TESI DI DOTTORATO

UNIVERSITÀ DEGLI STUDI DI NAPOLI "FEDERICO II"

DIPARTIMENTO DI INGEGNERIA ELETTRONICA
E DELLE TELECOMUNICAZIONI

DOTTORATO DI RICERCA IN
INGEGNERIA ELETTRONICA E DELLE TELECOMUNICAZIONI

MIMO MULTIPLE ACCESS CHANNELS WITH PARTIAL CHANNEL STATE INFORMATION

AUGUSTO AUBRY

Il Coordinatore del Corso di Dottorato
Ch.mo Prof. Niccoló RINALDI

I Tutori
Ch.ma Prof.ssa Antonia Maria TULINO
Ch.mo Prof. Ernesto CONTE

A. A. 2009–2010

”Meno Male
che ci sei tu!!!!!”

Acknowledgments

Firstly, I would like to express my sincere gratitude to Professor Tulino and Professor Conte for their assistance and patience. Special Thanks are for Serena and my family (Mom, Dad, Alison, Astrid and Alexia). Surely I can not forget to thnak Antonio, and my colleagues, specially Ivana and Marco. I also thank Giulia, Nino and Rosaria, and my friends Francesco, Simone, Fabio and Giacomo.

Contents

List of Figures

Introduction

The increasing demand for wireless communication systems has been the starting point for many research activities and discussions about the best way to make a cleaver usage of the available spectrum and the available power, in order to satisfy the requirements of the users. In these systems, many are the features that have to be taken into account, such as the mobility of the users, the presence of time varying obstacles, power and bandwidth limitations, in order to determine their Shannon-Capacity, i.e. the maximum amount of information that can be reliably delivered. Indeed, only in this way it is possible to understand the real quality of services that can be offered to the users, and consequently to determine the different types of applications that can be implemented on such communication systems.
Most of the efforts done in Information Theoretic research in the study of wireless communication channels, has been conducted assuming coherent reception of the signals at the receiver side. The key result obtained is that the better way to contrast the randomness of wireless channels (fading), which is the main contribution in the limitation of the performance of a wireless system, is the use of multiple antennas at both the transmitter and the receiver sides (spatial diversity). Specifically, [1] [2] show that Multiple-Input-Multiple-Output (MIMO) channels have a linear capacity growth with the number of antennas for point-to-point wireless channels with rich scattering. This promise of linear capacity growth has been generalized to the Multiple Access Channel (MAC), [3] [4] [5], and Broadcast Channel, [6] [7], where the transmitters and the receivers are equipped with multiple antennas. The potentiality of MIMO systems to contrast the fading, has been also exploited in the contest of RADAR systems, [8] [9] , showing that the probability of detection can be improved resorting to angle diversity, i.e. looking to the target from different perspective angles.
However, especially in the Communication field, the aforementioned analysis of MIMO systems requires the receiver to perform a noise-free, multi-

dimensional channel estimation, without using communication resources. In practice, any channel estimation is noisy and uses system resources, so it is very important to understand the performance of a communication system for which only a partial information of the channel state is available at the receiver. For a single user MIMO channel, the trade-off between a good channel estimation and an increase of achievable data rate has been considered in [10] where the analysis is conducted using a lower bound for the capacity. In [11] assuming gaussian input, a lower and an upper bound for the mutual information are provided under the assumption that imperfect channel state information is available at the receiver. The fundamental conclusion that can be obtained is that the quality of the estimation has a great impact on the performance limits of the system. Specifically, assuming that the quality of the estimation doesn't depend on the available power at the transmitter, the system is interference limited, i.e. the achievable rate saturates; instead, if it is possible to obtain a consistent estimation of the channel at high signal-to-noise ratio, the rate growth is essentially linear with the number of antennas, as long as it is inferior to the coherent time of the channel. Obviously, moving toward a multi-user setting, such as Multiple Access Channel (MAC), produces a higher amount of information that has to be known to the receiver, for coherent reception, and the assumption of perfect knowledge of the channel state information is totally unrealistic. Some efforts for two user MAC channels with partial channel state information (CSI), has been performed in [12] [13].

The aim of this Thesis, (cf.[14, 15, 16]), is to generalize the previous results to an arbitrary multiple-user scenario where receiver and transmitters are equipped with multiple, possibly correlated, antennas, and only a partial CSI of the users is available at the receiver. In particular, focusing on the mutual information conditioned on such a partial CSI, an inner and an outer bound on the rate region achievable with Gaussian inputs for a MIMO MAC channel with partial CSI, are provided. So the result about the sum-rate performance for a MIMO MAC with partial CSI, obtained in [17], has been generalized through this analysis. Moreover, the behavior of the gaps between the upper and lower bounds of the mutual information terms, which define the achievable rate region, is analyzed as function of: i) the number of the users, ii) the number of the receiving antennas, and iii) the signal to noise ratio. Furthermore, the problem of finding the precodings that attempt to maximize the lower bound to the sum-rate, is considered too. Then the low-SNR and high-SNR regime for a MIMO MAC with partial CSI at the receiver is described, focusing, respectively, on the minimum energy per bit and on the multiacces slope region for the former, and on the high-SNR slope for the latter.

The aforementioned framework turns out to be very useful to describe the performance limits of two relevant scenarios of MAC system, in which only a partial CSI can be used at the receiver: *Cooperative MIMO Networks* and *Training Based Systems*.

Concerning the Cooperative MIMO Networks, it is well known that the optimal management of cochannel interference is of great importance for any multiuser wireless network. In conventional cellular systems, each user is served by a single base station, according to a pre-fixed criteria, such as signal strength. Focusing on the uplink, each user then causes interference to users served by all other base stations. However, if the received signals at all the base stations could be jointly processed at a central processor (CP), there is no more interference and all signals are information-bearing. While challenging to implement, such joint processing is at least theoretically possible because the base stations are typically connected by a high-speed backhaul network. This decoding strategy has been referred as Cooperative MIMO network, multi-cell processing, base station coordination, macrodiversity, etc, [18], [19]. Obviously, the ability to suppress interference through receivers cooperation, as above, is crucially dependent on the availability and quality of the CSI for all the users at the central processor. In practice, the assumption of perfect CSI is particularly untenable in the context of multi-cell processing because of the need to estimate very weak channels (from faraway users). Therefore, it becomes important to understand the fundamental limits imposed by imperfect CSI on the performance of multi-cell processing. Consequently, the derived bounds can be applied to study Gaussian interference networks with cooperative receiver processing, assuming infinite-capacity backhaul links from the base stations to the CP. Each user is supposed to be decoded according to the received signals at Q base stations (Q the cluster size). Firstly, the case where the CP has full CSI for each user at the Q base stations that form his processing cluster, and no CSI (apart from the statistical characterization) for the other user-to-base links, is considered. Then the same situation with the further refinement that the CP has only quantized CSI instead of full CSI, is studied. In both cases, the impact of partial CSI through the derived upper and lower bounds on the sum rate (with Gaussian signaling) is analyzed, focusing in particular on the impact of Q.

Concerning the Training Based Systems, in most of telecommunication networks, before the transmission of the information, a synchronization phase occurs, during which the receiver estimates the channel matrix. In particular a standard technique to allow the receivers to estimate the channel matrix consists of the transmission from the users of training sequences among the data, [20]. Then, the founded inner and outer bounds can be applied to a MIMO MAC affected by block fading, where each block is divided into training and data transmission phases. Since the partial CSI available at the receiver is related to a minimum-mean-square-error (MMSE) channel estimation process, the attention is focused on the optimization of training signals, considering two performance metrics: the trace and the determinant, respectively, of the covariance matrix of the channel estimation error. Therefore, a robust design of the training sequences is considered, and the problem of the precoding for a Training Based System is solved. Finally, the impact of antennas' correlation at the transmitter and the receiver side on the performance of the system is analyzed, using the inner and outer bounds.

The Thesis is organized as follows. Chapter 1 contains the Information Theoretic results for a MIMO MAC with partial CSI; after describing the system model and the assumptions for the multiple access channel under consideration, an inner and an outer bound on the achievable rate region with partial CSI at the receiver and Gaussian input, are derived. Furthermore, the properties of the lower and upper bounds, characterizing the achievable rate region, are analyzed, and the problem of the optimal precoding applied by the users of the network is considered. In Chapter 2, a Cooperative MIMO Network and a Training Based System are characterized in an Information Theoretic sense. In particular, in Section 2.1, the derived bounds are applied for studing a Cooperative MIMO Network with incomplete and/or imperfect CSI, while Section 2.2 deals with the special case in which the CSI at the receiver is obtained through training sequences, and it contains the derivation of the optimal structure of such sequences. Chapter 3 discusses with the asymptotic behavior of a MIMO MAC with partial CSI in the low-SNR regime and high-SNR regime, then the Conclusions.

Notation: In the following, $\mathbf{X}^T$, $\mathbf{X}^\dagger$, $\|\mathbf{X}\|$, and $[\mathbf{X}]_{k,i}$ respectively denote the transpose, the Hermitian transpose, the Frobenius norm and the element at the k-th row and the i-th column of the matrix $\mathbf{X} \in \mathbb{C}^{m\times n}$. $\mathbf{X} = [\mathbf{x_1}, \ldots, \mathbf{x_n}] \in \mathbb{C}^{m\times n}$ is a matrix containing the $m-$dimensional vectors $\mathbf{x_1}, \ldots, \mathbf{x_n}$ as its columns. $\mathbf{A} \otimes \mathbf{B} \in \mathbb{C}^{kn\times kn}$ denotes the Kronecker (tensor) product of the matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ with the matrix $\mathbf{B} \in \mathbb{C}^{k\times k}$. $\mathrm{tr}\{\mathbf{A}\}$, $|\mathbf{A}|$ and $\lambda_{\mathsf{max}}(A)$ denote the trace, the determinant and the maximum eigenvalue of a square matrix $\mathbf{A}$. $\mathrm{vec}\{\mathbf{X}\} \in \mathbb{C}^{mn}$ indicates the vector obtained stacking up the rows of the matrix $\mathbf{X} \in \mathbb{C}^{m\times n}$. $\mathbf{I}_m$ indicates the identity matrix of order m. $\mathbf{I}_m^{(p)}$ denotes the $m \times p$ matrix defined as follows: $[\mathbf{I}_m^{(p)}]_{i,j} = \delta_{i,j}$. $\mathbf{e_{m,i}}$ indicates a row vector of dimension m such that $[\mathbf{e_{m,i}}]_j = \delta_{i,j}$. $\mathbf{1}_R$ is the R-dimensional vector with all entries equal to 1. Finally, $\mathbf{x}_{|\mathbf{y}}$ indicates the random variable $\mathbf{x}$ conditioned on $\mathbf{y}$, $\mathbb{E}[\cdot]$ denotes statistical expectation and $\mathbb{E}_{\mathbf{Y}}[\cdot]$ denotes conditional statistical expectation with respect to $\mathbf{Y}$.

Chapter 1

MIMO MAC with Partial CSI

In this chapter, the system model for a MIMO Multiple-Access Channel is presented and its Information-Theoretic characterization is considered, assuming a partial channel state information at the receiver side. In particular, an inner and an outer bound to the achievable rate region for a MIMO MAC with partial CSI, are derived. Some interesting properties about the lower and upper bounds derived are shown, and the behavior of the gap with respect to the number of the users and the receiving antennas, is analyzed. Finally, the optimization of precodings that have to be used by the users, is considered.

1.1 MIMO MAC Model

A MIMO MAC with a base station (BS) equipped with L[1] receiving antennas and K users, each with M transmitting antennas is considered (the generalization to the case of different numbers of antennas at the users is straightforward, and will therefore not be considered here).

Using the standard discrete-time equivalent channel representation for flat-flat fading channels [2], [21] [22] [23], the L-dimensional baseband complex

[1] Let us observe that in the case of cooperative MIMO networks, L is the sum of the receiving antennas of the different BSs distributed in the network.

[2] If the fading process is frequency selective, the channel can be decomposed into parallel noninteracting subchannels, each of which conforms to (1.1), considering simultaneous transmission of narrowband signals on not-overlapping frequency bands.

signal received at the base station, is given by:

$$\mathbf{y} = \sum_{k=1}^{K} \mathbf{H}^{(k)}\mathbf{x}^{(k)} + \mathbf{n}, \tag{1.1}$$

where $\mathbf{H}^{(k)}$ is the $L \times M$ channel matrix between the k-th user and the base station, $\mathbf{x}^{(k)}$ is the M-dimensional input vector for the k-th user, and $\mathbf{n}$ is the additive circularly symmetric zero-mean Gaussian noise. The single-sided spectral density of the noise is denoted by

$$N_0 = \frac{E[\|\mathbf{n}\|^2]}{L} \tag{1.2}$$

and we assume its normalized spatial covariance to be

$$\mathbf{\Phi}_{\mathbf{n}} \triangleq \frac{E[\mathbf{n}\mathbf{n}^\dagger]}{N_0} = \mathbf{I}_L. \tag{1.3}$$

The input signals are assumed to be zero-mean with normalized input covariance matrix

$$\mathbf{P}_k = \frac{\mathbb{E}\left[\mathbf{x}^{(k)}\mathbf{x}^{(k)^\dagger}\right]}{P}, \tag{1.4}$$

according to some power allocation policy, where P is the maximum available power per user. Moreover, we consider a normalized input power constraints $\mathrm{tr}\{\mathbf{P}_k\} \leq 1$ and in the following we will indicate with SNR the quantity $\mathsf{SNR} = \frac{P}{N_0}$. By stacking the transmitted K input vectors in (1.1) to form an MK-dimensional vector, $\mathbf{x} = [\mathbf{x}^{(1)^\dagger} \dots \mathbf{x}^{(K)^\dagger}]^\dagger$, we can write

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{1.5}$$

where $\mathbf{H} = \left[\mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \dots, \mathbf{H}^{(K)}\right]$ is the $L \times MK$ channel matrix whose k-th block is the matrix $\mathbf{H}^{(k)}$.

Let us now consider the statistical characterization of the channel matrices $\mathbf{H}^{(k)}$. We assume that $\mathbf{H}^{(k)}$ are statistically independent across users and with finite second order moments. Moreover, if explicitly stated, each channel $\mathbf{H}^{(k)}$ is supposed to be described by the UIU model [24], i.e. it turns out:

$$\mathbf{H}^{(k)} = \mathbf{U}_{\mathrm{R}}^{(k)}\mathbf{H}_w^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)^\dagger}, \quad k = 1, \dots, K, \tag{1.6}$$

where $\mathbf{U}_{\mathrm{R}}^{(k)}$ is a $L \times L$ unitary matrix, $\mathbf{U}_{\mathrm{T}}^{(k)}$ is a $M \times M$ unitary matrix and $\mathbf{H}_{w}^{(k)}$ is a zero-mean $L \times M$ Gaussian random matrix of independent elements with power profile $\boldsymbol{\Sigma}_{w}^{(k)}$, i.e. $E\left[\left|[\mathbf{H}_{w}^{(k)}]_{i,j}\right|^2\right] = [\boldsymbol{\Sigma}_{w}^{(k)}]_{i,j}$.

Such a statistical characterization covers most of the channel models that are considered in the literature. In particular, it encompasses the classical channel model of independent Rayleigh-faded gain, or the separable correlation model, for which the correlation between two transmitting antennas does not depend on the receiving antennas and, dually, the correlation between two receive antennas is independent of the transmitting antennas. It also includes the virtual representation, which is amply utilized in the literature to describe the channel gain for a MIMO system equipped with linear transmitting and receiving arrays, and it is characterized by two Fourier matrices for $\mathbf{U}_{\mathrm{R}}^{(k)}$ and $\mathbf{U}_{\mathrm{T}}^{(k)}$. Let us observe that with this model we can take into account the different path losses between the BS and the users, by choosing appropriately the set of values for $\mathrm{tr}\left\{\boldsymbol{\Sigma}_{w}^{(k)}\right\}$.

1.2 Information-Theoretic Analysis

In this section, we characterize in an information-theoretic sense a memoryless MIMO MAC channel when a statistic $\mathbf{S}$ of the channel realization $\mathbf{H}$, is available at the receiver[3]. For every memoryless MAC channel with K users, marginal power constraint $\mathbb{E}\left[||\mathbf{x}_k||^2\right] \leq \beta_k$, $k = 1, \ldots K$, and side information $\mathbf{S}$ at the receiver, the direct achievable rate region is given by, [25] and [26]:

$$\mathbf{A}(\boldsymbol{\beta}) = \bigcup_{\substack{\mathbf{x}_k :\, \forall k \\ \mathbb{E}\left[||\mathbf{x}_k||^2\right] \leq \beta_k}} \left\{ (R_1, \ldots, R_K) : \sum_{k \in \mathcal{A}} R_k \leq I(\mathbf{x}_{\mathcal{A}}; \mathbf{y} | \mathbf{S}, \mathbf{x}_{\bar{\mathcal{A}}}) \ \forall \ \mathcal{A} \subseteq \mathcal{U} \right\} \tag{1.7}$$

where $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_K]$, $\mathcal{A}$ is a generic subset of the set of users while $\bar{\mathcal{A}}$ is its complementary subset with respect to $\mathcal{U} \equiv \{1, \ldots, K\}$, and the union is over independent distributions on the input alphabets.

As proved in [25] and [26], assuming that each user has a power constraint

[3] With respect to channel use variable i, we assume that $\{(\mathbf{H}_i, \mathbf{S}_i)\}_{i \in \mathcal{N}}$ is an i.i.d. random process; the results still hold if $\{(\mathbf{H}_i, \mathbf{S}_i)\}_{i \in \mathcal{N}}$ is an ergodic process.

$\mathbb{E}\left[||\mathbf{x}_i||^2\right] \leq P$, the Capacity rate region $\mathbf{C}(P)$ is given by:

$$\mathbf{C}(P) = \left\{ (R_1, \ldots, R_K) : (R_1, \ldots, R_K, P, \ldots, P) \in co \bigcup_{\beta_k > 0, \forall k} (\mathbf{A}(\boldsymbol{\beta}), \boldsymbol{\beta}) \right\} \tag{1.8}$$

where $co\,(\mathcal{S})$, is the convex hull of the set of points that belong to $\mathcal{S}$, and corresponds to the time sharing of different codebooks.

Let us, now, consider the system model given in (1.1), in the case of perfect knowledge of the channel matrix $\mathbf{H}$ at the receiver side, i.e. $\mathbf{S} = \mathbf{H}$; it has been shown in [27] [22], that the capacity rate region $\mathbf{C}(P)$, defined in (1.8), can be particularized as:

$$\bigcup_{\substack{\mathbf{Q}_k \succeq 0: \\ \mathrm{tr}\{\mathbf{Q}_k\} = P, \\ \forall k}} \left\{ (R_1, \ldots, R_K) : \sum_{k \in \mathcal{A}} R_k \leq \mathbb{E}\left[\log \left| \mathbf{I}_L + \sum_{k \in \mathcal{A}} \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^{\dagger} \right| \right] \forall \mathcal{A} \subseteq \mathcal{U} \right\} \tag{1.9}$$

It can be observed that the capacity rate region, given in (1.9), corresponds to the union of the achievable rate region $\mathbf{A}(P)$ evaluated for independent zero mean Gaussian random vectors, with covariance matrices $\mathbf{Q}_k$, subjected to the power constraint $\mathrm{tr}\{\mathbf{Q}_k\} = P$.

From the above discussion, it is evident that for a complete characterization of the achievable rate regions $\mathbf{A}(\boldsymbol{\beta})$, and consequently for the capacity rate region of the system, the following quantities must be analyzed:

$$I(\mathbf{x}_{\mathcal{F}}; \mathbf{y} | \mathbf{S}, \mathbf{x}_{\mathcal{C}}) \tag{1.10}$$

where $\mathcal{F}$ and $\mathcal{C}$ are generic disjoint subsets of the set $\mathcal{U}$ of all users. Indeed, a special case of (1.10) is:

$$I(\mathbf{x}_{\mathcal{A}}; \mathbf{y} | \mathbf{S}, \mathbf{x}_{\bar{\mathcal{A}}}) \tag{1.11}$$

where $\mathcal{A}$ is a generic subset of the set of users while $\bar{\mathcal{A}}$ is its complementary subset with respect to $\mathcal{U}$, which are used to define achievable rate regions for the MIMO-MAC channel. The sum-rate of the MIMO MAC channel can be obtained by evaluating (1.11) when $\mathcal{A} \equiv \mathcal{U}$, and the other terms correspond to the corner points of the achievable rate region.
The evaluation of (1.10) is, in general, very difficult. In the following, we

assume that:

Hypothesis 1 *The input distribution of each user is zero mean Gaussian.*

Hypothesis 2 *Conditioned on* $\mathbf{S}$*, the random matrix* $\mathbf{H}_{|\mathbf{S}}$ *is circularly symmetric Gaussian with mean* $\hat{\mathbf{H}} = \mathbb{E}_{\mathbf{S}}[\mathbf{H}]$[4]*. Notice that* $\hat{\mathbf{H}}$ *can be seen as the estimate of the channel matrix* $\mathbf{H}$ *based on* $\mathbf{S}$*. Thus we are equivalently assuming that, conditioned on* $\mathbf{S}$*, the uncertainty*

$$\mathbf{Z} = \mathbf{H}_{|\mathbf{S}} - \mathbb{E}_{\mathbf{S}}[\mathbf{H}] = \mathbf{H}_{|\mathbf{S}} - \hat{\mathbf{H}} \tag{1.12}$$

is a circularly symmetric zero-mean Gaussian random matrix whose entries have a generic correlation structure.

Note that

$$\mathbf{Z} = \left[\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(K)}\right],$$

where $\mathbf{Z}^{(k)} = \mathbf{H}^{(k)}_{|\mathbf{S}} - \mathbb{E}_{\mathbf{S}}[\mathbf{H}^{(k)}] = \mathbf{H}^{(k)}_{|\mathbf{S}} - \hat{\mathbf{H}}^{(k)}$.

The assumption of zero mean Gaussian input distributions allows us to evaluate an achievable rate region for the MIMO-MAC with partial channel state information. Moreover, this analysis can be reinterpreted as a perturbation analysis of the ideal situation of coherent reception, for which it is well known the optimality of zero mean Gaussian distributions.

Let us now introduce the following notation which is going to be used in the subsequent derivations. Given a user subset $\mathcal{R} \subset \{1, \ldots, K\}$, we denote:

- by $|\mathcal{R}|$ the cardinality of the subset $\mathcal{R}$;
- by $\mathbf{x}_{\mathcal{R}}$ the $M|\mathcal{R}|$-dimensional vector obtained by concatenating the M-dimensional vectors $\mathbf{x}^{(k)}$, with $k \in \mathcal{R}$ (i.e. the input vectors of the users indexed by the elements of $\mathcal{R}$);
- by $\overline{\mathcal{R}} \subseteq \mathcal{U}$ the user subset such that $\overline{\mathcal{R}} \cup \mathcal{R} \equiv \mathcal{U}$ and $\overline{\mathcal{R}} \cap \mathcal{R} \equiv \emptyset$;
- by $\hat{\mathbf{H}}_{\mathcal{R}}$ and $\mathbf{Z}_{\mathcal{R}}$ the matrices made up deleting the blocks, not indexed by the elements of R, respectively of $\hat{\mathbf{H}} = \left[\hat{\mathbf{H}}^{(1)}, \ldots, \hat{\mathbf{H}}^{(K)}\right]$ and of $\mathbf{Z} = \left[\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(K)}\right]$;

[4] For technical conditions, the probability density function of $\mathbf{H}_{|\mathbf{S}}$ is assumed as a continuous function of S, and that $\mathbf{S}$ has a smooth probability density function: such conditions are satisfied, for example, when $(\mathbf{S}, \mathbf{H})$ are jointly Gaussian.

- by $\mathbf{P}_{\mathcal{R}}$ the $M|\mathcal{R}|$-dimensional block-diagonal matrix, containing along its diagonal the $M \times M$ normalized covariance matrices $\mathbf{P}_k$, with $k \in \mathcal{R}$. Note that, since $\mathrm{tr}\{\mathbf{P}_k\} \leq 1$, for all $k \in \mathcal{R}$, then $\mathrm{tr}\{\mathbf{P}_{\mathcal{R}}\} \leq |\mathcal{R}|$.

Under the previous hypotheses we derive lower and upper bounds for the mutual information in (1.10), which are given in Theorem 1 using the following definition.

Definition 1 *Let $\mathcal{A} \subseteq \mathcal{U}$ and $\mathcal{B} \subseteq \mathcal{U}$ be arbitrary disjoint user subsets. Then, $I_{\mathsf{L}(\mathcal{A},\mathcal{B})}$[5] is a function of the subsets $\mathcal{A}$, $\mathcal{B}$, defined as follows:*

$$I_{\mathsf{L}(\mathcal{A},\mathcal{B})} = \mathbb{E}\left[\log\left|\mathbf{I}_L + \left(\mathbf{\Omega}_{\overline{\mathcal{B}}} + \mathbf{\Gamma}_{\overline{\mathcal{A}\cup\mathcal{B}}} + \mathbf{\Phi}_{\mathcal{B}} + \frac{1}{\mathsf{snr}}\mathbf{I}_L\right)^{-1}\mathbf{\Gamma}_{\mathcal{A}}\right|\right] \tag{1.13}$$

where for any arbitrary user subset $\mathcal{R} \subseteq \mathcal{U}$, $\mathbf{\Gamma}_{\mathcal{R}}$ and $\mathbf{\Phi}_{\mathcal{R}}$ are defined as:

$$\mathbf{\Gamma}_{\mathcal{R}} = \hat{\mathbf{H}}_{\mathcal{R}}\mathbf{P}_{\mathcal{R}}\hat{\mathbf{H}}_{\mathcal{R}}^{\dagger},$$

and

$$\mathbf{\Phi}_{\mathcal{R}} = \frac{1}{P}\mathbb{E}_{\mathbf{S},\mathbf{x}_{\mathcal{R}}}\left[\mathbf{Z}_{\mathcal{R}}\mathbf{x}_{\mathcal{R}}\mathbf{x}_{\mathcal{R}}^{\dagger}\mathbf{Z}_{\mathcal{R}}^{\dagger}\right], \tag{1.14}$$

while $\mathbf{\Omega}_{\mathcal{R}} = \mathbb{E}_{\mathbf{S}}[\mathbf{\Phi}_{\mathcal{R}}]$.

Theorem 1 *Let $\mathcal{F} \subseteq \mathcal{U}$ and $\mathcal{C} \subseteq \mathcal{U}$ be disjoint user subsets. Then, for every fixed covariance matrix $\mathbf{P}_{\mathcal{U}}$, the mutual information in (1.10) is lower and upper bounded by:*

$$I_{\mathsf{L}(\mathcal{F},\mathcal{C})} \leq I(\mathbf{x}_{\mathcal{F}};\mathbf{y}|\mathbf{S},\mathbf{x}_{\mathcal{C}}) \leq I_{\mathsf{L}(\overline{\mathcal{C}},\mathcal{C})} - I_{\mathsf{L}(\overline{\mathcal{F}\cup\mathcal{C}},\mathcal{F}\cup\mathcal{C})} + \mathbf{\Delta}_{\overline{\mathcal{C}}} \tag{1.15}$$

with

$$\mathbf{\Delta}_{\overline{\mathcal{C}}} = \mathbb{E}\left[\log\frac{|\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Omega}_{\overline{\mathcal{C}}} + \mathsf{SNR}\,\mathbf{\Phi}_{\mathcal{C}}|}{|\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Phi}_{\mathcal{U}}|}\right] \tag{1.16}$$

The expectation in (1.16) is over the joint Gaussian distribution of the input vector $\mathbf{x}$ and over $\mathbf{S}$.
Proof: *See Appendix A* ***Proof of Theorem 1***

[5]When $\mathcal{A} = \emptyset$, the convention is to assume $I_{\mathsf{L}(\mathcal{A},\mathcal{B})} = 0$ for any $\mathcal{B}$.

Note that the lower bound to the mutual information can be interpreted in terms of a degraded channel, i.e., it is equivalent to the rate of a channel for which $\hat{\mathbf{H}}$ is perfectly known to the receiver and the noise covariance matrix is given by $\mathsf{SNR}\,\left(\mathbf{\Omega}_{\overline{\mathcal{C}}} + \mathbf{\Phi}_{\mathcal{C}}\right) + \mathbf{I}_L \succeq \mathbf{I}_L$, where the inequality is a matrix inequality, that correspond to an increase of the noise power in all the received directions. This generalizes the same interpretation given in [10] for a single-user channel.

The above analysis is related to the generic performance analysis of the system; we now give also a direct upper bound and lower bound to the sum rate performance, that is the most significant figure of merit for the MAC channels:

Corollary 1 *A lower bound, I_{Lower}, and an upper bound, I_{Upper}, of mutual information, $I(\mathbf{x};\mathbf{y}|\mathbf{S})$, for every fixed covariance matrix $\mathbf{P}_{\mathcal{U}}$, are given by:*

$$I_{\mathsf{Lower}} \leq I(\mathbf{x};\mathbf{y}|\mathbf{S}) \leq I_{\mathsf{Upper}} \tag{1.17}$$

where

$$I_{\mathsf{Lower}} = \mathbb{E}\left[\log\left|\mathbf{I}_L + \mathsf{SNR}\,\left(\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}}\right)^{-1}\hat{\mathbf{H}}\mathbf{P}_{\mathcal{U}}\hat{\mathbf{H}}^{\dagger}\right|\right]$$

and

$$I_{\mathsf{Upper}} \quad = \quad I_{\mathsf{Lower}} + \mathbf{\Delta}_{\mathcal{U}}$$

with

$$\mathbf{\Delta}_{\mathcal{U}} = \mathbb{E}\left[\log\frac{|\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}}|}{|\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Phi}_{\mathcal{U}}|}\right] \tag{1.18}$$

Proof: *It is easily obtained from Theorem 1 choosing the subset $\mathcal{F} \equiv \mathcal{U}$ and consequently $\mathcal{C} = \varnothing$.*

Note that, in Corollary 1, by definition $\mathbf{\Omega}_{\mathcal{U}}$ and $\mathbf{P}_{\mathcal{U}}$ boil down to:

$$\mathbf{\Omega}_{\mathcal{U}} = \mathbb{E}_{\mathbf{S}}\left[\mathbf{Z}\mathbf{P}_{\mathcal{U}}\mathbf{Z}^{\dagger}\right],$$

and to:

$$\mathbf{P}_{\mathcal{U}} = \mathrm{diag}\left(\mathbf{P}_1, \ldots, \mathbf{P}_K\right) \tag{1.19}$$

Let us observe that from the derived lower and upper bounds given in Theorem 1, it is immediate to evaluate an inner bound and an outer bound on the achievable rate region, with Gaussian input, starting from (1.7).

1.2.1 Lower Bound and Upper Bound: Properties and Behavior for Large Dimensions

In this subsection, some properties concerning the lower and upper bounds, previously derived, are presented. Some considerations about the obtained lower bounds are in order. First of all, applying the Jensen's inequality[6], it turns out that:

$$\begin{aligned}
I_{\mathsf{L}(\mathcal{A},\mathcal{B})} =& \mathbb{E}\left[\log\left|\mathbf{I}_L + \left(\boldsymbol{\Omega}_{\overline{\mathcal{B}}} + \boldsymbol{\Gamma}_{\overline{\mathcal{A}\cup\mathcal{B}}} + \boldsymbol{\Phi}_{\mathcal{B}} + \frac{\mathbf{I}_L}{\mathsf{snr}}\right)^{-1}\boldsymbol{\Gamma}_{\mathcal{A}}\right|\right] \\
>& \mathbb{E}\left[\log\left|\mathbf{I}_L + \left(\boldsymbol{\Omega}_{\overline{\mathcal{B}}} + \boldsymbol{\Gamma}_{\overline{\mathcal{A}\cup\mathcal{B}}} + \mathbb{E}_{\mathbf{S}}[\boldsymbol{\Phi}_{\mathcal{B}}] + \frac{\mathbf{I}_L}{\mathsf{snr}}\right)^{-1}\boldsymbol{\Gamma}_{\mathcal{A}}\right|\right] \\
=& \mathbb{E}\left[\log\left|\mathbf{I}_L + \left(\boldsymbol{\Omega}_{\mathcal{U}} + \boldsymbol{\Gamma}_{\overline{\mathcal{A}\cup\mathcal{B}}} + \frac{1}{\mathsf{snr}}\mathbf{I}_L\right)^{-1}\boldsymbol{\Gamma}_{\mathcal{A}}\right|\right].
\end{aligned} \tag{1.20}$$

For any possible assignment of the indices $\{1,\ldots,K\}$ to the K users, using (1.20), the following sequence of inequalities holds:

$$\begin{aligned}
I(\mathbf{x};\mathbf{y}|\mathbf{S}) =& \sum_{k=1}^{K} I(\mathbf{x}_{\{k\}};\mathbf{y}|\mathbf{S},\mathbf{x}_{\{1,\ldots,k-1\}}) && (1.21)\\
\geq& \sum_{k=1}^{K} I_{\mathsf{L}(\{k\},\{1,\ldots,k-1\})} && (1.22)\\
=& \sum_{k=1}^{K} \mathbb{E}\left[\log\left|\mathbf{I}_L + \left(\boldsymbol{\Omega}_{\overline{\mathcal{B}}} + \boldsymbol{\Gamma}_{\overline{\mathcal{A}\cup\mathcal{B}}} + \boldsymbol{\Phi}_{\mathcal{B}} + \frac{\mathbf{I}_L}{\mathsf{snr}}\right)^{-1}\boldsymbol{\Gamma}_{\mathcal{A}}\right|\right] && (1.23)\\
>& \sum_{k=1}^{K} R_k && (1.24)\\
=& I_{\mathsf{Lower}} && (1.25)
\end{aligned}$$

[6]Recall that the function $\log\left|\mathbf{I} + \mathbf{A}^{-1}\mathbf{B}\right|$, for a fixed $\mathbf{B}$, is a convex function with respect to $\mathbf{A}$ and that $f(\mathbf{C}\mathbf{x} + \mathbf{b})$ is a convex function in $\mathbf{x}$, if $f(\mathbf{y})$ is a convex function in $\mathbf{y}$, (affine transformation).

where (1.21) follows from the chain rule for the mutual information, (1.22) is consequence of Theorem 1, (1.23) follows from (1.13) with:

$$\overline{\mathcal{B}} = \{K, \ldots, k\} \qquad \overline{\mathcal{A} \cup \mathcal{B}} = \{K, \ldots, k+1\},$$

$$\mathcal{B} = \{1, \ldots, k-1\} \qquad \mathcal{A} = \{k\}$$

and finally (1.24) and (1.25) follows from (1.20) with

$$R_k = \mathbb{E}\left[\log\left|\mathbf{I}_L + \left(\mathbf{\Omega}_{\mathcal{U}} + \mathbf{\Gamma}_{\{k+1,\ldots,K\}} + \tfrac{1}{\mathsf{snr}}\mathbf{I}_L\right)^{-1}\mathbf{\Gamma}_{\{k\}}\right|\right] \tag{1.26}$$

and

$$I_{\mathsf{Lower}} = \mathbb{E}\left[\log\left|\mathbf{I}_L + \mathsf{SNR}\ (\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}})^{-1}\,\hat{\mathbf{H}}\mathbf{P}_{\mathcal{U}}\hat{\mathbf{H}}^{\dagger}\right|\right]. \tag{1.27}$$

From the above chain of inequalities, the following result is obtained:

Corollary 2 *Denote by π a permutation which belongs to the symmetric group, $\mathcal{S}_K$, (i.e. the set of all permutations of the indices $\{1, \ldots, K\}$). A tighter lower bound on the sum rate $I(\mathbf{x};\mathbf{y}|\mathbf{S})$, which can be achieved exploiting a successive interference cancelation strategy, is provided by:*

$$I(\mathbf{x};\mathbf{y}|\mathbf{S}) \geq \max_{\pi \in \mathcal{S}_K} \tilde{I}_{\mathsf{Lower}}(\pi) \geq I_{\mathsf{Lower}} \tag{1.28}$$

with

$$\tilde{I}_{\mathsf{Lower}}(\pi) = \sum_{k=1}^{K} \mathbb{E}\left[\log\left|\mathbf{I}_L + \left(\mathbf{Q}_{\pi} + \tfrac{1}{\mathsf{snr}}\mathbf{I}_L\right)^{-1}\mathbf{\Gamma}_{\{k\}}\right|\right] \tag{1.29}$$

where π is representative of the specific chosen ordering of the users, $\mathbf{Q}_{\pi}$ is given by:

$$\mathbf{Q}_{\pi} = \mathbf{\Omega}_{\{\pi(K),\ldots,\pi(k)\}} + \mathbf{\Gamma}_{\{\pi(K),\ldots,\pi(k+1)\}} + \mathbf{\Phi}_{\{\pi(1),\ldots,\pi(k-1)\}}$$

and finally I_{Lower} is given in Corollary 1.

It is worth emphasizing that, in many case of interest, the lower bound and the upper bound for the sum-rate are very tight. Consequently from Corollary 2, it follows that the successive interference cancelation is a useful decoding strategy to ensure good performance in terms of the achievable sum-rate.

Let us now observe that in the case of two user MIMO MAC channel,

$$I_{\mathsf{L}(\{1\},\{2\})} > \mathbb{E}\left[\log\left|\mathbf{I}_L + \left(\boldsymbol{\Omega}_{\mathcal{U}} + \tfrac{1}{\mathsf{snr}}\mathbf{I}_L\right)^{-1}\boldsymbol{\Gamma}_{\{1\}}\right|\right]$$

$$I_{\mathsf{L}(\{2\},\{1\})} > \mathbb{E}\left[\log\left|\mathbf{I}_L + \left(\boldsymbol{\Omega}_{\mathcal{U}} + \tfrac{1}{\mathsf{snr}}\mathbf{I}_L\right)^{-1}\boldsymbol{\Gamma}_{\{2\}}\right|\right]$$

and

$$\max\left(I_{\mathsf{L}(\{1\})} + I_{\mathsf{L}(\{2\},\{1\})}, I_{\mathsf{L}(\{2\})} + I_{\mathsf{L}(\{1\},\{2\})}\right) > \mathbb{E}\left[\log\left|\mathbf{I}_L + \left(\boldsymbol{\Omega}_{\mathcal{U}} + \tfrac{1}{\mathsf{snr}}\mathbf{I}_L\right)^{-1}\boldsymbol{\Gamma}_{\{1,2\}}\right|\right]$$

where the terms on the lefthand side define our achievable rate region, while the term terms on the righthand side define achievable rate region, derived in [12] [13]. Then, our inner bound is larger than the one given in [12] [13], describing better the performance limits of the system. The same conclusion still hold for higher number of users, i.e.: the developed tools better describe the achievable rate region, with respect to the multiple user extension of [12] [13].

Theorem 1 essentially provides an inner bound and an outer bound on the achievable rate region of a MIMO-MAC where a statistic $\mathbf{S}$ of the channel realization $\mathbf{H}$ is available at the receiver. In order to provide some analytical considerations concerning the tightness of such bounds, in the following the behavior of the the gaps between the lower bounds and the corresponding upper bounds in (1.30) is analyzed.

Given a subset $\mathcal{A} \subset \mathcal{U}$ of the user's set $\mathcal{U}$, by Theorem 1, we have that $I(\mathbf{x}_{\mathcal{A}};\mathbf{y}|\mathbf{S},\mathbf{x}_{\overline{\mathcal{A}}})$ is lower and upper bounded by:

$$I_{\mathsf{L}(\mathcal{A},\overline{\mathcal{A}})} \leq I(\mathbf{x}_{\mathcal{A}};\mathbf{y}|\mathbf{S},\mathbf{x}_{\overline{\mathcal{A}}}) \leq I_{\mathsf{L}(\mathcal{A},\overline{\mathcal{A}})} + \boldsymbol{\Delta}_{\overline{\mathcal{A}}} \tag{1.30}$$

with

$$\boldsymbol{\Delta}_{\mathcal{A}} = \mathbb{E}\left[\log\frac{|\mathbf{I}_L + \mathsf{SNR}\,\Omega_{\mathcal{A}} + \mathsf{SNR}\,\Phi_{\overline{\mathcal{A}}}|}{|\mathbf{I}_L + \mathsf{SNR}\,\boldsymbol{\Phi}_{\mathcal{U}}|}\right] \tag{1.31}$$

Proposition 1 *Let $\mathcal{A}_1 \subset \mathcal{U}$ and $\mathcal{A}_2 \subset \mathcal{U}$ subsets of $\mathcal{U}$ such that:*

$$\mathcal{A}_1 \subseteq \mathcal{A}_2.$$

Then

$$\mathbf{\Delta}_{\mathcal{A}_1} \leq \mathbf{\Delta}_{\mathcal{A}_2} \leq \mathbf{\Delta}_{\mathcal{U}}. \tag{1.32}$$

Proof: *See Appendix A **Proof of Proposition 1**.*

This implies that $\mathbf{\Delta}_{\mathcal{A}}$ is an increasing function of the cardinality of the user subset $\mathcal{A}$, otherwise stated as the number of users, indexed in the subset $\mathcal{A}$, increases, the gap increases. From the above proposition, it follows the subsequent result:

Theorem 2 *The achievable rate region of a MIMO-MAC where a statistic* $\mathbf{S}$ *of the channel realization* $\mathbf{H}$ *is available at the receiver, can be sandwiched between two polytopes, between which the gap can be upper bounded by the following quantity:*

$$\mathbf{\Delta}_{\mathcal{U}} = \mathbb{E}\left[\log \frac{|\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}}|}{|\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Phi}_{\mathcal{U}}|}\right] \tag{1.33}$$

Note that $\mathbf{\Delta}_{\mathcal{U}}$, given in (1.33), represents also the gap between the upper and the lower bound of the sum rate (cf. Corollary 1). Then this quantity can be used to characterize with a single parameter, at any operating SNR, the gaps between the upper bounds and the lower bounds. Another interesting result, about the behavior of the gap to the sum-rate, is the following one:

Theorem 3 *As function of* SNR, $\mathbf{\Delta}_{\mathcal{U}}(\mathsf{SNR})$, *given in (1.33), is an increasing function.*

Proof: See Appendix A **Proof of Theorem 3**. Consequently, from Theorem 2 and Theorem 3, the gaps between the upper and the lower bounds can be uniformly upper bounded, on every range of signal to noise ratio considered, evaluating the gap to the sum-rate at the highest SNR. Moreover, under some mild conditions, like the assumption that the matrix $\mathbf{Z}$ is full rank with probability one, the gap converges to a finite value when the SNR diverges: this value can be used to give an uniform upper bound, with respect to SNR, to the function $\mathbf{\Delta}_{\mathcal{U}}(\mathsf{SNR})$.

The next two theorems study the asymptotic (in the sense of number of receiving antennas or number of users) behavior of the gap between the inner and the outer bound on the achievable rate region of a MIMO-MAC with a statistic $\mathbf{S}$ of the channel realization available at the receiver, assuming that $\mathbf{Z}$ is statistically independent from $\mathbf{S}$[7]. Consequently, an insight about the behavior of the system can be obtained. Specifically, we can prove the following results:

Theorem 4 *As the number of receiving antennas, L, grows, the gap between the inner and the outer bound on the achievable rate region of a MIMO-MAC when a statistic* $\mathbf{S}$*, of the channel realization* $\mathbf{H}$*, is available at the receiver, goes to zero if:*

- $\mathbb{E}[|\mathbf{x}_k|^4]$ *is finite* $\forall k \in 1, \ldots, K$*,*
- $\mathbf{Z}$ *has independent elements*
- $\sigma^2_{i,j} = o\left(\frac{1}{L^{\frac{1}{2}+\epsilon}}\right)$ *where* $\sigma^2_{i,j}$ *is the variance of* $[\mathbf{Z}]_{i,j}$*,*

Equivalently, as $L \to \infty$*,*

$$\mathbf{\Delta}_{\mathcal{U}} \overset{L\to\infty}{\to} 0 \tag{1.34}$$

where $\mathbf{\Delta}_{\mathcal{U}}$ *is given as in (1.33).*

Proof: See Appendix A **Proof of Theorem 4**.
Let us observe that the condition given on the power profile of the estimation errors does not imply that the overall error norm goes to zero, whereas it can go to infinity.

Example 1 *A scenario, in which the conditions of Theorem 4 are satisfied, is the case of channel estimation with full rank training. In such scenario, which will be described in the subsequent Subsection 2.2, if the* k*-th user has one transmitting antenna, and the fading coefficients between the transmitting antenna of* k*-th user and the receiving antennas are modeled as zero-mean Gaussian random variables with variance* $\sigma_k^2 = \frac{\phi_k}{L}$ *(to take in account for the energy conservation),* $\mathbf{Z}$ *is a gaussian random matrix whose entries are independent and have common variance along the columns. Specifically, denoting by* $(\sigma^2_{\mathbf{Z}})_k$ *the common variance of the entries of the* k*-th column of* $\mathbf{Z}$*, we have:*

$$(\sigma^2_{\mathbf{Z}})_k = \frac{1}{(\sigma_k^2)^{-1} + P_T} = \frac{1}{(\frac{\phi_k}{L})^{-1} + P_T} = \frac{\phi_k}{L + \phi_k P_T}, \tag{1.35}$$

[7] If the conditions that we have considered are satisfied for each $\mathbf{S}$, the results are easily extended

with P_T denoting the power assigned by the k-th user with $k = 1, \ldots, K$ to the training. Then, defining $C = \max_{k \in 1,\ldots,K} \phi_k$, we have that $\sigma^2_{i,j} \leq \frac{C}{L}$.

Let us now consider the asymptotic behavior of $\mathbf{\Delta}_{\mathcal{U}}$ when the number of users, K, increases.

Theorem 5 *As the number of active users[8], $\bar{K}$, increases, the gap between the inner and the outer bound on the achievable rate region of a MIMO-MAC when a statistic $\mathbf{S}$, of the channel realization $\mathbf{H}$, is available at the receiver, goes to zero if:*

- $\mathbb{E}[|\mathbf{x}_k|^4]$ *is uniformly bounded, i.e.* $\exists\ G_x > 0\ : \mathbb{E}[|\mathbf{x}_k|^4] < G_x\ \forall k \in 1, \ldots, K$,
- $\mathbf{Z}$ *is constituted by independent elements with uniformly bounded variances, i.e.* $\exists\ G_\sigma > 0\ :\ \sigma^2_{i,j} < G_\sigma\ \ \forall (i,j)$, *where* $\sigma^2_{i,j}$ *is the variance of* $[\mathbf{Z}]_{i,j}$,

Equivalently, as $\bar{K} \to \infty$,

$$\mathbf{\Delta}_{\mathcal{U}} \overset{\bar{K}\to\infty}{\to} 0 \tag{1.36}$$

where $\mathbf{\Delta}_{\mathcal{U}}$ is given as in (1.33).

Proof: *See Appendix A **Proof of Theorem 5**.*

It is important to underline that the upper bound for the sum-rate proved in Corollary1 for Gaussian distribution can be easily extended to zero mean input distributions. Then, resorting to Theorem 5, we can conclude that the Gaussian distributions are asymptotically optimal for all the input distributions with finite fourth order moments; this is due to the convergence of the upper bound to the lower bound, which moreover is independent from input distribution and is achievable for Gaussian ones. This consideration has a very simple interpretation: when the number of users is high, for the Central Limit Theorem, the random vectors $\frac{1}{\sqrt{K}}\mathbf{Zx}$ converge to a random gaussian vector, and then the channel behaves as a Gaussian MIMO MAC channel with perfect channel state information $\hat{\mathbf{H}}$ and with an increased noise level, for which the optimality of the Gaussian distribution is well known.

[8] Active users are those users whose transmitting power is strictly larger than zero.

1.2.2 Precoding Optimization

Till now, we assumed fixed covariance matrices $\mathbf{P}_k$ of the input distributions. However, such parameters represent a degree of freedom that we have in the design of the system: we can therefore carefully choose the precodings matrices $\mathbf{P}_k$ in order to optimize the performance of the network. The more natural choice is to optimize the sum-rate of the system, i.e. the global amount of information reliably delivered. A closed form expression of the optimal power allocation strategy for our MIMO-MAC is, unfortunately, not available. A possible alternative is to consider as figure of merit the lower bound on the sum rate, which is given by

$$I_{\mathsf{Lower}}(\mathbf{P})=\mathbb{E}\left[\log\left|\mathbf{I}_L+\mathsf{SNR}\ (\mathbf{I}_L+\mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}})^{-1}\hat{\mathbf{H}}\mathbf{P}_{\mathcal{U}}\hat{\mathbf{H}}^{\dagger}\right|\right], \tag{1.37}$$

with $\mathbf{\Omega}_{\mathcal{U}} = \mathbb{E}_{\mathbf{S}}\left[\mathbf{Z}\mathbf{P}_{\mathcal{U}}\mathbf{Z}^{\dagger}\right]$, $\hat{\mathbf{H}} = [\hat{\mathbf{H}}^{(1)}\ldots\hat{\mathbf{H}}^{(K)}]$ and $\hat{\mathbf{H}}^{(k)} = \mathbb{E}_{\mathbf{S}}[\mathbf{H}^{(k)}]$, and finding the power allocation strategy that maximize such a lower bound. Specifically we can prove the following results:

Theorem 6 *Assume that* $\mathbf{S}$ *and* $\mathbf{H}$ *are such that:*

$\mathcal{H}.1$ $\hat{\mathbf{H}}$ *is modeled according to (1.6), i.e:*

$$\hat{\mathbf{H}}^{(k)} = \mathbf{U}_{\mathrm{R}}{}^{(k)}\hat{\mathbf{H}}_w^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)^{\dagger}},\ \ k=1,\ldots,K$$

$\mathcal{H}.2$ $\mathbf{\Omega}_{\mathcal{U}}$ *doesn't depend on* $\mathbf{S}$ *and*

$$\tilde{\mathbf{Z}}^{(k)} = \mathbf{Z}^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)},\ \ k=1,\ldots,K$$

are random matrices with independent columns, each column having entries whose joint distribution is symmetric with respect to zero.

Then the eigenvectors of the input covariance matrix $\mathbf{P}_k^*$ *that maximizes (1.37) are given by the columns of* $\mathbf{U}_{\mathrm{T}}^{(k)}$.

Proof: *See Appendix A* ***Proof of Theorem 6***

Let $\mathbf{\Lambda}_k^*$ be the eigenvalue matrix of the input covariance matrix, $\mathbf{P}_k^*$, that maximizes (1.37). From the block diagonal structure of $\mathbf{P}_{\mathcal{U}}$, as defined in

(1.19), it follows that the eigenvalue matrix, $\mathbf{\Lambda}_{\mathcal{U}}^*$, associated to the matrix $\mathbf{P}_{\mathcal{U}}^* = \text{diag}\left(\mathbf{P}_1^*, \ldots, \mathbf{P}_K^*\right)$ is given by:

$$\mathbf{\Lambda}_{\mathcal{U}}^* = \text{diag}\left(\mathbf{\Lambda}_1^*, \ldots, \mathbf{\Lambda}_K^*\right) \tag{1.38}$$

We know that under the hypothesis of Theorem 6:

$$\mathbf{P}_k^* = \mathbf{U}_{\text{T}}^{(k)} \mathbf{\Lambda}_k^* \mathbf{U}_{\text{T}}^{(k)^\dagger}, \qquad k = 1, \ldots, K \tag{1.39}$$

The next theorem provides the structure of $\mathbf{\Lambda}_k^*$ with $k = 1, \ldots, K$.

Theorem 7 *Assume that:*

- $\mathcal{H}$ .1-$\mathcal{H}$.2 *of Theorem 6 hold,*
- *the joint distribution of the channel matrix* $\mathbf{H}^{(k)}$ *and of the statistic* $\mathbf{S}$ *is such that:*

$$\mathsf{SNR}\,\mathbb{E}\left[\mathbf{Z}\mathbf{U}_T^\dagger \mathbf{\Lambda}_{\mathcal{U}} \mathbf{U}_T \mathbf{Z}^\dagger\right] = \alpha(\mathbf{\Lambda}_{\mathcal{U}})\mathbf{I}_L \tag{1.40}$$

with $\mathbf{\Lambda}_{\mathcal{U}}$ *denoting the* MK*-diagonal matrix whose diagonal elements are the eigenvalues of* $\mathbf{P}_{\mathcal{U}}$ *as defined in (1.19),* $\mathbf{U}_T = $ *diag* $\left(\mathbf{U}_{\text{T}}^{(1)}, \ldots, \mathbf{U}_{\text{T}}^{(K)}\right)$ *and* $\alpha = \sum_{i=1}^{KM} a_i [\mathbf{\Lambda}_{\mathcal{U}}]_{i,i}$.

Then, the input covariance matrices $\mathbf{P}_k^*$*, with* $k = 1, \ldots, K$*, maximizing the lower bound on the sum rate, are given by:*

$$\mathbf{P}_k^* = \mathbf{U}_{\text{T}}^k \mathbf{\Lambda}_k^* \mathbf{U}_{\text{T}}^{k^\dagger} \qquad k = 1, \ldots, K \tag{1.41}$$

where $\mathbf{\Lambda}_k^* = $ *diag*$(\lambda_1^{k*}, \ldots, \lambda_M^{k*})$ *with* $k = 1, \ldots K$ *are diagonal matrices, whose diagonal elements are given by:*

$$\begin{bmatrix} \lambda_1^{1*} \\ \vdots \\ \lambda_M^{1*} \\ \vdots \\ \lambda_1^{K*} \\ \vdots \\ \lambda_M^{K*} \end{bmatrix} = \left(\mathbf{I}_{KM} - \mathbf{d}\mathbf{a}^\dagger\right)^{-1} \mathbf{d} = \frac{1}{1 - \mathbf{a}^\dagger \mathbf{d}} \mathbf{d} \tag{1.42}$$

where

$$\mathbf{d} = [\mathbf{d}_1^\dagger, \ldots, \mathbf{d}_K^\dagger]^\dagger$$

is the KM-dimensional vector solution of the following convex optimization problem:

$$\begin{aligned} \max_{\{\mathbf{d}_k\}} \ & \mathbb{E}\left[\log\left|\mathbf{I}_L + \mathsf{SNR}\sum_{k=1}^{K}\hat{\mathbf{H}}^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)} diag\left(\mathbf{d}_k\right)\mathbf{U}_{\mathrm{T}}^{(k)^\dagger}\hat{\mathbf{H}}^{(k)^\dagger}\right|\right] \\ \text{s.t.} \ : \ & \sum_{i=1}^{M} d_{k,i} \leq 1 - \mathbf{a}^\dagger\mathbf{d}, \ \ k = 1, \ldots, K. \end{aligned} \tag{1.43}$$

where in (1.43) for all $i = 1, \ldots, M$ and $k = 1, \ldots, K$, with $d_{k,i}$ we have denoted the i-th element of $\mathbf{d}_k$.

Proof: See Appendix A **Proof of Theorem 7**.
Let us underline the importance of the convexity property of the optimization problem (1.43); indeed, this implies that the auxiliary vector $\mathbf{d}$ can be efficiently obtained through the interior point algorithm and then the optimal power allocations $\mathbf{\Lambda}_k^*$ can be easily found.

1.3 Appendix A

Proof of Theorem 1

In order to prove Theorem 1, let us firstly prove the lower bound. To this end, initially, we assume that $\overline{\mathbf{P}}_F$ is positive definite and expand the mutual information in the following way:

$$I(\mathbf{x}_{\mathcal{F}};\mathbf{y}|\mathbf{S},\mathbf{x}_{\mathcal{C}}) = h(\mathbf{x}_{\mathcal{F}}|\mathbf{S},\mathbf{x}_{\mathcal{C}}) - h(\mathbf{x}_{\mathcal{F}}|\mathbf{y},\mathbf{S},\mathbf{x}_{\mathcal{C}}). \tag{1.44}$$

Assuming $\mathbf{x}_{\mathcal{F}|\mathbf{S},\mathbf{x}_{\mathcal{C}}}$ Gaussian distributed with covariance matrix $\overline{\mathbf{P}}_F$, we have:

$$h(\mathbf{x}_{\mathcal{F}}|\mathbf{S},\mathbf{x}_{\mathcal{C}}) = \log|\pi e \overline{\mathbf{P}}_{\mathcal{F}}|. \tag{1.45}$$

Recall here that:

$$\overline{\mathbf{P}}^{(k)} = \mathbb{E}[\mathbf{x}^{(k)}\mathbf{x}^{(k)\dagger}] = P\,\mathbf{P}^{(k)}.$$

Moreover, we can upper bound the second term using the following inequalities chain:

$$\begin{aligned} h(\mathbf{x}_{\mathcal{F}}|\mathbf{y},\mathbf{S},\mathbf{x}_{\mathcal{C}}) &= h(\mathbf{x}_{\mathcal{F}}|\mathbf{y},\mathbf{S},\mathbf{x}_{\mathcal{C}},\hat{\mathbf{x}}_{\mathcal{F}}(\mathbf{y},\mathbf{S},\mathbf{x}_{\mathcal{C}})) \\ &\leq h(\mathbf{x}_{\mathcal{F}}|\mathbf{S},\mathbf{x}_{\mathcal{C}},\hat{\mathbf{x}}_{\mathcal{F}}(\mathbf{y},\mathbf{S},\mathbf{x}_{\mathcal{C}})) \\ &\leq \mathbb{E}[h(\epsilon_{\mathbf{x}_{\mathcal{F}}}|\mathbf{S}=\bar{S},\mathbf{x}_{\mathcal{C}}=\bar{x}_{\mathcal{C}})] \end{aligned} \tag{1.46}$$

where $\hat{\mathbf{x}}_{\mathcal{F}}(\mathbf{y},\mathbf{S},\mathbf{x}_{\mathcal{C}})$ denotes an estimate of the input data vector $\mathbf{x}_{\mathcal{F}}$, based on $\mathbf{y}$, $\mathbf{S}$ and $\mathbf{x}_C$, while $\epsilon_{\mathbf{x}_{\mathcal{F}}} = \mathbf{x}_{\mathcal{F}} - \hat{\mathbf{x}}_{\mathcal{F}}(\mathbf{y},\mathbf{S},\mathbf{x}_{\mathcal{C}})$ denotes the corresponding estimation-error vector.

Letting $\hat{\mathbf{x}}_{\mathcal{F}}(\mathbf{y},\mathbf{S},\mathbf{x}_{\mathcal{C}})$ be the conditional linear MMSE (LMMSE) estimator, we have:

$$\hat{\mathbf{x}}_{\mathcal{F}} = \overline{\mathbf{P}}_{\mathcal{F}}\hat{\mathbf{H}}_{\mathcal{F}}^{\dagger}\Big(\hat{\mathbf{H}}_{\mathcal{F}}\overline{\mathbf{P}}_{\mathcal{F}}\hat{\mathbf{H}}_{\mathcal{F}}^{\dagger}+\mathbf{\Theta}_{\mathcal{F},\mathcal{C}}\Big)^{-1}\Big(\mathbf{y}-\hat{\mathbf{H}}_{\mathcal{C}}\mathbf{x}_{\mathcal{C}}\Big) \tag{1.47}$$

where

$$\mathbf{\Theta}_{\mathcal{F},\mathcal{C}} = P\mathbb{E}_{\mathbf{S}}[\mathbf{\Phi}_{\overline{\mathcal{C}}}] + P\mathbf{\Gamma}_{\overline{\mathcal{F}\cup\mathcal{C}}} + P\mathbf{\Phi}_{\mathcal{C}} + N_0\mathbf{I} \tag{1.48}$$

while $\hat{\mathbf{H}}_{(\cdot)}$, $\mathbf{\Phi}_{(\cdot)}$, and $\mathbf{\Gamma}_{(\cdot)}$ are defined as in Section 1.2.

After some simple algebraic manipulations, the error covariance matrix of the estimator, conditioned on $\mathbf{S}$ and $\mathbf{x}_\mathcal{C}$, in (1.47), can be expressed as, [28]:

$$\overline{\mathbf{P}}_\mathcal{F} - \overline{\mathbf{P}}_\mathcal{F}\widehat{\mathbf{H}}_\mathcal{F}^\dagger \left(\widehat{\mathbf{H}}_\mathcal{F}\overline{\mathbf{P}}_\mathcal{F}\widehat{\mathbf{H}}_\mathcal{F}^\dagger + \mathbf{\Theta}_{\mathcal{F},\mathcal{C}}\right)^{-1} \widehat{\mathbf{H}}_\mathcal{F}\overline{\mathbf{P}}_\mathcal{F}. \tag{1.49}$$

In order to provide a lower bound of (1.44), we will provide an upper bound to the conditional differential entropy of the error in (1.46). For this purpose, the entropy maximizing property of the Gaussian distribution with the same covariance matrix is exploited. Specifically, applying the inversion Lemma [9] to the covariance matrix in (1.49), and exploiting (1.45) we obtain, that for arbitrary $\mathcal{F} \subseteq \mathcal{U}$ and $\mathcal{C} \subseteq \mathcal{U}$:

$$\begin{aligned} I(\mathbf{x}_\mathcal{F};\mathbf{y}|\mathbf{S},\mathbf{x}_\mathcal{C}) &\geq I_{\mathsf{L}(\mathcal{F},\mathcal{C})} \\ &= \mathbb{E}\left[\log\left|\mathbf{I}_L + (\mathbf{\Theta}_{\mathcal{F},\mathcal{C}})^{-1}\widehat{\mathbf{H}}_\mathcal{F}\overline{\mathbf{P}}_\mathcal{F}\widehat{\mathbf{H}}_\mathcal{F}^\dagger\right|\right] \end{aligned} \tag{1.50}$$

Let us now observe that $I(\mathbf{x}_\mathcal{F};\mathbf{y}|\mathbf{S},\mathbf{x}_\mathcal{C})$ and the lower bound in (1.50), are continuous function of $\mathbf{P}_\mathcal{F}$, in the set of the block diagonal matrices non-negative definite, [29]. Moreover, it is well known that the closure of the set of the block diagonal matrices positive definite is the set of the block diagonal matrices non-negative definite; furthermore, we know that (1.50) holds in the set of the block diagonal matrices positive definite. Then from sign permanence Theorem, (1.50) still hold for all block diagonal matrices non-negative definite.

Let us now move on to the derivation of the upper bound for $I(\mathbf{x}_\mathcal{F};\mathbf{y}|\mathbf{S},\mathbf{x}_\mathcal{C})$. From the chain rule of the mutual information, we have that:

$$\begin{aligned} I(\mathbf{x}_\mathcal{F};\mathbf{y}|\mathbf{S},\mathbf{x}_\mathcal{C}) \;=\; & I(\mathbf{x}_{\overline{\mathcal{C}}};\mathbf{y}|\mathbf{S},\mathbf{x}_\mathcal{C}) \\ & - I(\mathbf{x}_{\overline{\mathcal{F}\cup\mathcal{C}}};\mathbf{y}|\mathbf{S},\mathbf{x}_{\mathcal{F}\cup\mathcal{C}}) \end{aligned} \tag{1.51}$$

From (1.51), an upper bound for $I(\mathbf{x}_S;\mathbf{y}|\mathbf{S},\mathbf{x}_\mathcal{C})$ can be obtained by upper bounding $I(\mathbf{x}_{\overline{\mathcal{C}}};\mathbf{y}|\mathbf{S},\mathbf{x}_\mathcal{C})$ and lower bounding $I(\mathbf{x}_{\overline{\mathcal{F}\cup\mathcal{C}}};\mathbf{y}|\mathbf{S},\mathbf{x}_{\mathcal{F}\cup\mathcal{C}})$. Thus, in summary, the derivation of an upper bound for (1.51), boils down to deriving an upper and lower bound for $I(\mathbf{x}_\mathcal{A};\mathbf{y}|\mathbf{S},\mathbf{x}_{\overline{\mathcal{A}}})$, where $\mathcal{A}$ is an arbitrary subset of $\mathcal{U}$. Concerning the lower bound, we can use (1.50) where $\mathcal{A} \equiv \mathcal{F}$ and

[9]We exploit in our derivation the fact that for any square matrix $\mathbf{A} = \mathbf{B} - \mathbf{B}\mathbf{C}^\dagger(\mathbf{C}\mathbf{B}\mathbf{C}^\dagger + \mathbf{D})^{-1}\mathbf{C}\mathbf{B}$, we may express its inverse as $\mathbf{A}^{-1} = \mathbf{B}^{-1} + \mathbf{C}^\dagger\mathbf{D}^{-1}\mathbf{C}$ and the well known relation $|\mathbf{A}|^{-1} = |\mathbf{A}^{-1}|$.

$\overline{\mathcal{A}} \equiv \mathcal{C}$. For the upper bound, we expand $I(\mathbf{x}_{\mathcal{A}};\mathbf{y}|\mathbf{S},\mathbf{x}_{\overline{\mathcal{A}}})$ as:

$$I(\mathbf{x}_{\mathcal{A}};\mathbf{y}|\mathbf{S},\mathbf{x}_{\overline{\mathcal{A}}}) = h(\mathbf{y}|\mathbf{S},\mathbf{x}_{\overline{\mathcal{A}}}) - h(\mathbf{y}|\mathbf{S},\mathbf{x}). \tag{1.52}$$

From (1.5) and (1.12), it is obtained that the covariance matrix of $\mathbf{y}$, conditioned on $\mathbf{S},\mathbf{x}_{\overline{\mathcal{A}}}$, is:

$$\mathbb{E}_{\mathbf{S},\mathbf{x}_{\overline{\mathcal{A}}}}[\mathbf{y}\mathbf{y}^{\dagger}] = \hat{\mathbf{H}}_{\mathcal{A}}\overline{\mathbf{P}}_{\mathcal{A}}\hat{\mathbf{H}}_{\mathcal{A}}^{\dagger} + P\mathbf{\Omega}_{\mathcal{A}} + P\mathbf{\Phi}_{\overline{\mathcal{A}}} + N_0\mathbf{I}_L.$$

Thus, the first term in right side of (1.52) is upper bounded by:

$$\begin{aligned} h(\mathbf{y}|\mathbf{S},\mathbf{x}_{\overline{\mathcal{A}}}) \leq \mathbb{E}\Big[\log\Big|\widehat{\mathbf{H}}_{\mathcal{A}}\overline{\mathbf{P}}_{\mathcal{A}}\widehat{\mathbf{H}}_{\mathcal{A}}^{\dagger} + P(\mathbf{\Omega}_{\mathcal{A}} + \mathbf{\Phi}_{\overline{\mathcal{A}}}) + N_0\mathbf{I}_L\Big|\Big] \\ + L\log(\pi e) \end{aligned} \tag{1.53}$$

where the expectation is over $\hat{\mathbf{H}}$ and $\mathbf{x}_{\overline{\mathcal{A}}}$. Furthermore, from (1.5) and from Hypothesis 2, $\mathbf{y}_{|\mathbf{S},\mathbf{x}}$ is distributed as Gaussian vector with covariance matrix $N_0\mathbf{I}_L + P\mathbf{\Phi}_{\mathcal{U}}$ and its differential entropy is given by:

$$h(\mathbf{y}|\mathbf{S},\mathbf{x}) = \mathbb{E}\left[\log\left|\pi e\left(N_0\mathbf{I}_L + P\mathbf{\Phi}_{\mathcal{U}}\right)\right|\right] \tag{1.54}$$

with $\mathbf{\Phi}_{\mathcal{U}}$ defined as in Section 2.2.1. Combining (1.52), (1.53) and (1.54), we obtain:

$$\begin{aligned} I(\mathbf{x}_{\mathcal{A}};\mathbf{y}|\mathbf{S},\mathbf{x}_{\overline{\mathcal{A}}}) &\leq I_{\mathsf{U}(\mathcal{A},\overline{\mathcal{A}})} \\ &= I_{\mathsf{L}(\mathcal{A},\overline{\mathcal{A}})} + \mathbf{\Delta}_{\mathcal{A}} \end{aligned} \tag{1.55}$$

where the gap, $\mathbf{\Delta}_{\mathcal{A}}$, between the lower and the upper bound and is defined as:

$$\mathbf{\Delta}_{\mathcal{A}} = \mathbb{E}\left[\log\frac{|N_0\mathbf{I}_L + P\Omega_{\mathcal{A}} + P\Phi_{\overline{\mathcal{A}}}|}{|N_0\mathbf{I}_L + P\mathbf{\Phi}_{\mathcal{U}}|}\right] \tag{1.56}$$

Replacing (1.50) and (1.55), in (1.51), we have:

$$\begin{aligned} I(\mathbf{x}_{\mathcal{F}};\mathbf{y}|\mathbf{S},\mathbf{x}_{\mathcal{C}}) &\leq I_{\mathsf{U}(\mathcal{F},\mathcal{C})} \\ &= I_{\mathsf{L}(\overline{\mathcal{C}},\mathcal{C})} - I_{\mathsf{L}(\overline{\mathcal{F}\cup\mathcal{C}},\mathcal{F}\cup\mathcal{C})} + \mathbf{\Delta}_{\overline{\mathcal{C}}}. \end{aligned}$$

■

Proof of Proposition 1 Let us consider two subsets $\mathcal{A}_1 \subseteq \mathcal{A}_2 \subset \mathcal{U}$. Using (1.31) we have:

$$\begin{aligned}\mathbf{\Delta}_{\mathcal{A}_1} &= \mathbb{E}\left[\log \frac{|\mathbf{I}_L + \mathsf{SNR}\,\Omega_{\mathcal{A}_1} + \mathsf{SNR}\,\Phi_{\overline{\mathcal{A}_1}}|}{|\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Phi}_{\mathcal{U}}|}\right] \\ &= \mathbb{E}\left[\log |\mathbf{I}_L + \mathsf{SNR}\,\Omega_{\mathcal{A}_1} + \mathsf{SNR}\,\Phi_{\overline{\mathcal{A}_1}}|\right] \\ &\quad - \mathbb{E}\left[\log |\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Phi}_{\mathcal{U}}|\right]\end{aligned} \tag{1.57}$$

Using Jensen's inequality, taking into account that $\overline{\mathcal{A}_1}$ can be decomposed as

$$\overline{\mathcal{A}_1} \equiv \overline{\mathcal{A}_1} - \overline{\mathcal{A}_2} \bigcup \overline{\mathcal{A}_2}$$

and the fact that input signals from different users are zero mean and independent, it is obtained that:

$$\mathbb{E}_{\mathbf{S},\overline{\mathcal{A}_2}}\left[\Phi_{\overline{\mathcal{A}_1}}\right] = \mathbb{E}_{\mathbf{S},\overline{\mathcal{A}_2}}\left[\Phi_{\overline{\mathcal{A}_1}-\overline{\mathcal{A}_2}}\right] + \Phi_{\overline{\mathcal{A}_2}}$$

from which it follows that:

$$\begin{aligned}\mathbf{\Delta}_{\mathcal{A}_1} &< \mathbb{E}\left[\log |\mathbf{I}_L + \mathsf{SNR}\,\Omega_{\mathcal{A}_1} + \mathsf{SNR}\,\mathbb{E}_{\mathbf{S},\overline{\mathcal{A}_2}}\left[\Phi_{\overline{\mathcal{A}_1}-\overline{\mathcal{A}_2}}\right] + \mathsf{SNR}\,\Phi_{\overline{\mathcal{A}_2}}|\right] \\ &\quad - \mathbb{E}\left[\log |\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Phi}_{\mathcal{U}}|\right] \\ &= \mathbb{E}\left[\log |\mathbf{I}_L + \mathsf{SNR}\,\Omega_{\mathcal{A}_2} + \mathsf{SNR}\,\Phi_{\overline{\mathcal{A}_2}}|\right] - \mathbb{E}\left[\log |\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Phi}_{\mathcal{U}}|\right] \\ &= \mathbf{\Delta}_{\mathcal{A}_2}\end{aligned} \tag{1.58}$$

where we resort to the following equalities:

$$\mathbb{E}_{\mathbf{S},\overline{\mathcal{A}_2}}\left[\Phi_{\overline{\mathcal{A}_1}-\overline{\mathcal{A}_2}}\right] = \Omega_{\mathcal{A}_2-\mathcal{A}_1} \tag{1.59}$$

$$\Omega_{\mathcal{A}_1} + \Omega_{\mathcal{A}_2-\mathcal{A}_1} = \Omega_{\mathcal{A}_2} \tag{1.60}$$

From (1.58) it follows that $\mathbf{\Delta}_{\mathcal{A}}$ is an increasing function of the cardinality of the user subset $\mathcal{A}$. Thus the maximum gap is obtained when $\mathcal{A} \equiv \mathcal{U}$.

■

Proof of Theorem 3 Let us evaluate the derivative $\boldsymbol{\Delta}_{\mathcal{U}}(\mathsf{SNR})$ with respect to SNR;

$$\frac{d}{d\mathsf{SNR}}\boldsymbol{\Delta}_{\mathcal{U}}(\mathsf{SNR}) = \mathbb{E}\left[\frac{d}{d\mathsf{SNR}}\log|\mathbf{I}_L + \mathsf{SNR}\,\boldsymbol{\Omega}_{\mathcal{U}}| - \mathbb{E}\left[\frac{d}{d\mathsf{SNR}}\log|\mathbf{I}_L + \mathsf{SNR}\,\boldsymbol{\Phi}_{\mathcal{U}}|\right]\right]$$

$$= \mathbb{E}\left[\sum_{i=1}^{N}\frac{\lambda_i}{1+\mathsf{SNR}\,\lambda_i} - \mathbb{E}\left[\sum_{i=1}^{N}\frac{\beta_i}{1+\mathsf{SNR}\,\beta_i}\right]\right] \tag{1.61}$$

$$\geq \mathbb{E}\left[\sum_{i=1}^{N}\frac{\lambda_i}{1+\mathsf{SNR}\,\lambda_i} - \sum_{i=1}^{N}\frac{\mathbb{E}\left[\beta_i\right]}{1+\mathsf{SNR}\,\mathbb{E}\left[\beta_i\right]}\right] \tag{1.62}$$

$$\geq 0 \tag{1.63}$$

where in (1.61), λ_i are the eigenvalues of $\boldsymbol{\Omega}_{\mathcal{U}}$ in increasing order, while β_i are the random eigenvalues of $\boldsymbol{\Phi}_{\mathcal{U}}$, respect to $\mathbf{x}$, again in increasing order; in (1.62) we have applied Jensen's inequality and finally in (1.63) we have used the Schur concavity of the function $\sum_{i=1}^{N}\frac{x_i}{1+\mathsf{SNR}\,x_i}$ and the fact that:

$$\mathbb{E}[\boldsymbol{\lambda}(\boldsymbol{\Phi}_{\mathcal{U}})] \preceq \boldsymbol{\lambda}(\mathbb{E}[\boldsymbol{\Phi}_{\mathcal{U}}]) \tag{1.64}$$

$$= \boldsymbol{\lambda}(\boldsymbol{\Omega}_{\mathcal{U}}), \tag{1.65}$$

where in the above inequalities, the expectations are with respect to the input vector $\mathbf{x}$, and $\boldsymbol{\lambda}(\mathbf{B})$ denotes the vector containing the eigenvalues of the matrix $\mathbf{B}$. Now in order to prove the majorization given in (1.64), we demonstrate the following lemma:

Lemma 1 *Given a positive semi-definite random matrix* $\mathbf{B}$ *of dimension* $N \times N$, *then* $E[\boldsymbol{\lambda}(\mathbf{B})] \preceq \boldsymbol{\lambda}(E[\mathbf{B}])$, *where the eigenvalues are taken in increasing order.*

Proof: We have to prove that for all $r = 1, \dots, N-1$:

$$\mathbb{E}\left[\sum_{i=1}^{r} \lambda_i(\mathbf{B})\right] \leq \sum_{i=1}^{r} \lambda_i(\mathbb{E}[\mathbf{B}]) \quad r = 1, \dots, N-1 \tag{1.66}$$

and

$$\mathbb{E}\left[\sum_{i=1}^{N} \lambda_i(\mathbf{B})\right] = \sum_{i=1}^{N} \lambda_i(\mathbb{E}[\mathbf{B}]) \tag{1.67}$$

To this end, we have that:

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^{r} \lambda_i(\mathbf{B})\right] &= \mathbb{E}\left[\min_{\mathbf{U} \in \mathbb{D}^r} \operatorname{tr}\{\mathbf{U}^\dagger \mathbf{B} \mathbf{U}\}\right] & (1.68) \\ &\leq \min_{\mathbf{U} \in \mathbb{D}^r} \mathbb{E}[\operatorname{tr}\{\mathbf{U}^\dagger \mathbf{B} \mathbf{U}\}] \\ &= \sum_{i=1}^{r} \lambda_i(\mathbb{E}[\mathbf{B}]) & (1.69) \end{aligned}$$

where in (1.68) we use [30, Corollary 4.3.18], while in (1.69) we use the concavity of the minimum of a function. Finally, for $r = N$, we have:

$$\begin{aligned} \mathbb{E}\left[\sum_{i=1}^{N} \lambda_i(\mathbf{B})\right] &= \mathbb{E}\left[\operatorname{tr}\{\mathbf{B}\}\right] \\ &= \operatorname{tr}\{\mathbb{E}[\mathbf{B}]\} & (1.70) \\ &= \sum_{i=1}^{N} \lambda_i(\mathbb{E}[\mathbf{B}]) & (1.71) \end{aligned}$$

■

Proof of Theorem 4 In order to prove Theorem 4, let us recall here that:

$$\forall x \geq 0: \qquad x - \frac{x^2}{2} \leq \ln(1+x) \leq x \tag{1.72}$$

From Theorem 2, we have that the gap between the inner and outer bound on the rate region can be upper bounded by

$$\boldsymbol{\Delta}_{\mathcal{U}} = \mathbb{E}\left[\log \frac{|\mathbf{I}_L + \mathsf{SNR}\,\boldsymbol{\Omega}_{\mathcal{U}}|}{|\mathbf{I}_L + \mathsf{SNR}\,\boldsymbol{\Phi}_{\mathcal{U}}|}\right] \tag{1.73}$$

where:

$$\boldsymbol{\Omega}_{\mathcal{U}} = \frac{1}{P}\mathbb{E}\left[\mathbf{Z}\mathbf{x}\mathbf{x}^{\dagger}\mathbf{Z}^{\dagger}\right],$$

and

$$\boldsymbol{\Phi}_{\mathcal{U}} = \frac{1}{P}\mathbb{E}_{\mathbf{x}}\left[\mathbf{Z}\mathbf{x}\mathbf{x}^{\dagger}\mathbf{Z}^{\dagger}\right]$$

Particularizing this expression to our scenario, due to the assumption that $\mathbf{Z}$ has independent elements, we have that $\boldsymbol{\Omega}_{\mathcal{U}}$ and $\boldsymbol{\Phi}_{\mathcal{U}}$ are diagonal matrices with diagonal entries given by:

$$[\boldsymbol{\Omega}_{\mathcal{U}}]_{i,i} = \sum_{k=1}^{K}\sum_{j=1}^{M}\sigma_{i,k+j}^2[\mathbf{P}_k]_{j,j}$$

and

$$[\boldsymbol{\Phi}_{\mathcal{U}}]_{i,i} = \sum_{k=1}^{K}\sum_{j=1}^{M}\sigma_{i,k+j}^2\frac{|\mathbf{x}_{k+j}|^2}{P}.$$

Therefore, substituting these expressions in (1.73), we obtain:

$$\begin{aligned}\boldsymbol{\Delta} \;=\; & \sum_{i=1}^{L}\left(\log\left(1+\mathsf{SNR}\sum_{k=1}^{K}\sum_{j=1}^{M}\sigma_{i,k+j}^2[\mathbf{P}_k]_{j,j}\right)\right.\\ & \left. - \,\mathbb{E}\left[\log\left(1+\mathsf{SNR}\sum_{k=1}^{K}\sum_{j=1}^{M}\sigma_{i,k+j}^2\frac{|\mathbf{x}_{k+j}|^2}{P}\right)\right]\right)\end{aligned} \tag{1.74}$$

Applying (1.72) to (1.74), we have:

$$
\begin{aligned}
0 \le \boldsymbol{\Delta} \le & \sum_{i=1}^{L} \mathsf{SNR} \sum_{k=1}^{K} \sum_{j=1}^{M} \sigma_{i,k+j}^2 [\mathbf{P}_k]_{j,j} \\
& - \sum_{i=1}^{L} \mathbb{E}\left[\mathsf{SNR} \sum_{k=1}^{K} \sum_{j=1}^{M} \sigma_{i,k+j}^2 \frac{|\mathbf{x}_{k+j}|^2}{P} \right] \\
& + \sum_{i=1}^{L} \mathbb{E}\left[\frac{\left(\mathsf{SNR} \sum_{k=1}^{K} \sum_{j=1}^{M} \sigma_{i,k+j}^2 \frac{|\mathbf{x}_{k+j}|^2}{P} \right)^2}{2} \right] \\
\le & L\mathbb{E}\left[\frac{\left(\mathsf{SNR} \sum_{k=1}^{K} \sum_{j=1}^{M} \frac{C}{L^{\frac{1}{2}+\epsilon}} \frac{|\mathbf{x}_{k+j}|^2}{P} \right)^2}{2} \right] \\
\le & L^{-\epsilon} \frac{\mathsf{SNR}^2 C^2}{2} \sum_{k=1}^{K} \mathbb{E}\left[||\mathbf{x}_k||^4 \right] \\
\overset{L\to\infty}{\to} & \; 0
\end{aligned} \tag{1.75}
$$

where (1.75) follows from the fact that the fourth moment of the transmitted signal is finite.

■

Proof of Theorem 5 From Theorem 2, we have that gap between the inner and the outer bound on the rate region can be upper bounded by

$$
\boldsymbol{\Delta}_{\mathcal{U}} = \mathbb{E}\left[\log \frac{|\mathbf{I}_L + \mathsf{SNR}\, \boldsymbol{\Omega}_{\mathcal{U}}|}{|\mathbf{I}_L + \mathsf{SNR}\, \boldsymbol{\Phi}_{\mathcal{U}}|} \right] \tag{1.76}
$$

where:

$$
\boldsymbol{\Omega}_{\mathcal{U}} = \frac{1}{P} \mathbb{E}\left[\mathbf{Z}\mathbf{x}\mathbf{x}^\dagger \mathbf{Z}^\dagger \right],
$$

and

$$
\boldsymbol{\Phi}_{\mathcal{U}} = \frac{1}{P} \mathbb{E}_{\mathbf{x}}\left[\mathbf{Z}\mathbf{x}\mathbf{x}^\dagger \mathbf{Z}^\dagger \right]
$$

As in the proof of Theorem 4, due to the assumption that $\mathbf{Z}$ has independent elements, we have that $\boldsymbol{\Omega}_{\mathcal{U}}$ and $\boldsymbol{\Phi}_{\mathcal{U}}$ are diagonal matrices with diagonal entries given by:

$$[\mathbf{\Omega}_{\mathcal{U}}]_{i,i} = \sum_{k=1}^{\bar{K}} \sum_{j=1}^{M} \sigma_{i,k+j}^2 [\mathbf{P}_k]_{j,j}$$

and

$$[\mathbf{\Phi}_{\mathcal{U}}]_{i,i} = \sum_{k=1}^{\bar{K}} \sum_{j=1}^{M} \sigma_{i,k+j}^2 \frac{|\mathbf{x}_{k+j}|^2}{P}.$$

Therefore, substituting these expressions in (1.76), we obtain:

$$\begin{aligned} -\mathbf{\Delta} &= \sum_{i=1}^{L} \mathbb{E}\left[\log\left(\frac{1+\mathsf{SNR}\sum_{k=1}^{\bar{K}}\sum_{j=1}^{M}\sigma_{i,k+j}^2\frac{|\mathbf{x}_{k+j}|^2}{P}}{1+\mathsf{SNR}\sum_{k=1}^{\bar{K}}\sum_{j=1}^{M}\sigma_{i,k+j}^2[\mathbf{P}_k]_{j,j}}\right)\right]. \\ &= \sum_{i=1}^{L} \mathbb{E}\left[\log\left(1+\mathbf{y}_{\bar{K}}^i\right)\right] \end{aligned} \tag{1.77}$$

where

$$\mathbf{y}_{\bar{K}}^i = \frac{\frac{1}{\bar{K}}\mathsf{SNR}\sum_{k=1}^{\bar{K}}\sum_{j=1}^{M}\sigma_{i,k+j}^2\left(\frac{|\mathbf{x}_{k+j}|^2}{P}-[\mathbf{P}_k]_{j,j}\right)}{\frac{1}{\bar{K}}+\frac{1}{\bar{K}}\mathsf{SNR}\sum_{k=1}^{\bar{K}}\sum_{j=1}^{M}\sigma_{i,k+j}^2[\mathbf{P}_k]_{j,j}}$$

is a zero-mean random variable (i.e $\mathbb{E}\left[\mathbf{y}_{\bar{K}}^i\right]=0$) with variance

$$\mathbb{E}\left[(\mathbf{y}_{\bar{K}}^i)^2\right] = \frac{\frac{1}{\bar{K}^2}\mathsf{SNR}^2\sum_{k=1}^{\bar{K}}\mathbb{E}\left[\left(\sum_{j=1}^{M}\sigma_{1,k+j}^2\left(\frac{|\mathbf{x}_{k+j}|^2}{P}-[\mathbf{P}_k]_{j,j}\right)\right)^2\right]}{\left(\frac{1}{\bar{K}}+\frac{1}{\bar{K}}\mathsf{SNR}\sum_{k=1}^{\bar{K}}\sum_{j=1}^{M}\sigma_{1,k+j}^2[\mathbf{P}_k]_{j,j}\right)^2}.$$

Now notice that, using the Cauchy Schwartz inequality, $\mathbb{E}\left[(\mathbf{y}_{\bar{K}}^i)^2\right]$ can be upper bounded as follows:

$$\mathbb{E}\left[(\mathbf{y}_{\bar{K}}^i)^2\right] \leq \frac{\frac{1}{\bar{K}^2}\mathsf{SNR}^2 M \sum_{k=1}^{\bar{K}} \beta_k}{\left(\frac{1}{\bar{K}} + \frac{1}{\bar{K}}\mathsf{SNR}\,\sigma_{\mathsf{min}}^2 \sum_{k=1}^{\bar{K}} \mathrm{tr}\{\mathbf{P}_k\}\right)^2} \tag{1.78}$$

$$\leq \frac{\frac{1}{\bar{K}}\mathsf{SNR}^2 M\beta}{\left(\frac{1}{\bar{K}} + \mathsf{SNR}\,\sigma_{\mathsf{min}}^2 P_{\mathsf{min}}\right)^2} \tag{1.79}$$

where in (1.78)

$$\beta_k = \sum_{j=1}^{M} \mathbb{E}\left[\sigma_{1,k+j}^4 \left(\frac{|\mathbf{x}_{k+j}|^2}{P} - [\mathbf{P}_k]_{j,j}\right)^2\right],$$

while in (1.79) $\beta = \max_k\{\beta_k\}$, σ_{min}^2 denotes the minimum over all strictly positive values of $\sigma_{1,k+j}^2$, and $P_{\mathsf{min}} = \min_k \mathrm{tr}\{\mathbf{P}_k\}$ which is strictly positive since we are assuming $\bar{K} > 0$ active users. Recalling that $\sigma_{1,k+j}^2 < G_\sigma$ and $\mathbb{E}[|\mathbf{x}_k|^4] < G_x \; \forall k \in 1, \ldots, \bar{K}$ and $j = 1, \ldots, M$, we have that the $\{\beta_k\}$ are uniformly bounded, from which it follows that $\mathbf{y}_k^i$ converges in mean square sense to zero, equivalently:

$$\mathbb{E}\left[\mathbf{y}_{\bar{K}}^i\right] \to 0 \quad \text{m.s.} \tag{1.80}$$

Let us now lower bound (1.77); to this end, we need to lower bound the function $\log(1+x)$. Let us consider the function:

$$g(x) = \begin{cases} 1 - \frac{1}{1-\delta} & -1 < x < -\delta \\ 1 - \frac{1}{1+x} & x \geq -\delta \end{cases}$$

where δ is an arbitrary real number such that $0 < \delta < 1$. Then we have that

$$\log(1+x) \geq \begin{cases} \log(1+x) + g(x) & -1 < x < -\delta \\ g(x) & x \geq -\delta \end{cases} .$$

Let us observe that each random variable $\mathbf{y}^i_{\bar{K}} \geq a^i_{\bar{K}}$, where

$$-1 < a^i_{\bar{K}} = \frac{-\frac{1}{\bar{K}}\mathsf{SNR}\sum_{k=1}^{\bar{K}}\sum_{j=1}^{M}\sigma^2_{i,k+j}[\mathbf{P}_k]_{j,j}}{\frac{1}{\bar{K}} + \frac{1}{\bar{K}}\mathsf{SNR}\sum_{k=1}^{\bar{K}}\sum_{j=1}^{M}\sigma^2_{i,k+j}[\mathbf{P}_k]_{j,j}} < 0.$$

Thus we have

$$\begin{aligned} 0 \geq \mathbb{E}\left[\log\left(1+\mathbf{y}^i_{\bar{K}}\right)\right] &\geq \mathbb{E}\left[g\left(\mathbf{y}^i_{\bar{K}}\right)\right] + \int_{a^i_{\bar{K}}}^{-\delta} \log\left(1+\mathbf{y}^i_{\bar{K}}\right) d\mu_{\mathbf{y}^i_{\bar{K}}} \\ &\geq \mathbb{E}\left[g\left(\mathbf{y}^i_{\bar{K}}\right)\right] + \log\left(1+a^i_{\bar{K}}\right) P(|\mathbf{y}^i_{\bar{K}}| > \delta) \\ &\geq \mathbb{E}\left[g\left(\mathbf{y}^i_{\bar{K}}\right)\right] + \log\left(1+a^i_{\bar{K}}\right) \frac{\sigma^2_{\bar{K}}}{\delta^2} \end{aligned} \quad (1.81)$$

where $\sigma^2_{\bar{K}} = \frac{\frac{1}{\bar{K}}\mathsf{SNR}^2 M\beta}{\left(\frac{1}{\bar{K}} + \mathsf{SNR}\,\sigma^2_{\mathsf{min}}P_{\mathsf{min}}\right)^2}$.
Since $g(x)$ is a continuous and bounded function, we have

$$\mathbb{E}\left[g\left(\mathbf{y}^i_{\bar{K}}\right)\right] \to 0 \quad \forall i$$

(characterization of the convergence in distribution); moreover

$$\log\left(1+a^i_{\bar{K}}\right)\frac{\sigma^2_{\bar{K}}}{\delta^2} \to 0 \quad \forall i \text{ and } \forall \delta,$$

since σ^2_k goes to zero with $\bar{K}$, like $\frac{1}{\bar{K}}$, and $x\log(x) \to 0$ when $x \to 0$. Consequently, as $\bar{K} \to \infty$ we have that $\mathbf{\Delta}$ goes to zero.

■

Proof of Theorem 6 Starting from (1.37), we have that:

$$I_{\mathsf{Lower}}(\mathbf{P}) = \mathbb{E}\left[\log\left|\mathbf{I}_L + \mathsf{SNR}\,\mathbf{Q}\hat{\mathbf{H}}\mathbf{P}\hat{\mathbf{H}}^{\dagger}\right|\right] \tag{1.82}$$

$$= \mathbb{E}\left[\log\left|\mathbf{I}_L + \mathsf{SNR}\,\mathbf{Q}\sum_{k=1}^{K}\hat{\mathbf{H}}^{(k)}\mathbf{P}_k\hat{\mathbf{H}}^{(k)^{\dagger}}\right|\right] \tag{1.83}$$

$$= \mathbb{E}\left[\log\left|\mathbf{I}_L + \mathsf{SNR}\,\mathbf{Q}\sum_{k=1}^{K}\mathbf{U}_{\mathrm{R}}^{(k)}\hat{\mathbf{H}}_w^{(k)}\tilde{\mathbf{P}}_k\hat{\mathbf{H}}_w^{(k)^{\dagger}}\mathbf{U}_{\mathrm{R}}^{(k)\dagger}\right|\right] \tag{1.84}$$

where in (1.82)

$$\begin{aligned}\mathbf{Q} &= (\mathbf{I}_L + \mathsf{SNR}\,\boldsymbol{\Omega}_{\mathcal{U}})^{-1}\\ &= \left(\mathbf{I}_L + \mathsf{SNR}\,\mathbb{E}\left[\mathbf{Z}\mathbf{P}\mathbf{Z}^{\dagger}\right]\right)^{-1}\\ &= \left(\mathbf{I}_L + \mathsf{SNR}\,\mathbb{E}\left[\tilde{\mathbf{Z}}\tilde{\mathbf{P}}\tilde{\mathbf{Z}}^{\dagger}\right]\right)^{-1},\end{aligned} \tag{1.85}$$

in (1.83) we have used the fact that $\hat{\mathbf{H}}^{(k)}$ can be written as

$$\hat{\mathbf{H}}^{(k)} = \mathbf{U}_{\mathrm{R}}^{(k)}\hat{\mathbf{H}}_w^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)^{\dagger}}$$

and finally in (1.84) and in (1.85) we have denoted $\tilde{\mathbf{P}}_k = \mathbf{U}_{\mathrm{T}}^{(k)^{\dagger}}\mathbf{P}_k\mathbf{U}_{\mathrm{T}}^{(k)}$ and $\tilde{\mathbf{Z}}^{(k)} = \mathbf{Z}\mathbf{U}_{\mathrm{T}}^{(k)}$.

We want to show that nonzero off-diagonal entries in $\tilde{\mathbf{P}}_k$ can only reduce $I_{\mathsf{Lower}}(\mathbf{P})$. To reach this aim, we define $\boldsymbol{\Pi}_j$ as a diagonal matrix all of whose diagonal entries are 1 except for the (j,j)-th entry, which is -1. The entries of $\boldsymbol{\Pi}_j\tilde{\mathbf{P}}_k\boldsymbol{\Pi}_j^{\dagger}$ equal those of $\tilde{\mathbf{P}}_k$ except for the off-diagonals in the j-th row and j-th column, whose sign is reversed. Also we have $\mathrm{tr}\{\mathbf{P}_k\} = \mathrm{tr}\{\boldsymbol{\Pi}_j\tilde{\mathbf{P}}_k\boldsymbol{\Pi}_j^{\dagger}\}$ and

$$I_{\mathsf{Lower}}(\boldsymbol{\Pi}_j\tilde{\mathbf{P}}_k\boldsymbol{\Pi}_j^{\dagger}) = I_{\mathsf{Lower}}(\tilde{\mathbf{P}}_k) \tag{1.86}$$

where in (1.86) we have used the fact that, since the columns of $\hat{\mathbf{H}}_w^{(k)}$ and of $\tilde{\mathbf{Z}}$ are independent with a symmetric distribution, reversing the sign of the j-th column does not alter the distribution. The matrix $\frac{1}{2}\left(\boldsymbol{\Pi}_j\tilde{\mathbf{P}}_k\boldsymbol{\Pi}_j^{\dagger} + \tilde{\mathbf{P}}_k\right)$ has the same entries of $\tilde{\mathbf{P}}_k$ except for the off-diagonal in the $\jmath$-th row and j-th column, which are zero. Invoking Jensen's inequality:

$$I_{\mathsf{Lower}}\left(\frac{1}{2}\left(\mathbf{\Pi}_j\tilde{\mathbf{P}}_k\mathbf{\Pi}_j^\dagger+\tilde{\mathbf{P}}_k\right)\right)\geq\frac{I_{\mathsf{Lower}}(\mathbf{\Pi}_j\tilde{\mathbf{P}}_k\mathbf{\Pi}_j^\dagger)+I_{\mathsf{Lower}}(\tilde{\mathbf{P}}_k)}{2}$$
$$=I_{\mathsf{Lower}}(\tilde{\mathbf{P}}_k)$$

Hence, nullifying the off-diagonal entries of any column and corresponding row of $\tilde{\mathbf{P}}_k$ can only lead to an increase of $I_{\mathsf{Lower}}(\tilde{\mathbf{P}}_k)$. Repeating the same process KM times, we find that (1.37) is maximized when indeed $\tilde{\mathbf{P}}_k$ is diagonal for all $k = 1, \ldots, K$.

■

Proof of Theorem 7 Since the hypothesis of Theorem 6 are satisfied we know that the input covariance matrix, $\mathbf{P}_k^*$, that maximizes (1.37) is given by:

$$\mathbf{P}_k^* = \mathbf{U}_{\mathrm{T}}^{(k)}\mathbf{\Lambda}_k^*\mathbf{U}_{\mathrm{T}}^{(k)^\dagger}, \qquad k = 1, \ldots, K \tag{1.87}$$

with $\mathbf{\Lambda}_k^*$ denoting the eigenvalue matrix of $\mathbf{P}_k^*$. Thus, letting:

$$\tilde{\mathbf{H}} = [\hat{\mathbf{H}}^{(k)}\mathbf{U}_{\mathrm{T}}^{(1)}, \ldots, \hat{\mathbf{H}}^{(k)}\mathbf{U}_{\mathrm{T}}^{(K)}]$$

and

$$\mathbf{\Lambda}_{\mathcal{U}}^* = \mathrm{diag}\,(\mathbf{\Lambda}_1^*, \ldots, \mathbf{\Lambda}_K^*)$$

with $\mathbf{\Lambda}_{\mathcal{U}}^*$ denoting the eigenvalue matrix of $\mathbf{P}_{\mathcal{U}}^* = \mathrm{diag}\,(\mathbf{P}_1^*, \ldots, \mathbf{P}_K^*)$, the problem of maximizing (1.37) boils down to maximizing

$$\mathbb{E}\left[\log\left|\mathbf{I}_L + \frac{\mathsf{SNR}}{1+\alpha(\mathbf{\Lambda}_{\mathcal{U}}^*)}\sum_{k=1}^{K}\hat{\mathbf{H}}^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)}\mathbf{\Lambda}_k^*\mathbf{U}_{\mathrm{T}}^{(k)^\dagger}\hat{\mathbf{H}}^{(k)^\dagger}\right|\right], \tag{1.88}$$

over all possible positive semi-definite diagonal matrices $\{\mathbf{\Lambda}_k^*\}_{k=1}^K$ such that $\mathrm{tr}\{\mathbf{\Lambda}_k^*\} \leq 1$; notice that, in (1.88), we assume that the joint distribution of the channel matrix $\mathbf{H}^{(k)}$ and of the statistic $\mathbf{S}$ is such that:

$$\mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}} = \mathsf{SNR}\,\mathbb{E}\left[\mathbf{Z}\mathbf{P}_{\mathcal{U}}^*\mathbf{Z}^\dagger\right] \tag{1.89}$$
$$= \alpha(\mathbf{\Lambda}_{\mathcal{U}}^*)\mathbf{I}_L \tag{1.90}$$

with $\alpha = \sum_{i=1}^{KM} a_i[\mathbf{\Lambda}_{\mathcal{U}}^*]_{i,i}$.

Next, define $\mathbf{D}$ as the following KM-dimensional diagonal matrix:

$$\mathbf{D} = \frac{1}{1+\alpha(\mathbf{\Lambda}_{\mathcal{U}}^*)}\mathbf{\Lambda}_{\mathcal{U}}^*. \tag{1.91}$$

Then, (1.88) can be rewritten as:

$$\begin{aligned}
&\mathbb{E}\left[\log\left|\mathbf{I}_L + \frac{\mathsf{SNR}}{1+\alpha(\mathbf{\Lambda}_{\mathcal{U}}^*)}\sum_{k=1}^{K}\hat{\mathbf{H}}^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)}\mathbf{\Lambda}_k^*\mathbf{U}_{\mathrm{T}}^{(k)^\dagger}\hat{\mathbf{H}}^{(k)^\dagger}\right|\right] \\
&= \mathbb{E}\left[\log\left|\mathbf{I}_L + \frac{\mathsf{SNR}}{1+\alpha(\mathbf{\Lambda}_{\mathcal{U}}^*)}\tilde{\mathbf{H}}\mathbf{\Lambda}_{\mathcal{U}}^*\tilde{\mathbf{H}}^\dagger\right|\right] \\
&= \mathbb{E}\left[\log\left|\mathbf{I}_L + \mathsf{SNR}\,\tilde{\mathbf{H}}\mathbf{D}\hat{\mathbf{H}}^\dagger\right|\right]
\end{aligned} \tag{1.92}$$

Denoting by λ^* the vector of the diagonal elements of $\mathbf{\Lambda}_{\mathcal{U}}^* = \mathrm{diag}\left(\mathbf{\Lambda}_1^*, \ldots, \mathbf{\Lambda}_K^*\right)$:

$$\lambda^* = [\lambda_1^{1*}, \ldots, \lambda_M^{1*}, \ldots, \lambda_1^{k*}, \ldots, \lambda_M^{k*}, \ldots, \lambda_1^{K*} \ldots, \lambda_M^{K*}]^T$$

and by $\mathbf{d}$ the KM-dimensional vector:

$$\mathbf{d} = [d_{1,1}, \ldots, d_{1,M}, \ldots, d_{k,1}, \ldots, d_{k,M}, \ldots, d_{K,1} \ldots, d_{K,M}]^T,$$

from (1.91) it follows that:

$$\begin{aligned}
\boldsymbol{\lambda}^* &= \mathbf{d}\left(1+\alpha(\mathbf{\Lambda}_{\mathcal{U}}^*)\right) && (1.93)\\
&= \mathbf{d}\left(1+\sum_{i=1}^{KM} a_i \lambda^{\lceil \frac{i}{M}\rceil *}_{\mathrm{mod}[i,M]+1}\right) && (1.94)\\
&= \mathbf{d}\left(1+\mathbf{a}^T\boldsymbol{\lambda}^*\right) && (1.95)
\end{aligned}$$

where $\mathbf{a}$ is the KM-dimensional vector $\mathbf{a} = [a_1, \ldots, a_{KM}]^T$. From (1.95), after some algebraic manipulations , it follows that $\boldsymbol{\lambda}^*$ and $\mathbf{d}$ are related by:

$$\left(1-\mathbf{a}^T\mathbf{d}\right)\boldsymbol{\lambda}^* = \mathbf{d} \tag{1.96}$$

which is equivalent to:

$$\boldsymbol{\lambda}^* = \frac{1}{1-\mathbf{a}^T\mathbf{d}}\mathbf{d}. \tag{1.97}$$

Using (1.97), we can rewrite the power constraints, $\mathrm{tr}\{\mathbf{\Lambda}_k^*\} \leq 1$ in terms of elements of $\mathbf{d}$. Specifically, if $\boldsymbol{\lambda}$ satisfies (1.97), then the power constraints $\mathrm{tr}\{\mathbf{\Lambda}_k^*\} \leq 1$ with $k = 1, \ldots, K$ induce the following constraints on the elements of $\mathbf{d}$: for all $k = 1, \ldots, K$,

$$\sum_{i=1}^{M} d_{k,i} \leq 1 - \mathbf{a}^\dagger \mathbf{d}.$$

From the foregoing considerations and from (1.92), it follows that maximizing (1.88) over all possible positive semi-definite diagonal matrices $\mathbf{\Lambda}_{\mathcal{U}}^* = \mathrm{diag}(\lambda^*)$ such that $\mathrm{tr}\{\mathbf{\Lambda}_k^*\} \leq 1$, is tantamount to:

- solving the following optimization problem:

$$\mathbf{D}^* = \mathrm{diag}(\mathbf{d}^*) = \arg\max_{\mathbf{D}} \mathbb{E}\left[\log\left|\mathbf{I}_L + \mathsf{SNR}\,\tilde{\mathbf{H}}\mathbf{D}\hat{\mathbf{H}}^\dagger\right|\right]$$

  over all possible positive semi-definite diagonal matrices $\mathbf{D} = \mathrm{diag}(\mathbf{d}) = \mathrm{diag}([d_{1,1}, \ldots, d_{k,i}, \ldots d_{K,M}]^T)$ such that their diagonal entries $\{d_{k,i}\}$:

$$\sum_{i=1}^{M} d_{k,i} \leq 1 - \mathbf{a}^\dagger \mathbf{d}, \qquad \text{for all} \quad k = 1, \ldots, K,$$

- evaluating λ^* via (1.97) i.e:

$$\boldsymbol{\lambda}^* = \frac{1}{1 - \mathbf{a}^T \mathbf{d}^*} \mathbf{d}^*.$$

■

Chapter 2

Applications of MIMO MAC with Partial CSI

In chapter 1, a MIMO MAC system, in which an arbitrary statistic $\mathbf{S}$ of the channel realization $\mathbf{H}$ is available at the receiver, has been describer under an Information-Theoretic point of view. In this chapter possible setups, where such analysis is of interest, are described. In particular, two main scenarios where the perfect coherent reception of signals is totally unrealistic are considered; in this situation it is opportune to quantify the impact of partial channel state information. The first analyzed case concerns a cooperative MIMO network in which the received signals at several base stations are collected and jointly decoded at a CP, assuming that only a partial CSI is available to the CP. The second scenario, for which the analysis conducted in chapter 1 is applied, regards a MIMO MAC channel in which the channel knowledge at the receiver is obtained through training signals.

2.1 Application to Cooperative MIMO Networks

Coordinating the reception and transmission of signals across spatially distributed base stations has been shown to improve sum-rate performance by mitigating the effects of intercell interference in MIMO cellular networks, [18], [19]. Concerning the uplink, full network coordination can be interpreted as having a single base station receiver with spatially distributed antennas across the network, yielding an instance of a MIMO MAC.

More precisely, let us consider a cellular network with N base stations, each of them equipped with L receiving antennas, and serving in total K users, each of them equipped with M transmitting antennas. The received signal at the j-th base station is given by:

$$\mathbf{y}_j = \sum_{k=1}^{K} \mathbf{H}_j^{(k)} \mathbf{x}^{(k)} + \mathbf{n}^{(j)} \tag{2.1}$$

$$= \mathbf{H}_j \mathbf{x} + \mathbf{n}^{(j)} \qquad j = 1, \ldots, N, \tag{2.2}$$

where $\mathbf{H}_j^{(k)}$ is the $L \times M$ channel matrix between k-th user and j-th base, $\mathbf{x}^{(k)}$ is the M-dimensional input vector for the k-th user and $\mathbf{n}^{(j)}$ is the additive symmetric zero-mean Gaussian noise. In (2.2), $\mathbf{x} = [\mathbf{x}^{(1)^\dagger}, \ldots, \mathbf{x}^{(K)\dagger}]^\dagger$ is the MK-dimensional stacked input vector and $\mathbf{H}_j = \left[\mathbf{H}_j^{(1)}, \mathbf{H}_j^{(2)}, \ldots, \mathbf{H}_j^{(K)}\right]$, for all $j = 1, \ldots N$, is an $L \times MK$ channel matrix, statistically independent across j, whose blocks $\mathbf{H}_j^{(k)}$ are statistically independent across k. We suppose that each channel $\mathbf{H}_j^{(k)}$ is described by the UIU model [24], i.e.,

$$\mathbf{H}_j^{(k)} = \mathbf{U}_{\mathrm{R}}^{(k,j)} \mathbf{H}_w^{(k,j)} \mathbf{U}_{\mathrm{T}}^{(k,j)^\dagger}, \quad k = 1, \ldots, K, \quad j = 1, \ldots, N, \tag{2.3}$$

where $\mathbf{U}_{\mathrm{R}}^{(k,j)}$ is a $L \times L$ unitary matrix, $\mathbf{U}_{\mathrm{T}}^{(k,j)}$ is a $M \times M$ unitary matrix, and $\mathbf{H}_w^{(k,j)}$ is a zero-mean $L \times M$ Gaussian random matrix of independent elements with power profile $\mathbf{\Sigma}_w^{(k,j)}$, i.e. $E\left[|[\mathbf{H}_w^{(k,j)}]_{i,j}|^2\right] = [\mathbf{\Sigma}_w^{(k,j)}]_{i,j}$.

By stacking the N received signal vectors in (2.2) to form an LN-dimensional vector, $\mathbf{y} = [\mathbf{y}^{(1)^\dagger} \ldots \mathbf{y}^{(N)^\dagger}]^\dagger$, we can write

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2.4}$$

where $\mathbf{H}$ is the $NL \times MK$ channel matrix whose (j, k)-th block is the matrix $\mathbf{H}_j^{(k)}$, $\mathbf{x}$ is the MK-dimensional stacked input vector, and $\mathbf{n}$ is the NL-dimensional stacked noise vector.

In the subsequent subsections we analyze different types of statistic of the channels available at the CP. In particular we consider two main scenarios characterized by different quality of the side information, related to $\mathbf{H}$, that can be made available at the CP. In both cases the data vectors $\mathbf{y}_j$ are assumed as perfectly known to the CP.

2.1.1 Cooperative MIMO Networks with Incomplete CSI and infinite-capacity backhaul to a CP

In this subsection we analyze the performance limits of a cooperative uplink MIMO network in which N base stations send their received signals to a CP via backhaul links, for joint decoding. We assume that the backhaul links have infinite capacity. For an ideal joint decoding, the CP requires full CSI (amplitude and phase) for the link from every user to every base. In practice, achieving full CSI in a large network is very challenging because the low signal-to-noise ratio on the link between distant bases and users would prevent accurate channel estimation. Starting from these considerations we assume, denoting $\mathcal{U}$ the set of all users in the network, for all $j = 1, \ldots, N$, that the j-th base station has perfect knowledge of the channel matrices of only a given subset of users, which we denote by $\mathcal{C}_j \subseteq \mathcal{U}$, while for the remaining users $\overline{\mathcal{C}}_j = \mathcal{U} - \mathcal{C}_j$, only statistical channel state information (SCSI) is available at the j-th base. In order to define the subset $\mathcal{C}_j$ of users whose channel matrices are perfectly known at the j-th base station, we consider a user-based constraint: full CSI (phase and amplitude) for a given user is known at $Q \leq N$ base stations and each user independently chooses (based on a given criterion) its own set of Q base stations. We refer to Q as degree of cooperation [17]. The joint decoding process takes place at the CP wherein, in addition to the statistical CSI of the channel from all users, for all $j = 1, \ldots, N$, unquantized versions of the received data vectors $\mathbf{y}_j$ and the channels $\mathbf{H}_j^{(k)}$ with $k \in \mathcal{C}_j \subseteq \mathcal{U}$ are made available, thanks to the assumption of infinite-capacity links between each base station and the CP, that can be used for both the information that have to be processed.

Under this framework, the statistic of the channel $\mathbf{H}$ available at the CP is given by:

$$\mathbf{S} = \{\mathbf{H}_j^{(k)} : \quad j = 1, \ldots, N, \quad \text{and} \quad k \in \mathcal{C}_j\} \tag{2.5}$$

Based on the aforementioned consideration, for each $j = 1, \ldots, N$, we define the a $N \times K$ matrix $\bar{\mathbf{E}}$ whose (j,k)-th element, $[\bar{\mathbf{E}}]_{j,k}$ is 1 if if the channel $\mathbf{H}_j^{(k)}$ is known fully at the j-th base and 0 otherwise, which is equivalent to saying that for all $j = 1, \ldots, N$:

$$[\bar{\mathbf{E}}]_{j,k} = \begin{cases} 1 & \text{if } k \in \mathcal{C}_j, \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

Note that each column of $\bar{\mathbf{E}}$ has only Q ones and the other entries are all zeros.

Let

$$\mathbf{H}^{(k)} = [\mathbf{H}_1^{(k)\dagger}, \ldots, \mathbf{H}_N^{(k)\dagger}]^\dagger$$

denote the channel matrix between the k-th user and the N base stations. Then, using the notation introduced before, the set of channels, $\mathbf{H}^{(k\,\mathsf{c})}$, characterizing the links between the k-th user and its own set of Q base stations can be represented as:

$$\mathbf{H}^{(k\,\mathsf{c})} = \mathbf{H}^{(k)} \odot \left(\bar{\mathbf{E}}_k \otimes \mathbf{1}_L \mathbf{1}_M^\dagger\right), \tag{2.7}$$

where $\bar{\mathbf{E}}_k$ denotes the k-th column of $\bar{\mathbf{E}}$. Consequently, $\mathbf{H}^{(k)}$ can be re-written as

$$\mathbf{H}^{(k)} = \mathbf{H}^{(k\,\mathsf{c})} + \mathbf{H}^{(k\,\mathsf{i})},$$

with

$$\mathbf{H}^{(k\,\mathsf{i})} = \mathbf{H}^{(k)} \odot \left(\mathbf{1}_{LN} \mathbf{1}_M^\dagger - \left(\bar{\mathbf{E}}_k \otimes \mathbf{1}_L \mathbf{1}_M^\dagger\right)\right), \tag{2.8}$$

where c and i respectively denote *complete* and *incomplete* channel knowledge.

Furthermore, we denote by $\mathbf{E}$ the $LN \times MK$ matrix $\mathbf{E} = \bar{\mathbf{E}} \otimes \mathbf{1}_L \mathbf{1}_M^\dagger$. Using the foregoing definition and by introducing

$$\mathbf{H}^{\mathsf{c}} = \mathbf{H} \odot \mathbf{E}, \tag{2.9}$$

$$\mathbf{H}^{\mathsf{i}} = \mathbf{H} \odot (\mathbf{1}_{LN} \mathbf{1}_{MK}^\dagger - \mathbf{E}), \tag{2.10}$$

where again c and i respectively denote *complete* and *incomplete* channel knowledge, we can rewrite the channel matrix $\mathbf{H}$ in a compact form as

$$\mathbf{H} = \mathbf{H}^{\mathsf{c}} + \mathbf{H}^{\mathsf{i}}. \tag{2.11}$$

Since the matrices $\mathbf{H}^{(k\,\mathsf{c})}$ and $\mathbf{H}^{(k\,\mathsf{i})}$ are statistically independent, and by assumption independent across $k = 1, \ldots, K$, we have that $\mathbf{H}$, $\mathbf{S}$, $\mathbb{E}_{\mathbf{S}}[\mathbf{H}]$ and $\mathbf{Z}$ as defined in (1.12) satisfy **Hypothesis 2** given in Section 1.2. For this particular setting, we have

$$\hat{\mathbf{H}} = \mathbb{E}_{\mathbf{S}}[\mathbf{H}] \tag{2.12}$$

$$= [\mathbf{H}^{(1\,\mathsf{c})}, \ldots, \mathbf{H}^{(k\,\mathsf{c})}, \ldots, \mathbf{H}^{(K\,\mathsf{c})}] \tag{2.13}$$

$$= \mathbf{H}^{\mathsf{c}} \tag{2.14}$$

and

$$\begin{aligned}\mathbf{Z} &= \mathbf{H}_{|\mathbf{S}} - \mathbb{E}_{\mathbf{S}}[\mathbf{H}] &(2.15)\\ &= [\mathbf{H}^{(1\,\mathsf{i})}, \ldots, \mathbf{H}^{(k\,\mathsf{i})}, \ldots, \mathbf{H}^{(K\,\mathsf{i})}] &(2.16)\\ &= \mathbf{H}^{\mathsf{i}}. &(2.17)\end{aligned}$$

Particularizing Theorem 1 to this scenario, we obtain that for any disjoint user subsets $\mathcal{F} \subseteq \mathcal{U}$ and $\mathcal{C} \subseteq \mathcal{U}$,

$$I_{\mathsf{L}(\mathcal{F},\mathcal{C})} \leq I(\mathbf{x}_{\mathcal{F}}; \mathbf{y}|\mathbf{S}, \mathbf{x}_{\mathcal{C}}) \leq I_{\mathsf{L}(\overline{\mathcal{C}},\mathcal{C})} - I_{\mathsf{L}(\overline{\mathcal{F}\cup\mathcal{C}},\mathcal{F}\cup\mathcal{C})} + \mathbf{\Delta}_{\overline{\mathcal{C}}} \tag{2.18}$$

with

$$I_{\mathsf{L}(\mathcal{F},\mathcal{C})} = \mathbb{E}\left[\log\left|\mathbf{I}_{LN} + \left(\mathbf{\Omega}_{\overline{\mathcal{C}}} + \mathbf{\Gamma}_{\overline{\mathcal{F}\cup\mathcal{C}}} + \mathbf{\Phi}_{\mathcal{C}} + \tfrac{1}{\mathsf{snr}}\mathbf{I}_{LN}\right)^{-1}\mathbf{\Gamma}_{\mathcal{F}}\right|\right],$$

and

$$\mathbf{\Delta}_{\overline{\mathcal{C}}} = \mathbb{E}\left[\log\frac{|\mathbf{I}_{LN} + \mathsf{SNR}\,\mathbf{\Omega}_{\overline{\mathcal{C}}} + \mathsf{SNR}\,\mathbf{\Phi}_{\mathcal{C}}|}{|\mathbf{I}_{LN} + \mathsf{SNR}\,\mathbf{\Phi}_{\mathcal{U}}|}\right] \tag{2.19}$$

where for any arbitrary user subset $\mathcal{R} \subseteq \mathcal{U}$,

$$\mathbf{\Gamma}_{\mathcal{R}} = \mathbf{H}^{\mathsf{c}}_{\mathcal{R}}\mathbf{P}_{\mathcal{R}}\mathbf{H}^{\mathsf{c}\,\dagger}_{\mathcal{R}},$$

and

$$\mathbf{\Phi}_{\mathcal{R}} = \frac{1}{P}\mathbb{E}_{\mathbf{x}_{\mathcal{R}}}\left[\mathbf{H}^{i}_{\mathcal{R}}\mathbf{x}_{\mathcal{R}}\mathbf{x}^{\dagger}_{\mathcal{R}}\mathbf{H}^{i\,\dagger}_{\mathcal{R}}\right],$$

while $\mathbf{\Omega}_{\mathcal{R}} = \mathbb{E}[\mathbf{\Phi}_{\mathcal{R}}]$.

Computing explicitly the above expectations we have that: $\mathbf{\Omega}_{\overline{\mathcal{B}}}$ is a $NL \times NL$ block diagonal matrix with j-th block of dimension $L \times L$, given by:

$$\sum_{k \in \overline{\mathcal{B}} \cap \overline{\mathcal{C}}_j} \mathbf{U}_{\mathrm{R}}^{(k,j)}\mathbf{A}^{(k,j)}\mathbf{U}_{\mathrm{R}}^{(k,j)\,\dagger}, \quad \text{with} \quad j = 1, \ldots, N,$$

where $\mathbf{A}^{(k,j)}$ is a diagonal matrix, whose (i,i) entry is given by:

$$[\mathbf{A}^{(k,j)}]_{i,i} = \mathrm{tr}\left\{\mathbf{U}_{\mathrm{T}}^{(k,j)}\mathbf{P}_k\mathbf{U}_{\mathrm{T}}^{(k,j)\,\dagger}\mathbf{D}_i^{(k,j)}\right\} \tag{2.20}$$

with

$$\mathbf{D}_i^{(k,j)} = \text{diag}\left([\mathbf{\Sigma}_w^{(k,j)}]_{i,1}, \dots [\mathbf{\Sigma}_w^{(k,j)}]_{i,M}]\right)$$

Similarly, $\mathbf{\Phi}_{\mathcal{R}}$ is a $NL \times NL$ block diagonal matrix with j-th block, of dimension $L \times L$, given by:

$$\sum_{k \in \mathcal{R} \cap \overline{\mathcal{C}}_j} \mathbf{U}_{\text{R}}^{(k,j)} \mathbf{B}^{(k,j)} \mathbf{U}_{\text{R}}^{(k,j)\dagger} \qquad \text{with} \quad j = 1, \dots, N, \tag{2.21}$$

where $\mathbf{B}^{(k,j)}$ is a diagonal matrix, whose (i,i) entry is given by:

$$[\mathbf{B}^{(k,j)}]_{i,i} = \frac{1}{P} \text{tr} \left\{ \mathbf{U}_{\text{T}}^{(k,j)} \mathbf{x}_k \mathbf{x}_k^\dagger \mathbf{U}_{\text{T}}^{(k,j)\dagger} \mathbf{D}_i^{(k,j)} \right\}. \tag{2.22}$$

Let us, now, present a case of study to obtain insights about the behavior of the system. We consider a cluster of $N = 3$ base stations, as illustrated in Fig. 2.1, each at the center of an ideal hexagon, serving $K = 2$ users per cell at any one time and frequency resource. We assume that the users are placed uniformly inside each hexagon, of normalized ray $R = 1$, at a normalized distance greater than 0.1, from the center.

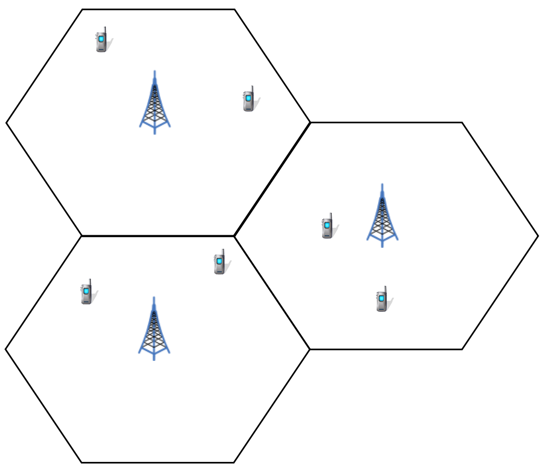

Figure 2.1: Cluster of $N = 3$ hexagonal cells.

Each base station is equipped with two receiving antennas, $L = 2$. Furthermore each user has $M = 1$ transmitting antenna. The channel matrices $\mathbf{H}_j^{(k)}$ are characterized according to (2.3):

$$\mathbf{H}_j^{(k)} = \mathbf{U}_{\mathrm{R}}^{(k,j)} \mathbf{H}_w^{(k,j)} \mathbf{U}_{\mathrm{T}}^{(k,j)^\dagger} \qquad (2.23)$$

$$= \sqrt{d_{j,k}^{-\gamma}} \mathbf{U}_{\mathrm{R}}^{(k,j)} \mathbf{\Lambda}_{\mathrm{R}}^{(k,j)} \mathbf{w}^{(k,j)} \qquad (2.24)$$

where $\mathbf{U}_{\mathrm{T}}^{(k,j)^\dagger}=1$ since $M=1$, $\mathbf{H}_w^{(k,j)}=\sqrt{d_{j,k}^{-\gamma}}\mathbf{\Lambda}_{\mathrm{R}}^{(k,j)}\mathbf{w}^{(k,j)}$, and $\mathbf{w}^{(k,j)}$ is an i.i.d. L-dimensional zero-mean Gaussian vector with unit variance entries. In (2.24), $\mathbf{U}_{\mathrm{R}}^{(k,j)}$ and $\mathbf{\Lambda}_{\mathrm{R}}^{(k,j)}$ count for the correlation at the receiving antennas, while $\sqrt{d_{j,k}^{-\gamma}}$ counts for the path loss attenuation of the signal with $d_{j,k}$ denoting the distance between the k-th user and the j-th base station. Specifically, denoting by $\mathbf{\Sigma}_{\mathrm{R}}^{j,k} = \mathbf{U}_{\mathrm{R}}^{(k,j)} \Lambda_{\mathrm{R}}^{(k,j)} \mathbf{U}_{\mathrm{R}}^{(k,j)}$ the $L \times L$ receive correlation matrix, we assume:

$$[\mathbf{\Sigma}_{\mathrm{R}}^{k}]_{l,l'} = \frac{1}{2} e^{-0.5 d_{\mathrm{R}}^2 |l-l'|} e^{i 2\pi f_0 / c \cos(\theta_{\mathrm{R}}^k)(l-l')} \qquad (2.25)$$

where f_0 is the carrier frequency, c is the speed of light, d_{R} is the receiving antennas spacing of the Uniform Linear Array (ULA) and θ_{R}^k is the angle between the direction defined by the receiving ULA of the base station and direction passing through the base station position and the k-th user position. For this example, we assume that each user selects the Q nearest BSs, taking in account for the distance-based pathloss. Specifically, for each user index $k \in \{1, \dots, K\}$, we can always find Q base-stations indices $(j_1, \dots, j_Q)$ such that:

- $k \in \bigcap_{\ell=1}^{Q} \mathcal{C}_{j_\ell}$,
- $j_\ell = \arg \min_{n \in \{1,\dots N\} - \{j_p\}_{p=1}^{\ell-1}} d_{n,k}$.

In Fig. 2.2 we analyze the average sum-rate of the system, with respect to the random position of the users, in terms of its lower and upper bound as given in (2.18), for two possible degree of cooperations $Q = 1$ and $Q = 2$.

We see that the lower and upper bounds are very tight in both case and thus they well describe the performance limits of the system. As expected, increasing the degree of cooperation produces an improvement of the performance. Moreover, in the case $Q = 1$, the system is interference limited since the matrix $\mathbf{\Omega}_{\mathcal{U}}$ is positive definite, for all user positions, while in the case $Q = 2$ the

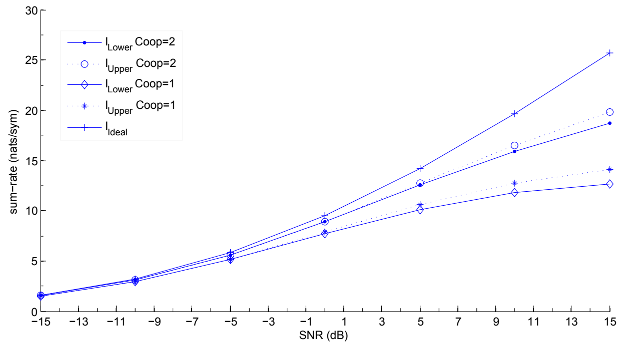

Figure 2.2: Lower and Upper bound of the sum-rate of a cooperative MIMO Network with incomplete CSI and infinite-capacity backhaul to a CP, with degree of cooperations $Q = 1$ and $Q = 2$.

sum-rate continues to increase, since, with a non-negligible probability, at least one base station knows all the user channels. Finally, note that the degree of Cooperation needed to obtain good performance compared with the ideal case, has to increase with the amount of power used to send information, underling the increasing importance of the quality of the side information, with respect to the operating SNR .

2.1.2 Cooperative MIMO Networks with Imperfect CSI and infinite-capacity backhaul to a CP

In this subsection, we extend the analysis of a cooperative MIMO network with degree of cooperation Q, as described subsection 2.1.1, to the case where only a quantized version of the channel matrix $\mathbf{H}^{\mathsf{c}} = [\mathbf{H}^{(1\,\mathsf{c})}, \ldots, \mathbf{H}^{(K\,\mathsf{c})}]$, with $\mathbf{H}^{(k\,\mathsf{c})}$ defined as in (2.7), is available at the CP. Denoting by $\mathbf{S}$ the statistic of the channel realizations available at the CP and using the same notation introduced in Subsection 2.1.1, we have that

$$\mathbf{S} \;=\; [\mathbf{S}^{(1)}, \ldots, \mathbf{S}^{(K)}], \tag{2.26}$$

where $\mathbf{S}^{(k)} = \mathbf{H}_{\mathsf{q}}^{(k\,\mathsf{c})}$, with $\mathbf{H}_{\mathsf{q}}^{(k\,\mathsf{c})}$ denoting the quantized version of the matrix $\mathbf{H}^{(k\,\mathsf{c})}$. This scenario can be a meaningful model for a time-division-duplex (TDD) MIMO network where each user perfectly estimates his channels with the own set of Q selected base stations, using powerful training signals transmitted by those base stations, and then sends back to them a quantized version of such channels, using system control channels [1].

Since the users are spatially separated, we have to consider the further constraint of distributed encoding: the k-th user, after perfectly estimating $\mathbf{H}^{(k\,\mathsf{c})}$, quantizes it independently from the other users, and consequently with independent quantization errors. We assume that each user compresses $\mathbf{H}^{(k\,\mathsf{c})}$ to within a prescribed distortion level D_k, determined by the available feedback bandwidth and the rate-distortion characteristics of the channel.

Given this setup, considering the optimal quantizer for Gaussian random vectors, we have that:

$$\hat{\mathbf{H}} = [\mathbf{H}_{\mathsf{q}}^{(1\,\mathsf{c})}, \ldots, \mathbf{H}_{\mathsf{q}}^{(K\,\mathsf{c})}] \tag{2.27}$$

and

$$\begin{aligned} \mathbf{Z} \;&=\; \mathbf{H}_{|\mathbf{S}} - \mathbb{E}_{\mathbf{S}}[\mathbf{H}] && (2.28)\\ &=\; [\mathbf{Z}^{(1)}, \ldots \mathbf{Z}^{(K)}] && (2.29) \end{aligned}$$

[1] A second setting, where the above scenario can be a meaningful model, is a setup where the base stations perfectly estimate their channels and forward a compressed version of the channels to the CP through a lossless link of finite capacity.

where

$$\mathbf{Z}^{(k)} = \mathbf{H}^{(k)}_{|\mathbf{S}^{(k)}} - \mathbb{E}_{\mathbf{S}^{(k)}}[\mathbf{H}^{(k)}] \quad (2.30)$$

$$= \mathbf{H}^{(k)}_{|\mathbf{H}_{\mathsf{q}}^{(k\,\mathsf{c})}} - \mathbf{H}_{\mathsf{q}}^{(k\,\mathsf{c})} \quad (2.31)$$

$$= \mathbf{H}^{(k\,\mathsf{i})} + \mathbf{Z}_{\mathsf{q}}^{(k)}. \quad (2.32)$$

In (2.30), $\mathbf{Z}_{\mathsf{q}}^{(k)}$ is a zero-mean Gaussian random matrix independent across $k = 1, \ldots, K$ and such that $\mathrm{vec}\{\mathbf{Z}_{\mathsf{q}}^{(k)}\} \in \mathbb{C}^{LNM}$ is an LMN-dimensional zero-mean Gaussian vector whose covariance matrix, $\mathbf{D}^{(k)*}$, is given by the solution of the following optimization problem, [31], [32]:

$$\min_{\mathbf{D}^{(k)}} \left(\log \left|\mathbf{\Omega}^{(k)}\right|^{+} - \log \left|\mathbf{D}^{(k)}\right|^{+} \right) \quad (2.33)$$

$$\text{s.t. :} \qquad 0 \preceq \mathbf{D}^{(k)} \preceq \mathbf{\Omega}^{(k)}, \mathrm{tr}\{\mathbf{D}\} \leq D^{(k)}$$

where $D^{(k)}$ is such that $\mathbb{E}\left[\|\mathbf{H}^{(k\,\mathsf{c})} - \mathbf{H}_{\mathsf{q}}^{(k\,\mathsf{c})}\|^2\right] \leq D_k$, while $\mathbf{\Omega}^{(k)}$ denotes the covariance matrix of LMN-dimensional random vector $\mathrm{vec}\{\mathbf{H}^{(k\,\mathsf{c})}\} \in \mathbb{C}^{LNM}$. From the solution of the optimization problem given in (2.33), it follows that $\mathbf{D}^{(k)*} = \mathbf{U}_{\Omega^{(k)}} \Lambda_{\mathbf{D}}^{(k)^*} \mathbf{U}^{\dagger}_{\Omega^{(k)}}$ where $\mathbf{U}_{\Omega^{(k)}}$ is the eigenvector matrix of $\mathbf{\Omega}^{(k)}$ while $\Lambda_{\mathbf{D}}^{(k)^*}$ is the solution of the following reverse water-filling

$$[\Lambda_{\mathbf{D}}^{(k)^*}]_{i,i} = \begin{cases} \eta & \text{if} \quad \eta < [\Lambda_{\mathbf{\Omega}^{(k)}}]_{i,i} \\ [\Lambda_{\mathbf{\Omega}^{(k)}}]_{i,i} & \text{if} \quad \eta \geq [\Lambda_{\mathbf{\Omega}^{(k)}}]_{i,i} \end{cases} \quad (2.34)$$

with η such that $\mathrm{tr}\{\Lambda_{\mathbf{D}}^{(k)}\} \leq D^{(k)}$.

Since $\mathbf{Z}_{\mathsf{q}}^{(k)}$ and $\mathbf{H}^{(k\,\mathsf{i})}$ are Gaussian random matrices statistical independent, independent across $k = 1, \ldots, K$ and independent from $\mathbf{S}$, then **Hypothesis 2** holds. Particularizing Theorem 1 to this scenario we have that: for any disjoint user subsets $\mathcal{F} \subseteq \mathcal{U}$ and $\mathcal{C} \subseteq \mathcal{U}$,

$$I_{\mathsf{L}(\mathcal{F},\mathcal{C})} \leq I(\mathbf{x}_{\mathcal{F}}; \mathbf{y}|\mathbf{S}, \mathbf{x}_{\mathcal{C}}) \leq I_{\mathsf{L}(\overline{\mathcal{C}},\mathcal{C})} - I_{\mathsf{L}(\overline{\mathcal{F}\cup\mathcal{C}},\mathcal{F}\cup\mathcal{C})} + \mathbf{\Delta}_{\overline{\mathcal{C}}} \quad (2.35)$$

with

$$I_{\mathsf{L}(\mathcal{F},\mathcal{C})} = \mathbb{E}\left[\log\left|\mathbf{I}_{LN} + \left(\mathbf{\Omega}_{\overline{\mathcal{C}}} + \mathbf{\Gamma}_{\overline{\mathcal{F}\cup\mathcal{C}}} + \mathbf{\Phi}_{\mathcal{C}} + \tfrac{1}{\mathsf{snr}}\mathbf{I}_{LN}\right)^{-1}\mathbf{\Gamma}_{\mathcal{F}}\right|\right],$$

and

$$\mathbf{\Delta}_{\overline{\mathcal{C}}} = \mathbb{E}\left[\log\frac{|\mathbf{I}_{LN} + \mathsf{SNR}\,\mathbf{\Omega}_{\overline{\mathcal{C}}} + \mathsf{SNR}\,\mathbf{\Phi}_{\mathcal{C}}|}{|\mathbf{I}_{LN} + \mathsf{SNR}\,\mathbf{\Phi}_{\mathcal{U}}|}\right] \tag{2.36}$$

where for any arbitrary user subset $\mathcal{R} \subseteq \mathcal{U}$,

$$\mathbf{\Gamma}_{\mathcal{R}} = \mathbf{Q}_{\mathcal{R}}\mathbf{P}_{\mathcal{R}}\mathbf{Q}_{\mathcal{R}^\dagger},$$

and

$$\mathbf{\Phi}_{\mathcal{R}} = \frac{1}{P}\mathbb{E}_{\mathbf{x}_{\mathcal{R}}}\left[\mathbf{H}^{\mathsf{i}}_{\mathcal{R}}\mathbf{x}_{\mathcal{R}}\mathbf{x}^\dagger_{\mathcal{R}}\mathbf{H}^{i\,\dagger}_{\mathcal{R}}\right] + \frac{1}{P}\mathbb{E}_{\mathbf{x}_{\mathcal{R}}}\left[\mathbf{E}_{\mathcal{R}}\mathbf{x}_{\mathcal{R}}\mathbf{x}^\dagger_{\mathcal{R}}\mathbf{E}_{\mathcal{R}}{}^\dagger\right].$$

with $\mathbf{\Omega}_{\mathcal{R}} = \mathbb{E}[\mathbf{\Phi}_{\mathcal{R}}]$, while $\mathbf{Q}_{\mathcal{R}}$ and $\mathbf{E}_{\mathcal{R}}$ represent matrices formed, respectively, from $\mathbf{H}^{\mathsf{c}}_{\mathsf{q}} = \left[\mathbf{H}^{(1\,\mathsf{c})}_{\mathsf{q}}, \ldots, \mathbf{H}^{(K\,\mathsf{c})}_{\mathsf{q}}\right]$ and $\mathbf{Z}_q = \left[\mathbf{Z}^{(1)}_q, \ldots, \mathbf{Z}^{(K)}_q\right]$, by deleting the blocks not indexed by the elements of $\mathcal{R}$.

As in Section 2.1.1, $\frac{1}{P}\mathbb{E}_{\mathbf{x}_{\mathcal{R}}}\left[\mathbf{H}^{\mathsf{i}}_{\mathcal{R}}\mathbf{x}_{\mathcal{R}}\mathbf{x}^\dagger_{\mathcal{R}}\mathbf{H}^{i\,\dagger}_{\mathcal{R}}\right]$ is a $NL \times NL$ block diagonal matrix with j-th block, of dimension $L \times L$, given by:

$$\sum_{k\in\mathcal{R}\cap\overline{\mathcal{C}}_j} \mathbf{U}^{(k,j)}_{\mathsf{R}}\mathbf{B}^{(k,j)}\mathbf{U}^{(k,j)\dagger}_{\mathsf{R}} \qquad \text{with} \quad j = 1, \ldots, N, \tag{2.37}$$

where $\mathbf{B}^{(k,j)}$ is a diagonal matrix, whose (i,i) entry is given in (2.22).

On the other hand, $\frac{1}{P}\mathbb{E}_{\mathbf{x}_{\mathcal{R}}}\left[\mathbf{E}_{\mathcal{R}}\mathbf{x}_{\mathcal{R}}\mathbf{x}^\dagger_{\mathcal{R}}\mathbf{E}_{\mathcal{R}}{}^\dagger\right]$ is a $NL \times NL$ matrix which depends on the set of matrices $\{\mathbf{D}^{(k)}\}_{k\in\mathcal{R}}$.

As case of study, we consider exactly the same scenario, the same propagation model and the same proximity-based cooperation protocol considered in Subsection 2.1.1. However, in this case we assume that the CP has only a quantized version of the channel matrix $\mathbf{H}^{\mathsf{c}}$. Specifically, we assume that each user perfectly estimates his channels with respect to his own set of Q base stations and then forwards a quantized version of the estimate to the base stations, which send them and the received data signals to the CP via infinite-capacity links.

For this scenario, we study the performance of the Cooperative MIMO

Network as function of the quantization rate dedicated by each user in forwarding the compressed version of his channels. We assume that all the users set a given common feedback rate $R(D)$ in order to send the quantized channel coefficients.

In Fig. 2.3, we illustrate the mean lower and upper bounds, with respect to the random position of the users, of the sum rate for a set of quantization rate given by $R(D) = \{1, 2, 4, 8, 16\}$, when the degree of cooperation, Q, is equal to 3. As we can see, increments of the quantization rate imply increments of the lower bound and decrements of the gap between lower and upper bound thanks to the reduction of the quantization error. Moreover, growing the operating SNR the quality of the channel state information, in terms of quantization rate, has to increase to obatin good performance compared to the ideal case. Let us, also, observe that the amounts of side information, appear to be high with respect to achievable sum-rate. However, we have no-take in account the coherent time of the system that reduce the effective quantization rate.

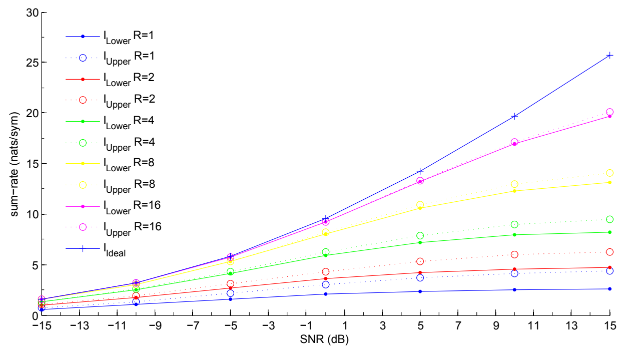

Figure 2.3: Lower and Upper bound of the sum rate, with degree of cooperation $Q = 3$, for several values of the quantization rate, and quantization performed by the users.

In Fig. 2.4, we illustrate the mean lower and upper bounds of the sum rate when the degree of cooperation $Q = 2$ and the quantization is performed by the BSs. Again, we consider as a set of quantization rate, given by $R(D) = \{2, 4, 8, 16, 32\}$. We can note that, for low values of the rate $R(D)$, there is an improvement of the performance with respect to the behavior of the system when the quantization is performed by the users. In fact, although there are some matrices that are not known, the base stations have more degrees of freedom in order to better allocate the resources. However for high values of the rate $R(D)$ we can see a reversed behavior: in this case it becomes more and more important to represent all the channel matrices. The limit behavior for $R \to \infty$ is the one illustrated in Fig. 2.2.

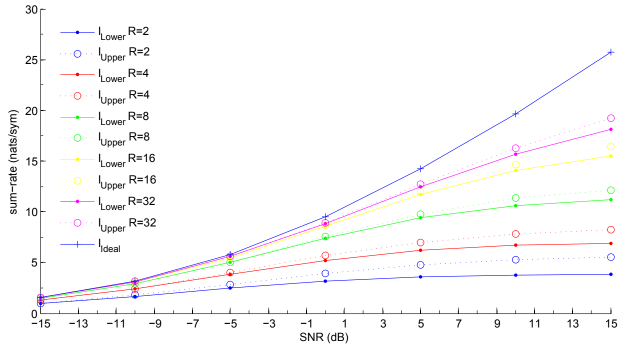

Figure 2.4: Lower and Upper bound of the sum rate, with degree of cooperation $Q = 2$, for several values of the quantization rate, and quantization performed by the BSs.

2.2 Training Based Systems

In most of telecommunication networks, before the transmission of the information, there is a synchronization phase during which the receiver estimate the channel matrix. In particular a standard technique to allow the receivers to estimate the channel matrix consists in the transmission from the users of training sequences among the data, i.e. a set of symbols whose location and values are known to the receivers, [20]. Therefore, in this section we describe a multiple-user, correlated MIMO channel where the channel statistic $\mathbf{S}$ is obtained through a dedicated training phase, i.e. a training base system. Specifically, we consider a MIMO MAC, with i.i.d. block fading, where each block is divided into training and data transmission phases. The length of the training phase is N_{T} channel uses. Denoting by $\sqrt{P_T}\mathbf{t}_i^{(k)}$ the training vector of dimension M sent by the k-th user at the i-th channel use, where P_T is the normalized (per antenna) available power for the training phase per user, from the uplink channel model (1.1), the received signal at the i-th channel use, $i = 1, \ldots, N_{\mathrm{T}}$, is given by:

$$\mathbf{y}_i = \sqrt{\mathsf{SNR}_T}\mathbf{H}\mathbf{t}_i + \mathbf{n}, \tag{2.38}$$

where $\mathbf{y}_i$ is the L-dimensional output vector at the i-th channel use, $\mathbf{t}_i = [\mathbf{t}_i^{(1)\dagger}, \ldots, \mathbf{t}_i^{(K)\dagger}]^\dagger$ is the MK-dimensional normalized training vector at the i-th channel use, $\mathsf{SNR}_T = \frac{P_T}{N_0}$, $\mathbf{H}$ is the Channel matrix, assumed zero mean Gaussian, and finally $\mathbf{n}$ is the additive circularly symmetric zero-mean Gaussian noise with unit variance. Compacting the entire training phase in a matrix, we have:

$$\mathbf{Y}_{\mathrm{T}} = \sqrt{\mathsf{SNR}_T}\mathbf{H}\mathbf{T} + \mathbf{N}, \tag{2.39}$$

where $\mathbf{Y}_{\mathrm{T}} = [\mathbf{y}_1, ..., \mathbf{y}_{N_{\mathrm{T}}}]$ has dimension $L \times N_{\mathrm{T}}$, $\mathbf{T} = [\mathbf{T}^{(1)^\dagger}, ..., \mathbf{T}^{(K)^\dagger}]^\dagger$ has dimension $MK \times N_{\mathrm{T}}$, with $\mathbf{T}^{(k)} = [\mathbf{t}_1^{(k)}, ..., \mathbf{t}_{N_{\mathrm{T}}}^{(k)}]$ denoting the $M \times N_{\mathrm{T}}$ normalized training matrix for the k-th user, whose ℓ-th row contains the normalized training sequence transmitted by the k-th user from its ℓ-th antenna. Furthermore $\mathbf{N}$ is a zero mean Gaussian random matrix of dimension $L \times N_{\mathrm{T}}$, with i.i.d. entries of unit variance. We impose a power constraint on $\mathbf{T}^{(k)}$, specifically

$$\frac{1}{MN_{\mathrm{T}}}\mathrm{tr}\left\{\mathbf{T}^{(k)}\mathbf{T}^{(k)^\dagger}\right\} \leq 1, \qquad k = 1, \ldots, K. \tag{2.40}$$

For this scheme, we have that $\mathbf{S} = \mathbf{Y}_{\mathrm{T}}$,

$$\hat{\mathbf{H}} = \mathbb{E}_{\mathbf{S}}[\mathbf{H}] = \mathbb{E}_{\mathbf{Y}_{\mathrm{T}}}[\mathbf{H}],$$

and the innovation matrix is given by

$$\mathbf{Z} = \mathbf{H} - \mathbb{E}_{\mathbf{S}}[\mathbf{H}] = [\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(K)}],$$

with

$$\mathbf{Z}^{(k)} = \mathbf{H}^{(k)} - \mathbb{E}_{\mathbf{S}}[\mathbf{H}^{(k)}],$$

and satisfies **Hypothesis 2**. Let us characterize $\mathbf{Z}$. Using a vector representation, (2.39) can be equivalently expressed as:

$$\mathbf{y}_{\mathrm{T}} = \sqrt{\mathsf{SNR}_T}\tilde{\mathbf{T}}\mathbf{h} + \mathbf{n}, \tag{2.41}$$

where $\tilde{\mathbf{T}} = \mathbf{I}_L \otimes \mathbf{T}^T$, $\mathbf{y}_{\mathrm{T}} = \mathrm{vec}\{\mathbf{Y}_{\mathrm{T}}\}$, $\mathbf{h} = \mathrm{vec}\{\mathbf{H}\}$ and $\mathbf{n} = \mathrm{vec}\{\mathbf{N}\}$ are the column vector obtained by stacking, respectively, the rows of $\mathbf{Y}_{\mathrm{T}}$, $\mathbf{H}$ and $\mathbf{N}$. Now, evaluating the MMSE-estimation of the channel coefficients based on observable $\mathbf{y}_{\mathrm{T}}$, given in (2.41), we have:

$$\hat{\mathbf{h}}_{\mathsf{mmse}} = \sqrt{\mathsf{SNR}_T}\mathbf{C}_{\mathbf{h}}\tilde{\mathbf{T}}^{\dagger}(\mathbf{I}_{LN_T} + \mathsf{SNR}_T\tilde{\mathbf{T}}\mathbf{C}_{\mathbf{h}}\tilde{\mathbf{T}}^{\dagger})^{-1}\mathbf{y}_{\mathrm{T}} \tag{2.42}$$

where $\mathbf{C}_{\mathbf{h}} = \mathbb{E}[\mathbf{h}\mathbf{h}^{\dagger}]$ is the covariance matrix of the channel random vector $\mathbf{h}$. Then, for a fixed training sequence $\mathbf{T}$, the corresponding error covariance matrix, $\mathbf{C}_{\mathbf{e}}(\mathbf{T})$, is:

$$\begin{aligned}\mathbf{C}_{\mathbf{e}}(\mathbf{T}) &= \mathbb{E}\left[(\mathbf{h} - \hat{\mathbf{h}}_{\mathsf{mmse}})(\mathbf{h} - \hat{\mathbf{h}}_{\mathsf{mmse}})^{\dagger}\right] \\ &= \mathbf{C}_{\mathbf{h}} - \mathsf{SNR}_T\mathbf{C}_{\mathbf{h}}\tilde{\mathbf{T}}^{\dagger}\left(\mathbf{I}_{LN_T} + \mathsf{SNR}_T\tilde{\mathbf{T}}\mathbf{C}_{\mathbf{h}}\tilde{\mathbf{T}}^{\dagger}\right)^{-1}\tilde{\mathbf{T}}\mathbf{C}_{\mathbf{h}}\end{aligned} \tag{2.43}$$

where $\mathbf{e}=\mathrm{vec}\{\mathbf{Z}\} = \mathbf{h} - \hat{\mathbf{h}}_{\mathsf{mmse}}$. If $\mathbf{C}_{\mathbf{h}}$ is invertible, using the inversion lemma, $\mathbf{C}_{\mathbf{e}}$ admits the following expression:

$$\mathbf{C}_{\mathbf{e}}(\mathbf{T}) = (\mathbf{C}_{\mathbf{h}}^{-1} + \mathsf{SNR}_T\tilde{\mathbf{T}}^{\dagger}\tilde{\mathbf{T}})^{-1}. \tag{2.44}$$

In the following, for sake of simplicity, with no loss of generality, we assume that $\mathrm{tr}\{\mathbf{C}_{\mathbf{h}}\} = 1$. In subsequent subsections, we discuss about the design of the training sequences, the precodings that have to be used by users, concluding the section with some numerical results.

2.2.1 Optimal Training-Sequences Design

In this subsection, we consider the optimization of the training sequences, focusing on two figures of merit:

- the trace of $\mathbf{C_e}(\mathbf{T})$, which is related to the expected square norm of the error vector,
- the determinant of $\mathbf{C_e}(\mathbf{T})$, which is related to the average volume in which is concentrated the error vector. Notice that $|\mathbf{C_e}(\mathbf{T})| \neq 0$ if and only if $\mathbf{C_h}$ is invertible.

For both the figures of merit we are interested in finding the minimum. In particular the optimization problem that we consider is:

$$\begin{cases} \min\limits_{\mathbf{T}\in C^{MK\times N_{\mathrm{T}}}} & f(\mathbf{T}) \\ \text{s.t.} \ \frac{1}{MN_{\mathrm{T}}}\mathrm{tr}\left\{\mathbf{T}_i\mathbf{T}_i^\dagger\right\} \leq 1 & i=1,\dots,K \end{cases} \tag{2.45}$$

wherein $f(\mathbf{T})$=tr$\{\mathbf{C_e}(\mathbf{T})\}$ or $f(\mathbf{T})$=$|\mathbf{C_e}(\mathbf{T})|$.

Let us now observe that in the case $N_{\mathrm{T}} \geq MK$ we have that the matrix $\mathbf{T}\mathbf{T}^\dagger$ can be full rank, while in case $N_{\mathrm{T}} < MK$ the matrix $\mathbf{T}\mathbf{T}^\dagger$ is rank-deficient. We will discuss separately the two cases.

Full rank case
When $N_{\mathrm{T}} \geq MK$, the two optimization problems can be transformed in convex optimization problem, which can be efficiently solved through the interior point method. However, an explicit expression for the optimal training sequences can be found as illustrated by the following theorems.

Theorem 8 *Assuming that each channel matrix $\mathbf{H}^{(k)}$, $k=1,\dots,K$, is characterized through the UIU model in (1.6), then the optimal training sequences solving (2.45) with $f(\mathbf{T})$=tr$\{\mathbf{C_e}(\mathbf{T})\}$, are given by:*

$$\mathbf{T}^{(k)} = (\mathbf{e_{K,k}} \otimes \mathbf{U}_{\mathrm{T}}^{(\mathbf{k})}\sqrt{\mathbf{\Upsilon}^{(\mathbf{k})*}})\mathbf{I}_{KM}{}^{(N_T)}\mathbf{V} \tag{2.46}$$

where $\mathbf{V}$ can be any arbitrary $N_T \times N_T$ unitary matrix, $\mathbf{U}_{\mathrm{T}}^{(k)}$ is defined as in (1.6), and $\mathbf{\Upsilon}^{(k)*}$ is a $M \times M$ diagonal matrix whose diagonal elements satisfy *the following conditions:*

$$\Upsilon_j^{(k)*} = \begin{cases} 0 & \text{if } \frac{1}{L}\sum_{l=1}^{L} [\boldsymbol{\Sigma}_w^{(k)}]_{l,j}^2 \leq \frac{1}{MN_T}\sum_{j=1}^{M} \beta_j^{(k)} \\ \frac{MN_T \beta_j^{(k)}}{\sum_{j=1}^{M} \beta_j^{(k)}} & \text{otherwise} \end{cases} \quad (2.47)$$

where

$$\beta_j^{(k)} = \frac{1}{L}\sum_{l=1}^{L} \frac{\mathsf{SNR}\,\Upsilon_j^{(k)*}\,[\boldsymbol{\Sigma}_w^{(k)}]_{l,j}^2}{(1+\mathsf{SNR}\,[\boldsymbol{\Sigma}_w^{(k)}]_{l,j}\Upsilon_j^{(k)*})^2}$$

.

Proof: See Appendix B **Proof of Theorem 8**.

Theorem 9 *Assuming that each channel matrix* $\mathbf{H}^{(k)}$, $k = 1, \ldots, K$, *is characterized through the UIU model in (1.6) with* $\mathbf{C_h}$ *having full rank, then the optimal training sequences solving (2.45) with* $f(\mathbf{T}) = |\mathbf{C_e}(\mathbf{T})|$ *are given by:*

$$\mathbf{T}^{(k)} = (\mathbf{e_{K,k}} \otimes \mathbf{U_T^{(k)}} \sqrt{\boldsymbol{\Upsilon}^{(\mathbf{k})*}}) \mathbf{I}_{KM}{}^{(N_T)} \mathbf{V} \quad (2.48)$$

where $\mathbf{V}$ *can be any arbitrary* $N_T \times N_T$ *unitary matrix,* $\mathbf{U}_\mathbf{T}^{(k)}$ *is defined as in (1.6), and* $\boldsymbol{\Upsilon}^{(k)*}$ *is a* $M \times M$ *diagonal matrix whose diagonal elements satisfy the following conditions:*

$$\Upsilon_j^{(k)*} = \begin{cases} 0 & \text{if } \frac{1}{L}\sum_{l=1}^{L} [\boldsymbol{\Sigma}_w^{(k)}]_{l,j} \leq \frac{1}{MN_T}\sum_{j=1}^{M} (1-\eta_j^{(k)}) \\ \frac{MN_T(1-\eta_j^{(k)})}{\sum_{j=1}^{M}(1-\eta_j^{(k)})} & \text{otherwise} \end{cases} \quad (2.49)$$

where

$$\eta_j^{(k)} = \frac{1}{L}\sum_{l=1}^{L} \frac{1}{1+\mathsf{SNR}\,[\boldsymbol{\Sigma}_w^{(k)}]_{l,j}\Upsilon_j^{(k)*}}$$

.

Proof: See Appendix B **Proof of Theorem 9**.

The key information that we obtain from the above Theorems are that, for both the figures of merit, different users have to send orthogonal training sequences and not all the channel coefficients are equally important. In practice, referring to $\mathbf{H}_w^{(k)}$, those channel coefficients whose power is small compared with the other channel coefficients and the background noise, are not important and the transmitted power is better utilized in channel coefficients that exhibit a higher mean square value. Obviously the way in which the power is allocated depends on the specific figure of merits, considered. Let us observe that the previous results generalize, to the multiple user setting, the design of the training sequences given in [33], for the point-to-point MIMO channel. Moreover we don't restrict the attention to the only virtual channel representation, as done in [33], and also find a fixed-point equation that have to be solved for the optimal power allocation.

A totally close-form expression for the singular-eigenvalues of $\mathbf{T}$ can be found for the setting illustrated in the subsequent theorem.

Corollary 3 *If for all $k = 1, \ldots, K$ $\mathbf{H}^{(k)}$ is characterized through the separable model with uncorrelated transmitting antennas, then uniform power allocation policy across the singular-values of $\mathbf{T}$ is optimal, i.e. $\lambda_j^{(k)} = N_T$ for all $j = 1, \ldots, M$ and $k = 1, \ldots, K$, for both $f(\mathbf{T})$=$\mathrm{tr}\{\mathbf{C_e}(\mathbf{T})\}$ and $f(\mathbf{T}) = |\mathbf{C_e}(\mathbf{T})|$.*

Let us observe that, as concerning the training dimensioning, if instead of considering the MMSE covariance matrix, we consider the error covariance matrix of maximum likelihood (ML) estimate, it is easy to prove that uniform power allocation policy across the singular-values of $\mathbf{T}$ is optimal in the sense of minimizing the trace or the determinant of the ML error covariance matrix.

Rank-deficient case

Let us now consider the case $N_\mathrm{T} < MK$. In general the problem (2.45), in the rank-deficient case, cannot be converted to convex problems, and so the classical Lagrangian method cannot be used to solve it. The general optimization problem in (2.45), when $\mathbf{C_e}(\mathbf{T})$ is an arbitrary matrix, is still an open problem. However it can be proved that:

Theorem 10 *If $\mathbf{H}^{(k)}$, with $k = 1, \ldots, K$, are statistically equivalent, with i.i.d. entries, then the optimal training sequences for the optimization problem (2.45) are given by*

$$\mathbf{T}^{(k)} = \sqrt{MK}\mathbf{U}^{(k)}\mathbf{V}, \tag{2.50}$$

where $\mathbf{V}$ can be any arbitrary $N_T \times N_T$ unitary matrix (for simplicity we could *assume* $\mathbf{V} = \mathbf{I}_{N_T}$*) and* $\mathbf{U}^{(k)}$ *is a* $M \times N_T$ *unitary matrix given by*

$$\mathbf{U}^{(k)} = (\mathbf{e_{K,k}} \otimes \mathbf{I_M})\mathbf{Q}_\pi \mathbf{I}_{MK}^{(N_T)}$$

where $\mathbf{Q}_\pi$ *is a per-column permuted version of a generic unitary matrix* $\mathbf{Q}$*, of dimension* $MK \times MK$*, such that, for all* $h = 1, \ldots, MK$*,*

$$\sum_{i=1}^{N_T} [\mathbf{Q}_\pi]_{h,i}^2 = \frac{N_T}{MK}. \tag{2.51}$$

Examples of such a $\mathbf{Q}$ *are the Fourier matrices and the Hadamard matrices, for which any permutation* $\mathbf{Q}_\pi$ *satisfies the condition (2.51).*

Proof: *See Appendix B* ***Proof of Theorem 10****.*

In the following we give an heuristic algorithm whose outcome represents a possible choice of training sequences in the case that $N_{\mathrm{T}} < MK$ under the assumption that $N_T = \vartheta K$ with $\vartheta \in \{1, \ldots, M\}$.

Specifically we will assume that the users send orthogonal training sequences, and ϑ orthogonal direction are allocated to each one. Then the training sequences that we consider are given by:

$$\mathbf{T}^{(k)} = \left(\mathbf{e_{K,k}} \otimes \mathbf{U}_{\mathrm{T}}^{(\mathbf{k})} \boldsymbol{\Upsilon}^{(\mathbf{k})\frac{1}{2}}\right) \mathbf{V} \tag{2.52}$$

where $\mathbf{U}_{\mathrm{T}}^{(k)}$ is defined as in (1.6), $\mathbf{V}$ is a $\vartheta K \times \vartheta K$ unitary matrix and $\sqrt{\boldsymbol{\Upsilon}^{(k)}}$ is a $M \times \vartheta$ matrix with zero off-diagonal elements and whose ϑ diagonal elements satisfy the following conditions,

$$\Upsilon_j^{(k)} = \begin{cases} 0 & \text{if } \frac{1}{L}\sum_{l=1}^{L} [\boldsymbol{\Sigma}_w^{(k)}]_{l,j}^2 \le \frac{1}{MN_{\mathrm{T}}} \sum_{j=1}^{\vartheta} \beta_j^{(k)} \\ \frac{MN_{\mathrm{T}}\beta_j^{(k)}}{\sum_{j=1}^{M} \beta_j^{(k)}} & \text{otherwise} \end{cases} \tag{2.53}$$

where

$$\beta_j^{(k)} = \frac{1}{L} \sum_{l=1}^{L} \frac{\mathsf{SNR}\, \Upsilon_j^{(k)} \left[\Sigma_w^{(k)}\right]_{l,j}^2}{(1 + \mathsf{SNR}\, [\Sigma_w^{(k)}]_{l,j} \Upsilon_j^{(k)})^2},$$

if $f(\mathbf{T})$=tr $\{\mathbf{C_e}(\mathbf{T})\}$, otherwise, if $f(\mathbf{T}) = |\mathbf{C_e}(\mathbf{T})|$,

$$\Upsilon_j^{(k)} = \begin{cases} 0 & \text{if } \frac{1}{L}\sum_{l=1}^{L}[\mathbf{\Sigma}_w^{(k)}]_{l,j} \leq \frac{1}{MN_{\mathrm{T}}}\sum_{j=1}^{\vartheta}(1-\eta_j^{(k)}) \\ \frac{MN_{\mathrm{T}}(1-\eta_j^{(k)})}{\sum_{j=1}^{M}(1-\eta_j^{(k)})} & \text{otherwise} \end{cases} \tag{2.54}$$

where

$$\eta_j^{(k)} = \frac{1}{L}\sum_{l=1}^{L}\frac{1}{1+\mathsf{SNR}\,[\Sigma_w^{(i)}]_{l,j}\Upsilon_j^{(k)}}.$$

Let us observe that the training sequences given in (2.52) achieve the same performance of the one given in (2.46) or (2.48), when $N_T = KM$, and, when are satisfied the hypotheses of Theorem 10, we obtain the same performance of the one given in (2.50).

2.2.2 Robust Training-Sequences Design

When the parameters characterizing the second order statistic of the channel matrices $\mathbf{H}^{(k)}$ are unknown to the transmitters, a robust training-sequence design is needed. More precisely, the training sequence can be designed following a max-min approach, so that $\mathbf{T}$ is designed to minimize the worst-case cost under all possible error covariance matrices. Specifically, in the case that $\mathbf{C_h}$ has full rank, $\mathbf{T}$ is designed as the solution of the following min-max problem:

$$\min_{\mathbf{T}}\max_{\mathbf{C_h}} f((\mathbf{C_h}^{-1} + \mathsf{SNR}_T\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}})^{-1}), \tag{2.55}$$

where the minimum is over all possible training sequence such that $\frac{1}{MN_T}\mathrm{tr}\{\mathbf{T}^{(k)}\mathbf{T}^{(k)\dagger}\} \leq 1,\ \forall k \in \mathcal{U}$ and the maximum is over all possible covariance matrices $\mathbf{C_h}$ such that $\mathrm{tr}\{\mathbf{C_h}\} = 1$. In (2.55) we assume either:

$$f((\mathbf{C_h}^{-1} + \mathsf{SNR}_T\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}})^{-1}) = \mathrm{tr}\left\{(\mathbf{C_h}^{-1} + \mathsf{SNR}_T\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}})^{-1}\right\},$$

or

$$f((\mathbf{C_h}^{-1} + \mathsf{SNR}_T\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}})^{-1}) = \frac{1}{|\mathbf{C_h}^{-1} + \mathsf{SNR}_T\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}}|}.$$

In the following we assume that the transmitters have complete lack of prior information on $\mathbf{C_h}$. We further assume that $N_{\mathrm{T}} \geq MK$. Under this hypothesis, the training-sequences matrix $\mathbf{T}$ designed based on the min-max approach, in (2.55), is given by the following theorem:

Theorem 11 *The training-sequences matrix* $\mathbf{T}$ *which solves (2.55) i.e.:*

$$\min_{\mathbf{T}} \max_{\mathbf{C_h}} f((\mathbf{C_h}^{-1} + \mathsf{SNR}_T \tilde{\mathbf{T}}^\dagger \tilde{\mathbf{T}})^{-1}),$$

where the minimum is over all possible training sequences satisfying the power constraints in (2.40), while the maximum is over all covariance matrices $\mathbf{C_h}$ *such that* $\mathrm{tr}\{\mathbf{C_h}\} = 1$*, is given by:*

$$\mathbf{T} = \sqrt{N_T} \mathbf{I}_{KM}^{(N_T)} \tag{2.56}$$

Proof: *See Appendix B* ***Proof of Theorems 11****.*

Theorem 11 still holds for the case that $\mathbf{C_h}$ can be not full rank. In this case the only option for a robust design of the training sequences is focus on the trace of the error covariance matrix $\mathbf{C_e}$ as given in (2.43). In this case we can prove the following result:

Theorem 12 *The training-sequences matrix* $\mathbf{T}$ *which solves:*

$$\min_{\mathbf{T}} \max_{\mathbf{C_h}} \mathrm{tr} \left\{ \mathbf{C_h} - \mathsf{SNR}_T \mathbf{C_h} \tilde{\mathbf{T}}^\dagger \left(\mathbf{I}_{LN_T} + \mathsf{SNR}_T \tilde{\mathbf{T}} \mathbf{C_h} \tilde{\mathbf{T}}^\dagger \right)^{-1} \tilde{\mathbf{T}} \mathbf{C_h} \right\},$$

where the minimum is over all possible training sequences satisfying the power constraint in (2.40), while the maximum is over all covariance matrices $\mathbf{C_h}$ *such that* $\mathrm{tr}\{\mathbf{C_h}\} = 1$*, is given by:*

$$\mathbf{T} = \sqrt{N_T} \mathbf{I}_{KM}^{(NT)} \tag{2.57}$$

Proof: *See Appendix B* ***Proof of Theorems 12****.*

Then the robust training sequences, correspond to orthogonal and equi-energetic sequences, as intuitively we could expect starting from an ML estimation approach.

2.2.3 Precoding Optimization for Training Based Systems

In this subsection we evaluate the optimal precoding for the lower bound to the sum-rate, given in (1.37), in the context of training-based statistic $\mathbf{S}$, specializing the results given in subsection 1.2.2. The attention is focused on the case when $\mathbf{T}$ has full rank, since in rank-deficient condition the optimal training-sequence structure is still an unsolved problem in general. Concerning the full rank case, we can prove the following results:

Theorem 13 *Assume that* $\mathbf{H}$ *is modeled according to (1.6), i.e:* $\mathbf{H}^{(k)} = \mathbf{U}_{\mathrm{R}}^{(k)}\mathbf{H}_w^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)\dagger}$ *for all* $k = 1, \dots, K$, *and that* $N_T \geq KM$.

If the training sequences are optimized based on Theorem 8 or Theorem 9, then the eigenvectors of the input covariance matrix $\mathbf{P}_k^*$ *that maximizes (1.37) are given by the columns of* $\mathbf{U}_{\mathrm{T}}^{(k)}$.

Proof: See Appendix B **Proof of Theorem 13**.

Theorem 14 *Assume that* $\mathbf{H}$ *is modeled according to (1.6), i.e:*

$$\mathbf{H}^{(k)} = \mathbf{U}_{\mathrm{R}}^{(k)}\mathbf{H}_w^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)\dagger}$$

where for all $k = 1, \dots, K$, $\mathbf{\Sigma}_w^{(k)}$ *is given by the outer product*

$$\mathbf{\Sigma}_w^{(k)} = \boldsymbol{\lambda}_{\mathrm{R}}^{(k)}\mathbf{1}_M{}^\dagger$$

with $\boldsymbol{\lambda}_{\mathrm{R}}^{(k)}$ *denoting a L-dimensional vector of non-negative entries* [2]. *If* $N_T \geq KM$ *and the training sequence are optimized based on Theorem 8 or Theorem 9, then the input covariance matrix* $\mathbf{P}_k^*$ *that maximize (1.37) is a scalar matrices, i.e.:*

$$\mathbf{P}_k^* = \frac{\varrho_k}{M}\mathbf{I}_M \quad k = 1, \dots, K \tag{2.58}$$

where $\varrho_k \leq 1 \quad k = 1, \dots, K$.

[2]Note that this assumption is equivalent to say that $\mathbf{H}$ is modeled according to a separable model, i.e: $\mathbf{H}^{(k)} = \mathbf{\Sigma}_{\mathrm{R}}^{k\,\frac{1}{2}}\mathbf{W}^{(k)}\mathbf{\Sigma}_{\mathrm{T}}^{k\,\frac{1}{2}}$ for all $k = 1, \dots, K$ where $\mathbf{\Sigma}_{\mathrm{T}}^k = \mathbf{I}_M$, $\mathbf{\Sigma}_{\mathrm{R}}^k = \mathbf{U}_{\mathrm{R}}^{(k)}\mathrm{diag}(\boldsymbol{\lambda}_{\mathrm{R}})\mathbf{U}_{\mathrm{R}}^{(k)\dagger}$ and $\mathbf{W}$ is a $L \times M$ matrix with i.i.d. entries.

Furthermore, the set of $\{\mathbf{P}_k^*\}_{k=1}^K$ *given as in (2.58) with* $\varrho_k = 1 \quad k = 1, \ldots, K$*, is a Nash equilibrium point for the payoff function:*

$$\mathfrak{L} = \{L_1, \ldots, L_K\}$$

where L_k *denotes the lower bound on the rate of the* k*-th user i.e.*

$$L_k = \mathbb{E}\left[\log\left|\mathbf{I} + \left(\mathbf{\Omega}_{\mathcal{U}} + \mathbf{\Gamma}_{\{\pi(k+1),\ldots,\pi(K)\}} + \tfrac{1}{\mathsf{snr}}\mathbf{I}\right)^{-1}\mathbf{\Gamma}_{\{\pi(k)\}}\right|\right]$$

for a fixed chosen decoding ordering, π*, of the users at the base station.*

Proof: See Appendix B **Proof of Theorem 14**.

Theorem 15 *Assume that* $\mathbf{H}$ *is modeled according to (1.6), i.e:*

$$\mathbf{H}^{(k)} = \mathbf{U}_{\mathrm{R}}^{(k)}\mathbf{H}_w^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)\dagger}$$

where for all $k = 1, \ldots, K$*,* $\mathbf{\Sigma}_w^{(k)}$ *is given by the outer product*

$$\mathbf{\Sigma}_w^{(k)} = \frac{1}{L}\mathbf{1}_L\boldsymbol{\lambda}_{\mathrm{T}}^{(k)\dagger}$$

with $\boldsymbol{\lambda}_{\mathrm{T}}^{(k)}$ *denoting a* M*-dimensional vector of non-negative entries.* [3]

If $N_T \geq KM$ *and the training sequence are optimized based on Theorem 8 or Theorem 9, then, the input covariance matrices* $\mathbf{P}^{k*}$*, with* $k = 1, \ldots, K$*, maximizing the lower bound on the sum rate (1.37), are given by:*

$$\mathbf{P}^{k*} = \mathbf{U}_{\mathrm{T}}^{k}\mathbf{\Lambda}_k^*\mathbf{U}_{\mathrm{T}}^{k\dagger} \qquad k = 1, \ldots, K \tag{2.59}$$

where $\mathbf{\Lambda}^*{}_k = diag(\lambda_1^{k*}, \ldots, \lambda_M^{k*})$*, with* $k = 1, \ldots K$*, are diagonal matrices, whose diagonal elements are given by:*

$$\boldsymbol{\lambda}^* = \frac{1}{1 - (\boldsymbol{\lambda}_{\mathrm{T}} - \boldsymbol{\psi})^{\dagger}\,\mathbf{d}}\mathbf{d} \tag{2.60}$$

where

[3] Note that this assumption is equivalent to say that $\mathbf{H}$ is modeled according to a separable model, i.e: $\mathbf{H}^{(k)} = \mathbf{\Sigma}_{\mathrm{R}}^{k\,\frac{1}{2}}\mathbf{W}^{(k)}\mathbf{\Sigma}_{\mathrm{T}}^{k\,\frac{1}{2}}$ for all $k = 1, \ldots, K$ where $\mathbf{\Sigma}_{\mathrm{T}}^{k} = \mathbf{U}_{\mathrm{T}}^{(k)}\mathrm{diag}(\boldsymbol{\lambda}_{\mathrm{T}})\mathbf{U}_{\mathrm{T}}^{(k)\dagger}$, $\mathbf{\Sigma}_{\mathrm{R}}^{k} = \frac{1}{L}\mathbf{I}_L$ and $\mathbf{W}$ is a $L \times M$ matrix with i.i.d. entries.

- $\boldsymbol{\psi} = [\Upsilon_1^{(1)*}, \ldots, \Upsilon_M^{(1)*}, \ldots, \Upsilon_1^{(K)*}, \ldots, \Upsilon_M^{(K)*}]$ *with* $\Upsilon_j^{(k)*}$ *given either by (2.47) or (2.48),*

- $\boldsymbol{\lambda}_{\mathrm{T}}$ *is the* KM*-dimensional vector defined as:* $\boldsymbol{\lambda}_{\mathrm{T}} = [\boldsymbol{\lambda}_{\mathrm{T}}^{(1)}, \ldots, \boldsymbol{\lambda}_{\mathrm{T}}^{(K)}]$.

- $\mathbf{d} = [\mathbf{d}_1^\dagger, \ldots, \mathbf{d}_K^\dagger]^\dagger$ *is the the* KM*-dimensional vector whose elements are the non-negative solution of the following convex optimization problem:*

$$\begin{aligned} &\max_{\{\mathbf{d}_k\}} \ \mathbb{E}\left[\log\left|\mathbf{I}_L + \mathsf{SNR}\sum_{k=1}^{K}\hat{\mathbf{H}}^{(k)}\mathbf{U}_T^{(k)} diag\left(\mathbf{d}_k\right)\mathbf{U}_T^{(k)^\dagger}\hat{\mathbf{H}}^{(k)^\dagger}\right|\right] \\ &\text{s.t.} \ : \quad \sum_{i=1}^{M} d_{k,i} \leq 1 - (\boldsymbol{\lambda}_{\mathrm{T}} - \boldsymbol{\psi})^\dagger\,\mathbf{d}, \ \ k = 1, \ldots, K. \end{aligned} \tag{2.61}$$

with $d_{k,i}$ *denoting the* i*-th element of* $\mathbf{d}_k$.

Proof: See Appendix B **Proof of Theorem 15**.

2.2.4 Numerical Results for Training Based Systems

In this subsection we present some numerical results to obtain insights about the behavior of a training based system. In our numerical evaluations, we consider a MIMO MAC with i.i.d. block fading, where each block is divided into a training phase and a data-transmission phase. The coherence time of the channel (i.e., the duration of each block) is denoted by T_c. The scenario analyzed is constituted by a base station, located at the center of a hexagonal cell of normalized unitary radius, and four users, $K = 4$, uniformly distributed in the cell, with a normalized distance from the center greater than 0.1. We assume that the base station is equipped with four receiving antennas, $L = 4$, and the users are equipped with two transmitting antennas, $M = 2$. Moreover we assume $\mathbf{H}$ modeled according to (1.6), precisely:

$$\mathbf{H}^{(k)} = \mathbf{U}_{\mathrm{R}}^{(k)}\mathbf{H}_w^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)^\dagger}$$

where for all $k = 1, \ldots, K$, $\boldsymbol{\Sigma}_w^{(k)}$ is given by an outer product of an L-dimensional vector $\boldsymbol{\lambda}_{\mathrm{R}}^{(k)^\dagger}$ and an M-dimensional vector $\boldsymbol{\lambda}_{\mathrm{T}}^{(k)^\dagger}$ i.e. $\boldsymbol{\Sigma}_w^{(k)} = \boldsymbol{\lambda}_{\mathrm{R}}^{(k)}\boldsymbol{\lambda}_{\mathrm{T}}^{(k)^\dagger}$. Under this model, which is commonly known as separable model,

$\mathbf{H}^{(k)}$ admits the following equivalent expression:

$$\mathbf{H}^{(k)} = \sqrt{d_k^{-\gamma}} \boldsymbol{\Sigma}_{\mathrm{R}}^{k\,\frac{1}{2}} \mathbf{W}^{(k)} \boldsymbol{\Sigma}_{\mathrm{T}}^{k\,\frac{1}{2}} \tag{2.62}$$

where $\mathbf{W}^{(k)}$ is a 4×2 matrix of i.i.d. zero mean Gaussian random variables with unit variance, d_k is the distance between the k-th user and the base station, $\sqrt{d_k^{-\gamma}}$ takes in account for the path loss attenuation of the signal, $\boldsymbol{\Sigma}_{\mathrm{T}}^k = \mathbf{U}_{\mathrm{T}}^{(k)} \mathrm{diag}(\boldsymbol{\lambda}_{\mathrm{T}}) \mathbf{U}_{\mathrm{T}}^{(k)\dagger}$ and $\boldsymbol{\Sigma}_{\mathrm{R}}^k = \mathbf{U}_{\mathrm{R}}^{(k)} \mathrm{diag}(\boldsymbol{\lambda}_{\mathrm{R}}) \mathbf{U}_{\mathrm{R}}^{(k)\dagger}$ are, respectively, the 2×2 correlation matrix at the transmitting antennas and the 4×4 correlation matrix at the receiving antennas, described by the following parametric models:

$$[\boldsymbol{\Sigma}_{\mathrm{T}}^k]_{l,l'} = e^{-0.5 d_T^2 |l-l'|} \tag{2.63}$$

$$[\boldsymbol{\Sigma}_{\mathrm{R}}^k]_{l,l'} = \frac{1}{4} e^{-0.5 d_R^2 |l-l'|} e^{i 2\pi f_0 / c \cos(\theta_R^k)(l-l')} \tag{2.64}$$

where f_0 is the carrier frequency, c is the speed of light, d_{T} and d_{R} are, respectively, the transmitting and receiving antennas spacing of the Uniform Linear Array (ULA) and θ_{R}^k is the angle between the direction defined by the receiving ULA of the base station and direction passing through the base-station position and the k-th user position. In all our numerical results, we suppose that all the users transmit at the maximum available power P, with covariance matrix $\mathbf{P}_k = \frac{1}{M} \mathbf{I}_M$ and that the training power P_T, defined in Section 2.2, equals $\frac{P}{M}$, i.e. we use the same power per antennas in both the phase.
In particular in Figs 2.5, 2.6, 2.7, and 2.8 we illustrate the behavior of the training based system, for different transmitting and the receiving antennas spacing, in terms of the average sum-rate (left plots):

$$I(\mathbf{x}; \mathbf{y} | \mathbf{S}), \tag{2.65}$$

and the average rate of the nearest user, which we denote by k^*, to the base station that we assume to be the first decoded using successive interference cancelation (right plots):

$$I(\mathbf{x}_{\mathcal{F}}; \mathbf{y} | \mathbf{S}, \mathbf{x}_{\mathcal{C}}). \tag{2.66}$$

with $\mathcal{F} = \{k^*\}$ and $\mathcal{C} = \emptyset$. In order to evaluate (2.65) and (2.66), we use the proposed bounds, given in Corollary 1 and Theorem 1. For each scenario we consider two different values of the training duration N_T, specifically:

$N_T = 8$, for which we are in the case of full rank condition, and $N_T = 4$. In the first case we use the results given in Subsection 2.2.1, Figs 2.5 and 2.7, and given in Subsection 2.2.2, Figs 2.6 and 2.8, to design the training sequences, while in the second case we assume that the users send Hadamard training sequences, although there is no optimality criteria. As already said, for all the plots we assume isotropic inputs (i.e. $\mathbf{P}_k = \frac{1}{M}\mathbf{I}_M$) due to different motivations. First of all, in the case of independent transmitting antennas it represents the Nash equilibrium solution and is the optimal one when we constraint the transmitted power to be the maximal one; moreover it is the optimal approach (in the min-max sense) when no prior information on the channel characterization is available at the transmitter; finally, since we are not able to optimize, in no-full rank case, the lower bound to the the sum rate, we can at least conduct a fair comparison between the performance of the full rank and no-full rank scenarios. From the plots, we see that, in all the situations, the

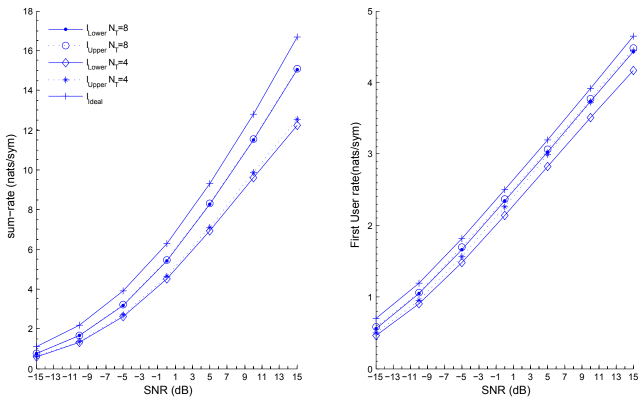

Figure 2.5: Sum-rate (left plot) and nearest user's rate (right plot) of a MIMO MAC with uncorrelated transmitting antennas, $d_R = \frac{\lambda}{2}$, and either $N_\mathrm{T} = 8$ or $N_\mathrm{T} = 4$.

upper and lower bound are tight, specially in the full rank condition, thus, the lower bound describes well the limiting performance of the system. Moreover the loss with respect to the ideal case of coherent reception is moderate in full rank condition. We can also observe that in the case of rank deficient condition, the correlation at the receive antennas has a large impact on the limiting performance of the system. More precisely, in the presence of high correlation between the receive antennas, the lower bound increases with SNR, in the analyzed range, while in the case of low correlation, we observe a quick saturation in the performance. Intuitively, high correlation between the receive antennas makes it easier to isolate the users by direction of arrival, which lowers inter-

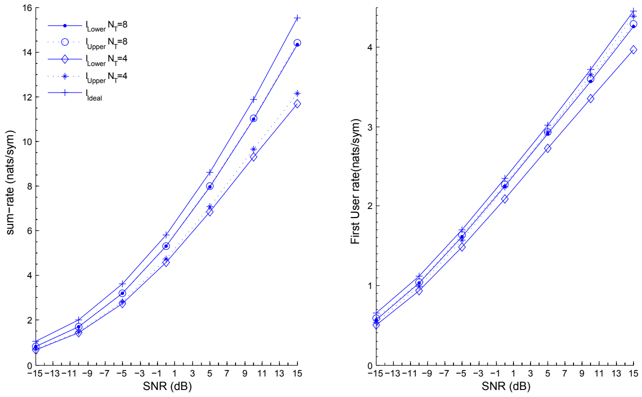

Figure 2.6: Sum-rate (left plot) and nearest user's rate (right plot) of a MIMO MAC with $d_T = \frac{\lambda}{2}$, $d_R = \frac{\lambda}{2}$, and either $N_{\mathrm{T}} = 8$ or $N_{\mathrm{T}} = 4$.

ference between users sending the same training sequences and permits better channel estimation. This in turn allows the sum rate to not quickly saturate with SNR. Another important figure of merit to analyze is the total throughput

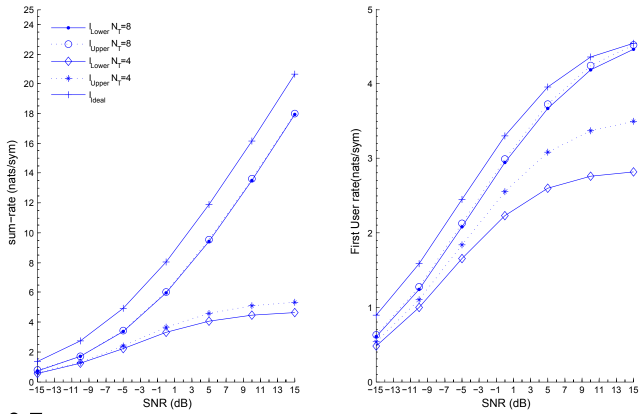

Figure 2.7: Sum-rate (left plot) and nearest user's rate (right plot) of a MIMO MAC with uncorrelated transmitting antennas, uncorrelated receiving antennas, and either $N_{\mathrm{T}} = 8$ or $N_{\mathrm{T}} = 4$.

of the MIMO MAC, which is defined as:

$$\mathcal{T} = \frac{T_c - N_{\mathrm{T}}}{T_c} I(\mathbf{x}; \mathbf{y}|\mathbf{S}), \tag{2.67}$$

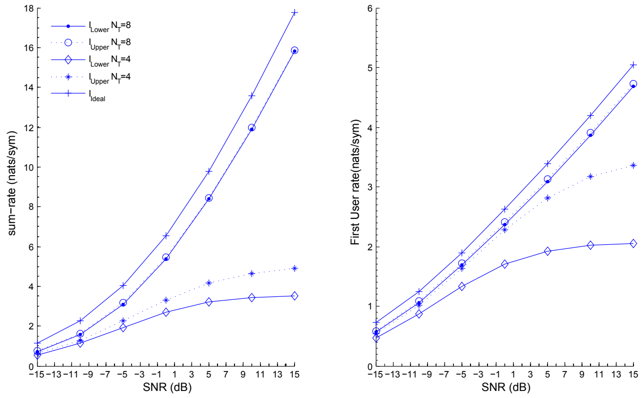

Figure 2.8: Sum-rate (left plot) and nearest user's rate (right plot) of a MIMO MAC with $d_T = \frac{\lambda}{2}$, uncorrelated receiving antennas, and either $N_{\rm T} = 8$ or $N_{\rm T} = 4$.

where T_c is the coherent time, the statistic $\mathbf{S}$ available at the receiver is obtained from training data as in (2.39), and finally the factor $\frac{T_c - N_{\rm T}}{T_c}$ takes in account the overhead introduced by the training. In Figs 2.9 and 2.10 we assume uncorrelated transmitting and receiving antennas. We also consider two different values of the training duration $N_{\rm T} = 8$ or $N_{\rm T} = 4$ and two different value of SNR : for the left plots we assume SNR $= -10dB$ and for the right plots we have SNR $= 10dB$. For comparison purposes, all the figures also illustrate the throughput for the case of perfect CSI. More precisely, in Fig. 2.9, we illustrate lower and upper bounds for the total throughput given in (2.67). In order to evaluate such lower and the upper bounds, we simply replace, in (2.67), $I(\mathbf{x}; \mathbf{y}|\mathbf{S})$ with its proposed lower and upper bounds, as given in Corollary 1. Figure 2.10, instead, illustrates the throughput of the nearest user that we assume to be the first decoded, using successive interference cancelation, at the base station. As previously seen, plots show that for both $N_{\rm T} = 8$ and $N_{\rm T} = 4$, the upper and lower bounds are tight. From the plots, it is also clear that a reduction factor of 2 in the training-sequence length translates into a reduction in the throughput that is more and more evident for high T_c. It is worth noticing that the tightness of the lower and upper bounds, as well as the beneficial effect of full-rank training matrix $\mathbf{T}$ (i.e., orthogonal training sequence), compared with rank-deficient training matrix $\mathbf{T}$, holds also for different values of SNR .

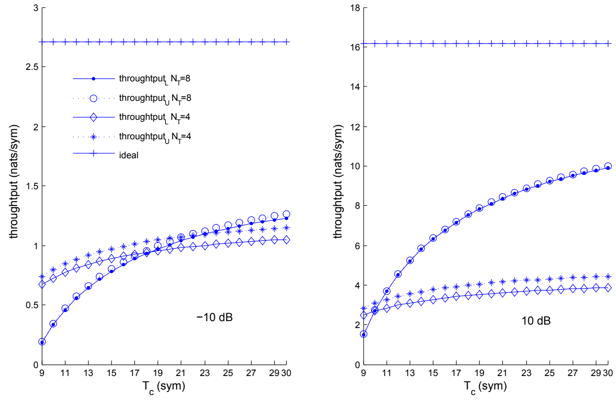

Figure 2.9: Total throughput of a MIMO MAC with $N_{\mathrm{T}} = 8$ and $N_{\mathrm{T}} = 4$: the left curve is for -10 dB and the right curve for 10dB.

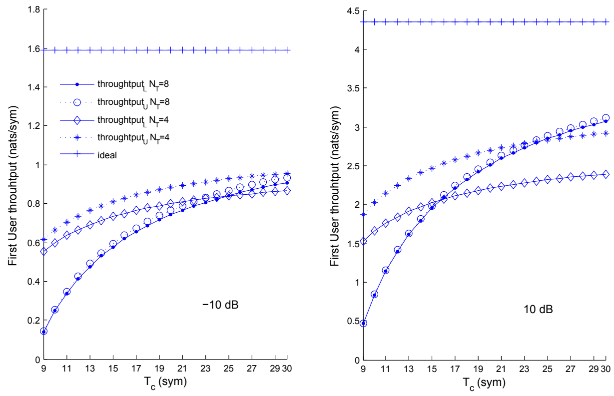

Figure 2.10: Nearest user's throughput of a MIMO MAC with $N_{\mathrm{T}} = 8$ and $N_{\mathrm{T}} = 4$: the left plot is for -10 dB and the right plot is for 10 dB.

2.3 Appendix B

Proof of Theorem 8 We first prove the result for the case when $\mathbf{C_h}$ is full rank. In proving Theorem 8 we proceed by steps:
Step 1: We start by finding a lower bound of the MMSE on the estimate of the channel matrix $\mathbf{H}$;
Step 2: We find a set of training sequences that achieve such a lower bound.

Step 1: Let us observe that a lower bound for the MMSE of $\mathbf{H}$ is given by:

$$\begin{aligned}
\mathrm{tr}\left\{\mathbf{C_e}(\mathbf{T})\right\} &= \sum_{k=1}^{K} mmse(\mathbf{H}^{(k)}|\mathbf{Y};\mathbf{T}) \\
&\geq \sum_{k=1}^{K} mmse(\mathbf{H}^{(k)}|\mathbf{Y},\{\mathbf{H}^{(j)}\}_{j\neq k};\mathbf{T}) \\
&= \sum_{k=1}^{K} \mathbb{E}\left[\left|\mathbf{H}^{(k)} - \mathbb{E}\left[\mathbf{H}^{(k)}|\mathbf{Y},\{\mathbf{H}^{(j)}\}_{j\neq k}\right]\right|^2\right] \\
&= \sum_{k=1}^{K} \mathbb{E}\left[\left|\mathbf{H}^{(k)} - \mathbb{E}\left[\mathbf{H}^{(k)}|\mathbf{H}^{(k)}\mathbf{T}^{(k)}+\mathbf{N}\right]\right|^2\right]
\end{aligned} \tag{2.68}$$

A further lower bound can be found by minimizing (2.68) with respect to $\mathbf{T}$. Given the fact that the power constraint on $\mathbf{T}$ is a disjoint power constraint over each $\mathbf{T}^{(k)}$ with $k = 1,\ldots,K$, then minimizing (2.68) is equivalent to minimizing for each $k = 1,\ldots K$, the term:

$$g(\mathbf{T}^{(k)}) = \mathbb{E}\left[\left|\mathbf{H}^{(k)} - \mathbb{E}\left[\mathbf{H}^{(k)}|\mathbf{H}^{(k)}\mathbf{T}^{(k)}+\mathbf{N}\right]\right|^2\right]$$

with respect to $\mathbf{T}^{(k)}$. To this end notice that, since $\mathbf{H}^{(k)} = \mathbf{U}_{\mathrm{R}}^{(k)}\mathbf{H}_w^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)\dagger}$, we have that:

$$\begin{aligned} g(\mathbf{T}^{(k)}) &= \mathbb{E}\left[\left|\mathbf{H}^{(k)} - \mathbb{E}\left[\mathbf{H}^{(k)}|\mathbf{H}^{(k)}\mathbf{T}^{(k)} + \mathbf{N}\right]\right|^2\right] \\ &= \mathbb{E}\left[\left|\mathbf{H}_w^{(k)} - \mathbb{E}\left[\mathbf{H}_w^{(k)}|\mathbf{H}^{(k)}\mathbf{T}^{(k)} + \mathbf{N}\right]\right|^2\right] \\ &= \mathbb{E}\left[\left|\mathbf{H}_w^{(k)} - \mathbb{E}\left[\mathbf{H}_w^{(k)}|\mathbf{H}_w^{(k)}\hat{\mathbf{T}}^{(k)} + \mathbf{N}\right]\right|^2\right] && (2.69) \\ &= \mathrm{tr}\{(\mathbf{C}_{\mathbf{h}_w^{(k)}}^{-1} + \mathsf{SNR}_T\tilde{\mathbf{T}}^{(k)\dagger}\tilde{\mathbf{T}}^{(k)})^{-1}\} && (2.70) \\ &\geq \mathrm{tr}\{(\mathbf{C}_{\mathbf{h}_w^{(k)}}^{-1} + \mathsf{SNR}_T\mathbf{D}^{(k)})^{-1}\} && (2.71) \end{aligned}$$

where in (2.69) $\hat{\mathbf{T}}^{(k)}=\mathbf{U}_{\mathrm{T}}^{(k)}\mathbf{T}^{(k)}$, while in (2.70) $\tilde{\mathbf{T}}^{(k)}=\mathbf{I}_L\otimes\ \hat{\mathbf{T}}^{(k)^T}$, $\mathbf{h}_w^{(k)}=\mathrm{vec}\{\mathbf{H}_w^{(k)}\}$, and $\mathbf{C}_{\mathbf{h}_w^{(k)}}=\mathbb{E}[\mathbf{h}_w^{(k)}\mathbf{h}_w^{(k)\dagger}]$ is the LMK-dimensional diagonal covariance matrix of $\mathbf{h}_w^{(k)}$. Finally in (2.71) we have used the fact that, given a $n \times n$ invertible matrix:

$$[\mathbf{A}^{-1}]_{i,i} \geq [\mathbf{A}]_{i,i}^{-1}$$

where the equality holds for the case of $\mathbf{A}$ diagonal and we have defined the diagonal matrix $\mathbf{D}^{(k)}$ in the following way:

$$[\mathbf{D}^{(k)}]_{i,i} = [\tilde{\mathbf{T}}^{(k)\dagger}\tilde{\mathbf{T}}^{(k)}]_{i.i}.$$

From (2.71), the problem of minimizing $g(\mathbf{T}_k)$ with respect to $\mathbf{T}_k$ boils down to solving:

$$\min \mathrm{tr}\{(\mathbf{C}_{\mathbf{h}_w^{(k)}}^{-1} + \mathsf{SNR}_T\mathbf{D}^{(k)})^{-1}\}$$

over all $LM \times LM$ diagonal matrices having the following block structure:

$$\mathbf{D}^{(k)} = \mathbf{I}_L \otimes \mathbf{\Upsilon}^{(\mathbf{k})}$$

and such that $\frac{1}{MN_T}\mathrm{tr}\left\{\mathbf{\Upsilon}^{(k)}\right\} \leq 1$ with $\mathbf{\Upsilon}^{(k)}$ denoting a $M \times M$ diagonal matrix. This is a classical convex problem, which can be solved considering the KKT condition, from which it follows that optimal $\mathbf{\Upsilon}^{(k)*}$ is a diagonal matrix with the diagonal elements given by:

$$\Upsilon_j^{(k)*} = \begin{cases} 0 & \text{if } \frac{1}{L}\sum_{l=1}^{L}[\mathbf{\Sigma}_w^{(k)}]_{l,j}^2 \leq \frac{1}{MN_\mathrm{T}}\sum_{j=1}^{M}\beta_j^{(k)} \\ \frac{MN_\mathrm{T}\beta_j^{(k)}}{\sum_{j=1}^{M}\beta_j^{(k)}} & \text{otherwise} \end{cases} \tag{2.72}$$

with

$$\beta_j^{(k)} = \frac{1}{L}\sum_{l=1}^{L}\frac{\mathsf{SNR}\,\Upsilon_j^{(k)*}\,[\mathbf{\Sigma}_w^{(k)}]_{l,j}^2}{(1+\mathsf{SNR}\,[\mathbf{\Sigma}_w^{(k)}]_{l,j}\Upsilon_j^{(k)*})^2}.$$

Step 2: From all the above considerations, using (2.68), (2.71), is follows that:

$$\mathrm{tr}\left\{\mathbf{C_e}(\mathbf{T})\right\} \geq \sum_{k=1}^{K}\mathrm{tr}\{(\mathbf{C}_{\mathbf{h}_w^{(k)}}^{-1} + \mathsf{SNR}\,_T\mathbf{D}^{(k)*})^{-1}\} \tag{2.73}$$

with $\mathbf{D}^{(k)*} = \mathbf{I}_L \otimes \mathbf{\Upsilon}^{(k)*}$ and $\mathbf{\Upsilon}^{(k)*}$ denoting a diagonal matrix with the diagonal elements given by (2.72). A sufficient condition for the equality in (2.73) being achieved is that the training sequences $\mathbf{T}^{(k)}$ are orthogonal and are such that:

$$\mathbf{D}^{(k)*} = \mathbf{I}_L \otimes (\mathbf{U}_\mathrm{T}^{(k)\dagger}\mathbf{T}^{(k)\dagger}\mathbf{T}^{(k)}\mathbf{U}_\mathrm{T}^{(k)}).$$

Putting those considerations together, it follows that the equality in (2.73) is achieved for

$$\mathbf{T} = [\mathbf{T}^{(1)\dagger}, \ldots, \mathbf{T}^{(K)\dagger}]^\dagger$$

such that:

$$\mathbf{T}^{(k)} = \left(\mathbf{e}_{K,k} \otimes \mathbf{U}_\mathrm{T}^{(k)}\sqrt{\mathbf{\Upsilon}^{(k)*}}\right)\mathbf{I}_{KM}^{(N_T)} \quad k = 1, \ldots, K \tag{2.74}$$

with $\mathbf{\Upsilon}^{(k)*}$ denoting a diagonal matrix with the diagonal elements given by (2.72). From (2.74), if follows that the $\tilde{\mathbf{T}}^{(k)*}$ with $k = 1, \ldots, K$ that minimize $\mathrm{tr}\{\mathbf{C_e}\}$ when $\mathbf{C_h}$ have full rank, are:

$$\tilde{\mathbf{T}}^{(k)*} = \mathbf{I}_L \otimes \left(\left(\mathbf{e}_{K,k} \otimes \sqrt{\mathbf{\Upsilon}^{(k)*}}\right)\mathbf{I}_{KM}^{(N_T)}\right)^T, \; k = 1, \ldots, K.$$

This proves Theorem 8 for the case when the covariance matrix, $\mathbf{C_h}$, of the channel vector $\mathbf{h} = \mathrm{vec}\{\mathbf{H}\}$ has full rank.

If the $\mathbf{C}_{\mathbf{h}}$ is not invertible, the same foregoing considerations made for the function in (2.71), apply to the following function

$$\mathrm{tr}\{\mathbf{C}_{\mathbf{h}_w^{(k)}}\} - \mathrm{tr}\left\{\mathbf{C}_{\mathbf{h}_w^{(k)}}\tilde{\mathbf{T}}^{(k)\dagger}\Big(\frac{1}{\mathsf{SNR}_T}\mathbf{I}+\tilde{\mathbf{T}}^{(k)}\mathbf{C}_{\mathbf{h}_w^{(k)}}\tilde{\mathbf{T}}^{(k)\dagger}\Big)^{-1}\tilde{\mathbf{T}}^{(k)}\mathbf{C}_{\mathbf{h}_w^{(k)}}\right\}$$

which, again, is maximized by $\tilde{\mathbf{T}}^{(k)*}$ having the following structure:

$$\tilde{\mathbf{T}}^{(k)*} = \mathbf{I}_L \otimes \left((\mathbf{e}_{K,k} \otimes \sqrt{\mathbf{A}^{(k)*}})\mathbf{I}_{KM}^{(N_T)}\right)^T, \; k = 1, \ldots, K.$$

where $\mathbf{A}$ is an $M \times M$ diagonal matrix satisfying the power constraint $\mathrm{tr}\{\mathbf{A}^{(k)}\} \le MM_T$. Again, solving the convex optimization problem via KKT conditions, we have that

$$\mathbf{A}^{(k)*} = \mathbf{\Upsilon}^{(k)*}$$

with $\mathbf{\Upsilon}^{(k)*}$ defined as in (2.72).

■

Proof of Theorem 9 We want to minimize the determinant of the error covariance matrix (2.44). This is equivalent to maximizing the log-det of the inverse of the error covariance matrix. Finally this is the same as maximizing the mutual information between the channel matrix $\mathbf{H}$ and the observable $\mathbf{Y}_T$, i.e.:

$$I(\mathbf{T}) = I\left(\mathbf{H}^{(1)}, \mathbf{H}^{(2)}, \ldots, \mathbf{H}^{(K)}; \mathbf{Y}_T \middle| \mathbf{T}\right). \tag{2.75}$$

Again, as in the proof of Theorem 8, in proving Theorem 8 we proceed by steps:
Step 1: We start by finding an upper bound for (2.75);
Step 2: We find a set of training sequences that achieve such a upper bound.

Step 1: Using the chain rule we obtain that an upper bound for (2.75) is given by:

$$\begin{aligned} I(\mathbf{T}) &\leq \sum_{k=1}^{K} I\left(\mathbf{H}^{(k)}; \mathbf{Y}_T \middle| \{\mathbf{H}^{(j)}\}_{j\neq k}, \mathbf{T}\right) && (2.76)\\ &= \sum_{k=1}^{K} I\left(\mathbf{H}^{(k)}; \mathbf{H}^{(k)}\mathbf{T}^{(k)} + \mathbf{N} \middle| \mathbf{T}^{(k)}\right) && (2.77)\\ &= \sum_{k=1}^{K} I\left(\mathbf{H}_w^{(k)}; \mathbf{H}_w^{(k)}\hat{\mathbf{T}}^{(k)} + \mathbf{N} \middle| \hat{\mathbf{T}}_k\right) && (2.78)\\ &= \sum_{k=1}^{K} \log\det(\mathbf{I}_{ML} + \tilde{\mathbf{T}}^{(k)}\mathbf{C}_{\mathbf{h}_w^{(k)}}\tilde{\mathbf{T}}^{(k)\dagger}) && (2.79) \end{aligned}$$

where in (2.78) $\hat{\mathbf{T}}_k = \mathbf{U}_{\mathrm{T}}^{(k)}\mathbf{T}^{(k)}$, while in (2.79) $\tilde{\mathbf{T}}^{(k)} = \mathbf{I}_L \otimes \hat{\mathbf{T}}^{(k)^T}$, $\mathbf{h}_w^{(k)} = \mathrm{vec}\{\mathbf{H}_w^{(k)}\}$, and $\mathbf{C}_{\mathbf{h}_w^{(k)}} = \mathbb{E}[\mathbf{h}_w^{(k)}\mathbf{h}_w^{(k)^\dagger}]$ is the LMK-dimensional diagonal covariance matrix of $\mathbf{h}_w^{(k)}$.
A further upper bound can be found maximizing (2.79) with respect to $\mathbf{T}$. Given the fact that the power constraint on $\mathbf{T}$ is a disjoint power constraint over each $\mathbf{T}^{(k)}$ with $k = 1, \ldots, K$, then maximizing (2.68) is equivalent to maximizing each term of the summation

$$g(\mathbf{T}_k) = \log\det(\mathbf{I}_{ML} + \tilde{\mathbf{T}}^{(k)}\mathbf{C}_{\mathbf{h}_w^{(k)}}\tilde{\mathbf{T}}^{(k)\dagger})$$

over all possible $\tilde{\mathbf{T}}^{(k)}$ having the following structure:

$$\tilde{\mathbf{T}}^{(k)} = \mathbf{I}_L \otimes \hat{\mathbf{T}}^{(k)^T}$$

with the constraint that $\frac{1}{MN_T}\mathrm{tr}\left\{\tilde{\mathbf{T}}^{(k)}\tilde{\mathbf{T}}^{(k)^\dagger}\right\} \leq 1$. This is a classical convex optimization problem and applying the classical KKT condition, we obtain that the optimal $\tilde{\mathbf{T}}^{(k)*}$ is given by:

$$\tilde{\mathbf{T}}^{(k)*} = \mathbf{I}_L \otimes \left((\mathbf{e}_{K,k} \otimes \sqrt{\mathbf{\Upsilon}^{(k)*}})\mathbf{I}_{KM}^{(N_T)}\right)^T$$

with $\boldsymbol{\Upsilon}^{(k)*}$ an $M \times M$ diagonal matrix whose diagonal elements are given by:

$$\Upsilon_j^{(k)*} = \begin{cases} 0 & \text{if } \frac{1}{L}\sum_{l=1}^{L}[\boldsymbol{\Sigma}_w^{(k)}]_{l,j} \leq \sum_{j=1}^{M} \frac{(1-\eta_j^{(k)})}{MN_T} \\ \frac{MN_T(1-\eta_j^{(k)})}{\sum_{j=1}^{M}(1-\eta_j^{(k)})} & \text{otherwise} \end{cases} \tag{2.80}$$

with

$$\eta_j^{(k)} = \frac{1}{L}\sum_{l=1}^{L} \frac{1}{1 + \mathsf{SNR}\,[\boldsymbol{\Sigma}_w^{(k)}]_{l,j}\Upsilon_j^{(k)}}.$$

Step 2: Using (2.79), we have that:

$$I(\mathbf{T}) \leq \sum_{k=1}^{K} \log\det(\mathbf{I}_{ML} + \tilde{\mathbf{T}}^{(k)*}\mathbf{C}_{\mathbf{h}_w^{(k)}}\tilde{\mathbf{T}}^{(k)*^{\dagger}}). \tag{2.81}$$

From all the above considerations, it follows that a sufficient condition for the equality in (2.81) being achieved is that the training sequences $\mathbf{T}^{(k)}$ are orthogonal and are such that:

$$\mathbf{I}_L \otimes \left(\mathbf{U}_{\mathrm{T}}^{(k)}\mathbf{T}^{(k)}\right) = \mathbf{I}_L \otimes \left(\sqrt{\boldsymbol{\Upsilon}^{(k)*}}\mathbf{I}_M^{(N_T)}\right)^T$$

Putting those considerations together, it follows that the optimal training set for which the determinant of the error covariance matrix (2.44) is minimized is:

$$\mathbf{T}_k = (\mathbf{e}_{K,k} \otimes \mathbf{U}_{\mathrm{T}}^{(k)}\sqrt{\boldsymbol{\Upsilon}^{(k)*}}) \;\; k = 1,\ldots,K. \tag{2.82}$$

■

Proof of Theorem 10

Let us first prove Theorem 10 for the case that $f(\mathbf{C_e}(\mathbf{T})) = |\mathbf{C_e}(\mathbf{T})|$. In this case we have $N_T < KM$ and $\mathbf{C_h} = \mathbf{I}_{KML}$. Thus, as already underline in the proof of Theorem 9, minimizing $|\mathbf{C_e}(\mathbf{T})|$ is equivalent to maximizing:

$$\max_{\tilde{\mathbf{T}}} \log\left|\mathbf{I}_{KML} + \mathsf{SNR}\,\tilde{\mathbf{T}}^{\dagger}\tilde{\mathbf{T}}\right|. \tag{2.83}$$

over all possible training sequences $\tilde{\mathbf{T}} = \mathbf{I}_L \otimes \mathbf{T}^T$ such that:

$$\mathbf{T} = [\mathbf{T}^{(1)\dagger},\ldots,\mathbf{T}^{(K)\dagger}]^{\dagger}$$

with $\frac{1}{MN_T}\mathrm{tr}\left\{\mathbf{T}^{(k)}\mathbf{T}^{(k)\dagger}\right\} \leq 1$.

From (2.83), it follows that the right eigenvector of $\mathbf{T}$ do not change the value of the objective function as well do not violated the constraints, and then can be arbitrarily chosen.

From the fact that $N_T < MK$, it follows that the non-zero eigenvalue of $\mathbf{T}\mathbf{T}^\dagger$ are at most N_T. Let indicate by $\lambda_i^+ \;\; i = 1, \ldots, N_T$, such non-zero eigenvalues We have that the objective function can be expressed as:

$$\log\left|\mathbf{I}_{KML} + \mathsf{SNR}\,\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}}\right| = L\sum_{i=1}^{N_T}\log\left(1 + \mathsf{SNR}\,\lambda_i^+\right) \tag{2.84}$$

Let us observe that the constraints $\frac{1}{MN_T}\mathrm{tr}\left\{\mathbf{T}_k\mathbf{T}_k^\dagger\right\} \leq 1$ for all $k = 1, \ldots, K$, implies that

$$\mathrm{tr}\left\{\mathbf{T}\mathbf{T}^\dagger\right\} = \sum_{i=1}^{N_T}\lambda_i^+ \tag{2.85}$$

$$\leq KMN_T \tag{2.86}$$

Thus, from (2.85), since the function in (2.84), is schur-concave with respect to $\lambda_1, \ldots, \lambda_{N_T}$, it follows that the maximum of $\log\left|\mathbf{I}_{KML} + \mathsf{SNR}\,\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}}\right|$ is achieved when the N_T non-zero eigenvalues of $\mathbf{T}\mathbf{T}^\dagger$ are given by:

$$\lambda_i^+ = MK \quad i = 1, \ldots, N_\mathrm{T} \tag{2.87}$$

In order to conclude the proof, we need to prove that there exist a training matrix:

$$\mathbf{T} = [\mathbf{T}^{(1)\dagger}, \ldots, \mathbf{T}^{(K)\dagger}]^\dagger$$

with $\frac{1}{MN_T}\mathbf{T}^{(k)}\mathbf{T}^{(k)\dagger} \leq 1$ for all $k = 1, \ldots, K$, such that its non-zero N_T singular values, $\sqrt{\lambda_i^+}$ satisfies:

$$\lambda_i^+ = MK \quad i = 1, \ldots, N_\mathrm{T} \tag{2.88}$$

To this end, denoting by: $\mathbf{\Lambda}_\mathrm{T} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{MK})$ the $MK \times MK$

diagonal matrix defined as follows:

$$\lambda_i = MK \quad i = 1, \ldots, N_T \tag{2.89}$$
$$\lambda_i = 0 \quad i \geq N_T, \tag{2.90}$$
$$\tag{2.91}$$

let us consider an unitary matrix $\mathbf{U}$ of dimension MK such that, for all $h = 1, \ldots, MK$:

$$\sum_{j=1}^{N_T} [\mathbf{U}]_{h,j}^2 = \frac{N_\mathrm{T}}{MK}, \tag{2.92}$$

Thus, we obtain that:

$$\left[\mathbf{U}\mathbf{\Lambda}_\mathrm{T}\mathbf{U}^\dagger\right]_{i,i} = N_\mathrm{T} \quad i = 1, \ldots MK,$$

Property (2.92) is equivalent to say that the partial sum of the square of the first N_T components of each row is equal to $\frac{N_\mathrm{T}}{MK}$. Note that Fourier matrices as well as Hadamard matrices of dimension MK are examples of orthogonal matrices that satisfy such a property. This concludes the proof for the case that $f(\mathbf{C_e}(\mathbf{T})) = |\mathbf{C_e}(\mathbf{T})|$.

Let us now move to the case when $f(\mathbf{C_e}(\mathbf{T})) = \mathrm{tr}\{\mathbf{C_e}(\mathbf{T})\}$. Note that since

$$\mathrm{tr}\{\mathbf{C_e}(\mathbf{T})\} = \mathrm{tr}\{\left(\mathbf{I}_{KML} + \mathsf{SNR}\,\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}}\right)^{-1}\}$$

is schur-convex with respect to the eigenvalues of $\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}}$ or equivalently with respect to the eigenvalues of $\mathbf{T}\mathbf{T}^\dagger$, then arguments similar to the one conducted above can be applied to this second case, from which Theorem 10 follows.

■

Proof of Theorems 11 and 12

Let us consider the following min-max problem

$$\min_{\mathbf{T}} \max_{\mathbf{C_h}} \mathrm{tr}\left\{\mathbf{C_e}(\mathbf{T}, \mathbf{C_h})\right\} \tag{2.93}$$

where $\mathbf{C_e}(\mathbf{T}, \mathbf{C_h})$ denotes the error covariance matrix of the estimation,

$\hat{\mathbf{h}}_{\mathsf{mmse}}$, of the channel vector $\mathbf{h}$, i.e.:

$$\begin{aligned}\mathbf{C_e} &= \mathbb{E}\left[(\mathbf{h}-\hat{\mathbf{h}}_{\mathsf{mmse}})(\mathbf{h}-\hat{\mathbf{h}}_{\mathsf{mmse}})^\dagger\right] \\ &= \mathbf{C_h} - \mathsf{SNR}_T \mathbf{C_h}\tilde{\mathbf{T}}^\dagger \left(\mathbf{I}_{LN_T} + \mathsf{SNR}_T \tilde{\mathbf{T}}\mathbf{C_h}\tilde{\mathbf{T}}^\dagger\right)^{-1} \tilde{\mathbf{T}}\mathbf{C_h}\end{aligned} \tag{2.94}$$

and where the minimum is over all possible training sequences satisfying the power constraint in (2.40), while the maximum is over all possible covariance matrices $\mathbf{C_h}$ such that $\mathrm{tr}\{\mathbf{C_h}\} = 1$. To solve this min-max problem let us consider the following Lemma:

Lemma 2 *Let us consider a min-max problem:*

$$\min_{\mathbf{x}\in D_x} \max_{\mathbf{y}\in D_y} g(\mathbf{x},\mathbf{y}). \tag{2.95}$$

Then, if $\exists\ (\overline{\mathbf{x}},\overline{\mathbf{y}})$*, such that:*

$$g(\overline{\mathbf{x}},\mathbf{y}) \leq g(\overline{\mathbf{x}},\overline{\mathbf{y}})\ \ \forall \mathbf{y} \in D_y \tag{2.96}$$

and

$$g(\mathbf{x},\overline{\mathbf{y}}) \geq g(\overline{\mathbf{x}},\overline{\mathbf{y}})\ \ \forall \mathbf{x} \in D_x \tag{2.97}$$

$(\overline{\mathbf{x}},\overline{\mathbf{y}})$ *is a min-max point.*

Proof:it is given in [34]; for completeness we also present the derivation

$$\min_{\mathbf{x}\in D_x}\left(\max_{\mathbf{y}\in D_y} g(\mathbf{x},\mathbf{y})\right) \leq \max_{\mathbf{y}\in D_y} g(\overline{\mathbf{x}},\mathbf{y}) = g(\overline{\mathbf{x}},\overline{\mathbf{y}}) \tag{2.98}$$

since $\min_{\mathbf{x}\in D_x} h(\mathbf{x}) \leq h(\overline{\mathbf{x}})$. Moreover we have:

$$\min_{\mathbf{x}\in D_x}\left(\max_{\mathbf{y}\in D_y} g(\mathbf{x},\mathbf{y})\right) \geq \min_{\mathbf{x}\in D_x} g(\mathbf{x},\overline{\mathbf{y}}) = g(\overline{\mathbf{x}},\overline{\mathbf{y}}) \tag{2.99}$$

since $\max_{\mathbf{y}\in D_y} g(\mathbf{x},\mathbf{y}) \geq g(\mathbf{x},\overline{\mathbf{y}})\ \forall \mathbf{x} \in D_x$. Then the minimum of $\left(\max_{\mathbf{y}\in D_y} g(\mathbf{x},\mathbf{y})\right)$ is blocked to be $g(\overline{\mathbf{x}},\overline{\mathbf{y}})$ and this value is obtained at saddle point $(\overline{\mathbf{x}},\overline{\mathbf{y}})$.

■

Let us now consider the function $\mathrm{tr}\left\{\mathbf{C_e}(\mathbf{T},\mathbf{C_h})\right\}$; we want to prove that $\overline{\mathbf{T}} =$

$\sqrt{N_T}\mathbf{I}_{KM}^{(N_T)}$ and $\overline{\mathbf{C_h}} = \frac{1}{KML}\mathbf{I}_{KML}$ satisfy Lemma 2, i.e. it is a saddle point. From Corollary 3, we know that:

$$\mathrm{tr}\left\{\mathbf{C_e}(\mathbf{T},\overline{\mathbf{C_h}})\right\} \geq \mathrm{tr}\left\{\mathbf{C_e}(\overline{\mathbf{T}},\overline{\mathbf{C_h}})\right\}. \tag{2.100}$$

Moreover we have that for all positive definite matrix $\mathbf{C_h}$:

$$\begin{aligned}\mathrm{tr}\left\{\mathbf{C_e}(\overline{\mathbf{T}},\mathbf{C_h})\right\} &= \mathrm{tr}\left\{(\mathbf{C_h}^{-1} + \mathsf{SNR}_{\mathrm{T}} N_T \mathbf{I}_{KML})^{-1}\right\} &(2.101)\\ &\leq \mathrm{tr}\left\{\mathbf{C_e}(\overline{\mathbf{T}},\overline{\mathbf{C_h}})\right\}. &(2.102)\end{aligned}$$

where (2.102) follows from the schur-concavity, with respect to the eigenvalue of $\mathbf{C_h}$, of the function $\mathrm{tr}\left\{(\mathbf{C_h}^{-1} + \mathsf{SNR}_{\mathrm{T}} N_T \mathbf{I}_{KML})^{-1}\right\}$; then Theorem 11, when the figure of merit is the mean square error, is proved. Moreover, since the function

$$\mathrm{tr}\left\{\mathbf{C_e}(\overline{\mathbf{T}},\mathbf{C_h})\right\}$$

is a continuous function with respect to $\mathbf{C_h}$, and the set of positive semi-definite matrix with $\mathrm{tr}\{\mathbf{C_h}\} = 1$ is the closure of the set of positive definite matrix with $\mathrm{tr}\{\mathbf{C_h}\} = 1$, (2.101) still old for every matrix with $\mathrm{tr}\{\mathbf{C_h}\} = 1$. Then Theorem 12 is proved. The same technique can also be used in the case that $\mathbf{C_h}$ has full rank and the min-max optimization problem is given by:

$$\min_{\mathbf{T}} \max_{\mathbf{C_h}} |\mathbf{C_e}(\mathbf{T},\mathbf{C_h})| \tag{2.103}$$

where $\mathbf{C_e}(\mathbf{T},\mathbf{C_h})$ is given in (2.94). Moreover we can also prove the results with a different derivation:

$$\begin{aligned}\min_{\mathbf{T}} \max_{\mathbf{C_h}} \log|\mathbf{C_e}| &= \min_{\mathbf{T}} \max_{\mathbf{C_h}} \log|\mathbf{C_e}(\mathbf{T},\mathbf{C_h})|\\ &= \min_{\mathbf{T}} \max_{\mathbf{C_h}} \log|(\mathbf{C_h}^{-1} + \mathsf{SNR}_T \tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}})^{-1}|\\ &\leq \max_{\mathbf{C_h}} \log|\mathbf{C_h}| +\\ &\quad \min_{\mathbf{T}} \max_{\mathbf{C_h}} \log|(\mathbf{I}_{KML} + \mathsf{SNR}_T \mathbf{C_h}\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}})^{-1}|.\end{aligned}$$

Let us note that the maximum of $\log|\mathbf{C_h}|$ is achieved when $\mathbf{C_h} = \frac{1}{KML}\mathbf{I}_{KML}$, for the schur concavity of the log-det function. Moreover notice that:

$$\log|(\mathbf{I}_{KML} + \mathsf{SNR}_T\mathbf{C_h}\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}})^{-1}| =$$

$$-\sum_{i=1}^{KML}\log(1+\mathsf{SNR}_T\lambda_i(\mathbf{C_h}\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}})) \quad (2.104)$$

$$\leq -\log(1+\mathsf{SNR}_T\mathrm{tr}\{\mathbf{C_h}\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}}\}) \quad (2.105)$$

$$\leq -\log(1+\mathsf{SNR}_T\mathrm{tr}\{\mathbf{C_h}\}\lambda_{\mathsf{min}}(\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}})). \quad (2.106)$$

Thus, as to the max-part, we have that (2.104) is maximized by a $\mathbf{C_h}$ whose eigenvectors coincide with the eigenvector or the eigenvectors (in the case of multiplicity larger than 1) corresponding the minimum eigenvector of $\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}}$

As to the min-part, starting from (2.106), we have that minimizing (2.104) is equivalent to maximizing

$$\lambda_{\mathsf{min}}(\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}})$$

which is Schur-concave function of the eigenvalues of $\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}}$ and it is, thus, again maximized by a $\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}}$ such that for all for all $i = 1,\ldots,KML$, $\lambda_i(\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}})$ equals a constant satisfying the power constraint $\mathrm{tr}\{\tilde{\mathbf{T}}^\dagger\tilde{\mathbf{T}}\} = N_T LKM$. Thus, (2.104) is minimized by $\mathbf{T} = \sqrt{N_T}\mathbf{I}_{KM}^{(N_T)}$, from which Theorem 11, for the case when the figure of merit is the determinant of the error covariance matrix, follows.

■

Proof of Theorem 13

In order to prove Theorem 13, it is enough to verify that the hypotheses $\mathcal{H}.1$ and $\mathcal{H}.2$ of Theorem 6 are satisfied.

To this end notice that, by assumption, $\mathbf{H}$ is modeled according to (1.6) i.e.

$$\mathbf{H}^{(k)} = \mathbf{U}_{\mathrm{R}}{}^{(k)}\mathbf{H}_w^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)^\dagger} \quad (2.107)$$

with $\mathbf{H}_w^{(k)}$ a zero-mean $L\times M$ Gaussian random matrix of independent elements. From (2.107), since the optimal training sequences are given either by (2.46) or by (2.48), we have that:

$$\hat{\mathbf{H}}^{(k)} = \mathbb{E}_{\mathbf{S}}[\mathbf{H}^{(k)}] = \mathbf{U}_{\mathrm{R}}{}^{(k)}\mathbb{E}_{\mathbf{Y}_T}[\mathbf{H}_w^{(k)}]\mathbf{U}_{\mathrm{T}}{}^{(k)^\dagger}$$

where $\hat{\mathbf{H}}_w^{(k)} = \mathbb{E}_{\mathbf{Y}_T}[\mathbf{H}_w^{(k)}]$ is a zero mean Gaussian random matrix with independent elements, then hypotheses $\mathcal{H}.1$ holds.

Furthermore, $\mathbf{\Omega}_{\mathcal{U}}$ doesn't depend from $\mathbf{S}$, since $\mathbf{Z}$ is statistically independent from $\mathbf{S}$, and it is easy to prove, after some algebraic manipulations, that:

$$\begin{aligned}\tilde{\mathbf{Z}}^{(k)} &= \mathbf{Z}^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)} \\ &= \mathbf{U}_{\mathrm{R}}^{(k)}\mathbf{Z}_{w}^{(k)}\end{aligned}$$

with $\mathbf{Z}_w^{(k)}$ the estimation error matrix of $\mathbf{H}_w^{(k)}$, are random matrices with independent columns, each column having entries whose joint distribution is symmetric with respect to zero. From this, it follows that $\mathcal{H}.2$ also holds. Thus, the hypotheses of Theorem 6 are satisfied and Theorem 13 follows immediately.

■

Proof of Theorem 14 In proving Theorem 14, we proceed by steps:
Step 1: we prove that (2.58) holds;
Step 2: we prove that $\{\mathbf{P}_k^* = \frac{\varrho}{M}\mathbf{I}_M \quad k = 1,\dots,K\}$ is a Nash equilibrium point.

Step 1: By assumption we have that $\mathbf{H}$ is modeled according to (1.6), i.e:

$$\mathbf{H}^{(k)} = \mathbf{U}_{\mathrm{R}}^{(k)}\mathbf{H}_{w}^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)^\dagger}$$

where for all $k = 1,\dots,K$, $\mathbf{\Sigma}_w^{(k)} = \boldsymbol{\lambda}_{\mathrm{R}}^{(k)}\mathbf{1}_M{}^\dagger$. This is equivalent to assume that $\mathbf{U}_{\mathrm{T}}^{(k)^\dagger} = \mathbf{I}_M$, from which it follows, using Theorem 13 that $\mathbf{P}_k^*$ has to be diagonal. Furthermore, from the assumptions done, we have that $\{\mathbf{Z}_w^{(k)}\}, \quad k = 1,\dots,K$, is a set of random matrices whose columns are independent and whose rows have independent and identically distributed entries along each row. Thus,

$$\begin{aligned}\mathbf{\Omega}_{\mathcal{U}} &= \mathbb{E}\left[\mathbf{Z}\mathbf{P}\mathbf{Z}^\dagger\right] = \sum_{k=1}^{K}\mathbb{E}\left[\mathbf{Z}^{(k)}\mathbf{P}_k\mathbf{Z}^{(k)^\dagger}\right] \\ &= \sum_{k=1}^{K}\mathbf{U}_{\mathrm{R}}^{k}\mathbb{E}\left[\mathbf{Z}_w^{(k)}\mathbf{P}_k\mathbf{Z}_w^{(k)^\dagger}\right]\mathbf{U}_{\mathrm{R}}^{k\,\dagger} \qquad (2.108)\end{aligned}$$

depends from $\mathbf{P} = \mathrm{diag}(\mathbf{P}_1,\dots,\mathbf{P}_K)$ only through $\{\mathrm{tr}\{\mathbf{P}_1\},\dots,\mathrm{tr}\{\mathbf{P}_K\}\}$.

From all above, it follows that the lower bound on the sum rate, $I_{\mathsf{Lower}}(\mathbf{P})$, given in (1.37) can be re-written as follows:

$$\mathbb{E}\left[\log\left|\mathbf{I}_L+\mathbf{Q}\sum_{k=1}^{K}\boldsymbol{\Theta}_{\mathrm{R}}^{(k)}\mathbf{W}^{(k)}\mathbf{P}^{*}{}_{k}\mathbf{W}^{(k)\dagger}\boldsymbol{\Theta}_{\mathrm{R}}^{(k)\dagger}\right|\right], \tag{2.109}$$

where $\boldsymbol{\Theta}_{\mathrm{R}}^{(k)}=\mathbf{U}_{\mathrm{R}}^{(k)}\sqrt{\mathrm{diag}(\hat{\boldsymbol{\lambda}}_{\mathrm{R}}^{(k)})}$, with $\hat{\boldsymbol{\lambda}}_{\mathrm{R}}^{(k)}$ the variance vector of an arbitrary column of $\hat{\mathbf{H}}_w^{(k)}$,

$$\mathbf{Q}=\mathsf{SNR}\left(\mathbf{I}_L+\mathsf{SNR}\sum_{k=1}^{K}\mathrm{tr}\{\mathbf{P}_k\}\mathbb{E}\left[\mathbf{Z}^{(k)}\mathbf{Z}^{(k)^{\dagger}}\right]\right)^{-1}$$

and $\{\mathbf{W}^{(k)}\}$ for $k=1,\ldots,K$ denote a set of Gaussian random matrices independent across k with i.i.d. zero-mean Gaussian entries. Note that, since $\{\mathbf{W}^{(k)}\}$ are i.i.d. zero-mean Gaussian matrices, (2.109) and consequently the lower bound on the sum rate, $I_{\mathsf{Lower}}(\mathbf{P})$, is invariant to circular shift of the elements of $\mathbf{P}_k$ for all $k=1,\ldots,K$, from which it follows, using Jensen's inequality, that the maximum of (2.109) is achieved when for all $k=1,\ldots,K$, the diagonal entries of $\mathbf{P}_k$ are equal i.e.

$$\mathbf{P}_k^{*}=\frac{\varrho_k}{M}\mathbf{I}_M$$

with ϱ_k subject to the power constraint $\varrho_k\leq 1$.

Step 2. Let:

$$\bar{\mathbf{p}}=[\underbrace{1,1,\ldots 1}_{\pi(k-1)},\varrho_{\pi(k)},\underbrace{1\ldots,1}_{K-\pi(k)}]$$

and

$$\mathbf{p}^{*}=[\underbrace{1,1,\ldots 1,1,1\ldots,1}_{K}]$$

with $\varrho_{\pi(k)}\in\,]0,1]$. We want to show that $\forall k=1,\ldots,K$:

$$\begin{aligned}L_k(\bar{\mathbf{p}}) &= \mathbb{E}\left[\log\left|\mathbf{I}+\varrho_{\pi(k)}\hat{\mathbf{H}}^{(\pi(k))}\hat{\mathbf{H}}^{(\pi(k))^{\dagger}}\Psi_{\pi}^{-1}(\bar{\mathbf{p}})\right|\right]\\ &\leq L_k(\mathbf{p}^{*})\\ &= \mathbb{E}\left[\log\left|\mathbf{I}+\hat{\mathbf{H}}^{(\pi(k))}\hat{\mathbf{H}}^{(\pi(k))^{\dagger}}\Psi_{\pi}^{-1}(\mathbf{p}^{*})\right|\right]\end{aligned} \tag{2.110}$$

where $\varrho_{\pi(k)} \in]0,1]$,

$$\Psi_\pi(\bar{\mathbf{p}}) = \varrho_{\pi(k)} \mathbb{E}\left[\mathbf{Z}^{(\pi(k))}\mathbf{Z}^{(\pi(k))^\dagger}\right] + \Xi_\pi,$$

$$\Psi_\pi(\mathbf{p}^*) = \mathbb{E}\left[\mathbf{Z}^{(\pi(k))}\mathbf{Z}^{(\pi(k))^\dagger}\right] + \Xi_\pi,$$

and

$$\Xi_\pi = \mathbf{\Omega}_{\overline{\pi(k)}} + \mathbf{\Gamma}_{\{\pi(k+1),\ldots,\pi(K)\}} + \tfrac{1}{\mathsf{snr}}\mathbf{I}.$$

In order to show (2.110), notice that:

$$\frac{1}{\varrho_{\pi(k)}}\Xi_\pi \succeq \Xi_\pi.$$

Consequently:

$$\left(\mathbb{E}\left[\mathbf{Z}^{(k)}\mathbf{Z}^{(k)^\dagger}\right] + \frac{1}{\varrho_{\pi(k)}}\Xi_\pi\right)^{-1} \preceq \left(\mathbb{E}\left[\mathbf{Z}^{(k)}\mathbf{Z}^{(k)^\dagger}\right] + \Xi_\pi\right)^{-1}.$$

Now recalling that for every $\mathbf{A} \succeq$, $\mathbf{B} \succeq 0$ and $\mathbf{C} \succeq 0$:

$$\log\left|\mathbf{I} + \mathbf{A}\mathbf{B}\right| \geq \log\left|\mathbf{I} + \mathbf{C}\mathbf{B}\right| \ \text{if} \mathbf{A} \succeq \mathbf{C}, \tag{2.111}$$

it follows that (2.110) holds for every $k = 1,\ldots,K$ and decoding order and so we obtain the thesis.

■

Proof of Theorem 15 By assumption we have that $\mathbf{H}$ is modeled according to (1.6), i.e:

$$\mathbf{H}^{(k)} = \mathbf{U}_{\mathrm{R}}^{(k)}\mathbf{H}_w^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)^\dagger} \tag{2.112}$$

where for all $k = 1,\ldots,K$, $\mathbf{\Sigma}_w^{(k)} = \frac{1}{L}\mathbf{1}_L\boldsymbol{\lambda}_{\mathrm{T}}^{(k)\dagger}$, which is equivalent to assume that $\mathbf{U}_{\mathrm{R}}^{(k)^\dagger} = \mathbf{I}_L$. Just from (2.112), through the same steps followed in the proof of Theorem 13, it can be proved that the hypotheses $\mathcal{H}.1$-$\mathcal{H}.2$ of Theorem 6 hold, from which it follows that:

$$\mathbf{P}_k^* = \mathbf{U}_{\mathrm{T}}^{(k)}\mathbf{\Lambda}_k^*\mathbf{U}_{\mathrm{T}}^{(k)\dagger} \tag{2.113}$$

From (2.113) and from the fact that $\mathbf{U}_{\mathrm{R}}^{(k)^\dagger} = \mathbf{I}_L$, it follows that

$\{\mathbf{Z}_w^{(k)}\}, \quad k = 1, \ldots, K$, is a set of random matrices whose columns are independent and each column has independent and identically distributed entries. Consequently:

$$\begin{aligned}
\boldsymbol{\Omega}_{\mathcal{U}} &= \mathbb{E}\left[\mathbf{Z}\mathbf{P}_{\mathcal{U}}^{*}\mathbf{Z}^{\dagger}\right] &(2.114)\\
&= \sum_{k=1}^{K}\mathbb{E}\left[\mathbf{Z}^{(k)}\mathbf{P}_{k}^{*}\mathbf{Z}^{(k)^{\dagger}}\right]\\
&= \sum_{k=1}^{K}\mathbb{E}\left[\mathbf{Z}_w^{(k)}\mathbf{U}_{\mathrm{T}}^{(k)}\mathbf{P}_{k}^{*}\mathbf{U}_{\mathrm{T}}^{(k)\dagger}\mathbf{Z}_w^{(k)^{\dagger}}\right]\\
&= \sum_{k=1}^{K}\mathbb{E}\left[\mathbf{Z}_w^{(k)}\boldsymbol{\Lambda}_{k}^{*}\mathbf{Z}_w^{(k)^{\dagger}}\right]\\
&= \left(\sum_{k=1}^{K}\sum_{i=1}^{M}\boldsymbol{\sigma}_{w_i}^{(k)}[\boldsymbol{\Lambda}_{k}^{*}]_{i,i}\right)\mathbf{I}_L &(2.115)
\end{aligned}$$

Thus, we can apply Theorem 7, from which Theorem 15 follows immediately.

■

Chapter 3

Asymptotic Behavior of MIMO MAC with Partial CSI

In this chapter, the performance of a MIMO MAC with partial CSI in asymptotic regimes, is analized. In particular, the low-SNR regime is described, analyzing both the minimum energy per bit required for reliable communication and the multiaccess slope region. Moreover the high SNR behavior is studied, evaluating the high SNR slope, i.e. the behavior of the sum rate with respect to $\log \mathsf{SNR}$ function. In the following sections, firstly the most important parameters used for the characterization of the system in these two asymptotic regimes are introduced, and they are particularized in the case of partial CSI.

3.1 Low SNR Characterization

As discussed in [35], the key performance measures in the low-SNR regime for MIMO point-to-point channel are $\frac{E_b}{N_0}_{\mathsf{min}}$, which is the minimum energy per information bit required to convey any positive rate reliably, and $\mathbf{S}_0$, which is the capacity slope therein in bits/s/Hz/(3 dB). These two quantity are defined [1] as:

$$\frac{E_b}{N_0}_{\mathsf{min}} = \inf_{\mathsf{SNR}} \frac{\mathsf{SNR}}{C(\mathsf{SNR})} \tag{3.1}$$

$$= \lim_{\mathsf{SNR} \downarrow 0} \frac{\mathsf{SNR}}{C(\mathsf{SNR})} \tag{3.2}$$

[1] In this section the capacity is assumed normalized to the number of receiving antennas, i.e. the spatial dimension , and $\mathsf{SNR} = \frac{\mathbb{E}[||\mathbf{x}||^2]}{N_0 L}$, as discussed in [35].

while the slope $\mathbf{S}_0$ of spectral efficiency in b/s/Hz/(3 dB) at the point $\frac{E_b}{N_0}_{\mathsf{min}}$ is given by:

$$\mathbf{S}_0 \triangleq \lim_{\frac{E_b}{N_0} \downarrow \frac{E_b}{N_0}_{\mathsf{min}}} \frac{C(\frac{E_b}{N_0})}{\frac{E_b}{N_0}\big|_{dB} - \frac{E_b}{N_0}_{\mathsf{min}}\big|_{dB}} 3dB \tag{3.3}$$

where

$$C\left(\frac{E_b}{N_0}\right) = C(\mathsf{SNR}),$$

with SNR such that

$$\frac{E_b}{N_0} = \frac{\mathsf{SNR}}{C(\mathsf{SNR})},$$

is the spectral efficiency. Let us observe that $\frac{E_b}{N_0}_{\mathsf{min}}$ and $\mathbf{S}_0$ determine the first-order behavior of the spectral efficiency as function of $\frac{E_b}{N_0}$ (in dB) via:

$$C\left(\frac{E_b}{N_0}\right) = \mathbf{S}_0 \frac{\frac{E_b}{N_0}\big|_{dB} - \frac{E_b}{N_0}_{\mathsf{min}}\big|_{dB}}{3dB} + \epsilon \tag{3.4}$$

where $\epsilon = o\left(\frac{E_b}{N_0} - \frac{E_b}{N_0}_{\mathsf{min}}\right)$.

Let us now consider the multiple access channel. Without loss of generality we consider the case of two users, since the generalization to a greater number of users is very straightforward. As described in [36] and [37], in this case the fundamental limits of interest, in the low-SNR regime, are described as function of the ratio $\theta = \frac{R_1}{R_2}$ with which the information rate of the first user, R_1, and the information rate of the second user R_2, go to zero. Specifically, indicating with $\mathbf{C}(\mathsf{SNR}_1, \mathsf{SNR}_2)$ the capacity rate region of the considered MIMO MAC channel, when the first user uses a power P_1=$\mathsf{SNR}_1 N_0 L$ and the second user uses a power P_2=$\mathsf{SNR}_2 N_0 L$, we have that:

$$\left(\frac{E_1}{N_0}_{\mathsf{min}}^{(\theta)}, \frac{E_2}{N_0}_{\mathsf{min}}^{(\theta)}\right) \triangleq \lim_{(\mathsf{SNR}_1, \mathsf{SNR}_2) \downarrow (0,0)} \left(\frac{\mathsf{SNR}_1}{R_1}, \frac{\mathsf{SNR}_2}{R_2}\right) \tag{3.5}$$

with the constraint that $(R_1, R_2) \in \mathbf{bd}\{\mathbf{C}(\mathsf{SNR}_1, \mathsf{SNR}_2)\}$ and that $\frac{R_1}{R_2}$=θ, where $\mathbf{bd}\{\mathbf{A}\}$ is the boundary set of the set $\mathbf{A}$. Moreover, we define the achievable segment of rates for user 1, when $\frac{R_1}{R_2} = \theta$ and for fixed $\frac{E_1}{N_0}$ and $\frac{E_2}{N_0}$ as:

$$\mathbf{R}_\theta \left(\frac{E_1}{N_0}, \frac{E_2}{N_0} \right) \triangleq \{ R_1 \geq 0 : \exists (\mathsf{SNR}_1, \mathsf{SNR}_2) \tag{3.6}$$

$$s.t. \left(R_1, \frac{R_1}{\theta} \right) \in \mathbf{C} \left(\mathsf{SNR}_1, \mathsf{SNR}_2 \right), \tag{3.7}$$

$$\left. \frac{\mathsf{SNR}_1}{R_1} = \frac{E_1}{N_0}, \frac{\mathsf{SNR}_2}{R_2} = \frac{E_2}{N_0} \right\} \tag{3.8}$$

The multiaccess slope region is, then, defined as the set of slope pairs $\mathbf{S}(\theta)$ that result from:

$$\mathbf{S}_1(\theta) \triangleq \lim_{\frac{E_1}{N_0} \downarrow \frac{E_1}{N_0}_{\min}} \frac{R_1}{\frac{E_1}{N_0}\big|_{dB} - \frac{E_1}{N_0}_{\min}(\theta)\big|_{dB}} 10 \log_{10} 2$$

$$\mathbf{S}_2(\theta) \triangleq \frac{1}{\theta} \lim_{\frac{E_2}{N_0} \downarrow \frac{E_2}{N_0}_{\min}} \frac{R_1}{\frac{E_2}{N_0}\big|_{dB} - \frac{E_2}{N_0}_{\min}(\theta)\big|_{dB}} 10 \log_{10} 2$$

for $\left(R_1, \frac{R_1}{\theta} \right)$ vanishing with $\frac{E_1}{N_0} \downarrow \frac{E_1}{N_0}_{\min}$ and $\frac{E_2}{N_0} \downarrow \frac{E_2}{N_0}_{\min}$, respecting the membership

$$R_1 \in \mathbf{R}_\theta \left(\frac{E_1}{N_0}, \frac{E_2}{N_0} \right).$$

3.1.1 Low SNR Analysis of MIMO MAC with Partial CSI

In this subsection, the analysis for low SNR of the MIMO MAC where only a channel-realization statistic $\mathbf{S}$ is available at the receiver, is considered. With no loss of generality we focus on the case of two-users MIMO MAC. Before analyzing the multi-user setting, some results from [35] [38] [39] , for the single user scenario[2] are here reported.

Theorem 16 *Let us consider a non-coherent point-to-point MIMO channel,*

$$\mathbf{y} = \sqrt{\mathsf{SNR}}\,\mathbf{H}\mathbf{x} + \mathbf{n}, \tag{3.9}$$

where $\mathbf{H}$ *is an* $L \times M$ *random matrix with mean* $\overline{\mathbf{H}}$ *and such that* $E[||\mathbf{H}||^{4+\alpha}] \leq \infty$ *for some* $\alpha > 0$, $\mathbf{x}$ *is the M-dimensional input vector with mean* $\overline{\mathbf{x}}$, *subject to the constraint* $\mathrm{tr}\{\mathbf{x}\mathbf{x}^\dagger\} \leq \mathbf{1}$ *and* $\mathbf{n}$ *is the L-dimensional*

[2] In the following we assume that each user only knows his own channel probabilistic characterization (Statistical CSI).

additive circularly symmetric zero mean Gaussian noise with i.i.d. entries, satisfying $\mathrm{tr}\{\mathbf{n}\mathbf{n}^\dagger\} = \mathbf{1}$..
Then we have [35]:

$$\frac{dC(\mathsf{SNR})}{d\mathsf{SNR}}\Big|_{\mathsf{SNR}=0} = \dot{C}(0) = \lambda_{\mathsf{max}}(\mathbb{E}[\mathbf{H}^\dagger\mathbf{H}]), \tag{3.10}$$

or, equivalently,

$$\frac{E_b}{N_0}_{\mathsf{min}} = \frac{\log_e 2}{\dot{C}(0)} = \frac{\log_e 2}{\lambda_{\mathsf{max}}(\mathbb{E}[\mathbf{H}^\dagger\mathbf{H}])}. \tag{3.11}$$

Further, on-off signaling is first order optimal, i.e. there exists an on-off signal such that (3.10) is achieved [26].
Moreover for a practical signaling, i.e. a proper-complex random vector [3] $\mathbf{x}$*, satisfying the following condition:*

$$\mathcal{P}\left\{||\mathbf{x}|| > \delta\right\} \leq \exp\left\{-\delta\nu\right\} \tag{3.12}$$

for all $\delta \geq \delta_0$*, where* $\delta_0 > 0$ *and* $\nu > 0$ *are some positive constants, we have [38]:*

$$\frac{dI(\mathsf{SNR})}{d\mathsf{SNR}}\Big|_{\mathsf{SNR}=0} = \dot{I}(0) = L\mathbb{E}\left[||\overline{\mathbf{H}}(\mathbf{x}-\overline{\mathbf{x}})||^2\right] \tag{3.13}$$

or equivalently the transmitted energy per information bit relative to the noise spectral level, for low SNR*, is given by,*

$$\frac{E_b}{N_0} = \frac{\log_e 2\, L}{\dot{I}(0)} = \frac{\log_e 2}{\mathbb{E}\left[||\overline{\mathbf{H}}(\mathbf{x}-\overline{\mathbf{x}})||^2\right]}. \tag{3.14}$$

Furthermore, particularizing the statistical characterization of $\mathbf{H}$ *to be an* $L \times M$ *circularly symmetric zero mean Gaussian random matrix with unit variance i.i.d. elements, we can relax the condition (3.12), to the following one [39]:*

- $\partial f(\mathbf{x}, \mathsf{SNR})/\partial\mathsf{SNR}$ *exists in* $\mathsf{SNR} = 0$*, where* $f(\mathbf{x}, \mathsf{SNR})$ *is the family of probability distribution function of the random vector* $\mathbf{x}$*, parameterized in* SNR*,*

- $\lim_{\mathsf{SNR}\to 0} \mathsf{SNR}\, \mathbb{E}[||\mathbf{x}||^4] = \mathbf{0}$*,*

[3] A complex valued random vector is said proper complex if $\mathbb{E}\left[\mathbf{x}\mathbf{x}^T\right] = \mathbb{E}\left[\mathbf{x}\right]\mathbb{E}\left[\mathbf{x}\right]^T$, [38].

and obtain:

$$\left.\frac{dI(\mathsf{SNR})}{d\mathsf{SNR}}\right|_{\mathsf{SNR}=0} = \dot{I}(0) = 0 \tag{3.15}$$

i.e. for low SNR *the required energy per bit goes to infinity.*

From the previous results we can conclude that the signals satisfying (3.12), assuming that the channel $\mathbf{H}$ has zero expected value, are power inefficient in low-SNR regime, since the required energy per bit goes to infinity.
Hereafter a point-to-point MIMO channel with partial CSI is characterized. First of all, some definitions from [35] are recalled:

Definition 2 *An input distribution parameterized by* SNR, $\mathbf{x}_{\mathsf{SNR}}$ *is first order optimal if:*
$\mathbb{E}\left[||\mathbf{x}_{\mathsf{SNR}}||^2\right] = \mathsf{SNR}\, N_0 L$ *and it achieves* $\frac{E_b}{N_0}_{\mathsf{min}}$*, i.e.:*

$$\lim_{\mathsf{SNR}\to 0} \frac{I(\mathbf{x};\mathbf{y}|\mathbf{S})}{L\,\mathsf{SNR}} = \dot{C}(0) \tag{3.16}$$

where $\mathbf{S}$ *is an arbitrary statistic of the channel* $\mathbf{H}$ *available at the receiver.*

Definition 3 *An input distribution parameterized by* SNR, $\mathbf{x}_{\mathsf{SNR}}$ *is said a Flash Signaling if, for all* $\delta > 0$*,*

$$\lim_{\mathsf{SNR}\downarrow 0} \frac{\mathbb{E}\left[||\mathbf{x}_{\mathsf{SNR}}||^2 1\{||\mathbf{x}_{\mathsf{SNR}}|| > \delta\}\right]}{\mathbb{E}\left[||\mathbf{x}_{\mathsf{SNR}}||^2\right]} = 1 \tag{3.17}$$

where $1\{.\}$ *is the indicator function.*

Essentially a Flash Signaling can be viewed as the mixture of two probability distribution one that asymptotically concentrates all its mass at 0 and the other that migrates to infinity. We can, then, prove the following result:

Theorem 17 *Assume that neither the receiver nor the transmitter know* $\mathbf{H}$*, but the receiver has at its disposal a statistic* $\mathbf{S}$ *of the channel* $\mathbf{H}$*. If*

$$\lambda_{min}\left(\mathbb{E}_{\mathbf{S}}\left[\mathbf{Z}^\dagger\mathbf{Z}\right]\right) \geq \alpha > 0 \quad a.s.$$

then $\mathbf{x}_{\mathsf{SNR}}$ *is first order optimal if and only if it is a Flash Signaling and*

$$\lim_{\mathsf{SNR}\to 0} \frac{\mathbb{E}\left[||\mathbf{H}\mathbf{x}_{\mathsf{SNR}}||^2\right]}{\mathbb{E}\left[||\mathbf{x}_{\mathsf{SNR}}||^2\right]} = \lambda_{max}\left(\mathbb{E}\left[\mathbf{H}^\dagger\mathbf{H}\right]\right) \tag{3.18}$$

Proof: *See Appendix C* ***Proof of Theorem 17***

This generalizes the result given in [35] for incoherent reception.

The following result, characterize, up to the second order[4] the capacity behavior of a MIMO point-to-point channel with imperfect channel state information:

Theorem 18 *Assume that neither the receiver nor the transmitter know* $\mathbf{H}$*, but the receiver has at its disposal a statistic* $\mathbf{S}$ *of the channel* $\mathbf{H}$*. If*

$$\lambda_{min}\left(\mathbb{E}_{\mathbf{S}}\left[\mathbf{Z}^{\dagger}\mathbf{Z}\right]\right) \geq \alpha > 0 \quad a.s.,$$

then,

$$\mathbf{S}_0 = 0$$

Proof: See Appendix C **Proof of Theorem 18**

Moving now to the multi-access scenario with imperfect channel state information, the transmitted and received energy per information bit relative to the noise spectral level, of user $k = 1, 2$, are defined by:

$$\frac{E_k}{N_0} = \frac{\mathsf{SNR}_k}{R_k}, \qquad \frac{E_k}{N_0}^r = \frac{\mathsf{SNR}_k}{R_k} g_k, \tag{3.19}$$

with g_k denoting the channel gain of the k-th user, defined as

$$\frac{\mathbb{E}\left[||\mathbf{H}^{(k)}\mathbf{x}^{(k)}||^2\right]}{\mathbb{E}\left[||\mathbf{x}^{(k)}||^2\right]}.$$

In general the maximum achievable channel gain is calculated over all possible choices for the input and it depends on the knowledge available at the transmitter, (see [35]). If the transmitter does not know the channel but it knows its distribution, then the maximum channel gain is given by $G_k = \lambda_{\mathsf{max}}(\mathbb{E}[\mathbf{H}^{(k)\dagger}\mathbf{H}^{(k)}])$, [35].

As discussed in Section 3.1, the two fundamental limits of interest in this section are the minimum energy per information bit, $\frac{E_k}{N_0}_{\mathsf{min}}$, which is obtained with asymptotically low power, and the multiaccess slope region. Let R_1 and R_2 go to zero while maintaining a fixed ratio $\theta = \frac{R_1}{R_2}$. The next theorems show how the performance measures of interest (i.e. the multiaccess minimum

---

[4]As shown in [35], $\mathbf{S}_0$ is related to second derivative of the capacity, in $\mathsf{SNR} = 0$.

energy per bit and optimum multiaccess slope region) do not depend on θ, for a MAC channel with partial CSI.

Theorem 19 *For all $\theta = R_1/R_2$, the minimum energies per information bit for a MIMO MAC, with a statistic $\mathbf{S}$ of the channel realization at receiver, are equal to:*

$$\frac{E_1}{N_0}_{\min} = \frac{\log_e 2}{\lambda_{\max}(\mathbb{E}[\mathbf{H}^{(1)\dagger}\mathbf{H}^{(1)}])}. \quad (3.20)$$

and

$$\frac{E_2}{N_0}_{\min} = \frac{\log_e 2}{\lambda_{\max}(\mathbb{E}[\mathbf{H}^{(2)\dagger}\mathbf{H}^{(2)}])}. \quad (3.21)$$

Furthermore, (3.20) is achieved by on-off signalling over orthogonal directions.

Proof: See Appendix C **Proof of Theorem 19**.

Theorem 20 *Let the rates vanish while keeping $\theta = R_1/R_2$. If the statistic $\mathbf{S}$ is such that:*

$$\lambda_{min}\left(\mathbb{E}_{\mathbf{S}}\left[\mathbf{Z}^{(k)^\dagger}\mathbf{Z}^{(k)}\right]\right) \geq \alpha_k > 0 \quad a.s., \quad k = 1, 2 \quad (3.22)$$

then the multiaccess slope region of MIMO MAC, with a statistic $\mathbf{S}$ of the channel realization at receiver, is:

$$\mathbf{S}(\theta) = (0, 0) \quad (3.23)$$

and is achieved by on-off signalling and TDMA.

Proof: See Appendix C **Proof of Theorem 20**.

From the previous Theorems, we can observe that, under the assumption (3.22), using on-off signalling and TDMA for the two users, we achieve the optimal performance in low SNR regime, independently from the available statistic $\mathbf{S}$. Since the availability of a statistic $\mathbf{S}$ to the receiver requires that the users spend some of their resources in terms of power and available dimensions, the optimal strategy, under the Hypothesis of Theorem 20, is to consider absence of $\mathbf{S}$, i.e. an incoherent reception. Moreover, from Theorem 17, all signals that are not flash signaling are power inefficient since they achieve an

higher $\frac{E_b}{N_0}$ with respect to the minimum one.
To conclude the low-SNR analysis, let us consider a global description of the system. Specifically, assuming that all the users transmit at the same power, it is possible to define a system energy per bit:

$$\frac{E_b}{N_0} = \frac{\mathsf{SNR}}{C^{sum}(\mathsf{SNR})} \tag{3.24}$$

where $C^{sum}(\mathsf{SNR})$, is the sum-rate of the system. It is easy to show, from the above Theorems, that

$$\frac{E_b}{N_0}_{\mathsf{min}} = \frac{1}{\sum_{K=1}^{2}\left(\frac{E_k}{N_0}_{\mathsf{min}}\right)^{-1}} \tag{3.25}$$

i.e. it is equal to the harmonic mean of the individual minimum energy per bit. Moreover, under the Hypotheses of Theorem 20, the wideband slope of the sum rate, defined as in (3.3), is equal to zero.
The case of statistic $\mathbf{S}$ obtained through a training phase, as described in Section2.2, is now considered.

Corollary 4 *Let the rates vanish while keeping $\theta = R_1/R_2$. Assuming that $\mathbf{C}_{\mathbf{h}^{(k)}}$, $k = 1, 2$ are full rank, then the multiaccess slope region of MIMO MAC with a channel-realization statistic obtained through a training phase, with arbitrary but finite SNR_T, is:*

$$\mathbf{S}(\theta) = (0, 0) \tag{3.26}$$

Proof: See Appendix C **Proof of Corollary 4**.
Then we can conclude that in low SNR regime, for a training based system, the optimal signaling strategy is a TDMA access to the channel, with on-off signaling sent from the users in the data phase. Moreover, it is evident that to improve the data rate at to low-SNR regime, we have not to consider the training phase, that only consume the available resources. Precisely, let $1-\alpha$ be the fraction of the total available power, used for the training phase, we can define the transmitted energy per information bit relative to the noise spectral level, with respect to the whole power available at the transmitter, i.e. including also the power spent for the training, obtaining:

$$\frac{E_k}{N_0} = \frac{1}{\alpha}\frac{E_k}{N_0}\bigg|^{d} \tag{3.27}$$

and consequently,

$$\frac{E_k}{N_0}_{\mathsf{min}} = \frac{1}{\alpha} \frac{E_k}{N_0}\bigg|^{d}_{\mathsf{min}} \tag{3.28}$$

where $\frac{E_k}{N_0}\Big|^{d}_{\mathsf{min}}$, is the minimum energy for information bit, characterized in the previous Theorem, with respect to the only power available in the transmission phase. Then the global minimum $\frac{E_k}{N_0}_{\mathsf{min}}$ is achieved for $\alpha = 1$, i.e. no training phase. Moreover, the multiaccess slope region, as proved in Corollary 4, does not depend on α and is always zero.

3.2 High SNR Characterization

As described in [40], the key performance measures in the high-SNR regime, for MIMO point-to-point channel, are $\mathcal{S}_\infty$, which quantifies the the high-SNR slope in bits/s/Hz/(3 dB) of the capacity as function of $\log \mathsf{SNR}$, and $\mathcal{L}_\infty$, which represents the zero-order term, or power offset, in 3-dB units, with respect to a reference channel having the same high-SNR slope but with unfaded and orthogonal dimensions whose expansion in (3.31) intersects the origin at $\mathsf{SNR}\,|_{dB} = 0$. These two quantity are defined as:

$$\mathcal{S}_\infty = \lim_{\mathsf{SNR}\to\infty} \frac{C(\mathsf{SNR})}{\log_2 \mathsf{SNR}} \tag{3.29}$$

and

$$\mathcal{L}_\infty = \lim_{\mathsf{SNR}\to\infty} \left(\log_2 \mathsf{SNR} - \frac{C(\mathsf{SNR})}{\mathcal{S}_\infty} \right) \tag{3.30}$$

Let us observe, that through $\mathcal{S}_\infty$ and $\mathcal{L}_\infty$, we can give the asymptotic behavior of the capacity, [40], as:

$$C\left(\mathsf{SNR}\right) = \mathcal{S}_\infty \left(\frac{\mathsf{SNR}\,|_{dB}}{3dB} - \mathcal{L}_\infty \right) + o\left(1\right). \tag{3.31}$$

Obviously, the previous quantity can also be defined for fixed signaling strategy, replacing $C(\mathsf{SNR})$ with the corresponding mutual information $I(\mathsf{SNR})$. Moving to the multiple access channel, the key performance measures in the high-SNR regime are defined with respect to asymptotic behavior of the sum-capacity, [23] [41]. Specifically, we can consider the same parameters introduced for a MIMO point-to-point channel, replacing the function $C(\mathsf{SNR})$ with the function $C^{sum}(\mathsf{SNR})$, for which we assume that all the users have the same power constraint. Specifically, the asymptotic behavior becomes:

$$C^{sum}\left(\mathsf{SNR}\right) = \mathcal{S}_\infty^{sum} \left(\frac{\mathsf{SNR}\,|_{dB}}{3dB} - \mathcal{L}_\infty^{sum} \right) + o\left(1\right). \tag{3.32}$$

3.2.1 High SNR Analysis of MIMO MAC with Partial CSI

In this section, the analysis for high SNR of the MIMO MAC where only a statistic $\mathbf{S}$ of the channel realization $\mathbf{H}$ is available at the receiver, is considered. Precisely, the high-SNR slope is studied. Before analyzing the multi-user setting with only partial CSI, one of the results for MIMO MAC channel with coherent reception, obtained in [23] [40] [41], is recalled.

Theorem 21 *Let us assume that* $\mathbf{H}$ *has full rank with probability one. Then the high-*SNR *slope of the sum-capacity is given by:*

$$\mathcal{S}_{\infty}^{sum} = D \tag{3.33}$$

where $D = \min(L, MK)$*, i.e. the rank of* $\mathbf{H}$.

Concerning the partial CSI, we focus on the asymptotic behavior of the lower bound to the sum-rate given in Corollary 1, which we know to be an achievable sum-rate with Gaussian signaling. In the following we assume that $(\mathbf{S}, \mathbf{H})$ is jointly Gaussian. Moreover, since the lower bound to the sum-rate depends on the probability characterization of static $\mathbf{S}$, that can be function of SNR, we proceed giving an asymptotic behavior, with respect to SNR, of the quantity related to $\mathbf{S}$.

Specifically, we assume that the matrix $\mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}}^{\mathsf{SNR}}$ is described through the following Eigenvalue Decomposition (EVD)

$$\mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}}^{\mathsf{SNR}} = \mathbf{U}_{\mathsf{SNR}}\,\mathbf{\Lambda}_{\mathsf{SNR}}\,\mathbf{U}_{\mathsf{SNR}}^{\dagger},$$

and describe the j-th entry of the diagonal matrix $\mathbf{\Lambda}_{\mathsf{SNR}}$ with one of the following asymptotic behavior with respect to SNR:

$$\lambda_{j,\mathsf{SNR}} = \begin{cases} a_j \frac{\mathsf{SNR}}{1+b_j\mathsf{SNR}} & \text{if } \alpha = 1 \\ a_j \frac{1}{|1-\alpha|}\mathsf{SNR}^{1-\alpha} & \text{if } \alpha \geq 0 \text{ and } \alpha \neq 1 \\ a_j \log(\mathsf{SNR}) & \text{otherwise} \end{cases} \tag{3.34}$$

where $a_j > 0$, for all $j \in 1, \ldots, L$. These behaviors essentially take in account the improving rate of the quality of the side information $\mathbf{S}$, with respect to the operating SNR. Furthermore, we assume that $\widehat{\mathbf{H}}$ is almost surely full rank for high SNR, that essentially means that $\mathbf{C}_h$ is a full rank matrix.
Under such Hypothesis, we can prove that:

Theorem 22 *The high-*SNR *slope* $\mathcal{S}_{\infty}^{sum}$*, of the lower bound to sum-rate, for* *every positive definite power allocation matrix* $\mathbf{P}$ *is given by:*

$$\mathcal{S}_{\infty}^{sum} = \begin{cases} D & \text{if } \alpha \geq 1 \\ D\alpha & \text{if } 0 \leq \alpha < 1 \\ D & \text{otherwise} \end{cases} \tag{3.35}$$

where $D = \min(L, KM)$ *is the degree of freedom of the system and represents the rank of the matrix* $\hat{\mathbf{H}}$.

Proof: See Appendix C **Proof of Theorem 22**.

Let us observe that in the case $\alpha \geq 1$, or for the logarithmic behavior, we obtain the same performance, in terms of the high-SNR slope, that we have for the coherent case, i.e. we achieve the maximum degree of freedom of the system, D. Moreover, we can observe that in terms of the high-SNR slope, too high improving rates, such as the case $\alpha > 1$, are not needed, while slow rates, i.e. $\alpha < 1$, produce a proportional penalty factor on the achievable high-SNR slope.

High-SNR Analysis for Training-Based System

In this paragraph, applying the previous results, we analyze the high-SNR behavior for a Training Based System. Specifically, we are interested in analyzing the influence on the performance of the system, from T_c, K, M and L.

We assume that $\mathsf{SNR}_T = \frac{\mathsf{SNR}}{M}$ for all the users and full rank training condition, i.e. $T_c > N_T \geq KM$, with training sequences given by Theorem 11. We also suppose that $\mathbf{C}_h$ is full rank and consider for simplicity isotropic inputs. Under such hypothesis, easily follows that the high SNR behavior of the lower bound to the effective throughput is given by:

$$\mathcal{S}_{\infty}^{\mathcal{T}} = \lim_{\mathsf{SNR} \to \infty} \frac{T_c - N_{\mathrm{T}}}{T_c} \frac{I_{\mathsf{Lower}}}{\log(\mathsf{SNR})} = \frac{T_c - N_{\mathrm{T}}}{T_c} \min(L, KM). \tag{3.36}$$

Let us study $\mathcal{S}_{\infty}^{\mathcal{T}}$ as function of N_{T}, K and M, for every fixed L and T_c[5]. Precisely, we assume that only a subset $\mathcal{E}_F$, with cardinality $M' \leq MK$, of total transmitting antennas of all the users is used, and we consider a training

[5]We assume that the coherent time T_c is an even number.

phase duration $N'_{\mathrm{T}} \geq M'$. The correspondent high-SNR slope is given by:

$$\mathcal{S}_{\infty}^{\mathcal{T}}\left(N'_T, \mathcal{E}_F\right) = \frac{T_c - N'_{\mathrm{T}}}{T_c} \min\left(M', L\right) \tag{3.37}$$

that depends on the chosen subset $\mathcal{E}_F$, only through M'. We are interested in optimizing $\mathcal{S}_{\infty}^{\mathcal{T}}$, with respect to N'_{T} and M', i.e. the number of the effective transmitting antennas in the system. Precisely we can prove that:

Theorem 23 *The optimal high-SNR slope $\mathcal{S}_{\infty}^{\mathcal{T}^{opt}}$ of the lower bound to throughput is given by:*

$$\mathcal{S}_{\infty}^{\mathcal{T}^{opt}} = \left(1 - \frac{K_T}{T_c}\right) K_T \tag{3.38}$$

where $K_T = \min\left(KM, L, \frac{T_c}{2}\right)$, with the optimal number of total transmitting antennas given by:

$$M' = K_T = \min\left(KM, L, \frac{T_c}{2}\right) \tag{3.39}$$

and training duration $N'_T = M'$.

Proof: See Appendix C **Proof of Theorem 23**.
The Proof of Theorem 23 follows the steps done in [10], in which the analysis is conducted for a point-to-point MIMO channel, where the channel matrix is characterized by i.i.d. zero mean Gaussian random variables. The final result is of the same type, i.e. $N'_{\mathrm{T}} = M' = K_T = \min\left(KM, L, \frac{T_c}{2}\right)$, but now this situation implies that we have to schedule the users and their transmitting antennas, in order to achieve the optimal high SNR slope associated with a total number of transmitting antennas given by K_T. Let us also observe that in the case of point-to-point MIMO channel with i.i.d. Gaussian channel coefficients and incoherent reception, it has been proved in [42] that

$$\left(1 - \frac{\min\left(KM, L, \frac{T_c}{2}\right)}{T_c}\right) \min\left(KM, L, \frac{T_c}{2}\right),$$

is the capacity high SNR slope, and then we can conclude that in the case of a MIMO MAC channel with symmetric users, each with i.i.d. channel matrix, the Training Based System achieves the optimal performance of the system in terms of high-SNR slope.

3.3 Appendix C

Proof of Theorem 17 First of all let us observe that all the flash signaling satisfying (3.18) are fist order optimal, [35]. We have to show the converse. To this end, let us observe that an upper bound to the interested quantity is given by:

$$\frac{I(\mathbf{x};\mathbf{y}|\mathbf{S})}{L\mathsf{SNR}} \leq \frac{\mathbb{E}\left[||\mathbf{H}\mathbf{x}_{\mathsf{SNR}}||^2\right]}{\mathbb{E}\left[||\mathbf{x}_{\mathsf{SNR}}||^2\right]} - \frac{No\mathbb{E}\left[\log|\mathbf{I}+\frac{1}{No}\mathrm{cov}(\mathbf{H}\mathbf{x}_{\mathsf{SNR}}|\mathbf{x}_{\mathsf{SNR}},\mathbf{S})|\right]}{\mathbb{E}\left[||\mathbf{x}_{\mathsf{SNR}}||^2\right]}$$

The objective is to prove that

$$\frac{No\mathbb{E}\left[\log|\mathbf{I}+\frac{1}{No}\mathrm{cov}(\mathbf{H}\mathbf{x}_{\mathsf{SNR}}|\mathbf{x}_{\mathsf{SNR}},\mathbf{S})|\right]}{\mathbb{E}\left[||\mathbf{x}_{\mathsf{SNR}}||^2\right]} > 0 \tag{3.40}$$

if the signaling is not a flash signaling, and then the only optimal signaling are the flash signaling.
To this end let us lower bound (3.40) observing that in general:

$$|\mathbf{I}+\mathbf{A}| \geq 1+\lambda_{max}(\mathbf{A}) \geq 1+\frac{1}{m}\mathrm{tr}\{\mathbf{A}\} \tag{3.41}$$

then

$$\begin{aligned}
|\mathbf{I}+&\frac{1}{No}\mathrm{cov}(\mathbf{H}\mathbf{x}_{\mathsf{SNR}}|\mathbf{x}_{\mathsf{SNR}},\mathbf{S})| \\
&\geq 1+\frac{1}{mNo}\mathbf{x}_{\mathsf{SNR}}^{\dagger}\mathbb{E}_{\mathbf{S}}\left[\mathbf{Z}^{\dagger}\mathbf{Z}\right]\mathbf{x}_{\mathsf{SNR}} \\
&\geq 1+\frac{1}{mNo}||\mathbf{x}_{\mathsf{SNR}}||^2\lambda_{min}\left(\mathbb{E}_{\mathbf{S}}\left[\mathbf{Z}^{\dagger}\mathbf{Z}\right]\right) \\
&\geq 1+\frac{1}{mNo}||\mathbf{x}_{\mathsf{SNR}}||^2\alpha \qquad a.s.
\end{aligned} \tag{3.42}$$

Using (3.42), we have that:

$$\mathbb{E}\left[\log|\mathbf{I}+\frac{1}{No}\text{cov}(\mathbf{H}\mathbf{x}_{\mathsf{SNR}}|\mathbf{x}_{\mathsf{SNR}},\mathbf{S})|\right]$$

$$\geq \mathbb{E}\left[log\left(1+\frac{1}{LNo}||\mathbf{x}_{\mathsf{SNR}}||^2\alpha\right)\right] \tag{3.43}$$

$$\geq \mathbb{E}\left[log\left(1+\frac{1}{LNo}||\mathbf{x}_{\mathsf{SNR}}||^2\alpha\right)1\{||\mathbf{x}_{\mathsf{SNR}}||<\nu\}\right] \tag{3.44}$$

$$\geq \frac{1}{\nu^2}log\left(1+\frac{1}{LNo}\nu^2\alpha\right)\mathbb{E}\left[||\mathbf{x}_{\mathsf{SNR}}||^2 1\{||\mathbf{x}_{\mathsf{SNR}}||<\nu\}\right] \tag{3.45}$$

where in (3.43), we have used (3.42), in (3.44) we have used the property $f(\mathbf{x}) \leq g(\mathbf{x}) \ \ \forall \mathbf{x}$ implies that $\mathbb{E}[f(\mathbf{x})] \leq \mathbb{E}[g(\mathbf{x})]$, and finally in (3.45) the fact that $\frac{1}{x}\log(1+x)$ is a decreasing function in x. Now, if the signaling is not a flash signaling, then exists a $\nu_o > 0$ such that

$$\frac{\mathbb{E}\left[||\mathbf{x}_{\mathsf{SNR}}||^2 1\{||\mathbf{x}_{\mathsf{SNR}}||<\nu_o\}\right]}{\mathbb{E}\left[||\mathbf{x}_{\mathsf{SNR}}||^2\right]} \geq \beta > 0, \tag{3.46}$$

then

$$\frac{No\mathbb{E}\left[\log|\mathbf{I}+\frac{1}{No}\text{cov}(\mathbf{H}\mathbf{x}_{\mathsf{SNR}}|\mathbf{x}_{\mathsf{SNR}}\mathbf{S})|\right]}{\mathbb{E}\left[||\mathbf{x}_{\mathsf{SNR}}||^2\right]} > 0 \tag{3.47}$$

since $\frac{1}{\nu_o^2}log\left(1+\frac{1}{mNo}\nu_o^2\alpha\right) > 0$.

■

Proof of Theorem 18 From Theorem 17 we know that the only optimal first order signallings are the Flash signallings. Then, we have to show that $\mathbf{S}_0 = 0$ for flash signaling. To this end we make use of the following Lemma:

Lemma 3 *Assume that neither the receiver nor the transmitter knows* $\mathbf{H}$*, but the receiver has at its disposal a statistic* $\mathbf{S}$ *of the channel* $\mathbf{H}$*. Then for all flash signaling,*

$$\mathbf{S}_0 = 0$$

Proof: Let us observe that $\mathbf{S}_0$ is a no-negative number, not greater than the slope in the case of coherent reception. Since in the case of coherent reception, is proved in [35] that the slope is zero, in our case the slope is again zero.

Then, using Lemma 3, follows the assert.

■

Proof of Theorem 19 Since the presence of interferers cannot lower the minimum energy per bit and (3.20) is the minimum transmitted energy per bit for the single-user case with imperfect channel estimation, the result will follow by showing that on-off signalling TDMA achieves the single-user transmitted energies per bit. This follows immediately from [26, Theorem 6], which claims that if both alphabets $\mathcal{A}_1$ and $\mathcal{A}_2$ of the two-user MIMO MAC contain free input symbols, i.e. $0 \in \mathcal{A}_k$ with $k = 1, 2$ such that $\|0\|^2 = 0$, then the following rectangle is achievable per unit cost:

$$\left\{0 \leq R_1 \leq \sup_{x \in \mathcal{A}_1} \frac{D(P_{Y|X_1=x, X_2=0} \parallel P_{y|X1=0,X_2=0})}{\|x\|^2}\right\}$$
$$\times \left\{0 \leq R_2 \leq \sup_{x \in \mathcal{A}_2} \frac{D(P_{Y|X_1=0, X_2=x} \parallel P_{y|X1=0,X_2=0})}{\|x\|^2}\right\}$$

Thus, for $k = 1$

$$\frac{E_b}{N_0}_{\mathsf{min}} = \lim_{\mathsf{SNR} \to 0} \frac{\mathsf{SNR}}{R_1} \tag{3.48}$$
$$= \frac{\log_e 2}{\lambda_{\mathsf{max}}(\mathbb{E}[\mathbf{H}^{(1)\dagger}\mathbf{H}^{(1)}])}, \tag{3.49}$$

where (3.49) follows from Theorem 16. Analogously for $k = 2$. Finally to keep the constraint $\frac{R_1}{R_2} = \theta$ it is enough to consider $\mathsf{SNR}_2 = \frac{G_1 \mathsf{SNR}_1}{G_2 \theta}$

■

Proof of Theorem 20 First of all, let us observe that $\mathcal{S}_k(\theta)$ for $k = 1, 2$ are non-negative quantity. Moreover, let us observe that both $\mathcal{S}_1(\theta)$ and $\mathcal{S}_2(\theta)$ are minus or equal to the related quantity $\mathcal{S}_0^k$ in the single user case, since the presence of the other user only increases the noise, and the minimum energy per bit is the same. Since in the single user case, under assumption (3.22), $\mathcal{S}_0^k = 0$ for $k = 1, 2$, from Theorem 18, we can conclude that $\mathcal{S}_k(\theta) = 0$ for $k = 1, 2$. Finally let us observe that using on-off signalling and TDMA for the two users, we achieve the optimal in low SNR regime.

■

Proof of Corollary 4 From Theorem 20, we have to prove that:

$$\lambda_{min}\left(\mathbb{E}_{\mathbf{S}}\left[\mathbf{Z}^{(k)\dagger}\mathbf{Z}^{(k)}\right]\right) \geq \alpha_k > 0 \quad a.s. \quad k = 1, 2. \tag{3.50}$$

First of all, let us observe that $\mathbb{E}_{\mathbf{S}}\left[\mathbf{Z}^{(k)\dagger}\mathbf{Z}^{(k)}\right]$ doesn't depend from $\mathbf{S}$, and that vec$\{\mathbf{Z}\}$=$\mathbf{e}$ is the estimation error vector defined in Section2.2. Moreover, from the hypotheses that $\mathbf{C}_{\mathbf{h}^{(k)}},\ k = 1, 2$ are full rank matrix, follows that

$$\mathbf{C}_{\mathbf{e}}(\mathbf{T}, \mathsf{SNR}_T) = (\mathbf{C}_{\mathbf{h}}^{-1} + \mathsf{SNR}_T\tilde{\mathbf{T}}^{\dagger}\tilde{\mathbf{T}})^{-1} \tag{3.51}$$

is full rank, since

$$\lambda_{min}\left(\mathbf{C}_{\mathbf{e}}(\mathbf{T}, \mathsf{SNR}_T)\right) \geq \frac{1}{\lambda_{min}\left(\mathbf{C}_{\mathbf{h}}\right) + \mathsf{SNR}_T\lambda_{max}\left(\tilde{\mathbf{T}}^{\dagger}\tilde{\mathbf{T}}\right)} \geq 0. \tag{3.52}$$

Then, indicating with SNR_T^* the maximum power available for the training phase, we obtain that

$$\lambda_{min}\left(\mathbf{C}_{\mathbf{e}}(\mathbf{T}, \mathsf{SNR}_T)\right) \geq \alpha^* = \frac{1}{\lambda_{min}\left(\mathbf{C}_{\mathbf{h}}\right) + \mathsf{SNR}_T^*\lambda_{max}\left(\tilde{\mathbf{T}}^{\dagger}\tilde{\mathbf{T}}\right)}$$

Let us now observe that all sub-vectors extracted from $\mathbf{e} = \text{vec}\{\mathbf{Z}\}$ have full rank covariance matrix with minimum eigenvalue greater or equal to α^*. Let us now characterize

$$\lambda_{min}\left(\mathbb{E}\left[\mathbf{Z}^{(k)\dagger}\mathbf{Z}^{(k)}\right]\right), \tag{3.53}$$

in particular, let us fix a unitary vector $\mathbf{u} \in \mathbb{C}^L$ and consider the quadratic form:

$$\begin{aligned}
\mathbf{u}^{\dagger}\mathbb{E}\left[\mathbf{Z}^{(k)\dagger}\mathbf{Z}^{(k)}\right]\mathbf{u}^{\dagger} &= \mathbb{E}\left[||\mathbf{Z}^{(k)}\mathbf{u}||^2\right] \\
&= \sum_{i=1}^{L}\mathbf{u}^{\dagger}\mathbb{E}\left[\mathbf{Z}_i^{(k)\dagger}\mathbf{Z}_i^{(k)}\right]\mathbf{u} \\
&\geq L\alpha^*
\end{aligned} \tag{3.54}$$

Then for $k = 1, 2$, the minimum eigenvalue is greater or equal to $L\alpha^*$, for all $\mathsf{SNR}_T \leq \mathsf{SNR}_{T*}$ and we conclude the proof.

Proof of Theorem 22

Let us analyze

$$I_{\mathsf{Lower}} = \mathbb{E}\left[\log\left|\mathbf{I}_L + \mathsf{SNR}\ (\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}})^{-1}\,\mathbf{\Gamma}_{\mathcal{U}}\right|\right] \tag{3.55}$$

for high SNR, where $\mathbf{\Gamma}_{\mathcal{U}} = \hat{\mathbf{H}}\mathbf{P}\hat{\mathbf{H}}^{\dagger}$. In particular let us expand (3.55) in the following way:

$$\mathbb{E}\left[\log\left|\mathbf{I}_L + \mathbf{U}\mathbf{\Lambda}_{\mathsf{SNR}}\,\mathbf{U}^{\dagger} + \mathsf{SNR}\,\mathbf{\Gamma}_{\mathcal{U}}\right|\right] - \log\left|\mathbf{I}_L + \mathbf{U}\mathbf{\Lambda}_{\mathsf{SNR}}\,\mathbf{U}^{\dagger}\right|.$$

Since $(\mathbf{S}, \mathbf{H})$ are jointly Gaussian, we can apply the Lebesgue Convergence Theorem, and study the asymptotic behavior of

$$\log\left|\mathbf{I}_L + \mathsf{SNR}\ (\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}})^{-1}\,\mathbf{\Gamma}_{\mathcal{U}}\right|,$$

for each $\mathbf{\Gamma}_{\mathcal{U}}$. For the Hypotheses done $\hat{\mathbf{H}}$ is full rank, and for the moment we assume that the matrix $\mathbf{\Gamma}_{\mathcal{U}}$ is full rank L with probability one. Let us now observe that using the Schur concave property of the function $\log|\mathbf{A}|$ with respect to the eigenvalues of $\mathbf{A}$, we have:

$$\sum_{j=1}^{L} \log(1+\lambda_{j,\mathsf{SNR}} + \mathsf{SNR}\,\lambda_{\mathbf{\Gamma}_{\mathcal{U}i}}) \leq \log\left|\mathbf{I}_L + \mathbf{\Lambda}_{\mathsf{SNR}} + \mathsf{SNR}\,\mathbf{U}^{\dagger}\mathbf{\Gamma}_{\mathcal{U}}\mathbf{U}\right| \tag{3.56}$$

where $\lambda_{j,\mathsf{SNR}}$ and $\lambda_{\mathbf{\Gamma}_{\mathcal{U}i}}$ are, respectively, the eigenvalues in increasing order of $\mathbf{\Lambda}_{\mathsf{SNR}}$ and $\mathbf{\Gamma}_{\mathcal{U}}$. Moreover from the Hadamard inequality we have:

$$\sum_{j=1}^{L} \log(1+\lambda_{j,\mathsf{SNR}} + \mathsf{SNR}\,\mathbf{\Gamma}_{\mathcal{U}i,i}) \geq \log\left|\mathbf{I}_L + \mathbf{\Lambda}_{\mathsf{SNR}} + \mathsf{SNR}\,\mathbf{U}^{\dagger}\mathbf{\Gamma}_{\mathcal{U}}\mathbf{U}\right| \tag{3.57}$$

where $\mathbf{\Gamma}_{\mathcal{U}i,i}$ are the diagonal elements of $\mathbf{U}^{\dagger}\mathbf{\Gamma}_{\mathcal{U}}\mathbf{U}$.
Furthermore we have:

$$\begin{aligned} \log \left| \mathbf{I}_L + \mathbf{U}\mathbf{\Lambda}_{\mathsf{SNR}}\mathbf{U}^{\dagger} \right| &= \log \left| \mathbf{I}_L + \mathbf{\Lambda}_{\mathsf{SNR}} \right| \\ &= \sum_{j=1}^{L} (\log 1 + \lambda_{j,\mathsf{SNR}}). \end{aligned} \tag{3.58}$$

Then using (3.56), (3.57) and (3.58) we obtain that:

$$\begin{aligned} &\sum_{j=1}^{L} \log \left(1 + \frac{\mathsf{SNR}}{1+\lambda_{j,\mathsf{SNR}}} \lambda \mathbf{\Gamma}_{\mathcal{U}i} \right) \\ &\quad \leq \log \left| \mathbf{I}_L + \mathsf{SNR} \left(\mathbf{I}_L + \mathsf{SNR}\, \mathbf{\Omega}_{\mathcal{U}} \right)^{-1} \mathbf{\Gamma}_{\mathcal{U}} \right| \\ &\quad \leq \sum_{j=1}^{L} \log \left(1 + \frac{\mathsf{SNR}}{1+\lambda_{j,\mathsf{SNR}}} \mathbf{\Gamma}_{\mathcal{U}i,i} \right) \end{aligned} \tag{3.59}$$

and then the asymptotic analysis of the lower bound to the sum-rate, can be carried on analyzing the behavior of the following function:

$$g(\mathsf{SNR}) = \log \left(1 + \frac{\mathsf{SNR}}{1+\lambda_{\mathsf{SNR}}} \beta \right) \tag{3.60}$$

where β is a positive number, for the hypothesis that $\mathbf{\Gamma}_{\mathcal{U}}$ is full rank with probability one.
Let us now describe the behavior of the function (3.60) for high SNR, as function of a_j and α.
If $\alpha \geq 1$:

$$\frac{g(\mathsf{SNR})}{log\,(\mathsf{SNR})} \overset{\mathsf{SNR}\to\infty}{\to} 1.$$

If $0 \leq \alpha < 1$:

$$\frac{g(\mathsf{SNR})}{log\,(\mathsf{SNR})} \overset{\mathsf{SNR}\to\infty}{\to} \alpha. \tag{3.61}$$

Finally, if $\lambda_{j,\mathsf{SNR}} = a_j \log(\mathsf{SNR})$

$$\frac{g(\mathsf{SNR})}{log\,(\mathsf{SNR})} \overset{\mathsf{SNR}\to\infty}{\longrightarrow} 1.$$

From the above analysis, we can conclude that:

$$\frac{I_{\mathsf{Lower}}}{\log\,(\mathsf{SNR})} \overset{\mathsf{SNR}\to\infty}{\longrightarrow} \begin{cases} L & \text{if } \alpha \geq 1 \\ L\alpha & \text{if } 0 \leq \alpha < 1 \\ L & \text{otherwise} \end{cases} \tag{3.62}$$

Let us now analyze the case in which $\mathbf{\Gamma}_{\mathcal{U}}$ is full rank KM with probability one. Concerning the lower bound, we obtain the same expression (3.56), where now the sum is done on the first KM eigenvalues, in decreasing order:

$$\sum_{j=1}^{KM} \log\left(1 + \frac{\mathsf{SNR}}{1+\lambda_{j,\mathsf{SNR}}}\lambda_{\mathbf{\Gamma}_{\mathcal{U}i}}\right) \leq \log\left|\mathbf{I}_L + \mathsf{SNR}\;(\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}})^{-1}\,\mathbf{\Gamma}_{\mathcal{U}}\right| \tag{3.63}$$

For the upper bound, let us indicate with $\mathbf{VDV}^\dagger$, the eigenvalue decomposition of $\mathbf{\Gamma}_{\mathcal{U}}$ with $\mathbf{D}$ a diagonal matrix of dimension $KM \times KM$, we have:

$$\begin{aligned} &\log\left|\mathbf{I}_L + \mathsf{SNR}\;(\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}})^{-1}\,\mathbf{\Gamma}_{\mathcal{U}}\right| \\ &= \log\left|\mathbf{I}_L + \mathsf{SNR}\;(\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}})^{-1}\,\mathbf{VDV}^\dagger\right| \\ &= \log\left|\mathbf{I}_{KM} + \mathsf{SNR}\,\mathbf{V}^\dagger\,(\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}})^{-1}\,\mathbf{VD}\right| \\ &= \log\left|\mathbf{I}_{KM} + \mathsf{SNR}\,\tilde{\mathbf{V}}\mathbf{D}_{\mathbf{\Omega}_{\mathcal{U}}}\tilde{\mathbf{V}}^\dagger\mathbf{D}\right| &(3.64) \\ &\leq \sum_{i=1}^{KM} \log\left(1 + \mathsf{SNR}\,\mathbf{D}_{\mathbf{\Omega}_{\mathcal{U}i}}\mathbf{K}_{i,i}\right) &(3.65) \\ &\leq \sum_{i=1}^{KM} \log\left(1 + \frac{\mathsf{SNR}}{1+\lambda_{i,\mathsf{SNR}}}\mathbf{K}_{i,i}\right) &(3.66) \end{aligned}$$

where in (3.64) $\tilde{\mathbf{V}}\mathbf{D}_{\mathbf{\Omega}_{\mathcal{U}}}\tilde{\mathbf{V}}^\dagger = \mathbf{V}^\dagger\,(\mathbf{I}_L + \mathsf{SNR}\,\mathbf{\Omega}_{\mathcal{U}})^{-1}\,\mathbf{V}$ (is its eigenvalue decomposition), in (3.65) we define the matrix $\mathbf{K} = \tilde{\mathbf{V}}\mathbf{D}\tilde{\mathbf{V}}^\dagger$ and apply the Hadamard inequality, and finally in (3.66) we apply The Eigenvalue Interlacement Theorem [30], and $\lambda_{i,\mathsf{SNR}}$ are taken in increasing order. From (3.63) and

(3.66), we have ,again, to study the behavior of the function:

$$g(\mathsf{SNR}) = \log\left(1 + \frac{\mathsf{SNR}}{1+\lambda_{\mathsf{SNR}}}\beta\right) \tag{3.67}$$

where β is a positive number. The analysis of the function (3.67), has been previously conducted for the case $L \leq KM$, and we can conclude that:

$$\frac{I_{\mathsf{Lower}}}{\log(\mathsf{SNR})} \overset{\mathsf{SNR}\to\infty}{\longrightarrow} \begin{cases} KM & \text{if } \alpha \geq 1 \\ KM\alpha & \text{if } 0 \leq \alpha < 1 \\ KM & \text{otherwise} \end{cases} \tag{3.68}$$

■

Proof of Theorem 23
We are interested to maximize

$$\mathcal{S}_{\infty}^{\mathcal{T}}\left(N'_{\mathrm{T}}, \mathcal{E}_F\right) = \frac{T_c - N'_{\mathrm{T}}}{T_c}\min\left(M', L\right) \tag{3.69}$$

with respect to N'_{T} and M'. First of all, it is evident that $N'_{\mathrm{T}} = M'$, since an higher N'_{T} only decreases the high-SNR slope. Then we have:

$$\mathcal{S}_{\infty}^{\mathcal{T}}\left(M'\right) = \begin{cases} \left(1 - \frac{M'}{T_c}\right)M' & \text{if } M' \leq L \\ \left(1 - \frac{M'}{T_c}\right)L & \text{if } M' \geq L \end{cases} \tag{3.70}$$

We maximize the term $\left(1 - \frac{M'}{T_c}\right)M'$, choosing M'=$\frac{T_c}{2}$ when $\min(KM, L) \geq \frac{T_c}{2}$, and choosing M'=$\min(KM, L)$ if $\min(KM, L) < \frac{T_c}{2}$. Therefore, the optimal choice is M'=$\min\left(KM, L, \frac{T_c}{2}\right)$. As concerned the second term $\left(1 - \frac{M'}{T_c}\right)L$, this is maximized when M'=L=$\min(KM, L)$. Hence, defining K_T=$\min\left(KM, L, \frac{T_c}{2}\right)$, we obtain that:

$$\mathcal{S}_{\infty}^{\mathcal{T}^{opt}} = \left(1 - \frac{K_T}{T_c}\right)K_T \tag{3.71}$$

with the optimal number of total transmitting antennas given by:

$$M' = K_T = \min\left(KM, L, \frac{T_c}{2}\right). \tag{3.72}$$

Conclusion

In this thesis, the impact of a partial CSI at the receiver on the achievable rate region of MIMO MAC with Gaussian Input, has been analyzed. The imperfect CSI is modeled as an arbitrary statistic conditioned on which the channel has a Gaussian distribution. Specifically, lower and upper bounds of the various mutual information terms, which define the achievable rate region for Gaussian inputs, have been derived. Furthermore, the tightness of these bounds have been numerically illustrated and it has also been shown that the gap goes to zero when the number of the users or the number of receiving antennas grows. Moreover, the low-SNR regime and high-SNR regime for a MIMO MAC with partial CSI, have been analyzed.

The developed tool has been applied to study the performance limits of two relevant scenarios: the Cooperative MIMO Networks and the Training Based System. Precisely, for a Cooperative MIMO Network, the partial CSI has been modeled in terms of knowledge of only some channel matrices; moreover the case in which the channel matrices are not perfectly known but only a quantized version is available has been treated. For a Training Based System, the CSI at the receiver is obtained through the transmission of training sequences in each coherence block of the channel. It has been shown that the training signals sent by the users, optimized according to two metrics, viz., the trace and determinant of the estimation error covariance, should be orthogonal in time if no constraints on the rank of the training sequence matrix are imposed. When there is a rank constraint, i.e., when the training sequences can not longer be orthogonal, the optimal structure for the training signals, has been derived under the assumption of symmetric network.

Owing to the above analysis, it can be concluded that the rates achievable over a MIMO MAC channel with partial CSI at the receiver side can be interpreted in terms of a coherent MIMO MAC channel with an increased noise level and that the successive interference cancelation strategy at the receiver is a good decoding strategy to achieve the derived lower bound on the sum-rate.

Concerning the Cooperative MIMO Network, the impact of incomplete CSI, in terms of the degree of cooperation, becomes more and more evident when the power used for the transmission of the information grows. Similarly, increasing the operating SNR, a finer quantization is needed to achieve good performance compared to the ideal case of no-quantization. For the Training Based System, the correlation at the receiver side has a great impact on the performance of the system in rank deficient case. In fact, when there are independent receiving antennas, the performance of the system rapidly saturates, since none channel matrix can be well estimated. Furthermore, when the coherent time is large enough, it is preferable to consider full rank training sequences to obtain higher achievable rates.
For a MIMO MAC with partial CSI, the minimum required energy per information bit for each user at the receiver equals $\ln 2$, which is the same of the one needed for coherent MIMO MAC. Moreover, assuming that none channel coefficient is perfectly known, TDMA access with on-off signalling sent by the users, achieves the optimal performance up to the second order of the system; this means that we have to decouple the users of the network in low SNR regime. Finally, it is enough that the mean square error of the channel estimation goes to zero like $\log(\mathsf{SNR})/\mathsf{SNR}$ to obtain the same high-SNR slope of the coherent case.
Possible future research guide lines could be different. One of these concerns the analysis of a Cooperative MIMO networks in which also the received data signals have to be quantized at the BSs, to take in account that the capacity of the backhaul links are finite. An additional research topic could concern, in order to optimize the achievable rates over the parameters of the training and data transmission phases, the use of the developed framework for a Training Based System. Finally it could be very interesting to extend the developed analysis to study the performance of relay channels, in which all the relays and the receiver have only partial CSI.

Bibliography

[1]G. J. Foschini and M. J. Gans, ”On Limits of Wireless Communications in a Fading Environment When Using Multiple Antennas,” *Wireless Pers. Commun.,* Vol. 6, No. 3, pp. 311-335, March 1998

[2]E. Telatar, ”Capacity of Multi-antenna Gaussian Channels,” *European Trans. of Telecommunication,* Vol. 10, No.6, pp. 586-595, Nov.-Dec. 1999.

[3]W. Yu, W. Rhee, S. Boyd and J. M. Cioffi,”Iterative Water-Filling for Gaussian Vector Multiple-Access Channels”*IEEE Trans. on Information Theory*, Vol.50,no. 1, pp.145-152, Jan 2004.

[4]B. Hochwald and S. Vishwanath,”Space-Time Multiple Access: Linear Growth in the Sum Rate”,*Allerton 2002*.

[5]D.N.C. Tse, P. Viswanath and L. Zheng,”Diversity Multiplexing Tradeoff in Multiple-Access Channels” *IEEE Trans. on Information Theory*, Vol.50,no. 9, pp.1859-1874, Sept 2004.

[6]G. Caire, and S. Shamai, ”On the achievable throughput of a multiantenna Gaussian broadcast channel”, *IEEE Trans. on Information Theory*, vol.49, no.7, pp.1691-1706, June 2003.

[7]N. Jindal, S. Vishwanath, and A. Goldsmith,”On the Duality of Gaussian Multiple-Access and Broadcast Channels”, *IEEE Trans. on Information Theory*, Vol.50,no. 5, pp.768-783, May 2004.

[8]A. De Maio, M. Lops, and L. Venturino, ”Diversity-integration trade-offs in MIMO detection,” *IEEE Trans. on Signal Processing* vol. 56, no. 10, pp. 5051-5061, Oct. 2008.

[9]A. Aubry, M. Lops, A. M. Tulino, and L. Venturino, ”On MIMO detection under non-Gaussian target scattering,” *IEEE Trans. on Information Theory*, vol. 56, no. 11, pp. 5822-5838, Nov. 2010.

[10]B. Hassibi and B. M. Hochwald, ”How Much Training is Needed in Multiple-Antenna Wireless Links?”, *IEEE Trans. on Information Theory*, Vol. 49, No. 4, pp. 951-963, April 2003

[11]T. Yoo and A. Goldsmith, ”Capacity and Power Allocation for Fading MIMO Channels With Channel Estimation Error”, *IEEE Trans. on Information Theory*, Vol. 52, No. 5, pp. 2203-2214, May 2006

[12]M. M édard, ”The Effect upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel”, *IEEE Trans. on Information Theory*, Vol. 46, No. 3, pp. 933-946, May 2000

[13]L. Musavian, M.R. Nakhai, M. Dohler and A.H. Aghvami, ”Effect of Channel Uncertainty on the Mutual Information of MIMO Fading Channels”*IEEE Trans. on ON Vehicular Tecnology*, Vol. 56, No. 5, pp. 2798-2806, Sept. 2007

[14]A. Aubry, A. M. Tulino, and S. Venkatesan, ”Multiple-Access Channel Capacity Region with Incomplete Channel State Information,”, *ISIT 2010*.

[15]A. Aubry, A. M. Tulino, and S. Venkatesan, ”Receiver Cooperation in Gaussian Interference Networks with Imperfect Channel State Information”, *Allerton 2010*.

[16]A. Aubry, A. M. Tulino, and S. Venkatesan, ” On the Achievable Rate Region of MIMO Multiple Access Channel with Partial Channel State Information,” *IEEE Trans. on Information Theory*, conditionally accepted Oct. 2010.

[17]G. Alfano, A. Aubry, H. Huang and A.M. Tulino, ”Analysis of cooperative MIMO networks with incomplete channel state information”, *PIMRC 2008*.

[18]G.J. Foschini, K. Karakayali and R.A. Valenzuela,” Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency,” *IEEE Wireless Communications*, pp. 56-61, Aug. 2006.

[19]S. Venkatesan, ”Coordinating Base Stations for Greater Uplink Spectral Efficiency in a Cellular Network”, *PIMRC 2008*.

[20]S. Sesia, I. Toufik and M. Baker, *LTE, The UMTS Long Term Evolution: From Theory to Practice* ,John Wiley & Sons, 2009.

[21]S. Verd ú, *Multiuser Detection*, Cambridge University Press, 1998.

[22]A. Goldsmith, S.A. Jafar, N. Jindal and S. Vishwanath, "Capacity Limits of MIMO Channels", *IEEE Journal on Selected Areas in Communications*, Vol. 21, No. 5, pp. 684-702, June 2003.

[23]S. Shamai and S. Verd ú, " The Impact of Frequency-Flat Fading on the Spectral Efficiency of CDMA", *IEEE Trans. on Information Theory*, Vol. 47, No. 4, pp. 1302-1327, May 2001.

[24]A. Tulino, A. Lozano and S. Verd ú, "Impact of Antenna Correlation on the Capacity of Multiantenna Channels", *IEEE Trans. on Information Theory*, Vol. 51, No. 7, pp. 2491-2509, July 2005.

[25]R.G. Gallager, "Energy limited channels; Coding, multiaccess and spread spectrum",*Conf Inform. Sci. Syst*, p. 372, Princeton, NJ, March 1988.

[26]S. Verd ú, "On Channel Capacity per Unit Cost," *IEEE Trans. on Information Theory*, Vol. 36, No. 5, pp. 1019-1030, Sept. 1990.

[27]S. Shamai and A.D. Wyner, "Information-theoretic considerations for symmetric, cellular, multiple-access fading channels,"*IEEE Trans. on Information Theory*, Vol. 43, No. 4, pp. 1877-1991, Nov. 1997.

[28]L.L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis,* Addison-Wesley Publishing, 1990.

[29]C.E. Shannon, "Some geometrical results in channel capacity," *Verband Deut. Elektrotechnik. fachber.*, vol. 19, pp.13-15, 1956.

[30]R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[31]T.M. Cover and J.A. Tomas, *Elements of Information Theory,*John Wiley & Sons, Inc., 1991.

[32]T. Berger, *Rate distortion theory,* Prentice-Hall, 1971.

[33]J.H. Kotecha and A.M. Sayeed,"Transmit Signal Design for Optimal Estimation of Correlated MIMO Channels," *IEEE Trans. on Signal ProcessingInformation Theory*, Vol. 52, No. 2, pp. 546-557, Feb. 2004.

[34]S. Boyd and L. Vandenberghe *Convex Optimization*Cambridge University Press & Sons, Inc., 1991.

[35]S. Verd ú, " Spectral Efficiency in the Wideband Regime," *IEEE Trans. on Information Theory*, Vol. 48, No. 6, pp. 1319–1343, June 2002.

[36]G. Caire, D. Tuninetti and S. Verd ú,"Suboptimality of TDMA in the Low-Power Regime," *IEEE Trans. on Information Theory*, Vol. 50, No. 4, pp. 608-620, April 2004.

[37]G. Caire, D. Tuninetti and S. Verd ú,"Variable-Rate Coding for Slowly Fading Gaussian Multiple-Access Channels," *IEEE Trans. on Information Theory*, Vol. 50, No. 10, pp. 2271-2292, Oct. 2004.

[38]V.V. Prelov and S. Verd ú, "Second-Order Asymptotics of Mutual Information," *IEEE Trans. on Information Theory*, Vol. 50, No. 8, pp. 1567-1580, Aug. 2004.

[39]C. Rao and B. Hassibi, "Analysis of multiple antenna wireless links at low SNR," *IEEE Trans. on Information Theory*, vol 50, no 9, pp. 2123-2130, Sep. 2004.

[40]A. Lozano, A. Tulino and S. Verd ú, "High-SNR Power Offset in Multi-antenna Communication," *IEEE Trans. on Information Theory*, vol 51, no 12, pp. 4134-4151, Dec. 2005.

[41]A. Sanderovich, O. Somekh, H.V. Poor and S. Shamai "Uplink Macro Diversity of Limited Backhaul Cellular Network," *IEEE Trans. on Information Theory*, Vol. 55, No. 8, pp. 3457-3478, Aug. 2009.

[42]L. Zheng and D.N.C. Tse, "Communication on the Grassmann Manifold: A Geometric Approach to the Noncoherent Multiple-Antenna Channel," *IEEE Trans. on Information Theory*, vol 48, no 2, pp. 359-383, Feb. 2002.