Cell-Free massive MIMO receiver design and channel estimation

Roya Gholamipourfard

Cell-Free Massive MIMO Receiver Design and Channel Estimation

Dissertation

submitted to

Sorbonne Université

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

Author

Roya Gholamipourfard

Publicly defended on the 17th December 2021, before a committee composed of

Reviewers

Prof.	**Constantinos Papadias**	The American College of Greece, Greece
Prof.	**Pascal Chevalier**	CNAM, France

Examiners

Prof.	**Merouane Debbah**	CentraleSupélec, France
Prof.	**Marios Kountouris**	Eurecom, France

Thesis Director

Prof.	**Dirk Slock**	Eurecom, France

Thesis Co-Director

Prof.	**Laura Cottatellucci**	Friedrich-Alexander University, Germany

Conception de Récepteurs et Estimation de Canal pour le MIMO Massif Distribué

Thèse

soumise à

Sorbonne Université

pour l'obtention du Grade de Docteur

présentée par

Roya Gholamipourfard

Soutenance de thèse prévue le 17 Décembre 2021 devant le jury composé de

Rapporteurs

Prof.	**Constantinos Papadias**	The American College of Greece, Grèce
Prof.	**Pascal Chevalier**	CNAM, France

Examinateurs

Prof.	**Merouane Debbah**	CentraleSupélec, France
Prof.	**Marios Kountouris**	Eurecom, France

Directeur de Thèse

Prof.	**Dirk Slock**	Eurecom, France

Codirectrice de Thèse

Prof.	**Laura Cottatellucci**	Friedrich-Alexander Université, Allemagne

To the memory of my father,
to my mother,
and to those who have inspired me.

Abstract

Next generation wireless systems shall satisfy the increasing demand of higher and higher data rates at very competitive prices as well as be able to efficiently accommodate for and adapt to a huge dynamic range of services, applications, and types of devices expected in the near future, e.g., see smart cities technologies, internet of things (IoT). Appealing architectural solutions have been leveraged on ultra-densification of antennas. Ultra-dense wireless systems envision ultra-dense distributed antenna systems (UD-DAS) based on remote distributed antennas empowered by the e-cloud for a centralized processing. However, neither DAS nor massive MIMO technology will meet the increasing data rate demands of the next generation wireless communications due to the inter-cell interference and large quality of service (QoS) variations. To overcome these issues, cell-free (CF) massive MIMO which combines the best aspects of DASs with the massive MIMO technology, has been introduced as a key solution. The new terminology is used for networks consisting of a massive number of geographically distributed access points (AP), which jointly serve a smaller number of users distributed over a wide area, in the absence of cell boundaries. All the APs are connected through a back-haul network to a central processing unit (CPU). The massive number of antennas improves spectral efficiency whereas energy efficiency and macro-diversity gain result from the distributed topology and ultra-densification. Additionally, since each user is surrounded by a large number of serving APs, with high probability all the users enjoy good channel conditions. Therefore, CF massive MIMO systems are expected to provide significant improvements in terms of spectral/energy efficiency and coverage probability.

One of the major issues in large-scale networks such as CF massive MIMO systems is complexity at the receivers. Interestingly, in centralized MIMO systems, the high complexity of centralized joint detectors has been successfully addressed by massive MIMO systems. In massive MIMO systems, as the number of receive antennas tends to infinity while the number of transmit antennas remains finite, the users' channels become almost orthogonal determining a phenomenon known as *favorable propagation* which makes low complexity linear processing almost optimal. In this regard, the first part of this thesis is devoted to analyzing the favorable propagation properties of CF massive MIMO systems in asymptotic conditions when the network dimensions go to infinity with given intensities of the transmit and receive antenna point processes

(PP). We study the analytical conditions of favorable propagation with two kinds of channels, namely, channels with path loss and transmit and receive antennas in line of sight (LoS) or in multipath Rayleigh fading. We show when the analytical conditions of favorable propagation are not satisfied, the use of low complexity linear multi-user detection becomes very appealing in practical systems.

Channel state information (CSI) in massive MIMO systems, both cellular and CF, plays a major role in improving the system performance. Ideally, training sequences or pilots should be selected to be mutually orthogonal in the channel estimation. However, in most practical scenarios the number of users is greater than the number of orthogonal training sequences and a given training sequence can be assigned to more than one user, therefore leading to the so-called *pilot contamination* which prevents the possibility of obtaining an adequate estimate of CSI. Recently, a wide research has been dedicated to the pilot assignment methods which address the pilot contamination problem via a careful assignment of pilots and do not exploit the inherent structure of channels and data in CF massive MIMO systems in contrast to blind or semi-blind estimation and detection techniques. Therefore, in the second part of this thesis, we address the pilot contamination problem in CF massive MIMO by exploiting the channel sparsity due to the strong attenuation with path loss. Specifically, we consider semi-blind methods for joint channel estimation and data detection and analyze the semi-blind approach with classical signal processing techniques such as Fisher information matrix (FIM), Cramer-Rao bound (CRB), and identifiability.

An extensive attention has been dedicated to the design of detectors relying on message passing (MP) algorithms in recent years. MIMO detection based on expectation propagation (EP) which is also a kind of MP algorithm, obtains near-optimal performance with acceptable complexity, under specific conditions. EP algorithms attempt to find the closest approximation within the exponential family of factors of the posterior distribution in an iterative refinement procedure. Finally, in the last part of this thesis, we propose an MP algorithm based on the EP principle to iteratively conduct Bayesian semi-blind methods for channel estimation and data detection in CF massive MIMO systems.

Abrégé

Les systèmes sans fil de la prochaine génération doivent répondre à la demande croissante de débits de données de plus en plus élevés à des prix très compétitifs et être capables de s'adapter efficacement à une vaste gamme dynamique de services, d'applications et de types de dispositifs attendus dans un avenir proche. Des solutions architecturales attrayantes ont été mises à profit pour l'ultra-densification des antennes. Les systèmes sans fil ultra-denses envisagent des systèmes d'antennes distribuées ultra-denses (UD-DAS) basés sur des antennes distribuées et qui s'appuient sur le cloud pour un traitement centralisé. Cependant, ni la technologie DAS ni la technologie MIMO massive ne pourront répondre aux demandes de débit de données croissantes de la prochaine génération de communications sans fil en raison des interférences intercellulaires et des grandes variations de qualité de service (QoS). Pour surmonter ces problèmes, le MIMO massif sans cellule (Cell-Free - CF), qui combine les meilleurs aspects des DAS avec la technologie MIMO massive, a été présenté comme une solution clé. Cette nouvelle terminologie est utilisée pour les réseaux constitués d'un nombre considérable de points d'accès (AP) répartis géographiquement, qui desservent ensemble un nombre beaucoup plus faible d'utilisateurs répartis sur une vaste zone, en l'absence de frontières entre les cellules. Tous les points d'accès sont reliés à une unité centrale de traitement (CPU) par un réseau de liaison. Le nombre massif d'antennes améliore l'efficacité spectrale tandis que l'efficacité énergétique et le gain de macro-diversité résultent de la topologie distribuée et de l'ultra-densification. En outre, comme chaque utilisateur est entouré d'un grand nombre de points d'accès de desserte, tous les utilisateurs bénéficient très probablement de bonnes conditions de canal. Par conséquent, les systèmes MIMO massifs CF devraient apporter des améliorations significatives en termes d'efficacité spectrale/énergétique et de probabilité de couverture.

L'un des problèmes majeurs dans les réseaux à grande échelle tels que les systèmes MIMO massifs CF est la complexité au niveau des récepteurs. Il est intéressant de noter que dans les systèmes MIMO centralisés, la complexité élevée des détecteurs conjoints centralisés a été résolue avec succès par les systèmes MIMO massifs. Dans les systèmes MIMO massifs, comme le nombre d'antennes de réception tend vers l'infini alors que le nombre d'antennes d'émission reste fini, les canaux des utilisateurs deviennent presque orthogonaux déterminant un phénomène connu sous le nom de *propagation favorable* qui rend le traitement linéaire à faible complexité presque optimal. À

cet égard, la première partie de cette thèse est consacrée à l'analyse des propriétés de propagation favorable des systèmes MIMO massifs CF dans des conditions asymptotiques lorsque les dimensions du réseau vont à l'infini avec des intensités données des processus ponctuels (PP) des antennes d'émission et de réception. Nous étudions les conditions analytiques de propagation favorable avec deux types de canaux, à savoir les canaux avec perte de chemin et les antennes d'émission et de réception en ligne de vue (LoS) ou en évanouissement de Rayleigh par trajets multiples. Nous montrons que lorsque les conditions analytiques de propagation favorable ne sont pas satisfaites, l'utilisation de la détection linéaire multi-utilisateurs à faible complexité devient très intéressante dans les systèmes pratiques.

L'information sur l'état du canal (CSI) dans les systèmes MIMO massifs, tant cellulaires que CF, joue un rôle majeur dans l'amélioration des performances du système. Idéalement, les séquences d'entraînement ou les pilotes devraient être sélectionnés pour être mutuellement orthogonaux dans l'estimation du canal. Cependant, dans la plupart des scénarios pratiques, le nombre d'utilisateurs est supérieur au nombre de séquences d'entraînement orthogonales et une séquence d'entraînement donnée peut être attribuée à plus d'un utilisateur, ce qui conduit à ce que l'on appelle la "contamination des pilotes", qui empêche d'obtenir une estimation adéquate du CSI. Récemment, de nombreuses recherches ont été consacrées aux méthodes d'affectation des pilotes qui traitent le problème de la contamination des pilotes par une affectation prudente des pilotes et n'exploitent pas la structure inhérente des canaux et des données dans les systèmes MIMO massifs CF, contrairement aux techniques d'estimation et de détection aveugles ou semi-aveugles. Par conséquent, dans la deuxième partie de cette thèse, nous abordons le problème de la contamination des pilotes dans les systèmes MIMO massifs CF en exploitant la sparsité du canal due à la forte atténuation avec perte de chemin. Plus précisément, nous considérons des méthodes semi-aveugles pour l'estimation conjointe du canal et la détection des données et nous analysons l'approche semi-aveugle avec des techniques classiques de traitement du signal telles que la matrice d'information de Fisher (FIM), la limite de Cramér-Rao (CRB) et l'identifiabilité.

Ces dernières années, une attention particulière a été accordée à la conception de détecteurs reposant sur des algorithmes de passage de messages (MP). Les détecteurs MIMO basés sur la propagation des moments (Expectation Propagation - EP), qui est également un type d'algorithme MP, obtiennent des performances quasi-optimales avec une complexité acceptable, dans des conditions spécifiques. Les algorithmes EP tentent de trouver l'approximation la plus proche dans la famille exponentielle des facteurs de la distribution postérieure dans une procédure de raffinement itérative. Enfin, dans la dernière partie de cette thèse, nous proposons un algorithme MP basé sur le principe EP pour conduire itérativement des méthodes bayésiennes semi-aveugles pour l'estimation de canal et la détection de données dans les systèmes CF.

Acknowledgements

I would like to express my sincere appreciation to my thesis advisors Prof. Dirk Slock and Prof. Laura Cottatellucci for their invaluable guidance during my PhD. Their motivation, enthusiasm and optimism, expertise and immense knowledge enabled me to grow not only as a researcher but especially as a person. Without their patience, guidance and persistent help this thesis would not have been possible.

I wish to present my sincere thanks to Prof. Constantinos Papadias and Prof. Pascal Chevalier for accepting to read the thesis and for their feedback on my work. I also wish to thank Prof. Merouane Debbah and Prof. Marios Kountouris for their participation in my committee. I am sure that their remarks and suggestions will be so precious to pursue my research.

I would also like to extend my warmest thanks to my colleagues at the Communication Systems department of Eurecom for maintaining a very enjoyable atmosphere and also the members of the Marie-Curie ITN project SPOTLIGHT.

Many thanks go to my friends who have been very supportive during the time at Sophia Antipolis. We all have shared joys and difficulties, concerns and hopes. I will always keep wonderful memories of these past years at Eurecom and my stay in France.

Last but not least, I wish to express my heartfelt gratitude to my family members who have always been encouraging me to pursue my passion, trusting in me, and showing me the value of knowledge and diligence. I dedicate this thesis to them.

Sophia Antipolis, France, December 2021
Roya Gholamipourfard

Contents

List of Figures

Acronyms and Abbreviations

The acronyms and abbreviations used throughout the manuscript are specified in the following. They are presented here in their singular form, and their plural forms are constructed by adding and s, e.g. BS (base stations) and BSc (Base stations). The meaning of an acronym is also indicated the first time that it is used. The English acronyms are also used for the abstract and summary in French.

5G	Fifth Generation.
1D	One-Dimensional.
2D	Two-Dimensional.
AP	Access Point.
AWGN	Additive White Gaussian Noise.
BER	Bit Error Rate.
BP	Belief Propagation.
BS	Base Station.
CF	Cell-Free.
CPU	Central Processing Unit.
CRB	Cramer-Rao Bound.
CSI	Channel State Information.
CSIT	Channel State Information at the Transmitter.
DAS	Distributed Antenna System.
DFT	Discrete Fourier Transform.
DL	Downlink.
EP	Expectation Propagation.
ERM	Euclidean Random Matrix.
FIM	Fisher Information Matrix.
i.i.d.	independent and identically distributed.

IoT	Internet of Things.
KL	Kullback-Leibler.
MAP	Maximum A Posteriori.
MF	Matched Filter.
ML	Maximum Likelihood.
LDPC	Low Density Parity Check.
LoS	Line of Sight.
MIMO	Multiple Input Multiple Output.
ML	Maximum Likelihood.
MMSE	Minimum Mean Square Error.
MP	Message Passing.
MSWF	Multi-Stage Wiener Filter.
Non-LoS	Non Line of Sight.
PP	Point Process.
PDF	Probability Density Function.
PA	Pilot Assignment.
QoS	Quality of Service.
RMT	Random Matrix Theory.
RV	Random Variable.
SIMO	Single Input Multiple Output.
SINR	Signal to Noise plus Interference Ratio.
SNR	Signal to Noise Ratio.
SVD	Singular Value Decomposition.
UL	Uplink.
w.r.t.	with respect to.

Notations

The next list describes an overview on the notation used throughout this manuscript. Boldface uppercase letters ($\mathbf{A}$) are used for matrices, boldface lowercase letters for vectors ($\mathbf{a}$), and regular letters for scalars (a or A). Sets are represented by calligraphic uppercase letters ($\mathcal{A}$).

$\mathbb{C}^{M\times N}$	Set of $M \times N$ complex matrices
$\mathbf{A}^T$	Transpose of matrix $\mathbf{A}$
$\mathbf{A}^H$	Hermitian transpose of matrix $\mathbf{A}$
$\mathbf{A}^*$	Conjugate of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	Inverse of matrix $\mathbf{A}$
$\mathbf{A}^{-1/2}$	Negative square root of matrix $\mathbf{A}$
$[\mathbf{A}]_{ij}$	Element in position (i, j) of matrix $\mathbf{A}$
$\lvert x\rvert$	Absolute value of variable x
$\lVert\mathbf{x}\rVert$	Euclidean norm of vector $\mathbf{x}$
$\lvert\mathcal{A}\rvert$	Cardinality of set $\mathcal{A}$
$\mathbb{E}\{\mathbf{A}\}$	Expectation of matrix $\mathbf{A}$
$\mathrm{tr}\{\mathbf{A}\}$	Trace of matrix $\mathbf{A}$
$\mathrm{diag}(\mathbf{x})$	Diagonal matrix with diagonal elements equal to the entries of vector $\mathbf{x}$
$\mathbf{I}_N$	N-dimensional identity matrix
$\mathbf{0}$	Zero vector or zero matrix with proper dimension
$\mathrm{vec}(\mathbf{A})$	Vectorization of matrix $\mathbf{A}$
$[\mathbf{x}_1 \dots \mathbf{x}_L]$	Horizontal concatenation of vectors $\mathbf{x}_1, \dots \mathbf{x}_L$
$\mathbf{A} \otimes \mathbf{B}$	Kronecker product of matrices $\mathbf{A}$ and $\mathbf{B}$
$\mathcal{N}(\mu, \sigma^2)$	Real Gaussian random variable with mean μ and variance σ^2

$\mathcal{CN}(\mu, \sigma^2)$	Circularly symmetric complex Gaussian random variable with mean μ and variance σ^2
$x \sim X$	Variable x follows distribution X
$\arg\min(.)$	Argument minimizing the expression in parentheses
$\arg\max(.)$	Argument maximizing the expression in parentheses
$\xrightarrow{a.s.}$	Convergence with probability 1 or almost sure convergence
$\mathbf{i} = \sqrt{-1}$	imaginary unit

Chapter 1

Introduction

Massive MIMO technology plays a key role in 5G systems providing the demand for higher data rates and traffic volumes far above previous technologies, and also reduces the latency of the data connections [1]. One of the major bottlenecks of massive MIMO systems as well as of any cellular network is the inter-cell interference, which significantly affects the cell-edge users performance, whose inter-cell performance is already degraded by the path attenuation. To address these limitations and provide uniformly service to all the users, beyond-5G networks need to enter the cell-free (CF) paradigm, where the absence of cell boundaries annihilates the inter-cell interference and handover issues but also causes new challenges [2].

The CF terminology was coined by Yang and Marzetta in [3], while the name CF Massive MIMO first appeared in [4]. CF massive MIMO systems have been drawing extensive research interests as an effective and promising approach for next generation wireless systems thanks to their potential to reap the benefit of both massive MIMO and distributed antenna systems (DAS) [4, 5]. In principle, CF massive MIMO is an embodiment of general ideas known as "virtual MIMO" [6], "network MIMO" [7], "distributed MIMO" [8], "(coherent) cooperative multi-point joint processing" (CoMP) [9] and "distributed antenna systems" [10]. The objective is to use advanced back-haul to achieve coherent processing across geographically distributed base station (BS) antennas, in order to provide uniformly good service for all users in the network

[5]. In [11], a generalized DAS was proposed and called as distributed MIMO system. Distributed MIMO system combines the advantages of point-to-point MIMO and DAS, and thus has the ability of exploiting both spatial micro and macro-diversities [11, 12]. The term "network MIMO" incorporates the principle that multiple cells cooperate to act as a single network and serve all the users in its joint coverage area. Hence, generally, network MIMO is a synonym of distributed MIMO, and includes all its subsequent variants [13]. More precisely, the concept of network MIMO emerged in the early 2000s mainly in the form of cooperative MIMO BSs serving all the users in their range of influence using multi-user MIMO processing techniques [14, 15]. It is worth noting that CoMP is a network MIMO with coordinated beamforming. In contrast to the network MIMO, the cellular or cell boundary concepts disappear in CF massive MIMO and the users are served simultaneously by all antennas, hence the name.

A CF massive MIMO system is composed of a massive number of antennas distributed over a wide area, called access points (APs). All the APs are connected through a back-haul network to a central processing unit (CPU) and jointly and coherently serve a relatively small number of single-antenna users over the same time-frequency resources. In CF massive MIMO, the massive number of antennas improves spectral efficiency [5] whereas energy efficiency [16, 17] and macro-diversity gain result from the distributed topology and ultra-densification [2]. Additionally, since each user is surrounded by a large number of serving APs, with high probability all the users enjoy good channel conditions [18]. Therefore, CF massive MIMO systems are expected to provide significant improvements in terms of spectral/energy efficiency and coverage probability. In [19], the energy efficiency of massive MIMO systems was analyzed for both cellular and CF scenarios. It was shown that the energy efficiency in CF massive MIMO can be improved by nearly ten times compared to traditional cellular massive MIMO. The achievable spectral efficiencies of CF massive MIMO have been analyzed in the early works [5], [20], considering single-antenna APs, single-antenna users, and Rayleigh fading channels. In successive works, more realistic scenarios such as single-

antenna APs with Rician fading channels [21], multi-antenna APs with correlated [22] or uncorrelated [23, 24] fading channels, and multi-antenna users [25] have been considered. Four different ways to divide the signal processing between the APs and CPU are considered in [22]. It has been shown that CF massive MIMO systems outperform conventional cellular massive MIMO and small-cell systems where each AP serves its own exclusive set of users, through various practical scenarios [5][20][22][26, 27]. In [5][20], the comparison was conducted under the assumption of employing maximum ratio (MR) processing, i.e., conjugate beamforming on the downlink and matched filtering on the uplink. In [22], [28–30], the authors advocated the use of more effective processing than MR processing in CF massive MIMO to guarantee superior performance of CF massive MIMO systems compared to traditional cellular massive MIMO and small-cell counterparts.

The capacity per unit area of DASs in uplink has been analyzed in [31,32] leveraging on a mathematical framework based on Euclidean random matrices (ERM) [33] and assuming that the network dimensions tend to infinity. To reap the benefits promised by this analysis completely, the use of a centralized optimal joint processing is crucial. However, an optimal maximum likelihood (ML) detector is essentially an exhaustive search method and has an unaffordable complexity for large systems.

1.1 Favorable Propagation

In centralized MIMO systems, the high complexity of centralized joint detectors has been successfully addressed by massive MIMO systems [34]. One of the key properties exploited in massive MIMO is that as the number of BS antennas grows the channel vectors associated with different users, i.e, channel vectors between the BS and the users, tend to become jointly orthogonal determining a phenomenon known as *favorable propagation* [35–37]. The favorable propagation makes linear signal processing schemes effective and nearly optimal, hence considerably simplifying the complexity. Under orthogonality conditions, low complexity matched filters (MF) are optimum

and asymptotically attain the same performance of ML detectors. More explicitly, on the uplink, with a simple linear detector such as MF, interference and noise can be canceled out. On the downlink, with linear beamforming techniques, the BS can simultaneously beamform multiple data streams to multiple user terminals without causing mutual interference [35, 36] [38]. The validity of the favorable conditions has been investigated for several massive MIMO settings [39].

With a massive number of APs, CF massive MIMO can exploit favorable propagation resulting from mutual orthogonality of the channels of different users [35]. In [5], favorable propagation was leveraged to derive closed-form expressions for the downlink and uplink achievable rates in CF massive MIMO systems. The spatial correlation resulting from the distributed deployment of APs may however have a detrimental impact on the favorable propagation. More specifically, users that are relatively closed to each other will incur high spatial correlation which will endanger the mutual orthogonality of the users' channel. In massive MIMO systems with centralized BSs, the assumption of Rayleigh fading provides realistic guidelines for system design. However, in CF massive MIMO systems where the APs are massively distributed and several of them could be very close to users and in direct line of sight (LoS), it becomes relevant to investigate the effects of LoS and path loss on the property of favorable propagation.

An initial numerical analysis for CF massive MIMO in Rayleigh fading was presented in [23]. In [23], a thorough investigation of the favorable propagation phenomenon in CF massive MIMO systems from a stochastic geometry perspective was provided. It was shown that one may not completely rely on favorable propagation when assessing the system performance since the derived bounds may not be tight due to the impact of spatial correlation between some users. In [40], the impact of the network configuration on the level of favorable propagation for a CF Massive MIMO network was investigated. They analyzed how spatial correlation between users' channels vector influences favorable propagation and explored how to improve orthogonality between users' channel by taking into account solely the large-scale fading and the number of available APs.

1.2 Pilot Contamination

Channel state information (CSI) in multiple antenna systems, both cellular and CF, is crucial for accomplishing successful transmission under various channel conditions [41]. Ideally, training sequences or pilots should be selected to be mutually orthogonal in the channel estimation. However, in most practical scenarios the number of users is greater than the number of orthogonal training sequences and a given training sequence can be assigned to more than one user, therefore leading to the so-called *pilot contamination* which prevents the possibility of obtaining an adequate estimate of CSI.

Pilot contamination is a major problem in massive MIMO, which is caused by non-orthogonality of pilot sequences used in adjacent cells. Usually, reusing pilots in multiple cells is the main cause of the problem. In this case, the estimated channel vector in any cell is the summation of all the channel vectors of users from the neighboring cells in addition to the original cell. As the number of interfering cells increases, the problem exponentially grows and eventually causes system malfunction [42]. The pilot contamination phenomenon in the context of centralized massive MIMO systems has been widely studied, see, e.g., [37] [43, 44]. Specific features of centralized massive MIMO channels such as channel hardening and favorable propagation or limited angular spread could be exploited to "separate" user channels in power domain [45], angular domain [46,47], or jointly in power and angular domain [48] and thus, mitigate or annihilate pilot contamination. However, these appealing properties of channels in centralized massive MIMO systems are destroyed in a distributed setting and pilot contamination is still an open and challenging problem in CF massive MIMO systems.

Pilot contamination has received a lot of research interests in the literature, since it can substantially compromise the system performance of CF massive MIMO. Several pilot assignment (PA) techniques for suppressing pilot contamination in a CF massive MIMO have been proposed recently. The problem of PA in CF massive MIMO was firstly investigated in [5], where, starting from a random PA, a greedy pilot assignment (GPA) based on the knowledge of the large-scale fading (LSF) channel coefficients was

proposed that iteratively updates the pilot of the user with the lowest achievable rate, i.e., the worst performing user, in order to enhance the system fairness. In [49], a location-based greedy (LBG) pilot assignment scheme was proposed to use the location information of the users to assign the pilot sequences instead of randomly assigning before using the GPA algorithm. Similarly, patent [50] proposed an iterative algorithm based on consecutive updates of the pilots for the worst and best performing users, again aiming at the maximization of the system fairness. In [51], a structured PA algorithm based on the knowledge of the users' positions was proposed that maximizes the minimum geographical distance between users sharing the same pilot sequences and in [52] a similar procedure was considered taking into account the large-scale fading channel coefficients between users and APs. Graph coloring based pilot assignment schemes proposed in [53] and [54]. In [53], using the AP selection method an interference graph was constructed to describe the interference relationship among users and an efficient pilot assignment scheme based on the graph coloring was proposed to significantly reduce the impact of pilot contamination on the system throughput. To avoid being trapped in a local optimum, tabu search is used in [55], where the tabu list records previous assignments to ensure the efficient search of the assignment solution space. The authors of [56] proposed an iterative approach based on the Hungarian algorithm. In each iteration, each users and its neighboring users are assigned mutual orthogonal pilots by exploiting the Hungarian algorithm, given the pilot assignment of the rest of the users is fixed. The final assignment is achieved when the performance measures reach convergence or the iterations reach the allowed maximum number.

All aforementioned techniques address the pilot contamination problem via a careful assignment of pilots and do not exploit the inherent structure of channels and data in CF massive MIMO systems in contrast to blind or semi-blind estimation and detection techniques. The blind estimation is fully based on the statistical properties of the transmitted data, whereas the semi-blind estimation depends on the joint use of pilots and data. A blind pilot decontamination approach was firstly proposed in [45] for centralized massive MIMO systems and utilized asymptotic orthogonality of

user channels to remove undesired interference including pilot contamination from the received signal. The same property was also exploited for semi-blind channel estimation, e.g., [48], in centralized massive MIMO but it does not hold in CF massive MIMO systems [29, 30]. Semi-blind channel estimation has been investigated in several papers, e.g., [57–60] and references therein. In [57], a semi-blind scheme based on a whitening-rotation decomposition of the channel matrix was proposed for MIMO flat-fading channel estimation. The MIMO channel matrix is decomposed into the product of a whitening matrix estimated blindly and a rotational unitary matrix estimated using training symbols. In [58], a semi-blind channel estimation technique for MIMO systems was introduced, which uses an iterative two-level optimization loop to jointly estimate channel coefficients and data symbols. In [59], a performance comparison between blind, semi-blind and training-sequence based channel estimation was proposed in terms of Cramer-Rao bound (CRB) for a deterministic and also a Gaussian symbol model. In [60], the authors studied two semi-blind channel estimators for SIMO systems based on ML estimation with deterministic and Gaussian models. The asymptotic performances of the estimators in [59, 60] were studied when the length of the training sequences and data sequences grow infinitely large. A semi-blind channel estimation method based on the expectation maximization (EM) algorithm was proposed and analyzed in [61]. A Gaussian distribution was considered for the unknown data symbols in the EM algorithm, which enables deriving a closed form solution for the expectation evaluation of the EM algorithm. They further derived deterministic and stochastic CRB for semi-blind channel estimation and studied their behavior in massive MIMO systems with unlimited number of antennas at the BS.

In this context, the concept of identifiability is very relevant since it guarantees the non-singularity of the Fisher information matrix (FIM) and thus, the existence of the CRB. The corresponding conditions provide fundamental insights into the feasibility of reliable communications in the analyzed system. Conditions under which channel and data signals are blindly and semi-blindly identifiable have been thoroughly studied in various settings for centralized systems, see, e.g., [62, 63]. In [63], the authors studied

the conditions under which the channel and the data signals are blindly and semi-blindly identifiable for an under-determined MIMO system. They obtained blind and semi-blind channel estimates based on the EM algorithm in the frequency domain and utilized a discrete random variable model for the unknown data.

1.3 Expectation Propagation

Recently, a wide research has been dedicated to the design of detectors relying on message passing (MP) algorithms. Expectation propagation (EP) which is a kind of MP algorithm is a generalization of sum-product belief propagation (BP). BP algorithm [64] which is a graph-based statistical inference technique approximates the posterior distributions of transmitted symbols by iterative updating and obtains good performance but still has high computational complexity. Compared to BP utilizing the sample values of a distribution, EP algorithm [65, 66] focusing on sufficient statistics, iteratively finds the best approximation for a computationally intractable target probability distribution from a tractable family of distributions [67]. Therefore, EP algorithm has a low computational complexity and high performance compared to BP algorithm [68]. The EP tries to find the closest approximation for the conditional marginal distribution of a desired variable in an iterative refinement procedure. The EP algorithm was first proposed in [69] and summarized in, e.g., [70] for approximate inference in probabilistic graphical models. The method of EP was firstly applied to MIMO detection in [68], where an EP-based MIMO detector shows near-optimal performance for all kinds of antenna configurations. With the EP-based MIMO detector, a Gaussian approximation is constructed for the posterior distribution of the transmitted symbols by an iterative procedure based on moment matching. The original EP detection algorithm for massive MIMO shows a near-optimal performance but still suffers from the high computational complexity caused by the matrix inversion in each iteration. In [71], an iterative successive updating scheme was proposed to reduce this computational complexity and improve the efficiency and accuracy of messages

updating to accelerate the convergence of the EP algorithm.

EP has been proposed as a low complexity algorithm for symbol detection in massive MIMO systems [72] [73]. In [72], the EP principle was exploited for designing efficient detector of extra-large-scale massive MIMO systems with the subarray-based processing architecture. They represented the a posteriori distribution as a factor graph and developed the iterative algorithm by computing and transferring messages among different nodes on the factor graph. A non-coherent detection scheme for SIMO systems based on the EP algorithm was proposed in [73]. The proposed EP detector iteratively searches for the best approximation of the joint probability density function of the channel coefficients and the transmitted symbols. The output probability density function is used for direct estimation of the channel coefficients, as well as the transmitted symbols.

An efficient data detection algorithm with affordable complexity to reach the optimum performance is of highly interest in large-scale networks, such as CF massive MIMO. In this aspect, some early works were focused on centralized algorithms where the detection is totally implemented at the CPU with the received pilots and data signals sent by all APs [5][74]. However, the computational overhead of such a centralized detection scheme is prohibitively high as the size of network becomes large. To address this challenge, distributed detectors have been recently investigated in CF massive MIMO systems. In [22], a centralized and three distributed receivers with different levels of cooperation among APs were compared in terms of spectral efficiency. However, the distributed receivers investigated in [22] are linear receivers and therefore highly sub-optimal in terms of the bit error rate (BER) performance. In [75], a non-linear detector for CF massive MIMO networks was proposed which is derived based on the EP principle [69] with a distributed approach [72][76]. It was shown that such detector can achieve better performance than other linear receivers for both original and scalable CF massive MIMO networks and compared to other distributed detectors, it can achieve a better BER performance.

1.4 Thesis Outline and Main Contributions

In this section, an outline of this dissertation is provided by highlighting the key concepts and contributions detailed in each chapter.

Chapter 2

In this chapter, we analyze favorable propagation conditions in CF massive MIMO systems through the characteristics of the channel eigenvalue moments for the two extreme cases with all transmit and receive antennas in LoS or in non-LoS and Rayleigh fading. We model APs and users as two independent uniform point processes (PPs) over a regular grid. We show analytically that the favorable propagation conditions are not satisfied in CF massive MIMO systems with APs and users in LoS. On the contrary, they hold in the case of path loss plus multipath Rayleigh fading. We analyze the performance of polynomial expansion detectors and multi-stage Wiener filters (MSWF) and show that when matched filtering is not almost optimum, the use of low complexity linear multi-user detection becomes very appealing in practical systems.

The work in this chapter has resulted in the following publications

1. **R. Gholami**, L. Cottatellucci, and D. Slock, "Channel Models, Favorable Propagation and Multi-Stage Linear Detection in Cell-Free Massive MIMO," *in 2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 2942–2947.

2. **R. Gholami**, L. Cottatellucci, and D. Slock, "Favorable Propagation and Linear Multi-User Detection for Distributed Antenna Systems," *in ICASSP 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5190–5194.

3. **R. Gholami**, S. F. Islam, S. Mahama, D. Slock, L. Cottatellucci, A. Burr, and D. Grace, "Cell-Free MIMO Systems for UDNs," in Enabling 6G Mobile Networks. Springer, 2022, pp. 39–69.

Chapter 3

In this chapter, we address the problem of pilot contamination in CF massive MIMO systems leveraging only the channel sparsity. We develop semi-blind techniques for joint channel estimation and data detection to mitigate pilot contamination. We analyze the potential of semi-blind approaches with classical signal processing techniques such as FIM, CRB, and identifiability. Additionally, we determine sufficient and necessary conditions for semi-blind identifiability under the assumption of deterministic parameters. We construct a bipartite graph that has APs and users as factor and variable nodes and propose an MP algorithm over this graph to verify identifiability and compute the channel coefficients.

The work in this chapter has resulted in the following publication

1. **R. Gholami**, L. Cottatellucci, and D. Slock, "Tackling Pilot Contamination in Cell-Free Massive MIMO by Joint Channel Estimation and Linear Multi-User Detection," *in 2021 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2021, pp. 2828-2833.

Chapter 4

In this chapter, we consider semi-blind methods for channel estimation in the presence of Gaussian i.i.d. data to resolve the pilot contamination. This task is further aided by exploiting prior channel information in a Bayesian formulation. We propose a new variable level expectation propagation (VL-EP) algorithm for MP style semi-blind channel estimation which provides an approximate minimum mean square error (MMSE) channel estimator which itself can not be found analytically.

The work in this chapter has resulted in the following publication

1. **R. Gholami**, L. Cottatellucci, and D. Slock, "Message Passing for a Bayesian Semi-Blind Approach to Cell-Free Massive MIMO," *Proc. of 55th Annual Asilomar Conference on Signals, Systems, and Computers (ACSSC)*. IEEE, 2021.

Chapter 5

This chapter concludes the dissertation discussing the contributions and drawing guidelines for future developments in this field.

Chapter 2

Favorable Propagation Analysis and Multi-Stage Linear Detection

2.1 Introduction

In centralized MIMO systems, the high complexity of centralized joint detectors has been successfully addressed by massive MIMO systems [34]. One of the key properties exploited in massive MIMO systems is that as the number of antennas at the BS grows, the channel vectors associated with different users, i.e., channel vectors between the BS and the users, tend to become jointly orthogonal determining a phenomenon known as *favorable propagation* [35, 36]. The favorable propagation makes linear processing achieve optimality and maximize the information rate. Under orthogonality conditions, low complexity MFs are optimum and asymptotically attain the same performance of ML detectors.

In massive MIMO systems with centralized BSs, the assumption of Rayleigh fading provides realistic guidelines for system design. However, in CF massive MIMO systems where the APs are massively distributed and several of them could be very close to users and in direct LoS, it becomes relevant to investigate the effects of LoS and path loss on the property of favorable propagation.

In this chapter, we analyze favorable propagation conditions in CF massive MIMO systems expressed in terms of channel eigenvalue moments for the two extreme cases with all transmit and receive antennas in LoS or in non-LoS and Rayleigh fading. We model APs and users as two independent uniform point processes (PPs) over a regular grid. Under this assumption, the inclusion of path loss and LoS or Rayleigh fading leads to classes of random matrices similar to the ERMs proposed in [31, 32]. We show analytically that the favorable propagation conditions are not satisfied in CF massive MIMO systems with APs and users in LoS. Conversely, they hold in the case of path loss plus multipath Rayleigh fading. When matched filtering is not almost optimum, the use of low complexity linear multi-user detection becomes very appealing in practical systems. Then, by extending the unified analytical framework proposed in [77], we analyze the performance of polynomial expansion detectors in [78] and multi-stage Wiener filters (MSWF) in [79].

2.2 System Model

Throughout this chapter, we will consider both networks over 1-dimensional (1D) and 2-dimensional (2D) spaces. The corresponding models and notation are discussed in parallel. We consider a distributed antenna system in uplink with users and APs equipped with a single antenna and independently and uniformly distributed over a squared box of side L and area $A = L^2$ in $\mathbb{R}^2$, denoted by $\mathcal{A}_L = \left[-\frac{L}{2}, +\frac{L}{2}\right) \times \left[-\frac{L}{2}, +\frac{L}{2}\right)$ and a segment of length L in $\mathbb{R}$, denoted by $\widetilde{\mathcal{A}}_L = \left[-\frac{L}{2}, +\frac{L}{2}\right)$.

For the sake of analytical tractability, we assume that users and APs are located on a grid in $\mathcal{A}_L$ ($\widetilde{\mathcal{A}}_L$). Let $\tau > 0$ be an arbitrary small real such that $L = \theta\,\tau$ with θ positive, even integer. Let $\mathbf{w} = \big((-\theta+2w_x)\tau/2, (-\theta+2w_y)\tau/2\big)$ or $w = (-\theta+2w_x)\tau/2$, with $w_x, w_y \in \mathbb{Z}$, be points of a regular grid in $\mathbb{R}^2$ or $\mathbb{R}$, respectively. We denote by $\mathcal{A}_L^{\#}$ ($\widetilde{\mathcal{A}}_L^{\#}$) the set of points regularly spaced in $\mathcal{A}_L$ ($\widetilde{\mathcal{A}}_L$) by τ, i.e., $\mathcal{A}_L^{\#} \equiv \big\{\mathbf{w} | \mathbf{w} \in \mathcal{A}_L,\, w_x, w_y = 0,\, 1, \ldots \theta - 1\big\}$ $\big(\widetilde{\mathcal{A}}_L^{\#} \equiv \big\{w | w \in \widetilde{\mathcal{A}}_L,\, w_x = 0,\, 1, \ldots \theta - 1\big\}\big)$. The set $\mathcal{A}_L^{\#}$ is illustrated in Fig. 2.1. We model the distributed users and APs as homogeneous

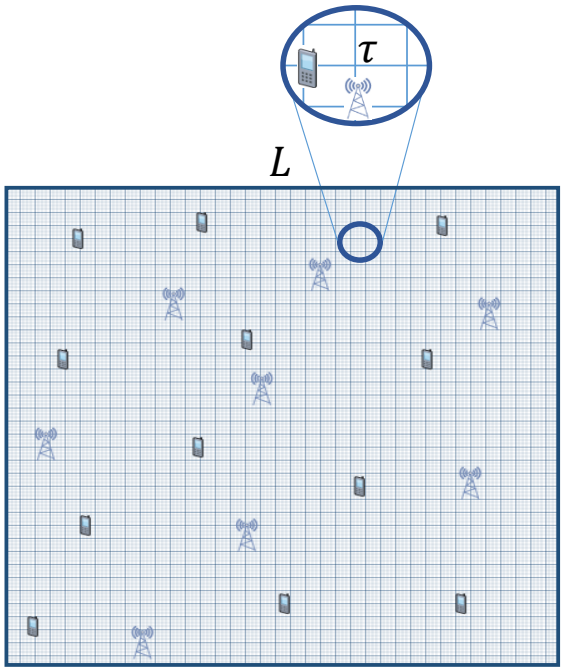

Figure 2.1: Representation of $\mathcal{A}_L^{\#}$. The large squared box represents the network surface.

PPs $\Phi_{\mathcal{T}}$ and $\Phi_{\mathcal{R}}$ in $\mathcal{A}_L^{\#}$ characterized by parameters $\beta_T = \rho_T \tau^2$ and $\beta_R = \rho_R \tau^2$, where ρ_T and ρ_R are the intensities, i.e., the number per unit area, of users and APs, respectively. Then, $N_T = \rho_T L^2 = \beta_T \theta^2$ and $N_R = \rho_R L^2 = \beta_R \theta^2$ are the number of users and APs, respectively. Similarly, in $\mathbb{R}$, the two homogeneous PPs $\widetilde{\Phi}_{\mathcal{T}}$ and $\widetilde{\Phi}_{\mathcal{R}}$ in $\widetilde{\mathcal{A}}_L^{\#}$ are characterized by parameters $\widetilde{\beta}_T = \widetilde{\rho}_T \tau$ and $\widetilde{\beta}_R = \widetilde{\rho}_R \tau$, and the number of users and APs are $\widetilde{N}_T = \widetilde{\rho}_T L = \widetilde{\beta}_T \theta$ and $\widetilde{N}_R = \widetilde{\rho}_R L = \widetilde{\beta}_R \theta$, respectively.

All the APs are connected to a CPU via a back-haul network such that detection is performed jointly. Users transmit at the same power p. At the central processing unit, the discrete time N_R-dimensional received signal vector is given by

$$\mathbf{y} = \sqrt{p}\, \mathbf{G}\, \mathbf{x} + \mathbf{n}, \tag{2.1}$$

where $\mathbf{x} = [x_1\ x_2\ \ldots\ x_{N_T}]^T$ is the N_T-dimensional column vector of independent and identically distributed (i.i.d.) transmitted symbols, x_j is the unitary energy symbol transmitted by user j, i.e., $\mathbb{E}\{|x_j|^2\} = 1$; $\mathbf{G}$ is the $N_R \times N_T$ matrix of channel coefficients whose (i,j)-element $g_{ij} = g(\mathbf{r}_i, \mathbf{t}_j)$ denotes the channel coefficient between transmitter j and receiver i with Euclidean coordinates $\mathbf{t}_j = (t_{x,j}, t_{y,j})$ and $\mathbf{r}_i = (r_{x,i}, r_{y,i})$, respec-

tively. The N_R-dimensional vector $\mathbf{n}$ denotes the complex additive white Gaussian noise (AWGN) vector with i.i.d. components having zero mean and variance σ^2.

In order to define the channel coefficients, we introduce the path loss matrix $\widehat{\mathbf{G}}$ with (i,j) element given by

$$\hat{g}_{ij} = \hat{g}(\mathbf{r}_i, \mathbf{t}_j) = \begin{cases} \dfrac{d_0^\alpha}{\|\mathbf{r}_i - \mathbf{t}_j\|_2^\alpha} & \text{if } \|\mathbf{r}_i - \mathbf{t}_j\|_2 > d_0 \\ 1 & \text{otherwise,} \end{cases} \tag{2.2}$$

where d_0 is a reference distance, α is the path loss exponent and $\|\mathbf{r}_i - \mathbf{t}_j\|_2$ denotes the Euclidean distance between transmit antenna j and receive antenna i. We ignore the shadowing effect and model the large-scale fading as pure path loss. At short distances between APs and users, i.e., for $\|\mathbf{r}_i - \mathbf{t}_j\|_2 \leq d_0$, the transmit signal is amplified beyond the transmit signal level and the amplification presents a vertical asymptote for $\|\mathbf{r}_i - \mathbf{t}_j\|_2 \to 0$. In order to remove this artifact while keeping the model simple [1], we assume negligible the signal attenuation in a close neighborhood of a transmitter and we fix the attenuation equal to $\hat{g}_{ij} = 1$.

In the case of antennas in LoS, the channel coefficients impaired by path loss are given by $g_{ij} = g(\mathbf{r}_i, \mathbf{t}_j) = \hat{g}(\mathbf{r}_i, \mathbf{t}_j) \exp(-\mathbf{i}2\pi\lambda^{-1}\|\mathbf{r}_i - \mathbf{t}_j\|_2)$, where the phase rotation depending on the distance $\|\mathbf{r}_i - \mathbf{t}_j\|_2$ is given by $\exp(-\mathbf{i}2\pi\lambda^{-1}\|\mathbf{r}_i - \mathbf{t}_j\|_2)$ and λ denotes the radio signal wavelength. In the case of non-LoS channels with Rayleigh fading, the channel coefficients are given by $g_{ij} = \hat{g}(\mathbf{r}_i, \mathbf{t}_j)\, h_{ij}$, where $h_{ij} \sim \mathcal{CN}(0,1)$ are i.i.d. complex Gaussian variables modeling the small-scale fading.

Similarly, in $\mathbb{R}$, the channel coefficient matrix denoted by $\mathbf{G}_{1D}$ is of size $\widetilde{N}_R \times \widetilde{N}_T$ whose (i,j)-element $g_{1D}(r_i, t_j)$ denoting the channel coefficient between transmitter j and receiver i with Euclidean distance $|r_i - t_j|$, in the case of LoS and non-LoS channels with Rayleigh fading is $g_{1D}(r_i, t_j) = \hat{g}_{1D}(r_i, t_j) \exp(-\mathbf{i}2\pi\lambda^{-1}|r_i - t_j|)$ and $g_{1D}(r_i, t_j) = \hat{g}_{1D}(r_i, t_j)\, h_{ij}$, respectively. $\hat{g}_{1D}(r_i, t_j)$ denotes the (i,j)-th element of the path loss matrix $\widehat{\mathbf{G}}_{1D}$ in $\mathbb{R}$, given by

[1] It is worth noticing that in contrast to the approach in [31, 32], the analysis proposed in this chapter can be applied to any path loss model which admits a Fourier transform and it is not restricted to (2.2).

$$\hat{g}_{ij,1D} = \hat{g}_{1D}(r_i, t_j) = \begin{cases} \dfrac{d_0^\alpha}{|r_i - t_j|^\alpha} & \text{if } |\mathbf{r}_i - \mathbf{t}_j| > d_0 \\ 1 & \text{otherwise,} \end{cases} \tag{2.3}$$

2.3 Preliminary Mathematical Tools

In this section, we introduce mathematical tools for analyzing the favorable propagation conditions in CF massive MIMO systems. Communication systems modeled by random channel matrices can be efficiently studied via their covariance eigenvalue spectrum [80, 81]. In the subsequent section, we derive a tight approximation of the eigenvalue moments of the channel covariance matrix that can be efficiently applied to the analysis of favorable propagation properties in CF massive MIMO systems and the design and analysis of multi-stage linear detectors.

In the following subsection, we recall some theoretical concepts necessary to derive the results in this chapter.

2.3.1 Free Probability Theory

Free probability is a mathematical theory that studies non-commutative random variables while classical probability theory is concerned with commutative random variables. Free probability theory was initiated by Dan Voiculescu in the 1980's in order to attack the free group factors isomorphism problem, an important unsolved problem in the theory of operator algebras [82–84]. A few years later, in 1991, Voiculescu discovered the relation between random matrices and free probability [85]. An interesting aspect and active research direction of free probability lies in its applications to random matrix theory (RMT) [86] [87]. Free probability theory provides a very efficient framework to study limiting distributions of some models of large dimensional random matrices. This theory introduces freeness between non-commutative random variables, which is analogous to the independence between classical commutative random variables. Free probability has developed many powerful tools from classical probability to provide new ideas to study random matrices.

In the following, we briefly describe some basic concepts which play important roles in free probability theory.

Definition 2.1 (Freeness Definition [88]) *Non-commutative random variables* $x_1, x_2, \ldots x_n$ *are called free or freely independent, if for any* m *polynomials* $p_k(x),\ 1 \leq k \leq m$, *with* $m \geq 2$,

$$\mathbb{E}\{p_1(x_{i_1})\ p_2(x_{i_2})\ \ldots p_m(x_{i_m})\} = 0, \tag{2.4}$$

when $\mathbb{E}\{p_k(x_{i_k})\} = 0$ *for all* k, $1 \leq k \leq m$, *and any two neighboring indices* i_l *and* i_{l+1} *are not equal, i.e.,* $1 \leq i_1 \neq i_2 \neq \ldots \neq i_m \leq n$.

Then, as random variables $x_1, x_2, \ldots x_n$ are freely independent, it implies a factorization rule for calculating any mixed moments in $x_1, x_2, \ldots x_n$ in terms of the moments of individual x_{i_k}'s.

One should note that free independence is a different rule from classical independence; free independence occurs typically for non-commuting random variables, like operators on random matrices. According to free probability theory, the random matrices of independent Gaussian random variables approximately become free when the matrix size goes to infinity.

Definition 2.2 (Non-Crossing Partitions [89]) *For a positive integer* n, *consider the ordered set* $\mathfrak{N}_n = \{1, 2, \ldots, n\}$. *A partition* π *of set* $\mathfrak{N}_n$ *means* $\pi = \{V_1, V_2, \ldots V_s\}$ *such that* $V_1, V_2, \ldots V_s \subset \mathfrak{N}_n$ *with*

$$V_i \neq \emptyset, \qquad V_i \cap V_j = \emptyset\ (1 \leq i \neq j \leq n), \qquad V_1 \cup \ldots \cup V_s = \mathfrak{N}_n. \tag{2.5}$$

Subsets $V_1, V_2, \ldots V_s$ *are called the blocks of* π. $\mathcal{P}(n)$ *denotes the set of all the partitions of* $\mathfrak{N}_n$.

Let $\pi \in \mathcal{P}(n)$. *If there exist* $i < j < k < l$ *such that* i *and* k *are in one block* V *of* π, *and* j *and* l *in another block* W *of* π, V *and* W *are called cross. If one cannot find any pair of blocks in* π *that cross, partition* π *is called non-crossing. The set of all non-crossing partitions of* $\mathfrak{N}_n$ *is denoted by* $NC(n)$.

A look at Fig. 2.2 should explain the terminology "non-crossing": one puts the

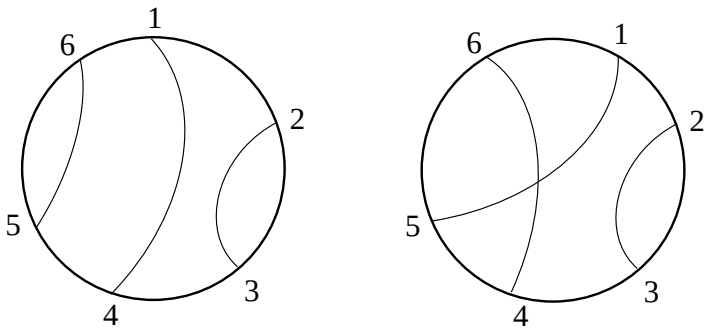

Figure 2.2: Non-crossing (left, $\pi = \big\{\{1,4\},\{2,3\},\{5,6\}\big\}$) and crossing (right, $\pi = \big\{\{1,5\},\{2,3\},\{4,6\}\big\}$) partitions of the set $\mathfrak{N}_6$.

points $1, 2, \ldots n$ on the circle, and connects each point with the next member of its block (in cyclic order) by an internal path. Then, the partition is non-crossing if this can be achieved without arcs crossing each other [90].

Definition 2.3 (Bipartite Graph) *In the mathematical field of graph theory, a bipartite graph is a graph whose vertices can be divided into two disjoint and independent sets U and V such that every edge connects a vertex in U to one in V. Vertex sets U and V are usually called the parts of the graph. Equivalently, a bipartite graph is a graph that does not contain any odd-length cycles.*

Definition 2.4 (Even Graph) *A sequence or graph is said to be even, whenever an edge from j to l appears in the graph (possibly a certain number of times), then it should also appear the same number of times in the opposite direction from l to j.*

In the following, we define Toeplitz and circulant matrices playing a fundamental role in developing the results of this chapter.

Definition 2.5 (Toeplitz Matrix [91]) *A Toeplitz matrix is an $n \times n$ matrix $\mathbf{T}_n = [t_{k,j};\ k, j = 0, 1, \ldots n-1]$ where $t_{k,j} = t_{k-j}$, i.e., a matrix of the form*

$$\mathbf{T}_n = \begin{bmatrix} t_0 & t_{-1} & t_{-2} & \cdots & t_{-(n-1)} \\ t_1 & t_0 & t_{-1} & \cdots & t_{-(n-2)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ t_{n-1} & & \cdots & & t_0 \end{bmatrix} \tag{2.6}$$

A common special case of Toeplitz matrices which results in significant simplification and play a fundamental role in developing more general results is circulant matrix when every row of the matrix is a right cyclic shift of the row above it so that $t_k = t_{-(n-k)} = t_{k-n},\ k = 1, 2, \ldots n-1$. *More specifically, a circulant matrix* $\mathbf{C}_n$ *is having the following form*

$$\mathbf{C}_n = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \cdots & \vdots \\ & c_{n-1} & c_0 & c_1 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ c_1 & c_2 & \cdots & & c_0 \end{bmatrix} \tag{2.7}$$

Block Toeplitz/circulant matrices are defined similarly, except that the structure refers to block, rather than elements.

Definition 2.6 (Block Toeplitz Matrix [92]) *An* $m \times n$ *block Toeplitz matrix with* $M \times N$ *blocks is an* $mM \times nN$ *matrix of the form*

$$\begin{bmatrix} \mathbf{A}_0 & \mathbf{A}_{-1} & \mathbf{A}_{-2} & \cdots & \mathbf{A}_{-(n-1)} \\ \mathbf{A}_1 & \mathbf{A}_0 & \mathbf{A}_{-1} & \cdots & \mathbf{A}_{-(n-2)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{A}_{m-1} & \mathbf{A}_{m-2} & \mathbf{A}_{m-3} & \cdots & \mathbf{A}_0 \end{bmatrix} \tag{2.8}$$

where $\mathbf{A}_k \in \mathbb{C}^{M \times N}$ *with* $1 - n \leq k \leq m - 1$.

When $M = N = 1$, the matrix in (2.8) is simply an $m \times n$ Toeplitz matrix.

2.4 Channel Eigenvalue Moments

The eigenvalue moments or, shortly, the moments of the channel covariance matrix $\mathbf{C} = \mathbf{G}^H\mathbf{G}$ are defined as follows

$$m_{\mathbf{C}}^{(n)} = \int \mu^n \mathrm{d}F_{\mathbf{C}}(\mu) = \frac{1}{N_T}\mathbb{E}\{\mathrm{tr}(\mathbf{C}^n)\} \qquad n \in \mathbb{N} \tag{2.9}$$

where μ and $F_{\mathbf{C}}(\mu)$ denote the eigenvalue and empirical eigenvalue distribution of matrix $\mathbf{C}$, respectively. The expectation $\mathbb{E}\{.\}$ is with respect to (w.r.t.) the two ho-

mogeneous point processes $\Phi_{\mathcal{T}}$ and $\Phi_{\mathcal{R}}$ ($\widetilde{\Phi}_{\mathcal{T}}$ and $\widetilde{\Phi}_{\mathcal{R}}$ in $\mathbb{R}$).

Following the approach in [31, 32][93], we decompose the path loss matrix, $\widehat{\mathbf{G}}$, as follows

$$\widehat{\mathbf{G}} = \mathbf{\Psi}_R \widehat{\mathbf{T}} \mathbf{\Psi}_T^H \tag{2.10}$$

where $\widehat{\mathbf{T}}$ is a $\theta^2 \times \theta^2$ matrix depending only on the function $\hat{g}(\mathbf{r}_i, \mathbf{t}_j)$, $\mathbf{\Psi}_R$ and $\mathbf{\Psi}_T$ are $N_R \times \theta^2$ and $N_T \times \theta^2$ random matrices depending only on random APs' and users' locations, respectively.

In order to define the matrices $\mathbf{\Psi}_T$, $\mathbf{\Psi}_R$, and $\widehat{\mathbf{T}}$, we consider the $\theta^2 \times \theta^2$ path loss matrix $\widehat{\boldsymbol{\mathcal{G}}}$ of a system with θ^2 transmit and receive antennas regularly spaced in $\mathcal{A}_L^{\#}$. It can be shown [94] that $\widehat{\boldsymbol{\mathcal{G}}}$ is a symmetric block Toeplitz matrix of $\theta \times \theta$ Toeplitz blocks and, asymptotically, for $\theta^2 \to \infty$, it admits an eigenvalue decomposition based on a $\theta^2 \times \theta^2$ 2D discrete Fourier transform (DFT) [2] matrix $\mathbf{F}$. Then, the spectral decomposition of $\widehat{\boldsymbol{\mathcal{G}}}$ is given as [94] [91]

$$\widehat{\boldsymbol{\mathcal{G}}} = \mathbf{F} \widehat{\mathbf{T}} \mathbf{F}^H \tag{2.11}$$

where the matrix $\widehat{\mathbf{T}}$ is a deterministic, asymptotically diagonal matrix whose diagonal elements are the DFT of the first row of $\widehat{\boldsymbol{\mathcal{G}}}$. The random matrices $\mathbf{\Psi}_R$ and $\mathbf{\Psi}_T$ are obtained by extracting independently and uniformly at random N_R and N_T rows of matrix $\mathbf{F}$.

The decomposition in (2.10) can also be regarded as a decomposition into two independent random Vandermonde matrices and one deterministic matrix. The computation of mixed moments of random Vandermonde matrices has been studied extensively in [95, 96] and methods have been provided to compute mixed moments for several classes of such matrices by exploiting combinatorial techniques. The analyzed classes share the common property that the mixed moments can be expressed in terms of the moments of the individual independent matrices as for free random matrices. However, in [95] this interesting property has been questioned for the class of matrices that includes the matrix $\widehat{\mathbf{G}}$ and the decomposition in (2.10) and the approach in [95]

[2] The 1D DFT matrix over N points is the $N \times N$ matrix with element in row i and column j given by $(\mathbf{F}_1)_{ij} = \frac{1}{\sqrt{N}} \mathrm{e}^{-2\pi \mathbf{i}(i-1)(j-1)/N}$. The definition can be extended to 2D and the 2D DFT matrix is given by $\mathbf{F} = \mathbf{F}_1 \otimes \mathbf{F}_1$.

is not applicable here. Therefore, to keep the problem tractable, we approximate the random Vandermonde matrices with Gaussian matrices with i.i.d. entries. With this approximation, our random matrices fall into the class of free matrices and the mixed moments can be expressed in terms of the moments of the individual independent matrices.

Similarly, for 1D systems, the path loss matrix $\widehat{\mathbf{G}}_{1D}$ admits the same decomposition as in (2.10), i.e., $\widehat{\mathbf{G}}_{1D} = \mathbf{\Psi}_{R,1D}\,\widehat{\mathbf{T}}_{1D}\,\mathbf{\Psi}_{T,1D}^{H}$. Accordingly, the matrix $\widehat{\mathbf{T}}_{1D}$ is a $\theta \times \theta$ matrix depending only on the function $\hat{g}_{1D}(r_i, t_j)$, $\mathbf{\Psi}_{R,1D}$ and $\mathbf{\Psi}_{T,1D}$ are $\widetilde{N}_R \times \theta$ and $\widetilde{N}_T \times \theta$ random matrices, respectively. In order to define the matrices $\mathbf{\Psi}_{R,1D}$, $\mathbf{\Psi}_{T,1D}$, and $\widehat{\mathbf{T}}_{1D}$, we consider the $\theta \times \theta$ path loss matrix $\widehat{\boldsymbol{\mathcal{G}}}_{1D}$ of a system with θ transmit and receive antennas regularly spaced in $\widetilde{\mathcal{A}}_L^{\#}$. The matrix $\widehat{\boldsymbol{\mathcal{G}}}_{1D}$ is a Toeplitz matrix and, asymptotically, for $\theta \to \infty$, it admits an eigenvalue decomposition based on a $\theta \times \theta$ 1D DFT matrix $\mathbf{F}_1$. The random matrices $\mathbf{\Psi}_{R,1D}$ and $\mathbf{\Psi}_{T,1D}$ are obtained by extracting independently and uniformly at random $\widetilde{N}_R$ and $\widetilde{N}_T$ rows of matrix $\mathbf{F}_1$.

In the following subsections, we obtain the eigenvalue moments of the channel covariance matrices for the two channel models considered in this work.

2.4.1 Eigenvalue Moments for Antennas in LoS

In this subsection, we derive the eigenvalue moments for DASs with transmitters and receivers in LoS and channel attenuation given by path loss. The matrix $\mathbf{G}$ ($\mathbf{G}_{1D}$) for transmitters and receivers in LoS admits a decomposition similar to $\widehat{\mathbf{G}}$ ($\widehat{\mathbf{G}}_{1D}$), i.e., $\mathbf{G} = \mathbf{\Psi}_R \mathbf{T}\, \mathbf{\Psi}_T^H$ ($\mathbf{G}_{1D} = \mathbf{\Psi}_{R,1D}\,\mathbf{T}_{1D}\,\mathbf{\Psi}_{T,1D}^{H}$). As in [32][93], the eigenvalue moments of the channel covariance matrix are obtained by approximating the random matrices $\mathbf{\Psi}_R$ ($\mathbf{\Psi}_{R,1D}$) and $\mathbf{\Psi}_T$ ($\mathbf{\Psi}_{T,1D}$) by the independent matrices $\mathbf{\Phi}_R$ ($\mathbf{\Phi}_{R,1D}$) and $\mathbf{\Phi}_T$ ($\mathbf{\Phi}_{T,1D}$), respectively, consisting of i.i.d. zero mean complex Gaussian elements with variance θ^{-2} (θ^{-1} for 1D systems) to obtain matrix $\widetilde{\mathbf{G}} = \mathbf{\Phi}_R \mathbf{T}\, \mathbf{\Phi}_T^H$ ($\widetilde{\mathbf{G}}_{1D} = \mathbf{\Phi}_{R,1D}\,\mathbf{T}_{1D}\,\mathbf{\Phi}_{T,1D}^{H}$). This approximation enables the application of classical techniques from RMT and free probability [86][97]. The derivation of the eigenvalue moments follows the techniques proposed in [77][98]. The results are summarized in the following proposition.

Proposition 2.1 *Let $g(\mathbf{r}_i, \mathbf{t}_j)$ be the function of channel coefficients in LoS, $T(f_1, f_2)$ with $(f_1, f_2) \in [-1/2, +1/2]^2$ be the 2D Fourier series of the sequence obtained by sampling $g(\mathbf{r}_i, \mathbf{t}_j)$ over the regular grid $\mathcal{A}_\infty^\#$, and $m_{\mathbf{T}}^{(2\ell)} = \int_{-1/2}^{+1/2}\int_{-1/2}^{+1/2} |T(f_1, f_2)|^{2\ell} \mathrm{d}f_1 \mathrm{d}f_2$. Consider the matrix $\widetilde{\mathbf{C}} = \widetilde{\mathbf{G}}^H \widetilde{\mathbf{G}}$ with $\widetilde{\mathbf{G}} = \mathbf{\Phi}_R \mathbf{T} \mathbf{\Phi}_T^H$. For $\theta^2, N_R, N_T \to +\infty$ with $N_T/\theta^2 \to \beta_T$ and $N_R/\theta^2 \to \beta_R$, $\widetilde{C}_{kk}^{(\ell)}$, the k-th diagonal element of matrix $\widetilde{\mathbf{C}}^\ell$ and $m_{\widetilde{\mathbf{C}}}^{(\ell)}$, the eigenvalue moment of order ℓ of the matrix $\widetilde{\mathbf{C}}$ converge to a deterministic value given by*

$$\widetilde{C}_{kk}^{(\ell)} = m_{\widetilde{\mathbf{C}}}^{(\ell)} = \sum_{n=0}^{\ell-1} \sigma^{(\ell-n)} m_{\widetilde{\mathbf{C}}}^{(n)} \qquad \text{for any } k \text{ and } \ell \geq 2$$

with

$$\sigma^{(\ell)} = \int\int \mathbb{P}^{(\ell)}(|T(f_1, f_2)|^2) \mathrm{d}f_1 \mathrm{d}f_2$$

and $\mathbb{P}^{(\ell)}(|T(f_1, f_2)|^2)$ polynomial in $|T(f_1, f_2)|^2$ recursively given by

$$\begin{aligned}\mathbb{P}^{(\ell)}(|T(f_1, f_2)|^2) = \beta_T m_{\widetilde{\mathbf{C}}}^{(\ell-1)} |T(f_1, f_2)|^2 + \beta_R \beta_T |T(f_1, f_2)|^2 \sum_{s=0}^{\ell-2} m_{\widetilde{\mathbf{C}}}^{(s)} \mathbb{P}^{(\ell-s-1)}(|T(f_1, f_2)|^2) \\ + \beta_T^2 |T(f_1, f_2)|^2 \sum_{s=0}^{\ell-2} \sum_{r=1}^{\ell-2-s} m_{\widetilde{\mathbf{C}}}^{(s)} m_{\widetilde{\mathbf{C}}}^{(r)} \mathbb{P}^{(\ell-s-r-1)}(|T(f_1, f_2)|^2).\end{aligned}$$

The initial values of the recursion are

$$\begin{aligned}\widetilde{C}_{kk}^{(0)} &= m_{\widetilde{\mathbf{C}}}^{(0)} = 1, \\ \mathbb{P}^{(1)}(|T(f_1, f_2)|^2) &= \beta_R |T(f_1, f_2)|^2, \\ \widetilde{C}_{kk}^{(1)} &= m_{\widetilde{\mathbf{C}}}^{(1)} = \sigma^{(1)} = \beta_R m_{\mathbf{T}}^{(2)}.\end{aligned} \tag{2.12}$$

Proposition 2.1 suggests a simple algorithm to determine $m_{\widetilde{\mathbf{C}}}^{(\ell)}$ and $\widetilde{C}_{kk}^{(\ell)}$ illustrated in Algorithm 1. For a detailed proof of the Algorithm 1, see Appendix A.1.

By applying the recursive Algorithm 1, we obtain the first three eigenvalue moments for antennas in LoS as follows

$$\begin{aligned}m_{\widetilde{\mathbf{C}}}^{(1)} &= \beta_R m_{\mathbf{T}}^{(2)}, \\ m_{\widetilde{\mathbf{C}}}^{(2)} &= \beta_R^2 \beta_T m_{\mathbf{T}}^{(4)} + \beta_R(\beta_R + \beta_T)(m_{\mathbf{T}}^{(2)})^2, \\ m_{\widetilde{\mathbf{C}}}^{(3)} &= \beta_R^3 \beta_T^2 m_{\mathbf{T}}^{(6)} + 3\beta_R^2 \beta_T (\beta_R + \beta_T) m_{\mathbf{T}}^{(2)} m_{\mathbf{T}}^{(4)} + \left[\beta_R \beta_T (3\beta_R + \beta_T) + \beta_R^3\right](m_{\mathbf{T}}^{(2)})^3.\end{aligned} \tag{2.13}$$

Algorithm 1 Eigenvalue Moments for Antennas in LoS

Initialization
Let $\mu_0 = \rho_0(x) = 1$, $\sigma^{(1)} = \mu_1 = \beta_R\, m_{\mathbf{T}}^{(2)}$, $\rho_1(x) = \beta_R x$, and $\ell = 2$.

Step ℓ

- Define polynomial $\rho_\ell(x)$ in x

$$\rho_\ell(x) = \beta_T \mu_{\ell-1} x + \beta_R \beta_T x \sum_{s=0}^{\ell-2} \mu_s \rho_{\ell-s-1}(x) + \beta_T^2 x \sum_{s=0}^{\ell-2} \sum_{r=1}^{\ell-2-s} \mu_s \mu_r \rho_{\ell-s-r-1}(x)$$

  and write it as a polynomial in x.
- In $\rho_\ell(x)$, replace the monomial $x, x^2, \ldots x^\ell$ by the moments $m_{\mathbf{T}}^{(2)}, m_{\mathbf{T}}^{(4)}, \ldots m_{\mathbf{T}}^{(2\ell)}$, respectively and assign the result to $\sigma^{(\ell)}$.
- Compute $\mu_\ell = \sum_{n=0}^{\ell-1} \sigma^{(\ell-n)} \mu_n$.
- Assign μ_ℓ to $m_{\widetilde{\mathbf{C}}}^{(\ell)}$ and $\widetilde{C}_{kk}^{(\ell)}$.
- Increase ℓ by a unit.

The Algorithm 1 also holds for 1D systems, as $\theta, \widetilde{N}_R, \widetilde{N}_T \to \infty$ with $\widetilde{N}_T/\theta \to \widetilde{\beta}_T$ and $\widetilde{N}_R/\theta \to \widetilde{\beta}_R$.

2.4.2 Eigenvalue Moments for Rayleigh Fading Channels

In this subsection, we consider the channel matrix for Rayleigh fading given by $\mathbf{G} = (\hat{g}_{ij} h_{ij})_{i=1,\ldots N_R}^{j=1,\ldots N_T}$ ($\mathbf{G}_{1D} = (\hat{g}_{ij,\,1D}\, h_{ij})_{i=1,\ldots N_R}^{j=1,\ldots N_T}$) and we determine an asymptotic approximation of its eigenvalue moments by approximating $\widehat{\mathbf{G}}$ ($\widehat{\mathbf{G}}_{1D}$), the path loss matrix, by $\breve{\mathbf{G}} = (\breve{g}_{ij})_{i=1,\ldots N_R}^{j=1,\ldots N_T} = \mathbf{\Phi}_R \widehat{\mathbf{T}}\, \mathbf{\Phi}_T^H$ $\big(\breve{\mathbf{G}}_{1D} = (\breve{g}_{ij,\,1D})_{i=1,\ldots N_R}^{j=1,\ldots N_T} = \mathbf{\Phi}_{R,1D}\, \widehat{\mathbf{T}}_{1D}\, \mathbf{\Phi}_{T,1D}^H\big)$. Then, the following result holds.

Proposition 2.2 *Let $\hat{g}(\mathbf{r}_i, \mathbf{t}_j)$ be the path loss function, $\widehat{T}(f_1, f_2)$, with $(f_1, f_2) \in [-1/2, 1/2]^2$, be the 2D discrete Fourier series of the sequence obtained by sampling $\hat{g}(\mathbf{r}_i, \mathbf{t}_j)$ over a regularly spaced grid $\mathcal{A}_\infty^{\#}$, and $m_{\widehat{\mathbf{T}}}^{(2\ell)} = \int_{-1/2}^{+1/2} \int_{-1/2}^{+1/2} |\widehat{T}(f_1, f_2)|^{2\ell}\, \mathrm{d}f_1\, \mathrm{d}f_2$. Consider the matrix $\widetilde{\mathbf{G}} = (\breve{g}_{ij} h_{ij})_{i=1,\ldots N_R}^{j=1,\ldots N_T}$. As $L \to +\infty$, the eigenvalue moment of order ℓ of the matrix $\widetilde{\mathbf{C}} = \widetilde{\mathbf{G}}^H \widetilde{\mathbf{G}}$ converges to the deterministic value given by*

$$m_{\widetilde{\mathbf{C}}}^{(\ell)} = (m_{\widehat{\mathbf{T}}}^{(2)})^\ell \sum_{k=0}^{\ell-1} \frac{1}{k+1} \binom{\ell-1}{k} \binom{\ell}{k} \beta_T^k\, \beta_R^{\ell-k} \tag{2.14}$$

Sketch of the proof. The ℓ-th eigenvalue moment of matrix $\widetilde{\mathbf{C}}$ is given by

$$\begin{aligned} m_{\widetilde{\mathbf{C}}}^{(\ell)} &= \mathbb{E}\Big\{\frac{1}{N_T}\mathrm{tr}(\widetilde{\mathbf{C}}^{\ell})\Big\} \\ &= \frac{1}{N_T}\sum_{j_1,\ldots j_\ell=1}^{N_T}\sum_{i_1,\ldots i_\ell=1}^{N_R}\mathbb{E}\big\{\breve{g}_{i_1j_1}^{*}h_{i_1j_1}^{*}\ldots\breve{g}_{i_\ell j_\ell}^{*}h_{i_\ell j_\ell}^{*}\breve{g}_{i_\ell j_1}h_{i_\ell j_1}\big\} \\ &= \frac{1}{N_T}\sum_{j_1,\ldots j_\ell=1}^{N_T}\sum_{i_1,\ldots i_\ell=1}^{N_R}\mathbb{E}\big\{h_{i_1j_1}^{*}\,h_{i_1j_2}\ldots h_{i_\ell j_\ell}^{*}\,h_{i_\ell j_1}\big\}\times\mathbb{E}\big\{\breve{g}_{i_1j_1}^{*}\,\breve{g}_{i_1j_2}\ldots\breve{g}_{i_\ell j_\ell}^{*}\,\breve{g}_{i_\ell j_1}\big\} \end{aligned} \tag{2.15}$$

where the last equality stems from the statistical independence of $\breve{g}_{ij}$ and h_{ij}. It is possible to show that for $L \to +\infty$, non-vanishing contributions to the eigenvalue moments in (2.15) are given by terms with indices $(j_1, i_1, j_2, i_2, \ldots, j_\ell, i_\ell)$ satisfying specific conditions defined in the following. Then, a closed-form expression of the eigenvalue moments is found by resorting to combinatorial techniques widely utilized in RMT to derive eigenvalue moments.

Given the sequence of $(j_1, i_1, j_2, i_2, \ldots, j_\ell, i_\ell)$, we consider the two sets of indices eventually repeated $\mathcal{J} = \{j_1, j_2, \ldots j_\ell\}$ and $\mathcal{I} = \{i_1, i_2, \ldots i_\ell\}$ and denote by p_1 and p_2 the number of distinct indices in $\mathcal{J}$ and $\mathcal{I}$. Additionally, we associate to the sequence a bipartite graph with directed edges from one to the other set given by each consecutive pair of indices in the sequence. The sequences of indices that contribute to non-vanishing terms of the eigenvalue moments satisfy the following conditions

(P_1) $p_1 + p_2 = \ell + 1$;

(P_2) Whenever an edge from $\mathcal{J}$ to $\mathcal{I}$ appears in the bipartite graph, then it should also appear in the opposite direction from $\mathcal{I}$ and $\mathcal{J}$,

(P_3) The bipartite graph is a tree.

Consider the sets obtained from $\mathcal{J}$ and $\mathcal{I}$ by removing repetitions of indices. It is possible to show that the number of distinct bipartite graphs satisfying the conditions above is given by [99]

$$\frac{1}{p_1}\binom{\ell-1}{p_1-1}\binom{\ell}{p_1-1} = \frac{1}{\ell-p_2+1}\binom{\ell-1}{\ell-p_2}\binom{\ell}{p_1-1} \tag{2.16}$$

Additionally, choosing the indices of $\mathcal{J}$ and $\mathcal{I}$ from the sets $\{1,\dots N_R\}$ and $\{1,\dots N_T\}$ respectively, there are

$$N_T(N_T-1)\dots(N_T-p_1-1)N_R(N_R-1)\dots(N_R-p_2-1)=\mathcal{O}(N_T^{p_1}N_R^{p_2}) \tag{2.17}$$

possible choices of indices yielding similar bipartite graphs. Then, in total, there are

$$N_T^{p_1}N_R^{p_2}\frac{1}{p_1}\binom{\ell-1}{p_1-1}\binom{\ell}{p_1-1} \tag{2.18}$$

sequences of indices satisfying the conditions P_1, P_2, and P_3 with p_1 vertices in $\{1,\dots N_T\}$ and $\ell+1-p_1$ vertices in $\{1,\dots N_R\}$.

Let us observe that the expectation of each term satisfying the conditions P_1, P_2, and P_3 with p_1 vertices in $\{1,\dots N_T\}$ and $\ell+1-p_2$ vertices in $\{1,\dots N_R\}$, gives the same contribution. In particular,

$$\mathbb{E}\big\{h^*_{i_1j_1}\,h_{i_1j_2}\dots h^*_{i_\ell j_\ell}\,h_{i_\ell j_1}\big\}=1 \tag{2.19}$$

since in the expectation appear ℓ independent factors $|h_{i,j}|^2$ with $h_{i,j}\sim\mathcal{CN}(0,1)$. Additionally,

$$\mathbb{E}\big\{\breve{g}^*_{i_1j_1}\,\breve{g}_{i_1j_2}\dots\breve{g}^*_{i_\ell j_\ell}\,\breve{g}_{i_\ell j_1}\big\}=\frac{(m^{(2)}_{\widehat{\mathbf{T}}})^\ell}{\theta^{2\ell}} \tag{2.20}$$

since in the expectations appear ℓ independent factors $|\breve{g}_{ij}|^2$ and $\mathbb{E}\{|\breve{g}_{ij}|^2\}=\frac{1}{\theta^2}\,m^{(2)}_{\widehat{\mathbf{T}}}$ where $m^{(2)}_{\widehat{\mathbf{T}}}$ denotes the eigenvalue moment of the diagonal matrix $\widehat{\mathbf{T}}\widehat{\mathbf{T}}^H$. Note that as $\theta^2\to+\infty$, $m^{(2)}_{\widehat{\mathbf{T}}}$ coincides with the moment defined in the statement of Proposition 2.2. From (2.16), (2.17), (2.19), and (2.20) we obtain

$$m^{(\ell)}_{\widetilde{\mathbf{C}}}=\frac{1}{N_T}\sum_{p_1=1}^{\ell}\frac{1}{p_1}\binom{\ell-1}{p_1-1}\binom{\ell}{p_1-1}N_R^{\ell+1-p_1}N_T^{p_1}\left(\frac{m^{(2)}_{\widehat{\mathbf{T}}}}{\theta^2}\right)^\ell$$

that leads to (2.14). This concludes our proof. See Appendix A.2 for more details. ■

The first three eigenvalue moments of the covariance matrix for the Rayleigh fading channel converge to the following values

$$\begin{aligned}
m^{(1)}_{\widetilde{\mathbf{C}}}&=\beta_R m^{(2)}_{\widehat{\mathbf{T}}}=\rho_R\tau^2 m^{(2)}_{\widehat{\mathbf{T}}},\\
m^{(2)}_{\widetilde{\mathbf{C}}}&=\beta_R(\beta_R+\beta_T)(m^{(2)}_{\widehat{\mathbf{T}}})^2=\rho_R\tau^4(\rho_R+\rho_T)(m^{(2)}_{\widehat{\mathbf{T}}})^2,\\
m^{(3)}_{\widetilde{\mathbf{C}}}&=\big[\beta_R\beta_T(3\beta_R+\beta_T)+\beta_R^3\big](m^{(2)}_{\widehat{\mathbf{T}}})^3=\tau^6\big[\rho_R\rho_T(3\rho_R+\rho_T)+\rho_R^3\big](m^{(2)}_{\widehat{\mathbf{T}}})^3.
\end{aligned} \tag{2.21}$$

Similarly, the analysis is carried out for the 1D system as follows.

Proposition 2.3 *Let $\hat{g}_{1D}(r_i, t_j)$ be the path loss function, $\widehat{T}_{1D}(f)$, with $f \in [-1/2, 1/2]$, be the 1D discrete Fourier series of the sequence obtained by sampling $\hat{g}_{1D}(r_i, t_j)$ over a regularly spaced grid $\widetilde{\mathcal{A}}_{\infty}^{\#}$, and $m_{\widehat{\mathbf{T}}_{1D}}^{(2\ell)} = \int_{-1/2}^{+1/2} |\widehat{T}_{1D}(f)|^{2\ell} \, \mathrm{d}f$. Consider the matrix $\widetilde{\mathbf{G}}_{1D} = (\breve{g}_{ij,\,1D}\, h_{ij})_{i=1,\ldots N_R}^{j=1,\ldots N_T}$. As $L \to +\infty$, the eigenvalue moment of order ℓ of the matrix $\widetilde{\mathbf{C}}_{1D} = \widetilde{\mathbf{G}}_{1D}^H \widetilde{\mathbf{G}}_{1D}$ converges to the deterministic value given by*

$$m_{\widetilde{\mathbf{C}}_{1D}}^{(\ell)} = (m_{\widehat{\mathbf{T}}_{1D}}^{(2)})^{\ell} \sum_{k=0}^{\ell-1} \frac{1}{k+1} \binom{\ell-1}{k} \binom{\ell}{k} \widetilde{\beta}_T^k \, \widetilde{\beta}_R^{\ell-k} \tag{2.22}$$

The derivation of (2.22) follows the same lines as the derivation done for 2D systems. The first three eigenvalue moments of the covariance matrix $\widetilde{\mathbf{C}}_{1D}$ converge to

$$\begin{aligned}
m_{\widetilde{\mathbf{C}}_{1D}}^{(1)} &= \widetilde{\beta}_R m_{\widehat{\mathbf{T}}_{1D}}^{(2)} = \widetilde{\rho}_R \tau m_{\widehat{\mathbf{T}}_{1D}}^{(2)}, \\
m_{\widetilde{\mathbf{C}}_{1D}}^{(2)} &= \widetilde{\beta}_R (\widetilde{\beta}_R + \widetilde{\beta}_T)(m_{\widehat{\mathbf{T}}_{1D}}^{(2)})^2 = \widetilde{\rho}_R \tau^2 (\widetilde{\rho}_R + \widetilde{\rho}_T)(m_{\widehat{\mathbf{T}}_{1D}}^{(2)})^2, \\
m_{\widetilde{\mathbf{C}}_{1D}}^{(3)} &= \big[\widetilde{\beta}_R \widetilde{\beta}_T (3\widetilde{\beta}_R + \widetilde{\beta}_T) + \widetilde{\beta}_R^3\big](m_{\widehat{\mathbf{T}}_{1D}}^{(2)})^3 = \tau^3 \big[\widetilde{\rho}_R \widetilde{\rho}_T (3\widetilde{\rho}_R + \widetilde{\rho}_T) + \widetilde{\rho}_R^3\big](m_{\widehat{\mathbf{T}}_{1D}}^{(2)})^3.
\end{aligned} \tag{2.23}$$

2.5 Favorable Propagation

To have favorable propagation, the channel vectors between the BS and the users should be orthogonal. It is said that the channel offers favorable propagation if the following condition holds [35]

$$\mathbf{g}_k^H \mathbf{g}_j = \begin{cases} 0, & k \neq j \\ \|\mathbf{g}_k\|^2 \neq 0, & k = j. \end{cases} \tag{2.24}$$

where $\mathbf{g}_k = [g_{1k}\, g_{2k} \ldots g_{Nk}]^T$ denotes the channel vector from the N antennas to the user k. The favorable propagation condition (2.24) offers the optimal performance with linear processing. However in practice, this condition is not fully satisfied, but can be approximately attained when the number of antennas grows large, in which case the channels are said to provide *asymptotically* favorable propagation. More precisely, the asymptotically favorable propagation condition can be defined as follows [35]

$$\frac{1}{N} \mathbf{g}_k^H \mathbf{g}_j \to 0, \quad N \to \infty, \; k \neq j. \tag{2.25}$$

In the following, we analyze the favorable propagation conditions for LoS and Rayleigh fading channel in CF massive MIMO systems through the characteristics of their channel eigenvalue moments. In a favorable propagation environment, when the users have almost orthogonal channels, the channel covariance matrix $\mathbf{R}$ of size $K \times K$ is almost diagonal and satisfies the following properties

$$\mathrm{MR}(\ell) = \frac{m_{\mathbf{R}}^{(\ell)}}{\frac{1}{K}\mathrm{tr}\big[\big(\mathrm{diag}(\mathbf{R})\big)^{\ell}\big]} \approx 1 \qquad \forall \ell \in \mathbb{N}^{+} \tag{2.26}$$

where $m_{\mathbf{R}}^{(\ell)}$ denotes the ℓ-order eigenvalue moment of matrix $\mathbf{R}$, and MR stands for moment ratio. These properties are asymptotically satisfied for centralized massive MIMO systems, in rich scattering environments, when the number of users stays finite while the number of antennas at the central base station tends to infinity.

By making use of the observation that in large DAS, as $L \to \infty$, $\widetilde{C}_{kk} = \beta_R m_{\mathbf{T}}^{(2)}$, we obtain that $\frac{1}{N_T}\mathrm{tr}\big[\big(\mathrm{diag}(\widetilde{\mathbf{C}})\big)^{\ell}\big] = \beta_R^{\ell}\,(m_{\mathbf{T}}^{(2)})^{\ell}$ such that (2.26) specializes for DAS with LoS channel and $\ell = 2, 3$ as follows

$$\frac{m_{\widetilde{\mathbf{C}}}^{(2)}}{\beta_R^2 (m_{\mathbf{T}}^{(2)})^2} = 1 + \frac{\beta_T}{\beta_R} + \beta_T \frac{m_{\mathbf{T}}^{(4)}}{(m_{\mathbf{T}}^{(2)})^2} \tag{2.27}$$

$$\frac{m_{\widetilde{\mathbf{C}}}^{(3)}}{\beta_R^3 (m_{\mathbf{T}}^{(2)})^3} = 1 + 3\frac{\beta_T}{\beta_R} + \frac{\beta_T^2}{\beta_R^2} + 3\beta_T \left(1 + \frac{\beta_T}{\beta_R}\right) \frac{m_{\mathbf{T}}^{(4)}}{(m_{\mathbf{T}}^{(2)})^2} + \beta_T^2 \frac{m_{\mathbf{T}}^{(6)}}{(m_{\mathbf{T}}^{(2)})^3} \tag{2.28}$$

As $\beta_R \to \infty$ while β_T is kept constant, i.e., for $\beta_T/\beta_R \to 0$ and $\beta_T > 0$, the ratios (2.27) and (2.28) converge to the following limiting values

$$\frac{m_{\widetilde{\mathbf{C}}}^{(2)}}{\beta_R^2 (m_{\mathbf{T}}^{(2)})^2} \to 1 + \beta_T \frac{m_{\mathbf{T}}^{(4)}}{(m_{\mathbf{T}}^{(2)})^2} \tag{2.29}$$

$$\frac{m_{\widetilde{\mathbf{C}}}^{(3)}}{\beta_R^3 (m_{\mathbf{T}}^{(2)})^3} \to 1 + 3\beta_T \frac{m_{\mathbf{T}}^{(4)}}{(m_{\mathbf{T}}^{(2)})^2} + \beta_T^2 \frac{m_{\mathbf{T}}^{(6)}}{(m_{\mathbf{T}}^{(2)})^3} \tag{2.30}$$

and conditions (2.26) are not satisfied.

For DASs with path loss and Rayleigh fading channel, the moment ratios in (2.26) converge to one for all $\ell \geq 1$, as $\beta_T/\beta_R \to 0$ and $\beta_R \to \infty$ as shown in the following

$$\frac{m_{\widetilde{\mathbf{C}}}^{(\ell)}}{\frac{1}{N_T}\mathrm{tr}\left[\left(\mathrm{diag}(\widetilde{\mathbf{C}})\right)^{\ell}\right]} = \frac{\beta_R^{\ell}(m_{\widehat{\mathbf{T}}}^{(2)})^{\ell}\sum_{k=0}^{\ell-1}\frac{1}{k+1}\binom{\ell-1}{k}\binom{\ell}{k}\left(\frac{\beta_T}{\beta_R}\right)^{k}}{\beta_R^{\ell}(m_{\widehat{\mathbf{T}}}^{(2)})^{\ell}}$$

$$=1+\sum_{k=1}^{\ell-1}\frac{1}{k+1}\binom{\ell-1}{k}\binom{\ell}{k}\left(\frac{\beta_T}{\beta_R}\right)^{k}\to 1 \qquad (2.31)$$

Then, conditions (2.26) are satisfied and Rayleigh fading channel offers favorable propagation.

Similarly, in the 1D system, it is straightforward to show that the favorable conditions are not satisfied in the case of antennas in LoS while Rayleigh fading channel offers favorable propagation.

2.6 Linear Multi-Stage Detectors

Systems with favorable propagation can efficiently utilize low complexity matched filters at the central processing unit since this filter achieves almost optimal performance in such environments. However, when conditions (2.26) are not satisfied, even linear multi-user detectors are expected to provide substantial gains compared to the matched filter. In the following, we consider low complexity multi-stage detectors including both polynomial expansion detectors, e.g., [78], and multi-stage Wiener filters [79] and we analyze their performance in terms of their signal to interference and noise ratio (SINR) by applying the unified framework proposed in [77][100]. In [77], it is shown that both design and analysis of multi-stage detectors with M stages can be described by a matrix $\mathbf{S}(X)$ defined as

$$\mathbf{S}(X)=\begin{bmatrix} X^{(2)}+\sigma^2X^{(1)} & \cdots & X^{(M+1)}+\sigma^2X^{(M)} \\ X^{(3)}+\sigma^2X^{(2)} & \cdots & X^{(M+2)}+\sigma^2X^{(M+1)} \\ \vdots & \ddots & \vdots \\ X^{(M+1)}+\sigma^2X^{(M)} & \cdots & X^{(2M)}+\sigma^2X^{(2M-1)} \end{bmatrix}$$

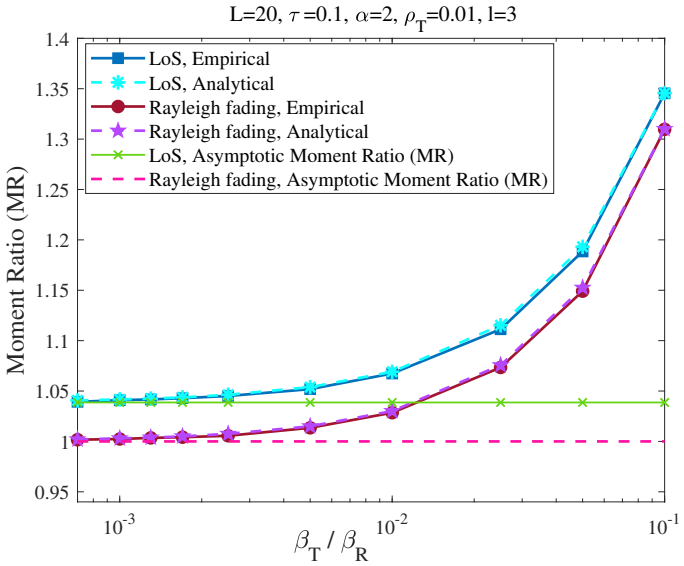

Figure 2.3: Favorable propagation conditions MR$= m_{\widetilde{\mathbf{C}}}^{(\ell)}/\mathrm{tr}[\big(\mathrm{diag}(\widetilde{\mathbf{C}})\big)^{\ell}]$ versus β_T/β_R.

and a vector $\mathbf{s}(X) = [X^{(1)}, X^{(2)}, \ldots X^{(M)}]^T$ where $X = m_{\widetilde{\mathbf{C}}}$ for polynomial expansion detectors and $X = \widetilde{C}_{kk}$ for multi-stage Wiener filters. From the asymptotic property that $\widetilde{C}_{kk}^{(\ell)} = m_{\widetilde{\mathbf{C}}}^{(\ell)}$ for any k and ℓ, we can conclude that multi-stage Wiener filters and polynomial expansion detectors are equivalent in DAS. Additionally, the performance of a centralized processor implementing an M-stage detector is given by [77]

$$\mathrm{SINR}_M = \frac{\mathbf{s}^T(m_{\widetilde{\mathbf{C}}})\mathbf{S}^{-1}(m_{\widetilde{\mathbf{C}}})\mathbf{s}(m_{\widetilde{\mathbf{C}}})}{1 - \mathbf{s}^T(m_{\widetilde{\mathbf{C}}})\mathbf{S}^{-1}(m_{\widetilde{\mathbf{C}}})\mathbf{s}(m_{\widetilde{\mathbf{C}}})} \tag{2.32}$$

For $M = 1$, a multi-stage detector reduces to a matched filter and (2.32) can be applied to determine its performance and SINR_1 yields the SINR at the output of a matched filter.

2.7 Simulation Results

In this section, we validate the analytical asymptotic results by simulations and analyze the performance of multi-stage detectors in large-scale systems. Throughout this section, we consider the following scenario. The channel is characterized by $\alpha = 2$ and reference distance $d_0 = 1$. The transmit antennas are distributed according to a homogeneous PP with intensity $\rho_T = 0.01$ over a finite network of area $A = L^2 = 400$ while

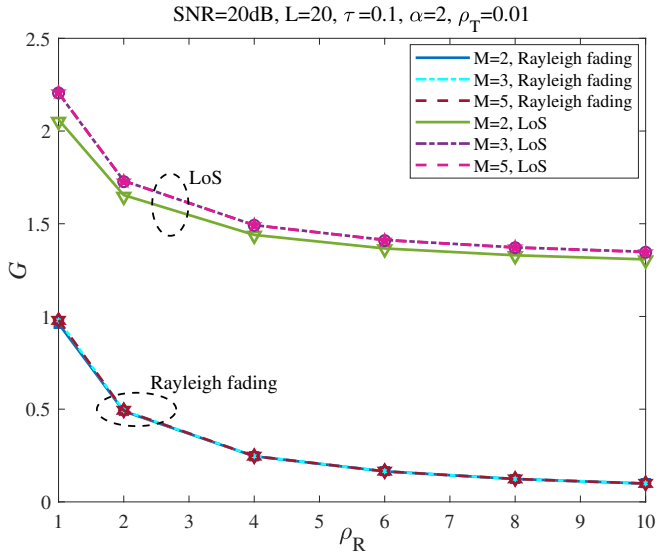

Figure 2.4: Asymptotic (solid lines) and empirical (markers) gains G versus ρ_R for multi-stage detectors ($M = 2, 3, 5$), with path loss plus LoS or plus Rayleigh fading.

the intensity of receivers varies in the range $\rho_R = [0.1, 15]$ in Fig. 2.3 and $\rho_R = [1, 10]$ in Fig. 2.4 and Fig. 2.5. In Fig. 2.6, $\rho_T = 0.5$ and $\rho = [1, 5]$.

Fig. 2.3 shows the moment ratio $m_{\widetilde{\mathbf{C}}}^{(\ell)}/(\mathrm{tr}[\big(\mathrm{diag}(\widetilde{\mathbf{C}})\big)^{\ell}]/N_T)$ versus $\beta_T/\beta_R = \rho_T/\rho_R$ for $\ell = 3$. The x-axis is plotted in logarithmic scale. The analytical moment ratios match almost perfectly the ratios for the simulated finite systems. As predicted analytically, the favorable propagation conditions are not satisfied for LoS channel while they hold in the case of the Rayleigh fading. For small ratios β_T/β_R, the curves of LoS and Rayleigh fading converge to the asymptotic moment ratios in (2.30) and (2.31), respectively. In Fig. 2.4 and Fig. 2.5, we consider a system with average signal to noise ratio (SNR) at the transmitters equal to 20dB and analyze the gain of a multi-stage Wiener filter or equivalently a polynomial expansion detector over a matched filter in terms of the its normalized increase in SINR defined as follows

$$G = \frac{\mathrm{SINR}_M - \mathrm{SINR}_1}{\mathrm{SINR}_1}.$$

Fig. 2.4 compares the performance of the two channel models and presents gain G versus ρ_R, the intensity of receivers for M-stage Wiener filters with $M = 2, 3, 5$. The analytical results in solid lines are obtained under the asymptotic assumption $L \to \infty$.

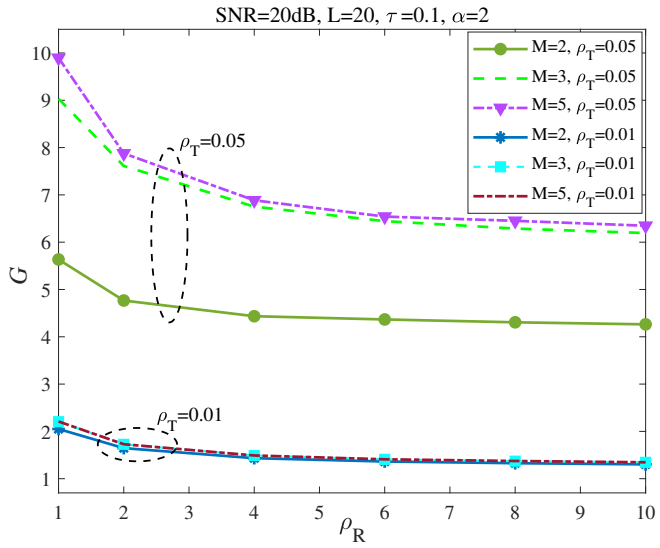

Figure 2.5: Gain G versus ρ_R of multi-stage detectors ($M = 2, 3, 5$) for path loss plus LoS channels and $\rho_T \in \{0.01, 0.05\}$.

The empirical results shown by markers are obtained for $L = 20$ and averaging over 100 network realizations. Simulations show an excellent match between the asymptotic performance and empirical results. In the case of Rayleigh fading, as favorable propagation conditions are satisfied, the performance gap between matched filter and multi-stage detectors tends to vanish and gain G becomes negligible as ρ_R increases while ρ_T is kept constant. Then, for ρ_R sufficiently large, the matched filter achieves almost optimal performance. On the contrary, in the LoS case, the performance gap between the matched filter and multi-stage detectors is dramatic with an increase in SINR of about 140% even for systems with 1000 receive antennas per transmitter per unit area. It is interesting to note that for the considered channel models, this dramatic performance enhancement can be attained already with a very simple 2-stage detector and higher complexity multi-stage detectors offer only incremental improvements at least at very low system loads.

In Fig. 2.5, we analyze the effect of ρ_T/ρ_R, the system load per unit area, in the case of transmit and receive antennas in LoS. Fig. 2.5 shows gain G for $\rho_T = 0.05$ and $\rho_T = 0.01$ as the intensity of receivers varies. Increasing the system load, the SINR increase offered by a 2-stage detector increases enormously and for higher load, also

the increments offered by higher order multi-stage detectors over a 2-stage detector become significant. More specifically, Fig. 2.6 illustrates the SINR [dB] of matched filters ($M = 1$) and multi-stage detectors with two and three stages versus the intensity of receive antennas ρ_R, for LoS channel. For increasing values of ρ_R the performance gap between the matched filter and the multi-stage detector is substantial and does not tend to vanish.

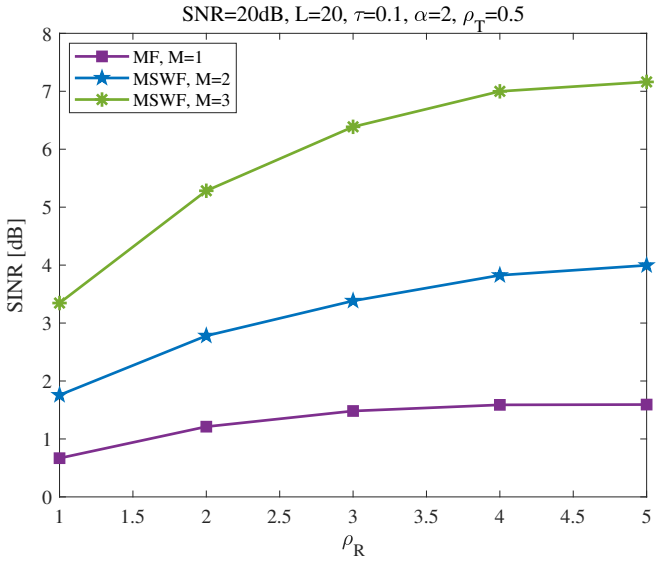

Figure 2.6: SINR [dB] versus ρ_R for matched filter ($M = 1$) and multi-stage detectors ($M = 2, 3$) with path loss plus LoS.

2.8 Conclusion

In this chapter, we considered a CF massive MIMO system in uplink, comprising a massive number of transmit and receive antennas distributed according to homogeneous PP. We analyzed the properties of the system in asymptotic conditions when the network dimensions go to infinity with given intensities of the transmit and receive antenna PPs. We studied the analytical conditions of favorable propagation in CF massive MIMO systems with two kinds of channels, namely, channels with path loss and transmit and receive antennas in LoS or in multipath Rayleigh fading. We showed that the analytical conditions of favorable propagation are satisfied for channels impaired by path loss and Rayleigh fading while they do not hold in the case of LoS

channels, motivating the use and analysis of low complexity linear multi-user detection. We analyzed the performance of polynomial expansion detectors and MSWFs and showed their equivalence in CF massive MIMO systems. Their performance analysis confirmed the expectations of the favorable propagation analysis and the substantial benefits of these detectors compared to MFs when the favorable propagation conditions are not satisfied. Simulation results of the favorable propagation conditions and the performance of multi-stage detectors for finite systems validated the asymptotic analytical results.

Chapter 3

Semi-Blind Pilot Decontamination

3.1 Introduction

The performance of CF massive MIMO systems is critically affected by the so-called *pilot contamination*. This impairment degrading the channel estimation performance, originates from the reuse of training sequences or pilots utilized in channel estimation. Pilot contamination prevents obtaining an adequate estimate of CSI.

CF massive MIMO channels are inherently sparse due to the distribution of APs over a large area and the natural path loss of wireless channels. In this chapter, we address the problem of pilot contamination in CF massive MIMO systems. we develop semi-blind techniques for joint channel estimation and data detection exploiting the sparsity of the channel support due to the strong attenuation with path loss to combat pilot contamination. We analyze the potential of semi-blind approaches with classical signal processing techniques such as FIM, CRB, and identifiability. We provide sets of sufficient and necessary conditions under which channels and data are identifiable. We define a graph that has APs and users as factor and variable nodes and propose an MP algorithm over this graph which computes the channel coefficients if the identifiability conditions are satisfied.

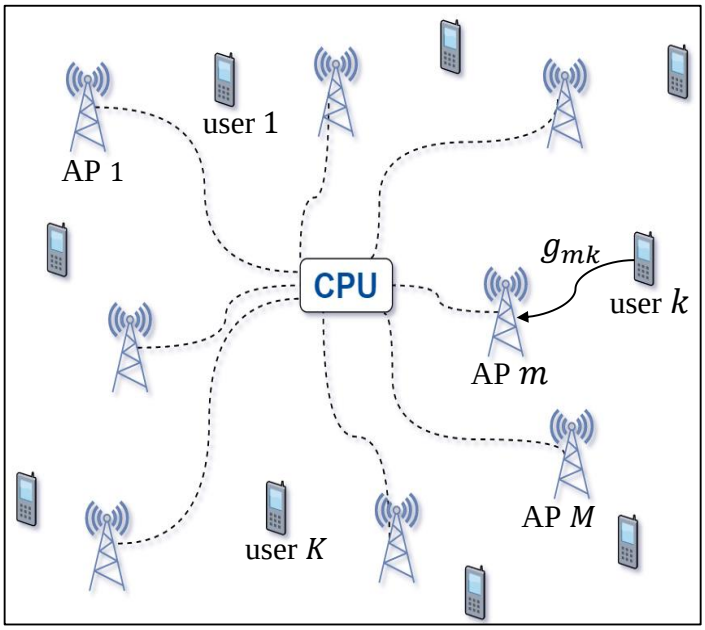

Figure 3.1: A cell-free massive MIMO system consisting of K users and M APs.

3.2 System Model

We consider the uplink of a CF massive MIMO system consisting of K users and M APs equipped with a single antenna and randomly distributed over a $D \times D$ square area. We assume that $M \geq K$. The M APs are connected to a central processing unit (CPU) via a back-haul network, as illustrated in Fig. 3.1. The channel matrix between the APs and users is given by $\mathbf{G} \in \mathbb{C}^{M \times K}$, whose (m, k)-element g_{mk} is the channel coefficient between AP m and user k and is modeled as follows

$$g_{mk} = \sqrt{\beta_{mk}}\, h_{mk}, \tag{3.1}$$

where β_{mk} represents the large-scale fading coefficient which accounts for path loss and shadowing effects and h_{mk} represents the small-scale fading. We assume that h_{mk}, $m = 1, \cdots M$, $k = 1, \cdots K$, are i.i.d. complex normal random variables, i.e., $h_{mk} \sim \mathcal{CN}(0, 1)$. Additionally, we assume perfect knowledge of the large-scale fading coefficients β_{mk}, $m = 1, \cdots M$, $k = 1, \cdots K$ at the CPU.

Due to the path loss, the channel coefficients are assumed to be negligible at distances higher than a given threshold γ. Then, for each AP m, the CPU is required to estimate only the channels of the users in a disc centered around AP m with radius

γ while the signals transmitted from users external to the disc are treated as additive noise. We denote by $\mathcal{K}_I(m)$ and $\mathcal{K}_0(m)$ the sets of users inside the disc centered around AP m and remaining users, respectively. At a global level, this determines a partition of the channel coefficients into two groups, the channel coefficients that have to be detected $\mathcal{K}_I$ and the complement set $\mathcal{K}_0$ given as follows

$$\mathcal{K}_I \equiv \{g_{mk} |\ m = 1, \dots M,\ k \in \mathcal{K}_I(m)\}$$

$$\mathcal{K}_0 \equiv \{g_{mk} |\ m = 1, \dots M,\ k \in \mathcal{K}_0(m)\}$$

Consistently with this partition, we decompose the channel matrix $\mathbf{G}$ into two matrices $\mathbf{G}_I$ and $\mathbf{G}_0$ such that $\mathbf{G} = \mathbf{G}_I + \mathbf{G}_0$. Then, $\mathbf{G}_I$ and $\mathbf{G}_0$ of size $M \times K$ denote the matrices of the relevant and negligible channel coefficients, respectively. Throughout this chapter, we assume that $\gamma \ll D$ and the APs are distributed over the whole region such that matrix $\mathbf{G}_I$ has a large number of zero elements.

In the uplink transmission, each user sends one of P pilot sequences known by the CPU followed by $L - P$ unknown data symbols. The pilot sequences are assumed to be ortho-normal, i.e., orthogonal with unit norm. The L received symbols at the M APs are given by

$$\mathbf{Y} = \sqrt{\rho}\, \mathbf{G}_I\, \mathbf{X} + \sqrt{\rho}\, \mathbf{G}_0\, \mathbf{X} + \mathbf{W}, \tag{3.2}$$

where ρ denotes the transmit power at each user terminal normalized by the noise variance. $\mathbf{Y} \in \mathbb{C}^{M \times L}$ is a matrix of the L received symbols at the M APs and $\mathbf{X} \in \mathbb{C}^{K \times L}$ is a matrix of the transmitted symbols. Note that the k-th row corresponds to the signals transmitted by user k. The matrix $\mathbf{W} \in \mathbb{C}^{M \times L}$ is the AWGN with i.i.d. components having zero mean and unit variance.

Let $\mathbf{X}_p \in \mathbb{C}^{K \times P}$ and $\mathbf{X}_d \in \mathbb{C}^{K \times (L-P)}$ denote the pilot sequences and data symbols, respectively. Then, $\mathbf{X} = [\mathbf{X}_p\ \mathbf{X}_d]$. Similarly, $\mathbf{Y} = [\mathbf{Y}_p\ \mathbf{Y}_d]$ where $\mathbf{Y}_p \in \mathbb{C}^{M \times P}$ and $\mathbf{Y}_d \in \mathbb{C}^{M \times (L-P)}$ represent the matrices of received training and data signals, respectively.

3.3 Cramer-Rao Bound Analysis

The performance of different semi-blind channel estimation algorithms has been evaluated and compared to a certain lower bound. One of these famous lower bounds extensively used as a benchmark for assessing the parameters estimation performance in the literature is CRB. Depending on how the symbols and the channel are treated, different versions of CRB have been derived. There are two possible cases on how to treat the symbols and/or the channel namely, deterministic unknowns or random [101]. Basically, the CRB is obtained as the inverse of the FIM [102].

In this section, we derive the CRB as a performance benchmark to analyze the performance of the semi-blind channel estimation. We obtain the CRB under the assumption of deterministic a priori knowledge on the unknown data symbols. In the deterministic approach both the data signal $\mathbf{X}_d$ and the channel $\mathbf{G}_I$ are modeled as unknown deterministic quantities.

If we denote by $\boldsymbol{\theta}$ the (complex) unknown parameter vector to be estimated, then it is given by

$$\boldsymbol{\theta} = [\mathbf{g}_I^H \;\; \mathrm{vec}^H(\mathbf{X}_d)]^H \tag{3.3}$$

where $\mathbf{g}_I$ is a vector deduced from the non-zero elements of the matrix $\mathbf{G}_I$, whose support is known. In the deterministic CRB, under the deterministic assumption for the data signal $\mathbf{X}_d$ and channel coefficients, we have

$$\mathbf{y} \sim \mathcal{CN}\big(m_{\mathbf{y}}(\boldsymbol{\theta}), \mathbf{C}_{\mathbf{yy}}\big) \tag{3.4}$$

where $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$. The mean $m_{\mathbf{y}}(\boldsymbol{\theta}) = \sqrt{\rho}\,\mathrm{vec}(\mathbf{G}_I\mathbf{X})$ and the covariance matrix $\mathbf{C}_{\mathbf{yy}} = \mathbf{I}_L \otimes \mathbf{C}_{\mathbf{YY}}$ with $\mathbf{C}_{\mathbf{YY}} = \mathbf{I}_M + \rho\,\mathbf{C}_0$ where the covariance matrix $\mathbf{C}_0$ is given by

$$\mathbf{C}_0 = \mathbb{E}\Big\{\mathbf{G}_0\mathbf{G}_0^H\Big\} = \mathrm{diag}\Big(\sum_{k\in\mathcal{K}_0}\beta_{1k}, \ldots, \sum_{k\in\mathcal{K}_0}\beta_{Mk}\Big) \tag{3.5}$$

The probability density function (pdf) of the observations $\mathbf{Y}$ in the parameter $\boldsymbol{\theta}$ is given by

$$f(\mathbf{Y}|\boldsymbol{\theta}) = \frac{1}{\pi^{ML}(\det(\mathbf{C}_{\mathbf{YY}}))^L} \exp\big(- \mathrm{tr}\big\{(\mathbf{Y} - \sqrt{\rho}\,\mathbf{G}_I\mathbf{X})^H \mathbf{C}_{\mathbf{YY}}^{-1}(\mathbf{Y} - \sqrt{\rho}\,\mathbf{G}_I\mathbf{X})\big\}\big)$$
$$= \frac{1}{\pi^{ML}(\det(\mathbf{C}_{\mathbf{YY}}))^L} \exp\Big(- \|\mathbf{C}_{\mathbf{yy}}^{-1/2}(\mathbf{y} - m_{\mathbf{y}})\|^2\Big), \tag{3.6}$$

Computing the Jacobian of $m_{\mathbf{y}}(\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$, the deterministic complex FIM denoted as $\mathcal{J}^d_{\boldsymbol{\theta},\boldsymbol{\theta}}$ on the basis of the data $\mathbf{Y}$ is given by

$$\mathcal{J}^d_{\boldsymbol{\theta},\boldsymbol{\theta}} = \Big(\frac{\partial m_{\mathbf{y}}^H}{\partial \boldsymbol{\theta}^*}\Big)\mathbf{C}_{\mathbf{yy}}^{-1}\Big(\frac{\partial m_{\mathbf{y}}^H}{\partial \boldsymbol{\theta}^*}\Big)^H$$
$$= \rho \begin{bmatrix}\mathbf{Q}' & \mathbf{R}'\end{bmatrix}^H \begin{bmatrix}\mathbf{Q}' & \mathbf{R}'\end{bmatrix} \tag{3.7}$$

where $\mathbf{Q}' = \mathbf{C}_{\mathbf{yy}}^{-1/2}\mathbf{Q}$, $\mathbf{Q} = \dfrac{1}{\sqrt{\rho}}\dfrac{\partial m_{\mathbf{y}}}{\partial \mathbf{g}_I^T}$ and $\mathbf{R}' = \mathbf{C}_{\mathbf{yy}}^{-1/2}\mathbf{R}$, $\mathbf{R} = \dfrac{1}{\sqrt{\rho}}\dfrac{\partial m_{\mathbf{y}}}{\partial \mathrm{vec}^T(\mathbf{X}_d)}$. Note that

$$\begin{aligned}\frac{1}{\sqrt{\rho}} m_{\mathbf{y}} &= \mathrm{vec}(\mathbf{G}_I\mathbf{X}) \\ &= \mathbf{Q}\,\mathbf{g}_I = \mathrm{vec}\big(\mathbf{G}_I \ [\mathbf{X}_p \ \mathbf{0}_{K\times(L-P)}]\big) + \mathbf{R}\,\mathrm{vec}(\mathbf{X}_d)\end{aligned} \tag{3.8}$$

where $\mathbf{0}_{m\times n}$ is an $m \times n$ matrix in which every entry is zero and the matrix $\mathbf{R}$ of size $ML \times K(L-P)$ is given by

$$\mathbf{R} = \begin{bmatrix}\mathbf{0}_{PM\times K(L-P)} \\ \\ \mathbf{I}_{L-P} \otimes \mathbf{G}_I\end{bmatrix} \tag{3.9}$$

and the matrix $\mathbf{Q}$ of size $ML \times n_{\mathbf{g}_I}$, where $n_{\mathbf{g}_I}$ denotes the dimension of column vector $\mathbf{g}_I$ or equivalently the number of non-zero elements of the matrix $\mathbf{G}_I$, is given by

$$\mathbf{Q} = [\mathbf{X}_{1,:}^T \otimes \mathbf{Q}_1 \ \ldots \ \mathbf{X}_{K,:}^T \otimes \mathbf{Q}_K] \tag{3.10}$$

where $\mathbf{X}_{k,:}$ denotes the k-th row of the matrix $\mathbf{X}$ and $\mathbf{Q}_k$ is a selection matrix that contains a subset of columns of identity matrix $\mathbf{I}_M$ corresponding to the consecutive positions of non-zero coefficients in the k-th column of the matrix $\mathbf{G}_I$.

The FIM $\mathcal{J}^d_{\boldsymbol{\theta},\boldsymbol{\theta}}$ is a 2×2 block matrix. The deterministic CRB^d is obtained as the inverse of the Fisher information matrix

$$\mathrm{CRB}^d = \big(\mathcal{J}^d_{\boldsymbol{\theta},\boldsymbol{\theta}}\big)^{-1}. \tag{3.11}$$

The blocks $(1,1)$ and $(2,2)$ of the CRB^d in (3.11) relative to the estimation of the channel coefficients $\mathbf{g}_I$ and data symbols $\mathrm{vec}(\mathbf{X}_d)$, respectively are given as follows

$$\mathrm{CRB}^d_{\mathbf{g}_I} = \frac{1}{\rho}\left(\mathbf{Q}'^H P^{\perp}_{\mathbf{R}'}\,\mathbf{Q}'\right)^{-1} \tag{3.12}$$

$$\mathrm{CRB}^d_{\mathrm{vec}(\mathbf{X}_d)} = \frac{1}{\rho}\left(\mathbf{R}'^H P^{\perp}_{\mathbf{Q}'}\,\mathbf{R}'\right)^{-1} \tag{3.13}$$

where $P_{\mathbf{A}} = \mathbf{A}\big(\mathbf{A}^H\mathbf{A}\big)^{-1}\mathbf{A}^H$ and $P^{\perp}_{\mathbf{A}} = \mathbf{I} - P_{\mathbf{A}}$ denote the projection matrices on the column space of matrix $\mathbf{A}$ and its orthogonal complement, respectively. The $\mathrm{CRB}^d_{\mathbf{g}_I}$ and $\mathrm{CRB}^d_{\mathrm{vec}(\mathbf{X}_d)}$ are derived in Appendix B.1. In the deterministic identifiability analysis that follows, we shall ignore $\mathbf{C}_0$ ($\mathbf{C}_0 = \mathbf{0}$) and hence $\mathbf{C}_{\mathbf{YY}} = \mathbf{I}_M$ and $\mathbf{C}_{\mathbf{yy}} = \mathbf{I}_{ML}$.

3.4 Identifiability

In this section, we derive sets of both sufficient and necessary conditions for the identifiability of vector parameter $\boldsymbol{\theta}$ under the assumption that $\boldsymbol{\theta}$ is a deterministic unknown parameter. Then, we propose an MP algorithm over a graph that determines the exact channel coefficients if the sufficient identifiability conditions are satisfied. Finally, we show that the system is identifiable via semi-blind algorithms if the Karp-Sipser algorithm applied to the same graph yields an empty core paving the way to an analysis of asymptotically large networks based on core percolation properties.

Definition 3.1 (Identifiability [62][103]) *Let ϑ be the parameter to be estimated and $\mathbf{Y}$ the observations. In the regular cases (i.e. in the non-blind cases), ϑ is called identifiable if*

$$\forall \mathbf{Y}, \quad f(\mathbf{Y}|\vartheta) = f(\mathbf{Y}|\vartheta') \;\Rightarrow\; \vartheta = \vartheta' \tag{3.14}$$

For both deterministic and Gaussian models, $f(\mathbf{Y}|\vartheta)$ is a Gaussian distribution, identifiability in this case means identifiability from the mean and the covariance of $\mathbf{Y}$.

In the framework of deterministic identifiability, we assume that vector parameter $\boldsymbol{\theta}$ is deterministic and consider channel $\mathbf{G}_0$ negligible. Then, the observation $\mathbf{y}$ is Gaussian distributed, i.e., $\mathbf{y} \sim \mathcal{CN}(m_{\mathbf{y}}(\boldsymbol{\theta}), \mathbf{I}_{ML})$ with covariance matrix independent of $\boldsymbol{\theta}$. The identifiability of $\boldsymbol{\theta}$ relies only on the known mean $m_{\mathbf{y}}(\boldsymbol{\theta})$ and, for semi-blind

methods, $\mathbf{X}_d$ and $\mathbf{g}_I$ are said to be *identifiable* if [62]

$$\mathbf{G}_I\mathbf{X} = \mathbf{G}_I^{'}\mathbf{X}^{'} \;\Rightarrow\; \mathbf{g}_I = \mathbf{g}_I^{'} \text{ and } \mathbf{X}_d = \mathbf{X}_d^{'} \tag{3.15}$$

Let $m_{\mathbf{Y}^{\sim 0}}$ be the expectation of $\mathbf{Y}$ in (3.2) obtained assuming $\mathbf{G}_0$ negligible. The identifiability problem reduces to analyze the following bi-linear system of equations in the unknowns $\mathbf{g}_I$ and $\mathbf{X}_d$

$$m_{\mathbf{Y}^{\sim 0}} = \sqrt{\rho}\,\mathbf{G}_I\mathbf{X}$$

and determine under which conditions this system admits a unique solution, which is assumed to exist. These identifiability conditions are summarized in the following propositions.

Proposition 3.1 ***Sufficient Identifiability Conditions*** – *Let $\mathcal{S}_k$ denote the support of the channel of user k, i.e., the set of all the indices m such that $\mathbf{G}_{I,m,k} \neq 0$, and let $|\mathcal{S}_k|$ be its cardinality. In a semi-blind joint data detection and channel estimation method, the unknown parameters $\mathbf{g}_I$ and $\mathbf{X}_d$ are identifiable if*

i) the $K \times L$ matrix $\mathbf{X}$, with $L \geq K$ has full row rank K,

ii) the channel of each user is sparse and $|\mathcal{S}_k| \leq M - K + 1$, and

iii) for each group of users $\mathcal{G}_p$ utilizing the same ortho-normal pilot sequence $\mathbf{x}_p^{(p)}$, it is possible to identify a sequence $\{\mathcal{G}_{p,1}, \mathcal{G}_{p,2}, \ldots \mathcal{G}_{p,s}\}$ satisfying the following properties:

1) $\bigcup_{j=1}^{s} \mathcal{G}_{p,j} \equiv \mathcal{G}_p$, i.e., the sequence of subsets is a partition of $\mathcal{G}_p$.

2) In the support of the channel of each user $k \in \mathcal{G}_{p,i}$, there exists at least an index $j \in \mathcal{S}_k$ that is not contained in any of the channel supports of other users in the same group $\mathcal{G}_{p,i}$ or in the following groups of the sequence $\mathcal{G}_{p,i+1}, \ldots \mathcal{G}_{p,s}$.

Remark 3.1 *Condition iii-2 implies that the signal transmitted by each user k in $\mathcal{G}_{p,i}$ impinges an AP in the disc $\mathcal{M}_k$ centered around user k with radius γ and no other signal transmitted by other users in $\mathcal{G}_{p,i}$ or subsequent subsets $\mathcal{G}_{p,i+1}, \mathcal{G}_{p,i+2}, \ldots \mathcal{G}_{p,s}$ impinges the same AP.*

Remark 3.2 *The assumption that $\mathbf{X}$ has full row rank K implies that $\mathbf{X}_d$ has at least*

rank $K - P$.

Proof. Observe that since in CF massive MIMO systems $M \gg K$, and the channel matrix $\mathbf{G}_I$ consists of independent channels, we can assume that it has full row rank equal to K with probability 1. Thanks to the assumptions of Proposition 3.1, also matrix $\mathbf{X}$ has full row rank equal to K as well as matrix $m_{\mathbf{Y}^{\sim 0}}$. Then, the singular value decomposition (SVD) of the noise-free system is given by

$$\frac{1}{\sqrt{\rho}}\, m_{\mathbf{Y}^{\sim 0}} = \mathbf{G}_I \mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H \tag{3.16}$$

where $\mathbf{U} \in \mathbb{C}^{M \times K}$ and $\mathbf{V} \in \mathbb{C}^{L \times K}$ are the matrices of the left and right singular-vectors and $\mathbf{\Sigma}$ is the $K \times K$ diagonal matrix of singular values. Additionally, the left and right singular value matrices $\mathbf{U}$ and $\mathbf{V}$ span the channel subspace $\mathbf{G}_I$ and the signal space $\mathbf{X}$, respectively. Then, the problem of identifiability reduces to determine a $K \times K$ non-singular matrix $\mathbf{T}$ such that $\mathbf{G}_I = \mathbf{U}\mathbf{T}$ and then, also matrix $\mathbf{X}$ is unequivocally given by $\mathbf{X} = \mathbf{T}^{-1}\mathbf{\Sigma}\mathbf{V}^H$. In order to determine matrix $\mathbf{T}$, we utilize the following properties and information:

- The support of each user channel is known and sparse and at least $K-1$ channel coefficients are zero.
- The *contaminated channel.* More specifically, let us consider the linear system of equations corresponding to the transmission of the pilot sequences, i.e., $\frac{1}{\sqrt{\rho}} m_{\mathbf{Y}_p^{\sim 0}} = \mathbf{G}_I \mathbf{X}_p$, where $m_{\mathbf{Y}_p^{\sim 0}}$ denotes the expectation of $\mathbf{Y}_p = \sqrt{\rho}\, \mathbf{G}_I \mathbf{X}_p$. By post-multiplying both sides of the system by the pilot sequence $\mathbf{x}_p^{(p)}$ and exploiting the ortho-normality of the training sequences, $\mathbf{X}_p\, \mathbf{x}_p^{(p)} = \mathbf{1}_{\mathcal{G}_p}$ where $\mathbf{1}_{\mathcal{G}_p}$ is the K-dimensional vector with elements with indices in $\mathcal{G}_p$, i.e., indices corresponding to users transmitting pilot $\mathbf{x}_p^{(p)}$, equal to one and and zero elsewhere. Then, it is apparent that this system of equations enables to determine exactly at each AP the sum of all the non-zero channel coefficients of the users in each group $\mathcal{G}_p$, $p = 1, \dots P$, i.e., $\frac{1}{\sqrt{\rho}}\, m_{\mathbf{Y}_p^{\sim 0}}\, \mathbf{x}_p^{(p)} = \mathbf{G}_I\, \mathbf{1}_{\mathcal{G}_p}$.

Then, let us focus on a user k in $\mathcal{G}_{p,1}$. Thanks to the assumptions on the partition of $\mathcal{G}_p$,

there exists at least an AP m such that $\mathbf{G}_{I,m,:}\,\mathbf{1}_{\mathcal{G}_p} = g_{m,k} = \dfrac{1}{\sqrt{\rho}}\, m_{\mathbf{Y}_p^{\sim 0}}\, \mathbf{x}_p^{(p)} \neq 0$, where $\mathbf{G}_{I,m,:}$ denotes the m-th row of the matrix $\mathbf{G}_I$. Furthermore, thanks to the assumption on the sparsity of the channels, we can obtain $K-1$ equations from the system of equations $\mathbf{G}_{I,:,k} = \mathbf{U}\mathbf{T}_{:,k}$ where the channel of user k is zero. Then, we can construct a non-homogeneous system of equations in the unknown $\mathbf{T}_{:,k}$ and the vector of constant terms consisting of zeros and at least the non-zero element h_{mk}. This system can be unequivocally solved to determine $\mathbf{T}_{:,k}$.

Thanks to the properties of the sequence $\mathcal{G}_{p,1}, \mathcal{G}_{p,2}, \ldots \mathcal{G}_{p,s}$, it is possible to determine sequentially, the columns of matrix $\mathbf{T}$ corresponding to a certain group, compute exactly the corresponding channels of the users in the group and cancel them from the contaminated channel for group $\mathcal{G}_p$ until the complete computation of all the columns of matrix $\mathbf{T}$ corresponding to all the users in $\mathcal{G}_p$ and the corresponding channels. This approach can be repeated for all the groups up to the complete computation of matrix $\mathbf{T}$ and channel $\mathbf{G}_I$. Then, we observe that $\mathbf{T}$ has full rank K since $\mathbf{G}_I$ has full row rank. The inverse of $\mathbf{T}$ exists and enables the computation of $\mathbf{X}_d$. This concludes the proof. ■

In the following, let $(\mathbf{G})_{\mathcal{G}_p}$ denote a reduced version of the matrix $\mathbf{G}$ containing only the columns corresponding to the users in $\mathcal{G}_p$.

Proposition 3.2 ***Necessary Identifiability Conditions*** – *Identification of* $\mathbf{g}_I$ *and* $\mathbf{X}_d$ *from the product* $\mathbf{G}_I\,\mathbf{X}$ *leads to the global necessary identifiability condition*

$$\frac{1}{K}\sum_{k=1}^{K} |\mathcal{S}_k| \;\leq\; M - K + P \tag{3.17}$$

or the per pilot necessary identifiability condition

$$\frac{1}{|\mathcal{G}_p|}\sum_{k\in\mathcal{G}_p} |\mathcal{S}_k| \;\leq\; M - K + \frac{K}{|\mathcal{G}_p|} \qquad\qquad p = 1, \ldots P\,. \tag{3.18}$$

Proof. Consider again the SVD in (3.16), $\mathbf{G}_I\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$, with $\mathbf{V}^H$ partitioned into P plus $L-P$ columns similar to $\mathbf{X}$, $\mathbf{V}^H = [\mathbf{V}_p^H \;\; \mathbf{V}_d^H]$. Introducing again the unknown $K \times K$ mixture $\mathbf{T}$, this leads to the following equations

$$\mathbf{G}_I = \mathbf{U}\,\mathbf{T}, \;\; \mathbf{T}\,\mathbf{X}_p = \mathbf{\Sigma}\,\mathbf{V}_p^H \tag{3.19}$$

which together represent $K(M+P)$ equations in the $\sum_{k=1}^{K} |\mathcal{S}_k|$ unknowns $\mathbf{g}_I$ and the K^2 unknowns $\mathbf{T}$. The proper conditions for solvability of the equations (3.19), that the number of equations needs to be at least equal to the number of unknowns, then leads to (3.17). If now we consider the equations for group of users $\mathcal{G}_p$, multiplying $\mathbf{T}\,\mathbf{X}_p = \mathbf{\Sigma}\,\mathbf{V}_p^H$ by $\mathbf{x}_p^{(p)}$ and exploiting $\mathbf{X}_p\,\mathbf{x}_p^{(p)} = \mathbf{1}_{\mathcal{G}_p}$, then we get

$$(\mathbf{G}_I)_{\mathcal{G}_p} = \mathbf{U}\,(\mathbf{T})_{\mathcal{G}_p} \tag{3.20}$$

$$\mathbf{T}\,\mathbf{1}_{\mathcal{G}_p} = (\mathbf{T})_{\mathcal{G}_p}\,\mathbf{1} = \mathbf{\Sigma}\,\mathbf{V}_p^H \mathbf{x}_p^{(p)} \tag{3.21}$$

which represents $M\,|\mathcal{G}_p| + K$ equations in the $\sum_{k\in\mathcal{G}_p} |\mathcal{S}_k| + K\,|\mathcal{G}_p|$ unknowns in $(\mathbf{G}_I)_{\mathcal{G}_p}$ and $(\mathbf{T})_{\mathcal{G}_p}$, hence leading to (3.18). ■

It is worth noting that the proof of Proposition 3.1 along with the sufficient conditions for identifiability of the deterministic parameters, provides also a constructive method to determine the unknown parameters $\mathbf{G}_I$ and $\mathbf{X}_d$ if for each $p = 1, \dots P$, the sequence $\{\mathcal{G}_{p,1}, \mathcal{G}_{p,2}, \dots \mathcal{G}_{p,s}\}$ partitioning set $\mathcal{G}_p$ were known. In the following, we address this problem and provide an MP algorithm that enables to identify at iteration i the set $\mathcal{G}_{p,i}$ and determine the channel coefficients of all users in the set.

3.4.1 Message Passing Algorithm

Let us focus on the set $\mathcal{G}_p$ and associate to each user k and AP m variable node k and factor node m, respectively. We construct a bipartite graph by connecting a variable node with a factor node if the distance between the corresponding user and AP is lower than γ. We further assume that the factor nodes are initialized with the values of the vector $\mathbf{g}_{I,p}^c = \mathbf{G}_I \mathbf{1}_{\mathcal{G}_p}$, i.e., the sum of all the channel coefficients of users in the corresponding γ-neighborhood. Each variable node knows the matrix $\mathbf{U}$ that spans the channel subspace.

The initial step of the MP algorithm starts at the factor nodes. Each factor node m that is a leaf transmits its initialization value $g_{I,p,m}^c$ to its neighbor. It transmits an erasure Δ if it is not a leaf. At iteration i, each variable node k that has received at least a message that is not an erasure solves the system of equations $\mathbf{U}\mathbf{T}_{:,k} = \mathbf{G}_{I,:,k}$

utilizing that value. The construction of a system of K equations to determine $\mathbf{T}_{:,k}$ is detailed in the proof of Proposition 3.1 and exploits the channel sparsity. Once $\mathbf{T}_{:,k}$ is known, it is possible to determine all the non-zero channel coefficients $\mathbf{G}_{I,:,k}$. Then, variable node k transmits to all its neighbors the corresponding channel coefficients. Variable node k transmits the same messages in all the following iterations. If variable node k receives all erasures it transmits erasures to all its neighbors. The second step of iteration i determines the messages at the factor nodes. A factor node m computes a message for the output edge $< m, k >$ as the difference between its initialization value $g^c_{I,p,m}$ and all the incoming messages. The resulting message is not an erasure if all the incoming messages are not erasures otherwise the factor node transmits an erasure.

The MP algorithm ends when all the channel coefficients have been determined and in this case the identifiability conditions are satisfied or when no additional erasure can be determined and thus the system is not identifiable. Set $\mathcal{G}_{p,i}$ includes all the users/variable nodes that compute their channel coefficients at iteration i.

Interestingly, this algorithm is closely related to the MP algorithm for decoding of low density parity check (LDPC) codes in transmissions through binary erasure channels in [104]. It is worth noting that also for random generated CF massive MIMO systems with nodes independently generated, the corresponding graphs have edges intrinsically correlated due to the underlying geometric constraints and the corresponding sparse graphs do not have tree-like neighborhoods in asymptotic conditions. Then, the performance analysis of LDPC codes based on density evolution, see [104], is not directly applicable although the graph is sparse and the message passing yields exact results thanks to the noiseless nature of the considered system and thus the absence of error propagation. Additionally, let us consider the Karp-Sipser or greedy leaf removal procedure [105–107] which consists in removing from a graph sequentially all the leaves and observe that sequential or simultaneous removal of leaves is equivalent in asymptotic conditions. Then, the sufficient identifiability conditions in Proposition 3.1 are satisfied if the greedy leaf removal procedure yields an empty core.

3.5 Bayesian Semi-Blind Iterative Algorithm

Whereas deterministic parameter identifiability allows for consistency in SNR in the approximated model which ignores $\mathbf{C}_0$, in practice performance can be improved by furthermore exploiting prior information. Hence, exploiting the Rayleigh fading channel prior and capturing the uncorrelatedness and constant variance of the data symbols with an i.i.d. Gaussian prior, we propose a Bayesian semi-blind iterative algorithm alternating between channel estimation and linear multi-user detection in this section.

Applying conditional probability, the joint probability density function of $\mathbf{Y}$ and $\boldsymbol{\theta}$ is

$$f(\mathbf{Y}, \boldsymbol{\theta}) = f(\mathbf{Y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) \tag{3.22}$$

where $f(\boldsymbol{\theta})$ stands for the pdf of $\boldsymbol{\theta}$ and $f(\mathbf{Y}|\boldsymbol{\theta})$ stands for the pdf of $\mathbf{Y}$ conditioned to $\boldsymbol{\theta}$. The data symbols and the channel coefficients are a priori independent of each other, therefore, by substituting $\boldsymbol{\theta}$ defined in (3.3), we get

$$f(\mathbf{Y}, \mathbf{g}_I, \mathbf{X}_d) = f(\mathbf{Y}|\mathbf{g}_I, \mathbf{X}_d)\, f(\mathbf{g}_I)\, f(\mathbf{X}_d) \tag{3.23}$$

where $f(\mathbf{g}_I)$ and $f(\mathbf{X}_d)$ are respectively given by

$$\begin{aligned} f(\mathbf{g}_I) &= (\pi)^{-n_{\mathbf{g}_I}} \left(\det \mathbf{C}_{\mathbf{g}_I\mathbf{g}_I}\right)^{-1} \exp\big(- \mathbf{g}_I^H \mathbf{C}_{\mathbf{g}_I\mathbf{g}_I}^{-1} \mathbf{g}_I \big) \\ f(\mathbf{X}_d) &= (\pi)^{-K(L-P)} \exp\big(- \mathrm{tr}\{\mathbf{X}_d^H \mathbf{X}_d\} \big) \end{aligned} \tag{3.24}$$

applying the log function on both sides of (3.23), we get the overall log-likelihood function as follows

$$\begin{aligned} \ln f(\mathbf{Y}, \mathbf{g}_I, \mathbf{X}_d) &= \ln f(\mathbf{Y}|\mathbf{g}_I, \mathbf{X}_d) + \ln f(\mathbf{g}_I) + \ln f(\mathbf{X}_d) \\ &= -\mathrm{tr}\{(\mathbf{Y} - \sqrt{\rho}\, \mathbf{G}_I \mathbf{X})^H \mathbf{C}_{\mathbf{YY}}^{-1} (\mathbf{Y} - \sqrt{\rho}\, \mathbf{G}_I \mathbf{X})\} - \mathbf{g}_I^H \mathbf{C}_{\mathbf{g}_I\mathbf{g}_I}^{-1} \mathbf{g}_I \\ &\quad -\mathrm{tr}\{\mathbf{X}_d^H \mathbf{X}_d\} + c^t \, . \end{aligned} \tag{3.25}$$

where c^t in (3.25) is a scalar constant. Alternating optimization w.r.t. $\mathbf{g}_I$ and $\mathbf{X}_d$ leads to the Bayesian semi-blind iterative algorithm illustrated in Algorithm 2. For a detailed proof of the Algorithm 2, see Appendix B.2. Note that the estimate of the matrix $\mathbf{G}_I$ is denoted by $\widehat{\mathbf{G}}_I$ and the relation between $\widehat{\mathbf{g}}_I$ and $\widehat{\mathbf{G}}_I$ is the same as the relation described for $\mathbf{g}_I$ and $\mathbf{G}_I$.

Algorithm 2 Bayesian Semi-Blind Iterative Algorithm

1. **Initialization** $\widehat{\mathbf{X}}_d^{(0)} = \mathbf{0}$

2. **Iteration** $(i+1)$
   - Minimization w.r.t. $\mathbf{g}_I$; $\mathbf{X}_d = \widehat{\mathbf{X}}_d^{(i)}$:
     $\Rightarrow \widehat{\mathbf{g}}_I^{(i+1)} = \sqrt{\rho}\left(\rho\,\mathbf{Q}^H(\widehat{\mathbf{X}}_d^{(i)})\,\mathbf{C}_{\mathbf{yy}}^{-1}\,\mathbf{Q}(\widehat{\mathbf{X}}_d^{(i)}) + \mathbf{C}_{\mathbf{g}_I\mathbf{g}_I}^{-1}\right)^{-1}\mathbf{Q}^H(\widehat{\mathbf{X}}_d^{(i)})\,\mathbf{C}_{\mathbf{yy}}^{-1}\,\mathbf{y}$
   - Minimization w.r.t. $\mathbf{X}_d$; $\mathbf{g}_I = \widehat{\mathbf{g}}_I^{(i+1)}$:
     $\Rightarrow \widehat{\mathbf{X}}_d^{(i+1)} = \sqrt{\rho}\left(\rho\,\widehat{\mathbf{G}}_I^{(i+1)^H}\mathbf{C}_{\mathbf{YY}}^{-1}\,\widehat{\mathbf{G}}_I^{(i+1)} + \mathbf{I}_K\right)^{-1}\widehat{\mathbf{G}}_I^{(i+1)^H}\mathbf{C}_{\mathbf{YY}}^{-1}\,\mathbf{Y}_d$

3. Repeat step 2 until $(\widehat{\mathbf{X}}_d^{(i+1)}, \widehat{\mathbf{g}}_I^{(i+1)}) \approx (\widehat{\mathbf{X}}_d^{(i)}, \widehat{\mathbf{g}}_I^{(i)})$.

3.6 Semi-Blind Approach with Gaussian Inputs

In this section, we consider approaches in which the unknown $\mathbf{X}_d$ are (still) modeled as i.i.d. Gaussian and hence can be eliminated, leading to the Gaussian distribution $f(\mathbf{Y}|\mathbf{g})$, where $\mathbf{g} = \mathrm{vec}(\mathbf{G})$. So, eliminating the i.i.d. Gaussian $\mathbf{X}_d$, the log-likelihood is given by

$$\begin{aligned} -\ln f(\mathbf{Y}|\mathbf{g}) &= -\ln f(\mathbf{Y}_p|\mathbf{g}) - \ln f(\mathbf{Y}_d|\mathbf{g}) \\ &= \mathrm{tr}\big\{\big(\mathbf{Y}_p - \sqrt{\rho}\,\mathbf{G}\mathbf{X}_p\big)^H\,\big(\mathbf{Y}_p - \sqrt{\rho}\,\mathbf{G}\mathbf{X}_p\big)\big\} \\ &\quad + (L-P)\ln\det\big(\rho\,\mathbf{G}\mathbf{G}^H + \mathbf{I}_M\big) + \mathrm{tr}\big\{\big(\rho\,\mathbf{G}\mathbf{G}^H + \mathbf{I}_M\big)^{-1}\mathbf{Y}_d\mathbf{Y}_d^H\big\} + c^t \end{aligned} \tag{3.26}$$

So, the per channel use covariance matrix in the blind data part is $\mathbf{C}_{\mathbf{YY}} = \rho\,\mathbf{G}\mathbf{G}^H + \mathbf{I}_M$. The non-quadratic appearance of $\mathbf{G}$ in the last two terms in (3.26) complicates the obtention of the posterior $f(\mathbf{g}|\mathbf{Y})$. The maximum a posteriori (MAP) estimator does not require the posterior, and can be obtained by maximizing $f(\mathbf{Y}|\mathbf{g})\,f(\mathbf{g})$. For this maximization, maybe the last two terms in (3.26) can be replaced by a quadratic minorizer (linear in $\mathbf{C}_{\mathbf{YY}}$, obtained by linearization).

The channel MAP estimator in the following subsection differs from the channel estimate obtained by the joint channel and data MAP estimator detailed in Algorithm 2 and can be expected to be closer to the MMSE estimate.

3.6.1 Joint Channel MAP for All Users

In the (3.26), the pilot part is convex. For the blind part $-\ln f(\mathbf{Y}_d|\mathbf{g})$, we construct a convex majorizer as in [108, Section V.A 4)], [109] which can actually also been derived with an expectation maximization (EM) approach. The construction of the majorizer is simply based on first-order Taylor series expansion of concave functions, either w.r.t. $\mathbf{G}$ directly or w.r.t. a covariance type expression (which is then quadratic in $\mathbf{G}$). Let $\mathbf{G}'$ be a current estimate of $\mathbf{G}$, then the term $\ln\det(\rho\,\mathbf{G}\,\mathbf{G}^H + \mathbf{I}_M)$ is upperbounded as

$$\begin{aligned} \ln\det(\rho\,\mathbf{G}\,\mathbf{G}^H + \mathbf{I}_M) &= \ln\det(\rho\,\mathbf{G}^H\,\mathbf{G} + \mathbf{I}_K) \\ &\leq \mathrm{tr}\{\rho\,(\rho\,\mathbf{G}'^H\,\mathbf{G}' + \mathbf{I}_K)^{-1}\mathbf{G}^H\mathbf{G}\,\}. \end{aligned} \tag{3.27}$$

On the other hand, using the matrix inversion lemma and Taylor series expansion, second term of the blind part can be upperbounded as

$$\begin{aligned} \mathrm{tr}\{(\rho\,\mathbf{G}\,\mathbf{G}^H + \mathbf{I}_M)^{-1}\,\mathbf{Y}_d\mathbf{Y}_d^H\} &= \mathrm{tr}\big\{[\mathbf{I}_M - \mathbf{G}(\mathbf{G}^H\,\mathbf{G} + \rho^{-1}\,\mathbf{I}_K)^{-1}\mathbf{G}^H]\,\mathbf{Y}_d\mathbf{Y}_d^H\big\} \\ &\leq \mathrm{tr}\{\mathbf{G}\,\mathcal{A}'\,\mathbf{G}^H - \mathcal{B}'\,\mathbf{G}^H - \mathbf{G}\,\mathcal{B}'^H\} + c^t \end{aligned} \tag{3.28}$$

where $\mathcal{A}'$ and $\mathcal{B}'$ are given by

$$\begin{aligned} \mathcal{A}' &= (\mathbf{G}'^H\,\mathbf{G}' + \frac{1}{\rho}\,\mathbf{I}_K)^{-1}\mathbf{G}'^H\mathbf{Y}_d\mathbf{Y}_d^H\mathbf{G}'(\mathbf{G}'^H\,\mathbf{G}' + \frac{1}{\rho}\,\mathbf{I}_K)^{-1} \\ \mathcal{B}' &= \mathbf{Y}_d\mathbf{Y}_d^H\mathbf{G}'(\mathbf{G}'^H\,\mathbf{G}' + \frac{1}{\rho}\,\mathbf{I}_K)^{-1}\,. \end{aligned} \tag{3.29}$$

Combining (3.27), (3.28) lead to the following quadratic majorizer in (3.26)

$$\begin{aligned} -\ln f(\mathbf{Y}_d|\mathbf{g}) &= (L-P)\ln\det\left(\rho\,\mathbf{G}\mathbf{G}^H + \mathbf{I}_M\right) + \mathrm{tr}\big\{\left(\rho\,\mathbf{G}\mathbf{G}^H + \mathbf{I}_M\right)^{-1}\mathbf{Y}_d\mathbf{Y}_d^H\big\} \\ &\leq \mathrm{tr}\big\{\big(\rho(L-P)(\rho\,\mathbf{G}'^H\,\mathbf{G}' + \mathbf{I}_K)^{-1} + \mathcal{A}'\big)\mathbf{G}^H\mathbf{G} - \mathcal{B}'\,\mathbf{G}^H - \mathbf{G}\,\mathcal{B}'^H\big\} \\ &\quad + c^t\,. \end{aligned} \tag{3.30}$$

Note that the quantities in (3.27), (3.29) have the following interpretation

$$\begin{aligned} \widehat{\mathbf{X}}_d &= \frac{1}{\sqrt{\rho}}\,(\mathbf{G}'^H\,\mathbf{G}' + \frac{1}{\rho}\,\mathbf{I}_K)^{-1}\mathbf{G}'^H\mathbf{Y}_d \\ \mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d} &= (L-P)\,(\rho\,\mathbf{G}'^H\,\mathbf{G}' + \mathbf{I}_K)^{-1} \end{aligned} \tag{3.31}$$

which are the LMMSE estimate and associated error covariance matrix of $\mathbf{X}_d$ (which is i.i.d. across channel uses). This means that the majorizer in (3.30) has the following

expectation maximization interpretation (which was not observed in [109], in spite of EM being discussed there also):

$$\begin{aligned}
-\ln f(\mathbf{Y}_d|\mathbf{G}) &= -\ln \mathbb{E}_{\mathbf{X}_d}\left\{f(\mathbf{Y}_d|\mathbf{X}_d,\mathbf{G})\right\} \\
&= -\ln \mathbb{E}_{\mathbf{X}_d|\mathbf{Y}_d,\mathbf{G}'}\left\{f(\mathbf{Y}_d|\mathbf{X}_d,\mathbf{G})\, f(\mathbf{X}_d)/f(\mathbf{X}_d|\mathbf{Y}_d,\mathbf{G}')\right\} \\
&= -\ln \mathbb{E}_{\mathbf{X}_d|\mathbf{Y}_d,\mathbf{G}'}\left\{f(\mathbf{Y}_d|\mathbf{X}_d,\mathbf{G})\right\} + c^t \\
&\leq \mathbb{E}_{\mathbf{X}_d|\mathbf{Y}_d,\mathbf{G}'}\{-\ln f(\mathbf{Y}_d|\mathbf{X}_d,\mathbf{G})\} + c^t \\
&= \mathbb{E}_{\mathbf{X}_d|\mathbf{Y}_d,\mathbf{G}'}\|\mathbf{Y}_d - \sqrt{\rho}\,\mathbf{G}\,\mathbf{X}_d\|^2 + c^t \\
&= \|\mathbf{Y}_d - \sqrt{\rho}\,\mathbf{G}\,\widehat{\mathbf{X}}_d\|^2 + \rho\,\mathbf{G}\,\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d}\mathbf{G}^H + c^t
\end{aligned} \tag{3.32}$$

where the inequality follows from Jensen's inequality and the convexity of $-\ln(.)$, and c^t denotes (various) terms that are constant w.r.t. $\mathbf{G}$. The MMSE estimate $\widehat{\mathbf{X}}_d$ and error covariance matrix $\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d}$ are defined in (3.31) and correspond to LMMSE estimation due to the joint Gaussianity of $f(\mathbf{Y}_d,\mathbf{X}_d|\mathbf{G}')$. Then, the quadratic majorizer is obtained as follows

$$\begin{aligned}
-\ln f(\mathbf{Y}|\mathbf{g}) &\leq f'(\mathbf{g}|\mathbf{g}') \\
&= \mathrm{tr}\big\{\big(\mathbf{Y}_p - \sqrt{\rho}\,\mathbf{G}\mathbf{X}_p\big)^H\,\big(\mathbf{Y}_p - \sqrt{\rho}\,\mathbf{G}\mathbf{X}_p\big)\big\} + \|\mathbf{Y}_d - \sqrt{\rho}\,\mathbf{G}\,\widehat{\mathbf{X}}_d\|^2 \\
&\quad + \rho\,\mathbf{G}\,\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d}\mathbf{G}^H + c^t \\
&= \rho\,\mathrm{tr}\{\mathbf{G}^H\mathbf{G}(\mathbf{X}_p\mathbf{X}_p^H + \widehat{\mathbf{X}}_d\widehat{\mathbf{X}}_d^H + \mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d})\} \\
&\quad -2\sqrt{\rho}\,\Re\,\mathrm{tr}\{\mathbf{G}(\mathbf{X}_p\mathbf{Y}_p^H + \widehat{\mathbf{X}}_d\mathbf{Y}_d^H)\} + c^t
\end{aligned} \tag{3.33}$$

where $\Re\{.\}$ denotes real part operator. The quadratic majorizer in (3.33) is separable between the channel use dimension and the receiver antenna dimension. When we add the channel prior, which contains different channel covariance matrices for different users, we need to switch from $\mathbf{G}$ to $\mathbf{g}$ and we get with $-\ln f(\mathbf{g}) = \mathbf{g}^H\,\mathbf{C}_{\mathbf{gg}}^{-1}\,\mathbf{g} + c^t$, where $\mathbf{C}_{\mathbf{gg}} = \mathrm{diag}(\beta_{11},\ldots,\beta_{M1}\ldots\beta_{1k},\ldots,\beta_{Mk}\ldots\beta_{1K},\ldots,\beta_{MK})$,

$$\begin{aligned}
-\ln f(\mathbf{Y}|\mathbf{g}) - \ln f(\mathbf{g}) &\leq f(\mathbf{g}|\mathbf{g}') \\
&= \rho\,\mathbf{g}^H\big(\mathbf{C}_{\mathbf{gg}}^{-1} + \rho\,((\mathbf{X}_p\mathbf{X}_p^H + \widehat{\mathbf{X}}_d\widehat{\mathbf{X}}_d^H + \mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d})^T \otimes \mathbf{I}_M)\big)\,\mathbf{g} \\
&\quad -2\sqrt{\rho}\,\Re\left\{\mathbf{g}^H\,\mathrm{vec}(\mathbf{Y}_p\mathbf{X}_p^H + \mathbf{Y}_d\widehat{\mathbf{X}}_d^H)\right\} + c^t
\end{aligned} \tag{3.34}$$

which leads to the following estimate

$$\widehat{\mathbf{g}} = \sqrt{\rho}\left(\mathbf{C}_{\mathbf{gg}}^{-1} + \rho\left((\mathbf{X}_p\mathbf{X}_p^H + \widehat{\mathbf{X}}_d\widehat{\mathbf{X}}_d^H + \mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d})^T \otimes \mathbf{I}_M\right)\right)^{-1}\mathrm{vec}(\mathbf{Y}_p\mathbf{X}_p^H + \mathbf{Y}_d\widehat{\mathbf{X}}_d^H)\,. \tag{3.35}$$

The estimate in (3.35) needs to be solved iteratively, with $\widehat{\mathbf{X}}_d$ and $\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d}$ in (3.31) computed with the previous channel estimate. The iterative process can be initialized with $\widehat{\mathbf{g}}^{(-1)} = \mathbf{0}$, which leads to a first iterate $\widehat{\mathbf{g}}^{(0)}$ being based only on pilots and prior information. The proposed joint channel MAP for all users is summarized in Algorithm 3.

Note that the Bayesian semi-blind MAP (Algorithm 2) can be obtained from the joint channel MAP by putting $\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d} = \mathbf{0}$ in (3.35). Note also that the performance of the joint channel MAP is bounded by the Gaussian inputs CRB, which should be attained asymptotically in L.

Algorithm 3 Joint Channel MAP for All Users

1: Initialize $\widehat{\mathbf{g}} = \mathbf{0}$, $\widehat{\mathbf{X}}_d = \mathbf{0}$, $\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d} = (L-P)\mathbf{I}_K$
2: **repeat**
3: $\quad t \leftarrow t+1$
4: $\quad$ compute $\widehat{\mathbf{g}}$ according to (3.35)
5: $\quad$ compute $\widehat{\mathbf{X}}_d$ and $\mathbf{C}_{\widetilde{\mathbf{X}}_d\widetilde{\mathbf{X}}_d}$ according to (3.31)
6: **until** convergence or $t = t_{\max}$
Return $\widehat{\mathbf{g}}$

3.7 Pilot Based Bayesian Performance Bounds

As the prior channel information is exploited, we consider estimating the whole channel $\mathbf{G}$ (limiting to $\mathbf{G}_I$ will not affect estimation performance much since $\mathbf{G}_0$ is small). We consider that pilots and data have the same power. Observe that $\mathrm{vec}(\mathbf{G}\,\mathbf{X}_p) = (\mathbf{X}_p^T \otimes \mathbf{I}_M)\,\mathbf{g}$. As a result the pilot portion leads to the following FIM

$$\mathrm{FIM}_p = \rho\,(\mathbf{X}_p^*\mathbf{X}_p^T) \otimes \mathbf{I}_M \tag{3.36}$$

for $\mathbf{g}$, which is singular, due to the pilot reuse. We can get a first idealized pilot only based CRB, by assuming that the pilots would somehow be orthogonal

$$\begin{aligned} \mathrm{CRB}_{p,o} &= \mathrm{FIM}_{p,o}^{-1}\,, \\ \mathrm{FIM}_{p,o} &= \rho\,\mathrm{diag}(\mathbf{X}_p^*\mathbf{X}_p^T) \otimes \mathbf{I}_M = \rho\,P\,I_{MK}\,. \end{aligned} \tag{3.37}$$

The pilot contamination can be alleviated by prior channel information, leading to the Bayesian pilot based CRB

$$\mathrm{CRB}_{p,B} = (\mathrm{FIM}_p + \mathbf{C}_{\mathbf{gg}}^{-1})^{-1}\,. \tag{3.38}$$

For the semi-blind approaches considered here, another ideal MSE lower bound can be considered, considering to the ideal scenario in which the data $\mathbf{X}_d$ would be detected exactly, hence becoming also pilots for the channel estimation, leading to the genie-aided Bayesian semi-blind CRB

$$\mathrm{CRB}_{p+d,B} = \left(\rho\,(\mathbf{X}^*\mathbf{X}^T) \otimes \mathbf{I}_M + \mathbf{C}_{\mathbf{gg}}^{-1}\right)^{-1} \tag{3.39}$$

where $\mathbf{X} = [\mathbf{X}_p\ \mathbf{X}_d]$.

For any of the CRBs considered, we get a corresponding normalized mean square error (NMSE) bound in the form of $\mathrm{NMSE} = \mathrm{tr}\{\mathrm{CRB}\}/\mathrm{tr}\{\mathbf{C}_{\mathbf{gg}}\}$, where asymptotically $\|\mathbf{g}\|^2 = \mathbb{E}\,\|\mathbf{g}\|^2 = \mathrm{tr}\{\mathbf{C}_{\mathbf{gg}}\}$. Note that we get for all Bayesian approaches $\mathrm{NMSE} < 1$.

3.8 Gaussian Inputs Bayesian Semi-Blind CRB

Eliminating the i.i.d. Gaussian $\mathbf{X}_d$, we get the log-likelihood as follows

$$\begin{aligned}\ln f(\mathbf{Y}|\mathbf{g}) \;=\; & -\mathrm{tr}\left\{\left(\mathbf{Y}_p - \sqrt{\rho}\,\mathbf{G}\mathbf{X}_p\right)^H \left(\mathbf{Y}_p - \sqrt{\rho}\,\mathbf{G}\mathbf{X}_p\right)\right\} \\ & - (L-P)\ln\det\left(\rho\,\mathbf{G}\mathbf{G}^H + \mathbf{I}_M\right) - \mathrm{tr}\left\{\left(\rho\,\mathbf{G}\mathbf{G}^H + \mathbf{I}_M\right)^{-1}\mathbf{Y}_d\mathbf{Y}_d^H\right\} + c^t\end{aligned} \tag{3.40}$$

The blind FIM can be shown to be

$$\mathrm{FIM}_b = \rho^2\,(\mathbf{G}^H\mathbf{C}^{-1}\mathbf{G})^* \otimes \mathbf{C}^{-1}\,,\ \mathbf{C} = \rho\,\mathbf{G}\mathbf{G}^H + \mathbf{I}_M \tag{3.41}$$

which results in the deterministic semi-blind CRB

$$\mathrm{CRB}_{SB,d} = (\mathrm{FIM}_p + (L-P)\mathrm{FIM}_b)^{-1}\,. \tag{3.42}$$

which depends on the true channel. $\mathrm{CRB}_{SB,d}$ could be compared to its genie-aided version $\mathrm{CRB}_{p+d,d} = \dfrac{1}{\rho}\,(\mathbf{X}^*\mathbf{X}^T)^{-1} \otimes \mathbf{I}_M$. The corresponding Bayesian semi-blind CRB

$$\begin{aligned}\mathrm{CRB}_{SB,B} &= (\mathrm{FIM}_p + (L-P)\,\mathbb{E}_{\mathbf{g}}\{\mathrm{FIM}_b\} + \mathbf{C}_{\mathbf{gg}}^{-1})^{-1} \\ &= (\mathrm{CRB}_{p,B}^{-1} + (L-P)\,\mathbb{E}_{\mathbf{g}}\{\mathrm{FIM}_b\})^{-1}\end{aligned} \tag{3.43}$$

is difficult to compute analytically (except at low/high SNR) but more importantly, can be expected to be quite loose. In any case, assuming that $\mathbf{G}$ is tall ($M > K$), at

high SNR we get the dominating term

$$\mathrm{FIM}_b^{hSNR} = \rho\, \mathbf{I}_K \otimes \mathbf{P}_{\mathbf{G}}^{\perp} \tag{3.44}$$

where $\mathbf{P}_{\mathbf{G}}^{\perp}$ denotes the projection on the orthogonal complement of the column space of $\mathbf{G}$. Then we get approximately

$$\mathbb{E}_{\mathbf{g}}\{\mathrm{FIM}_b^{hSNR}\} \approx \rho\,(1 - \frac{K}{M})\,\mathbf{I}_{MK} \tag{3.45}$$

which would be exact if the elements of $\mathbf{G}$ were i.i.d. On the other hand, at low SNR we get $\mathbf{C} \approx \mathbf{I}_M$ and hence $\mathrm{FIM}_b^{lSNR} = \rho^2\,(\mathbf{G}^H\mathbf{G})^* \otimes \mathbf{I}_M$ from which

$$\mathbb{E}_{\mathbf{g}}\{\mathrm{FIM}_b^{lSNR}\} \approx \rho^2\,\mathrm{tr}_M\{\mathbf{C}_{\mathbf{gg}}\} \otimes \mathbf{I}_M \tag{3.46}$$

where $\mathrm{tr}_M\{\mathbf{A}\}$ is a diagonal matrix obtained by taking the trace of $\mathbf{A}$ over consecutive diagonal element portions of size M. In other words $(\mathrm{tr}_M\{\mathbf{C}_{\mathbf{gg}}\})_{k,k} = \mathbb{E}\|\mathbf{g}_k\|^2$. Remains to find an interpolation between low and high SNR. In this regard, consider the SVD of $\mathbf{G} = \mathbf{U}\,[\mathbf{\Sigma}\ \mathbf{0}]^T\,\mathbf{V}^H$ (where $\mathbf{U}$, $\mathbf{V}$ and the diagonal $\mathbf{\Sigma}$ are square) and note that $\mathbf{C}^{-1} = \mathbf{I} - \mathbf{G}(\mathbf{G}^H\mathbf{G} + \frac{1}{\rho}\mathbf{I})^{-1}\mathbf{G}^H$, then we get

$$\begin{aligned}
\mathrm{FIM}_b &= \rho\left(\mathbf{G}^H\mathbf{G}(\mathbf{G}^H\mathbf{G} + \frac{1}{\rho}\mathbf{I})^{-1}\right)^* \otimes \mathbf{C}^{-1} \\
&= \rho\left(\mathbf{V}\mathbf{\Sigma}^2(\mathbf{\Sigma}^2 + \frac{1}{\rho}\mathbf{I})^{-1}\mathbf{V}^H\right)^* \otimes \left(\mathbf{U}\begin{bmatrix}(\rho\mathbf{\Sigma}^2+\mathbf{I})^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{M-K}\end{bmatrix}\mathbf{U}^H\right).
\end{aligned} \tag{3.47}$$

If $\mathbf{G}$ would have had i.i.d. elements then $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{\Sigma}$ would be independent. This incites us to take the expectation of the two factors in the Kronecker product separately. With some further approximation, we then get

$$\begin{aligned}
\mathbb{E}_{\mathbf{g}}\{\mathrm{FIM}_b\} \approx \rho\,(1 - \frac{K}{M} + \frac{1}{M}\,\mathrm{tr}\{(\rho\,\mathrm{tr}_M\{\mathbf{C}_{\mathbf{gg}}\} + \mathbf{I}_K)^{-1}\}) \\
\mathrm{tr}_M\{\mathbf{C}_{\mathbf{gg}}\}(\mathrm{tr}_M\{\mathbf{C}_{\mathbf{gg}}\} + \frac{1}{\rho}\mathbf{I}_K)^{-1} \otimes \mathbf{I}_M
\end{aligned} \tag{3.48}$$

which is consistent with the high and low SNR limits in (3.45), (3.46), and which needs to be plugged into (3.43).

3.9 Gaussian-Gaussian Extrinsic Information Lower Bound

Another performance bound can be obtained by considering $\mathbf{G}_{\overline{k}}\mathbf{X}_{d,\overline{k}}$ as Gaussian. Since for the estimation of the signal of user k, considering the interfering signals to be Gaussian corresponds to a worst case interference for given interference covariance, the resulting mutual information lower bound should lead to an information matrix lower bound and hence to an error covariance (MSE) upper bound. For the signal k in the resulting model, the Gaussian input $\mathbf{x}_{d,k}$, i.e., the data symbols sent by user k, can then be eliminated, and the CRB for $\mathbf{g}_k$ can be computed. In the absence of a prior on $\mathbf{g}_k$, this would correspond to extrinsic information on $\mathbf{g}_k$.

Let $\mathbf{g}_k \sim \mathcal{CN}(0, \mathbf{C}_k)$ with $\mathbf{C}_k = \mathrm{diag}(\beta_{1k}, \ldots, \beta_{Mk})$, $\mathbf{y} = [\mathbf{y}_p^T \ \mathbf{y}_d^T]^T$, and $\mathbf{x}_k = [\mathbf{x}_{p,k} \ \mathbf{x}_{d,k}]$. Eliminating the Gaussian $\mathbf{G}_{\overline{k}}$, $\mathbf{x}_k$ and $\mathbf{G}_{\overline{k}}\mathbf{X}_{d,\overline{k}}$, we get

$$\begin{aligned} \ln f(\mathbf{y}|\mathbf{g}_k) \ &= \ -(\mathbf{y}_p - \sqrt{\rho}\,\mathbf{x}_{p,k}^T \otimes \mathbf{g}_k)^H (\mathbf{I}_{MP} + \rho \textstyle\sum_{i\neq k} \mathbf{x}_{p,i}^H \mathbf{x}_{p,i} \otimes \mathbf{C}_i)^{-1} (\mathbf{y}_p - \sqrt{\rho}\,\mathbf{x}_{p,k}^T \otimes \mathbf{g}_k) \\ &\quad -(L-P) \ln\det\left(\mathbf{I}_{M(L-P)} + \rho\, \mathbf{I}_{L-P} \otimes (\mathbf{g}_k\mathbf{g}_k^H + \textstyle\sum_{i\neq k} \mathbf{C}_i)\right) \\ &\quad -\mathbf{y}_d^H \left(\mathbf{I}_{M(L-P)} + \rho\, \mathbf{I}_{L-P} \otimes (\mathbf{g}_k\mathbf{g}_k^H + \textstyle\sum_{i\neq k} \mathbf{C}_i)\right)^{-1} \mathbf{y}_d + c^t. \end{aligned} \tag{3.49}$$

The data portion of (3.49) can be simplified as follows

$$\ln p(\mathbf{y}_d|\mathbf{g}_k) = -(L-P)\ \ln\det(\mathbf{\Sigma}) - \mathrm{tr}\{\mathbf{Y}_d^H \mathbf{\Sigma}^{-1} \mathbf{Y}_d\} + c^t, \tag{3.50}$$

where

$$\mathbf{\Sigma} = \mathbf{I}_M + \rho\,(\mathbf{g}_k\mathbf{g}_k^H + \sum_{i\neq k} \mathbf{C}_i)$$

and we used $\det(\mathbf{A}\otimes\mathbf{B}) = \det(\mathbf{A})^m \det(\mathbf{B})^n$, where $\mathbf{A}$ is an $n\times n$ matrix and $\mathbf{B}$ is an $m\times m$ matrix, and $\mathrm{vec}^T(\mathbf{A})\,(\mathbf{D}\otimes\mathbf{B})\,\mathrm{vec}(\mathbf{C}) = \mathrm{tr}\{\mathbf{A}^T\,\mathbf{B}\,\mathbf{C}\,\mathbf{D}^T\}$. Using the FIM for a circularly complex Gaussian pdf, we get

$$\begin{aligned} \mathrm{FIM}_{\mathbf{g}_k}^{GGei} \ &= \ \rho^2\,(L-P)\,\mathbf{g}_k^H \mathbf{\Sigma}^{-1} \mathbf{g}_k\, \mathbf{\Sigma}^{-*} \\ &\quad + \rho(\mathbf{x}_{p,k}\otimes\mathbf{I}_M)(\mathbf{I}_{MP} + \rho \textstyle\sum_{i\neq k} \mathbf{x}_{p,i}^H \mathbf{x}_{p,i} \otimes \mathbf{C}_i)^{-1} (\mathbf{x}_{p,k}\otimes\mathbf{I}_M)^H \end{aligned} \tag{3.51}$$

where $\mathbf{\Sigma}^{-*} = (\mathbf{\Sigma}^*)^{-1}$. Then, the extrinsic information CRB upper bound is obtained

a the inverse of the FIM

$$\mathrm{CRB}_{\mathbf{g}_k}^{\mathrm{GGei}} = (\mathrm{FIM}_{\mathbf{g}_k}^{\mathrm{GGei}})^{-1}$$

where GGei stands for Gaussian-Gaussian extrinsic information.

3.10 Simulation Results

In this section, we provide some numerical results verifying the analytical derivations in this chapter. The M APs and K users are uniformly distributed at random over a square area of size $D \times D$. The large-scale fading coefficient β_{mk} in (3.1) models the path loss and shadow fading as follows

$$\beta_{mk} = 10^{\frac{\mathrm{PL}_{mk}}{10}} \, 10^{\frac{\sigma_{sh} z_{mk}}{10}} \tag{3.52}$$

where PL_{mk} represents the path loss (expressed in dB), and $10^{\frac{\sigma_{sh} z_{mk}}{10}}$ represents the shadow fading with standard deviation σ_{sh}, and $z_{mk} \sim \mathcal{N}(0,1)$, i.e., we assume uncorrelated shadow fading.

The three-slope model in [110] is adopted for the path loss. Three values of loss exponent can be distinguished according to the distance between the k-th user and the m-th AP, denoted by d_{mk}. More particularly, the path-loss exponent equals 3.5 if $d_{mk} > d_1$, equals 2 if $d_0 < d_{mk} \leq d_1$, and equals 0 if $d_{mk} \leq d_0$. More precisely, the path loss PL_{mk} (in dB) is given by

$$\mathrm{PL}_{mk} = \begin{cases} -L_0 - 10\log_{10}(d_{mk}^{3.5}) & \text{if } d_{mk} > d_1 \\ -L_0 - 10\log_{10}(d_1^{1.5} d_{mk}^2) & \text{if } d_0 < d_{mk} \leq d_1 \\ -L_0 - 10\log_{10}(d_1^{1.5} d_0^2) & \text{if } d_{mk} \leq d_0 \end{cases} \tag{3.53}$$

where

$$L_0 \triangleq 46.3 + 33.9\log_{10}(f) - 13.82\log_{10}(h_{AP}) - [1.11\log_{10}(f) - 0.7]h_{UE} + 1.56\log_{10}(f) - 0.8,$$

where f is the carrier frequency (in MHz), h_{AP} and h_{UE} denotes the AP and user antenna heights (in meter), respectively. The path loss PL_{mk} is a continuous function of d_{mk}. In our simulation setup, we consider a communication bandwidth of $W = 20$

MHz centered over the carrier frequency $f = 1.9$ GHz. The antenna height at the AP is $h_{AP} = 15$ m and at the user is $h_{UE} = 1.65$ m. The standard deviation of the shadow fading is $\sigma_{sh} = 8$ dB, the parameters for the three slope path loss model in (3.53) are $d_1 = 50$ m and $d_0 = 10$ m.

We take pilot sequences to be rows of the identity matrix $\mathbf{I}_P$. First, we fit the big square with the repetition of the pattern of small squares. In a first instance, we consider a pattern of 4 small squares, 2×2. Then, we partition the pilots in 4 groups of $P/4$ and assign each portion of $P/4$ pilots to each of the 4 small squares and keep this assignment as the pattern of 4 small squares gets repeated to fill the big square. In this way, no two neighboring small squares have common pilots.

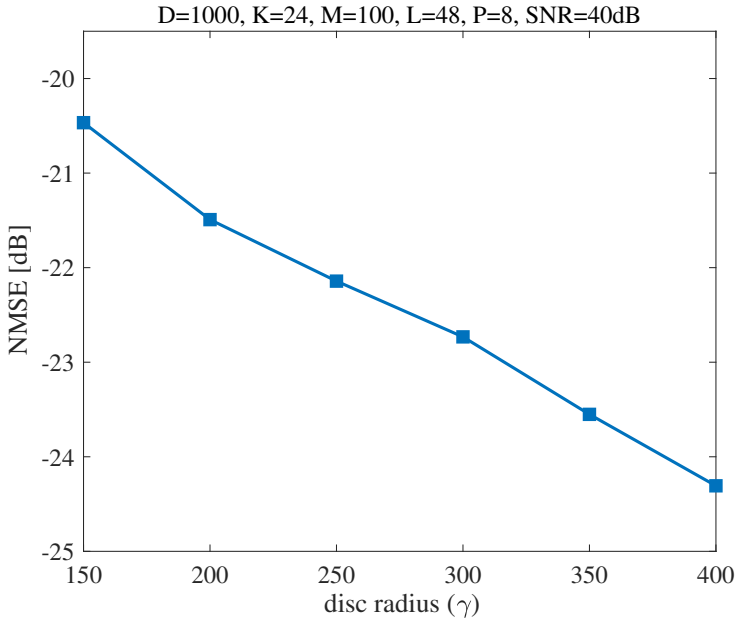

Figure 3.2: NMSE [dB] versus disc radius (γ) for Bayesian semi-blind estimation.

Throughout this section, we consider the following scenario. The $M = 100$ APs and $K = 24$ users are uniformly distributed at random over a square area of side $D = 1000$ and we consider $P = 8$ and $L = 48$. The performance of the Bayesian estimation is assessed by the normalized mean square error (NMSE) defined as

$$\text{NMSE} = \frac{avg\|\mathbf{g}_I - \widehat{\mathbf{g}}_I\|^2}{avg\|\mathbf{g}_I\|^2} \tag{3.54}$$

where avg stands for average. Fig. 3.2 illustrates NMSE [dB] versus the disc radius

varying in the range $\gamma \in [150, 400]$, for Bayesian semi-blind estimation in Algorithm 2. The performance of Bayesian semi-blind approach becomes better as the disc radius increases. Fig. 3.3 compares the performance of Bayesian semi-blind estimation and deterministic CRB and presents NMSE [dB] versus SNR [dB] for different values of disc radius γ. The NMSE for the deterministic CRB is defined as $\frac{avg\,\mathrm{tr}\{\mathrm{CRB}^{d}_{\mathbf{g}_I}\}}{avg\|\mathbf{g}_I\|^2}$. Fig. 3.3 corroborates the analytical derivations and the non-singularity of the FIM and thus, the existence of the CRB. The Bayesian estimation outperforms the deterministic CRB and increasing the SNR and disc radius γ, the Bayesian estimation performance improves. On the contrary, the deterministic CRB behaves differently, the performance becomes worse as the disc radius γ increases. In the deterministic CRB, increasing the radius γ more parameters have to get estimated and they are difficult to estimate whereas in the Bayesian estimation we exploit priors so the estimation error has posterior variances which is at most as large as the prior variances and increasing the radius there are more coefficients for which the prior variance gets reduced to posterior variance.

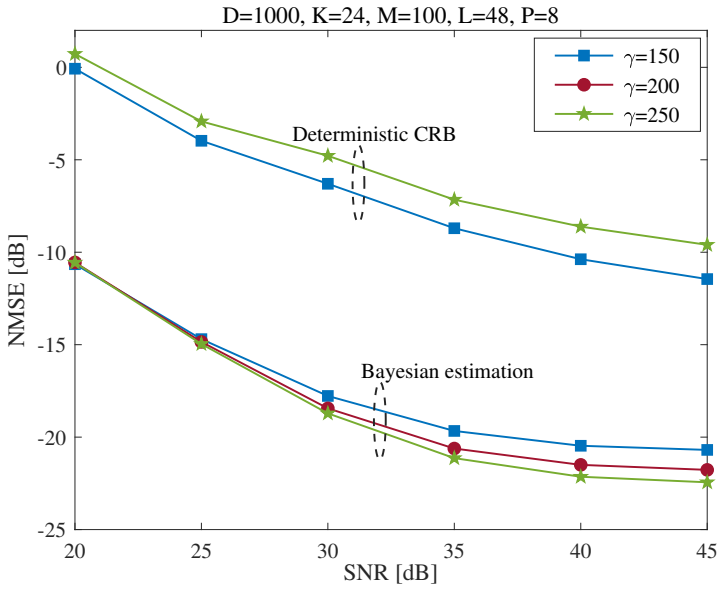

Figure 3.3: NMSE [dB] versus SNR [dB] for Bayesian semi-blind estimation and deterministic CRB.

Fig. 3.4 compares the performance of Bayesian semi-blind channel estimation in Algorithm 2 and channel MAP estimation introduced in subsection (3.6.1), and presents

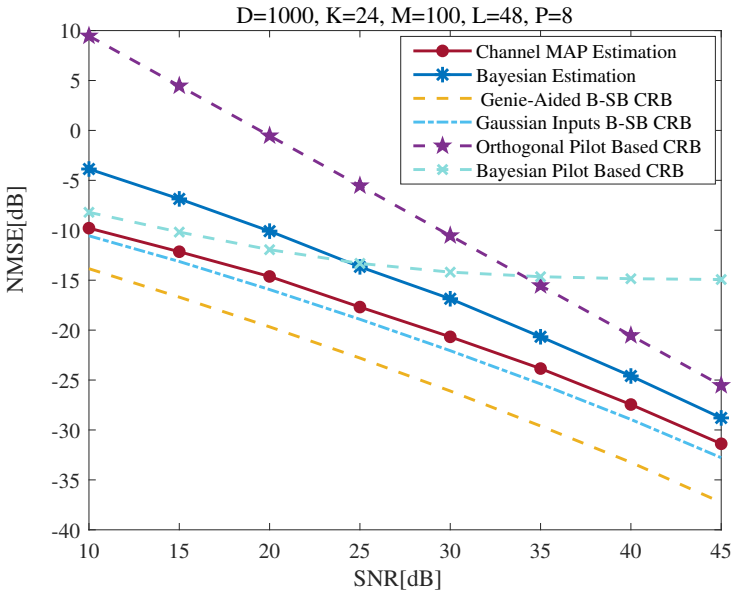

Figure 3.4: NMSE [dB] versus SNR [dB] for Bayesian semi-blind estimation and channel MAP estimation.

the NMSE [dB], NMSE $= \dfrac{avg\|\mathbf{g}-\widehat{\mathbf{g}}\|^2}{avg\|\mathbf{g}\|^2}$, versus SNR [dB]. In this figure, the performance of different CRBs, i.e., the genie-aided Bayesian semi-blind CRB ($\text{CRB}_{p+d,B}$), orthogonal pilot based CRB ($\text{CRB}_{p,o}$), Bayesian pilot based CRB ($\text{CRB}_{p,B}$), and Gaussian inputs Bayesian semi-blind CRB ($\text{CRB}_{SB,B}$), is evaluated as well. As expected, the channel MAP estimation outperforms Bayesian semi-blind channel estimation. The Bayesian semi-blind algorithm alternatingly estimates the channel or data as if the estimate for the other quantity is perfect, whereas the channel MAP estimation takes into account the data error covariance matrix. The curve corresponding to the channel MAP estimation is quite close to the genie-aided Bayesian semi-blind CRB, which shows that the channel MAP estimation works well.

3.11 Conclusion

In this chapter, we addressed the problem of pilot contamination in CF massive MIMO systems leveraging only the channel sparsity. Exploiting the sparsity of channels due to the strong attenuation with the path loss, we developed semi-blind joint channel estimation and data detection methods to combat pilot contamination. We analyzed

the potential of semi-blind approaches with classical signal processing techniques such as FIM, CRB, and identifiability. Additionally, we determined sufficient and necessary conditions for semi-blind identifiability under the assumption of deterministic parameters. We constructed a bipartite graph that has APs and users as factor and variable nodes and proposed an MP algorithm over this graph which computes the channel coefficients if the identifiability conditions are satisfied. We assumed that both the channel and the unknown symbols are random with Gaussian distribution supposed to be estimated jointly and proposed a Bayesian semi-blind approach resulting in an algorithm which alternates between channel estimation and linear multi-user detection. We also considered semi-blind approaches for channel estimation based on treating the unknown symbols as random with known prior distribution to be eliminated and proposed a semi-blind channel MAP estimation in the presence of Gaussian i.i.d. data. We derived FIM and CRBs under different assumptions to evaluate the performance of semi-blind approaches introduced in this chapter. We verified the analytical derivations via numerical simulations.

Chapter 4

Expectation Propagation Based Bayesian Semi-Blind Approach

4.1 Introduction

An extensive attention has been dedicated to the design of detectors relying on MP algorithms in recent years. The EP which is a kind of MP algorithm attempts to find the closest approximation for a computationally intractable target probability distribution from a tractable family of distributions in an iterative refinement procedure by minimizing a Kullback-Leiber (KL) distance [65–68]. The method of EP was firstly applied to MIMO detection in [68], where an EP-based MIMO detector shows near-optimal performance with acceptable complexity under specific conditions. With the EP-based MIMO detector, a Gaussian approximation is constructed for the posterior distribution of the transmitted symbols by an iterative procedure based on moment matching.

In this chapter, we consider semi-blind methods for channel estimation in the presence of Gaussian i.i.d. data to tackle the pilot contamination problem in CF massive MIMO systems. This task is further aided by exploiting prior channel information in a Bayesian formulation. We propose a variable level expectation propagation (VL-EP)

algorithm for MP style semi-blind channel estimation which provides an approximate MMSE channel estimator which itself can not be found analytically.

4.2 System Model

We consider a CF massive MIMO system in uplink in which M APs serve K users in the same time-frequency resource. All APs and users equipped with a single antenna are uniformly distributed in a given area. Furthermore, all APs are connected to a CPU via a back-haul network. The channel is assumed to remain constant over L consecutive symbol intervals, i.e., a block. The first P symbols of the block of L symbols serve as the pilot sequences known at the CPU and the remaining $L - P$ symbols are used for the data transmissions. The received signal $\mathbf{Y} \in \mathbb{C}^{M \times L}$ at the M APs over the block interval is given by

$$\mathbf{Y} = \sqrt{\rho}\, \mathbf{G}\, \mathbf{S}^T + \mathbf{V} \tag{4.1}$$

where $\mathbf{S} = [\mathbf{s}_1 \dots \mathbf{s}_K] \in \mathbb{C}^{L \times K}$ denotes the transmitted symbols in the block where $\mathbf{s}_k \in \mathbb{C}^{L \times 1}$ is the signal vector sent by the user k. The channel vector between user k and M APs is denoted by $\mathbf{g}_k = [g_{1k} \dots g_{Mk}]^T \in \mathbb{C}^{M \times 1}$, then the channel matrix between the APs and users is given by $\mathbf{G} = [\mathbf{g}_1 \dots \mathbf{g}_K] \in \mathbb{C}^{M \times K}$, whose (m, k)-element g_{mk} is the channel coefficient between AP m and user k and is modeled same as the one in (3.1). The matrix $\mathbf{V} \in \mathbb{C}^{M \times L}$ represents the AWGN with i.i.d. components having zero mean and unit variance.

Let the matrices $\mathbf{S}_p \in \mathbb{C}^{P \times K}$ and $\mathbf{S}_d \in \mathbb{C}^{(L-P) \times K}$ denote the pilot sequences and data symbols, respectively. Then, $\mathbf{S} = [\mathbf{S}_p^T \ \mathbf{S}_d^T]^T$ and $\mathbf{s}_k = [\mathbf{s}_{p,k}^T \ \mathbf{s}_{d,k}^T]^T$. Similarly, $\mathbf{Y} = [\mathbf{Y}_p \ \mathbf{Y}_d]$ where $\mathbf{Y}_p \in \mathbb{C}^{M \times P}$ and $\mathbf{Y}_d \in \mathbb{C}^{M \times (L-P)}$ represent the matrices of received training and data signals, respectively.

4.3 Expectation Propagation Algorithm

EP is an iterative algorithm for finding the best approximation to a desired distribution from a tractable family of distributions. In this section, first we briefly review EP principle [111].

Following the proposed algorithm in [65] and [69], suppose the parameter ϑ must be estimated from some independent measurements $x_1, x_2, \dots x_n$. As is common in Bayesian estimation, the prior distribution of ϑ is assumed to be known. Therefore the posterior distribution is given by

$$p(\vartheta|x_1, \dots, x_n) \propto p(\vartheta) \prod_{i=1}^{n} p(x_i|\vartheta) \triangleq \prod_{i=0}^{n} p_i(\vartheta) \tag{4.2}$$

where $\propto$ denotes equality up to a scale factor, $p_0 \triangleq p(\vartheta)$, and $p_i(\vartheta) \triangleq p(x_i|\vartheta)$ for $i = 1, 2, \dots n$. EP exploits this factorized structure to construct a tractable approximation to the above conditional distribution by a distribution from the exponential family, $q(\vartheta)$, of the form

$$q(\vartheta) \propto \prod_{i=0}^{n} q_i(\vartheta) \tag{4.3}$$

where $q_i(\vartheta), i = 0, 1, \dots n$ is from an exponential family. Several properties of the exponential family are helpful in simplifying the computations. Two of these properties are extensively used in the computations involved in EP. First is that as in (4.3), multiplication (or division) of two exponential distributions results in an exponential distribution. Moreover, the parameters of the resulting distribution are easily computed from the parameters of the constituent distributions. Next, the EP algorithm tries to iteratively find the closest $q(\vartheta)$ to the distribution $p(\vartheta|x_1, \dots, x_n)$ where closeness is in terms of the KL divergence. Therefore, $q(\vartheta)$ is the solution of the following optimization problem

$$q^*(\vartheta) = \arg\min_{q \in \mathcal{F}} \mathrm{KL}\big(p(\vartheta|x_1, \dots, x_n)||q(\vartheta)\big) \tag{4.4}$$

where $\mathcal{F}$ denotes a family of exponential distributions. It turns out that when $\mathcal{F}$ is the exponential family with sufficient statistics $\phi_1(\vartheta), \phi_2(\vartheta), \dots, \phi_S(\vartheta)$, then the solution of (4.4) is obtained from the moment matching condition, namely

$$\mathbb{E}_q[\phi_i(\vartheta)] = \mathbb{E}_p[\phi_i(\vartheta)], \quad i = 1, 2, \ldots S \tag{4.5}$$

where $\mathbb{E}_q[.]$ denotes expectation w.r.t. the distribution $q(\vartheta)$. In other words, in each step of the optimization we need to match the moments between $q(\vartheta)$ and $p(\vartheta|x_1, \ldots, x_n)$. For example if we choose $q(\vartheta)$ from the family of normal distributions, this is equivalent to equating the mean and variance of $q(\vartheta)$ and $p(\vartheta|x_1, \ldots, x_n)$. However, EP implements this process in a subtle way, in which instead of finding the best $q(\vartheta)$ at once, it finds the best factors of $q(\vartheta)$ one by one and refines them through successive iterations. At first, the algorithm starts by initializing all the factors $q_i(\vartheta)$ and consequently $q(\vartheta)$ itself [111].

4.4 Variable Level Expectation Propagation(VL-EP)

Consider more generally a data model with parameters $\boldsymbol{\theta}$. In our system model, $\boldsymbol{\theta} = [\mathbf{g}^T \mathbf{s}_d^T]^T$, where $\mathbf{g} = \text{vec}(\mathbf{G})$ and $\mathbf{s}_d = \text{vec}(\mathbf{S}_d)$, and $\mathbf{y} = \text{vec}(\mathbf{Y})$. We partition $\boldsymbol{\theta}$ into groups such that $\boldsymbol{\theta} = \{\boldsymbol{\theta}_i\}$ and we assume the prior factors at the level of these groups, which we call the variables in factor graph terminology. Then, the true posterior is given by

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{1}{Z}\, p(\mathbf{y}|\boldsymbol{\theta}) \prod_i p(\boldsymbol{\theta}_i) \tag{4.6}$$

where the factors on the RHS are called the factors in a factor graph and Z is a normalization factor given by

$$Z = p(\mathbf{y}) = \int_{\boldsymbol{\theta}} p(\mathbf{y}|\boldsymbol{\theta}) \prod_i p(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta} \tag{4.7}$$

The problem is the computation of the normalization factor Z involving a high-dimensional integral, therefore the mean and covariance of the posterior is not computationally tractable.

In many applications, $p(\mathbf{y}|\boldsymbol{\theta})$ corresponds to a noisy measurement of a signal $\mathbf{z}$. Very often, this signal is decomposed as a superposition of signals, $\mathbf{z} = \sum_k \mathbf{z}_k$. Each of the $\mathbf{z}_k$ is then parameterized by a subset $\boldsymbol{\theta}_k$ of $\boldsymbol{\theta}$. If one introduces these intermediate variables $\mathbf{z}_k$, as is typically done in the space-alternating generalized expectation-maximization (SAGE) algorithm, then we can decompose $p(\mathbf{y}|\boldsymbol{\theta}) \to p(\mathbf{y}|\mathbf{z}) \prod_k p(\mathbf{z}_k|\boldsymbol{\theta}_k)$.

However, here we propose to not introduce these variables $\mathbf{z}_k$.

In EP, as in several other MP and Variational Bayes variations, it is proposed to approximate the posterior distribution. In EP, the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$ is approximated by a distribution from the exponential family $\mathcal{F}$. Here we consider Gaussian pdfs. To this end, we propose the following approximation

$$\frac{1}{Z}\, p(\mathbf{y}|\boldsymbol{\theta}) \approx \prod_i m(\boldsymbol{\theta}_i) \quad \Rightarrow \quad p(\boldsymbol{\theta}|\mathbf{y}) \approx q(\boldsymbol{\theta}) = \prod_i q(\boldsymbol{\theta}_i) \tag{4.8}$$

where $q(\boldsymbol{\theta}) \in \mathcal{F}$ is the approximate posterior in factored form at the variable level, and $q(\boldsymbol{\theta}_i) = m(\boldsymbol{\theta}_i)\, p(\boldsymbol{\theta}_i)$ where the $m(\boldsymbol{\theta}_i)$ are the extrinsic pdfs. Additionally, we assume the prior pdfs $p(\boldsymbol{\theta}_i)$ to be simple (typically Gaussian or other members of the exponential family) so that they do not need approximation. Therefore, only the data pdf $p(\mathbf{y}|\boldsymbol{\theta})$ requires approximation. The EP approach adjusts the approximate posterior by minimizing a KL distance.

In the original EP algorithm [112], the approximate posterior factors get approximated alternatively at the factor level, with each factor being optimized completely. In the original EP, the approximate factors are in the exponential family, but not constrained any further. Hence the factors can involve possibly all variables. However, it is possible to introduce constraints in approximate pdfs (e.g. Gaussians with a block diagonal covariance). In the EP variation considered in [113], the approximate factors are also factorizable at variable level. However, here we propose to optimize the factors not at factor level but at variable level. Hence the name variable level EP (VL-EP), as opposed to the classical factor level EP (FL-EP), however the updating follows exactly the EP principle. We optimize a factor $m(\boldsymbol{\theta}_i)$ by minimizing the KL distance

$$\mathrm{KL}\Big(\tfrac{1}{Z}\, p(\mathbf{y}|\boldsymbol{\theta}) q(\boldsymbol{\theta}_{\bar{i}}) \;||\; m(\boldsymbol{\theta}_i)\, q(\boldsymbol{\theta}_{\bar{i}})\Big) = \frac{1}{Z} \int p(\mathbf{y}|\boldsymbol{\theta})\, q(\boldsymbol{\theta}_{\bar{i}}) \ln \frac{\frac{1}{Z}\, p(\mathbf{y}|\boldsymbol{\theta})\, q(\boldsymbol{\theta}_{\bar{i}})}{m(\boldsymbol{\theta}_i)\, q(\boldsymbol{\theta}_{\bar{i}})}\, d\boldsymbol{\theta} \tag{4.9}$$

w.r.t. a Gaussian $m(\boldsymbol{\theta}_i)$. Note that $q(\boldsymbol{\theta}) = q(\boldsymbol{\theta}_i)\, q(\boldsymbol{\theta}_{\bar{i}})$. The minimization of the KL distance leads to (see section 2 in [112] which exposes the original EP):

$$\begin{aligned} \widehat{p}(\boldsymbol{\theta}_i) &= \int q(\boldsymbol{\theta}_{\bar{i}})\, p(\mathbf{y}|\boldsymbol{\theta}) d\boldsymbol{\theta}_{\bar{i}} \\ Z_i &= \int \widehat{p}(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta}_i \\ \boldsymbol{\mu}_i &= \frac{1}{Z_i} \int \boldsymbol{\theta}_i\, \widehat{p}(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta}_i \\ \boldsymbol{\Sigma}_i &= \frac{1}{Z_i} \int \boldsymbol{\theta}_i\, \boldsymbol{\theta}_i^H\, \widehat{p}(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta}_i - \boldsymbol{\mu}_i \boldsymbol{\mu}_i^H \end{aligned} \tag{4.10}$$

where Z_i is a normalization constant, $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ are the mean and covariance of the Gaussian $m(\boldsymbol{\theta}_i)$. Note that $\widehat{p}(\boldsymbol{\theta}_i)$ integrates out all other variables and produces the (un-normalized) target pdf for $\boldsymbol{\theta}_i$ that we approximate by the Gaussian $m(\boldsymbol{\theta}_i)$. It is this integration which produces the cleaned $\mathbf{y}$, cleaned from the interference of other variables $\boldsymbol{\theta}_{\bar{i}}$. Actually, the proof of (4.10) is fairly straightforward. Since the KL distance in (4.9) needs to be minimized w.r.t. (the parameters of) $m(\boldsymbol{\theta}_i)$, we can write

$$\begin{aligned} \mathrm{KL} &= c^t - \frac{1}{Z_i} \int \ln(m(\boldsymbol{\theta}_i))\, \widehat{p}(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta}_i \\ &= c^t + n_{\boldsymbol{\theta}_i} \ln(\pi) + \ln\det(\boldsymbol{\Sigma}_i) + \frac{1}{Z_i} \mathrm{tr}\{\boldsymbol{\Sigma}_i^{-1} \int (\boldsymbol{\theta}_i - \boldsymbol{\mu}_i)(\boldsymbol{\theta}_i - \boldsymbol{\mu}_i)^H \widehat{p}(\boldsymbol{\theta}_i)\, d\boldsymbol{\theta}_i\} \\ &\geq c^t + n_{\boldsymbol{\theta}_i} \ln(\pi) + \ln\det(\boldsymbol{\Sigma}_i) + \frac{n_{\boldsymbol{\theta}_i}}{Z_i} \end{aligned} \tag{4.11}$$

where $n_{\boldsymbol{\theta}_i}$ denotes the dimension of $\boldsymbol{\theta}_i$. The minimization over $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_i$ leads to the solution in (4.10) and the minimal value in the last line in (4.11). As this minimal value is decreasing in Z_i, which itself is linear in $\widehat{p}(\boldsymbol{\theta}_i)$, we can majorize the KL distance by replacing $\widehat{p}(\boldsymbol{\theta}_i)$ by a minorizer and still retain a valid KL distance minimization strategy. We follow this strategy below when the moments of $\widehat{p}(\boldsymbol{\theta}_i)$ cannot be computed analytically.

4.5 VL-EP for Gaussian-Gaussian Semi-Blind

4.5.1 Channel VL-EP for GG-SB with Eliminated Inputs

For the application to the Gaussian inputs Gaussian channel semi-blind (GG-SB) channel estimation problem, we shall consider the problem formulation which eliminates the Gaussian $\mathbf{s}_d$. Then, we have the correspondence $\boldsymbol{\theta} = \mathbf{g}$, $\boldsymbol{\theta}_i = \mathbf{g}_i$. The FL-EP of [113] introduces the auxiliary hidden variables $\mathbf{z}_k = \mathbf{s}_{d,k} \otimes \mathbf{g}_k$, which we avoid here.

We have $\mathbf{Y} = [\mathbf{Y}_p \ \mathbf{Y}_d]$. We shall consider here the development of $\mathbf{Y}_p$ and $\mathbf{Y}_d$ separately. The alternating updating of the posterior factors loops over the K users, i.e., the update for user k updates the posterior factor for $\mathbf{g}_k$. Let us consider first of all the pilot part, and let user k use the n^{th} pilot, $k \in \mathcal{G}_n$ so that $\mathbf{s}_{p,k} = \mathbf{s}_p^{(n)}$. The pilot signal model for user k can be written as

$$\begin{aligned} \mathbf{Y}_p &= \sqrt{\rho}\,\mathbf{g}_k\,\mathbf{s}_{p,k}^T + \sqrt{\rho}\,\textstyle\sum_{i\neq k}\mathbf{g}_i\,\mathbf{s}_{p,i}^T + \mathbf{V}_p \\ \mathbf{Y}_p\,\mathbf{s}_p^{(n)*} &= \sqrt{\rho}\,P\,\mathbf{g}_k + \sqrt{\rho}\,P\sum_{i\in\mathcal{G}_n\setminus\{k\}}(\widehat{\mathbf{g}}_i + \widetilde{\mathbf{g}}_i) + \mathbf{V}_p\,\mathbf{s}_p^{(n)*} \end{aligned} \tag{4.12}$$

where $\mathbf{V} = [\mathbf{V}_p \ \mathbf{V}_d]$ and $\mathbf{g}_i$ has (approximate) posterior pdf $q(\mathbf{g}_i) \sim \mathcal{CN}(\widehat{\mathbf{g}}_i, \mathbf{C}_i)$ and $\mathbf{C}_i = \mathbb{E}\{\widetilde{\mathbf{g}}_i\,\widetilde{\mathbf{g}}_i^H\}$. All variables whose pdf appears in different factors in the approximate posterior are treated as independent. Hence, $\sqrt{\rho}\,P\,\mathbf{g}_k$ has a Gaussian pdf with mean $\mathbf{Y}_p\,\mathbf{s}_p^{(n)*} - \sqrt{\rho}\,P\sum_{i\in\mathcal{G}_n\setminus\{k\}}\widehat{\mathbf{g}}_i$ and covariance

$$\mathbf{C}_{\widetilde{\mathbf{Y}}_{p,k}} = P\,\mathbf{I}_M + \rho\,P^2\sum_{i\in\mathcal{G}_n\setminus\{k\}}\mathbf{C}_i\,. \tag{4.13}$$

The likelihood from the pilot part needs to be combined with the data likelihood, where $\mathbf{s}_{d,i}$ has prior pdf $p_{\mathbf{s}_{d,i}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{L-P})$. The signal for user k can be written as

$$\mathbf{Y}_d = \sqrt{\rho}\,\mathbf{g}_k\,\mathbf{s}_{d,k}^T + \widetilde{\mathbf{Y}}_{d,k} \tag{4.14}$$

$$\begin{aligned} \widetilde{\mathbf{Y}}_{d,k} &= \sqrt{\rho}\,\textstyle\sum_{i\neq k}\mathbf{g}_i\,\mathbf{s}_{d,i}^T + \mathbf{V}_d \\ &= \sqrt{\rho}\,\textstyle\sum_{i\neq k}(\widehat{\mathbf{g}}_i + \widetilde{\mathbf{g}}_i)\,\mathbf{s}_{d,i}^T + \mathbf{V}_d \end{aligned}$$

which has zero mean and covariance matrix $\mathbf{C}_{\mathbf{Y}_d} = \rho\,\mathbf{g}_k\mathbf{g}_k^H + \mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}$, where

$$\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}} = \mathbf{I}_M + \rho\sum_{i\neq k}(\widehat{\mathbf{g}}_i\,\widehat{\mathbf{g}}_i^H + \mathbf{C}_i)\,. \tag{4.15}$$

We shall model here $\widetilde{\mathbf{g}}_i\,\mathbf{s}_{d,i}^T$ also as Gaussian, just as an EP with variables $\mathbf{z}_i$ would do, since we need to go towards Gaussian approximations $q(\boldsymbol{\theta}_k)$ in any case. So we associate a Gaussian pdf to (4.14) by moment matching (whereas there are actually products of Gaussian variables). This leads to (apart from a constant)

$$-\ln p(\mathbf{Y}_d|\mathbf{g}_k, \mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}) = \|\mathbf{Y}_d\|^2_{(\rho\,\mathbf{g}_k\mathbf{g}_k^H + \mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}})^{-1}} + c^t \tag{4.16}$$

with the squared weighted Frobenius norm $\|\mathbf{Y}\|^2_{\mathbf{A}} = \mathrm{tr}\{\mathbf{A}\mathbf{Y}\mathbf{Y}^H\}$, and where the knowl-

edge of $\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}$ comprises the knowledge of $\widehat{\mathbf{g}}_{\overline{k}}$ and $\mathbf{C}_{\overline{k}}$. According to the majorization step in (3.32), we can majorize the negative log-likelihood of (4.15) or (4.16) by

$$\begin{aligned}
-\ln p(\mathbf{Y}_d|\mathbf{g}_k, \mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}) &= -\ln \mathbb{E}_{\mathbf{s}^p_{d,k}, \mathbf{G}^q_{\overline{k}}\mathbf{S}^p_{d,\overline{k}}|\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}} \{p(\mathbf{Y}_d|\mathbf{g}_k, \mathbf{g}_{\overline{k}}, \mathbf{s}_d)\} \\
&= -\ln \mathbb{E}_{\mathbf{s}^p_{d,k}|\mathbf{Y}_d, \widehat{\mathbf{g}}_k, \mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}} \mathbb{E}_{\mathbf{G}^q_{\overline{k}}\mathbf{S}^p_{d,\overline{k}}|\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}} \frac{p(\mathbf{Y}_d|\mathbf{g}_k, \mathbf{g}_{\overline{k}}, \mathbf{s}_d)\, p(\mathbf{s}_{d,k})}{p(\mathbf{s}_{d,k}|\mathbf{Y}_d, \widehat{\mathbf{g}}_k, \mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}})} \\
&= -\ln \mathbb{E}_{\mathbf{s}^p_{d,k}|\mathbf{Y}_d, \widehat{\mathbf{g}}_k, \mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}} \mathbb{E}_{\mathbf{G}^q_{\overline{k}}\mathbf{S}^p_{d,\overline{k}}|\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}} p(\mathbf{Y}_d|\mathbf{g}_k, \mathbf{g}_{\overline{k}}, \mathbf{s}_d) + c^t \\
&\leq \mathbb{E}_{\mathbf{s}^p_{d,k}|\mathbf{Y}_d, \widehat{\mathbf{g}}_k, \mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}} \{-\ln \mathbb{E}_{\mathbf{G}^q_{\overline{k}}\mathbf{S}^p_{d,\overline{k}}|\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}} p(\mathbf{Y}_d|\mathbf{g}_k, \mathbf{g}_{\overline{k}}, \mathbf{s}_d)\} + c^t \\
&= \mathbb{E}_{\mathbf{s}^p_{d,k}|\mathbf{Y}_d, \widehat{\mathbf{g}}_k, \mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}} \|\mathbf{Y}_d - \sqrt{\rho}\, \mathbf{g}_k\, \mathbf{s}^T_{d,k}\|^2_{\mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}}} + c^t \\
&= \|\mathbf{Y}_d - \sqrt{\rho}\, \mathbf{g}_k\, \widehat{\mathbf{s}}^T_{d,k}\|^2_{\mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}}} + \rho\, \mathrm{tr}\{\mathbf{R}_k\}\, \mathbf{g}^H_k \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}} \mathbf{g}_k + c^t \\
&= -\ln \widehat{p}(\mathbf{g}_k)
\end{aligned} \tag{4.17}$$

where

$$\begin{aligned}
&\widehat{\mathbf{s}}^T_{d,k} = \sqrt{\rho}\, (1 + \rho\, \widehat{\mathbf{g}}^H_k \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}} \widehat{\mathbf{g}}_k)^{-1} \widehat{\mathbf{g}}^H_k \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}} \mathbf{Y}_d \\
&\mathbf{R}_k = \mathbf{C}_{\widetilde{\mathbf{s}}_{d,k}\widetilde{\mathbf{s}}_{d,k}} = \sigma^2_{\widetilde{\mathbf{s}}_{d,k}}\, \mathbf{I}_{L-P}, \quad \sigma^2_{\widetilde{\mathbf{s}}_{d,k}} = (1 + \rho\, \widehat{\mathbf{g}}^H_k\, \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}}\, \widehat{\mathbf{g}}_k)^{-1}
\end{aligned} \tag{4.18}$$

are the LMMSE estimate and associated error covariance matrix, based on the current estimate $\widehat{\mathbf{g}}_k$. Note that $\mathbb{E}_{\mathbf{G}^q_{\overline{k}}\mathbf{S}^p_{d,\overline{k}}|\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}}$ in (4.17) uses a Gaussian distribution for $\mathbf{G}_{\overline{k}}\mathbf{S}_{d,\overline{k}}$ which is based on moment matching from $p(\mathbf{S}_{d,\overline{k}})$ and $q(\mathbf{G}_{\overline{k}})$ (as VL-EP requires). Note also that because of the Gaussian approximation of $\mathbf{G}_{\overline{k}}\mathbf{S}_{d,\overline{k}}$ and the EM majorization step, the target pdf $\widehat{p}(\mathbf{g}_k)$ is Gaussian. This Gaussian blind information pdf needs to be combined with the Gaussian pilot part and the Gaussian prior to yield

$$\begin{aligned}
-\ln q(\mathbf{g}_k) &= \|\mathbf{Y}_p \mathbf{s}^{(n)*}_p - \sqrt{\rho}\, P \textstyle\sum_{i\in\mathcal{G}_n\setminus\{k\}} \widehat{\mathbf{g}}_i - \sqrt{\rho}\, P\, \mathbf{g}_k\|^2_{\mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{p,k}}} + \|\mathbf{Y}_d - \sqrt{\rho}\, \mathbf{g}_k\, \widehat{\mathbf{s}}^T_{d,k}\|^2_{\mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}}} \\
&\quad + \rho\, (\|\widehat{\mathbf{s}}_{d,k}\|^2 + (L-P)\, \sigma^2_{\widetilde{\mathbf{s}}_{d,k}})\, \mathbf{g}^H_k \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}} \mathbf{g}_k + \mathbf{g}^H_k \mathbf{C}^{-1}_{\mathbf{g}_k\mathbf{g}_k} \mathbf{g}_k + c^t \\
&= -2\, \Re\{\widehat{\mathbf{g}}^H_k \mathbf{C}^{-1}_k \mathbf{g}_k\} + \mathbf{g}^H_k \mathbf{C}^{-1}_k \mathbf{g}_k + c^t
\end{aligned} \tag{4.19}$$

which is Gaussian with mean and covariance

$$\begin{aligned}
\widehat{\mathbf{g}}_k &= \sqrt{\rho}\, \mathbf{C}_k\, [P\, \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{p,k}} (\mathbf{Y}_p \mathbf{s}^{(n)*}_p - \sqrt{\rho}\, P \textstyle\sum_{i\in\mathcal{G}_n\setminus\{k\}} \widehat{\mathbf{g}}_i) + \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}} \mathbf{Y}_d\, \widehat{\mathbf{s}}^*_{d,k}] \\
\mathbf{C}_k &= \big(\rho\, [P^2\, \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{p,k}} + (\|\widehat{\mathbf{s}}_{d,k}\|^2 + (L-P)\, \sigma^2_{\widetilde{\mathbf{s}}_{d,k}})\, \mathbf{C}^{-1}_{\widetilde{\mathbf{Y}}_{d,k}}] + \mathbf{C}^{-1}_{\mathbf{g}_k\mathbf{g}_k}\big)^{-1}
\end{aligned} \tag{4.20}$$

where now $\mathbf{g}_k$ is the new estimate, and $\mathbf{C}_{\widetilde{\mathbf{Y}}_{p,k}}$, $\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}$ are defined in (4.13), (4.15).

This is an iterative procedure that cycles through the $\mathbf{g}_k,\ k = 1, \dots K$, and can be initialized with $\widehat{\mathbf{g}}_k^{(-1)} = \mathbf{0}$ or with the channel MAP estimate for $\widehat{\mathbf{g}}_k$ with associated

$$\mathbf{C}_k = \left(\rho\,[P\,\mathbf{I}_M + (L-P)\,\left(\mathbf{I}_M + \rho \sum_{i \neq k} \widehat{\mathbf{g}}_i\,\widehat{\mathbf{g}}_i^H)^{-1}\right] + \mathbf{C}_{\mathbf{g}_k \mathbf{g}_k}^{-1}\right)^{-1} \tag{4.21}$$

where we used $\|\mathbf{s}_{p,k}\|^2 = P$ and $\|\widehat{\mathbf{s}}_{d,k}\|^2 + (L-P)\,\sigma^2_{\widetilde{\mathbf{s}}_{d,k}}) \approx (L-P)\,\sigma^2_{\widetilde{\mathbf{s}}_{d,k}} = L-P$. Note that if parallel updating of the users is performed, one can reduce complexity in the computation of sums of the form $\sum_{i \neq k} A_i = \sum_{i=1}^{K} A_i - A_k$, so by computing a sum only once and then performing single term corrections.

The proposed channel estimation VL-EP for GG-SB is summarized in Algorithm 4. This channel VL-EP for GG-SB algorithm can be considered as an iterative version for the channel MAP algorithm (Algorithm 3) if one puts $\mathbf{C}_i = \mathbf{0}$ in $\mathbf{C}_{\widetilde{\mathbf{Y}}_{p,k}}$ and $\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}$ (alternatingly optimizing the $\mathbf{g}_k$ instead of trying to optimize w.r.t. all of $\mathbf{G}$ at once).

Algorithm 4 Iterative Channel Estimation VL-EP for GG-SB

Initialize $\widehat{\mathbf{g}}_k = \mathbf{0}_{M\times 1}$, $\mathbf{C}_k = \mathrm{diag}(\beta_{1k}, \dots, \beta_{Mk})$, $\widehat{\mathbf{s}}_{d,k} = \mathbf{0}_{(L-P)\times 1}$,
$\mathbf{C}_{\widetilde{\mathbf{Y}}_{p,k}} = P\,\mathbf{I}_M + \rho\,P^2 \sum_{i \in \mathcal{G}_n \setminus \{k\}} \mathbf{C}_i$, $\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}} = \mathbf{I}_M + \rho \sum_{i \neq k} \mathbf{C}_i$, for $k = 1, \dots K$.
repeat
  $t \leftarrow t + 1$
  for $k \leftarrow 1$ **to** K **do**
    compute $\sigma^2_{\widetilde{\mathbf{s}}_{d,k}}$ according to (4.18) and then compute $\mathbf{C}_k$ according to (4.20)
    compute $\widehat{\mathbf{g}}_k$ according to (4.20)
    compute $\widehat{\mathbf{s}}_{d,k}$ according to (4.18)
    compute $\mathbf{C}_{\widetilde{\mathbf{Y}}_{p,k}}$ and $\mathbf{C}_{\widetilde{\mathbf{Y}}_{d,k}}$ according to (4.13) and (4.15)
  end for
until convergence or $t = t_{\max}$
Return $\widehat{\mathbf{g}}_k$ for $k = 1, \dots K$

4.6 Simulation Results

In this section, we assess numerically the performance of the analytical derivations in this chapter. The M APs and K users are uniformly distributed at random over a square area of size $D \times D$. We adopt the same large-scale fading coefficients modeled in (3.52) and (3.53). Throughout this section, we consider the following scenario. The $M = 100$ APs and $K = 24$ users are uniformly distributed at random over a square

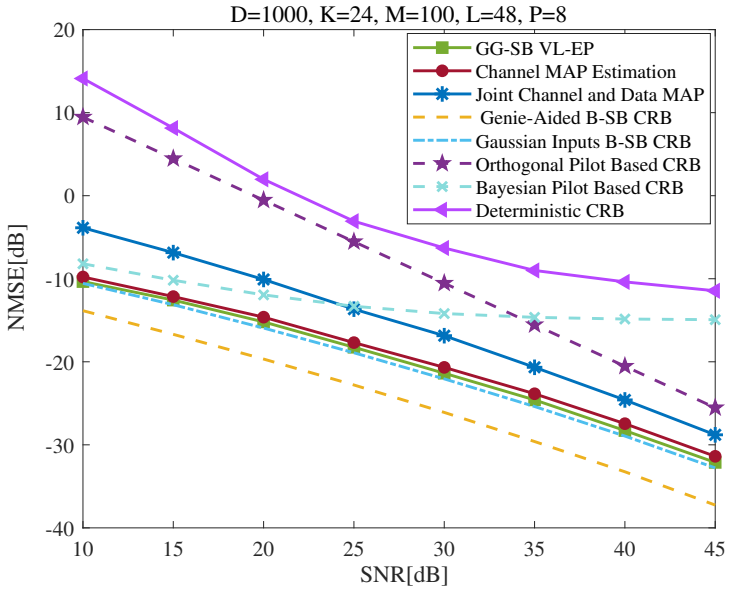

Figure 4.1: NMSE [dB] versus SNR [dB].

area of side $D = 1000$ and we consider $P = 8$ and $L = 48$.

The performance of the different channel estimators is assessed by NMSE versus SNR. The NMSE is defined as NMSE $= \dfrac{avg\|\mathbf{g} - \widehat{\mathbf{g}}\|^2}{avg\|\mathbf{g}\|^2}$ where avg stands for average. Fig. 4.1 compares the performance of the proposed channel estimation VL-EP for GG-SB and channel MAP estimation and presents NMSE [dB] versus SNR [dB]. The proposed algorithm outperforms the channel MAP estimation and the joint channel and data MAP algorithm, termed Bayesian semi-blind approach in the chapter 3. The joint channel and data MAP alternatingly estimates the channel or data as if the estimate for the other quantity is perfect, whereas the channel MAP estimation takes into account the data error covariance matrix. Therefore the channel MAP estimation outperforms the Bayesian semi-blind iterative algorithm.

The performance of these three different semi-blind channel estimation algorithms is compared to the different CRBs. For the semi-blind approaches one can consider the genie-aided scenario in which the data would be detected exactly, hence becoming also pilots for the channel estimation, leading to the genie-aided Bayesian semi-blind (B-SB) CRB. For our VL-EP or channel MAP scenario, we consider Gaussian channels with the Gaussian input symbols eliminated, leading to the Gaussian inputs B-SB CRB. The

deterministic CRB curve in the figure corresponds to a deterministic framework introduced in the chapter 3 in which both data signal and channel coefficients are modeled as unknown deterministic quantities. The performance of the different CRBs is evaluated by NMSE = $\mathrm{tr}\{\mathrm{CRB}\}/\mathrm{tr}\{\mathbf{C_{gg}}\}$, where $\mathbf{C_{gg}} = \mathrm{diag}(\beta_{11}, \ldots, \beta_{M1} \ldots \beta_{1K}, \ldots, \beta_{MK})$. The simulations show that exploiting prior information gives significant performance gains. Compared to a fictitious scenario of just orthogonal pilot based channel estimation (pilots still of length P), deterministic semi-blind does not do as well whereas Bayesian semi-blind still does much better. On the other hand, the Bayesian pilot based CRB shows that just adding channel prior information to the contaminating pilots allows already to significantly improve MSE at low to moderate SNR, but floors at higher SNR. Adding the blind channel information from the data second-order statistics breaks this flooring, and both channel MAP and especially VL-EP allow to get performance close to the corresponding CRB, which behaves with just an SNR offset compared to the genie aided CRB.

4.7 Conclusion

In this chapter, we considered semi-blind methods for channel estimation in the presence of Gaussian i.i.d. data, exploiting prior channel information to mitigate the pilot contamination which originates from reusing pilot sequences, in CF massive MIMO systems. We proposed a VL-EP algorithm for semi-blind channel estimation which provides an approximate MMSE channel estimator. Numerical simulations corroborated the analytical derivations and the proposed VL-EP algorithm.

Chapter 5

Conclusions and Future work

In this dissertation, we have mainly focused on providing insights and fundamental understanding of critical, unknown aspects of DASs, exploring the fundamental limits and potential of this architecture, studying fundamental metrics playing a key role in the system behavior, namely system parameters, such as the system load, i.e, the ratio between number of users and antennas, their geographical distributions, concentration of antennas at each wireless access port, and physical parameters, such as exponent loss, analyzing the performance of DASs through the channel eigenvalue spectrum and moments. Special attention has been paid to study and analyze the performance of a special class of DASs, the CF Massive MIMO system, has been attracting a wide interest recently.

More precisely, in chapter 2, we considered a distributed antenna system in uplink, comprising a massive number of distributed transmit and receive antennas. In our DAS, transmit and receive antennas are distributed according to homogeneous PP and the received signals are processed jointly at a CPU. In centralized massive MIMO systems, the phenomenon of favorable propagation has been observed: when the number of receive antennas tends to infinity while the number of transmit antennas remains finite, the users' channels become almost orthogonal and low complexity detection via matched filtering is almost optimal. We analyzed the properties of DASs in asymptotic conditions when the network dimensions go to infinity with given intensities of

the transmit and receive antenna PPs. We studied the analytical conditions of favorable propagation in CF massive MIMO systems with two kinds of channels, namely, channels with path loss and transmit and receive antennas in LoS or in multipath Rayleigh fading. We showed that the analytical conditions of favorable propagation are satisfied for channels impaired by path loss and Rayleigh fading while they do not hold in the case of LoS channels, motivating the use and analysis of multi-stage receivers. We validated the asymptotic analytical results by simulation results of the favorable propagation conditions and the performance of multi-stage detectors for finite systems. In future works, we will study how to extend the analytical results of favorable propagation to multiple-antenna APs scenarios.

In chapter 3, we addressed the problem of pilot contamination in CF massive MIMO systems. We exploited the channel sparsity to tackle pilot contamination, which originates from the reuse of pilot sequences. Specifically, we considered semi-blind methods for joint channel estimation and data detection. Under the challenging assumption of deterministic parameters, we determined sufficient and necessary conditions for semi-blind identifiability, which guarantee the non-singularity of the FIM and the existence of the CRB. We proposed an MP algorithm which determines the exact channel coefficients in the case of semi-blind identifiability. We proposed a semi-blind joint channel and data MAP estimator alternating between channel estimation and linear multi-user detection. Additionally, we proposed a semi-blind MAP estimation of channel in presence of Gaussian i.i.d. data. We also derived FIM and CRBs under different assumptions to evaluate the performance of semi-blind approaches introduced in this chapter. In this chapter, as already mentioned, we determined the identifiability conditions under the assumption of deterministic parameters, hence developing the identifiability conditions in a Bayesian framework could be carried out in the future.

In chapter 4, we considered semi-blind methods for channel estimation based on treating the unknown symbols as random with known prior distribution to be eliminated, to resolve the pilot contamination originating from the reuse of pilot sequences, in CF massive MIMO systems. This task is further aided by exploiting prior channel

information in a Bayesian formulation. We proposed a VL-EP algorithm for MP style semi-blind channel estimation which provides an approximate MMSE channel estimator which itself can not be found analytically. In future works, we will study how to construct approximate posteriors for both channels and data symbols for semi-blind joint channel estimation and data detection approaches based on the EP algorithm. Moreover, the distributed version of the techniques introduced in this dissertation could be explored in order to further reduce the computational complexity of receivers in CF massive MIMO systems.

Appendices

Appendix A

Appendices of Chapter 2

A.1 Proof of Algorithm 1

Let us adopt the following notation

- $\boldsymbol{\xi}_{Tk}$ and $\boldsymbol{\Phi}_{Tk}$ denote the k-th row and column of the matrix $\boldsymbol{\Phi}_T$, respectively.
- $\boldsymbol{\xi}_{Rk}$ and $\boldsymbol{\Phi}_{Rk}$ denotes the k-th row and column of the matrix $\boldsymbol{\Phi}_R$, respectively.
- $\boldsymbol{\Xi}_{T\sim k}$ is the $(N_T-1)\times\theta^2$ matrix obtained from the matrix $\boldsymbol{\Phi}_T$ by removing the k-th row.
- $\boldsymbol{\Xi}_{R\sim k}$ is the $(N_R-1)\times\theta^2$ matrix obtained from the matrix $\boldsymbol{\Phi}_R$ by removing the k-th row.
- $\boldsymbol{\Phi}_{T\sim k}$ is the $N_T\times(\theta^2-1)$ matrix obtained from the matrix $\boldsymbol{\Phi}_T$ by removing the k-th column.
- $\boldsymbol{\Phi}_{R\sim k}$ is the $N_R\times(\theta^2-1)$ matrix obtained from the matrix $\boldsymbol{\Phi}_R$ by removing the k-th column.
- $\mathbf{T}_{\sim k}$ is the $(\theta^2-1)\times(\theta^2-1)$ matrix obtained from the matrix $\mathbf{T}$ by removing the k-th row and k-th column.
- $\mathbf{B}_{kk}$ denotes k-th diagonal element of the square matrix $\mathbf{B}$.

For any k and $\ell \geq 2$, k-th diagonal element of the channel covariance matrix

$\widetilde{\mathbf{C}}^{\ell} = (\widetilde{\mathbf{G}}^{H}\widetilde{\mathbf{G}})^{\ell} = \mathbf{\Phi}_T\mathbf{T}\,\mathbf{\Phi}_R^H\Big(\mathbf{\Phi}_R\mathbf{T}\,\mathbf{\Phi}_T^H\mathbf{\Phi}_T\mathbf{T}\,\mathbf{\Phi}_R^H\Big)^{\ell-1}\mathbf{\Phi}_R\mathbf{T}\,\mathbf{\Phi}_T^H$ of size $N_T \times N_T$ is given by

$$\widetilde{\mathbf{C}}_{kk}^{(\ell)} = \boldsymbol{\xi}_{Tk}\mathbf{T}\mathbf{\Phi}_R^H\,\mathbf{D}^{\ell-1}\,\mathbf{\Phi}_R\mathbf{T}\boldsymbol{\xi}_{Tk}^H \tag{A.1}$$

where $\mathbf{D}$ denotes an $N_R \times N_R$ matrix defined as follows,

$$\mathbf{D} = \widetilde{\mathbf{G}}\widetilde{\mathbf{G}}^{\mathbf{H}} = \mathbf{\Phi}_R\mathbf{T}\mathbf{\Phi}_T^H\mathbf{\Phi}_T\mathbf{T}\mathbf{\Phi}_R^H \tag{A.2}$$

The matrix $\mathbf{\Phi}_T^H\mathbf{\Phi}_T$ can be written as follows

$$\mathbf{\Phi}_T^H\mathbf{\Phi}_T = \boldsymbol{\xi}_{Tk}^H\boldsymbol{\xi}_{Tk} + \mathbf{\Xi}_{T\sim k}^H\mathbf{\Xi}_{T\sim k} \tag{A.3}$$

Making use of (A.3), the k-th diagonal element of matrix $\widetilde{\mathbf{C}}^{\ell}$ in (A.1) can be written as

$$\begin{aligned}\widetilde{\mathbf{C}}_{kk}^{(\ell)} &= \boldsymbol{\xi}_{Tk}\mathbf{T}\mathbf{\Phi}_R^H\,\Big(\mathbf{\Phi}_R\mathbf{T}(\boldsymbol{\xi}_{Tk}^H\boldsymbol{\xi}_{Tk} + \mathbf{\Xi}_{T\sim k}^H\mathbf{\Xi}_{T\sim k})\mathbf{T}\mathbf{\Phi}_R^H\Big)\mathbf{D}^{\ell-2}\,\mathbf{\Phi}_R\mathbf{T}\boldsymbol{\xi}_{Tk}^H \\ &= \boldsymbol{\xi}_{Tk}\mathbf{T}\mathbf{\Phi}_R^H\,\mathbf{\Phi}_R\mathbf{T}\boldsymbol{\xi}_{Tk}^H\boldsymbol{\xi}_{Tk}\mathbf{T}\mathbf{\Phi}_R^H\mathbf{D}^{\ell-2}\,\mathbf{\Phi}_R\mathbf{T}\boldsymbol{\xi}_{Tk}^H + \\ &\qquad \boldsymbol{\xi}_{Tk}\mathbf{T}\mathbf{\Phi}_R^H\,\mathbf{\Phi}_R\mathbf{T}\mathbf{\Xi}_{T\sim k}^H\mathbf{\Xi}_{T\sim k}\mathbf{T}\mathbf{\Phi}_R^H\mathbf{D}^{\ell-2}\,\mathbf{\Phi}_R\mathbf{T}\boldsymbol{\xi}_{Tk}^H\end{aligned} \tag{A.4}$$

Expanding the product, we can rewrite the first term of (A.4) as follows

$$\boldsymbol{\xi}_{Tk}\mathbf{T}\mathbf{\Phi}_R^H\,\mathbf{\Phi}_R\mathbf{T}\boldsymbol{\xi}_{Tk}^H\boldsymbol{\xi}_{Tk}\mathbf{T}\mathbf{\Phi}_R^H\mathbf{D}^{\ell-2}\,\mathbf{\Phi}_R\mathbf{T}\boldsymbol{\xi}_{Tk}^H = \boldsymbol{\xi}_{Tk}\mathbf{\Gamma}^{(1)}\boldsymbol{\xi}_{Tk}^H\tilde{\mathbf{C}}_{kk}^{(\ell-1)} \tag{A.5}$$

where $\mathbf{\Gamma}^{(\ell)} = \mathbf{T}\mathbf{\Phi}_R^H\,\mathbf{D}^{\ell-1}\,\mathbf{\Phi}_R\mathbf{T}$ is a matrix of size $\theta^2 \times \theta^2$. The second term in (A.4) can be further decomposed as

$$\begin{aligned}\boldsymbol{\xi}_{Tk}\mathbf{T}\mathbf{\Phi}_R^H\,\mathbf{\Phi}_R\mathbf{T}\mathbf{\Xi}_{T\sim k}^H\mathbf{\Xi}_{T\sim k}\mathbf{T}\mathbf{\Phi}_R^H\mathbf{D}^{\ell-2}\,\mathbf{\Phi}_R\mathbf{T}\boldsymbol{\xi}_{Tk}^H &= \boldsymbol{\xi}_{Tk}\,\mathbf{\Gamma}_{\sim k}^{(2)}\,\boldsymbol{\xi}_{Tk}^H\widetilde{\mathbf{C}}_{kk}^{(\ell-2)} + \\ &\qquad \boldsymbol{\xi}_{Tk}\mathbf{T}\mathbf{\Phi}_R^H\mathbf{D}_{\sim k}^2\,\mathbf{D}^{\ell-3}\mathbf{\Phi}_R\mathbf{T}\boldsymbol{\xi}_{Tk}^H\end{aligned} \tag{A.6}$$

where the matrices $\mathbf{D}_{\sim k}$ and $\mathbf{\Gamma}_{\sim k}^{(\ell)}$ are defined as follows

$$\begin{aligned}\mathbf{D}_{\sim k} &= \mathbf{\Phi}_R\mathbf{T}\,\mathbf{\Xi}_{T\sim k}^H\mathbf{\Xi}_{T\sim k}\,\mathbf{T}\mathbf{\Phi}_R^H \\ \mathbf{\Gamma}_{\sim k}^{(\ell)} &= \mathbf{T}\,\mathbf{\Phi}_R^H\mathbf{D}_{\sim k}^{\ell-1}\mathbf{\Phi}_R\,\mathbf{T}\end{aligned}$$

Iterating the expansion (A.6), we get

$$\widetilde{\mathbf{C}}_{kk}^{(\ell)} = \sum_{n=0}^{\ell-1} \boldsymbol{\xi}_{Tk}\, \boldsymbol{\Gamma}_{\sim k}^{(\ell-n)}\, \boldsymbol{\xi}_{Tk}^{H} \widetilde{\mathbf{C}}_{kk}^{(n)} \tag{A.7}$$

The almost sure convergence as $\theta^2 \to \infty$

$$\boldsymbol{\xi}_{Tk}\, \boldsymbol{\Gamma}_{\sim k}^{(\ell-n)}\, \boldsymbol{\xi}_{Tk}^{H} \xrightarrow{a.s.} m_{\boldsymbol{\Gamma}}^{(\ell-n)} \tag{A.8}$$

follows along the same lines as the proof of Lemma 4.1 in [114]. The eigenvalue moment of order ℓ, of the channel covariance matrix $\widetilde{\mathbf{C}}$ converges to a deterministic value given by

$$\begin{aligned} m_{\widetilde{\mathbf{C}}}^{(\ell)} &= \mathbb{E}\big\{\frac{1}{N_T}\mathrm{tr}(\widetilde{\mathbf{C}}^{\ell})\big\} \\ &= \mathbb{E}\big\{\frac{1}{N_T}\sum_{k=1}^{N_T}\widetilde{\mathbf{C}}_{kk}^{(\ell)}\big\} = \sum_{n=0}^{\ell-1} m_{\boldsymbol{\Gamma}}^{(\ell-n)} m_{\widetilde{\mathbf{C}}}^{(n)} \end{aligned} \tag{A.9}$$

Now, we need to compute the eigenvalue moments of the matrix $\boldsymbol{\Gamma}$ in (A.9). To this end, let us write the matrix $\mathbf{D}$ as follows

$$\mathbf{D} = \big(\boldsymbol{\Phi}_{Rk}\mathbf{T}_{kk}\boldsymbol{\Phi}_{Tk}^{H} + \mathbf{A}\big)\big(\boldsymbol{\Phi}_{Tk}\mathbf{T}_{kk}\boldsymbol{\Phi}_{Rk}^{H} + \mathbf{A}^{H}\big) \tag{A.10}$$

where $\mathbf{A} = \boldsymbol{\Phi}_{R\sim k}\, \mathbf{T}_{\sim k}\, \boldsymbol{\Phi}_{T\sim k}^{H}$ is an $N_R \times N_T$ matrix and $\mathbf{T}_{kk}$ denotes the k-th diagonal element of matrix $\mathbf{T}$. Making use of (A.10), the k-th diagonal element of the matrix $\boldsymbol{\Gamma}^{(\ell)}$ is given by

$$\begin{aligned} \boldsymbol{\Gamma}_{kk}^{(\ell)} &= \mathbf{T}_{kk}^{2}\, \boldsymbol{\Phi}_{Rk}^{H}\, \mathbf{D}^{\ell-1}\, \boldsymbol{\Phi}_{Rk} \\ &= \mathbf{T}_{kk}^{2}\, \boldsymbol{\Phi}_{Rk}^{H}\, \big(\boldsymbol{\Phi}_{Rk}\mathbf{T}_{kk}\boldsymbol{\Phi}_{Tk}^{H} + \mathbf{A}\big)\big(\boldsymbol{\Phi}_{Tk}\mathbf{T}_{kk}\boldsymbol{\Phi}_{Rk}^{H} + \mathbf{A}^{H}\big)\mathbf{D}^{\ell-2}\boldsymbol{\Phi}_{Rk} \end{aligned} \tag{A.11}$$

Expanding the product, we can rewrite the first term as

$$\mathbf{T}_{kk}^{4}\, \boldsymbol{\Phi}_{Rk}^{H}\boldsymbol{\Phi}_{Rk}\, \boldsymbol{\Phi}_{Tk}^{H}\boldsymbol{\Phi}_{Tk}\boldsymbol{\Phi}_{Rk}^{H}\, \mathbf{D}^{\ell-2}\, \boldsymbol{\Phi}_{Rk} = \mathbf{T}_{kk}^{2}\, \boldsymbol{\Phi}_{Rk}^{H}\boldsymbol{\Phi}_{Rk}\, \boldsymbol{\Phi}_{Tk}^{H}\boldsymbol{\Phi}_{Tk}\boldsymbol{\Gamma}_{kk}^{(\ell-1)}$$

The third term in (A.11), i.e., $\boldsymbol{\Phi}_{Rk}^{H}\, \mathbf{A}\, \boldsymbol{\Phi}_{Tk}$, goes to zero because the random matrix $\mathbf{A}$ is independent of both random vectors $\boldsymbol{\Phi}_{Rk}$ and $\boldsymbol{\Phi}_{Tk}$. The rest two terms in (A.11) can be further decomposed as follows

$$\begin{aligned}
&\mathbf{T}_{kk}^3\,\mathbf{\Phi}_{Rk}^H\mathbf{\Phi}_{Rk}\mathbf{\Phi}_{Tk}^H\,\mathbf{A}^H\,\mathbf{D}^{\ell-2}\mathbf{\Phi}_{Rk}+\mathbf{T}_{kk}^2\,\mathbf{\Phi}_{Rk}^H\,(\mathbf{A}\mathbf{A}^H)\,\mathbf{D}^{\ell-2}\mathbf{\Phi}_{Rk}\\
&\quad=\mathbf{T}_{kk}^2\,\mathbf{\Phi}_{Rk}^H\mathbf{\Phi}_{Rk}\,\mathbf{\Phi}_{Tk}^H(\mathbf{A}^H\mathbf{A})\mathbf{\Phi}_{Tk}\mathbf{\Gamma}_{kk}^{(\ell-2)}+\mathbf{T}_{kk}^2\,\mathbf{\Phi}_{Rk}^H(\mathbf{A}\mathbf{A}^H)\mathbf{\Phi}_{Rk}\,\mathbf{\Phi}_{Tk}^H\mathbf{\Phi}_{Tk}\mathbf{\Gamma}_{kk}^{(\ell-2)}+\\
&\quad\mathbf{T}_{kk}^3\,\mathbf{\Phi}_{Rk}^H\mathbf{\Phi}_{Rk}\,\mathbf{\Phi}_{Tk}^H(\mathbf{A}^H\mathbf{A})\mathbf{A}^H\mathbf{D}^{\ell-3}\mathbf{\Phi}_{Rk}+\mathbf{T}_{kk}^3\,\mathbf{\Phi}_{Rk}^H(\mathbf{A}\mathbf{A}^H)\mathbf{\Phi}_{Rk}\,\mathbf{\Phi}_{Tk}^H\mathbf{A}^H\mathbf{D}^{\ell-3}\mathbf{\Phi}_{Rk}+\\
&\quad\mathbf{T}_{kk}^2\mathbf{\Phi}_{Rk}^H(\mathbf{A}\mathbf{A}^H)^2\,\mathbf{D}^{\ell-3}\mathbf{\Phi}_{Rk}
\end{aligned}\tag{A.12}$$

The last three terms in (A.12) can be further decomposed. Iterating the expansion, we get the following expression

$$\begin{aligned}
\mathbf{\Gamma}_{kk}^{(\ell)}=\sum_{s=0}^{\ell-2}\;\sum_{r=0}^{\ell-2-s}\mathbf{T}_{kk}^2\,\mathbf{\Phi}_{Rk}^H(\mathbf{A}\mathbf{A}^H)^r\mathbf{\Phi}_{Rk}\,\mathbf{\Phi}_{Tk}^H(\mathbf{A}^H\mathbf{A})^s\mathbf{\Phi}_{Tk}\mathbf{\Gamma}_{kk}^{(\ell-(s+r)-1)}\\
+\mathbf{T}_{kk}^2\,\mathbf{\Phi}_{Rk}^H\,(\mathbf{A}\mathbf{A}^H)^{\ell-1}\mathbf{\Phi}_{Rk}
\end{aligned}$$

The almost sure convergence as $\theta^2\longrightarrow\infty$

$$\mathbf{\Phi}_{Rk}^H(\mathbf{A}\mathbf{A}^H)^r\mathbf{\Phi}_{Rk}\longrightarrow\beta_R\,m_{\mathbf{A}\mathbf{A}^H}^{(r)}\xrightarrow{a.s.}\beta_R\,m_{\mathbf{D}}^{(r)}\tag{A.13}$$

$$\mathbf{\Phi}_{Tk}^H(\mathbf{A}^H\mathbf{A})^s\mathbf{\Phi}_{Tk}\longrightarrow\beta_T\,m_{\mathbf{A}^H\mathbf{A}}^{(s)}\xrightarrow{a.s.}\beta_T\,m_{\widetilde{\mathbf{C}}}^{(s)}\tag{A.14}$$

where the moments $m_{\mathbf{A}\mathbf{A}^H}^{(r)}$ and $m_{\mathbf{A}^H\mathbf{A}}^{(s)}$ are approximated by $m_{\mathbf{D}}^{(r)}$ and $m_{\widetilde{\mathbf{C}}}^{(s)}$ respectively, therefore k-th diagonal element of matrix $\mathbf{\Gamma}^{(\ell)}$ for any $\ell\geq 2$ is given by

$$\mathbf{\Gamma}_{kk}^{(\ell)}=\beta_R\beta_T\mathbf{T}_{kk}^2\sum_{s=0}^{\ell-2}\;\sum_{r=0}^{\ell-2-s}m_{\widetilde{\mathbf{C}}}^{(s)}m_{\mathbf{D}}^{(r)}\mathbf{\Gamma}_{kk}^{(\ell-(s+r)-1)}+\beta_R m_{\mathbf{D}}^{(\ell-1)}\mathbf{T}_{kk}^2\tag{A.15}$$

and the eigenvalue moment of order ℓ of the matrix $\mathbf{\Gamma}$ is given by

$$\begin{aligned}
m_{\mathbf{\Gamma}}^{(\ell)}&=\mathbb{E}\big\{\frac{1}{\theta^2}\mathrm{tr}(\mathbf{\Gamma}^{(\ell)})\big\}\\
&=\beta_R\beta_T\sum_{s=0}^{\ell-2}\;\sum_{r=0}^{\ell-2-s}m_{\widetilde{\mathbf{C}}}^{(s)}m_{\mathbf{D}}^{(r)}\,\mathbb{E}(\frac{\mathbf{T}_{kk}^2}{\theta^2}\mathbf{\Gamma}_{kk}^{(\ell-(s+r)-1)})+\beta_R m_{\mathbf{D}}^{(\ell-1)}m_{\mathbf{T}}^{(2)}
\end{aligned}\tag{A.16}$$

In order to compute the eigenvalue moments of the matrix $\mathbf{D}$, we define a $\theta^2\times\theta^2$ matrix $\mathbf{\Delta}^{(\ell)}=\mathbf{T}\mathbf{\Phi}_T^H\,\widetilde{\mathbf{C}}^{\ell-1}\,\mathbf{\Phi}_T\mathbf{T}$. For any $\ell\geq 2$, k-th diagonal element of the matrix $\mathbf{D}^\ell$ of size $N_R\times N_R$ is given by

$$\mathbf{D}_{kk}^{(\ell)}=\boldsymbol{\xi}_{Rk}\mathbf{T}\mathbf{\Phi}_T^H\,\widetilde{\mathbf{C}}^{\ell-1}\,\mathbf{\Phi}_T\mathbf{T}\boldsymbol{\xi}_{Rk}^H\tag{A.17}$$

following the similar approach, we get

$$\mathbf{D}_{kk}^{(\ell)} = \sum_{n=0}^{\ell-1} \boldsymbol{\xi}_{Rk} \boldsymbol{\Delta}_{\sim k}^{(\ell-n)} \, \boldsymbol{\xi}_{Rk}^{H} \, \mathbf{D}_{kk}^{n} \tag{A.18}$$

where the matrix $\boldsymbol{\Delta}_{\sim k}^{(\ell)}$ is defined as follows

$$\boldsymbol{\Delta}_{\sim k}^{(\ell)} = \mathbf{T} \, \boldsymbol{\Phi}_{T}^{H} \, \widetilde{\mathbf{C}}_{\sim k}^{\ell-1} \boldsymbol{\Phi}_{T} \, \mathbf{T} \tag{A.19}$$

$$\widetilde{\mathbf{C}}_{\sim k} = \boldsymbol{\Phi}_{T} \mathbf{T} \boldsymbol{\Xi}_{R\sim k}^{H} \boldsymbol{\Xi}_{R\sim k} \mathbf{T} \boldsymbol{\Phi}_{T}^{H} \tag{A.20}$$

The almost sure convergence as $\theta^2 \to \infty$

$$\boldsymbol{\xi}_{Rk} \, \boldsymbol{\Delta}_{\sim k}^{(\ell-n)} \, \boldsymbol{\xi}_{Rk}^{H} \xrightarrow{a.s.} m_{\boldsymbol{\Delta}}^{(\ell-n)} \tag{A.21}$$

The eigenvalue moment of order ℓ of the matrix $\mathbf{D}$ is given by

$$\begin{aligned} m_{\mathbf{D}}^{(\ell)} &= \mathbb{E}\big\{ \frac{1}{N_R} \mathrm{tr}(\mathbf{D}^{\ell}) \big\} \\ &= \mathbb{E}\big\{ \frac{1}{N_R} \sum_{k=1}^{N_R} \mathbf{D}_{kk}^{(\ell)} \big\} = \sum_{n=0}^{\ell-1} m_{\boldsymbol{\Delta}}^{(\ell-n)} m_{\mathbf{D}}^{(n)} \end{aligned} \tag{A.22}$$

In order to compute the eigenvalue moments of matrix $\boldsymbol{\Delta}$, Similar to (A.10) the matrix $\widetilde{\mathbf{C}}$ can be written as follows

$$\widetilde{\mathbf{C}} = \big(\boldsymbol{\Phi}_{Tk} \mathbf{T}_{kk} \boldsymbol{\Phi}_{Rk}^{H} + \mathbf{A}^{H} \big) \big(\boldsymbol{\Phi}_{Rk} \mathbf{T}_{kk} \boldsymbol{\Phi}_{Tk}^{H} + \mathbf{A} \big) \tag{A.23}$$

following the similar approach, the k-th diagonal element of matrix $\boldsymbol{\Delta}^{(\ell)}$ for any $\ell \geq 2$ is given by

$$\boldsymbol{\Delta}_{kk}^{(\ell)} = \beta_R \beta_T \mathbf{T}_{kk}^{2} \sum_{s=0}^{\ell-2} \sum_{r=0}^{\ell-2-s} m_{\widetilde{\mathbf{C}}}^{(s)} m_{\mathbf{D}}^{(r)} \boldsymbol{\Delta}_{kk}^{(\ell-(s+r)-1)} + \beta_T m_{\widetilde{\mathbf{C}}}^{(\ell-1)} \mathbf{T}_{kk}^{2} \tag{A.24}$$

and ℓ-th eigenvalue moment of matrix $\boldsymbol{\Delta}$ is given by

$$\begin{aligned} m_{\boldsymbol{\Delta}}^{(\ell)} &= \mathbb{E}\big\{ \frac{1}{\theta^2} \mathrm{tr}(\boldsymbol{\Delta}^{(\ell)}) \big\} \\ &= \beta_R \beta_T \sum_{s=0}^{\ell-2} \sum_{r=0}^{\ell-2-s} m_{\widetilde{\mathbf{C}}}^{(s)} m_{\mathbf{D}}^{(r)} \, \mathbb{E}\big(\frac{\mathbf{T}_{kk}^{2}}{\theta^2} \boldsymbol{\Delta}_{kk}^{(\ell-(s+r)-1)} \big) + \beta_T m_{\widetilde{\mathbf{C}}}^{(\ell-1)} m_{\mathbf{T}}^{(2)} \end{aligned} \tag{A.25}$$

Making use of the relation $m_{\mathbf{D}}^{(\ell)} = \dfrac{\beta_T}{\beta_R} \, m_{\widetilde{\mathbf{C}}}^{(\ell)}$, the expressions (A.16), and (A.25), the

recursion yields $m_{\widetilde{\mathbf{C}}}^{(\ell)}$ and the Algorithm 1.

In order to compute $m_{\widetilde{\mathbf{C}}}^{(\ell)}$, it is necessary to determine $m_{\mathbf{\Gamma}}^{(\ell)}$ and $m_{\mathbf{\Delta}}^{(\ell-1)}$. It is easy to verify that the diagonal elements $\widetilde{C}_{kk}^{(\ell)}$ and $D_{kk}^{(\ell)}$ are independent of the index k and all equal.

A.2 Proof of Rayleigh Fading Eigenvalue Moments

The eigenvalue moment of order ℓ of the matrix $\widetilde{\mathbf{C}}$ in Rayleigh fading channel is given by

$$m_{\widetilde{\mathbf{C}}}^{(\ell)} = \frac{1}{N_T} \sum_{j_1,\dots j_\ell=1}^{N_T} \sum_{i_1,\dots i_\ell=1}^{N_R} \mathbb{E}\{\breve{g}_{i_1 j_1}^* h_{i_1 j_1}^* \dots \breve{g}_{i_\ell j_\ell}^* h_{i_\ell j_\ell}^* \breve{g}_{i_\ell j_1} h_{i_\ell j_1}\} \tag{A.26}$$

$$= \frac{1}{N_T} \sum_{j_1,\dots j_\ell=1}^{N_T} \sum_{i_1,\dots i_\ell=1}^{N_R} \mathbb{E}\{h_{i_1 j_1}^* h_{i_1 j_2} \dots h_{i_\ell j_\ell}^* h_{i_\ell j_1}\} \times \mathbb{E}\{\breve{g}_{i_1 j_1}^* \breve{g}_{i_1 j_2} \dots \breve{g}_{i_\ell j_\ell}^* \breve{g}_{i_\ell j_1}\}$$

Observe that the contribution of the terms with indices $(j_1, i_1, j_2, i_2, \dots, j_\ell, i_\ell)$ which do not correspond to even graphs is zero. In fact, assume that there is an edge between the pair of indices (i, j) from $\mathcal{I}$ to $\mathcal{J}$ with multiplicity t and an edge between the same pair of indices in the opposite direction with multiplicity $t^{'} \neq t$. Then, these terms contain the factor $\mathbb{E}\{h_{ij}^t \; h_{ij}^{*t'}\} = 0$ and do not contribute to the eigenvalue moment [1]. Thus, we can restrict the sum in (A.26) to even sequences of indices. We consider now the contributions of even sequences with p_1 and p_2 distinct indices in $\mathcal{I}$ and $\mathcal{J}$, respectively, with $p_1 + p_2 = r$. Since the graph is connected it is straightforward to recognize that $r \leq \ell + 1$. The contribution of the terms with $r = \ell + 1$ has been derived in subsection 2.4.2, equations (2.16)-(2.20). Thus, in the following we focus on the cases as $r \leq \ell$. A sequence with p_1 and p_2 distinct indices in $\mathcal{I}$ and $\mathcal{J}$ corresponds to

$$N_T(N_T-1)\dots(N_T-p_1-1)N_R(N_R-1)\dots(N_R-p_2-1) = \mathcal{O}(N_T^{p_1} N_R^{p_2}) = \mathcal{O}(\theta^{2(p_1+p_2)})$$

identical terms. It is straightforward to verify that the term $\mathbb{E}\{h_{i_1 j_1}^* h_{i_1 j_2} \dots h_{i_\ell j_\ell}^* h_{i_\ell j_1}\}$ is bounded [2] since the moments of a Gaussian distribution are bounded. Additionally,

$$\mathbb{E}\{\breve{g}_{i_1 j_1}^* \breve{g}_{i_1 j_2} \dots \breve{g}_{i_\ell j_\ell}^* \breve{g}_{i_\ell j_1}\} = O\left(\frac{(m_{\widehat{\mathbf{T}}}^{(2)})^\ell}{\theta^{2\ell}}\right) \tag{A.27}$$

[1] Note that this is true for complex Gaussian random h_{ij}. The proof can be readily extended to real Gaussian random variables considering terms where a pair (i, j) appears only once in the sequence $(j_1, i_1, j_2, i_2, \dots, j_\ell, i_\ell)$.

[2] A rough upper bound is given by $\mathbb{E}\{h_{i_1 j_1}^* h_{i_1 j_2} \dots h_{i_\ell j_\ell}^* h_{i_\ell j_1}\} \leq \left((\frac{\ell}{2} - 1)!!\right)^{\ell/2}$, where !! denotes the double factorial.

Then, the contributions of terms corresponding to a given even graph $(j_1, i_1, j_2, i_2, \ldots, j_\ell, i_\ell)$ with $p_1 + p_2 = r \leq \ell$ is proportional to

$$\frac{1}{N_T}\, \theta^{2(p_1+p_2)}\, \frac{(m_{\widehat{\mathbf{T}}}^{(2)})^\ell}{\theta^{2\ell}} = \frac{1}{N_T}\, \theta^{-2(\ell-r)}\, (m_{\widehat{\mathbf{T}}}^{(2)})^\ell \tag{A.28}$$

In the worst case, as $r = \ell$, this term vanishes as $\dfrac{1}{N_T}$. Since the number of distinct graphs $(j_1, i_1, j_2, i_2, \ldots, j_\ell, i_\ell)$ is finite and independent of N_T, the contributions of terms corresponding to even graphs with $r < \ell + 1$ vanishes as $L \to \infty$.

Appendix B

Appendices of Chapter 3

B.1 Derivation of Deterministic CRB

A non-singular square matrix $\mathbf{R}$ and its inverse $\mathbf{R}^{-1}$ can be partitioned into 2×2 blocks as

$$\mathbf{R} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}$$

and

$$\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{bmatrix}$$

when the diagonal partitions of $\mathbf{R}$ and $\mathbf{R}^{-1}$ are square, i.e., the matrices $\mathbf{A}$, $\mathbf{D}$, $\mathbf{E}$, $\mathbf{H}$ are square, $\mathbf{A}$ and $\mathbf{E}$ have the same size, and so do $\mathbf{D}$ and $\mathbf{H}$.

Assume $\mathbf{A}$ and $\mathbf{D}$ are non-singular; then the matrix $\mathbf{R}$ is invertible if and only if the Schur complement $\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$ of $\mathbf{A}$ and the Schur complement $\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ of $\mathbf{D}$ are invertible, then the diagonal partitions of matrix $\mathbf{R}^{-1}$ are given by [115]

$$\mathbf{E} = (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} \tag{B.1}$$

$$\mathbf{H} = (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \tag{B.2}$$

So, according to $\mathcal{J}^{d}_{\boldsymbol{\theta},\boldsymbol{\theta}}$ in (3.7)

$$\mathcal{J}^d_{\boldsymbol{\theta},\boldsymbol{\theta}} = \rho \begin{bmatrix} \mathbf{Q}'^H\mathbf{Q}' & \mathbf{Q}'^H\mathbf{R}' \\ \mathbf{R}'^H\mathbf{Q}' & \mathbf{R}'^H\mathbf{R}' \end{bmatrix}$$

and making use of (B.1), the block $(1,1)$ of $\left(\mathcal{J}^d_{\boldsymbol{\theta},\boldsymbol{\theta}}\right)^{-1}$ relative to the estimation of the channel coefficients $\mathbf{g}_I$ is given by

$$\begin{aligned} \mathrm{CRB}^d_{\mathbf{g}_I} &= \frac{1}{\rho}\Big(\mathbf{Q}'^H\mathbf{Q}' - \mathbf{Q}'^H\mathbf{R}'(\mathbf{R}'^H\mathbf{R}')^{-1}\mathbf{R}'^H\mathbf{Q}'\Big)^{-1} \\ &= \frac{1}{\rho}\Big(\mathbf{Q}'^H\big[\mathbf{I} - \mathbf{R}'(\mathbf{R}'^H\mathbf{R}')^{-1}\mathbf{R}'^H\big]\mathbf{Q}'\Big)^{-1} \\ &= \frac{1}{\rho}\left(\mathbf{Q}'^H P^{\perp}_{\mathbf{R}'}\,\mathbf{Q}'\right)^{-1} \end{aligned} \tag{B.3}$$

where $P^{\perp}_{\mathbf{R}'} = \mathbf{I} - \mathbf{R}'(\mathbf{R}'^H\mathbf{R}')^{-1}\mathbf{R}'^H$. Similarly, making use of (B.2), the block $(2,2)$ of $\left(\mathcal{J}^d_{\boldsymbol{\theta},\boldsymbol{\theta}}\right)^{-1}$relative to data symbols $\mathrm{vec}(\mathbf{X}_d)$ can be obtained as follows

$$\begin{aligned} \mathrm{CRB}^d_{\mathrm{vec}(\mathbf{X}_d)} &= \frac{1}{\rho}\Big(\mathbf{R}'^H\mathbf{R}' - \mathbf{R}'^H\mathbf{Q}'(\mathbf{Q}'^H\mathbf{Q}')^{-1}\mathbf{Q}'^H\mathbf{R}'\Big)^{-1} \\ &= \frac{1}{\rho}\Big(\mathbf{R}'^H\big[\mathbf{I} - \mathbf{Q}'(\mathbf{Q}'^H\mathbf{Q}')^{-1}\mathbf{Q}'^H\big]\mathbf{R}'\Big)^{-1} \\ &= \frac{1}{\rho}\left(\mathbf{R}'^H P^{\perp}_{\mathbf{Q}'}\,\mathbf{R}'\right)^{-1} \end{aligned} \tag{B.4}$$

where $P^{\perp}_{\mathbf{Q}'} = \mathbf{I} - \mathbf{Q}'(\mathbf{Q}'^H\mathbf{Q}')^{-1}\mathbf{Q}'^H$.

B.2 Derivation of Algorithm 2

The channel vector $\mathbf{g}_I$ estimation and data matrix $\mathbf{X}_d$ detection can be carried out jointly by maximizing the posterior probability density function $f(\mathbf{g}_I, \mathbf{X}_d|\mathbf{Y})$ leading to the joint MAP estimation of channel and data symbols as follows:

$$\begin{aligned}
(\widehat{\mathbf{g}}_I, \widehat{\mathbf{X}}_d) &= \arg\max_{\mathbf{g}_I, \mathbf{X}_d} \ln f(\mathbf{g}_I, \mathbf{X}_d|\mathbf{Y}) \\
&= \arg\max_{\mathbf{g}_I, \mathbf{X}_d} \ln f(\mathbf{Y}|\mathbf{g}_I, \mathbf{X}_d) + \ln f(\mathbf{g}_I) + \ln f(\mathbf{X}_d) \\
&= \arg\max_{\mathbf{g}_I, \mathbf{X}_d} \Big(- \mathrm{tr}\{(\mathbf{Y} - \sqrt{\rho}\,\mathbf{G}_I\mathbf{X})^H \mathbf{C}_{\mathbf{YY}}^{-1}(\mathbf{Y} - \sqrt{\rho}\,\mathbf{G}_I\mathbf{X})\} - \mathbf{g}_I^H \mathbf{C}_{\mathbf{g}_I\mathbf{g}_I}^{-1}\mathbf{g}_I - \mathrm{tr}\{\mathbf{X}_d^H\mathbf{X}_d\}\Big) \\
&= \arg\min_{\mathbf{g}_I, \mathbf{X}_d} \Big(\underbrace{\mathrm{tr}\{(\mathbf{Y} - \sqrt{\rho}\,\mathbf{G}_I\mathbf{X})^H \mathbf{C}_{\mathbf{YY}}^{-1}(\mathbf{Y} - \sqrt{\rho}\,\mathbf{G}_I\mathbf{X})\} + \mathbf{g}_I^H \mathbf{C}_{\mathbf{g}_I\mathbf{g}_I}^{-1}\mathbf{g}_I + \mathrm{tr}\{\mathbf{X}_d^H\mathbf{X}_d\}}_{\boldsymbol{\phi}} \Big) \\
&= \arg\min_{\mathbf{g}_I, \mathbf{X}_d} \Big(\underbrace{(\mathbf{y} - \sqrt{\rho}\,\mathbf{Q}\mathbf{g}_I)^H \mathbf{C}_{\mathbf{yy}}^{-1}(\mathbf{y} - \sqrt{\rho}\,\mathbf{Q}\mathbf{g}_I) + \mathbf{g}_I^H \mathbf{C}_{\mathbf{g}_I\mathbf{g}_I}^{-1}\mathbf{g}_I + \mathrm{tr}\{\mathbf{X}_d^H\mathbf{X}_d\}}_{\boldsymbol{\psi}} \Big)
\end{aligned} \tag{B.5}$$

The optimization in (B.5) can be solved by alternating between minimization w.r.t. $\mathbf{g}_I$ and $\mathbf{X}_d$ which yields to

- Minimization w.r.t. $\mathbf{g}_I$: Setting the derivative of $\boldsymbol{\psi}$ w.r.t. $\mathbf{g}_I$ equal to zero, we get

$$\frac{\partial \boldsymbol{\psi}}{\partial \mathbf{g}_I} = 0 \tag{B.6}$$

$$-\sqrt{\rho}\,\mathbf{Q}^H \mathbf{C}_{\mathbf{yy}}^{-1}\big(\mathbf{y} - \sqrt{\rho}\,\mathbf{Q}\mathbf{g}_I\big) + \mathbf{C}_{\mathbf{g}_I\mathbf{g}_I}^{-1}\mathbf{g}_I = 0 \tag{B.7}$$

$$\widehat{\mathbf{g}}_I = \sqrt{\rho}\left(\rho\,\mathbf{Q}^H\,\mathbf{C}_{\mathbf{yy}}^{-1}\,\mathbf{Q} + \mathbf{C}_{\mathbf{g}_I\mathbf{g}_I}^{-1}\right)^{-1}\mathbf{Q}^H\,\mathbf{C}_{\mathbf{yy}}^{-1}\,\mathbf{y} \tag{B.8}$$

- Minimization w.r.t. $\mathbf{X}_d$: Setting the derivative of $\boldsymbol{\phi}$ w.r.t. $\mathbf{X}_d$ equal to zero, we get

$$\frac{\partial \boldsymbol{\phi}}{\partial \mathbf{X}_d} = 0 \tag{B.9}$$

$$-\sqrt{\rho}\,\mathbf{G}_I \mathbf{C}_{\mathbf{YY}}^{-1}\big(\mathbf{Y}_d - \sqrt{\rho}\,\mathbf{G}_I\mathbf{X}_d\big) + \mathbf{X}_d = 0 \tag{B.10}$$

$$\widehat{\mathbf{X}}_d = \sqrt{\rho}\left(\rho\,\mathbf{G}_I\mathbf{C}_{\mathbf{YY}}^{-1}\,\mathbf{G}_I + \mathbf{I}_K\right)^{-1}\mathbf{G}_I\,\mathbf{C}_{\mathbf{YY}}^{-1}\,\mathbf{Y}_d \tag{B.11}$$

Bibliography

[1] V. W. Wong, R. Schober, D. W. K. Ng, and L.-C. Wang, *Key technologies for 5G wireless systems*. Cambridge university press, 2017.

[2] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5g," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1637–1660, 2020.

[3] H. Yang and T. L. Marzetta, "Capacity performance of multicell large-scale antenna systems," in *2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2013, pp. 668–675.

[4] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO: uniformly great service for everyone," in *2015 IEEE 16th international workshop on signal processing advances in wireless communications (SPAWC)*. IEEE, 2015, pp. 201–205.

[5] ——, "Cell-free massive MIMO versus small cells," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1834–1850, 2017.

[6] W. Feng, Y. Wang, N. Ge, J. Lu, and J. Zhang, "Virtual mimo in multi-cell distributed antenna systems: Coordinated transmissions with large-scale csit," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 10, pp. 2067–2081, 2013.

[7] S. Venkatesan, A. Lozano, and R. Valenzuela, "Network mimo: Overcoming intercell interference in indoor wireless systems," in *2007 Conference Record of the Forty-First Asilomar Conference on Signals, Systems and Computers*. IEEE, 2007, pp. 83–87.

[8] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai, "Distributed mimo in multi-cell wireless systems via finite-capacity links," in *2008 3rd International Symposium on Communications, Control and Signal Processing*. IEEE, 2008, pp. 203–206.

[9] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance,

and field trial results," *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102–111, 2011.

[10] X.-H. You, D.-M. Wang, B. Sheng, X.-Q. Gao, X.-S. Zhao, and M. Chen, "Cooperative distributed antenna systems for mobile communications," *IEEE Wireless Communications*, vol. 17, no. 3, pp. 35–43, 2010.

[11] W. Roh and A. Paulraj, "Mimo channel capacity for the distributed antenna," in *Proceedings IEEE 56th Vehicular Technology Conference*, vol. 2. IEEE, 2002, pp. 706–709.

[12] W. Choi and J. G. Andrews, "Downlink performance and capacity of distributed antenna systems in a multicell environment," *IEEE Transactions on Wireless Communications*, vol. 6, no. 1, pp. 69–73, 2007.

[13] G. Interdonato, *Cell-Free Massive MIMO: Scalability, Signal Processing and Power Control*. Linköping University Electronic Press, 2020, vol. 2090.

[14] S. Shamai and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *IEEE VTS 53rd Vehicular Technology Conference, Spring 2001. Proceedings (Cat. No. 01CH37202)*, vol. 3. IEEE, 2001, pp. 1745–1749.

[15] S. Zhou, M. Zhao, X. Xu, J. Wang, and Y. Yao, "Distributed wireless communication system: a new architecture for future public wireless access," *IEEE Communications Magazine*, vol. 41, no. 3, pp. 108–113, 2003.

[16] D. Wang, J. Wang, X. You, Y. Wang, M. Chen, and X. Hou, "Spectral efficiency of distributed MIMO systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 10, pp. 2112–2127, 2013.

[17] J. Joung, Y. K. Chia, and S. Sun, "Energy-efficient, large-scale distributed-antenna system (L-DAS) for multiple users," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 954–965, 2014.

[18] J. Zhang, S. Chen, Y. Lin, J. Zheng, B. Ai, and L. Hanzo, "Cell-free massive mimo: A new next-generation paradigm," *IEEE Access*, vol. 7, pp. 99 878–99 888, 2019.

[19] H. Yang and T. L. Marzetta, "Energy efficiency of massive mimo: Cell-free vs. cellular," in *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*. IEEE, 2018, pp. 1–5.

[20] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive mimo systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4445–4459, 2017.

[21] Ö. Özdogan, E. Björnson, and J. Zhang, "Performance of cell-free massive mimo with rician fading and phase shifts," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5299–5315, 2019.

[22] E. Björnson and L. Sanguinetti, "Making cell-free massive mimo competitive with mmse processing and centralized implementation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 77–90, 2019.

[23] Z. Chen and E. Björnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5205–5219, 2018.

[24] M. Bashar, H. Q. Ngo, A. G. Burr, D. Maryopi, K. Cumanan, and E. G. Larsson, "On the performance of backhaul constrained cell-free massive mimo with linear receivers," in *2018 52nd Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2018, pp. 624–628.

[25] T. C. Mai, H. Q. Ngo, and T. Q. Duong, "Cell-free massive mimo systems with multi-antenna users," in *2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 2018, pp. 828–832.

[26] P. Liu, K. Luo, D. Chen, and T. Jiang, "Spectral efficiency analysis of cell-free massive mimo systems with zero-forcing detector," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 795–807, 2019.

[27] A. Papazafeiropoulos, P. Kourtessis, M. Di Renzo, S. Chatzinotas, and J. M. Senior, "Performance analysis of cell-free massive mimo systems: A stochastic geometry approach," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 4, pp. 3523–3537, 2020.

[28] Z. Chen and E. Bjoernson, "Can we rely on channel hardening in cell-free massive MIMO?" in *2017 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2017, pp. 1–6.

[29] R. Gholami, L. Cottatellucci, and D. Slock, "Favorable propagation and linear multiuser detection for distributed antenna systems," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 5190–5194.

[30] ——, "Channel models, favorable propagation and multistage linear detection in cell-free massive mimo," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 2942–2947.

[31] L. Cottatellucci, "Spectral efficiency of extended networks with randomly distributed transmitters and receivers," in *2014 IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP)*. IEEE, 2014, pp. 673–677.

[32] ——, "Capacity per unit area of distributed antenna systems with centralized processing," in *2014 IEEE Global Communications Conference*. IEEE, 2014, pp. 1746–1752.

[33] M. Mézard, G. Parisi, and A. Zee, "Spectra of Euclidean random matrices," *Nuclear Physics B*, vol. 559, no. 3, pp. 689–701, 1999.

[34] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[35] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Aspects of favorable propagation in massive MIMO," in *2014 22nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2014, pp. 76–80.

[36] ——, "Energy and spectral efficiency of very large multiuser mimo systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.

[37] T. L. Marzetta, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.

[38] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE signal processing magazine*, vol. 30, no. 1, pp. 40–60, 2012.

[39] X. Wu, N. C. Beaulieu, and D. Liu, "On favorable propagation in massive mimo systems and different antenna configurations," *IEEE Access*, vol. 5, pp. 5578–5593, 2017.

[40] S. E. Hajri, J. Denis, and M. Assaad, "Enhancing favorable propagation in cell-free massive mimo through spatial user grouping," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2018, pp. 1–5.

[41] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive mimo networks: Spectral, energy, and hardware efficiency," *Foundations and Trends in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[42] N. Hassan and X. Fernando, "Massive mimo wireless networks: An overview," *Electronics*, vol. 6, no. 3, p. 63, 2017.

[43] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE transactions on wireless communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

[44] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive mimo—5g system," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 905–923, 2015.

[45] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 773–786, 2014.

[46] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on selected areas in communications*, vol. 31, no. 2, pp. 264–273, 2013.

[47] H. Yin, D. Gesbert, and L. Cottatellucci, "Dealing with interference in distributed large-scale mimo systems: A statistical approach," *IEEE Journal of selected topics in signal processing*, vol. 8, no. 5, pp. 942–953, 2014.

[48] H. Yin, L. Cottatellucci, D. Gesbert, R. R. Müller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *IEEE Transactions on Signal Processing*, vol. 64, no. 11, pp. 2990–3003, 2016.

[49] Y. Zhang, H. Cao, P. Zhong, C. Qi, and L. Yang, "Location-based greedy pilot assignment for cell-free massive mimo systems," in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*. IEEE, 2018, pp. 392–396.

[50] A. Ashikhmin, H. Q. Ngo, T. L. Marzetta, and H. Yang, "Pilot assignment in cell free massive mimo wireless systems," Apr. 4 2017, uS Patent 9,615,384.

[51] M. Attarifar, A. Abbasfar, and A. Lozano, "Random vs structured pilot assignment in cell-free massive mimo wireless networks," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2018, pp. 1–6.

[52] G. Femenias and F. Riera-Palou, "Cell-free millimeter-wave massive mimo systems with limited fronthaul capacity," *IEEE Access*, vol. 7, pp. 44 596–44 612, 2019.

[53] H. Liu, J. Zhang, S. Jin, and B. Ai, "Graph coloring based pilot assignment for cell-free massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 9180–9184, 2020.

[54] W. H. Hmida, V. Meghdadi, A. Bouallegue, and J.-P. Cances, "Graph coloring based pilot reuse among interfering users in cell-free massive mimo," in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, 2020, pp. 1–6.

[55] H. Liu, J. Zhang, X. Zhang, A. Kurniawan, T. Juhana, and B. Ai, "Tabu-search-based pilot assignment for cell-free massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 2286–2290, 2019.

[56] S. Buzzi, C. D'Andrea, M. Fresia, Y.-P. Zhang, and S. Feng, "Pilot assignment in cell-free massive mimo based on the hungarian algorithm," *IEEE Wireless Communications Letters*, vol. 10, no. 1, pp. 34–37, 2020.

[57] A. K. Jagannatham and B. D. Rao, "Whitening-rotation-based semi-blind mimo channel estimation," *IEEE Transactions on Signal Processing*, vol. 54, no. 3, pp. 861–869, 2006.

[58] M. Abuthinien, S. Chen, and L. Hanzo, "Semi-blind joint maximum likelihood channel estimation and data detection for mimo systems," *IEEE Signal Processing Letters*, vol. 15, pp. 202–205, 2008.

[59] E. De Carvalho and D. T. Slock, "Cramer-rao bounds for semi-blind, blind and training sequence based channel estimation," in *First IEEE signal processing workshop on signal processing advances in wireless communications*. IEEE, 1997, pp. 129–132.

[60] E. de Carvalho and D. Slock, "Asymptotic performance of ml methods for semi-blind channel estimation," in *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems and Computers (Cat. No. 97CB36136)*, vol. 2. IEEE, 1997, pp. 1624–1628.

[61] E. Nayebi and B. D. Rao, "Semi-blind channel estimation for multiuser massive mimo systems," *IEEE Transactions on Signal Processing*, vol. 66, no. 2, pp. 540–553, 2017.

[62] E. de Carvalho and D. T. Slock, "Blind and semi-blind fir multichannel estimation:(global) identifiability conditions," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 1053–1064, 2004.

[63] C. H. Aldana, E. de Carvalho, and J. M. Cioffi, "Channel estimation for multicarrier multiple input single output systems using the em algorithm," *IEEE Transactions on Signal Processing*, vol. 51, no. 12, pp. 3280–3292, 2003.

[64] J. Yang, W. Song, S. Zhang, X. You, and C. Zhang, "Low-complexity belief propagation detection for correlated large-scale mimo systems," *Journal of Signal Processing Systems*, vol. 90, no. 4, pp. 585–599, 2018.

[65] T. P. Minka, "Expectation propagation for approximate bayesian inference," *arXiv preprint arXiv:1301.2294*, 2013.

[66] M. Seeger, "Bayesian gaussian process models: Pac-bayesian generalisation error bounds and sparse approximations," University of Edinburgh, Tech. Rep., 2003.

[67] K. Ghavami and M. Naraghi-Pour, "Blind channel estimation and symbol detection for multi-cell massive mimo systems by expectation propagation," *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 943–954, 2017.

[68] J. Cespedes, P. M. Olmos, M. Sánchez-Fernández, and F. Perez-Cruz, "Expectation propagation detection for high-order high-dimensional mimo systems," *IEEE Transactions on Communications*, vol. 62, no. 8, pp. 2840–2849, 2014.

[69] T. P. Minka, "A family of algorithms for approximate bayesian inference," Ph.D. dissertation, Massachusetts Institute of Technology, 2001.

[70] T. Heskes, M. Opper, W. Wiegerinck, O. Winther, and O. Zoeter, "Approximate inference techniques with expectation constraints," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2005, no. 11, p. P11015, 2005.

[71] G. Yao, G. Yang, J. Hu, and C. Fei, "A low complexity expectation propagation detection for massive mimo system," in *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2018, pp. 1–6.

[72] H. Wang, A. Kosasih, C.-K. Wen, S. Jin, and W. Hardjawana, "Expectation propagation detector for extra-large scale massive mimo," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2036–2051, 2020.

[73] K. Ghavami and M. Naraghi-Pour, "Noncoherent massive mimo detection by expectation propagation," in *GLOBECOM 2017-2017 IEEE Global Communications Conference*. IEEE, 2017, pp. 1–6.

[74] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and B. D. Rao, "Performance of cell-free massive mimo systems with mmse and lsfd receivers," in *2016 50th Asilomar Conference on Signals, Systems and Computers*. IEEE, 2016, pp. 203–207.

[75] H. He, H. Wang, X. Yu, J. Zhang, S. Song, and K. B. Letaief, "Distributed expectation propagation detection for cell-free massive mimo," *arXiv preprint arXiv:2108.07498*, 2021.

[76] C. Jeon, K. Li, J. R. Cavallaro, and C. Studer, "Decentralized equalization with feedforward architectures for massive mu-mimo," *IEEE Transactions on Signal Processing*, vol. 67, no. 17, pp. 4418–4432, 2019.

[77] L. Cottatellucci and R. R. Muller, "A systematic approach to multistage detectors in multipath fading channels," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3146–3158, 2005.

[78] S. Moshavi, "Multi–user detection for DS–CDMA communications," *IEEE Communications Magazine*, vol. 34, no. 10, pp. 124–136, Oct. 1996.

[79] J. S. Goldstein, I. S. Reed, and L. L. Scharf, "A multistage representation of the Wiener filter based on orthogonal projections," *IEEE Transactions on Information Theory*, vol. 44, no. 7, Nov. 1998.

[80] A. Tulino and S. Verdú, *Random Matrix Theory and Wireless Communications*, ser. Foundations and Trends in Communications and Information Theory. Now Publishers Inc., Jun. 2004, vol. 1.

[81] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*. Cambridge University Press, 2011.

[82] D. Voiculescu, "Symmetries of some reduced free product c*-algebras," in *Operator algebras and their connections with topology and ergodic theory*. Springer, 1985, pp. 556–588.

[83] ——, "Addition of certain non-commuting random variables," *Journal of functional analysis*, vol. 66, no. 3, pp. 323–346, 1986.

[84] ——, "Multiplication of certain non-commuting random variables," *Journal of Operator Theory*, pp. 223–235, 1987.

[85] ——, "Limit laws for random matrices and free products," *Inventiones mathematicae*, vol. 104, no. 1, pp. 201–220, 1991.

[86] J. A. Mingo and R. Speicher, *Free probability and random matrices*. Springer, 2017, vol. 35.

[87] R. Speicher, "Free probability and random matrices," *arXiv preprint arXiv:1404.3393*, 2014.

[88] ——, "Free probability theory," *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 119, no. 1, pp. 3–30, 2017.

[89] ——, "Free probability theory and non-crossing partitions," *Sém. Lothar. Combin*, vol. 39, p. 38, 1997.

[90] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*. Cambridge university press, 2010, no. 118.

[91] R. M. Gray *et al.*, "Toeplitz and circulant matrices: A review," *Foundations and Trends® in Communications and Information Theory*, vol. 2, no. 3, pp. 155–239, 2006.

[92] J. Gutiérrez-Gutiérrez, P. M. Crespo *et al.*, *Block Toeplitz matrices: Asymptotic results and applications*. Now, 2012.

[93] S. Skipetrov and A. Goetschy, "Eigenvalue distributions of large Euclidean random matrices for waves in random media," *arXiv preprint arXiv:1007.1379*, 2010.

[94] A. Nyberg, "The Laplacian spectra of random geometric graphs," Ph.D. dissertation, 2014.

[95] Ø. Ryan and M. Debbah, "Asymptotic behavior of random vandermonde matrices with entries on the unit circle," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3115–3147, 2009.

[96] ——, "Convolution operations arising from vandermonde matrices," *IEEE transactions on information theory*, vol. 57, no. 7, pp. 4647–4659, 2011.

[97] X.-G. Xia, "A simple introduction to free probability theory and its application to random matrices," *arXiv preprint arXiv:1902.10763*, 2019.

[98] L. Cottatellucci, R. R. Muller, and M. Debbah, "Asynchronous CDMA systems with random spreading—Part II: design criteria," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1498–1520, 2010.

[99] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*. Springer, 2010, vol. 20.

[100] L. Cottatellucci and R. R. Müller, "CDMA systems with correlated spatial diversity: A generalized resource pooling result." *IEEE Transactions on Information Theory*, vol. 53, no. 3, pp. 1116–1136, Mar. 2007.

[101] S.-M. Omar, D. T. Slock, and O. Bazzi, "Bayesian and deterministic crbs for semi-blind channel estimation in simo single carrier cyclic prefix systems," in *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*. IEEE, 2011, pp. 1682–1686.

[102] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.

[103] B. Hochwald and A. Nehorai, "On identifiability and information-regularity in parametrized normal distributions," *Circuits, Systems and Signal Processing*, vol. 16, no. 1, pp. 83–89, 1997.

[104] T. Richardson and R. Urbanke, *Modern coding theory*. Cambridge university press, 2008.

[105] R. M. Karp and M. Sipser, "Maximum matching in sparse random graphs," in *22nd Annual Symposium on Foundations of Computer Science (sfcs 1981)*. IEEE, 1981, pp. 364–375.

[106] J. Aronson, A. Frieze, and B. G. Pittel, "Maximum matchings in sparse random graphs: Karp–sipser revisited," *Random Structures & Algorithms*, vol. 12, no. 2, pp. 111–177, 1998.

[107] J.-H. Zhao and H.-J. Zhou, "Two faces of greedy leaf removal procedure on graphs," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2019, no. 8, p. 083401, 2019.

[108] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 3, pp. 794–816, 2016.

[109] Y. Sun, A. Breloy, P. Babu, D. P. Palomar, F. Pascal, and G. Ginolhac, "Low-complexity algorithms for low rank clutter parameters estimation in radar systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 8, pp. 1986–1998, 2015.

[110] A. Tang, J. Sun, and K. Gong, "Mobile propagation loss with a low base station antenna for NLOS street microcells in urban area," in *IEEE VTS 53rd Vehicular Technology Conference, Spring 2001. Proceedings (Cat. No. 01CH37202)*, vol. 1. IEEE, 2001, pp. 333–336.

[111] K. Ghavami and M. Naraghi-Pour, "Mimo detection with imperfect channel state information using expectation propagation," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8129–8138, 2017.

[112] S. Barthelmé and N. Chopin, "Expectation propagation for likelihood-free inference," *Journal of the American Statistical Association*, vol. 109, no. 505, pp. 315–333, 2014.

[113] K.-H. Ngo, M. Guillaud, A. Decurninge, S. Yang, and P. Schniter, "Multi-user detection based on expectation propagation for the non-coherent simo multiple access channel," *IEEE Transactions on Wireless Communications*, vol. 19, no. 9, pp. 6145–6161, 2020.

[114] J. Zhang, E. K. Chong, and D. N. C. Tse, "Output mai distributions of linear mmse multiuser receivers in ds-cdma systems," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 1128–1144, 2001.

[115] T.-T. LU and S.-H. SHIOU, "Inverses of 2× 2 block matrices," *Computers and Mathematics with Applications*, vol. 43, pp. 119–129, 2002.