{H},\overline{\mathcal{G}}_{\mathcal{K}}}\big(\mathrm{h}, \overline{g}_{\mathcal{K}}\big)}{f_{\overline{\mathcal{G}}_{\mathcal{K}}}\big(\overline{g}_{\mathcal{K}}\big)}}_{f_{\mathcal{H}|\overline{\mathcal{G}}_{\mathcal{K}}}(\mathrm{h}|\overline{g}_{\mathcal{K}})} \frac{f_{\bar{P}^{-\alpha^{(K+1)}}\Delta^{(K+1)}}\big(\hat{\mathrm{h}}^{(K+1)} - \mathrm{h}\big)}{f_{\hat{\mathcal{H}}^{(K+1)}|\overline{\mathcal{G}}_{\mathcal{K}}}\big(\hat{\mathrm{h}}^{(K+1)}|\overline{g}_{\mathcal{K}}\big)}. \tag{D.33}$$

From the induction hypothesis, it holds that $\max f_{\mathcal{H}|\overline{\mathcal{G}}_{\mathcal{K}}}(\mathrm{h} \mid \overline{g}_{\mathcal{K}}) = \Theta\big(\bar{P}^{\alpha^{(1)}}\big)$. Thus, we need to prove that

$$0 < \lim_{P\to\infty} \frac{f_{\bar{P}^{-\alpha^{(K+1)}}\Delta^{(K+1)}}\big(\hat{\mathrm{h}}^{(K+1)} - \mathrm{h}\big)}{f_{\hat{\mathcal{H}}^{(K+1)}|\overline{\mathcal{G}}_{\mathcal{K}}}\big(\hat{\mathrm{h}}^{(K+1)}|\overline{g}_{\mathcal{K}}\big)} < \infty. \tag{D.34}$$

Let us denote $\Delta' = \bar{P}^{-\alpha^{(K+1)}}\Delta^{(K+1)} - \bar{P}^{-\alpha^{(1)}}\Delta^{(1)}$. From the equivalence

$$\hat{\mathrm{h}}^{(K+1)} - \hat{\mathrm{h}}^{(1)} = \bar{P}^{-\alpha^{(K+1)}}\delta^{(K+1)} - \bar{P}^{-\alpha^{(1)}}\delta^{(1)}, \tag{D.35}$$

the denominator of the expression in (D.34) can be rewritten as

$$\begin{aligned} f_{\hat{\mathcal{H}}^{(K+1)}|\overline{\mathcal{G}}_{\mathcal{K}}}\big(\hat{\mathrm{h}}^{(K+1)}|\overline{g}_{\mathcal{K}}\big) &= f_{\hat{\mathcal{H}}^{(1)}+\Delta'|\overline{\mathcal{G}}_{\mathcal{K}}}\big(\hat{\mathrm{h}}^{(K+1)} - \hat{\mathrm{h}}^{(1)} + \hat{\mathrm{h}}^{(1)} \mid \overline{g}_{\mathcal{K}}\big) \\ &= f_{\Delta'|\overline{\mathcal{G}}_{\mathcal{K}}}\big(\bar{P}^{-\alpha^{(K+1)}}\delta^{(K+1)} - \bar{P}^{-\alpha^{(1)}}\delta^{(1)} \mid \overline{g}_{\mathcal{K}}\big). \end{aligned} \tag{D.36}$$

Hence, expressing $f_{\Delta'|\overline{\mathcal{G}}_{\mathcal{K}}}$ as convolution of PDFs yields

$$f_{\hat{\mathcal{H}}^{(K+1)}|\overline{\mathcal{G}}_{\mathcal{K}}}\big(\hat{\mathrm{h}}^{(K+1)}|\overline{g}_{\mathcal{K}}\big) = f_{\bar{P}^{-\alpha^{(K+1)}}\Delta^{(K+1)}} * f_{-\bar{P}^{-\alpha^{(1)}}\Delta^{(1)}|\overline{\mathcal{G}}_{\mathcal{K}}}\big(\hat{\mathrm{h}}^{(K+1)} - \hat{\mathrm{h}}^{(1)}|\overline{g}_{\mathcal{K}}\big). \tag{D.37}$$

Let us introduce the notation $\delta'_y \triangleq \delta^{(K+1)} - \bar{P}^{\alpha^{(K+1)}-\alpha^{(1)}}(\delta^{(1)} + y)$ for ease of reading. Thus, by applying the same steps as in (D.26)-(D.30), we obtain

$$\frac{f_{\bar{P}^{-\alpha^{(K+1)}}\Delta^{(K+1)}}\big(\hat{\mathrm{h}}^{(K+1)} - \mathrm{h}\big)}{f_{\hat{\mathcal{H}}^{(K+1)}|\overline{\mathcal{G}}_{\mathcal{K}}}\big(\hat{\mathrm{h}}^{(K+1)}|\overline{g}_{\mathcal{K}}\big)} = \frac{f_{\Delta^{(K+1)}}\big(\delta^{(K+1)}\big)}{\int_{-\infty}^{\infty} f_{\Delta^{(K+1)}}\big(\delta'_y\big) f_{-\Delta^{(1)}|\overline{\mathcal{G}}_{\mathcal{K}}}\big(y|\overline{g}_{\mathcal{K}}\big)\,\mathrm{d}y}, \tag{D.38}$$

We can see that (D.38) is equivalent to (D.15) with $\Delta^{(K+1)}$ in place of $\Delta^{(2)}$. Then, by following the same derivation as in the $K = 2$ case, i.e., using Corollary D.1 and Lebesgue's Dominated Convergence Theorem, we conclude the induction step. From the base case and the induction step, Lemma 4.1 is proven. ■

D.4 Proof of Lemma D.1

Suppose that $\alpha^{(j)} > 0$. Then,

$$\begin{aligned}
\lim_{P\to\infty} f_{\hat{\mathcal{H}}^{(j)}}(\hat{\mathrm{h}}^{(j)}) &= \lim_{P\to\infty} f_{\mathcal{H}+\bar{P}^{-\alpha^{(j)}}\Delta^{(j)}}(\mathrm{h}+\bar{P}^{-\alpha^{(j)}}\delta^{(j)}) \\
&= \lim_{P\to\infty} \int_{-\infty}^{\infty} f_{\mathcal{H}}(\mathrm{h}+\bar{P}^{-\alpha^{(j)}}\delta^{(j)}-x) f_{\bar{P}^{-\alpha^{(j)}}\Delta^{(j)}}(x)\,\mathrm{d}x \\
&\overset{(a)}{=} \lim_{P\to\infty} \int_{-\infty}^{\infty} f_{\mathcal{H}}(\mathrm{h}+\bar{P}^{-\alpha^{(j)}}\delta^{(j)}-\bar{P}^{-\alpha^{(j)}}y) f_{\Delta^{(j)}}(y)\,\mathrm{d}y \\
&\overset{(b)}{=} \int_{-\infty}^{\infty} f_{\mathcal{H}}(\mathrm{h}) f_{\Delta^{(j)}}(y)\,\mathrm{d}y \\
&= f_{\mathcal{H}}(\mathrm{h}),
\end{aligned} \tag{D.39}$$

where (a) comes from applying Proposition D.1 to express $f_{\bar{P}^{-\alpha^{(j)}}\Delta^{(j)}}$ as function of $f_{\Delta^{(j)}}$, and from the change of integration variable $y = \bar{P}^{\alpha^{(j)}}x$. Finally, (b) follows from applying Lebesgue's Dominated Convergence Theorem. Hence, $f_{\hat{\mathcal{H}}^{(j)}}$ converges almost surely to $f_{\mathcal{H}}$ and hence Lemma D.1 is proven. In order to prove Corollary D.1, i.e., that

$$\lim_{P\to\infty} f_{\Delta^{(1)}|\hat{\mathcal{H}}^{(1)},\ldots,\hat{\mathcal{H}}^{(K)}}(y|\hat{\mathrm{h}}^{(1)},\ldots,\hat{\mathrm{h}}^{(K)}) = f_{\Delta^{(1)}}(y), \tag{D.40}$$

we apply Bayes' formula such that

$$\begin{aligned}
f_{\Delta^{(1)}|\hat{\mathcal{H}}^{(1)},\ldots,\hat{\mathcal{H}}^{(K)}}(y|\hat{\mathrm{h}}^{(1)},\ldots,\hat{\mathrm{h}}^{(K)}) &= \frac{f_{\hat{\mathcal{H}}^{(1)},\ldots,\hat{\mathcal{H}}^{(K)}|\Delta^{(1)}}(\hat{\mathrm{h}}^{(1)},\ldots,\hat{\mathrm{h}}^{(K)}|y) f_{\Delta^{(1)}}(y)}{f_{\hat{\mathcal{H}}^{(1)},\ldots,\hat{\mathcal{H}}^{(K)}}(\hat{\mathrm{h}}^{(1)},\ldots,\hat{\mathrm{h}}^{(K)})} \\
&= \frac{f_{\mathcal{H},\hat{\mathcal{H}}^{(2)},\ldots,\hat{\mathcal{H}}^{(K)}}(\mathrm{h},\hat{\mathrm{h}}^{(2)},\ldots,\hat{\mathrm{h}}^{(K)})}{f_{\hat{\mathcal{H}}^{(1)},\ldots,\hat{\mathcal{H}}^{(K)}}(\hat{\mathrm{h}}^{(1)},\ldots,\hat{\mathrm{h}}^{(K)})} f_{\Delta^{(1)}}(y).
\end{aligned} \tag{D.41}$$

From (D.39) and the fact that $\alpha^{(1)} > 0$ we obtain that

$$\lim_{P\to\infty} \frac{f_{\mathcal{H},\hat{\mathcal{H}}^{(2)},\ldots,\hat{\mathcal{H}}^{(K)}}(\mathrm{h},\hat{\mathrm{h}}^{(2)},\ldots,\hat{\mathrm{h}}^{(K)})}{f_{\hat{\mathcal{H}}^{(1)},\ldots,\hat{\mathcal{H}}^{(K)}}(\hat{\mathrm{h}}^{(1)},\ldots,\hat{\mathrm{h}}^{(K)})} = 1. \tag{D.42}$$

From (D.42), the limit of (D.41) as $P \to \infty$ is obtained as

$$\lim_{P\to\infty} f_{\Delta^{(1)}|\hat{\mathcal{H}}^{(1)},\ldots,\hat{\mathcal{H}}^{(K)}}(y|\hat{\mathrm{h}}^{(1)},\ldots,\hat{\mathrm{h}}^{(K)}) = f_{\Delta^{(1)}}(y), \tag{D.43}$$

what concludes the proof. ■

Appendix E

CSI Allocation for the Distributed CSIT Setting

Let us introduce a value A such that the sum of accuracy scaling coefficients α satisfies that

$$\sum_{j=1}^{K} \alpha^{(j)} = A. \tag{E.1}$$

We can write (4.51) of Lemma 4.3 as $\max(\mathrm{DoF}^{\max}_{\mathrm{single}}, \mathrm{DoF}_{\mathrm{2ph}})$, where the $\mathrm{DoF}^{\max}_{\mathrm{single}}$ is the maximum DoF obtained by using only one Transmission Mode and $\mathrm{DoF}_{\mathrm{2ph}}$ the maximum DoF with two Transmission Modes. Specifically, $\mathrm{DoF}^{\max}_{\mathrm{single}} = \max_k \mathrm{DoF}^{k}_{\mathrm{single}}$ and $\mathrm{DoF}^{k}_{\mathrm{single}}$

$$\mathrm{DoF}^{k}_{\mathrm{single}} = \underset{n,\ \boldsymbol{\alpha}}{\text{maximize}}\ 1 + (k-1)\alpha^{(n)} \tag{E.2}$$

$$\text{subject to}\ \ n \in \{1, \dots, k-1\}, \tag{E.3}$$

$$d_{n,k} \geq 0, \quad \sum_{j=1}^{k-1} \alpha^{(j)} = A. \tag{E.4}$$

Consider now the one-mode case. Therefore, the best allocation is always to give equal level of CSIT to the n active TXs and no CSIT to the other $k-n$ TXs. Then,

$$\alpha^{(j)} = \begin{cases} \frac{A}{n} & \text{if } j \leq n \\ 0 & \text{if } j > n \end{cases} \tag{E.5}$$

Moreover, from $d_{n,k} \geq 0$ and $\alpha^{(n)} = \dfrac{A}{n}$ we have that

$$1 - \alpha^{(n)} - k(k-n-1)\alpha^{(n)} \geq 0 \quad \Rightarrow \quad n \geq \frac{A + k(k-1)A}{kA+1} \tag{E.6}$$

From(E.6), $\mathrm{DoF}^{k}_{\mathrm{single}}$ is obtained as

$$\mathrm{DoF}^{k}_{\mathrm{single}} = \begin{cases} k & \text{if } A \geq k-1 \\ 1 + (k-1)\frac{A}{\left\lceil \frac{A+k(k-1)A}{kA+1} \right\rceil} & \text{if } A \leq k-1 \end{cases} \tag{E.7}$$

And therefore $\mathrm{DoF}^{\max}_{\mathrm{single}}$ is found by choosing the biggest constant among $K-1$ possible choices. On the other hand, we have that $\mathrm{DoF}_{\mathrm{2ph}}$ is given by

$$\begin{aligned} \mathrm{DoF}_{\mathrm{2ph}} = \underset{\substack{k_1,n_1,\\ k_2,n_2,\\ \boldsymbol{\alpha}}}{\text{maximize}} \; & 1 + \frac{d_{n_2,k_2}}{d_{n_2,k_2} - d_{n_1,k_1}}(k_1-1)\alpha^{(n_1)} - \frac{d_{n_1,k_1}}{d_{n_2,k_2} - d_{n_1,k_1}}(k_2-1)\alpha^{(n_2)} \\ \text{subject to} \; & k_1, k_2 \in \mathbb{N}_K, \\ & n_1 \in \mathbb{N}_{k_1} \mid d_{n_1,k_1} \leq 0, \\ & n_2 \in \mathbb{N}_{k_2} \mid d_{n_2,k_2} \geq 0, \\ & \sum_{j=1}^{K-1} \alpha^{(j)} = A. \end{aligned} \tag{E.8}$$

where $d_{n,k} \triangleq \quad 1 - \alpha^{(n)} - k(k-n-1)\alpha^{(n)}$.

We present in the following some particular cases of Lemma 4.3. We have obtained the maximum DoF for a given budget A, for any possible allocation $\boldsymbol{\alpha}$. For this, we sample the continuous value of $\alpha^{(j)}$ with a *precision*, such that a precision=1/100 means that we compute the DoF for all $\alpha^{(j)} = \frac{i}{\text{precision}}$ such that

$$\forall i \in \mathbb{N} \mid \alpha^{(j)}_{\min} < \frac{i}{\text{precision}} < \alpha^{(j)}_{\max}, \tag{E.9}$$

where $\alpha^{(j)}_{\min}, \alpha^{(j)}_{\max}$ are the extreme values of $\alpha^{(j)}$ for each case (given $\alpha^{(\ell<j)}, n, k, K, A...$).

Due to the multidimensional nature of the problem –there exist K different $\alpha^{(j)}$ coefficients–, it is not simple to choose a good way of representing all the possible cases. We have chosen to do it by hierarchy of TXs: The axis x is ordered such as, for any possible value of $\alpha^{(1)}$, we draw all the possible values of $\alpha^{(2)}$ in increasing order. Again, for any of those values of $\alpha^{(2)}$, we draw in increasing order all the possible values of $\alpha^{(3)}$, and so on. This is represented in Fig. E.1.

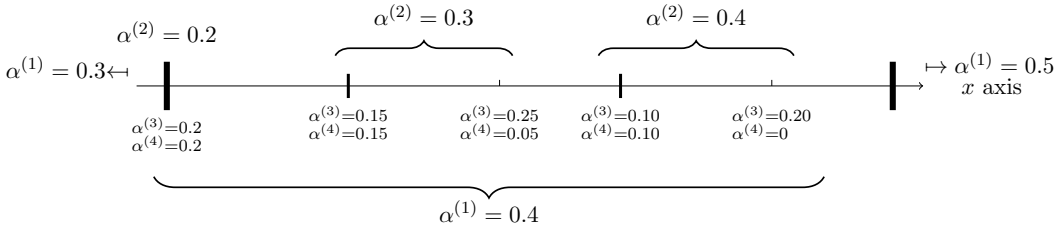


Figure E.1 – Example of the defined axis x for a total CSIT of $A = 1$.

We have simulated 4 possible cases:

1. $K = 4$, $A = 0.5$, precision $= 1/100$
2. $K = 4$, $A = 1.5$, precision $= 1/100$
3. $K = 4$, $A = 2.5$, precision $= 1/100$
4. $K = 5$, $A = 2.5$, precision $= 1/50$

In Fig. E.2 we present the first case ($K = 4,\ A = 0.5,\ precision = 1/100$). In this picture, as well as in the others, we plot the DoF value in blue, surrounded by the upper envelope in green (given by the local maxima) and the lower envelope in red (given by the local minima). Moreover, we plot in vertical orange lines the values of the horizontal x where $\alpha^{(1)}$ increases. Finally, we plot the total square-difference $\sum_{i=1}^{K-2}(\alpha^{(i)} - \alpha^{(i+1)})^2$.

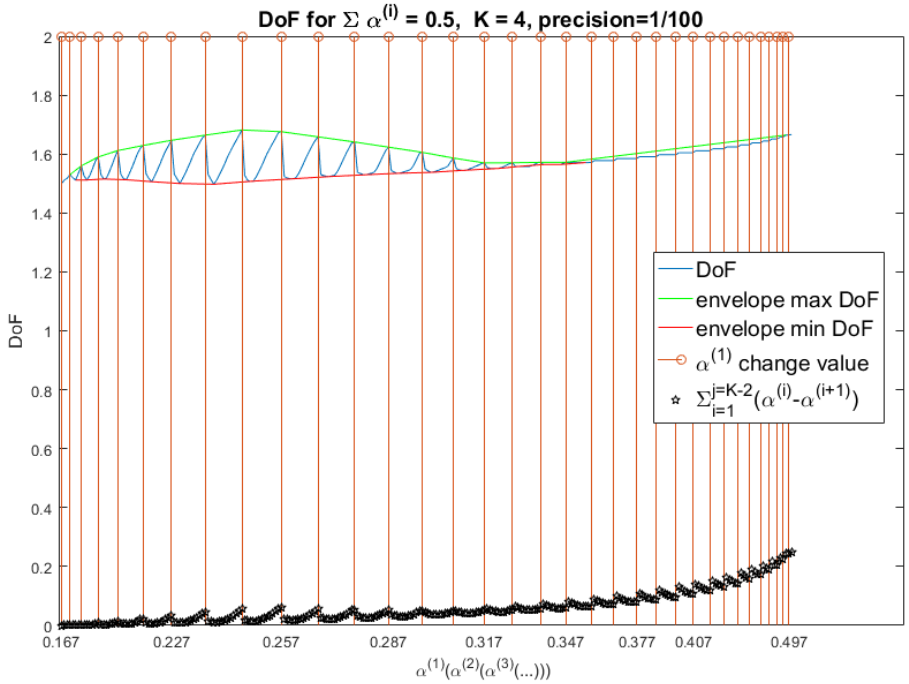


Figure E.2 – DoF for the case $K = 4$, $A = 0.5$, precision $= 1/100$

Fig. E.2 shows that the DoF is cyclic for every value of $\alpha^{(1)}$: For each $\alpha^{(1)}$ we obtain the maximum DoF just before it gets the new value. This means –from our definition of the axis x in Fig. E.1– that the maximum is obtained when $\alpha^{(2)}$ gets the maximum value. We see the same cyclic pattern in Fig. E.3 ($A = 1.5$), and Fig. E.4 ($A = 2.5$), but while in Fig. E.3 the local maximum is also at the end of the cycle, for Fig. E.4 the maximum is at the beginning, i.e., for smallest values of $\alpha^{(2)}$.

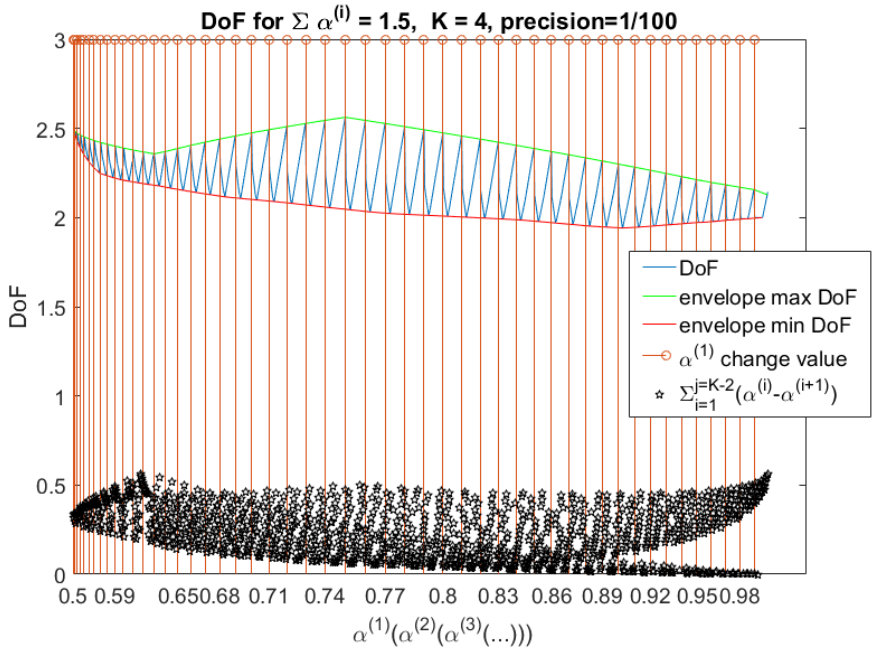


Figure E.3 – DoF for the case $K = 4$, $A = 1.5$, precision $= 1/100$

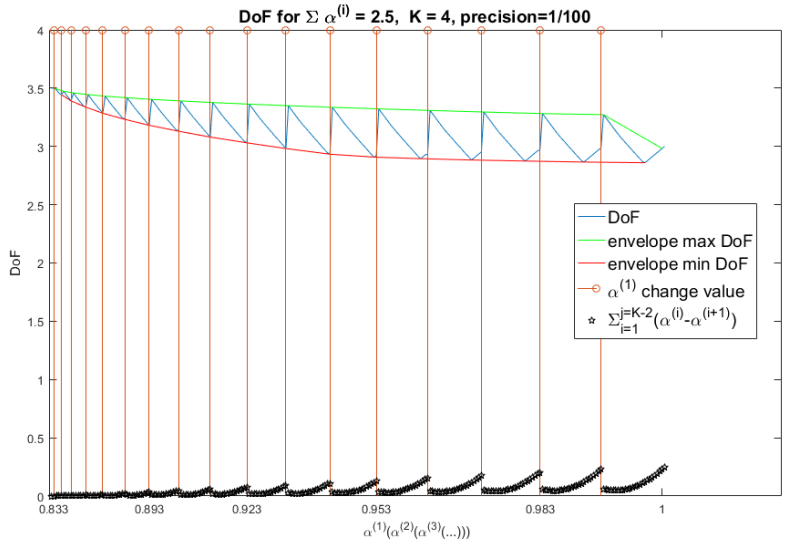


Figure E.4 – DoF for the case $K = 4$, $A = 2.5$, precision $= 1/100$

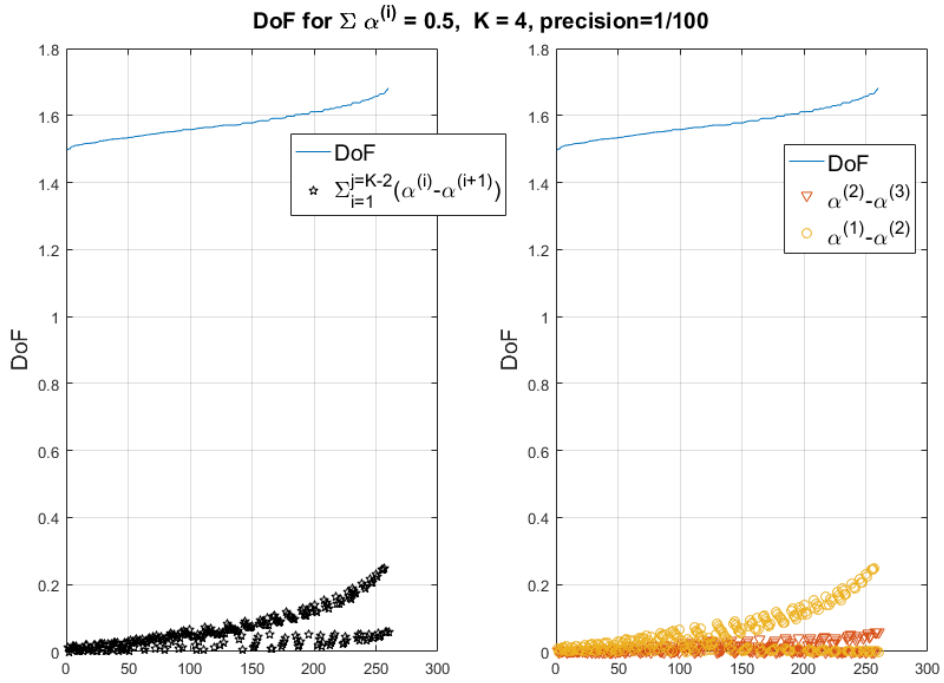


Figure E.5 – Sorted DoF for the case $K = 4$, $A = 0.5$, precision $= 1/100$

We observe that there exists a pattern in the difference between $\alpha^{(j)}$. To analyze this behavior, we depict the sorted DoF values together with $\sum_{i=1}^{K-2}(\alpha^{(i)} - \alpha^{(i+1)})^2$. This representation is shown for the $K = 4$ case in Fig. E.5 ($A = 0.5$), Fig. E.6 ($A = 1.5$), and Fig. E.7 ($A = 2.5$). We show two graphs in each of the figures: One depicting the sum $\sum_{i=1}^{K-2}(\alpha^{(i)} - \alpha^{(i+1)})^2$, and the other with the decoupled differences between successive $\alpha^{(i)}$, i.e., $(\alpha^{(1)} - \alpha^{(2)})^2$ and $(\alpha^{(2)} - \alpha^{(3)})^2$. It is interesting to note that, as the DoF grows, each difference term tends to either increase or decrease. In Fig. E.6 ($A = 1.5$) we can easily see that, for the cases with higher DoF, $(\alpha^{(1)} - \alpha^{(2)})^2 \to 0$ while $(\alpha^{(2)} - \alpha^{(3)})^2 \to (0.75)^2$. This behavior implies that the optimal allocation is

$$\begin{aligned} \alpha^{(1)} = \alpha^{(2)} &= 0.75, \\ \alpha^{(3)} = \alpha^{(4)} &= 0. \end{aligned}$$

In Fig. E.7 ($A = 2.5$) this behavior is not so clear because the quantity of CSIT is too big compared with the number of TXs, i.e.,

$$\begin{aligned} 0.833 \leq \alpha^{(1)} &\leq 1, \\ 0.75 \leq \alpha^{(2)} &\leq 1, \\ 0.5 \leq \alpha^{(3)} &\leq 0.833. \end{aligned}$$

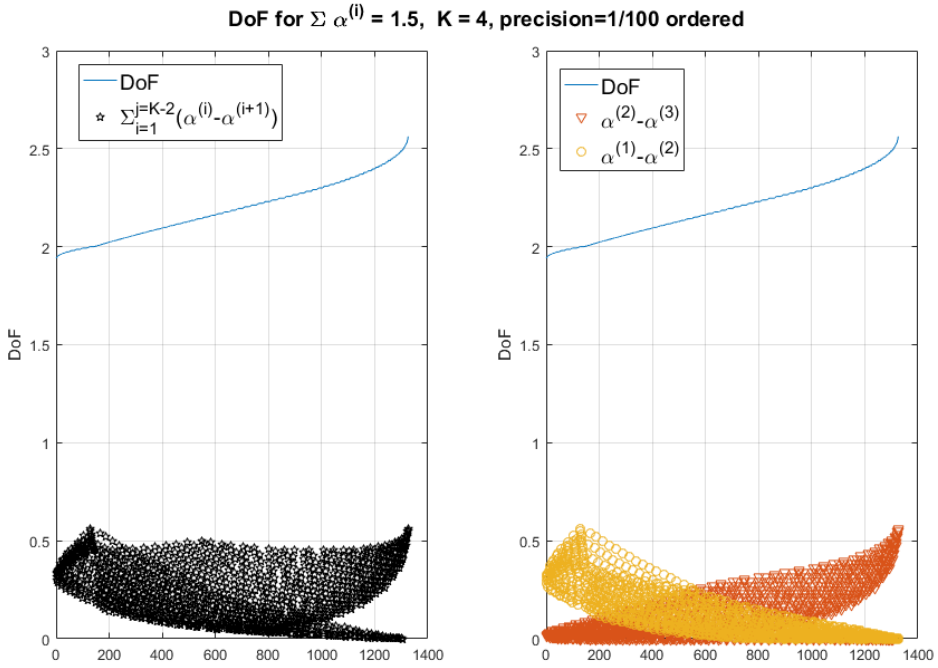


Figure E.6 – DoF for the case $K = 4$, $A = 1.5$, precision $= 1/100$

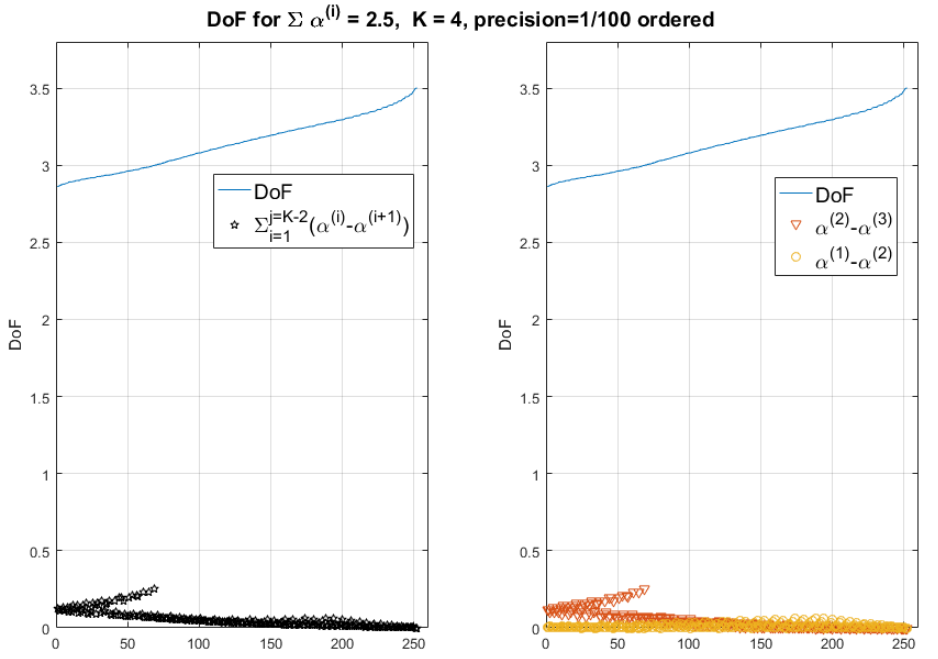


Figure E.7 – Sorted DoF for the case $K = 4$, $A = 2.5$, precision $= 1/100$

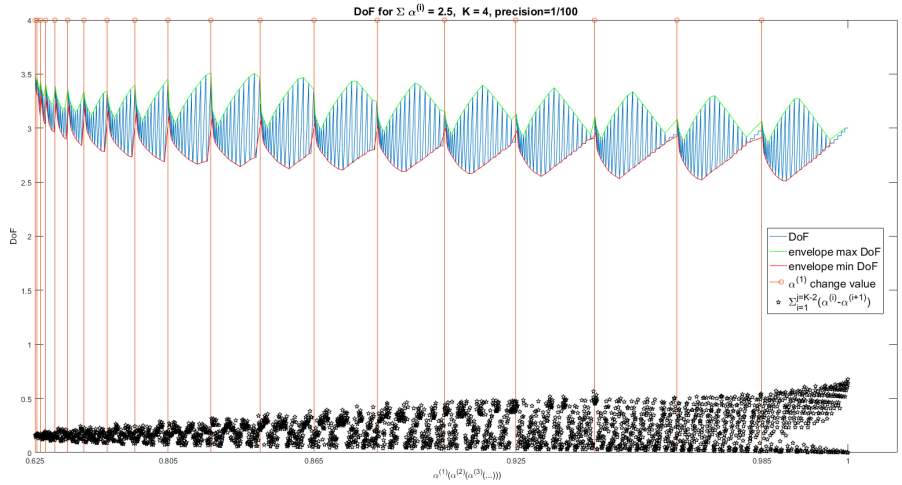


Figure E.8 – DoF for the case $K = 5$, $A = 2.5$.

In Fig. E.8 and Fig. E.9, we show the results for $K = 5$, $A = 2.5$. We observe the same insights:

- Fig. E.8: There exists a cyclic pattern for each value of $\alpha^{(1)}$. In this setting, since we have more TXs, there are sub-cycles, one for each value of $\alpha^{(2)}$. We conjecture that this holds for all the $\alpha^{(j)}$s with $j < K - 2$.
- Fig. E.9: We observe the same *divergence* of differences. Moreover, the difference that increases as the DoF in this case is $(\alpha^{(3)} - \alpha^{(4)})^2$. This implies that the best performance is obtained when the first three TXs obtain a CSI with the same accuracy and the other two TXs do not receive any CSI.

Main insights

We have seen that the allocation of CSIT for a given budget is a complex optimization. However, some interesting patterns turn out from the analysis. The behavior of difference between values of $\alpha^{(j)}$ is probably due to the proposed structure. Since Corollary 4.1 implies that the maximum DoF is obtained with only two Transmission Modes, it holds that the best strategy is to allocate the CSIT such that we have a *two-steps* staggered CSIT allocation; first n_1 active TXs, second n_2 active TXs and passive TXs with $\alpha^{(p)} = 0$.

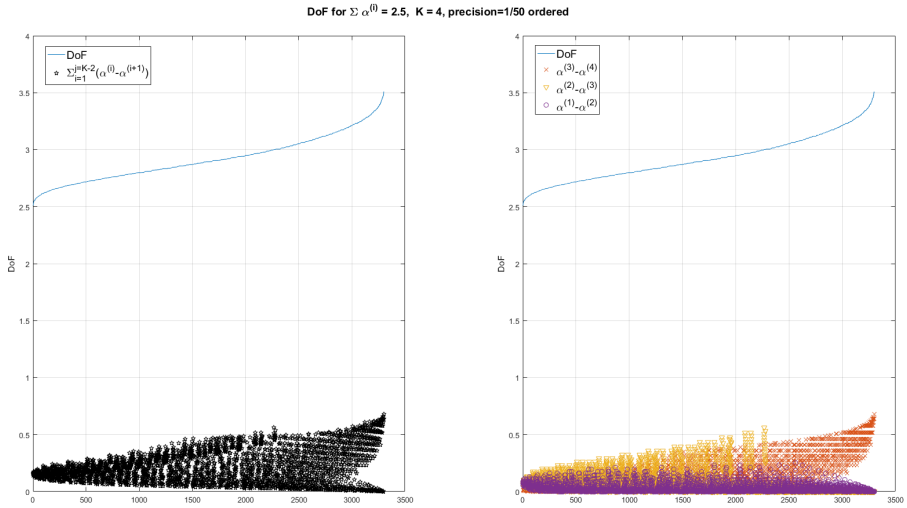


Figure E.9 – Sorted DoF for the case $K = 5$, $A = 2.5$.

Appendix F

Asymptotic Properties of Quantization: Proof of Lemma 6.1

In the following, we prove Lemma 6.1, i.e., that for any $c_q > 1$ the quantizer

$$\mathcal{Q}_u(x) \triangleq \bar{P}^{\frac{-\alpha_{\min}}{c_q}} \left\lfloor \bar{P}^{\frac{\alpha_{\min}}{c_q}} x \right\rfloor \tag{F.1}$$

satisfies properties (P0), (P1) and (P2). We first prove property (P1). Afterward, we demonstrate (P2) and finally (P0). We define α_q as $\alpha_q \triangleq \frac{\alpha_{\min}}{c_q}$ so as to simplify the notation. The quantization step is then $q = \bar{P}^{-\alpha_q}$.

F.1 Proof of (P1): Convergence

In order to prove that $\mathcal{Q}_u$ satisfies (P1), i.e., that

$$\lim_{P\to\infty} \mathcal{Q}_u(\lambda_i^{(j)}) - \lambda_i^\star = 0 \quad \text{a.s.} \quad \forall i, j \in \mathbb{N}_2, \tag{F.2}$$

we demonstrate (F.2) for $j = 2$, as the case with $j = 1$ is straightforwardly proved following the same derivation. Let $\mathbf{v}_{\mathbf{H}}^{(j)} \in \mathbb{R}^{8\times 1}$ be the column vector obtained by stacking the real and imaginary parts of the elements of $\hat{\mathbf{H}}^{(j)}$ one on top of another, such that

$$\mathbf{v}_{\mathbf{H}}^{(j)} = \begin{bmatrix} \mathrm{Re}\left(\hat{\mathrm{h}}_{1,1}^{(j)}\right) \\ \mathrm{Im}\left(\hat{\mathrm{h}}_{1,1}^{(j)}\right) \\ \cdots \\ \mathrm{Im}\left(\hat{\mathrm{h}}_{2,2}^{(j)}\right) \end{bmatrix}, \tag{F.3}$$

where $\mathrm{Re}(x)$ and $\mathrm{Im}(x)$ denote the real and imaginary part of $x \in \mathbb{C}$, respectively. Let $\mathbf{v}_{\mathbf{H}}^{\star} \in \mathbb{R}^{8\times 1}$ be the analogous expanded vector for the genie-aided best-estimate channel $\hat{\mathbf{H}}^{\star}$. We recall the definition of λ_i as a function in (6.21), and we rewrite it in terms of $\mathbf{v}_{\mathbf{H}}^{\star}$ as $\lambda_i \triangleq \Lambda_i(\mathbf{v}_{\mathbf{H}}^{\star})$. Therefore, consider the Taylor's expansion of $\lambda_i^{(2)}$ centered in $\lambda_i^{\star}$. We can write that

$$\lambda_i^{(2)} - \lambda_i^{\star} = \left(\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{\star}\right)^{\mathrm{T}} \nabla\Lambda_i\left(\mathbf{v}_{\mathbf{H}}^{\star}\right) + o\left(\|\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{\star}\|\right), \tag{F.4}$$

where $\left(\nabla\Lambda_i(\cdot)\right)^{\mathrm{T}}$ is the i-th row of the Jacobian Matrix $\mathbf{J}_{\Lambda}$ introduced in Section 6.3.2. Let ϑ be defined as

$$\vartheta \triangleq \lambda_i^{(2)} - \lambda_i^{\star}. \tag{F.5}$$

From (F.5) and the definition of $\mathcal{Q}_u$ in (F.1), it follows that

$$\mathcal{Q}_u(\lambda_i^{(2)}) - \lambda_i^{\star} = \bar{P}^{-\alpha_q} \left\lfloor \bar{P}^{\alpha_q}(\lambda_i^{\star} + \vartheta) \right\rfloor - \lambda_i^{\star}. \tag{F.6}$$

Since for any $c \in \mathbb{R}^{+}$ it holds that $c\left\lfloor \frac{1}{c}(x+y)\right\rfloor - x \leq y$, we obtain that

$$\mathcal{Q}_u(\lambda_i^{(2)}) - \lambda_i^{\star} \leq \vartheta. \tag{F.7}$$

Similarly, since for any $c \in \mathbb{R}^{+}$ it holds that $c\left\lfloor \frac{1}{c}(x+y)\right\rfloor - x \geq c\left\lfloor \frac{y}{c}\right\rfloor \geq y - c$, we can bound (F.6) from below as

$$\mathcal{Q}_u(\lambda_i^{(2)}) - \lambda_i^{\star} \geq \vartheta - \bar{P}^{-\alpha_q}. \tag{F.8}$$

From (F.7) and (F.8), it is sufficient to prove that

$$\lim_{P\to\infty} \vartheta = 0 \qquad \text{a.s.} \tag{F.9}$$

to demonstrate that $\lim_{P\to\infty} \mathcal{Q}_u(\lambda_i^{(2)}) = \lambda_i^{\star}$ almost surely. To do so, we make use of the following lemma, whose proof is relegated to Appendix F.4.

Lemma F.1. *Let $\alpha_{i,k}^{(j)} > 0$ for any $i, j, k \in \mathbb{N}_2$. Then, it holds that*

$$\lim_{P\to\infty} \|\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{\star}\| = 0 \qquad \textit{a.s.} \tag{F.10}$$

Since we assume that $\|\nabla\Lambda_i\| \leq \|\mathbf{J}_{\Lambda}\| \leq M_{\mathbf{J}}$, –see Section 6.3.2–, it holds that

$$|\vartheta| \leq \|\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{\star}\| M_{\mathbf{J}} + |o\left(\|\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{\star}\|\right)|. \tag{F.11}$$

Thus, from Lemma F.1 we obtain that $\lim_{P\to\infty} \vartheta = 0$ almost surely. Consequently, $\mathcal{Q}_u$ satisfies (P1). ■

F.2 Proof of (P2): Probability of Agreement

We want to prove that $\mathcal{Q}_u$ satisfies

$$\Pr\left(\Omega^{\mathsf{c}}\right) = o\Big(\frac{1}{\log_2(P)}\Big), \tag{F.12}$$

where $\Pr\left(\Omega^{\mathsf{c}}\right) = 1 - \Pr\Big(\forall i \in \mathbb{N}_2,\ \mathcal{Q}\big(\lambda_i^{(1)}\big) = \mathcal{Q}\big(\lambda_i^{(2)}\big) \in \mathbb{R}^+\Big)$. Note that, for any two events A, B, it holds that

$$1 - \Pr(A \wedge B) \leq 1 - \Pr(A) + 1 - \Pr(B). \tag{F.13}$$

Suppose w.l.o.g. that the probability of agreement is smaller for λ_1 than for λ_2. Therefore,

$$\Pr\left(\Omega^{\mathsf{c}}\right) \leq 2\left(1 - \Pr\left(\mathcal{Q}_u(\lambda_1^{(1)}) = \mathcal{Q}_u(\lambda_1^{(2)}) \in \mathbb{R}^+\right)\right). \tag{F.14}$$

Moreover, it holds that

$$\begin{aligned} 1 - \Pr\Big(\mathcal{Q}_u(\lambda_1^{(1)}) = \mathcal{Q}_u(\lambda_1^{(2)}) \in \mathbb{R}^+\Big) \leq{}& \Pr\left(\mathcal{Q}_u(\lambda_1^{(1)}) \neq \mathcal{Q}_u(\lambda_1^{(2)})\right) \\ &+ \Pr\left(\mathcal{Q}_u(\lambda_1^{(1)}) = 0\right). \end{aligned} \tag{F.15}$$

Consider the last term of (F.15). It follows that

$$\begin{aligned} \Pr\left(\mathcal{Q}_u(\lambda_1^{(1)}) = 0\right) &\overset{(a)}{=} \Pr\left(\lambda_1^{(1)} \leq \bar{P}^{-\alpha_q}\right) \\ &\overset{(b)}{\leq} f_{\Lambda_1}^{\max} \bar{P}^{-\alpha_q}, \end{aligned} \tag{F.16}$$

where (a) follows from the quantization step size of $\mathcal{Q}_u$, and (b) follows from the bounded density assumption of (6.22). This leads to

$$\Pr\Big(\mathcal{Q}_u(\lambda_1^{(1)}) = 0\Big) = o\left(\frac{1}{\log_2(P)}\right). \tag{F.17}$$

Consider now the other term of (F.15), the probability of disagreement $\Pr\left(\mathcal{Q}_u(\lambda_1^{(1)}) \neq \mathcal{Q}_u(\lambda_1^{(2)})\right)$. In order to prove that this probability is $o(1/\log_2(P))$, we need to introduce first some notation related to the quantization levels.

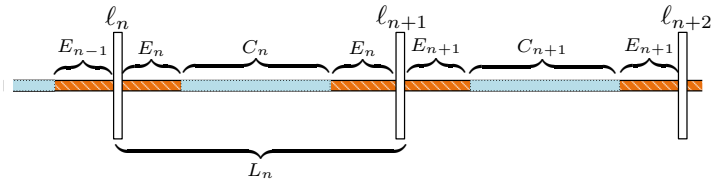


Figure F.1 – Illustration of a reconstruction level L_n of the quantizer and the two sub-areas in which we divide it: The central area C_n and the edge area E_n.

F.2.1 Edge and Center of the Reconstruction Level

Let ℓ_n be the n-th quantization level of $\mathcal{Q}_u$, $n \in \mathbb{N}_N$ with $N = \left\lceil \bar{P}^{\alpha_q} \right\rceil$. We assume that $\bar{P}^{\alpha_q} \in \mathbb{N}$ in order to ease the notation, although the result holds for any $\bar{P}^{\alpha_q} \in \mathbb{R}$. Let us define L_n as the input interval that outputs ℓ_n, i.e.,

$$L_n \triangleq \{x \mid \mathcal{Q}_u(x) = \ell_n\}. \tag{F.18}$$

L_n has a range $[L_n^{\min}, L_n^{\max})$ such that $|L_n| \triangleq L_n^{\max} - L_n^{\min} = \bar{P}^{-\alpha_q}$ (and $\ell_{N+1} = 1$). We split L_n in two areas, the *edge* area E_n and the *center* area C_n, as depicted in Fig. F.1. The *edge* area is defined as the part of L_n that is at most at distance $\bar{P}^{-c_e \alpha_q}$ of the boundary of the cell, with $c_e > 1$. Thus,

$$E_n \triangleq \left\{ x \in L_n \mid x - L_n^{\min} < \bar{P}^{-c_e \alpha_q} \ \vee \ L_n^{\max} - x < \bar{P}^{-c_e \alpha_q} \right\}. \tag{F.19}$$

The *center* area hence is given by $C_n \triangleq \{x \in L_n \backslash E_n\}$. Intuitively, the probability of disagreement is high if one estimate lies in the edge area E_n, whereas this probability vanishes in the central area C_n. Mathematically, we have that

$$\begin{aligned} \Pr\left(\mathcal{Q}_u(\lambda_1^{(1)}) \neq \mathcal{Q}_u(\lambda_1^{(2)})\right) \leq \Pr\left(\lambda_1^{(1)} \in \bigcup_{n \in \mathbb{N}_N} E_n\right) \\ + \Pr\left(\mathcal{Q}_u(\lambda_1^{(1)}) \neq \mathcal{Q}_u(\lambda_1^{(2)}) \mid \lambda_1^{(1)} \in \bigcup_{n \in \mathbb{N}_N} C_n\right). \end{aligned} \tag{F.20}$$

Let us analyze separately the two probabilities in the RHS of (F.20).

F.2.2 Probability of Belonging to the Edge Area

Consider an arbitrary quantization level ℓ_n. From the bounded density assumption of (6.22), the probability that a computed value $\lambda_1^{(1)}$ is in E_n is

$$\begin{aligned} \Pr\left(\lambda_1^{(1)} \in E_n\right) &\leq f_{\Lambda_1}^{\max} |E_n| \\ &= f_{\Lambda_1}^{\max} 2 \bar{P}^{-c_e \alpha_{\min}}, \end{aligned} \tag{F.21}$$

where $|E_n|$ denotes the length of E_n. Since there are $N = \bar{P}^{\alpha_q}$ cells, the probability of being in the *edge* of any cell is

$$\begin{aligned}\Pr\left(\lambda_1^{(1)} \in \bigcup_{n\in\mathbb{N}_N} E_n\right) &\leq \bar{P}^{\alpha_q} f_{\Lambda_1}^{\max} 2\bar{P}^{-c_e\alpha_q} \\ &= 2f_{\Lambda_1}^{\max}\bar{P}^{(1-c_e)\alpha_q} \\ &= o\left(\frac{1}{\log_2(P)}\right).\end{aligned} \tag{F.22}$$

F.2.3 Probability of Disagreement in the Center Area

Let us focus now on the probability of disagreement in C_n. The minimum distance from any point of C_n to the border of L_n is $\bar{P}^{-c_e\alpha_q}$. Therefore,

$$\Pr\left(\mathcal{Q}_u(\lambda_1^{(1)}) \neq \mathcal{Q}_u(\lambda_1^{(2)})\middle|\lambda_1^{(1)} \in \bigcup_{n\in\mathbb{N}_N} C_n\right) \leq \Pr\left(\left|\lambda_1^{(1)} - \lambda_1^{(2)}\right| \geq \bar{P}^{-c_e\alpha_q}\middle|\lambda_1^{(1)} \in \bigcup_{n\in\mathbb{N}_N} C_n\right).$$

Given that, for two events A, C, $\Pr(A \mid C) \leq \Pr(A)/\Pr(C)$, it follows that

$$\begin{aligned}\Pr\left(\left|\lambda_1^{(1)} - \lambda_1^{(2)}\right| \geq \bar{P}^{-c_e\alpha_q} \;\middle|\; \lambda_1^{(1)} \in \bigcup_{n\in\mathbb{N}_N} C_n\right) &\leq \frac{\Pr\left(|\lambda_1^{(1)} - \lambda_1^{(2)}| \geq \bar{P}^{-c_e\alpha_q}\right)}{\Pr\left(\lambda_1^{(1)} \in \bigcup_{n\in\mathbb{N}_N} C_n\right)} \\ &\overset{(a)}{\leq} \frac{1}{1 - 2f_{\Lambda_1}^{\max}\bar{P}^{(1-c_e)\alpha_q}} \Pr\left(|\lambda_1^{(1)} - \lambda_1^{(2)}| \geq \bar{P}^{-c_e\alpha_q}\right) \\ &\overset{(b)}{\leq} \frac{1}{1 - 2f_{\Lambda_1}^{\max}\bar{P}^{(1-c_e)\alpha_q}} \frac{\mathbb{E}\left[\left|\lambda_1^{(1)} - \lambda_1^{(2)}\right|^2\right]}{\bar{P}^{-2c_e\alpha_q}},\end{aligned} \tag{F.23}$$

where (a) follows from (F.22) and (b) from Chebyshev's Inequality. In the following, we obtain the expectation $\mathbb{E}\left[\left|\lambda_1^{(1)} - \lambda_1^{(2)}\right|^2\right]$. In a similar manner to (F.4), Taylor's Theorem leads to

$$\begin{aligned}\mathbb{E}\left[\left|\lambda_1^{(1)} - \lambda_1^{(2)}\right|^2\right] &\leq \mathbb{E}\left[\left|\left(\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{(1)}\right)^{\mathrm{T}}\nabla\Lambda_1\left(\mathbf{v}_{\mathbf{H}}^{(1)}\right)\right|^2\right] \\ &\quad + \mathbb{E}\left[\left|o\left(\|\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{(1)}\|\right)\right|^2\right] \\ &\leq M_{\mathbf{J}}^2\mathbb{E}\left[\|\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{(1)}\|^2\right] + \mathbb{E}\left[o\left(\|\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{(1)}\|^2\right)\right],\end{aligned} \tag{F.24}$$

where (F.24) comes from the fact that $\|\nabla\Lambda_i\| \leq \|\mathbf{J}_\Lambda\| \leq M_{\mathbf{J}}$. We present in the following a useful lemma whose proof is relegated to Appendix F.5.

Lemma F.2. *Suppose* $\alpha_{i,k}^{(j)} > 0$ *for any* $i, j, k \in \mathbb{N}_2$. *Consider* $\alpha_{\min} \triangleq \min_{i,j,k\in\mathbb{N}_2} \alpha_{i,k}^{(j)}$. *Let* κ *be a positive real constant. Then, it holds that*

$$\mathbb{E}\left[\|\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{(1)}\|^2\right] = \kappa P^{-\alpha_{\min}}. \tag{F.25}$$

It follows from Lemma F.1 and Lemma F.2 than

$$\mathbb{E}\left[o\left(\|\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{(1)}\|^2\right)\right] = o\big(P^{-\alpha_{\min}}\big). \tag{F.26}$$

Including Lemma F.2 and (F.26) in (F.24) yields

$$\mathbb{E}\left[\left|\lambda_1^{(1)} - \lambda_1^{(2)}\right|^2\right] \leq \kappa M_{\mathbf{J}}^2 P^{-\alpha_{\min}} + o\big(P^{-\alpha_{\min}}\big). \tag{F.27}$$

Recall that $\bar{P} = \sqrt{P}$ and $\alpha_q = \frac{\alpha_{\min}}{c_q}$. Thus, by substituting (F.27) in (F.23) we obtain

$$\Pr\left(\mathcal{Q}_u(\lambda_1^{(1)}) \neq \mathcal{Q}_u(\lambda_1^{(2)}) \mid \lambda_1^{(1)} \in \bigcup_{n\in\mathbb{N}_N} C_n\right) \leq \frac{\kappa M_{\mathbf{J}}^2 P^{-\alpha_{\min}} + o\left(P^{-\alpha_{\min}}\right)}{\left(1 - 2f_{\Lambda_1}^{\max}\bar{P}^{(1-c_e)\frac{\alpha_{\min}}{c_q}}\right)P^{-c_e\frac{\alpha_{\min}}{c_q}}}. \tag{F.28}$$

It follows that

$$\frac{\kappa M_{\mathbf{J}}^2 P^{-\alpha_{\min}} + o\left(P^{-\alpha_{\min}}\right)}{\left(1 - 2f_{\Lambda_1}^{\max}\bar{P}^{(1-c_e)\frac{\alpha_{\min}}{c_q}}\right)P^{-c_e\frac{\alpha_{\min}}{c_q}}} = \mathcal{O}\big(P^{(\frac{c_e}{c_q}-1)\alpha_{\min}}\big). \tag{F.29}$$

Let us select an *edge* size such that $c_e < c_q$. Hence,

$$\Pr\left(\mathcal{Q}_u(\lambda_1^{(1)}) \neq \mathcal{Q}_u(\lambda_1^{(2)}) \mid \lambda_1^{(1)} \in \bigcup_{n\in\mathbb{N}_N} C_n\right) = o\left(\frac{1}{\log_2(P)}\right) \tag{F.30}$$

for any $c_q > 1$.

F.2.4 Assembling Probabilities

Plugging (F.17), (F.22) and (F.30) into (F.14) yields

$$\begin{aligned}\Pr\left(\Omega^{\mathsf{c}}\right) &\leq 2\left(\Pr\left(\mathcal{Q}_u(\lambda_1^{(1)}) \neq \mathcal{Q}_u(\lambda_1^{(2)})\right) + \Pr\left(\mathcal{Q}_u(\lambda_1^{(1)}) = 0\right)\right) \\ &= o\left(\frac{1}{\log_2(P)}\right),\end{aligned} \tag{F.31}$$

what concludes the proof for property (P2). ■

F.3 Proof of (P0): Bounded Expectation

We show now that there exists a constant $M < \infty$ such that for all P it holds that

$$\left|\mathbb{E}_{|\mathcal{Q}_u(\lambda_i^{(j)})>0}\left[\log_2\left(\mathcal{Q}_u(\lambda_i^{(j)})\right)\right]\right| \leq M. \tag{F.32}$$

Let us denote $x \triangleq \lambda_i^{(j)}$, with $i, j \in \mathbb{N}_2$, and the quantization step size as $q \triangleq \bar{P}^{\frac{-\alpha_{\min}}{c_q}}$. First, we upper bound the expectation from the fact that $0 \leq \lambda_i^{(j)} \leq 1$ as

$$\mathbb{E}_{|\mathcal{Q}_u(x)>0}\left[\log_2\left(\mathcal{Q}_u(x)\right)\right] \leq 0. \tag{F.33}$$

In order to lower bound it, note that

$$\mathbb{E}_{|\mathcal{Q}_u(x)>0}\left[\log_2\left(\mathcal{Q}_u(x)\right)\right] \triangleq \sum_{i=1}^{M} \log_2(iq)\Pr\left(\mathcal{Q}_u(x) = iq \mid \mathcal{Q}_u(x) > 0\right), \tag{F.34}$$

where $M \triangleq \left\lceil \frac{1}{q} \right\rceil - 1$ because the quantization level $\mathcal{Q}_u(x) = 0$ $(i = 0)$ is excluded from $\mathcal{Q}_u(x) > 0$. Besides this, $\Pr\left(\mathcal{Q}_u(x) = iq\right) = \Pr\left(iq \leq x \leq (i+1)q\right)$. The expectation in (F.34) is bounded for a given finite P because $q = \bar{P}^{-\alpha_{\min}/c_q} > 0$. In the following we prove that it is bounded also when $P \to \infty$. We can write that

$$\begin{aligned}\Pr\left(\mathcal{Q}_u(x) = iq \mid \mathcal{Q}_u(x) > 0\right) &= \frac{\Pr\left(\mathcal{Q}_u(x) = iq\right)}{\Pr\left(\mathcal{Q}_u(x) > 0\right)} \\ &\leq \frac{f_\Lambda^{\max}}{\max\left(0,\ 1 - f_\Lambda^{\max}\bar{P}^{\frac{-\alpha_{\min}}{c_q}}\right)} q.\end{aligned} \tag{F.35}$$

The last inequality comes from (F.16) as $1 - \Pr\left(\mathcal{Q}_u(x) > 0\right) \leq f_\Lambda^{\max}\bar{P}^{\frac{-\alpha_{\min}}{c_q}}$ and from the fact that $\Pr\left(\mathcal{Q}_u(x) = iq\right) \leq f_\Lambda^{\max} q$. There exists a $P_{\min}$ such that for all $P > P_{\min}$ it holds that $1 - f_\Lambda^{\max}\bar{P}^{\frac{-\alpha_{\min}}{c_q}} > 0$. Thus, as we focus on the limit as $P \to \infty$, we assume hereinafter that $1 - f_\Lambda^{\max}\bar{P}^{\frac{-\alpha_{\min}}{c_q}} > 0$. We introduce the notation

$$p'_{\max} \triangleq \frac{f_\Lambda^{\max}}{1 - f_\Lambda^{\max}\bar{P}^{\frac{-\alpha_{\min}}{c_q}}}. \tag{F.36}$$

Hence, since $M \leq \frac{1}{q}$ and for any $i \leq \frac{1}{q}$ it holds that $\log_2(iq) \leq 0$, we obtain that

$$\mathbb{E}_{|\mathcal{Q}_u(x)>0}\left[\log_2\left(\mathcal{Q}_u(x)\right)\right] \geq \frac{p'_{\max}}{M}\left(\sum_{i=1}^{M}\log_2(i) - \sum_{i=1}^{M}\log_2(M)\right). \tag{F.37}$$

Then, computing the summations yields

$$\mathbb{E}_{|\mathcal{Q}_u(x)>0}\left[\log_2\left(\mathcal{Q}_u(x)\right)\right] \geq p'_{\max}\left(\frac{\log_2(M!)}{M} - \log_2(M)\right). \tag{F.38}$$

We have that

$$\lim_{M\to\infty}\left(\frac{\log_2(M!)}{M} - \log_2(M)\right) = \frac{-1}{\ln(2)}, \tag{F.39}$$

what together with the fact that $\lim_{P\to\infty} p'_{\max} = f_{\Lambda}^{\max}$ implies that

$$\lim_{P\to\infty} \mathbb{E}_{|\mathcal{Q}_u(x)>0}\left[\log_2\left(\mathcal{Q}_u(x)\right)\right] \geq \frac{-f_{\Lambda}^{\max}}{\ln(2)}, \tag{F.40}$$

what concludes the proof. ■

F.4 Proof of Lemma F.1

Let us start noting that

$$\|\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{\star}\| \xrightarrow{a.s.} 0 \quad \Longleftrightarrow \quad |\{\mathbf{v}_{\mathbf{H}}^{(2)}\}_{\ell} - \{\mathbf{v}_{\mathbf{H}}^{\star}\}_{\ell}|^2 \xrightarrow{a.s.} 0, \quad \forall \ell \in \mathbb{N}_8. \tag{F.41}$$

It turns out that the only condition needed for this proof is that all $\alpha_{i,k}^{(j)}$ are strictly positive –as we will see in the following–. We recall that we have assumed that the estimates satisfy that $\alpha_{i,k}^{(j)} > 0$, for any $i, j, k \in \mathbb{N}_2$. Hence, we allow ourselves to focus on the first element of the vector $\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{\star}$. Let us denote the first element of the vector $\mathbf{v}_{\mathbf{H}}^{(j)} \in \mathbb{R}^{8\times 1}$ as $\hat{\mathrm{h}}_{\Re}^{(j)}$, i.e., $\hat{\mathrm{h}}_{\Re}^{(j)} = \mathrm{Re}\left(\hat{\mathrm{h}}_{1,1}^{(j)}\right)$. Similarly, $\tilde{\mathrm{h}}_{\Re}$ denotes the real part of the normalized channel coefficient, $\tilde{\mathrm{h}}_{\Re} = \mathrm{Re}\left(\tilde{\mathrm{h}}_{1,1}\right)$. From the feedback model it follows that

$$|\hat{\mathrm{h}}_{\Re}^{(2)} - \hat{\mathrm{h}}_{\Re}^{\star}|^2 \xrightarrow{a.s.} 0 \iff \hat{\mathrm{h}}_{\Re}^{(j)} - \tilde{\mathrm{h}}_{\Re} \xrightarrow{a.s.} 0 \quad \forall j \in \mathbb{N}_2. \tag{F.42}$$

Let $A_n = \{|X_n - X| > \varepsilon\}$. Then, the definition of almost sure convergence says that

$$X_n \xrightarrow{a.s.} X \quad \Longleftrightarrow \quad \Pr\left(A_n \ i.o.\right) = 0 \quad \forall \varepsilon > 0, \tag{F.43}$$

where

$$\begin{aligned} A_n \ i.o. &\triangleq \{w : w \in A_n \text{ for infinitely many } n\} \\ &= \limsup_n \ A_n. \end{aligned} \tag{F.44}$$

Let $X_n = \hat{\mathrm{h}}_{\Re}^{(j)} - \tilde{\mathrm{h}}_{\Re}$ and $X = 0$. We obtain in the following $\Pr(A_n) = \Pr\left(|\hat{\mathrm{h}}_{\Re}^{(j)} - \tilde{\mathrm{h}}_{\Re}| > \varepsilon\right)$. The absolute value of the difference can be bounded as

$$\begin{aligned} |\hat{\mathrm{h}}_{\Re}^{(j)} - \tilde{\mathrm{h}}_{\Re}| &= |(1 - \breve{z}_1^{(j)})\hat{\mathrm{h}}_{\Re}^{(j)} - z_1^{(j)}\delta_{\Re}^{(j)}| \\ &\leq (1 - \breve{z}_1^{(j)}) + z_1^{(j)}, \end{aligned} \tag{F.45}$$

what comes from the estimate model in (6.8) and because $|\hat{\mathrm{h}}_{\Re}^{(j)}| \leq 1$ and $|\delta_{\Re}^{(j)}| \leq 1$. The absolute value is omitted because $0 \leq z_1^{(j)} \leq 1$. Let us remind that $\breve{z}_1^{(j)} = \sqrt{1 - (z_1^{(j)})^2}$ and $z_1^{(j)} = \sqrt{Z_1^{(j)}}$. The fact that $1 - \sqrt{1 - x^2} \leq x$ for $0 \leq x \leq 1$ yields

$$|\hat{\mathrm{h}}_{\Re}^{(j)} - \tilde{\mathrm{h}}_{\Re}| \leq 2z_1^{(j)}. \tag{F.46}$$

As a result,

$$\begin{aligned} \Pr\left(|\hat{\mathrm{h}}_{\Re}^{(2)} - \tilde{\mathrm{h}}_{\Re}| > \varepsilon\right) &\leq \Pr\left(2z_1^{(j)} > \varepsilon\right) \\ &= \Pr\left(Z_1^{(j)} > \frac{\varepsilon^2}{4}\right). \end{aligned} \tag{F.47}$$

The quantization error $Z_1^{(j)}$ is distributed as the minimum of $n = 2^{B_1^{(j)}} = P^{\alpha_1^{(j)}}$ standard uniform random variables [43, 132]. Upon denoting $\varepsilon' = \frac{\varepsilon^2}{4}$, we obtain

$$\Pr\left(Z_1^{(j)} > \varepsilon'\right) = \left(1 - \varepsilon'\right)^n. \tag{F.48}$$

By definition –see (F.44)–, $\Pr(A_n \; i.o.)$ satisfies

$$\Pr(A_n \; i.o.) \leq \lim_{n\to\infty} \sum_{m=n}^{\infty} \Pr(A_n). \tag{F.49}$$

Introducing (F.48) in (F.49) leads to

$$\begin{aligned} \Pr(A_n \; i.o.) &\leq \lim_{n\to\infty} \sum_{m=n}^{\infty} \left(1 - \varepsilon'\right)^n \\ &\overset{(a)}{=} \lim_{n\to\infty} \frac{(1 - \varepsilon')^{n-1}}{\varepsilon'} \\ &= 0. \end{aligned} \tag{F.50}$$

where (a) comes from the application of the geometric series' formula. This implies that $\Pr(A_n \; i.o.) = 0$ for any $\varepsilon > 0$. We can repeat the process for all the elements in the RHS of(F.41), and thus Lemma F.1 is proven. ■

F.5 Proof of Lemma F.2

In this section we prove Lemma F.2, i.e., that

$$\mathbb{E}\left[\|\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{(1)}\|^2\right] \leq \kappa P^{-\alpha_{\min}}. \tag{F.51}$$

Let us consider w.l.o.g. that $\hat{\mathrm{h}}_{1,1}^{(2)}$ is the estimate whose accuracy scales as $\alpha_{\min}$. As defined in the previous section, let $\hat{\mathrm{h}}_{\Re}^{(j)} = \mathrm{Re}\left(\hat{\mathrm{h}}_{1,1}^{(j)}\right)$ and $\tilde{\mathrm{h}}_{\Re} = \mathrm{Re}\left(\tilde{\mathrm{h}}_{1,1}\right)$. We start by noting that

$$\mathbb{E}\left[\|\mathbf{v}_{\mathbf{H}}^{(2)} - \mathbf{v}_{\mathbf{H}}^{(1)}\|^2\right] \leq 8\,\mathbb{E}\left[|\hat{\mathrm{h}}_{\Re}^{(2)} - \hat{\mathrm{h}}_{\Re}^{(1)}|^2\right]. \tag{F.52}$$

The absolute value of the difference can be bounded as

$$\begin{aligned} |\hat{\mathrm{h}}_{\Re}^{(2)} - \hat{\mathrm{h}}_{\Re}^{(1)}| &\leq |\hat{\mathrm{h}}_{\Re}^{(2)} - \tilde{\mathrm{h}}_{\Re}| + |\tilde{\mathrm{h}}_{\Re} - \hat{\mathrm{h}}_{\Re}^{(1)}| \\ &\leq 2z_1^{(2)} + 2z_1^{(1)}, \end{aligned} \tag{F.53}$$

what follows from (F.46). Since $z_1^{(2)}$ is drawn from the same distribution as $z_1^{(1)}$ but with higher variance, it holds that

$$\mathbb{E}\left[\left(2z_1^{(2)} + 2z_1^{(1)}\right)^2\right] \leq \mathbb{E}\left[\left(4z_1^{(2)}\right)^2\right] \tag{F.54}$$

and consequently

$$\begin{aligned} \mathbb{E}\left[|\hat{\mathrm{h}}_{\Re}^{(2)} - \hat{\mathrm{h}}_{\Re}^{(1)}|^2\right] &\leq 16\,\mathbb{E}\left[\left(z_1^{(2)}\right)^2\right] \\ &\leq 16P^{-\alpha_{\min}}, \end{aligned} \tag{F.55}$$

where (F.55) is obtained from [132, Lemma 1]. This concludes the proof of Lemma F.2. ■

Appendix G

Analysis of the Power Normalization Parameters λ_i

We prove in this appendix the results presented in Section 6.6. As the parameters λ_i are obtained separately for each RX, we consider here only one of them. Consequently, we omit the sub-index $_i$ in the following. As we consider only one of the two parameters, the set Ω is defined as

$$\Omega \triangleq \left\{ (\hat{\mathbf{H}}^{(1)}, \hat{\mathbf{H}}^{(2)}) \middle| \mathcal{Q}\big(\lambda^{(1)}\big) = \mathcal{Q}\big(\lambda^{(2)}\big) \wedge \mathcal{Q}\big(\lambda^{(1)}\big) \neq 0 \right\}. \tag{G.1}$$

Let us consider the extended set $\Omega_{\cup 0}$ as the set including the cases in which the quantized agreed parameter is 0. It is hence defined as

$$\Omega_{\cup 0} \triangleq \left\{ (\hat{\mathbf{H}}^{(1)}, \hat{\mathbf{H}}^{(2)}) \middle| \ \mathcal{Q}\big(\lambda^{(1)}\big) = \mathcal{Q}\big(\lambda^{(2)}\big) \right\}. \tag{G.2}$$

We will assume when needed that

$$f_{\lambda|\Omega_{\cup 0}} = f_{\lambda}. \tag{G.3}$$

This assumption is done so as to simplify the result. It has not been proven, but it is expected to hold because of the uniformity of the quantizer and the isotropy of the CSIT. Moreover, it holds that $f_{\lambda|\Omega} \xrightarrow{a.s.} f_{\lambda}$ as P approaches infinity. The equality in (G.3) has been verified by simulations.

We consider the uniform quantizer of Lemma 6.1, denoted by $\mathcal{Q}$, with step size q. There are $N = \lceil 1/q \rceil$ quantization levels. Let define ℓ_n as the n-th quantization point.

Let $\ell_0 = 0$, $\ell_n = nq$, $n \in \{0, \dots, N-1\}$. The quantizer is given by the function

$$\mathcal{Q}_u(x) \triangleq q \left\lfloor \frac{1}{q} x \right\rfloor. \tag{G.4}$$

Consequently, $\Pr(\mathcal{Q}_u(x) = \ell_n) = \int_{\ell_n}^{\ell_{n+1}} f_X(x)\,\mathrm{d}x$, where $f_X(x)$ is the probability density function of the input x.

G.1 Maximum TX norm Normalization

Let us first consider the case in which $\lambda_i = |\tilde{\mathrm{h}}_{\bar{i},1}|\mu_i$, with $\mu_1 = \mu_2 = \mu$, and

$$\mu \triangleq \frac{1}{\max(\|\mathbf{t}^o_{\mathrm{TX}\,1}\|, \|\mathbf{t}^o_{\mathrm{TX}\,2}\|)}. \tag{G.5}$$

In this case, (6.49) implies that $f_{\max(\|\mathbf{t}^o_{\mathrm{TX}\,1}\|, \|\mathbf{t}^o_{\mathrm{TX}\,2}\|)} = 4 - 2x$, $x \in [1, 2]$, and μ satisfies that, $\forall x \in [\frac{\sqrt{2}}{2}, 1]$,

$$F_\mu = 1 - 4x^{-2} - x^{-4}, \qquad f_\mu = 8x^{-3} - 4x^{-5}. \tag{G.6}$$

Expected value of λ

In this case λ satisfies that, $\forall x \in [0, 1]$,

$$F_\lambda = \frac{x^2(1-x^2)}{1+x^2} + \begin{cases} \frac{x^2}{2-2x^2} & \text{if } x \le 1/\sqrt{2} \\ 1.5 - \frac{1}{2x^2} & \text{if } x \ge 1/\sqrt{2} \end{cases} \tag{G.7}$$

$$f_\lambda = \frac{4x}{(1+x^2)^2} - 2x + \min\left(\frac{x}{(1-x^2)^2}, \frac{1}{x^3}\right) \tag{G.8}$$

$$\Rightarrow \mathbb{E}[\lambda] = \frac{1}{6}\Big(-16 + 9\sqrt{2} + 3\pi - 3\tanh^{-1}(\frac{1}{\sqrt{2}})\Big) \approx 0.584763. \tag{G.9}$$

Moreover λ^2 satisfies that, $\forall x \in [0, 1]$,

$$F_{\lambda^2} = \frac{x(1-x)}{1+x} + \begin{cases} \frac{x}{2-2x} & \text{if } x \le 1/2 \\ 1.5 - \frac{1}{2x} & \text{if } x \ge 1/2 \end{cases} \tag{G.10}$$

$$f_{\lambda^2} = \frac{2}{(1+x)^2} - 1 + \min\left(\frac{1}{2(1-x)^2}, \frac{1}{2x^2}\right) \tag{G.11}$$

$$\Rightarrow \mathbb{E}\big[\lambda^2\big] = -1 + \tanh^{-1}(2) - \tanh^{-1}(3) - \frac{\ln(2)}{2} - \frac{\ln(3)}{2} + \ln(8) \approx 0.386294. \tag{G.12}$$

Consider now $\log_2(\lambda^2)$. From the previous results, it satisfies, $\forall x \in (-\infty, 0]$,

$$f_{\log_2(\lambda^2)} = \ln(2)\left(2^x\left(\frac{2}{(2^x+1)^2} - 1\right) + \min\left(\frac{2^{x-1}}{(1-2^x)^2}, 2^{-x-1}\right)\right) \tag{G.13}$$

$$\Rightarrow \mathbb{E}\left[\log_2(\lambda^2)\right] = \frac{3 - 4\ln(4)}{2\ln(2)} \approx -1.83596. \tag{G.14}$$

Thus, $\mathbb{E}[\log_2(\lambda)] = \frac{3-4\ln(4)}{4\ln(2)} \approx -0.91798$, what proves Lemma 6.2. ■

Expected value of λ conditioned on Ω

We have analyzed the expected values for the λ parameter. Let us consider now the expectation conditioned on Ω. First, note that $\Pr(\lambda^{\mathcal{Q}} = \ell_n) = F_\lambda(\ell_{n+1}) - F_\lambda(\ell_n)$. Hence, the assumption in (G.3) yields

$$\Pr\left(\lambda^{\mathcal{Q}} = 0 | \Omega_{\cup 0}\right) = F_\lambda(q). \tag{G.15}$$

Then, from (G.7), we obtain

$$\Pr\left(\lambda^{\mathcal{Q}} = 0 | \Omega_{\cup 0}\right) = \frac{q^2(1-q^2)}{1+q^2} + \begin{cases} \frac{q^2}{2-2q^2} & \text{if } q \leq 1/\sqrt{2} \\ 1.5 - \frac{1}{2q^2} & \text{if } q \geq 1/\sqrt{2} \end{cases} \tag{G.16}$$

From the fact that $f_{\lambda|\Omega}(x) = \frac{f_{\lambda|\Omega_{\cup 0}}(x)}{1-\Pr(\lambda^{\mathcal{Q}}=0|\Omega_{\cup 0})}$ if $x \geq q$, (G.3), and (G.15) it follows that

$$f_{\lambda|\Omega}(x) = = \frac{f_\lambda(x)}{1 - F_\lambda(q)} \qquad \text{if } x \geq q. \tag{G.17}$$

Recalling (G.3), we can write

$$\mathbb{E}_{|\Omega}[\log_2\left(\lambda^2\right)] = \frac{1}{1-F_\lambda(q)} \underbrace{\int_{\log_2(q^2)}^{0} x f_{\log_2(\lambda^2)}(x)}_{\triangleq \Phi}. \tag{G.18}$$

Let us introduce the notation $a \triangleq \log_2(q^2)$. Then,

$$\Phi = \begin{cases} \frac{1}{\ln(2)}\Big(\frac{1}{2} + 2^{-a-1}(a\ln(2)+1) \\ \qquad + 2^a(a\ln(2)-1) + 2\ln\left(2^a+1\right) - \frac{2a\ln(2)}{1+2^{-a}}\Big) - 2 & \text{if } a \geq -1 \\ \frac{3-12\ln\left(\frac{3}{2}\right)-\ln(2)}{\ln(64)} - \frac{\frac{1+2^a-(2^a-1)a\ln(2)}{1+2^{-a}} - 2\ln(1+2^a)}{\ln(2)} - 1 \\ \qquad + \frac{\frac{2^a a\ln(2)}{2^a-1} - \ln(1-2^a)}{2\ln(2)} + \frac{2-\ln(2)}{\ln(4)} - \frac{\ln\left(\frac{8192}{729}\right)}{\ln(8)} & \text{if } a \leq -1 \end{cases} \tag{G.19}$$

Expected values for the quantized parameter $\lambda^{\mathcal{Q}}$

For the quantized parameter, it follows in a similar manner as in (G.17) that, for $n \geq 1$,

$$\Pr\left(\lambda^{\mathcal{Q}} = \ell_n | \Omega\right) = \frac{\Pr\left(\lambda^{\mathcal{Q}} = \ell_n | \Omega_{\cup 0}\right)}{1 - F_{\lambda}(q)}. \tag{G.20}$$

We denote the mass density function (MDF) of a discrete variable x as f_x. From the law of the unconscious statistician and (G.3) we can write that

$$\begin{aligned} \mathbb{E}_{|\Omega}\left[\log_2\left((\lambda^{\mathcal{Q}})^2\right)\right] &= \sum_{n=1}^{N-1} \log_2((nq)^2) f_{\lambda^{\mathcal{Q}}|\Omega}(nq) \\ &= \frac{2}{1 - F_{\lambda}(q)} \underbrace{\sum_{n=1}^{N-1} \log_2(nq) f_{\lambda^{\mathcal{Q}}}(nq)}_{\triangleq \Xi}. \end{aligned} \tag{G.21}$$

It follows that

$$\begin{aligned} \Xi &= \sum_{n=1}^{N-1} \log_2(nq)\big(F_{\lambda}((n+1)q) - F_{\lambda}(nq)\big) \\ &= \sum_{n=2}^{N-1} F_{\lambda}(nq) \log_2\left(\frac{n-1}{n}\right) + \log_2(N-1) + \log_2(q)\left(1 - F_{\lambda}(q)\right), \end{aligned} \tag{G.22}$$

what concludes the proof for Lemma 6.3. ■

G.2 Unit-norm per RX Normalization

In this case, we transmit $\frac{P}{2}$ power for each RX stream (data symbol), and the precoder of each symbol is unit-norm. Hence, $\forall i \in \mathbb{N}_2$, $\mu_i = \mu$ and $\mu = 1$.

Expected value of λ

Therefore, $\lambda_i \triangleq |h_{\bar{i},1}|$ and thus

$$\lambda^2 \sim \text{Uniform}(0,1), \qquad \lambda \sim \text{Triangular}(0,1). \tag{G.23}$$

Note that, under such assumptions, Λ_1^* and Λ_2^* are i.i.d. The distributions in (G.23) imply that, $\forall x \in [0,1]$,

$$F_{\lambda} = x^2, \qquad f_{\lambda} = 2x, \qquad \Rightarrow \mathbb{E}[\lambda] = \frac{2}{3}. \tag{G.24}$$

Regarding λ^2, it follows that for any $x \in [0,1]$

$$F_{\lambda^2} = x, \qquad f_{\lambda^2} = 1, \qquad \Rightarrow \mathbb{E}\big[\lambda^2\big] = \frac{1}{2}. \tag{G.25}$$

Furthermore, $\log_2(\lambda^2)$ satisfies that, $\forall x \in (-\infty, 0]$,

$$F_{\log_2(\lambda^2)} = 2^x, \qquad f_{\log_2(\lambda^2)} = 2^x \ln(2) \tag{G.26}$$

$$\Rightarrow \mathbb{E}\big[\log_2(\lambda^2)\big] = \frac{-1}{\ln(2)} \approx -1.4427. \tag{G.27}$$

and also $\log_2(\lambda) = 0.5\log_2(\lambda^2)$ and then $\mathbb{E}[\log_2(\lambda)] = \frac{-1}{2\ln(2)} \approx -0.72135$. Hence, equation (6.58) of Lemma 6.4 is proven.

Expected value of λ conditioned on Ω

In a similar manner as in (G.16), it holds that $\Pr(\lambda^{\mathcal{Q}} = \ell_n) = F_\lambda(\ell_{n+1}) - F_\lambda(\ell_n)$ and hence

$$\Pr(\lambda^{\mathcal{Q}} = \ell_n) = \begin{cases} 1 - (N-1)^2 q^2 & \text{if } n = N-1 \\ (2n+1)q^2 & \text{otherwise.} \end{cases} \tag{G.28}$$

From (G.3), it follows that

$$\Pr\left(\lambda^{\mathcal{Q}} = \ell_0 | \Omega_{\cup 0}\right) = q^2, \qquad \Pr\left(\lambda^{\mathcal{Q}} \neq \ell_0 | \Omega_{\cup 0}\right) = 1 - q^2. \tag{G.29}$$

Likewise (G.17), by considering that $\Pr\left(\lambda^{\mathcal{Q}} \neq \ell_0 | \Omega_{\cup 0}\right) = 1 - q^2$ and (G.3) we obtain

$$\begin{aligned} f_{\lambda|\Omega}(x) &= \frac{f_{\lambda|\Omega_{\cup 0}}(x)}{1-q^2} \\ &= \frac{2x}{1-q^2}, \end{aligned} \tag{G.30}$$

and thus

$$\begin{aligned} \mathbb{E}_{|\Omega}\left[\log_2\left(\lambda^2\right)\right] &= \int_q^1 2\log_2(x)\frac{2x}{1-q^2}dx \\ &= \frac{1}{(1-q^2)}\left[\frac{x^2(2\ln(x)-1)}{\ln(2)}\right]_q^1 \\ &= \frac{-1}{\ln(2)} - \frac{q^2\log_2(q^2)}{1-q^2}, \end{aligned} \tag{G.31}$$

what concludes the proof of Lemma 6.4. ■

Expected values for the quantized parameter $\lambda^{\mathcal{Q}}$

Similar to (G.17) and (G.30), the quantized parameter $\lambda^{\mathcal{Q}}$ satisfies that

$$\Pr\left(\lambda^{\mathcal{Q}} = \ell_n | \Omega\right) = \frac{\Pr\left(\lambda^{\mathcal{Q}} = \ell_n\right)}{1-q^2}. \tag{G.32}$$

Thus,

$$\Pr\left(\lambda^{\mathcal{Q}} = \ell_n | \Omega\right) = \begin{cases} \frac{1-(N-1)^2q^2}{1-q^2} & \text{if } n = N-1 \\ \frac{(2n+1)q^2}{1-q^2} & \text{otherwise.} \end{cases} \tag{G.33}$$

Moreover, from the Law of the Unconscious Statistician, it holds that

$$\mathbb{E}_{|\Omega}\left[\log_2\left((\lambda^{\mathcal{Q}})^2\right)\right] = 2\sum_{n=1}^{N-1} \log_2(nq) \Pr\left(\lambda^{\mathcal{Q}} = nq | \Omega\right). \tag{G.34}$$

Therefore,

$$\begin{aligned} \mathbb{E}_{|\Omega}\left[\log_2\left((\lambda^{\mathcal{Q}})^2\right)\right] &= 2\left(\sum_{n=1}^{N-2} \log_2(nq)\frac{(2n+1)q^2}{1-q^2} + \log_2((N-1)q)\frac{1-(N-1)^2q^2}{1-q^2}\right) \\ &= 2\Big(\frac{q^2}{1-q^2}\underbrace{\sum_{n=1}^{N-2} \log_2(nq)(2n+1)}_{\Psi} + \log_2((N-1)q)\frac{1-(N-1)^2q^2}{1-q^2}\Big). \end{aligned} \tag{G.35}$$

To conclude, the sum Ψ can be expressed as

$$\begin{aligned} \Psi &= \sum_{n=1}^{N-2} 2n\log_2(n) + \sum_{n=1}^{N-2} \log_2(n) + \log_2(q)\sum_{n=1}^{N-2}(2n+1) \\ &= 2\log_2\left(\prod_{n=1}^{N-2} n^n\right) + \log_2((N-2)!) + N(N-2)\log_2(q) \\ &= 2\sum_{n=1}^{N-2} n\log_2(n) + \log_2((N-2)!) + N(N-2)\log_2(q). \end{aligned} \tag{G.36}$$

Appendix H

Proofs for Chapter 7

This appendix encloses all the proofs of the lemmas, corollaries and propositions of Chapter 7. We present them in demonstrative order.

H.1 Proof of Lemma 7.3 (Probability of Power Outage)

We denote the event of power outage as $\mathcal{P}_o$. Note that

$$\Pr\left(\mathcal{P}_o\right) \leq N_1 \Pr\left(\|\mathbf{T}_{1,1}\| > 1\right), \tag{H.1}$$

and $\mathbf{T}_{1,1} = \mu[\mathbf{w}_{1,1,1}, \mathbf{w}_{2,1,1}, \ldots, \mathbf{w}_{K,1,1}]$, where $\mathbf{w}_{i,j,n}$ represents the n-th element of the precoding vector at TX j for the data symbols of RX i. Therefore,

$$\begin{aligned}
\Pr\left(\mathcal{P}_o\right) &\leq N_1 \Pr\left(\|\mu[\mathbf{w}_{1,1,1}, \mathbf{w}_{2,1,1}, \ldots, \mathbf{w}_{K,1,1}]\| > 1\right) \\
&\overset{(a)}{\leq} N_1 \Pr\left(\bigcup_{i\in\mathbb{N}_K} \|\mu\mathbf{w}_{i,1,1}\| > \|\mathbf{v}_{i,1,1}\|\right) \\
&\overset{(b)}{\leq} N_1 K \Pr\left(\|\mu\mathbf{w}_{1,1,1}\| > \|\mathbf{v}_{1,1,1}\|\right) \\
&\overset{(c)}{\leq} N_1 K \Pr\left(\mu\|\mathbf{v}_{1,1,1}\| + \mu\|\boldsymbol{\phi}_1\| > \|\mathbf{v}_{1,1,1}\|\right) \\
&= N_1 K \Pr\left(\|\boldsymbol{\phi}_1\| > \frac{1-\mu}{\mu}\|\mathbf{v}_{1,1,1}\|\right),
\end{aligned} \tag{H.2}$$

where (a) is obtained from the precoder definition –since $\|[\mathbf{v}_{1,1,1} \ \ldots \ \mathbf{v}_{K,1,1}]\| \leq 1$–, (b) because $\mathbf{w}_{i,1,1}$ (resp. $\mathbf{v}_{i,1,1}$) is equally distributed for any $i \in \mathbb{N}_K$, and (c) from (7.22). Now, we obtain the probability by conditioning on $\|\mathbf{v}_{1,1,1}\|$ and then averaging over the

distribution of $\|\mathbf{v}_{1,1,1}\|$. Let us denote $\mu' \triangleq \frac{1-\mu}{\mu}$. Hence,

$$\Pr\left(\mathcal{P}_o\right) \leq N_1 K \int_{-\infty}^{\infty} \Pr\left(\|\boldsymbol{\phi}_1\| > \mu'\nu\right) f_{\|\mathbf{v}_{1,1,1}\|}(\nu)\,\mathrm{d}\nu. \tag{H.3}$$

Using Markov's inequality we obtain that

$$\begin{aligned}\Pr\left(\mathcal{P}_o\right) &\leq N_1 K \int_{-\infty}^{\infty} \frac{\mathbb{E}[\|\boldsymbol{\phi}_1\|]}{\mu'\nu} f_{\|\mathbf{v}_{1,1,1}\|}(\nu)\,\mathrm{d}\nu \\ &= N_1 K\, \mathbb{E}[\|\boldsymbol{\phi}_1\|] \frac{1}{\mu'} \mathbb{E}\left[\|\mathbf{v}_{1,1,1}\|^{-1}\right],\end{aligned} \tag{H.4}$$

where $\mathbb{E}\left[\|\mathbf{v}_{1,1,1}\|^{-1}\right]$ exists from property (ZF2). Let us focus on the first expectation term of (H.4) ($\mathbb{E}[\|\boldsymbol{\phi}_1\|]$). Recalling (7.22), $\boldsymbol{\phi}_i$ is defined as

$$\boldsymbol{\phi}_i = \hat{\mathbf{H}}_{\bar{i},1}^{\dagger} \hat{\mathbf{H}}_{\bar{i},\bar{1}} (\mathbf{v}_{i,\bar{1}} - \mathbf{w}_{i,\bar{1}}). \tag{H.5}$$

Then,

$$\begin{aligned}\mathbb{E}\left[\|\boldsymbol{\phi}_i\|\right] &= \mathbb{E}\left[\|\hat{\mathbf{H}}_{\bar{i},1}^{\dagger} \hat{\mathbf{H}}_{\bar{i},\bar{1}} (\mathbf{v}_{i,\bar{1}} - \mathbf{w}_{i,\bar{1}})\|\right] \\ &\overset{(a)}{\leq} \mathbb{E}\left[\|\hat{\mathbf{H}}_{\bar{i},1}^{\dagger} \hat{\mathbf{H}}_{\bar{i},\bar{1}}\|_{\mathrm{F}} \|\mathbf{v}_{i,\bar{1}} - \mathbf{w}_{i,\bar{1}}\|\right] \\ &\overset{(b)}{\leq} \sqrt{\mathbb{E}\left[\|\hat{\mathbf{H}}_{\bar{i},1}^{\dagger} \hat{\mathbf{H}}_{\bar{i},\bar{1}}\|_{\mathrm{F}}^2\right] \mathbb{E}\left[\|\mathbf{v}_{i,\bar{1}} - \mathbf{w}_{i,\bar{1}}\|^2\right]},\end{aligned} \tag{H.6}$$

where (a) comes from the sub-multiplicative property of Frobenius norm and (b) from Cauchy–Schwarz inequality. Let us denote $g_m \triangleq \sqrt{\mathbb{E}\left[\|\hat{\mathbf{H}}_{\bar{i},1}^{\dagger} \hat{\mathbf{H}}_{\bar{i},\bar{1}}\|_{\mathrm{F}}^2\right]}$, which is a value that does not depend on P because the channel estimates are equally distributed for any estimation error variance. Then, we have that

$$\mathbb{E}\left[\|\boldsymbol{\phi}_i\|\right] \leq g_m \sqrt{\mathbb{E}\left[\|\mathbf{v}_{i,\bar{1}} - \mathbf{w}_{i,\bar{1}}\|^2\right]}. \tag{H.7}$$

Lemma 7.2 and the fact that $\mathbb{E}\left[\|\mathbf{v}_{i,\bar{1}} - \mathbf{w}_{i,\bar{1}}\|^2\right] = \sum_{j=2}^{M} \mathbb{E}\left[\|\mathbf{v}_{i,j} - \mathbf{w}_{i,j}\|^2\right]$ yield

$$\mathbb{E}\left[\|\boldsymbol{\phi}_1\|\right] = \mathcal{O}\big(\bar{P}^{-\alpha_q}\big). \tag{H.8}$$

Since $\mu = 1 - \varepsilon$, with $\varepsilon = \Theta(\bar{P}^{-\alpha_\mu})$ and $\varepsilon > 0$, the term $\frac{1}{\mu'} = \frac{\mu}{1-\mu}$ satisfies $\frac{1}{\mu'} = \Theta(\bar{P}^{\alpha_\mu})$. From (ZF2), $\mathbb{E}\left[\|\mathbf{v}_{1,1,1}\|^{-1}\right]$ is $\Theta(1)$. Hence, recalling (H.4),

$$\begin{aligned}\Pr\left(\mathcal{P}_o\right) &\leq N_1 K \mathbb{E}\left[\|\mathbf{v}_{1,1,1}\|^{-1}\right] \mathbb{E}[\|\boldsymbol{\phi}_1\|] \frac{1}{\mu'} \\ &= \Theta(1)\mathcal{O}\left(\bar{P}^{-\alpha_q}\right) \Theta(\bar{P}^{\alpha_\mu}),\end{aligned} \tag{H.9}$$

what implies that $\Pr(\mathcal{P}_o) = \mathcal{O}\left(\bar{P}^{\alpha_\mu - \alpha_q}\right)$. By selecting $\alpha_\mu < \alpha_q$, the probability of power outage vanishes and it holds that

$$\Pr(\mathcal{P}_o) = o\left(\frac{1}{\log_2(P)}\right), \tag{H.10}$$

what concludes the proof. ■

H.2 Proof of Corollary 7.2 (Error on Distributed Precoder)

In order to prove that $\mathbb{E}\left[\|\mathbf{v}_i - \mathbf{w}_i\|^2\right] = \mathcal{O}\left(P^{-\alpha_q}\right)$, let us recall that the vector $\mathbf{w}_i - \mathbf{v}_i$ can be written as

$$\mathbf{w}_i - \mathbf{v}_i = \begin{bmatrix} \boldsymbol{\phi}_i \\ \mathbf{w}_{i,\bar{1}} - \mathbf{v}_{i,\bar{1}} \end{bmatrix}, \tag{H.11}$$

where $\boldsymbol{\phi}_i$, which has been defined in (7.22), is the difference at TX 1 and $\mathbf{w}_{i,\bar{1}} - \mathbf{v}_{i,\bar{1}}$ denotes the difference for the coefficients of all the TXs but TX 1, i.e., $\mathbf{w}_{i,\bar{1}} - \mathbf{v}_{i,\bar{1}} = [(\mathbf{w}_{i,2} - \mathbf{v}_{i,2})^{\mathrm{T}}, \ldots, (\mathbf{w}_{i,K} - \mathbf{v}_{i,K})^{\mathrm{T}}]^{\mathrm{T}}$. Let us define $\mathbf{G_I}$ as $\mathbf{G_I} \triangleq \begin{bmatrix} \hat{\mathbf{H}}_{\bar{i},1}^{\dagger} \hat{\mathbf{H}}_{\bar{i},\bar{1}} \\ \mathbf{I} \end{bmatrix}$. From (H.5), we can rewrite (H.11) as

$$\mathbf{w}_i - \mathbf{v}_i = \mathbf{G_I}(\mathbf{w}_{i,\bar{1}} - \mathbf{v}_{i,\bar{1}}). \tag{H.12}$$

From Cauchy-Schwarz inequality, (H.12) implies that

$$\mathbb{E}\left[\|\mathbf{w}_i - \mathbf{v}_i\|^2\right] \leq \sqrt{\mathbb{E}[\|\mathbf{G_I}\|^4]\,\mathbb{E}[\|\mathbf{w}_{i,\bar{1}} - \mathbf{v}_{i,\bar{1}}\|^4]}. \tag{H.13}$$

Let g_s be $g_s \triangleq \sqrt{\mathbb{E}[\|\mathbf{G_I}\|^4]}$, which is $\Theta(1)$. Moreover, the instantaneous power constraint for the precoder ensures that $\|\mathbf{v}_{i,j} - \mathbf{w}_{i,j}\|^2 \leq 4N_j^2$. Hence,

$$\mathbb{E}\left[\|\mathbf{w}_{i,\bar{1}} - \mathbf{v}_{i,\bar{1}}\|^4\right] \leq \sum_{j=2}^{M} \mathbb{E}\left[4N_j^2 \|\mathbf{w}_{i,j} - \mathbf{v}_{i,j}\|^2\right], \tag{H.14}$$

what, together with Lemma 7.2, means that

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{w}_i - \mathbf{v}_i\|^2\right] &\leq 4N_j^2 g_s \sum_{j=2}^{M} \mathbb{E}\left[\|\mathbf{v}_{i,j} - \mathbf{w}_{i,j}\|^2\right] \\ &= \mathcal{O}(\bar{P}^{-\alpha_q}), \end{aligned} \tag{H.15}$$

what concludes the proof. ■

H.3 Proof of Proposition 7.1

We prove in the following that $\Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell| < \bar{P}^{-\gamma}\right) = o\left(\frac{1}{\log_2(P)}\right)$, for any $\gamma > 0$ and $i, \ell \in \mathbb{N}_K : \ell \neq i$. Let us denote the precoder for RX ℓ obtained with perfect knowledge of $\mathbf{H}$ as $\mathbf{u}_\ell$. Then,

$$\begin{aligned}\Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell| < \bar{P}^{-\gamma}\right) &= \Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell + \boldsymbol{\delta}_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)| < \bar{P}^{-\gamma}\right) \\ &\leq \Pr\left(\left||\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell| - |\boldsymbol{\delta}_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)|\right| < \bar{P}^{-\gamma}\right),\end{aligned} \tag{H.16}$$

where we have applied the inverse triangle inequality. In order to prove Proposition 7.1 we capitalize the intuition that the term $|\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell|$ is independent of the quality of the estimate and P, but the value of $|\boldsymbol{\delta}_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)|$ is directly proportional to the quality $P^{-\alpha^{(1)}}$. Before applying this intuition to (H.16), we first analyze the term $|\boldsymbol{\delta}_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)|$ to obtain which is the probability $\Pr\left(|\boldsymbol{\delta}_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)| > \bar{P}^{-\beta}\right)$, for $\beta < \alpha^{(1)}$. Let us define the scalar $\varepsilon > 0$ such that $\beta < \beta + \varepsilon < \alpha^{(1)}$. By means of the Cauchy-Schwarz inequality and the law of total probability we obtain

$$\begin{aligned}\Pr\left(|\boldsymbol{\delta}_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)| > \bar{P}^{-\beta}\right) &\leq \Pr\left(\|\boldsymbol{\delta}_i^{(1)}\|\|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta}\right) \\ &\leq \Pr\left(\|\boldsymbol{\delta}_i^{(1)}\|\|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta} \mid \|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta-\varepsilon}\right)\Pr\left(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta-\varepsilon}\right) \\ &\quad + \Pr\left(\|\boldsymbol{\delta}_i^{(1)}\|\|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta} \mid \|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon}\right)\Pr\left(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon}\right) \\ &\leq \Pr\left(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta-\varepsilon}\right) + \Pr\left(\|\boldsymbol{\delta}_i^{(1)}\| > \bar{P}^{\varepsilon} \mid \|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon}\right).\end{aligned} \tag{H.17}$$

The first term $\Pr\left(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta-\varepsilon}\right)$ can be upper-bounded by means of the Markov's inequality, such that

$$\begin{aligned}\Pr\left(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| > \bar{P}^{-\beta-\varepsilon}\right) &\leq \bar{P}^{\beta+\varepsilon}\,\mathbb{E}[\|\mathbf{v}_\ell - \mathbf{u}_\ell\|] \\ &= \mathcal{O}(\bar{P}^{\beta+\varepsilon-\alpha^{(1)}}),\end{aligned} \tag{H.18}$$

where the last step follows directly after applying Lemma 7.2 to vectors whose respective input estimates differ by a $\mathcal{O}(\bar{P}^{-\alpha^{(1)}})$ additive error term. For the last term in (H.17), $\Pr\left(\|\boldsymbol{\delta}_i^{(1)}\| > \bar{P}^{\varepsilon} \mid \|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon}\right)$, it follows that

$$\Pr\left(\|\boldsymbol{\delta}_i^{(1)}\| > \bar{P}^{\varepsilon} \mid \|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon}\right) \leq \frac{\Pr\left(\|\boldsymbol{\delta}_i^{(1)}\| > \bar{P}^{\varepsilon}\right)}{\Pr\left(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon}\right)}. \tag{H.19}$$

From (H.18), it holds that $\Pr\left(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon}\right) = 1 - \mathcal{O}(\bar{P}^{\beta+\varepsilon-\alpha^{(1)}})$. Besides this, $\|\boldsymbol{\delta}_i^{(1)}\|^2 = \sum_{n=1}^{N_T} |\delta_{i,n}^{(1)}|^2$, where $\delta_{i,n}^{(1)}$ are i.i.d. as $\mathcal{N}_{\mathbb{C}}(0,1)$. Consequently, $|\delta_{i,n}^{(1)}|^2$ is distributed following a Rayleigh distribution and $\|\boldsymbol{\delta}_i^{(1)}\|^2 \sim \Gamma_d(N_T, 1)$, where $\Gamma_d(N_T, 1)$ denotes the

Gamma distribution. Moreover, $\Gamma_d(N_T, 1)$ is also called the Erlang distribution, and it satisfies that

$$\Pr\left(X \sim \Gamma_d(N_T, 1) < x\right) = 1 - \sum_{n=0}^{N_T-1} \frac{1}{n!} e^{-x} x^n. \tag{H.20}$$

Hence, it follows that

$$\begin{aligned} \Pr\left(\|\boldsymbol{\delta}_i^{(1)}\| > \bar{P}^{\varepsilon}\right) &= \Pr\left(\|\boldsymbol{\delta}_i^{(1)}\|^2 > \bar{P}^{2\varepsilon}\right) \\ &= \sum_{n=0}^{N_T-1} \frac{1}{n!} e^{-\bar{P}^{2\varepsilon}} \bar{P}^{2n\varepsilon}. \end{aligned} \tag{H.21}$$

Since $\varepsilon > 0$, it follows that $\Pr\left(\|\boldsymbol{\delta}_i^{(1)}\| > \bar{P}^{\varepsilon}\right) = o(\bar{P}^x)$, $\forall x \in \mathbb{R}$, and hence

$$\frac{\Pr\left(\|\boldsymbol{\delta}_i^{(1)}\| > \bar{P}^{\varepsilon}\right)}{\Pr\left(\|\mathbf{v}_\ell - \mathbf{u}_\ell\| \leq \bar{P}^{-\beta-\varepsilon}\right)} = \frac{1}{1 - \mathcal{O}(\bar{P}^{\beta+\varepsilon-\alpha^{(1)}})}\, o\left(\frac{1}{\log_2(P)}\right), \tag{H.22}$$

what together with (H.18)and (H.17) leads to

$$\Pr\left(|\boldsymbol{\delta}_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)| > \bar{P}^{-\beta}\right) = o\left(\frac{1}{\log_2(P)}\right) \tag{H.23}$$

for any $\beta < \alpha^{(1)}$. With this result, we can focus back on (H.16), that can be expanded by means of the Law of total probability such that

$$\begin{aligned} &\Pr\left(\left||\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell| - |\boldsymbol{\delta}_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)|\right| < \bar{P}^{-\gamma}\right) \\ &\quad = \Pr\left(\left||\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell| - |\boldsymbol{\delta}_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)|\right| < \bar{P}^{-\gamma} \mid |\boldsymbol{\delta}_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)| \leq \bar{P}^{-\beta}\right) \\ &\qquad + o\left(\frac{1}{\log_2(P)}\right) \\ &\quad \leq \Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell| < \bar{P}^{-\gamma} + \bar{P}^{-\beta}\right) + o\left(\frac{1}{\log_2(P)}\right). \end{aligned} \tag{H.24}$$

Let us assume w.l.o.g. that $\beta < \gamma$, such that $\Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell| < \bar{P}^{-\gamma} + \bar{P}^{-\beta}\right) \leq \Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell| < 2\bar{P}^{-\beta}\right)$. Therefore, it remains to prove that $\Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell| < 2\bar{P}^{-\beta}\right) = o\left(\frac{1}{\log_2(P)}\right)$. Let ϵ_β be a scalar such that $0 < \epsilon_\beta < \beta$ and let us define ψ as the angle satisfying

$$\cos(\psi) \triangleq \frac{|\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell|}{\|\boldsymbol{\delta}_i^{(1)}\|\|\mathbf{u}_\ell\|}. \tag{H.25}$$

Then, we use again the Law of Total Probability to obtain

$$\begin{aligned}\Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell| < 2\bar{P}^{-\beta}\right) &= \Pr\left(\|\boldsymbol{\delta}_i^{(1)}\|\|\mathbf{u}_\ell\|\cos(\psi) < 2\bar{P}^{-\beta} \,|\, \|\mathbf{u}_\ell\| \leq \bar{P}^{-\epsilon_\beta}\right)\Pr\left(\|\mathbf{u}_\ell\| \leq \bar{P}^{-\epsilon_\beta}\right) \\ &\quad + \Pr\left(\|\boldsymbol{\delta}_i^{(1)}\|\|\mathbf{u}_\ell\|\cos(\psi) < 2\bar{P}^{-\beta} \mid \|\mathbf{u}_\ell\| > \bar{P}^{-\epsilon_\beta}\right)\Pr\left(\|\mathbf{u}_\ell\| > \bar{P}^{-\epsilon_\beta}\right) \qquad \text{(H.26)} \\ &\leq \Pr\left(\|\mathbf{u}_\ell\| \leq \bar{P}^{-\epsilon_\beta}\right) + \Pr\left(\|\boldsymbol{\delta}_i^{(1)}\|\cos(\psi) < 2\bar{P}^{-\beta}\bar{P}^{\epsilon_\beta} \mid \|\mathbf{u}_\ell\| > \bar{P}^{-\epsilon_\beta}\right).\end{aligned}$$

Importantly, $\boldsymbol{\delta}_i^{(1)}$ is isotropically distributed (i.e., the normalized value $\boldsymbol{\delta}_i^{(1)}/\|\boldsymbol{\delta}_i^{(1)}\|$ is uniformly distributed in the sphere surface). Besides this, $\mathbf{u}_\ell$ is a function of $\mathbf{H}$. Since $\mathbf{H}$ and $\boldsymbol{\delta}_i^{(1)}$ are mutually independent, so $\boldsymbol{\delta}_i^{(1)}$ and $\mathbf{u}_\ell$ are. Hence, from isotropy of $\boldsymbol{\delta}_i^{(1)}$, $\cos(\psi)$ is independent of $\mathbf{u}_\ell$. On this basis, we can select $\mathbf{u}_\ell = [1, \mathbf{0}_{1\times N_T-1}]$ to obtain that

$$\Pr\left(\|\boldsymbol{\delta}_i^{(1)}\|\cos(\psi) < 2\bar{P}^{\epsilon_\beta-\beta} \mid \|\mathbf{u}_\ell\| > \bar{P}^{-\epsilon_\beta}\right) = \Pr\left(|\boldsymbol{\delta}_{i,1,1}^{(1)}| < 2\bar{P}^{\epsilon_\beta-\beta}\right), \qquad \text{(H.27)}$$

where $\boldsymbol{\delta}_{i,1,1}^{(1)}$ denotes the first element of the vector $\boldsymbol{\delta}_i^{(1)}$, and it is distributed as $\mathcal{N}_{\mathbb{C}}(0,1)$. Then,

$$\begin{aligned}\Pr\left(|\boldsymbol{\delta}_{i,1,1}^{(1)}| < 2\bar{P}^{\epsilon_\beta-\beta}\right) &= \frac{2}{\sqrt{2\pi}}\int_0^{2\bar{P}^{\epsilon_\beta-\beta}} e^{-x^2/2}\,\mathrm{d}x \\ &\leq \frac{4}{\sqrt{2\pi}}\bar{P}^{\epsilon_\beta-\beta}.\end{aligned} \qquad \text{(H.28)}$$

On the other hand, the term $\Pr\left(\|\mathbf{u}_\ell\| \leq \bar{P}^{-\epsilon_\beta}\right)$ can be bounded by

$$\begin{aligned}\Pr\left(\|\mathbf{u}_\ell\| \leq \bar{P}^{-\epsilon_\beta}\right) &= \int_0^{\bar{P}^{-\epsilon_\beta}} f_{\|\mathbf{u}_i\|}(x)\,\mathrm{d}x \\ &\leq f_{\|\mathbf{u}_i\|}^{\max}\bar{P}^{-\epsilon_\beta},\end{aligned} \qquad \text{(H.29)}$$

what follows from (ZF3). By introducing (H.28) and (H.29) in (H.26) we obtain that

$$\Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell| < 2\bar{P}^{-\beta}\right) = \mathcal{O}(\bar{P}^{\max(-\epsilon_\beta,\ \epsilon_\beta-\beta)}). \qquad \text{(H.30)}$$

Note that ϵ_β satisfies $0 < \epsilon_\beta < \beta$. Hence,

$$\begin{aligned}\Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{v}_\ell| < \bar{P}^{-\gamma}\right) &\leq \Pr\left(\left||\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell| - |\boldsymbol{\delta}_i^{(1)}(\mathbf{v}_\ell - \mathbf{u}_\ell)|\right| < \bar{P}^{-\gamma}\right) \\ &\leq \Pr\left(|\boldsymbol{\delta}_i^{(1)}\mathbf{u}_\ell| < 2\bar{P}^{-\beta}\right) + o\left(\frac{1}{\log_2(P)}\right) \\ &= o\left(\frac{1}{\log_2(P)}\right),\end{aligned} \qquad \text{(H.31)}$$

what concludes the proof of Proposition 7.1. ■

H.4 Proof of Proposition 7.2

We prove in the following that $\mathbb{E}[|\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell - \mathbf{v}_\ell)|] = \mathcal{O}(\bar{P}^{-\alpha_q})$ for any $i, \ell \in \mathbb{N}_K : \ell \neq i$. It follows that

$$\begin{aligned}\mathbb{E}\big[|\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell - \mathbf{v}_\ell)|\big] &\leq \mathbb{E}\big[\|\boldsymbol{\delta}_i^{(1)}\|\|\mathbf{w}_\ell - \mathbf{v}_\ell\|\big] \\ &= \operatorname{cov}\Big(\|\boldsymbol{\delta}_i^{(1)}\|, \|\mathbf{w}_\ell - \mathbf{v}_\ell\|\Big) + \mathbb{E}\Big[\|\boldsymbol{\delta}_i^{(1)}\|\Big]\,\mathbb{E}[\|\mathbf{w}_\ell - \mathbf{v}_\ell\|] \\ &\leq \sqrt{\mathbb{E}\Big[\|\boldsymbol{\delta}_i^{(1)}\|^2\Big]}\sigma_{\|\mathbf{w}_\ell - \mathbf{v}_\ell\|} + \mathbb{E}\Big[\|\boldsymbol{\delta}_i^{(1)}\|\Big]\,\mathbb{E}[\|\mathbf{w}_\ell - \mathbf{v}_\ell\|]\,,\end{aligned} \tag{H.32}$$

where $\operatorname{cov}(X, Y) \triangleq \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$ is the covariance between X and Y and σ_X^2 represents the variance of the random variable X. The last inequality comes from the fact that $\operatorname{cov}(x, y) \leq \sigma_x \sigma_y$ and $\sigma_x^2 \leq \mathbb{E}\big[x^2\big]$. Besides this, it holds from $\|\boldsymbol{\delta}_i^{(1)}\|^2 \sim \Gamma_d(N_T, 1)$ that $\mathbb{E}\Big[\|\boldsymbol{\delta}_i^{(1)}\|^2\Big] = N_T$. From this and the fact that $\mathbb{E}[x] \leq \sqrt{\mathbb{E}[x^2]}$ we can write

$$\begin{aligned}\mathbb{E}\big[|\boldsymbol{\delta}_i^{(1)}(\mathbf{w}_\ell - \mathbf{v}_\ell)|\big] &\leq \sqrt{N_T}\left(\sigma_{\|\mathbf{w}_\ell - \mathbf{v}_\ell\|} + \mathbb{E}[\|\mathbf{w}_\ell - \mathbf{v}_\ell\|]\right) \\ &\overset{(a)}{\leq} \sqrt{N_T}\; 2\sqrt{\mathbb{E}[\|\mathbf{w}_\ell - \mathbf{v}_\ell\|^2]} \\ &\overset{(b)}{=} \mathcal{O}(\bar{P}^{-\alpha_q}),\end{aligned} \tag{H.33}$$

where (a) comes from the fact that $\sigma_x + \mathbb{E}[x] \leq 2\sqrt{\mathbb{E}[x^2]}$ and (b) from Corollary 7.2. ■

H.5 Proof of Lemma 7.1 (Quantizer Consistency)

The proof of Lemma 7.1 is analogous to the proof of Property P2 in Chapter 6, although the estimate model and the quantizer are different. Let $q \triangleq \bar{P}^{-\alpha_q}$ be the quantization step size of the quantizer $\mathcal{Q}_u$. Then, $\mathcal{Q}_u$ is defined such that, for a scalar value $x \in \mathbb{R}$,

$$\mathcal{Q}_u(x) \triangleq q\left\lfloor \frac{x}{q} + \frac{1}{2} \right\rfloor. \tag{H.34}$$

We extend the notation for any complex matrix $\mathbf{A} \in \mathbb{C}^{n \times m}$ such that $\mathbf{A}_q = \mathcal{Q}_u(\mathbf{A})$ denotes the element-wise quantization, i.e.,

$$(\mathbf{A}_q)_{i,k} = \mathcal{Q}_u\big(\operatorname{Re}(\mathbf{A}_{i,k})\big) + \imath\mathcal{Q}_u\big(\operatorname{Im}(\mathbf{A}_{i,k})\big), \tag{H.35}$$

where $\operatorname{Re}(x)$ and $\operatorname{Im}(x)$ stand for the real imaginary part of $x \in \mathbb{C}$, and $\imath \triangleq \sqrt{-1}$. In this appendix we prove that, for a scalar uniform quantizer $\mathcal{Q}_u$ with $q = \bar{P}^{-\alpha_q}$ and

$\alpha^{(j)} > \alpha_q > 0$, $\forall j \in \mathbb{N}_M$, it follows that

$$\Pr\left(\hat{\mathbf{H}}_q^{(j)\leftarrow(1)} \neq \hat{\mathbf{H}}_q^{(j)}\right) = o\left(\frac{1}{\log_2(P)}\right), \tag{H.36}$$

where $\hat{\mathbf{H}}_q^{(j)} = \mathcal{Q}_u(\hat{\mathbf{H}}^{(j)})$ and $\hat{\mathbf{H}}_q^{(j)\leftarrow(1)}$ is the MAP estimator of $\hat{\mathbf{H}}_q^{(j)}$ given $\hat{\mathbf{H}}^{(1)}$. We start by noting that, by definition of the MAP estimator,

$$\Pr\left(\hat{\mathbf{H}}_q^{(j)\leftarrow(1)} \neq \hat{\mathbf{H}}_q^{(j)}\right) \leq \Pr\left(\mathcal{Q}_u(\hat{\mathbf{H}}^{(1)}) \neq \hat{\mathbf{H}}_q^{(j)}\right). \tag{H.37}$$

Since $\mathrm{Re}(\hat{\mathbf{H}}_{i,k}^{(1)})$ and $\mathrm{Im}(\hat{\mathbf{H}}_{i,k}^{(1)})$ are i.i.d. for any i, k, it follows that

$$\Pr\left(\mathcal{Q}_u(\hat{\mathbf{H}}^{(1)}) \neq \hat{\mathbf{H}}_q^{(j)}\right) \leq 2KN_T \Pr\left(\mathcal{Q}_u(\mathrm{Re}(\hat{\mathbf{H}}_{1,1}^{(1)})) \neq \mathcal{Q}_u(\mathrm{Re}(\hat{\mathbf{H}}_{1,1}^{(j)}))\right), \tag{H.38}$$

where we have selected w.l.o.g. the real part of the (1,1) channel element. Hence, it is sufficient to obtain the probability of disagreement for $\mathrm{Re}(\hat{\mathbf{H}}_{1,1}^{(j)})$. For that purpose, we follow the same approach of Appendix F.2 and we split each reconstruction level in two parts: The *edge* of the cell and the *center* of the cell. This is done in order to show that, as P increases, the probability of disagreement vanishes if $\hat{\mathbf{H}}_{1,1}^{(1)}$ is in the *center* of the quantization level and, besides this, that the probability that $\hat{\mathbf{H}}_{1,1}^{(1)}$ is in the *edge* area also vanishes. We rigorously show that in the following. Before starting, we introduce the simplified notation $\mathrm{h}^{(j)} \triangleq \mathrm{Re}(\hat{\mathbf{H}}_{1,1}^{(j)})$ to ease the readability. Accordingly, we also introduce the notation $\mathrm{h} \triangleq \mathrm{Re}(\mathbf{H}_{1,1})$ and $\delta \triangleq \mathrm{Re}(\boldsymbol{\delta}_{1,1}^{(j)})$ such that $\mathrm{h}^{(j)} = \breve{z}^{(j)}\mathrm{h} + z^{(j)}\delta^{(j)}$, with $z^{(j)} = \bar{P}^{-\alpha^{(j)}}$ and $\breve{z}^{(j)} = \sqrt{1-(z^{(j)})^2}$. Furthermore, similar to Section 7.5, we introduce the notations $\breve{z}_{\mathrm{inv}}^{(j)} = \frac{1}{\breve{z}^{(j)}}$ and $z_n^{(j)} = \frac{z^{(j)}}{\breve{z}^{(j)}}$.

H.5.1 Egde and Center of the Reconstruction Level

This division is the same we applied in Appendix F.2.1. We recall it here for convenience. Let ℓ_n be the n-th quantization level of $\mathcal{Q}_u$, $n \in \mathbb{Z}$, with $\ell_0 = 0$. Let us define L_n as the input interval that outputs ℓ_n, i.e.,

$$L_n \triangleq \{x \mid \mathcal{Q}_u(x) = \ell_n\}. \tag{H.39}$$

L_n has a range $[L_n^{\min}, L_n^{\max})$ such that $|L_n| \triangleq L_n^{\max} - L_n^{\min} = \bar{P}^{-\alpha_q}$. We split L_n in two areas, the *edge* area E_n and the *center* area C_n, depicted in Fig. F.1. The *edge* area is defined as the part of L_n that is at most at distance $\bar{P}^{-c_e\alpha_q}$ of the boundary of the cell, with $c_e > 1$, i.e.,

$$E_n \triangleq \left\{x \in L_n \,|\, x - L_n^{\min} < \bar{P}^{-c_e\alpha_q} \ \vee \ L_n^{\max} - x < \bar{P}^{-c_e\alpha_q}\right\}. \tag{H.40}$$

The *center* area is given by $C_n \triangleq \{x \in L_n \backslash E_n\}$. Intuitively, the probability of disagreement is very high if $\mathrm{h}^{(1)}$ lies in the edge area E_n, whereas this probability vanishes in the central area C_n. Mathematically, we have that

$$\begin{aligned}\Pr\left(\mathcal{Q}_u(\mathrm{h}^{(1)}) \neq \mathcal{Q}_u(\mathrm{h}^{(j)})\right) \leq \Pr\left(\mathrm{h}^{(1)} \in \bigcup_{n\in\mathbb{Z}} E_n\right) \\ + \Pr\left(\mathcal{Q}_u(\mathrm{h}^{(1)}) \neq \mathcal{Q}_u(\mathrm{h}^{(j)}) \mid \mathrm{h}^{(1)} \in \bigcup_{n\in\mathbb{Z}} C_n\right).\end{aligned} \tag{H.41}$$

H.5.2 Probability of Belonging to the Edge Area

Suppose an arbitrary quantization level ℓ_n. Let $f_{L_n}^{\max}$ be the maximum value of the PDF of $\mathrm{h}^{(1)}$ in $L_n = \{x \mid \mathcal{Q}_u(x) = \ell_n\}$. Then, the probability that $\mathrm{h}^{(1)}$ is in E_n is bounded by

$$\begin{aligned}\Pr\left(\mathrm{h}^{(1)} \in E_n\right) &\leq f_{L_n}^{\max}|E_n| \\ &= 2f_{L_n}^{\max}\bar{P}^{-c_e\alpha_q},\end{aligned} \tag{H.42}$$

where $|E_n|$ denotes the length of E_n. The standard normal distribution has a derivative that is, at most, $1/\sqrt{2\pi e}$. Thus, the probability of being in L_n satisfies

$$\begin{aligned}\Pr\left(\mathrm{h}^{(1)} \in L_n\right) &\geq \left(f_{L_n}^{\max} - 1/\sqrt{2\pi e}|L_n|\right)|L_n| \\ &= \left(f_{L_n}^{\max} - 1/\sqrt{2\pi e}\bar{P}^{-\alpha_q}\right)\bar{P}^{-\alpha_q}.\end{aligned} \tag{H.43}$$

Hence, the probability that $\mathrm{h}^{(1)}$ is in E_n, given that it is in L_n, satisfies $\forall n$ that

$$\begin{aligned}\Pr\left(\mathrm{h}^{(1)} \in E_n \mid L_n\right) &= \frac{\Pr\left(\mathrm{h}^{(1)} \in E_n\right)}{\Pr\left(\mathrm{h}^{(1)} \in L_n\right)} \\ &\leq \frac{2f_{L_n}^{\max}}{\left(f_{L_n}^{\max} - 1/\sqrt{2\pi e}\bar{P}^{-\alpha_q}\right)}\bar{P}^{-(c_e-1)\alpha_q}.\end{aligned} \tag{H.44}$$

Let us define $g_{\max}$ as $g_{\max} \triangleq \max_{n\in\mathbb{Z}} \frac{2f_{L_n}^{\max}}{\left(f_{L_n}^{\max} - 1/\sqrt{2\pi e}\bar{P}^{-\alpha_q}\right)}$. Note that $g_{\max} = \Theta(1)$. Hence, from (H.44) and the fact that $\sum_{n\in\mathbb{Z}} \Pr\left(\mathrm{h}^{(1)} \in L_n\right) = 1$ we can write

$$\begin{aligned}\Pr\left(\mathrm{h}^{(1)} \in \bigcup_{n\in\mathbb{Z}} E_n\right) &= \sum_{n\in\mathbb{Z}} \frac{\Pr\left(\mathrm{h}^{(1)} \in E_n\right)}{\Pr\left(\mathrm{h}^{(1)} \in L_n\right)} \Pr\left(\mathrm{h}^{(1)} \in L_n\right) \\ &\leq g_{\max}\bar{P}^{-(c_e-1)\alpha_q} \sum_{n\in\mathbb{Z}} \Pr\left(\mathrm{h}^{(1)} \in L_n\right) \\ &= \mathcal{O}\left(\bar{P}^{-(c_e-1)\alpha_q}\right).\end{aligned} \tag{H.45}$$

Consequently, it holds that

$$\Pr\left(\mathrm{h}^{(1)} \in \bigcup_{n\in\mathbb{Z}} E_n\right) = o\left(\frac{1}{\log_2(P)}\right). \tag{H.46}$$

H.5.3 Probability of Disagreement in the Center Area

From (F.23) in Appendix F.2.3, it follows that

$$\Pr\left(\mathcal{Q}_u(\mathrm{h}^{(1)}) \neq \mathcal{Q}_u(\mathrm{h}^{(j)}) \;\middle|\; \mathrm{h}^{(1)} \in \bigcup_{n\in\mathbb{Z}} C_n\right) \leq \frac{1}{\Pr\left(\mathrm{h}^{(1)} \in \bigcup_{n\in\mathbb{Z}} C_n\right)} \frac{\mathbb{E}\left[\left|\mathrm{h}^{(1)} - \mathrm{h}^{(j)}\right|\right]}{\bar{P}^{-c_e\alpha_q}}.$$

In the following, we obtain the expectation $\mathbb{E}\left[\left|\mathrm{h}^{(1)} - \mathrm{h}^{(j)}\right|\right]$. Then,

$$\mathrm{h}^{(1)} - \mathrm{h}^{(j)} = \mathrm{h}(\breve{z}^{(1)} - \breve{z}^{(j)}) + (z^{(1)}\delta^{(1)} - z^{(j)}\delta^{(j)}). \tag{H.47}$$

From the assumption of Gaussian variables, it holds that

$$\mathrm{h}(\breve{z}^{(1)} - \breve{z}^{(j)}) \sim \mathcal{N}\left(0,\ (\breve{z}^{(1)} - \breve{z}^{(j)})^2\right), \tag{H.48}$$

$$z^{(1)}\delta^{(1)} - z^{(j)}\delta^{(j)} \sim \mathcal{N}\left(0,\ (z^{(1)})^2 + (z^{(j)})^2\right). \tag{H.49}$$

Since $\mathrm{h}^{(1)}$ is independent of $\delta^{(1)}$ and $\delta^{(j)}$, it follows that

$$\mathrm{h}^{(1)} - \mathrm{h}^{(j)} \sim \mathcal{N}\left(0, \sigma_d^2\right), \tag{H.50}$$

where σ_d^2 is given by

$$\sigma_d^2 = (\breve{z}^{(1)} - \breve{z}^{(j)})^2 + (z^{(1)})^2 + (z^{(j)})^2. \tag{H.51}$$

Substituting the variables for their values yields

$$\sigma_d^2 = 2\left(1 - \sqrt{1 - P^{-\alpha^{(1)}} - P^{-\alpha^{(j)}} + P^{-\alpha^{(1)}-\alpha^{(j)}}}\right). \tag{H.52}$$

Furthermore, if $\mathrm{h}^{(1)} - \mathrm{h}^{(j)}$ is drawn from a zero-mean Normal distribution of variance σ_d^2, $|\mathrm{h}^{(1)} - \mathrm{h}^{(j)}|$ is distributed as a half-normal distribution of mean

$$\mathbb{E}\left[\left|\mathrm{h}^{(1)} - \mathrm{h}^{(j)}\right|\right] = \sigma_d\sqrt{\frac{2}{\pi}}. \tag{H.53}$$

From (H.52) and the fact that, $\forall\ 0 \leq x \leq 1$, $\sqrt{1-x} \geq 1-x$, it follows that

$$\begin{aligned}\mathbb{E}\left[\left|\mathrm{h}^{(1)} - \mathrm{h}^{(j)}\right|\right] &\leq \sqrt{\frac{4}{\pi}\left(P^{-\alpha^{(1)}} + P^{-\alpha^{(j)}} - P^{-\alpha^{(1)}-\alpha^{(j)}}\right)} \\ &= \mathcal{O}(\bar{P}^{-\alpha^{(j)}}).\end{aligned} \tag{H.54}$$

Besides this, it holds from (H.46) that

$$\begin{aligned}\Pr\left(\mathrm{h}^{(1)} \in \bigcup_{n\in\mathbb{Z}} C_n\right) &= 1 - \Pr\left(\mathrm{h}^{(1)} \in \bigcup_{n\in\mathbb{Z}} E_n\right) \\ &= 1 - \mathcal{O}(\bar{P}^{-(c_e-1)\alpha_q}).\end{aligned} \tag{H.55}$$

Both (H.54) and (H.55) lead to

$$\begin{aligned}\Pr\left(\mathcal{Q}_u(\mathrm{h}^{(1)}) \neq \mathcal{Q}_u(\mathrm{h}^{(j)}) \;\middle|\; \mathrm{h}^{(1)} \in \bigcup_{n\in\mathbb{Z}} C_n\right) &\leq \frac{1}{\Pr\left(\mathrm{h}^{(1)} \in \bigcup_{n\in\mathbb{Z}} C_n\right)} \frac{\mathbb{E}\left[\left|\mathrm{h}^{(1)} - \mathrm{h}^{(j)}\right|\right]}{\bar{P}^{-c_e\alpha_q}} \\ &= \mathcal{O}\left(\bar{P}^{c_e\alpha_q - \alpha^{(j)}}\right) \\ &= o\left(\frac{1}{\log_2(P)}\right).\end{aligned} \tag{H.56}$$

The last inequality is obtained only if $c_e\alpha_q < \alpha^{(j)}$, $\forall j \in \mathbb{N}_M$. Thus, it follows from (H.56) that it is necessary to satisfy that $c_e\alpha_q < \alpha^{(j)}$, $\forall j \in \mathbb{N}_M$. Since for any $\alpha_q < \alpha^{(j)}$ we can find a $c_e > 1$ such that $c_e\alpha_q < \alpha^{(j)}$, any $\alpha_q < \alpha^{(j)}$ will satisfy (H.56) as long as a correct c_e is selected.

H.5.4 Assembling Probabilities

Recalling (H.41), we make use of (H.46) and (H.56) to show that

$$\Pr\left(\mathcal{Q}_u(\mathrm{h}^{(1)}) \neq \mathcal{Q}_u(\mathrm{h}^{(j)})\right) = o\left(\frac{1}{\log_2(P)}\right), \tag{H.57}$$

what concludes the proof of Lemma 7.1. ■

H.6 Proof of Lemma 7.2 (Error on Naive Precoder)

In this appendix we show that

$$\mathbb{E}\left[\|\mathbf{v}_{i,2} - \mathbf{w}_{i,2}\|^2\right] = \mathcal{O}\left(P^{-\alpha_q}\right). \tag{H.58}$$

Then, (H.58) is straightforwardly generalized for any $\mathbb{E}\left[\|\mathbf{v}_{i,k}-\mathbf{w}_{i,k}\|^2\right]$, $k\in\mathbb{N}_M\backslash 1$. In order to prove (H.58), we make use of the fact that, as presented in Section 7.2.1, we assume that the precoding vectors and matrices can be expressed as a combination of summations, products, and generalized inverses of the channel estimate. Note that with the previous operations, it is also possible to compute divisions and norms of the channel estimate coefficients.

First, note that both $\mathbf{w}_{i,2}$ and $\mathbf{v}_{i,2}$ are obtained following the same algorithm but based on different information (input). Specifically, $\mathbf{w}_{i,2}$ is computed on the basis of $\hat{\mathbf{H}}_q^{(2)}=\mathcal{Q}_u(\hat{\mathbf{H}}^{(2)})$, where $\mathcal{Q}_u$ is a scalar uniform quantizer with quantization step $q=\bar{P}^{-\alpha_q}$, and $\mathbf{v}_{i,2}$ on the basis of $\hat{\mathbf{H}}^{(1)}$. Similar to the previous appendix, let $\mathrm{h}_q^{(2)}$ (resp. $\mathrm{h}^{(j)}$ and h) denote the real or imaginary part of an arbitrary element of the matrix $\hat{\mathbf{H}}_q^{(2)}$ (resp. $\hat{\mathbf{H}}^{(j)}$ and $\mathbf{H}$). Let us define $\mathrm{h}_\varsigma^{(2)}$ as

$$\mathrm{h}_\varsigma^{(2)}\triangleq \mathrm{h}^{(2)}+\varsigma_q, \tag{H.59}$$

where ς_q is distributed as a binary symmetric distribution with points $[-q,\,q]$, independent of the other variables, such that $\sigma_{\varsigma_q}^2=q^2$. Note that the error $\mathrm{h}_q^{(2)}-\mathrm{h}^{(1)}$ has smaller or equal variance than $\mathrm{h}_\varsigma^{(2)}-\mathrm{h}^{(1)}=\mathrm{h}^{(2)}-\mathrm{h}^{(1)}+\varsigma_q$. Hence, we can assume that $\mathbf{w}_{i,2}$ is computed on the basis of the estimate $\mathrm{h}_\varsigma^{(2)}$ as increasing the error variance can only hurt. Consequently, the error $\xi\triangleq \mathrm{h}_\varsigma^{(2)}-\mathrm{h}^{(1)}$ has a variance

$$\begin{aligned}\sigma_\xi^2&\leq \sigma_d^2+\sigma_{\varsigma_q}^2+2\sigma_d\sigma_{\varsigma_q}\\&=\mathcal{O}(P^{-\alpha_q}),\end{aligned} \tag{H.60}$$

where σ_d^2 is given in (H.52). Therefore, we can write that

$$\mathrm{h}_\varsigma^{(2)}=\mathrm{h}^{(1)}+\bar{P}^{-\alpha_q}\delta_\xi, \tag{H.61}$$

where δ_ξ is a variable with variance $\Theta(1)$ and bounded density. We continue by showing that the error variance remains being at most $\Theta(P^{-\alpha_q})$ after applying addition, product, inverse or pseudo-inverse operations. Afterward, based on those results, we prove (H.58).

H.6.1 Error in the Addition

Let $\mathrm{a}_\xi^{(2)}$, $\mathrm{b}_\xi^{(2)}$, be distributed as (H.61), i.e., $\mathrm{a}_\xi^{(2)}\triangleq \mathrm{a}^{(1)}+\bar{P}^{-\alpha_q}\delta_\xi^{\mathrm{a}}$, $\mathrm{b}_\xi^{(2)}\triangleq \mathrm{b}^{(1)}+\bar{P}^{-\alpha_q}\delta_\xi^{\mathrm{b}}$, where $\delta_\xi^{\mathrm{a}},\delta_\xi^{\mathrm{b}}$, are variables with variance $\Theta(1)$ and bounded density. It is easy to see that, for any $\mathrm{a}_\xi^{(2)}$, $\mathrm{b}_\xi^{(2)}$,

$$\begin{aligned}\mathrm{a}_\xi^{(2)}+\mathrm{b}_\xi^{(2)}&=\mathrm{a}^{(1)}+\bar{P}^{-\alpha_q}\delta_\xi^{\mathrm{a}}+\mathrm{b}^{(1)}+\bar{P}^{-\alpha_q}\delta_\xi^{\mathrm{b}}\\&=\mathrm{a}^{(1)}+\mathrm{b}^{(1)}+\bar{P}^{-\alpha_q}(\delta_\xi^{\mathrm{a}}+\delta_\xi^{\mathrm{b}}).\end{aligned} \tag{H.62}$$

This implies that the error variance of the sum is also $\mathcal{O}(P^{-\alpha_q})$.

H.6.2 Error in the Product

It follows that

$$\begin{aligned} \mathrm{a}_{\xi}^{(2)}\mathrm{b}_{\xi}^{(2)} &= \left(\mathrm{a}^{(1)} + \bar{P}^{-\alpha_q}\delta_{\xi}^{\mathrm{a}}\right)\left(\mathrm{b}^{(1)} + \bar{P}^{-\alpha_q}\delta_{\xi}^{\mathrm{b}}\right) \\ &= \mathrm{a}^{(1)}\mathrm{b}^{(1)} + \bar{P}^{-\alpha_q}\left(\mathrm{a}^{(1)}\delta_{\xi}^{\mathrm{b}} + \mathrm{b}^{(1)}\delta_{\xi}^{\mathrm{a}} + \bar{P}^{-\alpha_q}\delta_{\xi}^{\mathrm{a}}\delta_{\xi}^{\mathrm{b}}\right), \end{aligned} \tag{H.63}$$

what implies that the product also maintains the scaling of the variance as $\mathcal{O}(P^{-\alpha_q})$. Moreover, as the sum and product of matrices is a composition of sums and products of its coefficients, the result extends to any two matrices of suitable dimension.

H.6.3 Error in the Inverse

Let us first assume that $\mathbf{A}_{\xi}^{(2)}$ and $\mathbf{A}^{(1)}$ are square matrices with full rank with probability one, and with coefficients following (H.61). We can then write that

$$(\mathbf{A}_{\xi}^{(2)})^{-1} = \left(\mathbf{A}^{(1)} + \bar{P}^{-\alpha_q}\boldsymbol{\Delta}_{\xi}^{\mathbf{A}}\right)^{-1} \tag{H.64}$$

$$= (\mathbf{A}^{(1)})^{-1} - \bar{P}^{-\alpha_q}(\mathbf{A}^{(1)})^{-1}\boldsymbol{\Delta}_{\xi}^{\mathbf{A}}\left(\mathbf{A}^{(1)} + \bar{P}^{-\alpha_q}\boldsymbol{\Delta}_{\xi}^{\mathbf{A}}\right)^{-1} \tag{H.65}$$

which is obtained from the Woodbury matrix identity [164]. Hence, the error variance of the inverse is again $\mathcal{O}(P^{-\alpha_q})$. Once that it is proved that the inverse operation generates an error with variance $\mathcal{O}(P^{-\alpha_q})$, we extend it for the Moore–Penrose inverse (pseudo-inverse) $(\mathbf{A}_{\xi}^{(2)})^{\dagger}$. We assume (as throughout the rest of the document) that each sub-matrix has maximum rank, i.e.,

$$\operatorname{rank}\left(\mathbf{A}_{\xi}^{(2)} \in \mathbb{C}^{N\times M}\right) = \min(N, M). \tag{H.66}$$

Let us assume that $\mathbf{A}_{\xi}^{(2)}$ is full row-rank matrix, i.e., $N \leq M$. Under this assumption, the pseudo-inverse is given by

$$(\mathbf{A}_{\xi}^{(2)})^{\dagger} = (\mathbf{A}_{\xi}^{(2)})^{\mathrm{H}}\left(\mathbf{A}_{\xi}^{(2)}(\mathbf{A}_{\xi}^{(2)})^{\mathrm{H}}\right)^{-1}. \tag{H.67}$$

The case in which $\mathbf{A}_{\xi}^{(2)}$ is full column-rank matrix ($N \geq M$) will follow the same steps and thus we omit it. It follows from (H.63) that $\mathbf{A}_{\xi}^{(2)}(\mathbf{A}_{\xi}^{(2)})^{\mathrm{H}} = \mathbf{A}^{(1)}(\mathbf{A}^{(1)})^{\mathrm{H}} + \bar{P}^{-\alpha_q}\boldsymbol{\Delta}_{\mathrm{eq}}$, where $\boldsymbol{\Delta}_{\mathrm{eq}}$ has variance $\Theta(1)$. This, together with (H.65), implies that

$$\left(\mathbf{A}_{\xi}^{(2)}(\mathbf{A}_{\xi}^{(2)})^{\mathrm{H}}\right)^{-1} = \left(\mathbf{A}^{(1)}(\mathbf{A}^{(1)})^{\mathrm{H}}\right)^{-1} + \bar{P}^{-\alpha_q}\boldsymbol{\Delta}'_{\mathrm{eq}}, \tag{H.68}$$

and by applying again (H.63) it holds

$$(\mathbf{A}_{\xi}^{(2)})^{\dagger} = (\mathbf{A}^{(1)})^{\mathrm{H}} \left(\mathbf{A}^{(1)} (\mathbf{A}^{(1)})^{\mathrm{H}} \right)^{-1} + \bar{P}^{-\alpha_q} \boldsymbol{\Delta}_{\mathrm{eq}}^{''}, \tag{H.69}$$

where $\boldsymbol{\Delta}_{\mathrm{eq}}^{'}$ and $\boldsymbol{\Delta}_{\mathrm{eq}}^{''}$ have variance $\Theta(1)$. As explained in [145], under the assumption that $\hat{\mathbf{X}}$ is a full row-rank matrix, any generalized inverse may be expressed as $\hat{\mathbf{X}}^{-} = \hat{\mathbf{X}}^{\dagger} + \mathbf{P}_{\perp}\mathbf{U}$. Hence, similar result could be obtained for a broad set of generalized inverse.

H.6.4 Error Variance of the Difference of Precoders

The centralized precoder $\mathbf{v}_{i,2}$ is based on $\hat{\mathbf{H}}^{(1)}$, i.e., $\mathbf{V} = \mathcal{V}(\hat{\mathbf{H}}^{(1)})$. The distributed precoder at TX 2 is based on its own quantized CSIT $\hat{\mathbf{H}}_q^{(2)}$, then $\mathbf{w}_{i,2}$ is obtained from $\mathbf{W} = \mathcal{V}(\hat{\mathbf{H}}_q^{(2)})$. Based on the previous results and the definition of linear precoders, it follows that we can write the distributed precoder based on the CSIT of TX 2 ($\hat{\mathbf{H}}_q^{(2)}$) as

$$\mathbf{w}_{i,2} = \mathbf{v}_{i,2} + \bar{P}^{-\alpha_q} \mathbf{e}_{\mathbf{w}}, \tag{H.70}$$

where $\mathbb{E}\left[\|\mathbf{e}_{\mathbf{w}}\|^2\right] = \mathcal{O}(1)$. Consequently,

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{v}_{i,2} - \mathbf{w}_{i,2}\|^2\right] &= \mathbb{E}\left[\|\mathbf{v}_{i,2} - (\mathbf{v}_{i,2} + \bar{P}^{-\alpha_q}\mathbf{e}_{\mathbf{w}})\|^2\right] \\ &= P^{-\alpha_q}\, \mathbb{E}\left[\|\mathbf{e}_{\mathbf{w}}\|^2\right] \\ &= \mathcal{O}\left(P^{-\alpha_q}\right). \end{aligned} \tag{H.71}$$

Moreover, since $\mathbb{E}[\|\mathbf{x}\|] \leq \sqrt{\mathbb{E}[\|\mathbf{x}\|^2]}$, it follows that

$$\mathbb{E}\left[\|\mathbf{v}_{i,j} - \mathbf{w}_{i,j}\|\right] = \mathcal{O}\left(\bar{P}^{-\alpha_q}\right), \tag{H.72}$$

what concludes the proof of Lemma 7.2. ■

H.7 Proof of Corollary 7.4

We show in the following that the use of CD-ZF or AP-ZF in a setting with $\alpha^{(M)} = 0$ and instantaneous power constraint leads to $\lim_{P\to\infty} R^{\star}(\alpha^{(1)}) - R(\boldsymbol{\alpha}_M) = \infty$. The main limitation that leads to this result is that the power outage probability does not vanish at high SNR. We assume a per-TX power constraint, i.e., $\|\mathbf{T}_j\| = \|\mu[\mathbf{w}_{1,j}, \ldots, \mathbf{w}_{K,j}]\| \leq 1$, so as to simplify the notation. The proof for a per-antenna power constraint follows directly.

Let $\mathbf{w}_i^o$ represent the distributed precoder (CD-ZF or AP-ZF) before taking into account the instantaneous power constraint and let $\mathbf{w}_i$ represent the precoder after power

normalization. This means that

$$\mathbf{w}_i = \begin{cases} \mathbf{w}_i^o & \text{if } \|\mathbf{T}_1\| \leq 1 \\ \left[\frac{\mathbf{w}_{i,1}^o}{\|\mathbf{T}_1\|},\ \mathbf{w}_{i,2}^o,\ \ldots,\ \mathbf{w}_{i,K}^o\right]^{\mathrm{T}} & \text{otherwise.} \end{cases} \tag{H.73}$$

We focus on the rate of RX 1, and the analysis of the rate of other RXs follows by symmetry. Let $\mathcal{P}_C$ denote the event that TX 1 is in power outage *and* that $|\mu \mathbf{h}_1 \mathbf{w}_2|^2 \geq |\mu \mathbf{h}_1 \mathbf{w}_2^o|^2 + c$, where c is a constant and $c > 0$. The rate gap $\Delta \mathrm{R}_1$ can be decomposed as

$$\begin{aligned} \Delta \mathrm{R}_1 &= \Pr\left(\mathcal{P}_C\right) \Delta \mathrm{R}_{1|\mathcal{P}_C} + \Pr\left(\mathcal{P}_C^{\mathsf{c}}\right) \Delta \mathrm{R}_{|\mathcal{P}_C^{\mathsf{c}}} \\ &\geq \Pr\left(\mathcal{P}_C\right) \Delta \mathrm{R}_{1|\mathcal{P}_C}, \end{aligned} \tag{H.74}$$

where we have assumed that the ideal centralized rate is achieved conditioned on $\mathcal{P}_C^{\mathsf{c}}$. The achievable rate conditioned on $\mathcal{P}_C$ yields

$$\begin{aligned} R_{1|\mathcal{P}_C}(\boldsymbol{\alpha}) &\leq \mathbb{E}_{|\mathcal{P}_C}\left[\log_2\left(1 + \frac{\frac{P}{K}\left|\mu \mathbf{h}_1 \mathbf{w}_1\right|^2}{1 + \frac{P}{K}\left|\mu \mathbf{h}_1 \mathbf{w}_2^o\right|^2 + \frac{P}{K}c}\right)\right] \\ &\leq \mathbb{E}_{|\mathcal{P}_C}\left[\log_2\left(1 + \frac{1}{c}\left|\mu \mathbf{h}_1 \mathbf{w}_1\right|^2\right)\right]. \end{aligned} \tag{H.75}$$

From the instantaneous power constraint assumption, this term does not scale as function of P and it is upper-bounded. Let us define the constant $m_c < \infty$ such that it satisfies that $\mathbb{E}_{|\mathcal{P}_C}\left[\log_2\left(1 + c^{-1}\left|\mu \mathbf{h}_1 \mathbf{w}_1\right|^2\right)\right] \leq m_c$. Hence, the rate gap conditioned on $\mathcal{P}_C$ is

$$\begin{aligned} \Delta \mathrm{R}_{1|\mathcal{P}_C} &= R^{\star}_{1|\mathcal{P}_C}(\alpha^{(1)}) - R_{1|\mathcal{P}_C}(\boldsymbol{\alpha}) \\ &\geq R^{\star}_{1|\mathcal{P}_C}(\alpha^{(1)}) - m_c \\ &= \infty. \end{aligned} \tag{H.76}$$

It remains to prove that $\Pr\left(\mathcal{P}_C\right) = \Theta(1)$, i.e., that it does not vanish at high SNR, since together with (H.76) and (H.74) it implies that the rate gap is unbounded. In the following, we compute the probability of $\mathcal{P}_C$, i.e., of having $|\mu \mathbf{h}_1 \mathbf{w}_2|^2 \geq |\mu \mathbf{h}_1 \mathbf{w}_2^o|^2 + c$. Let $\boldsymbol{\theta}_{\mathbf{w}}$ be defined as $\boldsymbol{\theta}_{\mathbf{w}} \triangleq \mathbf{w}_2 - \mathbf{w}_2^o$. Hence,

$$\Pr(|\mu \mathbf{h}_1 \mathbf{w}_2|^2 \geq |\mu \mathbf{h}_1 \mathbf{w}_2^o|^2 + c) = \Pr\left(|\bar{P}^{-\alpha^{(1)}} \boldsymbol{\delta}_1 \mathbf{w}_2^o + \mathbf{h}_1 \boldsymbol{\theta}_{\mathbf{w}}|^2 - |\bar{P}^{-\alpha^{(1)}} \boldsymbol{\delta}_1 \mathbf{w}_2^o|^2 \geq \frac{c}{\mu^2}\right),$$

which comes from the fact that $\mathbf{h}_1 \mathbf{w}_2^o = \bar{P}^{-\alpha^{(1)}} \boldsymbol{\delta}_1^{(1)} \mathbf{w}_2^o$. Note that $\bar{P}^{-\alpha^{(1)}} \boldsymbol{\delta}_1 \mathbf{w}_2^o$ converges almost surely to 0 [101], what leads to

$$\lim_{P \to \infty} \Pr\left(|\mu \mathbf{h}_1 \mathbf{w}_2|^2 \geq |\mu \mathbf{h}_1 \mathbf{w}_2^o|^2 + c\right) = \Pr\left(|\mathbf{h}_1 \boldsymbol{\theta}_{\mathbf{w}}|^2 \geq \frac{c}{\mu^2}\right). \tag{H.77}$$

From (H.73) and the fact that $\mathcal{P}_C \Rightarrow \|\mathbf{T}_1\| > 1$,

$$\Pr\left(|\mathbf{h}_1\boldsymbol{\theta}_{\mathbf{w}}|^2 \geq \frac{c}{\mu^2}\right) = \Pr\left(\left|\mathbf{h}_{1,1}\frac{\mathbf{w}_{1,1}^o}{\|\mathbf{T}_1\|}\right|^2 \geq \frac{c}{\mu^2}\right). \tag{H.78}$$

We recall that the CD-ZF precoder is given in (7.22) by

$$\mathbf{w}_{\ell,1} = \mathbf{v}_{\ell,1} + \hat{\mathbf{H}}_{\bar{\ell},1}^{\dagger}\hat{\mathbf{H}}_{\bar{\ell},\bar{1}}(\mathbf{v}_{\ell,\bar{1}} - \mathbf{w}_{\ell,\bar{1}}) \tag{H.79}$$

and the AP-ZF precoder [82, 100] is given by

$$\mathbf{w}_{\ell,1} = -\hat{\mathbf{H}}_{\bar{\ell},1}^{\dagger}\hat{\mathbf{H}}_{\bar{\ell},\bar{1}}\mathbf{w}_{\ell,\bar{1}}. \tag{H.80}$$

Since the accuracy of $\mathbf{w}_{\ell,\bar{1}}$ does not scale with the SNR, the difference $\mathbf{v}_{\ell,\bar{1}} - \mathbf{w}_{\ell,\bar{1}}$ of (H.79) has the same probability distribution for any value of P –conversely to the case with $\alpha^{(M)} > 0$, where it vanishes–. This implies that the probability distribution of $|\mathbf{h}_{1,1}\frac{\mathbf{w}_{1,1}^o}{\|\mathbf{T}_1\|}|^2$ in (H.78) is not affected by the value of P, and since $\mathbf{h}_{1,1}$ is unbounded, there exists an $\varepsilon > 0$ such that

$$\Pr\left(|\mathbf{h}_1\boldsymbol{\theta}_{\mathbf{w}}|^2 \geq \frac{c}{\mu^2}\right) \geq \varepsilon, \qquad \forall \frac{c}{\mu^2} > 0. \tag{H.81}$$

Hence, the probability does not vanish when $P \to \infty$ and consequently Corollary 7.4 is proven. ■

H.8 Proof of Theorem 7.2

In this section we prove that, if $N_1 > K - 1$, and $\alpha^{(j)} = 0$ for any $j > 1$, then, under average power constraint $\mathbb{E}\left[\|\mathbf{T}_{j,n}\|^2\right] \leq 1$, $\forall n \in \mathbb{N}_{N_j}$, $j \in \mathbb{N}_M$, it holds that

$$\lim_{P\to\infty} R^{\star}(\alpha^{(1)}) - R(\boldsymbol{\alpha}_M) \leq \log_2\left(\mathbb{E}\left[\|\hat{\mathbf{H}}_{K,1}^{\dagger}\|^2\right]\right) + \log_2(K^2) + \log_2\left(4(K-1)\right). \tag{H.82}$$

Let us assume that the TXs from TX 2 to TX M precode with a known, fixed precoder, for example

$$\mathbf{w}_{i,j} = \frac{1}{\sqrt{K}}\mathbf{1}_{N_j\times 1}, \tag{H.83}$$

for any $i \in \mathbb{N}_K$, $j \in \mathbb{N}_M \backslash 1$. The final precoding vector will be $\mu\mathbf{w}_{i,j}$, where $\mu \leq 1$ is an average parameter to satisfy the power constraint. Note that (H.83) is chosen for sake of simplicity, but the bound in Theorem 7.2 would be modified if $\mathbf{w}_{i,j}$ is defined differently.

In a similar way as in (7.22), TX 1 computes its precoder so as to cancel out the

interference generated by the other TXs. If $|\hat{\mathbf{h}}_i^{(1)}\mathbf{w}_i|^2$ were statistically equivalent to $|\hat{\mathbf{h}}_i^{(1)}\mathbf{v}_i|^2$, it would be enough to ensure that $\hat{\mathbf{h}}_i^{(1)}\mathbf{w}_\ell = \hat{\mathbf{h}}_i^{(1)}\mathbf{v}_\ell \ \forall \ell \in \mathbb{N}_K \backslash i$. However, the proof of that analogy is still an open and challenging problem. Hence, we design the precoder so as to ensure that $\hat{\mathbf{h}}_i^{(1)}\mathbf{w}_i = \hat{\mathbf{h}}_i^{(1)}\mathbf{v}_i$. The precoder at TX 1 for RX ℓ is

$$\mathbf{w}_{\ell,1} = \hat{\mathbf{H}}_{K,1}^{\dagger}\left(\begin{bmatrix} \hat{\mathbf{h}}_\ell^{(1)}\mathbf{v}_\ell \\ \mathbf{0}_{K-1\times 1} \end{bmatrix} - \hat{\mathbf{H}}_{K,\bar{1}}\mathbf{w}_{\ell,\bar{1}} \right), \tag{H.84}$$

where $\hat{\mathbf{H}}_{K,1}$ represents the first N_1 columns of the channel matrix –channel from TX 1 towards all the RXs– and $\hat{\mathbf{H}}_{K,\bar{1}}$ denotes the remaining $N_T - N_1$ columns –channel coefficients from the rest of TXs towards all RXs–. The precoder at TX 1 is denoted by $\mathbf{T}_1 \in \mathbb{C}^{N_1\times K}$, and the precoding coefficients at the n-th antenna of TX 1 as $\mathbf{T}_{1,n} \in \mathbb{C}^{1\times K}$. Let $\mathbf{T}_1^o$ be defined as $\mathbf{T}_1^o \triangleq [\mathbf{w}_{1,1}, \ \ldots, \ \mathbf{w}_{K,1}]$. The precoding coefficients are equally distributed for every RX ($\forall \ell \in \mathbb{N}_K$) and for every antenna at TX 1 ($\forall n \in N_1$). Hence, it follows that

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{T}_{1,n}^o\|^2\right] &= \frac{1}{N_1}\mathbb{E}\left[\|\mathbf{T}_1^o\|^2\right] \\ &= \frac{K}{N_1}\mathbb{E}\left[\|\mathbf{w}_{1,1}\|^2\right]. \end{aligned} \tag{H.85}$$

With per-antenna average power constraint, the precoder $\mathbf{T}_1 \in \mathbb{C}^{N_1\times K}$ will be given by

$$\mathbf{T}_1 = \mu \begin{bmatrix} \mathbf{w}_{1,1} & \ldots & \mathbf{w}_{K,1} \end{bmatrix}, \tag{H.86}$$

with μ selected to satisfy the power constraint as

$$\mu = \frac{\sqrt{N_1}}{\sqrt{K\,\mathbb{E}\left[\|\mathbf{w}_{1,1}\|^2\right]}}. \tag{H.87}$$

Note that we can rewrite the rate gap as

$$\begin{aligned} \Delta \mathrm{R}_i &= \mathbb{E}\left[\log_2\left(1 + \frac{\frac{P}{K}|\mathbf{h}_i\mathbf{v}_i|^2}{1+\frac{P}{K}\sum_{\ell\neq i}|\mathbf{h}_i\mathbf{v}_\ell|^2}\right)\right] - \mathbb{E}\left[\log_2\left(1 + \frac{\frac{P}{K}|\mu\mathbf{h}_i\mathbf{w}_i|^2}{1+\frac{P}{K}\sum_{\ell\neq i}|\mu\mathbf{h}_i\mathbf{w}_\ell|^2}\right)\right] \\ &\leq \mathbb{E}\left[\log_2\left(1 + \frac{\frac{P}{K}|\mathbf{h}_i\mathbf{v}_i|^2}{1+\frac{P}{K}\sum_{\ell\neq i}|\mathbf{h}_i\mathbf{v}_\ell|^2}\right)\right] - \mathbb{E}\left[\log_2\left(1 + \frac{\frac{P}{K}|\mathbf{h}_i\mathbf{w}_i|^2}{1+\frac{P}{K}\sum_{\ell\neq i}|\mathbf{h}_i\mathbf{w}_\ell|^2}\right)\right] - \log_2(\mu^2) \\ &= \log_2(1/\mu^2), \end{aligned} \tag{H.88}$$

where the last step follows from the definition of the precoder (since $\hat{\mathbf{h}}_i^{(1)}\mathbf{w}_\ell = \hat{\mathbf{h}}_i^{(1)}\mathbf{v}_\ell$, for all $i, \ell \in \mathbb{N}_K$) and the independence between the channel and the estimation noises. It

remains to bound $\log_2(1/\mu^2)$. Note that

$$\begin{aligned}\log_2(1/\mu^2) &= \log_2(K) + \log_2\left(\mathbb{E}\left[\|\mathbf{w}_{1,1}\|^2\right]\right) \\ &\leq \log_2\left(\frac{K}{N_1}\right) + \log_2\left(\mathbb{E}\left[\|\hat{\mathbf{H}}_{K,1}^{\dagger}\|^2\right]\right) \\ &\quad + \log_2(\mathbb{E}[\|\hat{\mathbf{h}}_{\ell}^{(1)}(\mathbf{v}_\ell - \mathbf{w}_\ell)\|^2]) + \log_2(\mathbb{E}[\|\hat{\mathbf{H}}_{\bar{\ell},\bar{1}}\mathbf{w}_{\ell,\bar{1}}\|^2]).\end{aligned} \tag{H.89}$$

Let $\hat{\mathbf{h}}_{i,n}^{(1)}$ represent the channel coefficient for the n-th antenna, with $n \in \mathbb{N}_{N_T}$. Given that $\mathbf{w}_{\ell,j} = \frac{1}{\sqrt{K}}\mathbf{1}_{N_j \times 1}$, for any $j > 2$,

$$\log_2\left(\mathbb{E}[\|\hat{\mathbf{H}}_{\bar{\ell},\bar{1}}\mathbf{w}_{\ell,\bar{1}}\|^2]\right) = \log_2\left(\frac{1}{K}\,\mathbb{E}\left[\sum_{\substack{i=1\\ i\neq\ell}}^{K}\left\|\sum_{n=N_1+1}^{N_T}\hat{\mathbf{h}}_{i,n}^{(1)}\right\|^2\right]\right). \tag{H.90}$$

Note that $g_i \triangleq \sum_{n=N_1+1}^{N_T}\hat{\mathbf{h}}_{i,n}^{(1)} \sim \mathcal{N}_{\mathbb{C}}(0, N_{\bar{1}})$. Hence, $\|g_i\|^2 = |g_i|^2 \sim \text{Exp}(\frac{1}{2N_{\bar{1}}})$. This, together with the fact that the channel coefficients are i.i.d., leads to

$$\log_2\left(\mathbb{E}[\|\hat{\mathbf{H}}_{\bar{\ell},\bar{1}}\mathbf{w}_{\ell,\bar{1}}\|^2]\right) \leq \log_2\left(2N_{\bar{1}}\frac{K-1}{K}\right). \tag{H.91}$$

Moreover,

$$\begin{aligned}\mathbb{E}[\|\hat{\mathbf{h}}_{\ell}^{(1)}(\mathbf{v}_\ell - \mathbf{w}_\ell)\|^2] &\leq 2K\,\mathbb{E}[\|\hat{\mathbf{h}}_{\ell}^{(1)}\|^2] \\ &\leq 2K^2.\end{aligned} \tag{H.92}$$

Introducing (H.91) and (H.92) in (H.89) leads to

$$\log_2(1/\mu^2) \leq \log_2\left(\mathbb{E}\left[\|\hat{\mathbf{H}}_{K,1}^{\dagger}\|^2\right]\right) + \log_2(2K^2) + \log_2\big(2(K-1)\big). \tag{H.93}$$

Note that if we remove the condition for the intended received signal $\hat{\mathbf{h}}_i^{(1)}\mathbf{w}_i = \hat{\mathbf{h}}_i^{(1)}\mathbf{v}_i$ and we assume that $\hat{\mathbf{h}}_i^{(1)}\mathbf{w}_i$ and $\hat{\mathbf{h}}_i^{(1)}\mathbf{v}_i$ are statistically equivalent, we would obtain a tighter bound as

$$\log_2(1/\mu^2) \leq \log_2\left(\mathbb{E}\left[\|\hat{\mathbf{H}}_{K,1}^{\dagger}\|^2\right]\right) + \log_2\big(2(K-1)\big). \tag{H.94}$$

This concludes the proof of Theorem 7.2. ■

Résumé[Français]

a Introduction

Les technologies de communication sans fil ont considérablement évolué au cours de la dernière décennie. Il est prévu que le déploiement imminent de la technologie de réseau cellulaire de cinquième génération (5G) [1] entraîne non seulement débits de données plus élevés, mais aussi un variété de cas d'utilisation inédits et diversifiés. En effet, cette diversification des services est l'un des principaux objectifs du développement 5G [2]. Trois cas d'utilisation principaux sont envisagés dans la génération à venir : Communications robustes et ultra fiables à faible latence (URLLC) [3–5], services à large bande améliorés (eMBB) [6], et communication machine à machine massive (mMTC) [7,8]. La conjonction de ces trois côtés est destinée à contribuer à l'épanouissement de fonctionnalités jamais vues auparavant telles que l'Internet tactile [9, 10], les réseaux de drones [11, 12], les réseaux automobiles [13, 14], ou l'Internet des objets (IoT) [15]. Afin de pouvoir fournir ces nouvelles applications, le réseau s'appuiera sur des technologies innovantes [16], telles que les communications à ondes millimétriques [17–19], la mise en cache [20, 21], les communications entre dispositifs [22], ou les communications avec un nombre massive d'antennes (massive MIMO) [23–26]. En outre, l'augmentation attendue de la densité des réseaux - en termes de cellules et d'appareils - fait de la gestion des interférences l'un des problèmes essentiels dans des transmissions sans fil [2].

L'une des conséquences qui découle de la description du réseau antérieurement décrite est l'augmentation de l'hétérogénéité du réseau, soit entre des nœuds communicants ou entre différents réseaux partageant les mêmes ressources. Cette hétérogénéité affecte également les capacités de *backhaul* des différents nœuds. Par ailleurs, les situations dans lesquelles les nœuds communicants se déplacent à grande vitesse sont en plein essor. Les scénarios hétérogènes et avec mobilité élevée conduisent à l'impossibilité d'avoir une gestion centralisée des communications sans fil, d'où la nécessité de comprendre comment les systèmes distribués se comportent et quelles sont leurs limites fondamentales.

a.1 Transmission Coopérative

La densification significative du réseau entraîne une nécessité de coopération pour éviter l'encombrement du canal sans fil. Les réseaux coopératifs multi-utilisateurs et l'ampleur de ses capacités théoriques ont été analysés en détail dans la littérature [27–31]. La coopération dans les réseaux sans fil peut prendre de nombreuses formes différentes. À l'origine, la coopération était réduite à des politiques statiques qui assuraient un certain fonctionnement, comme la réutilisation partielle des fréquences. Avec l'escalade de la complexité du réseau, les méthodes de coopération ont évolué pour répondre aux exigences croissantes [32]. Intuitivement, les gains de coopération sont subordonnés à l'échange d'informations entre les nœuds de coopération. On peut distinguer deux catégories d'informations essentielles : L'information de canal –ou de système– et les données de l'utilisateur.

En ce qui concerne les données d'utilisateur, les mécanismes de coopération dépendent de la disponibilité ou non de ces informations à tous les nœuds. Si les données de l'utilisateur ne sont pas partagées, étant chaque nœud doté d'informations de données différentes, la coopération peut être réalisée par le biais de la formation coordonnée de faisceaux [33, 34] ou de la planification coordonnée [35]. Dans le scénario opposé, avec partage des données de l'utilisateur, nous pouvons appliquer des stratégies avec une coopération plus forte que les précédentes. L'une des principales stratégies de coopération est la transmission coordonnée multipoint (CoMP) [36]. Cette stratégie de transmission, également connue sous le nom de MIMO coopératif, transmission conjointe, ou réseau-MIMO, bénéficie du partage des données de sorte que l'interférence peut être annulée ou même transformée en signal utile.

La transmission conjointe multi-utilisateur dans des réseaux sans fil est connue pour apporter des améliorations multiplicatives dans les débits des réseaux [37], mais seulement dans l'hypothèse d'avoir une information d'état de canal (CSI) parfaite. Ce scénario idéal a été étudié en profondeur [37–41]. Malheureusement, une acquisition parfaite de la CSI n'est pas possible dans la plupart des applications réseau actuelles en raison de leur complexité et de leur consommation de ressources. Par conséquent, la littérature s'est efforcée d'élucider l'incidence de la CSI imparfaite ou quantifiée sur le rendement des émetteurs.

a.2 Précodage sous Coopération et Information non-Idéales

Information de l'État du Canal Imparfaite

En raison de l'impossibilité d'appliquer l'hypothèse idéale précédente, les cas dans lesquels l'information disponible aux nœuds communicants ne répond pas à l'hypothèse de CSI parfaite ont été étudiés en profondeur. Ainsi, la communauté s'est concentrée sur les

milieux où l'information disponible est opportune mais imparfaite [31, 42–50], ou lorsque l'information est retardée [51–57]. Ce sujet de recherche est resté actif au cours de la dernière décennie, et un grand nombre d'ouvrages ont développé des schémas généralisés pour le cas de CSIT partialement obsolète [54, 58, 59], basculant [60], or changeant [61]. La tentative de compréhension du comportement des réseaux actuels a conduit à l'étude de milieux élaborés et complexes. Par exemple, le canal d'interférence cognitive, dans lequel seulement certains nœuds ont accès aux informations des autres nœuds [62, 63], ou le réseau de relais [64–68].

Appréciablement, même si les travaux précités supposaient une acquisition ou une estimation imparfaite du CSI, ils considèrent que tous les nœuds coopérants partagent la même information imparfaite. Nous nous référons ci-après au scénario où la transmission est optimisée sur la base d'une *unique* estimation de canal imparfaite/dépassée, laquel est commune à toutes les antennes d'émission, comme réglage CSIT centralisé (C-CSIT). Néanmoins, les caractéristiques actuelles et futures des réseaux sans fil rendent cette hypothèse de partage parfait impraticable dans de nombreuses applications. Ceci est dû, par exemple, à la prolifération de réseaux hétérogènes pour lesquels certains nœuds ont une liaison sans fil, fluctuante ou limitée [69–71], ou applications URLLC dans lesquelles le partage parfait des informations entraînerait un retard inopérable [3, 4, 72]. Les environnements dans lesquels des dispositifs simples à faibles capacités visent à communiquer dans un environnement dense –comme dans les applications IoT– tombent également dans les cas d'utilisation dans lesquels le partage de l'information est à la fois souhaitable et difficile. Cette évolution des différents cas d'utilisation augmente l'intérêt sur les systèmes avec d'information distribuée, dans lesquels l'information disponible aux nœuds communicants est non seulement imparfaite mais différente d'un nœud à l'autre. Ce type de scénarios peut être inclus dans les problèmes dits de Décision d'Équipe [73, 74], dans lesquels différents agents visant le même but tentent de coopérer en l'absence de communication parfaite entre eux.

Récemment, l'importance croissante de la coopération des émetteurs non co-localisés –comme, par exemple, dans les réseaux assistés par des dispositifs aériens sans pilote (UAV) [11]– a conduit à un nombre croissant de travaux contestant cette hypothèse de centralisation de la CSIT. Dans [75, 76], des méthodes ont été mises au point pour réduire la CSIT nécessaire à l'alignement des interférences MIMO (IA), et le régime à haut rapport signal-à-bruit (SNR) avec CSIT retardée et locale dans le canal interférence (IC) est également étudié dans plusieurs travaux [77–79]. L'hypothèse d'une CSIT centralisé a également été contestée dans l'analyse de la capacité du canal d'accès multiple [80] et du canal relais [81], entre autres.

Précodage avec Information du Canal Décentralisée

L'hypothèse du modèle C-CSIT peut modéliser un émetteur multi-antenne ou une transmission conjointe de différents émetteurs non co-localisés si nous supposons un réseau d'accès radio dans le nuage (C-RAN) idéal [16]. Dans ces cas, on peut supposer que la CSI imparfaite est parfaitement partagée entre les antennes d'émission non co-localisées. Néanmoins, les réseaux hétérogènes à venir comprennent une grande variété d'appareils, tels que les terminaux d'utilisateurs, les relais drones, les pico-stations, etc., qui cherchent à coopérer malgré l'absence d'une liaison de retour idéale entre eux. D'autres scénarios mettant en scène des liens backhaul existants peuvent favoriser un traitement local plutôt que centralisé afin de répondre aux fortes contraintes de latence dérivées de la 5G et des applications pour l'Internet Tactile [9].

Cet aspect permet d'analyser ce qui se passe lorsque la CSI n'est pas parfaitement partagé entre les dispositifs, c'est-à-dire lorsque chaque nœud peut avoir un CSI différent. Ce nouveau configuration s'appelle configuration avec *CSIT Distribuée* (D-CSIT) [82]. Dans cette configuration réseau, chaque nœud est doté d'une information imparfaite sur l'état du système. Cette information peut être différente d'un nœud à l'autre, et en outre la précision d'un nœud peut varier d'un paramètre à l'autre. L'hétérogénéité des réseaux sans fil actuels est donc correctement contenue dans ce modèle. Dans cette thèse, nous nous concentrons sur le réseau MIMO distribué pour étudier l'impact de ces divergences entre les nœuds coopérants. Bien qu'il ait été suggéré dans la littérature passée que les réseaux avec CSIT distribuée pourraient souffrir une réduction sévère de la performance par rapport aux scénarios classiques de CSIT centralisé [82], un problème crucial est comprendre comment les émetteurs peuvent coopérer pour combattre le manque de cohérence entre les différents informations afin de réduire l'écart par rapport aux performance centralisée du système.

Plusieurs travaux se sont concentrés sur cette configuration avec CSIT distribuée [83], par exemple, l'analyse des performances d'alignement d'interférence [76] ou l'étude de la performance de Zéro-Forçage conventionnel dans la limite de grand système [84]. Toutefois, bon nombre des questions et des défis posés par ce contexte demeurent des problèmes non résolus. Par conséquent, il y a un intérêt évident à examiner le scénario dans lequel chaque émetteur peut avoir une information différente sur le canal [85]. Il existe un grand nombre de systèmes distribués différents [83, 86–91]. Néanmoins, cette thèse s'adresse à ce que l'on appelle le Réseau Distribué MIMO, dans lequel les émetteurs ont accès à toute l'information à fournir aux utilisateurs, mais ils ne partagent pas la même CSIT [82]. Ce modèle survient dans les cas où les données peuvent être mises en mémoire tampon ou en antémémoire [92, 93, 93, 94], mais la CSI doit être disponible avec un très faible retard, comme dans des scénarios à haute mobilité, IoT, réseaux V2X, ou communications sur un

canal rapidement variant mais avec faible débit de transmission [15, 95, 96]. En général, le modèle comprend tout cas d'utilisation dans lequel les contraintes de latence entravent le partage efficace de la CSIT pendant le temps de cohérence des canaux.

a.3 Organisation de la Mémoire de la Thèse

Cette mémoire est divisée dans trois parties. Avant de présenter les contributions de cette thèse, nous introduisons sa motivation et sa portée dans la première partie, qui est composée de deux chapitres. Le premier chapitre s'engage à motiver l'étude du sujet traité, ainsi qu'à fournir un aperçu de l'état de l'art et de nos principales contributions. Le deuxième chapitre comprend une description complète du modèle et des outils considérés. En particulier, il décrit le modèle mathématique pour l'hypothèse de l'information décentralisée, les facteurs de mérite et la notation employée. En outre, le deuxième chapitre présente également quelques scénarios pratiques qui motivent et mettent en œuvre le modèle théorique. Dans chacune des deux autres parties, nous voulons mettre en lumière les limites fondamentales de la communication coopérative et décentralisée avec une perspective différente. La Partie II cherche à caractériser les degrés de liberté (DoF) optimaux des réseaux MIMO avec CSIT décentralisée. La Partie III aborde le problème d'un point de vue différent, car nous analysons la performance des schémas de pré-codage Zero-Forcing dans le cadre d'une coopération distribuée. Le choix du Zéro-Forçage est motivé par sa simplicité et par le fait qu'il est surtout utilisé dans les transmissions pratiques avec multiplexage spatial.

Partie II

La caractérisation DoF réalisée dans la Partie II nécessite généralement une double analyse : L'analyse de faisabilité, dans laquelle nous développons des schémas qui peuvent atteindre une certaine performance, et l'inverse, dans laquelle nous établissons des limites supérieures sur la performance réalisable. Cette partie comprend trois chapitres.

Dans Chapitre 3, nous considérons une système simple avec 2 émetteurs (TX) et 2 récepteurs (RX), d'une antenne chacun, et nous étudions la métrique des degrés de liberté généralisés (GDoF) d'une transmission conjointe dans laquelle les deux TX sont dotés d'une information de canal différente. Récemment, le GDoF du paramètre centralisé dans lequel les deux TX partagent les informations de canal a été obtenu dans [50]. La contribution principale de ce chapitre est la caractérisation GDoF pour son réseau homologue mais décentralisé dans lequel chaque TX peut avoir une estimation de canal différente. Nous montrons que la performance GDoF centralisée est atteinte pour n'importe quelle topologie de pertes de canal et quel que soit le TX qui a la meilleure estimation pour chaque coefficient de canal. Ce résultat intéressant est obtenu grâce à

un nouveau schéma de précodage qui s'adapte à la configuration décentralisée, et qui se base sur l'idée qu'un TX n'utilise ses informations de canal instantanées que si ces informations sont les plus précises parmi les TXs.

Le chapitre précédent bénéficie de la structure du réseau considérée, dans la mesure où il n'existe qu'un seul RX brouillé, et donc un seul TX peut gérer l'interférence si nous concevons un schéma approprié. De ce fait, l'extension de ces résultats à un réseau étendu avec plus de nœuds n'est pas simple. Dans le Chapitre 4, nous étudions le réseau MISO $K \times K$ avec CSIT distribuée afin de déterminer dans quelle mesure les résultats précédents sont généralisables. Nos principales contributions sont doubles : Premièrement, nous dérivons une limite supérieure centralisée pour le cadre avec CSIT distribuée. Cette limite supérieure est basée sur un cadre assisté par un génie dans lequel les TXs sont capables de partager leur CSIT local avec les autres TXs. Par conséquent, le système assisté par le génie est un scénario centralisé dans lequel chaque TX obtient K estimations différentes. Nous montrons que ce réglage atteint le même DoF qu'un réglage centralisé dans lequel les TX ne sont dotés que de l'estimation la plus précise parmi les K estimations disponibles. Deuxièmement, nous développons un schéma réalisable qui augmente considérablement le DoF atteint par rapport aux approches connues dans la littérature. Ce schéma montre que, pour un certain régime de précision CSI, il est encore possible d'atteindre le DoF du réglage assisté par génie pour toute taille de réseau. La clé pour atteindre ce résultat est d'exploiter l'idée que le brouillage inévitable peut être utilisé comme information collatérale sur le récepteur. Le schéma proposé illustre à quel point il est important de choisir de manière appropriée *qui* transmet et de *à qui* il transmet dans les systèmes coopératifs décentralisés, car il s'avère que le DoF maximum est parfois obtenu seulement si une partie des TXs ne transmet aucun signal.

Les deux chapitres précédents sont plus axés sur l'analyse de faisabilité, car la limite supérieure est obtenue à partir d'un cadre centralisé assisté par un génie. Dans Chapitre 5 nous nous intéressons aux bornes supérieures distribuées.

Partie III

L'analyse de la Partie III est motivée par les résultats de la Partie II, car une des principales questions soulevées dans les chapitres précédents est de savoir si ces résultats s'étendent à des métriques plus fines que DoF et GDoF. Pour répondre à cette question, nous limitons notre analyse à de simples schémas de transmission dy type Zéro-Forçage. L'objectif est de révéler la perte de performance à cause du fait de ne partager parfaitement la CSI. Nous étudions donc l'écart de débit du scénario décentralisé par rapport à le scénario centralisé dans lequel l'information sur les canaux des émetteurs est parfaitement partagée. Nous considérons dans cette partie que le précodeur satisfait à une contrainte de puissance instantanée. Ceci est important en raison de la structure décentralisée du réseau

considéré, car un émetteur ne peut pas connaître la normalisation de puissance appliquée à l'autre émetteur parce que chacun le calcule sur la base de ses propres informations de canal, qui peuvent être différentes. La Partie III comprend deux chapitres différents.

Dans le Chapitre 6, nous considérons le réglage avec 2 TX et 2 RX utilisant une seule antenne, comme dans le chapitre initial de la Partie II. Cependant, nous analysons maintenant l'écart de débit du réglage distribué lorsque la transmission conjointe utilise des schémas de Zéro-Forçage pour atténuer le brouillage. La contribution de ce chapitre est multiple. Tout d'abord, nous montrons que le taux obtenu avec la transmission avec Zéro-Forçage dans le cadre de CSIT distribuée converge vers le taux atteint dans le cadre du système centralisé assisté par un génie où la meilleure estimation du canal est partagée entre les émetteurs. Ce résultat implique qu'il n'y a pas d'écart de débit sur le régime asymptotique. Deuxièmement, nous développons un schéma de précodage de type Zéro-Forçage adapté à l'environnement distribué. Ce schéma de précodage s'appuie sur les principales conclusions du Chapitre 3 pour l'analyse DoF. Troisièmement, nous proposons de nouvelles stratégies de précodage qui permettent d'augmenter considérablement la performance à un rapport SNR faible ou moyen. Parmi ces stratégies, il convient de noter qu'une réduction de la précision de l'information de canal à un TX peut améliorer les performances. Ce comportement résulte d'un compromis implicite entre l'exactitude de la décision prise localement par un certain émetteur et la cohérence –ou l'accord– entre les décisions des deux émetteurs.

Dans le Chapitre 7, nous étendons l'analyse du chapitre précédent pour le réseau MISO généralisé avec M émetteurs à antennes multiples et K utilisateurs. Dans le même ordre d'idées qu'au Chapitre 4, l'objectif de ce chapitre est de comprendre les principaux aperçus des résultats du cadre simple en confrontant l'analyse à un cas plus général. La contribution de ce chapitre est de montrer que le débit de la configuration centralisée assistée par le génie est atteint asymptotiquement pour toutes les configurations d'information de canal décentralisée dont le DoF centralisé est atteint. En d'autres termes, le réseau décentralisé permet non seulement d'obtenir le même gain de multiplexage que le réseau centralisé, mais aussi le gain de formation de faisceau. Nous étudions ensuite comment ce résultat s'étend au régime SNR non asymptotique. Nous développons un schéma de transmission pour atteindre ces résultats. Ce schéma utilise également l'idée que la réduction de la précision de l'information à certains nœuds améliore l'accord entre tous les émetteurs. Cependant, il diffère du schéma du cas 2×2 par le fait que les émetteurs disposant d'informations plus précises tentent maintenant de corriger le brouillage généré par les autres émetteurs.

Pour conclure, nous discutons dans le Chapitre 8 des principales conclusions qui se dégagent des travaux développés au cours de cette thèse.

b Énoncé du problème et modèle de système

Nous considérons un réseau sans fil coopératif dans lequel plusieurs émetteurs visent à desservir conjointement plusieurs utilisateurs. Pour appliquer cette transmission coopérative multipoint (CoMP), tous les nœuds émetteurs partagent les symboles d'information destinés à être décodés sur les récepteurs. Par conséquent, la transmission conjointe a pour but d'annuler ou d'éviter que le brouillage causé par d'autres utilisateurs n'affecte la performance. La principale particularité de notre modèle est l'hypothèse d'une information décentralisée sur l'état des canaux. Par conséquent, chaque émetteur possède une estimation de canal particulière, éventuellement différente de celle qui est disponible pour les autres émetteurs.

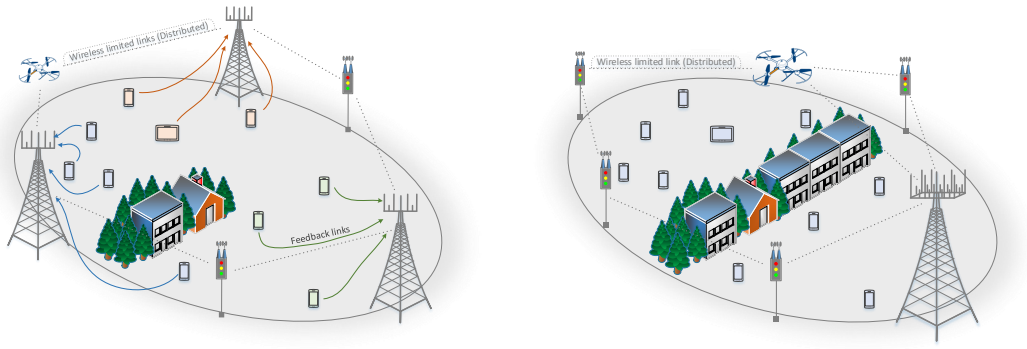
b.1 Partage Idéal des Données et Acquisition de CSI Imparfaite

Nous nous concentrons dans cette thèse sur une configuration sans fil dans laquelle les symboles de données de l'utilisateur sont disponibles et transmis conjointement à partir de tous les TX, alors que les estimations du canal n'ont pu être obtenues que de manière imparfaite aux TX. Telles hypothèses, bien qu'apparemment contradictoires à première vue, sont en fait très pertinentes dans les réseaux sans fil actuels, et encore plus dans les futurs réseaux 5G et au-delà. La raison principale en est que, dans de nombreux scénarios intéressants, la contrainte de latence pour la transmission des données est nettement plus lâche que la contrainte d'obsolescence de la CSI, car cette dernière est liée au temps de cohérence et est donc très courte dans de nombreux scénarios de mobilité pertinents. Cette propriété a pour conséquence que le partage de données (ou la mise en cache) entre les TX peut être réalisé dans la pratique tandis que l'acquisition et le partage des CSI deviennent le principal goulet d'étranglement.

Nous supposons qu'une coopération limitée entre les TX a eu lieu avant la phase de transmission, conduisant à une certaine configuration de précision CSI. Par conséquent, nous supposons que lap précision moyenne de la CSIT reste constante pendant un certain temps. Le problème de l'étude de la meilleure stratégie de partage de la CSI dans une communication limitée et restreinte est un problème de recherche très intéressant en soi, et nous ne discutons donc pas du mécanisme d'acquisition exact des CSI.

b.2 Coopération entre les émetteurs sur la liaison descendante

Nous considérons le réseau MISO dans lequel M TXs servent conjointement à K récepteurs (RXs) à antenne unique, et où TX j a N_j antennes. Nous désignons le nombre total d'antennes d'émission par $N_T \triangleq \sum_{j=1}^{M} N_j$. Les TXs cherchent à délivrer des données s_i à chaque RX i. Les données s_i sont indépendants et distribués de manière identique i.i.d.),



(a) Plusieurs stations de base. Chacun connaît mieux le CSI d'un sous-ensemble de RXs. Ils obtiennent une estimation moins précise des autres RX via les liaisons sans fil TX-TX.

(b) Station de base principale avec têtes-radio télécommandées. Il obtient une estimation de la matrice du canal, puis il transmet les CSI bruyants ou compressés aux TX auxiliaires.

FIGURE I.1 – Exemples de réseau avec CSIT Distribuée.

chacun étant tiré d'une distribution gaussienne complexe à symétrie circulaire $\mathcal{N}_{\mathbb{C}}(0,1)$. Nous introduisons le vecteur $\mathbf{s}$ as $\mathbf{s} \triangleq [s_1, \ldots, s_K]^{\mathrm{T}}$, qui est connu par tous les TXs.

Le canal allant des M TXs aux K RXs est représenté par la matrice de canaux $\mathbf{H} \in \mathbb{C}^{K \times N_T}$. Le vecteur des coefficients de canal de TX k à RX i est représenté par $\mathbf{h}_{i,k}^{\mathrm{H}} \in \mathbb{C}^{1 \times N_k}$. Le vecteur des coefficients de canal depuis tous les TXs à RX i est représenté par $\mathbf{h}_i^{\mathrm{H}} \in \mathbb{C}^{1 \times N_T}$. Les coefficients de canal sont supposés être tirés d'une distribution avec une densité de sorte que toutes les sous-matrices de canal soient de rang complet presque sûrement. Nous supposons que tous les TXs sont dotés du vector de données $\mathbf{s}$. Ils précodent le vecteur $\mathbf{s}$ avec un précodeur $\mathbf{T}$. Le signal reçu aux RXs est alors donné par .

$$\mathbf{y} \triangleq \sqrt{P}\mathbf{H}\mathbf{T}\mathbf{s} + \mathbf{n}, \tag{b.1}$$

où P est la puissance de transmission, $\mathbf{y} \triangleq [\mathrm{y}_1, \ldots, \mathrm{y}_K]^{\mathrm{T}}$ est le vecteur de signal reçu et y_i est le signal reçu au RX i. Le vecteur $\mathbf{n} \in \mathbb{C}^K$ représente le bruit additive Gaussien distribué comme $\mathcal{N}_{\mathbb{C}}(0,1)$.

La matrice de précodage $\mathbf{T}$ peut être décomposée en plusieurs sous-matrices d'intérêt. Ainsi, $\mathbf{T}_k \in \mathbb{C}^{N_k \times K}$ désigne la matrice de précodage appliquée à TX k. Le vecteur de précodage global appliqué aux données de RX i est désigné par $\mathbf{t}_i \in \mathbb{C}^{N_T \times 1}$. Le vecteur de précodage appliqué à TX k pour les données de RX i est représenté par $\mathbf{t}_{i,k}$. Tout au long de ce manuscrit, nous supposons différentes contraintes de puissance pour le signal d'émission. En particulier, nous considérons une contrainte de puissance moyenne, telle qu'il existe une constante $c \in \mathbb{R}^+$ satisfaisant $\mathbb{E}[\|\mathbf{T}_k\|] \leq c$, et une contrainte de norme unitaire instantanée par TX, telle que $\|\mathbf{T}_k\| \leq c$.

b.3 Facteurs de Mérite

La principale mesure de performance considérée est la valeur attendue du débit par utilisateur. En particulier, dans l'hypothèse où les symboles de données sont distribués de manière independant et identique comme $\mathcal{N}_{\mathbb{C}}(0,1)$, le débit attendu pour RX i est donné par

$$R_i(P) \triangleq \mathbb{E}\left[\log_2\left(1+\frac{\frac{P}{K}|\mathbf{h}_i^{\mathrm{H}}\mathbf{t}_i|^2}{1+\sum_{\ell\neq i}\frac{P}{K}|\mathbf{h}_i^{\mathrm{H}}\mathbf{t}_\ell|^2}\right)\right]. \tag{b.2}$$

Il a été démontré que trouver les limites fondamentales du taux dans les systèmes multi-utilisateurs complexes est un problème difficile à résoudre. Pour cette raison, plusieurs mesures asymptotiques ont été largement utilisées dans la littérature. Nous présentons ci-après les paramètres asymptotiques que nous considérons.

Le premièr est la métrique des "Degrés-de-Liberté" (DoF), aussi connue comme gain de multiplexage. Le DoF est defini comme

$$\mathrm{DoF} \triangleq \lim_{P\to\infty}\frac{R(P)}{\log_2(P)}. \tag{b.3}$$

Intuitivement, le DoF est l'approximation du premier ordre du débit, et il représente la pente du débit en fonction du logarithme du SNR P lorsque P approche l'infini. Fig. I.2 illustre sa signification. Malgré le fait que le DoF présente plusieurs limites en tant que facteur de mérite, il s'est avéré utile dans la caractérisation de problèmes complexes, tels que CSIT retardée [31, 112], CSIT distribuée [113, 114], CSIT mixte [59, 115], Alignement des interférences [40, 116], *caching* [94, 117–119], etc. L'une des faiblesses du DoF est qu'il ne tient pas compte de la topologie du réseau –nous appelons "topologie du réseau" la caractérisation de la perte de chemin du système–.

Le concept des Degrés de Liberté Généralisés (GDoF) a été introduit dans le but de surmonter cette limitation et de prendre en compte la topologie de la perte de chemin. GDoF est une extension du modèle DoF où l'affaiblissement sur le trajet est modélisé en fonction du SNR. En effet, GDoF a la même définition du DoF, mais avec un modèle de canal différent : Considérons un coefficient de canal arbitraire $\mathrm{h}_{i,k}$. Dans le modèle de canal GDoF, il est défini comme $\mathrm{h}_{i,k} \triangleq \sqrt{P^{\gamma_{i,k}-1}}\mathrm{g}_{i,k}$, où $\mathrm{g}_{i,k}$ est tiré d'une distribution qui ne dépend pas du paramètre P, c'est-à-dire, comme $\mathrm{h}_{i,k}$ dans le modèle de canal précédent. Le paramètre $\gamma_{i,k} \in [0,1]$ représente la force relative du canal du lien. GDoF s'est avéré une approche intéressante puisque les schémas optimaux réalisables pour l'analyse GDoF permettent également d'obtenir des résultats de capacité avec une erreur borné par un nombre constant de bits [41, 123, 124].

Le GDoF subit encore l'autre limitation principale de DoF : Que la métrique ne fournit aucune connaissance bornée sur le débit atteignable. Dans la Figure I.2, nous

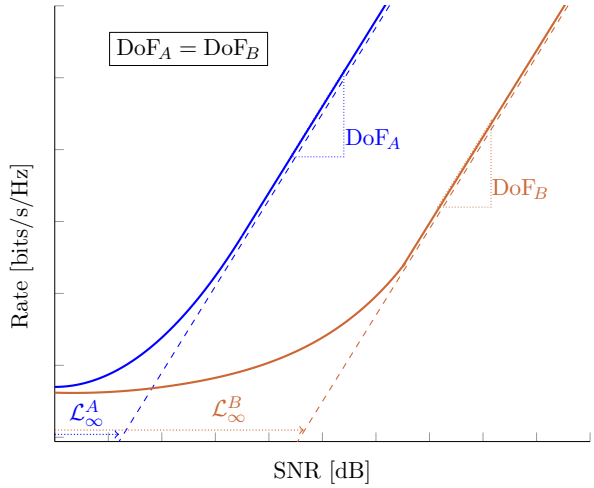


FIGURE I.2 – Illustration qualitative de l'approximation affine de deux systèmes différents avec le même DoF (pente) mais un décalage de débit différent $\mathcal{R}_\infty = \text{DoF}\,\mathcal{L}_\infty$.

montrons comment deux réglages avec le même DoF –qui représente la pente du débit– peuvent atteindre des débits considérablement différents. Néanmoins, la métrique peut être affinée pour offrir les résultats d'un débit réalisable avec un écart limité. En particulier, nous considérons l'approximation affine du débit à un rapport SNR élevé, introduite in [130]. Selon cette approximation, le taux réalisable peut être exprimé comme suit [29]

$$R = \text{DoF}\log_2(P) - \mathcal{R}_\infty + o(1), \tag{b.4}$$

où $\mathcal{R}_\infty$ désigne le *décalage de débit*. On définit aussi le *décalage de puissance* $\mathcal{L}_\infty$, où $\mathcal{R}_\infty = \text{DoF}\,\mathcal{L}_\infty$. Une visualisation illustrative est présentée dans la Fig. I.2.

b.4 Modèle de CSIT Distribuée

La principale particularité de cette thèse est la considération que les TX ne partagent pas parfaitement leur CSI. Dans le scénario avec CSIT centralisée (C-CSIT), il y a une seule estimation de la matrice de canal $\mathbf{H} \in \mathbb{C}^{K\times N_T}$, partagée par tous les TXs. Ensuite, l'hypothèse de information imparfaite est modélisée de telle sorte que

$$\hat{\mathrm{h}}_{i,k} \triangleq \sqrt{1 - Z_{i,k}}\ \mathrm{h}_{i,k} + \sqrt{Z_{i,k}}\ \delta_{i,k}, \tag{b.5}$$

ou $\delta_{\ i,k}$ est la variable de bruit additif et Z indique l'échelle de variance de ce bruit. Soient $\mathbf{1}_{n\times m}$, $\mathbf{0}_{n\times m}$, la matrice à uns et la matrice à zéros de taille $n \times m$. Basé sur (b.5), l'estimation de la matrice de canal peut être écrite comme $\hat{\mathbf{H}} \triangleq \sqrt{\mathbf{1}_{K\times N_T} - \mathbf{Z}} \odot \mathbf{H} + \sqrt{\mathbf{Z}} \odot \mathbf{\Delta}$, ou $\odot$ représente le produit par élément. La variance du bruit est donnée par la

matrice deterministique $\mathbf{Z}$.

Le modèle de CSIT distribuée (D-CSIT) se caractérise par le fait que chaque TX est doté d'une estimation éventuellement différente. Ainsi, la singularité clé de ce réglage est que, pour n'importe quel coefficient de canal, il existe autant d'estimations que de TXs, chacune d'entre elles *localement disponible à un seul TX*. Notons que l'estimation à TX j est égale à $\hat{\mathbf{H}}^{(j)}$, étant donnée par $\hat{\mathbf{H}}^{(j)} \triangleq \sqrt{\mathbf{1}_{K\times N_T} - \mathbf{Z}^{(j)}} \odot \mathbf{H} + \sqrt{\mathbf{Z}^{(j)}} \odot \mathbf{\Delta}^{(j)}$. Ainsi, chaque TX a un bruit d'estimation différent ($\mathbf{\Delta}^{(j)}$) avec une puissance différente ($\mathbf{Z}^{(j)}$).

On sait que, dans le cas avec CSIT centralisée, le rapport signal/bruit de l'estimation doit être proportionel à P^{α}, avec $\alpha > 0$, afin d'éviter l'effondrement du gain du multiplexage [50, 132]. Par conséquent, nous supposons que la matrice $\mathbf{Z}^{(j)}$ est définie de telle sorte que son coefficient (i,k) est donné par $\mathbf{Z}^{(j)}_{i,k} = P^{-\alpha^{(j)}_{i,k}}$, avec $0 \leq \alpha^{(j)}_{i,k} \leq 1$. Le coefficient $\alpha^{(j)}_{i,k}$ est le paramètre de *mis en schélle de la précision* qui mesure la qualité de l'estimation du canal à TX j. Nous définissons l'ensemble des paramètres comme $\boldsymbol{\alpha}$, de sorte que $\boldsymbol{\alpha} \triangleq \{\alpha^{(j)}_{i,k}\}_{i\in\mathbb{N}_K,\ j,k\in\mathbb{N}_M}$. Les paramètres $\alpha^{(j)}_{i,k}$ sont supposés des coefficients à long terme qui varient lentement. Sur cette base, il est supposé que chaque TX connaît le jeu complet $\boldsymbol{\alpha}$. Nous appellerons le TX avec le paramètre $\alpha^{(j)}_{i,k}$ le plus grand comme "le plus informé" ou "le mieux informé".

b.5 Scénario Centralisé Assisté par un Génie

De plus, nous introduisons la notion de "scénario centralisé assisté par un génie" qui sera utilisée tout au long de cette thèse. Cette scénario idéal est motivé du fait qu'il est une borne supérieure. Un scénario centralisé assisté par un génie est un réglage C-CSIT –dans lequel toutes les TX sont dotées du même CSI– qui est obtenu à partir d'un réglage D-CSIT en fournissant des informations supplémentaires aux TXs, soit tout la CSI disponible dans d'autres TX, soit la CSI la plus précise. De cette façon, nous sommes en mesure d'analyser quel est l'impact d'avoir l'information décentralisée ou, en d'autres termes, quel est le coût de ne pas partager la CSI.

c Résultats principaux de l'Analyse DoF et GDoF

Notre premier résultat est l'obtention d'une borne supérieure intuitive pour le cas distribué. Cette borne est dérivée du scénario centralisé assisté par un génie où chaque TX a accès aux estimations de tous les TXs.

Theorem I.1. *Dans le réseau MISO distribué avec K TXs et K RXs, tous ayant une seule antenne, le GDoF optimal est borné en haut par le GDoF d'un scénario*

centralisé dans lequel tous les TXs ont accès à l'ensemble des estimations $\{\hat{\mathbf{H}}^{(j)}\}_{j\in\mathbb{N}_K}$. *Définissons l'ensemble des paramètres de mise à l'échelle comme*

$$\boldsymbol{\alpha} \triangleq \{\alpha_{i,k}^{(j)}\}_{i,j,k\in\mathbb{N}_K} \qquad et \qquad \boldsymbol{\alpha}^\star \triangleq \{\max_{j\in\mathbb{N}_2} \alpha_{i,k}^{(j)}\}_{i,k\in\mathbb{N}_K}. \tag{c.6}$$

Alors, $\mathrm{GDoF}^{\mathrm{DCSIT}}(\boldsymbol{\alpha}) \leq \mathrm{GDoF}^{\mathrm{CCSIT}}(\boldsymbol{\alpha}^\star)$.

Une fois le GDoF est délimité dans sa limité supérieure, nous développons des schémas de transmission pour essayer d'atteindre ce GDoF optimal. Le premier résultat s'applique dans le réseau 2×2.

Lemma I.1. *Dans le réseau MISO* 2×2, *le réglage distribué atteint le GDoF du scénario idéal centralisé avec partage parfait de la CSI, de tel sorte que tous les deux TXs ont accès à l'ensemble* $\{\hat{\mathbf{H}}^{(1)}, \hat{\mathbf{H}}^{(2)}\}$. *En définissant l'ensemble des paramètres de mise à l'échelle comme*

$$\boldsymbol{\alpha} \triangleq \{\alpha_{i,k}^{(j)}\}_{i,j,k\in\mathbb{N}_2} \qquad et \qquad \boldsymbol{\alpha}^\star \triangleq \{\max_{j\in\mathbb{N}_2} \alpha_{i,k}^{(j)}\}_{i,k\in\mathbb{N}_2}, \tag{c.7}$$

on obtient que $\mathrm{GDoF}^{\mathrm{DCSIT}}(\boldsymbol{\alpha}) = \mathrm{GDoF}^{\mathrm{CCSIT}}(\boldsymbol{\alpha}^\star)$.

Nous présentons ci-dessous un exemple simple pour illustrer l'intuition principale qui se cache derrière les résultats précédents. Nous considérons la métrique DoF, c'est-à-dire que la perte de chemin de canal n'est pas mise à l'échelle comme le SNR P, et on considère une répartition de CSI telle que, pour $\rho \in [0,1]$,

$$\begin{aligned}
\text{TX } 1 \rightarrow &\left\{\alpha_{1,1}^{(1)} = 0.25,\ \alpha_{1,2}^{(1)} = 0.25,\ \alpha_{2,1}^{(1)} = 0.5,\ \alpha_{2,2}^{(1)} = 0.5\right\},\\
\text{TX } 2 \rightarrow &\left\{\alpha_{1,1}^{(2)} = \rho,\ \alpha_{1,2}^{(2)} = \rho,\ \alpha_{2,1}^{(2)} = 1-\rho,\ \alpha_{2,2}^{(2)} = 1-\rho\right\}.
\end{aligned}$$

Dans la Figure I.3 nous montrons le DoF obtenu par le schéma proposé (S-ZF) en fonction de ρ. Nous comparons ce DoF avec celui d'un scénario centralisé avec la qualité de CSIT $\alpha_{i,k} = \max(\alpha_{i,k}^{(1)}, \alpha_{i,k}^{(2)})$, $\forall i,k \in \mathbb{N}_2$, dont DoF est calculé dans [50], ainsi que avec Zéro-Forçage (ZF) conventionel et Multiplexage par Répartition dans le Temps (TDM). Comme indiqué auparavant, le schéma proposé atteint le DoF du cas centralisé assisté par génie, alors que le schéma basé sur ZF conventionnel, qui est optimal dans le scénario CSIT centralisé, fonctionne mal face aux différences CSI entre les TXs.

Considérons maintenant un réseau avec K TXs et K RXs. Nous considérons dans ce cas une structure rangée, dans lequel nous pouvons ordonner les TXs de telle sorte que $1 \geq \alpha_{i,k}^{(1)} \geq \alpha_{i,k}^{(2)} \geq \cdots \geq \alpha_{i,k}^{(M)} \geq 0$. Cette hypothèse est formulée de manière à éviter

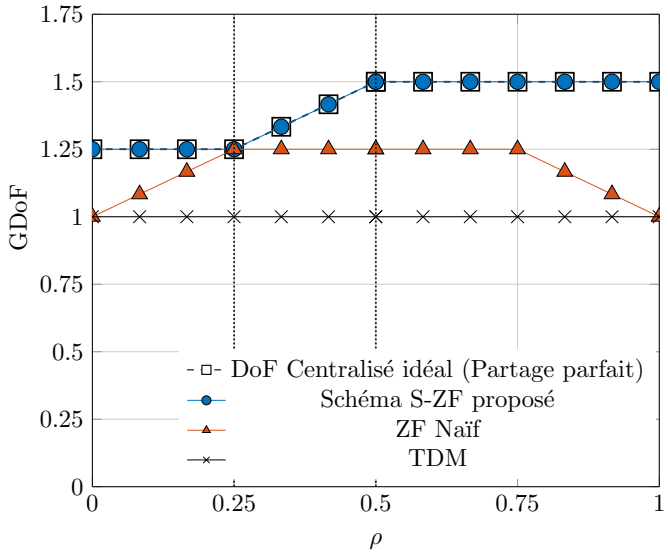


FIGURE I.3 – DoF pour l'exemple illustratif en fonction du paramètre ρ.

l'augmentation incontrôlable des régimes possibles de la CSIT. Puisque la structure triée s'applique à n'importe quel coefficient de canal dans le même ordre, on peut supposer que $\alpha_{i,k}^{(j)} = \alpha^{(j)}$. Par conséquent, $1 \geq \alpha^{(1)} \geq \cdots \geq \alpha^{(M)} \geq 0$, et l'estimation à TX j peut être écrite comme $\hat{\mathbf{H}}^{(j)} = \mathbf{H} + \bar{P}^{-\alpha^{(j)}} \mathbf{\Delta}^{(j)}$. La configuration CSIT distribuée multi-utilisateurs peut donc être représentée par le vecteur de mise à l'échelle multi-TX $\boldsymbol{\alpha} \in \mathbb{R}^K$ défini comme $\boldsymbol{\alpha} \triangleq \{\alpha^{(1)}, \ldots, \alpha^{(K)}\}$. Nous montrons ci-après les résultats obtenus pour cette configuration. Nous divisons la limite inférieure en deux régimes CSIT différents, selon qu'elle correspond à la limite supérieure centralisée ou non.

Theorem I.2. *Supposons que les premières m TXs ont la même précision de CSIT, i.e., $\alpha^{(1)} = \cdots = \alpha^{(m)}$, $m < K$. Définissons α_m^{Weak} comme*

$$\alpha_m^{\text{Weak}} \triangleq \frac{1}{1 + K(K - m - 1)}. \tag{c.8}$$

Alors, si $\alpha^{(1)} \leq \alpha_m^{\text{Weak}}$, le DoF du réseau MISO $K \times K$ avec CSIT distribuée satisfait que $\text{DoF}^{\text{DCSIT}}(\boldsymbol{\alpha}) = \text{DoF}^{\text{CCSIT}}(\alpha^{(1)})$.

Étonnamment, pour $m = 1$, le cas le plus hétérogène, le DoF ne dépend que de la qualité CSI à TX 1, mais avec l'inconvénient de réduire la gamme des configurations CSIT possibles. Le dernier théorème montre que les configurations CSIT pour lesquelles la limite supérieure est étroite. Dans ce qui suit, nous présentons un schéma de transmission robuste qui s'appuie sur le résultat antérieur et qui est étendu pour s'adapter sans

restriction à toute configuration CSIT.

Theorem I.3. *Dans le $K \times K$ D-CSIT réseau MISO avec paramètres de précision $\boldsymbol{\alpha}$, le DoF est borné en bas par $\mathrm{DoF}^{\mathrm{APZF}}(\boldsymbol{\alpha})$, obtenu en résolvant ce problème linéaire :*

$$\mathrm{DoF}^{\mathrm{APZF}}(\boldsymbol{\alpha}) = \max_{\gamma_{n,k}} \sum_{k=2}^{K} \sum_{n=1}^{k-1} \gamma_{n,k} \left(1 + (k-1)\alpha^{(n)}\right) \tag{c.9}$$

$$\text{soumis à} \quad \sum_{k=2}^{K} \sum_{n=1}^{k-1} \gamma_{n,k} = 1, \quad \gamma_{n,k} \geq 0, \tag{c.10}$$

$$\sum_{k=2}^{K} \sum_{n=1}^{k-1} d_{n,k} \gamma_{n,k} \geq 0, \tag{c.11}$$

où $\gamma_{n,k}$ est une variable de répartition des temps et $d_{n,k} \triangleq 1 - \alpha^{(n)} - k(k-n-1)\alpha^{(n)}$.

Considérons maintenant le cas avec 2 RXs, avec N_1 et N_2 antennes respectivement, où il y a k antennes de transmission avec CSI parfaite et $M-k$ antennes de transmission sans CSI. Nous avons obtenu la région DoF por ce scénario en fonction de la taille du réseau et la valeur de k.

Theorem I.4. *La région DoF du cas avec 2 utilisateurs, avec $m \geq N_1$ antennes de TX avec CSIT parfaite est donnée par*

$$(d_1, d_2) \in \begin{cases} d_1 \leq \min(M, N_1) \\ d_2 \leq \min(M, N_2) \\ \frac{d_1}{\min(M, N_1+N_2)-m} + \frac{d_2-m}{N_2-m} \leq 1 \end{cases} \tag{c.12}$$

si $N_1 \leq m < \min(N_2, M-N_1)$, et

$$(d_1, d_2) \in \begin{cases} d_1 \leq \min(M, N_1) \\ d_2 \leq \min(M, N_2) \\ d_1 + d_2 \leq M \end{cases} \tag{c.13}$$

si $m \geq \min(N_2, M-N_1)$.

d Résultats principaux de l'Analyse de ZF Distribué

Nous considérons maintenant une transmission simple avec précodage de Zéro-Forçage. Nous sommes intéressés à analyser comment les résultat théoriques précédents peuvent être étendus à d'autres métriques que le DoF. En particulier, nous voulons caractériser le débit atteignable dans le régime de SNR haut. Dénotons le débit atteignable dans le cas distribué comme $R^{\mathrm{HAP}}(\boldsymbol{\alpha})$. Ce débit est obtenu avec une transmission utilisant un schéma de précodage distribué développé dans cette thèse. De manière similaire, le débit moyen atteint par le précodeur ZF centralisé sur la base des meilleures estimations est noté comme $R^{\mathrm{ZF}}(\boldsymbol{\alpha}^{\star})$. Notez que $\boldsymbol{\alpha}^{\star} = [\max_{j\in\mathbb{N}_2} \alpha_1^{(j)}, \quad \max_{j\in\mathbb{N}_2} \alpha_2^{(j)}]$. Donc, l'écart de débit entre ces deux scénarios est défini comme

$$\Delta \mathrm{R} \triangleq R^{\mathrm{ZF}}(\boldsymbol{\alpha}^{\star}) - R^{\mathrm{HAP}}(\boldsymbol{\alpha}). \tag{d.14}$$

D'abord, nous considérons le cas avec 2 TXs et 2 RXs, tous ayant une seule antenne. Nous supposons aussi que l'information de canal est obtenue aux TXs par voie des RXs : Les RXs ont une information parfaite de leur canal. Donc, ils quantifient cette information avec un nombre de bits proportionnel au SNR, en utilisant quantification vectorielle avec des livres-codes aléatoires.

Theorem I.5. *Pour tout $\alpha_{i,k}^{(j)} > 0$, le schéma proposé atteint un débit tel que*

$$\lim_{P\to\infty} \Delta \mathrm{R} \leq 0. \tag{d.15}$$

Donc, le débit dans le cas décentralisé atteint asymptotiquement le débit du cas idéal centralisé où tous les deux TXs partagent la meilleure estimation. Concentrons-nous à titre d'exemple sur le cas où l'information de canal du RX 1 est distribuée de manière locale et celle de RX 2 de manière opposée :

$$\text{À TX 1} \longrightarrow \begin{array}{cc|cc} \alpha_{1,1}^{(1)} = 1 & \alpha_{1,2}^{(1)} = 0.4 & \alpha_{1,1}^{(2)} = 0.4 & \alpha_{1,2}^{(2)} = 1 \\ \alpha_{2,1}^{(1)} = 0.4 & \alpha_{2,2}^{(1)} = 1 & \alpha_{2,1}^{(2)} = 0.4 & \alpha_{2,2}^{(2)} = 1 \end{array} \longleftarrow \text{À TX 2}$$

Notez que cette configuration correspond au cas où chaque TX a l'estimation la plus précise pour son propre coefficient vers RX 1, et chaque TX a l'estimation la plus précise pour le coefficient de l'autre TX vers RX2. Nous constatons dans la Fig. I.4 que le schéma proposé offre une performance nettement supérieure par rapport au ZF naïf et au TDM. De plus, il converge vers le cas idéal où les deux TXs partagent la meilleure estimation.

On présente dans la Fig. I.5 le pourcentage du débit du ZF centralisé idéal qui est atteint par le schéma de transmission proposé pour toutes les configurations d'information de canal possibles. On montre donc côte à côte le pourcentage atteint à $P = 30,\ 50$ et 80dB.

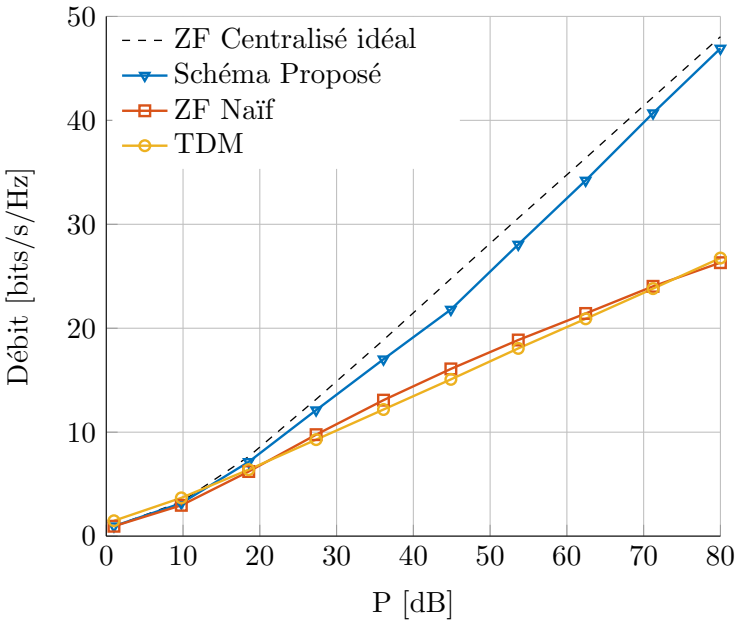


FIGURE I.4 – Débit total pour le cas où l'information de canal du RX 1 est « locale » et celle de RX 2 est « opposée ».

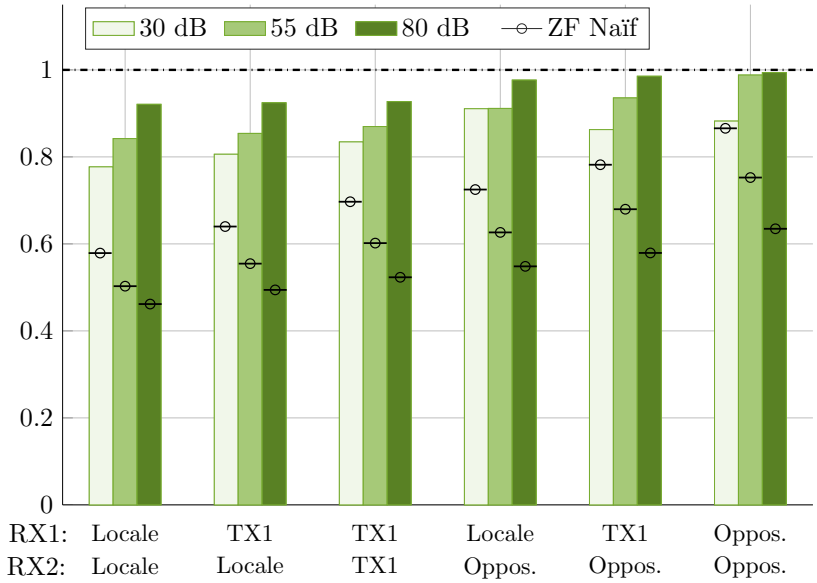


FIGURE I.5 – Pourcentage du débit centralisé « idéal » atteint par le schéma proposé et par zéro-forçage naïf pour P $\in \{30, 55, 80\}$ dB.

Motivés par ce résultat, nous étendons le résultat au cas général avec M TXs –où TX j a N_j antennes– et RXs avec qu'un seule antenne. L'expansion à ce scénario général n'est pas triviale, dans la mesure que dans ce cas il faut aussi récupérer le gain de formation de faisceaux.

Nos principales contributions s'appuient sur un nouveau schéma de précodage de type ZF appelé *ZF Décentralisé et Cohérent* (CD-ZF). En bref, le précodage appliqué à chaque TX est différent si le TX est le mieux informé ou non. Soit $R(\boldsymbol{\alpha}_M)$ le débit atteignable pour notre scénario distribué. De même, supposons que $R^\star(\alpha^{(1)})$ soit le débit obtenu par un schéma ZF sur le scénario idéal assisté par génie comme décrit auparavant. Par conséquent, l'écart de débit entre ces deux configurations est défini comme $\Delta\mathrm{R} \triangleq R^\star(\alpha^{(1)}) - R(\boldsymbol{\alpha}_M)$.Nous pouvons maintenant énoncer notre résultat principal.

Theorem I.6. *Dans le réseau MISO avec CSIT distribué, où TX 1 a $N_1 \geq K-1$ antennes et $\alpha^{(M)} > 0$, le débit moyen atteint par les schémas de type ZF dans le cadre du CSIT centralisé assisté par génie est atteint asymptotiquement. Alors,*

$$\lim_{P\to\infty} R^\star(\alpha^{(1)}) - R(\boldsymbol{\alpha}_M) = 0. \tag{d.16}$$

Remarquablement, ce théorème implique qu'il est possible d'atteindre non seulement le gain de multiplexage mais aussi le gain de formation de faisceau obtenu par le cas centralisé avec N_T antennes, même si seulement N_1 antennes sont dotées de la meilleure précision. La contrainte $N_1 \geq K-1$, c'est-à-dire que le TX avec la CSI la plus précise a un nombre d'antennes au moins égal au nombre de RXs brouillés, provient du fait que si $N_1 < K-1$ l'utilisation de ZF n'est pas suffisant pour atteindre le DoF du cas centralisé, et donc $\lim_{P\to\infty} R^\star(\alpha^{(1)}) - R(\boldsymbol{\alpha}_M) = \infty$.

e Conclusions et Perspectives

Dans cette thèse, nous nous sommes concentrés sur la caractérisation du régime SNR élevé du réseau distribué MIMO, dans lequel un ensemble de TXs servent conjointement un ensemble d'utilisateurs avec la particularité que les informations disponibles à chaque TX peuvent être différentes ou peuvent avoir une précision différente. Nous avons montré que l'environnement distribué atteint la performance de l'environnement centralisé idéal dans lequel les TXs peuvent parfaitement partager leur CSIT. La manière d'obtenir ce résultat asymptotique est intéressante en soi. Il convient de noter que ces performances optimales sont obtenues en réduisant la précision du pré-codeur. En particulier, la principale conclusion tirée du schéma proposé est qu'il existe un compromis implicite entre la

précision locale à chaque TX et la cohérence entre les TXs. Ainsi, la réduction de la précision à certains nœuds –normalement à ceux dont la précision n'est pas la meilleure– permet d'améliorer la cohérence de la décision prise par tous les TXs, puisque les TXs les mieux informés sont capables d'estimer ces paramètres de précision réduite, et donc de corriger le brouillage généré.

Nous avons montré que les gains de coopération sont moins sensibles aux déficiences du CSIT que ce que l'on suppose habituellement. L'idée clé est que nous devons développer des systèmes qui sont conscients et réactifs à l'allocation CSIT, puisque les systèmes communs induisent une diminution significative de la performance. Bien que nous soyons capables de compenser la décentralisation de l'information avec des algorithmes appropriés dans certains cas, atteindre la performance centralisée est impossible pour de nombreux scénarios distribués. Une conclusion intéressante de cette thèse est l'idée que l'ajout d'une structure à la configuration CSIT permet d'augmenter les performances réalisables. Motivé par ce comportement, une analyse intéressante à faire est d'étudier comment cette réduction de performance face à la réduction d'*overhead* implicite dans les environnements distribués. Ainsi, une performance réduite peut être avantageuse par rapport à une performance centralisée idéale si celle-ci implique une quantité d'informations impossible à partager. Les approches mises au point pour atteindre les résultats susmentionnés sont jugées utiles dans de nombreux contextes décentralisés ou pour prendre des décisions d'équipe. En effet, le compromis entre la cohérence et l'exactitude est un compromis inhérent aux problèmes de décision en équipe.

Il est nécessaire de comprendre à quel point les résultats exposés sont dépendants de l'hypothèse d'un partage parfait des données d'information de l'utilisateur. Un autre sujet intéressant est de déplacer la perspective considérée ici –celle du développement de la meilleure transmission pour une allocation CSI donnée– vers l'analyse inverse de l'optimisation de l'allocation CSI soumise à une certaine stratégie de transmission et à un protocole de rétroaction. De plus, la prise en compte du partage de la CSIT dans la mesure de la performance est une étape ultérieure dans l'analyse des réseaux décentralisés.

En conclusion, on s'attend à ce que les futurs réseaux sans fil 5G et au-delà s'adaptent à des scénarios très hétérogènes et à des spécifications inédites, comme l'IoT, des réseaux à très haute vitesse (réseaux V2X), ou à des contraintes de délais extrêmement exigeantes (URLLC). Ces scénarios surviennent à la suite de l'élargissement des cas d'utilisation, à mesure que de nouvelles applications primordiales sont envisagées, par exemple les communications haptiques, les services médicaux à distance ou l'automatisation industrielle. Donc, l'analyse d'autres scénarios coopératifs, ainsi que des scénarios non coopératifs où l'inclusion d'une coopération partielle est possible, est un sujet très intéressant qui peut être analysé dans la recherche future.

f Publications

Les publications suivantes sont le résultat des travaux réalisés au cours du doctorat.

f.1 Conférences

- Antonio Bazco, Paul de Kerret, David Gesbert, and Nicolas Gresset, "Generalized Degrees-of-Freedom of the 2-user MISO Broadcast Channel with Distributed CSIT," dans Proc. IEEE International Symposium on Information Theory (ISIT), Juin 2017, pp. 1092–1096.

- Antonio Bazco, Paul de Kerret, David Gesbert, and Nicolas Gresset, "Méthode de transmission robuste au partage imparfait de l'information de canal entre transmetteurs," dans Proc. Colloque GRETSI, Septembre 2017.

- Antonio Bazco-Nogueras, Lorenzo Miretti, Paul de Kerret, David Gesbert, and Nicolas Gresset, "Achieving Vanishing Rate Loss in Decentralized Network MIMO," dans Proc. IEEE International Symposium on Information Theory (ISIT), Juillet 2019, pp. 1457-1461.

- Antonio Bazco-Nogueras, Lorenzo Miretti, Paul de Kerret, David Gesbert, and Nicolas Gresset, "Transmission Robuste de Zéro-Forçage Asymptotiquement Optimale pour Coopération Imparfaite de Transmetteurs," dans Proc. Colloque GRETSI, Août 2019.

f.2 Journaux

- Antonio Bazco-Nogueras, Paul de Kerret, David Gesbert, and Nicolas Gresset, "Distributed CSIT does not reduce the Generalized DoF of the 2-user MISO Broadcast Channel," dans IEEE Wireless Communications Letters, Juin 2019, pp. 685-688.

- Antonio Bazco-Nogueras, Paul de Kerret, David Gesbert, and Nicolas Gresset, "On the Degrees-of-Freedom of the K-user Distributed Broadcast Channel," soumis à IEEE Transactions on Information Theory, 2018.

- Antonio Bazco-Nogueras, Paul de Kerret, David Gesbert, and Nicolas Gresset, "Asymptotically Achieving Centralized Rate on the $M \times K$ Decentralized Network MISO," soumis à IEEE Transactions on Information Theory, 2019.

Bibliography

[1] G. P. A. W. Group, "View on 5G Architecture," Jun. 2019, version 3.0. [Online]. Available: https://5g-ppp.eu/wp-content/uploads/2019/07/5G-PPP-5G-Architecture-White-Paper_v3.0_PublicConsultation.pdf

[2] A. Gupta and R. K. Jha, "A survey of 5g network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.

[3] P. Popovski, J. J. Nielsen, C. Stefanovic, E. d. Carvalho, E. Strom, K. F. Trillingsgaard, A. Bana, D. M. Kim, R. Kotaba, J. Park, and R. B. Sorensen, "Wireless access for ultra-reliable low-latency communication: Principles and building blocks," *IEEE Network*, vol. 32, no. 2, pp. 16–23, March 2018.

[4] M. Bennis, M. Debbah, and H. V. Poor, "Ultrareliable and low-latency wireless communication: Tail, risk, and scale," *Proceedings of the IEEE*, vol. 106, no. 10, pp. 1834–1853, Oct 2018.

[5] M. Kountouris, P. Popovski, I. Hou, S. Buzzi, A. Müller, S. Sesia, and R. W. Heath, "Guest Editorial Ultra-Reliable Low-Latency Communications in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 701–704, Apr. 2019.

[6] P. Rost, A. Banchs, I. Berberana, M. Breitbach, M. Doll, H. Droste, C. Mannweiler, M. A. Puente, K. Samdanis, and B. Sayadi, "Mobile network architecture evolution toward 5g," *IEEE Communications Magazine*, vol. 54, no. 5, pp. 84–91, May 2016.

[7] C. Bockelmann, N. Pratas, H. Nikopour, K. Au, T. Svensson, C. Stefanovic, P. Popovski, and A. Dekorsy, "Massive machine-type communications in 5g: physical and mac-layer solutions," *IEEE Communications Magazine*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[8] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 120–128, February 2017.

[9] M. Simsek, A. Aijaz, M. Dohler, J. Sachs, and G. Fettweis, "5G-Enabled Tactile Internet," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 460–473, March 2016.

[10] A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos, and M. Frodigh, "Realizing the tactile internet: Haptic communications over next generation 5G cellular networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 82–89, Apr. 2017.

[11] Z. Becvar, M. Vondra, P. Mach, J. Plachy, and D. Gesbert, "Performance of mobile networks with UAVs: Can flying base stations substitute ultra-dense small cells?" in *Proc. 23th European Wireless Conference (EW)*, May 2017, pp. 1–7.

[12] B. Li, Z. Fei, and Y. Zhang, "UAV communications for 5G and Beyond: Recent advances and future trends," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2241–2263, Apr. 2019.

[13] Z. Ning, F. Xia, N. Ullah, X. Kong, and X. Hu, "Vehicular social networks: Enabling smart mobility," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 16–55, May 2017.

[14] Z. Su, Y. Hui, and Q. Yang, "The next generation vehicular networks: A content-centric framework," *IEEE Wireless Communications*, vol. 24, no. 1, pp. 60–66, February 2017.

[15] C. Wu, Z. Liu, D. Zhang, T. Yoshinaga, and Y. Ji, "Spatial intelligence toward trustworthy vehicular iot," *IEEE Commun. Mag.*, vol. 56, no. 10, pp. 22–27, oct 2018.

[16] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, "Networks and Devices for the 5G Era," *IEEE Communications Mag.*, vol. 52, no. 2, pp. 90–96, February 2014.

[17] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!" *IEEE Access*, vol. 1, pp. 335–349, 2013.

[18] J. Qiao, X. S. Shen, J. W. Mark, Q. Shen, Y. He, and L. Lei, "Enabling device-to-device communications in millimeter-wave 5g cellular networks," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 209–215, January 2015.

[19] M. Xiao, S. Mumtaz, Y. Huang, L. Dai, Y. Li, M. Matthaiou, G. K. Karagiannidis, E. Björnson, K. Yang, C. I, and A. Ghosh, "Millimeter wave communications for future mobile networks," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 9, pp. 1909–1935, Sep. 2017.

[20] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug 2014.

[21] M. Chen, Y. Qian, Y. Hao, Y. Li, and J. Song, "Data-driven computing and Caching in 5G networks: Architecture and delay analysis," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 70–75, Feb. 2018.

[22] A. Asadi, Q. Wang, and V. Mancuso, "A survey on device-to-device communication in cellular networks," *IEEE Communications Surveys Tutorials*, vol. 16, no. 4, pp. 1801–1819, Fourthquarter 2014.

[23] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.

[24] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive mimo for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

[25] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive mimo: ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, February 2016.

[26] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive mimo: A survey," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[27] M. Costa and A. E. Gamal, "The capacity region of the discrete memoryless interference channel with strong interference (Corresp.)," *IEEE Trans. Inf. Theory*, vol. 33, no. 5, pp. 710–711, Sep. 1987.

[28] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian Broadcast Channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[29] A. Lozano, A. M. Tulino, and S. Verdú, "High-SNR power offset in multiantenna communication," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4134–4151, Dec 2005.

[30] B. Hassibi and M. Sharif, "Fundamental Limits in MIMO Broadcast Channels," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 7, pp. 1333–1344, Sep. 2007.

[31] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[32] E. Pateromichelakis, M. Shariat, A. u. Quddus, and R. Tafazolli, "On the evolution of multi-cell scheduling in 3gpp lte / lte-a," *IEEE Communications Surveys Tutorials*, vol. 15, no. 2, pp. 701–717, Second 2013.

[33] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. on Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, 2010.

[34] F. Maschietti, D. Gesbert, P. de Kerret, and H. Wymeersch, "Robust location-aided beam alignment in millimeter wave massive mimo," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Singapore, SINGAPOUR, Dec 2017.

[35] Suman Das, Harish Viswanathan, and G. Rittenhouse, "Dynamic load balancing through coordinated scheduling in packet data systems," in *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No.03CH37428)*, vol. 1, March 2003, pp. 786–796.

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

[37] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "The capacity region of the gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, Sep. 2006.

[38] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian Broadcast Channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, Aug. 2003.

[39] N. Jindal and A. Goldsmith, "Dirty-paper coding versus tdma for mimo broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 5, pp. 1783–1794, May 2005.

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

[41] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[42] M. Kountouris, R. de Francisco, D. Gesbert, D. T. M. Slock, and T. Salzer, "Multiuser diversity - multiplexing tradeoff in mimo broadcast channels with limited feedback," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Oct 2006, pp. 364–368.

[43] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE J. Sel. Areas Commun.*, vol. 6, no. 2, pp. 458–462, Feb. 2007.

[44] P. Ding, D. J. Love, and M. D. Zoltowski, "Multiple antenna Broadcast Channels with shape feedback and limited feedback," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3417–3428, Jul. 2007.

[45] A. S. Motahari and A. K. Khandani, "Capacity Bounds for the Gaussian Interference Channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 2, pp. 620–643, Feb 2009.

[46] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[47] M. Min, Y. Jeon, and G. Im, "On Achievable Multiplexing Gain of BD in MIMO Broadcast Channels With Limited Feedback," *IEEE Trans. on Wireless Communications*, vol. 15, no. 2, pp. 871–885, Feb 2016.

[48] E. Piovano and B. Clerckx, "Optimal DoF Region of the K-User MISO BC With Partial CSIT," *IEEE Commun. Lett.*, vol. 21, no. 11, pp. 2368–2371, Nov 2017.

[49] A. G. Davoodi and S. A. Jafar, "CSIT Thresholds for Collapse of Degrees of Freedom in Wireless Networks," in *2018 IEEE International Conference on Communications (ICC)*, May 2018, pp. 1–6.

[50] A. G. Davoodi, B. Yuan, and S. A. Jafar, "GDoF Region of the MISO BC: Bridging the Gap Between Finite Precision and Perfect CSIT," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 7208–7217, Nov 2018.

[51] M. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4418–4431, Jul. 2012.

[52] J. Xu, J. G. Andrews, and S. A. Jafar, "MISO Broadcast Channels with delayed finite-rate feedback: Predict or observe?" *IEEE Trans. on Wireless Commun.*, vol. 11, no. 4, pp. 1456–1467, Apr. 2012.

[53] R. Tandon, S. Mohajer, V. Poor, and S. Shamai (Shitz), "Degrees of freedom region of the MIMO interference channel with output feedback and delayed CSIT," *IEEE Trans. Inf. Theory*, vol. PP, no. 99, p. 1, 2012.

[54] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, "Degrees of freedom of time correlated MISO Broadcast Channel with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315–328, Jan. 2013.

[55] M. J. Abdoli, A. Ghasemi, and A. K. Khandani, "On the Degrees of Freedom of K-User SISO interference and X channels with Delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6542–6561, Oct. 2013.

[56] M. Dai and B. Clerckx, "Transmit beamforming for MISO Broadcast Channels with Statistical and Delayed CSIT," *IEEE Trans. on Commun.*, vol. 63, no. 4, pp. 1202–1215, 2015.

[57] A. Vahid, M. A. Maddah-Ali, and A. S. Avestimehr, "Approximate Capacity Region of the MISO Broadcast Channels With Delayed CSIT," *IEEE Trans. on Commun.*, vol. 64, no. 7, pp. 2913–2924, July 2016.

[58] T. Gou and S. Jafar, "Optimal use of current and outdated channel state information: Degrees of freedom of the MISO BC with mixed CSIT," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1084–1087, Jul. 2012.

[59] P. de Kerret, D. Gesbert, J. Zhang, and P. Elia, "Optimally bridging the gap from delayed to perfect CSIT in the K-user MISO BC," in *Proc. IEEE Information Theory Workshop (ITW)*, Sep. 2016, pp. 300–304.

[60] R. Tandon, S. A. Jafar, S. Shamai (Shitz), and H. V. Poor, "On the synergistic benefits of alternating CSIT for the MISO BC," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4106–4128, 2013.

[61] J. Chen and P. Elia, "Toward the performance versus feedback tradeoff for the two-user MISO Broadcast Channel," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8336–8356, Dec. 2013.

[62] C. S. Vaze and M. K. Varanasi, "Dirty paper coding for the mimo cognitive radio channel with imperfect csit," in *2009 IEEE International Symposium on Information Theory*, June 2009, pp. 2532–2536.

[63] H. Du and T. Ratnarajah, "Robust joint signal and interference alignment for mimo cognitive radio network," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2012, pp. 448–452.

[64] O. Munoz-Medina, J. Vidal, and A. Agustin, "Linear transceiver design in nonregenerative relays with channel state information," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2593–2604, June 2007.

[65] X. Tang and Y. Hua, "Optimal design of non-regenerative mimo wireless relays," *IEEE Transactions on Wireless Communications*, vol. 6, no. 4, pp. 1398–1407, April 2007.

[66] H. W. Je, D. H. Kim, and K. B. Lee, "Joint precoding for mimo-relay systems with partial channel state information," in *2009 IEEE International Conference on Communications*, June 2009, pp. 1–5.

[67] Y. Huang, L. Yang, M. Bengtsson, and B. Ottersten, "A limited feedback joint precoding for amplify-and-forward relaying," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1347–1357, March 2010.

[68] J. Zhang and K. B. Letaief, "Interference management with relay cooperation in two-hop interference channels," *IEEE Wireless Communications Letters*, vol. 1, no. 3, pp. 165–168, June 2012.

[69] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Downlink multicell processing with limited-backhaul capacity," *EURASIP Journal on Advances in Signal Processing*, May 2009.

[70] J. Zhao, T. Q. S. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2762–2775, Jun. 2013.

[71] M. Jaber, M. A. Imran, R. Tafazolli, and A. Tukmanov, "5g backhaul challenges and emerging research directions: A survey," *IEEE Access*, vol. 4, pp. 1743–1766, 2016.

[72] P. Popovski, "Ultra-reliable communication in 5g wireless systems," in *1st International Conference on 5G for Ubiquitous Connectivity*, Nov 2014, pp. 146–151.

[73] R. Radner, "Team decision problems," *The Annals of Mathematical Statistics*, 1962.

[74] Y. C. Ho, "Team decision theory and information structures," *Proceedings of the IEEE*, vol. 68, no. 6, pp. 644–654, 1980.

[75] X. Rao, L. Ruan, and V. K. N. Lau, "CSI feedback reduction for MIMO interference alignment," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4428–4437, Sept. 2013.

[76] P. de Kerret and D. Gesbert, "Interference alignment with incomplete CSIT sharing," *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2563–2573, May. 2014.

[77] C. Hao and B. Clerckx, "Degrees-of-Freedom of the K-user MISO interference channel with delayed local CSIT," in *Proc. IEEE International Conference on Communications (ICC)*, June 2015, pp. 4217–4222.

[78] N. Lee, R. Tandon, and R. W. Heath, "Distributed space-time Interference Alignment with moderately Delayed CSIT," *IEEE Trans. Wireless Commun.*, vol. 14, no. 2, pp. 1048–1059, Feb. 2015.

[79] A. Vahid and R. Calderbank, "Two-user erasure Interference Channels with local delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 4910–4923, Sept. 2016.

[80] A. Lapidoth and Y. Steinberg, "The multiple access channel with two independent states each known causally to one encoder," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2010, pp. 480–484.

[81] R. Kolte, A. Özgür, and H. Permuter, "Cooperative binning for semideterministic channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 3, pp. 1231–1249, Mar. 2016.

[82] P. de Kerret and D. Gesbert, "Degrees of freedom of the network MIMO channel with distributed CSI," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6806–6824, Nov. 2012.

[83] R. Zakhour and D. Gesbert, "Team decision for the cooperative MIMO channel with imperfect CSIT sharing," in *Proc. Information Theory and Applications Workshop (ITA)*, 2010.

[84] Q. Li, P. de Kerret, D. Gesbert, and N. Gresset, "Robust regularized ZF in decentralized Broadcast Channel with correlated CSI noise," in *Proc. Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2015, pp. 329–336.

[85] D. Gesbert and P. de Kerret, "Team Methods for Device Cooperation in Wireless Networks," in *Cooperative and Graph Signal Processing*, ch. 18, pp. 469 – 487, Ed. by P. M. Djurić and C. Richard, Academic Press, 2018.

[86] S. Grandhi, R. Vijayan, and D. Goodman, "Distributed power control in cellular radio systems," *IEEE Trans. on Commun.*, vol. 42, no. 234, pp. 226–228, Feb 1994.

[87] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, 2008.

[88] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov 2010.

[89] V. Aggarwal, Y. Liu, and A. Sabharwal, "Sum capacity of interference channels with a local view: Impact of distributed decisions," *IEEE Trans. on Inf. Theory*, vol. 58, no. 3, pp. 1630–1659, 2012.

[90] A. Vahid and R. Calderbank, "Two-User Erasure Interference Channels With Local Delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 4910–4923, Sept 2016.

[91] G. Sartoretti, Y. Wu, W. Paivine, T. K. S. Kumar, S. Koenig, and H. Choset, "Distributed reinforcement learning for multi-robot decentralized collective construction," *Correll N., Schwager M., Otte M. (eds) Distributed Autonomous Robotic Systems. Springer Proceedings in Advanced Robotics*, vol. 9, 2019.

[92] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of Caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[93] J. Zhang, F. Engelmann, and P. Elia, "Coded caching for reducing csit-feedback in wireless communications," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2015, pp. 1099–1105.

[94] J. Zhang and P. Elia, "Fundamental Limits of Cache-Aided Wireless BC: Interplay of Coded-Caching and CSIT Feedback," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, pp. 3142–3160, May 2017.

[95] M. A. Hail, M. Amadeo, A. Molinaro, and S. Fischer, "Caching in named data networking for the wireless internet of things," in *2015 International Conference on Recent Advances in Internet of Things (RIoT)*, April 2015, pp. 1–6.

[96] S. Wang, X. Zhang, Y. Zhang, L. Wang, J. Yang, and W. Wang, "A survey on mobile edge networks: Convergence of computing, caching and communications," *IEEE Access*, vol. 5, pp. 6757–6779, 2017.

[97] A. Bazco, P. de Kerret, D. Gesbert, and N. Gresset, "Generalized degrees-of-freedom of the 2-user case MISO broadcast channel with distributed CSIT," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 1092–1096.

[98] A. Bazco-Nogueras, P. de Kerret, D. Gesbert, and N. Gresset, "Distributed CSIT Does Not Reduce the Generalized DoF of the 2-user MISO Broadcast Channel," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 685–688, June 2019.

[99] A. Bazco, P. de Kerret, D. Gesbert, and N. Gresset, "Méthode de transmission robuste au partage imparfait de l'information de canal entre transmetteurs," in *2017 GRETSI Colloque*, Septembre 2017.

[100] A. Bazco-Nogueras, P. de Kerret, D. Gesbert, and N. Gresset, "On the Degrees-of-Freedom of the K-user Distributed Broadcast Channel," 2018, submitted to IEEE Trans. Inf. Theory.

[101] A. Bazco-Nogueras, L. Miretti, P. de Kerret, D. Gesbert, and N. Gresset, "Achieving Vanishing Rate Loss in Decentralized Network MIMO," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1457–1461.

[102] ——, "Transmission robuste de zéro-forçage asymptotiquement optimale pour coopération imparfaite de transmetteurs," in *2019 GRETSI Colloque*, August 2019.

[103] A. Bazco-Nogueras, P. de Kerret, D. Gesbert, and N. Gresset, "Asymptotically Achieving Centralized Rate on the MxK Decentralized Network MISO," 2019, submitted to IEEE Trans. Information Theory.

[104] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec 2013.

[105] N. Golrezaei, A. G. Dimakis, and A. F. Molisch, "Scaling Behavior for Device-to-Device Communications With Distributed Caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4286–4298, 2014.

[106] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. Workshop on Mobile Cloud Computing (MCC)*, 2012, pp. 13–16.

[107] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges," *IEEE Wireless Commun.*, vol. 22, no. 2, pp. 152–160, Apr. 2015.

[108] *Final definition and evaluation of PHY layer approaches for RANaaS and joint backhaul-access layer.* Available: http://www.ict-ijoin.eu/deliverables/, [Accessed: 21/09/2015], May. 2011.

[109] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[110] T. Cover and A. Thomas, *Elements of information theory.* Wiley-Interscience, Jul. 2006.

[111] S. A. Jafar and M. J. Fakhereddin, "Degrees of Freedom for the MIMO Interference Channel," *IEEE Trans. Inf. Theory*, vol. 53, no. 7, pp. 2637–2642, July 2007.

[112] M. A. Maddah-Ali and D. N. C. Tse, "Completely stale transmitter channel state information is still very useful," in *Proc. Allerton Conference on Communication, Control, and Computing (Allerton)*, 2010.

[113] P. de Kerret, "Transmitter cooperation with distributed feedback in wireless networks," Ph.D. dissertation, TELECOM ParisTech, 2013. [Online]. Available: http://tel.archives-ouvertes.fr/tel-00952820

[114] P. de Kerret, M. Guillaud, and D. Gesbert, "Degrees of freedom of certain interference alignment schemes with distributed CSI," in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2013.

[115] J. Chen, S. Yang, and P. Elia, "On the fundamental feedback-vs-performance tradeoff over the MISO-BC with imperfect and delayed CSIT," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2013.

[116] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo, "On the degrees of freedom achievable through interference alignment in a MIMO interference channel," in *Proc. IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2011.

[117] M. A. Maddah-Ali and U. Niesen, "Cache-aided interference channels," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 809–813.

[118] A. Sengupta, R. Tandon, and O. Simeone, "Cache aided wireless networks: Tradeoffs between storage and latency," in *2016 Annual Conference on Information Science and Systems (CISS)*, March 2016, pp. 320–325.

[119] J. Hachem, U. Niesen, and S. N. Diggavi, "Degrees of freedom of cache-aided wireless interference networks," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 5359–5380, July 2018.

[120] A. Gholami Davoodi and S. A. Jafar, "Aligned Image Sets Under Channel Uncertainty: Settling Conjectures on the Collapse of Degrees of Freedom Under Finite Precision CSIT," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, pp. 5603–5618, Oct 2016.

[121] A. Lozano, R. W. Heath, and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5213–5226, 2013.

[122] V. V. Veeravalli and A. El Gamal, *Interference Management in Wireless Networks: Fundamental Bounds and the Role of Cooperation*. Cambridge University Press, 2018.

[123] C. Huang, S. A. Jafar, S. Shamai (Shitz), and S. Vishwanath, "On degrees of freedom region of MIMO networks without channel state information at transmitters," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 849–857, Feb. 2012.

[124] G. Bresler and D. Tse, "The two-user gaussian interference channel: a deterministic view," *European Transactions on Telecommunications*, vol. 19, no. 4, pp. 333–354, 2008. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ett.1287

[125] D. Tuninetti, "On interFerence channel with generalized feedback (IFC-GF)," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2007.

[126] W. Wu, S. Vishwanath, and A. Arapostathis, "Capacity of a class of cognitive radio channels: Interference channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, Nov. 2007.

[127] P. Mohapatra and C. Murthy, "Inner bound on the GDOF of the K-user MIMO Gaussian symmetric interference channel," *IEEE Trans. Commun.*, vol. PP, no. 99, pp. 1–10, 2012.

[128] C. S. Vaze and M. K. Varanasi, "The degree-of-freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5354–5374, Aug. 2012.

[129] A. Chaaban and A. Sezgin, "On the generalized degrees of freedom of the Gaussian interference relay channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4432–4461, Jul. 2012.

[130] S. Shamai and S. Verdú, "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.

[131] J. Lee and N. Jindal, "High SNR Analysis for MIMO Broadcast Channels: Dirty Paper Coding Versus Linear Precoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4787 – 4792, Dec. 2007.

[132] N. Jindal, "MIMO Broadcast Channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[133] S. Lin and I.-H. Wang, "Gaussian Broadcast Channels With Intermittent Connectivity and Hybrid State Information at the Transmitter," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, pp. 6362–6383, Sept 2018.

[134] S. Lashgari, R. Tandon, and S. Avestimehr, "MISO Broadcast Channel With Hybrid CSIT: Beyond Two Users," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7056–7077, Dec 2016.

[135] E. Piovano, H. Joudeh, and B. Clerckx, "Overloaded multiuser MISO transmission with imperfect CSIT," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov 2016, pp. 34–38.

[136] P. de Kerret, R. Fritzsche, D. Gesbert, and U. Salim, "Robust precoding for network MIMO with hierarchical CSIT," in *2014 11th Int. Symp. Wireless Commun. Systems (ISWCS)*, Aug 2014, pp. 987–991.

[137] P. de Kerret, A. Bazco, and D. Gesbert, "Enforcing coordination in network MIMO with unequal CSIT," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, Nov 2016, pp. 39–43.

[138] I. Atzeni and D. Gesbert, "Cooperative mimo precoding with distributed csi: A hierarchical approach," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2018, pp. 1–5.

[139] G. Hardy, E. Wright, D. Heath-Brown, and J. Silverman, *An Introduction to the Theory of Numbers*, ser. Oxford mathematics. Oxford University Press, 2008. [Online]. Available: https://books.google.fr/books?id=d3wpAQAAMAAJ

[140] D. E. Knuth, "Big Omicron and Big Omega and Big Theta," *SIGACT News*, vol. 8, no. 2, pp. 18–24, Apr. 1976. [Online]. Available: http://doi.acm.org/10.1145/1008328.1008329

[141] S. Janson, "Probability asymptotics: notes on notation," *arXiv preprint arXiv:1108.3924*, 2011.

[142] C. Hao, B. Rassouli, and B. Clerckx, "Achievable DoF Regions of MIMO Networks With Imperfect CSIT," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6587–6606, Oct 2017.

[143] M. W. Amos Lapidoth, Shlomo Shamai, "On the capacity of fading MIMO Broadcast Channels with imperfect transmitter side-information," 2006. [Online]. Available: http://arxiv.org/pdf/cs/0605079.pdf

[144] C. Hao, Y. Wu, and B. Clerckx, "Rate analysis of two-receiver MISO Broadcast Channel with finite rate feedback: A rate-splitting approach," *IEEE Trans. on Commun.*, vol. 63, no. 9, pp. 3232–3246, Sept. 2015.

[145] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, 2008.

[146] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, "Rate splitting for MIMO wireless networks: a promising PHY-layer strategy for LTE evolution," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 98–105, May 2016.

[147] A. Gholami Davoodi and S. A. Jafar, "Aligned Image Sets and the GDoF of Symmetric MIMO Interference Channel with Partial CSIT," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec 2017, pp. 1–6.

[148] ——, "Generalized Degrees of Freedom of the Symmetric K User Interference Channel Under Finite Precision CSIT," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, pp. 6561–6572, Oct 2017.

[149] A. G. Davoodi and S. Jafar, "Aligned Image Sets and the Generalized Degrees of Freedom of Symmetric MIMO Interference Channel With Partial CSIT," *IEEE Trans. Inf. Theory*, vol. 65, no. 1, pp. 406–417, Jan 2019.

[150] A. G. Davoodi and S. A. Jafar, "K-User Symmetric $M \times N$ MIMO Interference Channel Under Finite Precision CSIT: A GDoF Perspective," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 1126–1136, Feb 2019.

[151] E. Piovano, H. Joudeh, and B. Clerckx, "Generalized Degrees of Freedom of the Symmetric Cache-Aided MISO Broadcast Channel With Partial CSIT," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5799–5815, Sep. 2019.

[152] Veniam. (2019, Jul.) Moving Terabytes of data between vehicles and the cloud. [Online]. Available: https://veniam.com/

[153] C. T. K. Ng, D. Gunduz, A. J. Goldsmith, and E. Erkip, "Distortion minimization in Gaussian layered broadcast coding with successive Refinement," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5074–5086, 2009.

[154] O. E. Ayach and R. W. Heath, "Interference alignment with analog channel state feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 626–636, Feb. 2012.

[155] A. G. Davoodi and S. A. Jafar, "Sum-set inequalities from aligned image sets: Instruments for robust gdof bounds," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, June 2017, pp. 684–688.

[156] A. Gholami Davoodi and S. A. Jafar, "Generalized Degrees of Freedom of the Symmetric K-User Interference Channel under Finite Precision CSIT," *CoRR*, vol. abs/1601.06463, 2016. [Online]. Available: http://arxiv.org/abs/1601.06463

[157] M. Madiman, "On the entropy of sums," in *Proc. IEEE Information Theory Workshop (ITW)*, May 2008, pp. 303–307.

[158] P. Billingsley, *Probability and Measure*, ser. Wiley Series in Probability and Statistics. Wiley, 1995.

[159] R. Couillet and M. Debbah, *Random matrix methods for wireless Communications*. Cambridge University Press, 2011.

[160] D. G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*. Springer Publishing Company, Incorporated, 2015.

[161] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[162] A. Lapidoth, *A foundation in Digital communication*. Cambridge University Press, 2009.

[163] K. B. Petersen and M. S. Pedersen, *The Matrix Cookbook*, Nov. 2012. [Online]. Available: http://matrixcookbook.com

[164] H. V. Henderson and S. R. Searle, "On deriving the inverse of a sum of matrices," *Siam Review*, vol. 23, no. 1, pp. 53–60, Jan. 1981.