ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2015-ENST-0080

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Communication et Electronique »

présentée et soutenue publiquement par

Haifan YIN

le 14 décembre 2015

L'atténuation des interférences dans les systèmes MIMO massifs

Directeur de thèse : **David GESBERT**
Co-encadrement de la thèse : **Laura COTTATELLUCCI**

Jury

M. Constantinos PAPADIAS,	Professeur,	Athens Information Technology	Rapporteur
M. Wolfgang UTSCHICK,	Professeur,	Technische Universität München	Rapporteur
M. Bruno CLERCKX,	Lecteur,	Imperial College London	Examinateurs
M. Gaoning HE,	Docteur,	Huawei France Research Center	Examinateurs

THÈSE

TELECOM ParisTech
école de l'Institut Télécom - membre de ParisTech

DISSERTATION

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
from TELECOM ParisTech

Specialization : Communication and Electronics

Haifan Yin

Interference Mitigation in Massive MIMO Systems

Successfully defended the 14th of December 2015 before a committee
composed of :

Reviewer and President of the Jury		
Professor	Constantinos Papadias	Athens Information Technology
Reviewer		
Professor	Wolfgang Utschick	Technische Universität München
Examiners		
Professor	Bruno Clerckx	Imperial College London
Doctor	Gaoning He	Huawei France Research Center
Thesis Supervisor		
Professor	David Gesbert	EURECOM
Thesis Co-supervisor		
Professor	Laura Cottatellucci	EURECOM

THÈSE

présentée pour l'obtention du grade de

Docteur de TELECOM ParisTech

Spécialité : Communication et Electronique

Haifan Yin

L'atténuation des interférences dans les systèmes MIMO massifs

Soutenance de thèse validée le 14 Décembre 2015 devant le jury composé de :

Président du jury		
Professeur	Constantinos Papadias	Athens Information Technology
Rapporteurs		
Professeur	Wolfgang Utschick	Technische Universität München
Professeur	Bruno Clerckx	Imperial College London
Docteur	Gaoning He	Huawei France Research Center
Directeur de thèse		
Professeur	David Gesbert	EURECOM
Co-encadrement de la thèse		
Professeur	Laura Cottatellucci	EURECOM

Abstract

Massive MIMO is an emerging technique considered for use in future mobile networks that might enable much higher throughput and energy efficiency compared to traditional multiuser MIMO (MU-MIMO) systems. The gain is achieved by adding a large number of inexpensive low-power antennas at the base stations, instead of having small number of high-cost, high-power antenna elements at the base stations, e.g., 8 antennas per base station in currently standardized LTE-Advanced systems. In less than five years, massive MIMO has sparked tremendous research activities. It is considered to be a potential key technology in future 5G standard. Despite its potential of huge improvements, there are still plenty of open problems and challenges which limit the potential of massive MIMO. Among them, this thesis focuses on two of the challenges of massive MIMO systems, namely pilot interference reduction in Time-Division Duplex (TDD) mode and Channel State Information (CSI) feedback reduction in Frequency Division Duplex (FDD) mode.

Channel estimation in massive MIMO networks, which is known to be hampered by the pilot contamination effect, constitutes a bottleneck for overall performance. We present novel approaches which tackle this problem by enabling a low-rate coordination between cells during the channel estimation phase itself. The coordination makes use of the additional second-order statistical information about the user channels, which are shown to offer a powerful way of discriminating across interfering users with even strongly correlated pilot sequences. Importantly, we demonstrate analytically that in the large-number-of-antennas regime, the pilot contamination effect is made to vanish completely under certain condition on the channel covariance. This condition is identified as a non-overlapping condition, which states that when the support of multipath angle-of-arrival (AoA) of interference is non-overlapping with the AoA support of desired channel, then for a base station equipped with a uniform linear array (ULA), the pilot contamination can be made to vanish completely using a minimum mean square error

(MMSE) estimator. This phenomenon is mainly owing to the low-rankness property of channel covariance, which is identified and proved theoretically in this thesis. Furthermore, we show that such a low-rank property is not inherently related to ULA. It can be generalized to non-uniform array, and more surprisingly, to two dimensional distributed large scale arrays. In addition to pilot decontamination, we demonstrate that such a property has other promising applications such as statistical interference filtering.

Although the proposed MMSE-based estimator leads to full pilot decontamination under the strict condition that the desired and interference channel do not overlap in their AoA regions, in practice this condition is unlikely to hold at all times, owing to the random user location and scattering effects. To this end, we propose novel robust channel estimation schemes that combine the merits of MMSE estimator and the known amplitude based projection method. Asymptotic analysis shows that the proposed methods require weaker conditions compared to the known methods to achieve full decontamination.

Finally, we tackle the CSI feedback problem for massive MIMO operating in FDD mode by novel cooperative feedback mechanisms. We exploit synergies between massive MIMO systems and inter-user communications based on Device-to-Device (D2D) communications. The exchange of local CSI among users, enabled by D2D communications, allows to construct more *informative* forms of feedback based on this shared knowledge. Two feedback variants are highlighted : 1) cooperative CSI feedback, and 2) cooperative precoder index feedback. For a given feedback overhead, the sum-rate performance is assessed and the gains compared with a conventional massive MIMO setup without D2D are shown.

Acknowledgements

It has been ten years since the I stepped into my first university. A decade of university life brings me plenty of happiness from success as well as sorrow from failure. I do not remember how many times I stayed up all night writing programs or designing circuits. Nor do I recall the bitterness of the desperate struggles in many electronic design contests. I quite enjoyed myself in practical implementation, until the resolution of pursuing a Ph.D. pushed me out of it. As people told me that it would not be easy to publish papers on the LTE hardware platform I built. As a result I decided to venture into the field of theoretical research.

However things did not go on so well in the beginning. I was wandering around the vast sea of research, like a sailboat without a rudder, for a year. The situation began to improve when I met David Gesbert, my current supervisor, who introduced me to a relatively new research topic - massive MIMO, a topic which was not so hot as it is today. With his precise intuition, he predicted that this topic would soon be very competitive. Time proved him right. I owe him a lot of gratitude for such a good research direction, as well as his profound knowledge and careful guidance. He is a scholar with razor-sharp wit and gentle personality. Without him this work could not have been accomplished. I would also like to thank Laura Cottatellucci, my co-supervisor, who has been offering me so much help. She is a very devoted and meticulous researcher, always willing to help me solve difficult mathematical problems.

I feel very lucky to have the opportunity to work together with talented people. Discussing with them always gives me enlightenments and inspirations. I would like to express my thankfulness to them : Xinping Yi, Qianrui Li, Junting Chen, Miltiades Filippou, Ralf Müller, Gaoning He, Maxime Guillaud, Dirk Slock, etc.

My wife, Sujuan Hu, has been giving me courage and unconditional support during my Ph.D. research. Without her, life could never be so colorful and lovely.

Finally, I would like to express my gratitude to my parents. They help me realize my potential. Their respect for knowledge planted deep into my mind since I was a child. It was the impetus that made me decide to pursue a Ph.D. degree.

Table des matières

Abstract i
Acknowledgements iii
Contents v
List of Figures ix
List of Tables xiii
Acronyms xiv
Notations xvii

1 Résumé [Français] 1
1.1 Abstract 1
1.2 Introduction 2
1.3 Propriétés des canaux en MIMO massifs 2
1.3.1 Modèle de réseau 3
1.3.2 Modèle de rang faible pour les antennes ULA 3
1.3.3 Modèle de rang faible pour les réseaux d'antennes aléatoires 6
1.3.4 Rang faible pour les réseaux aléatoires linéaires 6
1.3.5 Modèle de rang faible pour les systèmes DAS 7
1.3.6 Propriété de covariance uniformement bornée 8
1.3.7 Conclusions 9
1.4 Estimation de canal basée sur la covariance 10
1.4.1 Apprentissage du canal pour la connection montante . 10
1.4.2 Estimation bayésienne du canal 10
1.4.3 Affectation coordonnée des pilotes 13
1.4.4 Conclusions 16
1.5 Décontamination basée sur la puissance et les angles 17
1.5.1 Transmission de données 17
1.5.2 Méthode basée sur la projection dans le domaine de l'amplitude 18
1.5.3 Conclusions 23

1.6 Feedback coopératif pour le FDD . . . 23
1.6.1 Signal et modèles de canal pour FDD . . . 23
1.6.2 Conception de l'acquisition de l'information de canal sans D2D . . . 24
1.6.3 Acquisistion coopérative de l'information de canal avec D2D . . . 25
1.6.4 Acquisition coopérative de l'indice du précodeur avec D2D . . . 27
1.6.5 Conclusions . . . 27
1.7 Publications . . . 28
1.7.1 Conférences . . . 28
1.7.2 Journaux . . . 29
1.7.3 Brevets . . . 30

2 Introduction **31**
2.1 Motivations . . . 33
2.2 CSI Acquisition: Challenges and Avenues . . . 34
2.2.1 CSI Acquisition in TDD Massive MIMO . . . 34
2.2.2 CSI Acquisition in FDD Massive MIMO . . . 35
2.2.3 Massive MIMO and D2D . . . 36
2.3 Contributions and Publications . . . 37
2.3.1 MMSE-based pilot decontamination . . . 37
2.3.2 Generalized low-rankness of channel covariance and its applications . . . 38
2.3.3 Robust Angle/Power based Pilot Decontamination . . . 39
2.3.4 Cooperative Feedback for FDD Massive MIMO . . . 40

3 Properties of Massive MIMO Channels **41**
3.1 Network Model . . . 41
3.2 Low-Rank Model in ULA . . . 42
3.2.1 Channel Model . . . 42
3.2.2 Low-rankness property of ULA . . . 42
3.3 Low-Rank Model in Random Linear Arrays . . . 44
3.3.1 Channel Model . . . 44
3.3.2 Low-rankness Property of Random Linear Array . . . 45
3.4 Low-Rank Model in DAS . . . 47
3.4.1 Channel Model . . . 47
3.4.2 Low-rankness Property of DAS . . . 48
3.5 Uniformly Boundedness of Channel Covariance in ULA . . . 50
3.6 Conclusions . . . 50

4 Covariance based Channel Estimation **53**
4.1 Introduction . . . 53
4.2 UL training . . . 54
4.3 Pilot Contamination Problem . . . 55
4.4 Covariance-aided Channel Estimation . . . 55
4.4.1 Bayesian Estimation . . . 56
4.4.2 Channel Estimation with Full Pilot Reuse . . . 57
4.4.3 Large-scale Analysis . . . 59
4.5 Coordinated Pilot Assignment . . . 65
4.6 Interference filtering via subspace projection . . . 67
4.7 Numerical Results . . . 68
4.7.1 Co-located Antenna Array . . . 69
4.7.2 Distributed Antenna Array . . . 73
4.8 Discussions . . . 76
4.9 Conclusions . . . 76

5 Joint Angle/Power based Decontamination **79**
5.1 Introduction . . . 79
5.2 UL Training/Data Transmission . . . 81
5.3 A review of LMMSE estimation . . . 82
5.3.1 Asymptotic performance of MMSE . . . 82
5.4 A review of power domain discrimination . . . 83
5.4.1 Generalized amplitude projection . . . 83
5.5 Covariance-aided amplitude based projection . . . 84
5.5.1 Single user per cell . . . 84
5.5.2 Asymptotic performance of the proposed CA estimator 87
5.5.3 Generalization to multiple users per cell . . . 91
5.6 Low-complexity alternatives . . . 93
5.6.1 Subspace and amplitude based projection . . . 93
5.6.2 MMSE + amplitude based projection . . . 95
5.7 Numerical Results . . . 96
5.8 Conclusions . . . 101

6 Cooperative Feedback Design in FDD **103**
6.1 Introduction . . . 103
6.2 Signal and Channel Models for FDD . . . 105
6.3 Feedback Design without D2D . . . 107
6.4 Cooperative Feedback of CSI with D2D . . . 108
6.5 Cooperative Feedback of Precoder Index with D2D . . . 109
6.6 Numerical Results . . . 110

6.6.1 Cooperative CSI Feedback 111
6.6.2 Cooperative Precoder Index Feedback 112
6.7 Conclusions 114

7 Conclusion **115**

Appendices **117**
.1 Proof of Lemma 1: 119
.2 Proof of Theorem 1: 120
.3 Proof of Proposition 5: 121
.4 Proof of Proposition 3: 122
.5 Proof of Theorem 2: 123
.6 Proof of Lemma 2: 126
.7 Proof of Proposition 5: 127
.8 Proof of Lemma 4: 128
.9 Proof of Lemma 6: 128
.10 Proof of Lemma 7: 129
.11 Proof of Lemma 8: 130
.12 Proof of Theorem 4: 132
.13 Proof of Theorem 5: 133

Table des figures

1.1 Support borné de trajets multiples AoA. 5
1.2 Canal composé de $Q = 2$ groupements de trajets multiples. . 5
1.3 Rang de la covariance du canalde pour modèle en forme fermée vs. rang actuel. 7
1.4 Configuration distribuée dans le cadre d'une diffusion avec modèle circulaire. 8
1.5 Rank vs. r, $M = 2000$, $\lambda = 0.15$m, $R_c = 500$m. 9
1.6 Estimation MSE vs. nombre BS antennes , 2-cellule network, deux utilisateurs en position fixée, AOA uniformément distribuées avec $\theta_\Delta = 20$ degrés, nans chevauchement trajets multiples. 12
1.7 Estimation MSE vs. nombre de BS antennes, AOA uniformément distribuées avec $\theta_\Delta = 10$ degrés, 2-cellules network. . . . 15
1.8 Estimation MSE vs. number of BS antennas, AOA distribués gaussienne $\sigma = 10$ degrees, 2-cell network. 15
1.9 Taux par cellule vs. nombre de BS antennes, 2-cellules network, AOA distribués gaussienne avec $\sigma = 10$ degrés. 16
1.10 Estimation des performance versus M pour un réseau à deux-cellules avec un utilisateur par cellule, un exposant d'atténuation $\gamma = 0$ et un support angulaire recouvrant partiellement les angles d'arrivés à hauteur de 60 degrés, $C = 500$, SNR = 0 dB. 20
1.11 Estimation performance vs. M, 7-cellules network, 1 utilisateurs par cellule, AoA réparties 60 degrés, exposant de perte de trajet $\gamma = 2$, cellule-bord SNR = 0 dB. 21
1.12 DL sum-rates with/without feedback cooperation, feedback overhead : 16 bits per user, $K = 3$, $M = 50$. 26

1.13 DL Sum-taux de sélection coopérative de pré-codage et non-coopérative CSI feedback, feedback overhead : 4 bits par utilisateur. 28

3.1 Desired channel composed of $Q = 2$ clusters of multipath. . . 44
3.2 Closed-form rank model for the channel covariance vs. actual rank. 46
3.3 The distributed large-scale antenna setting with a one-ring model. 48
3.4 Rank vs. r, $M = 2000$, $\lambda = 0.15$m, $R_c = 500$m. 49

4.1 Estimation MSE vs. number of BS antennas , 2-cell network, fixed positions of two users, uniformly distributed AoAs with $\theta_\Delta = 20$ degrees, non-overlapping multipath. 62
4.2 Channel Estimation MSE vs. M, $\overline{D} = \lambda/2$, 2-cell network, angle spread 30 degrees, $\overline{\Phi}_d \cap \overline{\Phi}_i = \emptyset$, cell-edge SNR is 20dB. We compare the standard LS to MMSE estimators, in interference and interference-free scenarios. 65
4.3 Estimation MSE vs. number of BS antennas, uniformly distributed AoAs with $\theta_\Delta = 10$ degrees, 2-cell network. 71
4.4 Estimation MSE vs. number of BS antennas, Gaussian distributed AoAs with $\sigma = 10$ degrees, 2-cell network. 71
4.5 Estimation MSE vs. standard deviation of Gaussian distributed AoAs with $M = 10$, 7-cell network. 72
4.6 Per-cell rate vs. number of BS antennas, 2-cell network, Gaussian distributed AoAs with $\sigma = 10$ degrees. 72
4.7 Per-cell rate vs. standard deviation of AoA (Gaussian distribution) with $M = 10$, 7-cell network. 73
4.8 Estimation performance vs. distance between two users, $M = 2000$, $r = 15$m, single-cell network. 74
4.9 Uplink sum-rate vs. distance between 2 users, $M = 500$, $r = 15$m, cell-edge SNR 20dB, single-cell network. 75
4.10 Uplink per-cell rate vs. r, cell-edge SNR 20dB, 7-cell network, each cell has $M = 500$ distributed antennas. 76

5.1 Estimation performance vs. M, 2-cell network, 1 user per cell, path loss exponent $\gamma = 0$, partially overlapping angular support, AoA spread 60 degrees, SNR = 0 dB. 97

5.2 Estimation performance vs. M, 7-cell network, one user per cell, AoA spread 30 degrees, path loss exponent $\gamma = 2$, cell-edge SNR = 0 dB. 98
5.3 Uplink per-cell rate vs. M, 7-cell network, one user per cell, AoA spread 30 degrees, path loss exponent $\gamma = 2$, cell-edge SNR = 0 dB. 99
5.4 Estimation performance vs. M, 7-cell network, 4 users per cell, AoA spread 30 degrees, path loss exponent $\gamma = 2$, cell-edge SNR = 0 dB. 100
5.5 Uplink per-cell rate vs. M, 7-cell network, 4 users per cell, AoA spread 30 degrees, path loss exponent $\gamma = 2$, cell-edge SNR = 0 dB. 100

6.1 DL sum-rates with/without feedback cooperation, feedback overhead: 16 bits per user. 112
6.2 DL sum-rates of cooperative precoder selection and non-cooperative CSI feedback, feedback overhead: 4 bits per user. 113

1 Illustration of a line of scatterers. 124

Liste des tableaux

4.1 Basic simulation parameters 69

Acronyms

Here are the main acronyms used in this document. The meaning of an acronym is also indicated when it is first used.

3GPP	3rd Generation Partnership Project
5G	5th Generation
AoA	Angle of Arrival
AoD	Angle of Departure
AWGN	Additive White Gaussian Noise
BC	Broadcast Channel
BS	Base Station
C-RAN	Cloud-enabled Radio Access Networks
CB	Covariance-aided Bayesian estimation
CPA	Coordinated Pilot Assignment
CDF	Cumulative Density Function.
CLT	Central Limit Theorem
CSI	Channel State Information
CSIT	Channel State Information at the Transmitter
CoMP	Coordinated Multi Point
D2D	Device-to-Device
DAS	Distributed Antenna System
DFT	Discrete Fourier Transform
EVD	Eigenvalue Decomposition
DL	Downlink.
FDD	Frequency Division Duplexing
i.i.d.	independent and identically distributed
JSDM	Joint Spatial Division and Multiplexing
ML	Maximum Likelihood
MRC	Maximum Ratio Combining.
LSAS	Large Scale Antenna Systems

LOS	Line Of Sight
LTE	Long Term Evolution.
LS	Least Squares
LSAS	Large-Scale Antenna Systems
M2M	Machine-to-Machine
MAP	Maximum a Posteriori
MIMO	Multiple Input Multiple Output
MISO	Multiple Input Single Output
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
PC	Pilot Contamination
PDF	Probability Density Function
PSD	Positive Semi-Definite
RobustICA	Robust Independent Component Analysis
RRH	Remote Radio Head
SIMO	Single Input Multiple Output
SINR	Signal to Noise and Interference Ratio
SIR	Signal to Interference ratio
SNR	Signal to Noise Ratio
SISO	Single-Input Single-Output
s.t.	such that
SVD	Singular Value Decomposition.
TDD	Time Division Duplex
UL	Uplink
ULA	Uniform Linear Array
ZF	Zero-Forcing

Notations

Here are the main notations used in this document. The meaning of a notation is also indicated when it is first used.

$\|x\|$	Magnitude of the scalar x
$\mathbf{I}_M$	$M \times M$ identity matrix
$\mathbf{X}^T$	Transpose of the matrix $\mathbf{X}$
$\mathbf{X}^*$	Conjugate of the matrix $\mathbf{X}$
$\mathbf{X}^H$	Conjugate transpose of the matrix $\mathbf{X}$
$\mathbf{X}^{-1}$	Inverse of the matrix $\mathbf{X}$
$\mathbf{X}^\dagger$	Moore-Penrose pseudoinverse of the matrix $\mathbf{X}$
$\mathbf{X}(m,n)$	(m,n)-th entry of the matrix $\mathbf{X}$
$\mathrm{tr}\{\mathbf{X}\}$	Trace of the square matrix $\mathbf{X}$
$\det\{\mathbf{X}\}$	Determinant of the square matrix $\mathbf{X}$
$\mathbf{X} \otimes \mathbf{Y}$	Kronecker product of the two matrices $\mathbf{X}$ and $\mathbf{Y}$
$\mathrm{vec}(\mathbf{X})$	Vectorization of the matrix $\mathbf{X}$
$\|\mathbf{x}\|_2$	ℓ^2 norm of the vector $\mathbf{x}$
$\|\mathbf{X}\|_2$	Spectral norm of the matrix $\mathbf{X}$
$\|\mathbf{X}\|_F$	Frobenius norm of the matrix $\mathbf{X}$
$\mathrm{rank}(\mathbf{X})$	Rank of the matrix $\mathbf{X}$
$\mathbb{E}\{\cdot\}$	Expectation
$\mathrm{diag}\{\mathbf{x_1}, ..., \mathbf{x_N}\}$	(Block) diagonal matrix with $\mathbf{x_1}, ..., \mathbf{x_N}$ at the main diagonal
$\triangleq$	Used for definition
$\xrightarrow{\text{a.s.}}$	Almost sure convergence
$\xrightarrow{\text{d}}$	Convergence in distribution
$\lambda_k\{\mathbf{X}\}$	The k-th largest eigenvalue of the Hermitian matrix $\mathbf{X}$
$\mathbf{e}_k\{\mathbf{A}\}$	The eigenvector of $\mathbf{X}$ corresponding to the eigenvalue $\lambda_k\{\mathbf{X}\}$
$\mathcal{CN}(\mathbf{0}, \mathbf{R})$	Zero-mean complex Gaussian distribution with covariance matrix $\mathbf{R}$
$\mathbb{C}^{K \times C}$	$K \times C$ complex matrix

$\text{span}\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$	Span of linear vector space on the basis of $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$
$\dim\{\mathcal{A}\}$	Dimension of a linear space $\mathcal{A}$
$\text{null}\{\mathbf{R}\}$	Null space of the matrix $\mathbf{R}$
$\lceil x \rceil$	Smallest integer not less than x
$\lfloor x \rfloor$	Largest integer not greater than x
$J_0(x)$	Zero-order Bessel function of the first kind
$\text{Var}(x)$	Variance of the random variable x
$\binom{n}{k}$	Number of k-combinations from a given set of n elements
$f(x) = o(g(x))$	Represents the fact that $\lim_{x\to\infty} f(x)/g(x) = 0$
$\mathcal{U}(a, b)$	Continuous uniform distribution on the interval $[a, b]$
ess inf$f$	Essential minimum of f, i.e., the infimum of f up to within a set of measure zero
ess sup$f$	Essential maximum of f, i.e., the supremum of f up to within a set of measure zero

Chapitre 1

Résumé [Français]

1.1 Abstract

Les systèmes d'antennes MIMO ayant un nombre élevé d'antennes, appelés "MIMO massifs", devraient permettre un débit et une efficacité énergétique beaucoup plus élevée par rapport aux systèmes MIMO traditionnels. Leur utilisation est fortement envisagée pour les futurs réseaux 5G. Malgré un fort potentiel, il y a encore de nombreux obstacles qui limitent le potentiel des systèmes MIMO massifs. Cette thèse se concentre sur certains de ces défis liés à l'acquisition de l'information de l'etat du canal (CSI) à la fois dans le mode de multiplexage temportel (TDD) et dans le mode de multiplexage fréquentiel (FDD).

Dans la phase d'estimation du canal en mode TDD, l'effet de contamination des pilotes constitue la limitation principale pour la performance globale. Nous présentons de nouvelles approches qui abordent ce problème en exploitant les statistiques de second ordre des canaux d'utilisateur. Nous démontrons analytiquement que dans la limite d'un grand nombre d'antennes, l'effet de contamination des pilotes disparaît complètement sous une certaine condition sur la matrice de covariance du canal. Cette condition stipule que le support des trajets multiples d'angle d'arrivée (AoA) des interférences doit être distinct du support des AoA du canal désiré. Cette condition est étroitement liée à la propriété de rang faible de la matrice de covariance. En outre, nous montrons qu'une telle propriété de faible rang n'est pas intrinsèquement liée à la géométrie du système d'antennes. Il peut être généralisé à matrice non uniforme, et de façon plus surprenante, à une répartition spatiale en deux dimensions.

Bien que l'estimateur de MMSE proposé conduise à la décontamination

complète sous la condition de non-chevauchement dans le domaine angulaire, dans la pratique, il est peu probable que cette condition soit tout le temps vérifiée, en raison de la distribution aléatoire des utilisateurs et des effets de diffusion. En conséquence, nous proposons de nouveaux systèmes robustes d'estimation de canal qui combinent les mérites de l'estimateur MMSE et la méthode de projection basée sur l'amplitude connue. L'analyse asymptotique lorsque le nombre d'antennes devient large, montre que les méthodes proposées exigent des conditions plus faibles par rapport aux méthodes connues pour réaliser une décontamination complète.

Enfin, nous abordons le problème de l'acquisition de l'information de canal pour les systèmes MIMO massifs fonctionnant en mode FDD par de nouveaux mécanismes de coopération entre utilisateurs qui sont activés par communications d'appareil à appareil (D2D). L'échange de l'information de canal entre les utilisateurs permet de d'exploiter cette connaissance partagée pour transmettre plus efficacement au transmetteur l'information de canal. L'impact de cette amélioration sur le débit total est évaluée et les gains vis-à-vis d'une configuration MIMO massifs classique sans D2D sont présentées.

1.2 Introduction

Les travaux présentés dans cette thèse s'articulent autour de défis liés à l'acquisition de l'information de canal en mode TDD et en mode FDD dans le cadre de systèmes MIMO massif. Plus précisément, le chapitre 3 présente quelques propriétés utiles pour les canaux en MIMO massifs. Dans les chapitres 4 et 5, nous abordons ensuite l'acquisition de l'information de canal des systèmes TDD, notamment le problème de la contamination des pilotes. Afin d'améliorer la qualité de l'estimation, de nouvelles méthodes d'estimation du canal sont proposées en se basant sur des propriétés du canal nouvellement identifiés dans le chapitre 3. Enfin, le chapitre 6 se concentre sur l'acquisition de l'information de canal dans le cadre de systèmes MIMO massif en mode FDD. Nous développons de nouvelles approches permettant de réduire le cout de l'acquisition de l'information de canal dans des systèmes FDD.

1.3 Propriétés des canaux en MIMO massifs

Une approche prometteuse pour faciliter l'acquisition de l'information de canal réside dans l'exploitation des propriétes de ce canal. En effet, dans les systèmes MIMO massif, certaines propriétés particulières du canal sans

fil peuvent être exploitées afin d'améliorer l'estimation de canal et réduire les interférences. Dans ce chapitre, nous mettons en évidence ces propriétés, qui seront ensuite exploitées dans les chapitres suivants.

1.3.1 Modèle de réseau

Considérons un réseau de L celulles où chaque station de base est équipée de M antennes. K utilisateurs ayant une seule antenne chacun sont servis simultannément dans chaque cellule par leur propre station de base. L'estimation de canal des utilisateurs vers la station de base est obtenue à la station de base par le biais de pilotes spécifiques permettant l'estimation de ce canal et éventuellement, par le biais de données émis par les utilisateurs. Le canal entre le k-ième utilisateur situé dans la l-ème cellule et la j-ème station de base est noté par $\mathbf{h}_{lk}^{(j)}$. Dans cette section, nous étudions les propriétés particulières des canaux MIMO massifs. En particulier, nous mettons en évidence la faible dimension de la matrice de covariance du canal, ce qui se montrera très utile pour les techniques d'atténuation d'interférences proposées dans cette thèse.

1.3.2 Modèle de rang faible pour les antennes ULA

Modèle de canal

Dans cette section, les canaux sans fil de l'ULA largement déployée seront étudiés. Le modèle classique [1] du canal à trajets multiples est adopté dans cette thèse :

$$\mathbf{h}_{lk}^{(j)} = \frac{\beta_{lk}^{(j)}}{\sqrt{P}} \sum_{p=1}^{P} \mathbf{a}(\theta_{lkp}^{(j)}) e^{i\varphi_{lkp}^{(j)}}, \tag{1.1}$$

où P est le nombre de chemins i.i.d., $e^{i\varphi_{lkp}^{(j)}}$ est la phase aléatoire i.i.d., et $\mathbf{a}(\theta)$ représente le vecteur de direction de la route en provenance de l'angle d'arrivée θ :

$$\mathbf{a}(\theta) \triangleq \begin{bmatrix} 1 \\ e^{-j2\pi \frac{D}{\lambda} \cos(\theta)} \\ \vdots \\ e^{-j2\pi \frac{(M-1)D}{\lambda} \cos(\theta)} \end{bmatrix}, \tag{1.2}$$

où λ est la longueur d'onde du signal et D l'espacement d'antenne. $\beta_{lk}^{(j)}$ est le coefficient de perte de trajet

$$\beta_{lk}^{(j)} = \sqrt{\frac{\alpha}{d_{lk}^{(j)^{\gamma}}}}, \tag{1.3}$$

γ dénote l'exposant d'attenuation, $d_{lk}^{(j)}$ est la distance géographique entre l'utilisateur et la j-ième station de base et α est une constante.

Propriété de rang faible pour les antennes ULA

Considérons un certain utilisateurs servi avec un angle ayant le support $\Phi = [\theta_{\min}, \theta_{\max}]$, ce qui signifie la fonction de densité de probabilité $p(\theta)$ de l'AoA pour ce canal d'utilisateur $\mathbf{h}$ satisfait $p(\theta) > 0$ si $\theta \in \Phi$ et $p(\theta) = 0$ si $\theta \notin \Phi$. Nous pouvons alors montrer le résultat suivant.

Théorème 1. *Le rang de la matrice de covariance de canal* $\mathbf{R}$ *satisfait :*

$$\frac{\text{rank}(\mathbf{R})}{M} \leqslant d, \ \textit{quand } M \textit{ est suffisamment grand} \ ,$$

où d *est défini par*

$$d \triangleq \big(\cos(\theta^{min}) - \cos(\theta^{max})\big)\frac{D}{\lambda}.$$

Théorème 1 montre que pour grand M, le noyau de la matrix de covariance, dénoté par null($\mathbf{R}$), est de dimension $(1-d)M$, ce qui peut être exploitée pour éliminer des interférences. Considérons maintenant un modèle de multi-trajets plus générale lorsque les AoAs correspondant au canal d'un utilisateur sont encore limitées, mais proviennent de plusieurs grappes disjointes.

Q dénote le nombre de groupes et $[\theta_q^{\min}, \theta_q^{\max}]$ représente le support des AoAs pour le q-ième groupement des chemins dans l'intervalle $[0, \pi]$. Ce scénario de transmission est représenté schématiquement dans Le graphique3.1 pour $Q = 2$. Nous pouvons ensuite définir l'union des supports :

$$\overline{\Phi} \triangleq \cup_{q=1}^{Q}[\theta_q^{\min}, \theta_q^{\max}], \tag{1.4}$$

En conséquence, la densité de probabilité $p(\theta)$ des AoAs satisfait $p(\theta) > 0$ sous la condition que $\theta \in \overline{\Phi}$ et $p(\theta) = 0$ quand $\theta \notin \overline{\Phi}$. Nous pouvons ainsi obtenir la généralisation du résultat précédent dans ce cadre plus général.

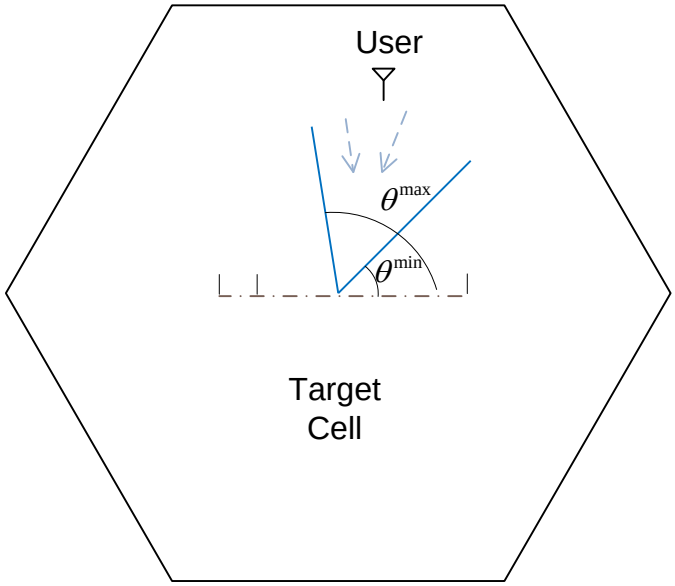

FIGURE 1.1 – Support borné de trajets multiples AoA.

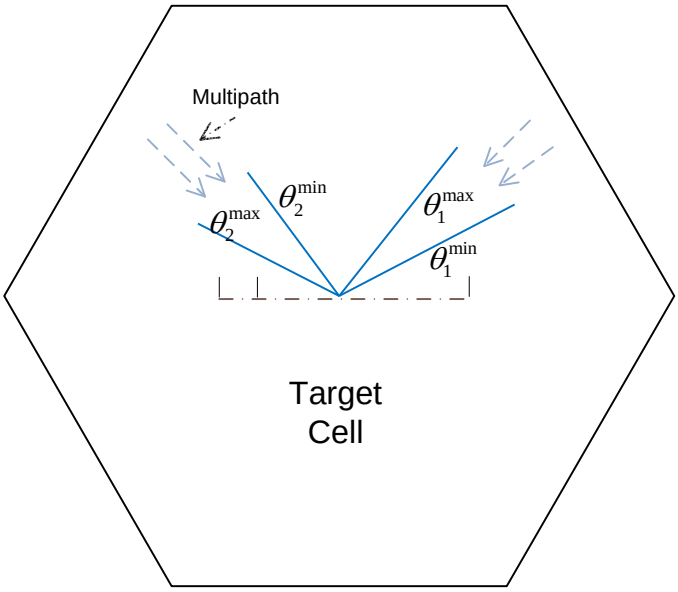

FIGURE 1.2 – Canal composé de $Q = 2$ groupements de trajets multiples.

Corollaire 1. *Le rang de la matrice de covariance de canal* $\mathbf{R}$ *satisfait :*

$$\frac{\text{rank}(\mathbf{R})}{M} \leqslant d, \text{ si } M \text{ est suffisamment grand },$$

où d *est défini par*

$$d \triangleq \min(1, \sum_{q=1}^{Q} \big(\cos(\theta_q^{min}) - \cos(\theta_q^{max})\big) \frac{D}{\lambda}).$$

1.3.3 Modèle de rang faible pour les réseaux d'antennes aléatoires

Nous considérons maintenant un ensemble d'antennes situées au hasard sur une ligne, et couvrant une ouverture totale de $\mathcal{D}$ mètres. Dans ce cas, un chemin élémentaire avec un angle θ peut être représenté par le vecteur de réponse tel que :

$$\mathbf{a}(\theta) \triangleq \begin{bmatrix} e^{-j2\pi \frac{d_1}{\lambda} \cos(\theta)} \\ \vdots \\ e^{-j2\pi \frac{d_M}{\lambda} \cos(\theta)} \end{bmatrix}, \tag{1.5}$$

où la position de la m-ième antenne ($1 \leq m \leq M$), d_m, suit une distribution uniforme, i.e., $d_m \sim \mathcal{U}(0, \mathcal{D})$.

1.3.4 Rang faible pour les réseaux aléatoires linéaires

Nous étudions ensuite le rang de la matrice de covariance de canal pour un réseau d'antenne linéaire aléatoire. L'espacement d'antenne moyen est alors dénoté par $\overline{D} \triangleq \mathcal{D}/M$. En supposant que l'ouverture du réseau d'antenne $\mathcal{D}$ augmente linéairement avec M, i.e., $\overline{D}$ est constant, les résultats sur la propriété de faible dimension sont montrés ci-dessous :

Proposition 1. *Avec un support borné des AoAs* $\overline{\Phi}$ *comme dans (3.4), le rang de la matrice de covariance de canal* $\mathbf{R}$ *satisfait :*

$$\text{rank}(\mathbf{R}) \leq \sum_{q=1}^{Q} \big(\cos(\theta_q^{min}) - \cos(\theta_q^{max})\big) \frac{M\overline{D}}{\lambda} + o(M), \tag{1.6}$$

Ce résultat ci-dessus montre que la propriété de rang faible dans les systèmes MIMO massif n'est pas dépendante à la structure de Fourier des vecteurs propres de la matrice de covariance. En outre, la borne supérieure

ci-dessus est en fait très serré pour les grandes valeurs de M, comme en témoignent les résultats des simulations dans le graphique 3.2, où nous prenons $Q = 1, \overline{D} = \lambda/2$. L'étallement des AoA est de 40 degrés, et le modèle se réfère à

$$f(M) \triangleq \sum_{q=1}^{Q} \left(\cos(\theta_q^{\min}) - \cos(\theta_q^{\max})\right) \frac{M\overline{D}}{\lambda}.$$

On peut observer que rank($\mathbf{R}$) est bien approximée par $f(M)$.

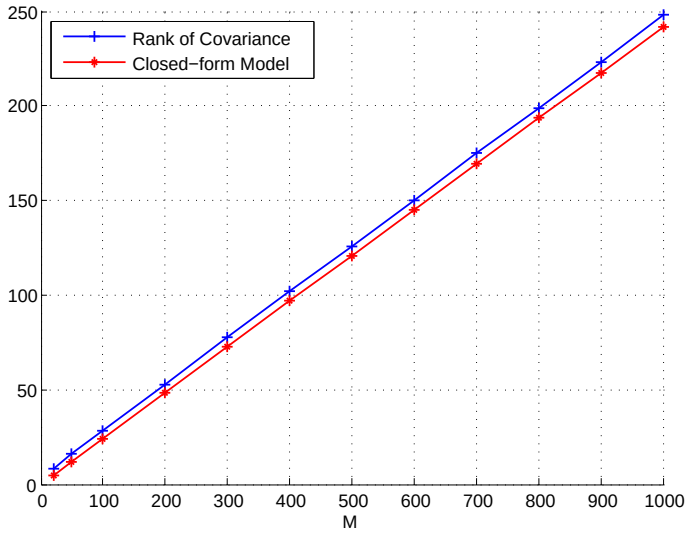

FIGURE 1.3 – Rang de la covariance du canalde pour modèle en forme fermée vs. rang actuel.

La Proposition 4 et le graphique 3.2 mettent en évidence l'additivité des rangs lorsque plusieurs groupes disjoints d'utilisateurs co-existent dans des systèmes MIMO massif, i.e., pour $M \to +\infty$.

1.3.5 Modèle de rang faible pour les systèmes DAS

Nous nous tournons maintenant vers les systèmes d'antennes distribuées. On suppose que les M antennes de la station de base se situent uniformément et aléatoirement dans un réseau de taille fixe, servant des utilisateurs équippés d'une seule antenne. Nous considérons une seule cellule en forme de disque de rayon R_c. Nous adoptons le classique modèle circulaire [2,3] où les utilisateurs sont entourées par un anneau de diffuseurs locaux (Voir Figure 1.4) situées à R mètres de l'utilisateur. Les positions des diffuseurs sont

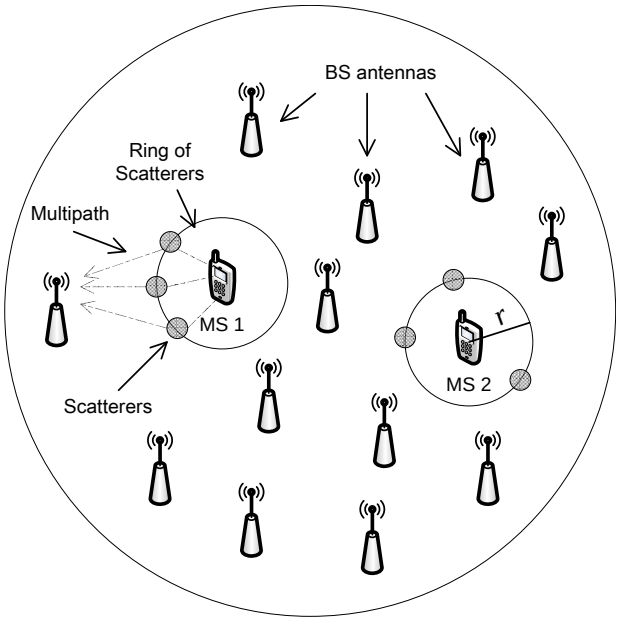

FIGURE 1.4 – Configuration distribuée dans le cadre d'une diffusion avec modèle circulaire.

considérés suivre une distribution uniforme sur l'anneau. De plus, chaque chemin de propagation rebondit une fois sur le diffuseur avant d'atteindre toutes les M destinations.

On considère ci-après un réseau (dense) dans laquelle la perte de trajet est non négligeable.

Théorème 2. *Le rang de la matrice de covariance de canal pour un système d'antenne distribués satisfait :*

$$rank(\mathbf{R}) \leq \frac{4\pi r}{\lambda} + o(r). \tag{1.7}$$

Le graphique3.4 montre le comportement du rang de la matrice de covariance par rapport au rayon de diffusion r. Nous pouvons voir que le rang varie linéairement avec la pente $4\pi/\lambda$. Toutefois, en raison du nombre limité d'antennes, le rang finit par saturer vers M quand r ne cesse d'augmenter.

1.3.6 Propriété de covariance uniformement bornée

Nous édutions maintenans comment borner la norme spectrale de la matrice de covariance du canal dans cadre d'un réseau d'antenne ULA. Cette

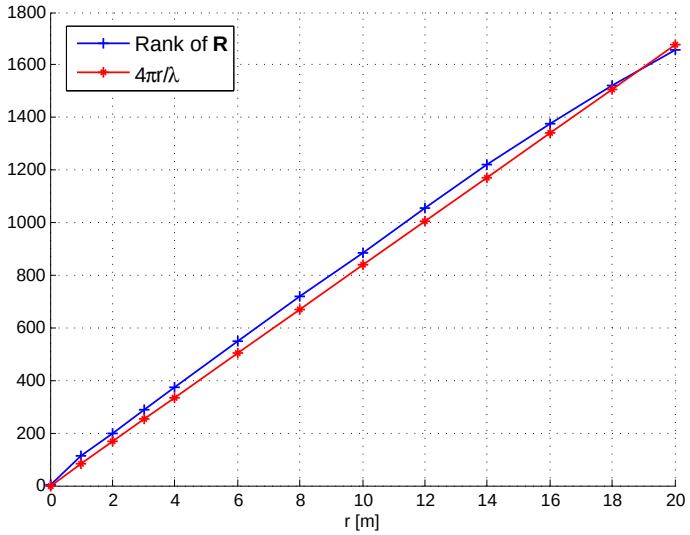

FIGURE 1.5 – Rank vs. r, $M = 2000$, $\lambda = 0.15$m, $R_c = 500$m.

propriété sera utile dans l'analyse du chapitre 5.

Proposition 2. *Denotons par* Φ *le support des AoA sd'un certain utilisateur et par* $p(\theta)$ *la fonction de densité de probabilité de AoA de l'utilisateur. si* $p(\theta)$ *est est uniformément bornée, i.e.,* $p(\theta) < +\infty, \forall \theta \in \Phi$, *et* Φ *se situe dans un intervalle fermé qui ne comprend pas les directions parallèles par rapport à la matrice, i.e.,* $0, \pi \notin \Phi$, *alors, la norme spectrale de covariance de l'utilisateur* $\mathbf{R}$ *est uniformément bornée :*

$$\forall M, \|\mathbf{R}\|_2 < +\infty. \tag{1.8}$$

1.3.7 Conclusions

Dans cette section, nous soulignons plusieurs propriétés fondamentales des chaînes des utilisateurs de MIMO massifs. Nous étudions les propriétés de covariance de sous-espaces de signaux en basse dimension dans des topologies générales de massifss array,y compris array uniforme linéaire, array linéaire aléatoire, et array distribuée en 2D.Une autre propriété révélé est le bornitude uniforme de la norme spectrale de covariance de canal pour ULA. Ces propriétés seront utiles dans les sections suivantes pour réduire les interférences et pour réduire la quantité de feedback CSI.

1.4 Estimation de canal basée sur la covariance

Dans cette section, nous exploitons le rang faible de la matrice de covariance afin de faciliter l'acquisition de l'information de canal dans les systèmes MIMO massif, dans le cadre d'un fonctionnement en TDD. Nous proposons une méthode d'estimation conduisant à une amélioration considérable de la performance. En particulier, nous mettons en évidence que la connaissance de la matrice de covariance peut conduire à une élimination complète des effets de la contamination de pilote dans la limite d'un grand nombre d'antennes à la station de base. Un algorithme permettant l'exploitation de ce concepte dans des scénarios pratiques est ensuite présenté. L'idée principale est d'utiliser la connaissance des matrices de covariance lors de l'affectation des pilotes d'estimation du canal.

1.4.1 Apprentissage du canal pour la connection montante

Pour faciliter la compréhension, nous considérons dans le reste de la section qu'un seul utilisateur par cellule est présent. La suite de pilotes utilisée dans la l-ième cellule est dénotée par :

$$\mathbf{s}_\ell = [\ s_{\ell 1} \quad s_{l2} \quad \cdots \quad s_{\ell\tau}\ ]^T. \tag{1.9}$$

Les puissances allouées aux séquences de pilotes sont supposées égales de telle sorte que $|s_{l1}|^2+\cdots+|s_{l\tau}|^2 = \tau, l = 1, 2, \ldots, L$. Sans perte de généralité, nous supposons que la première cellule est la cellule cible de telle sorte que nous puissions omettre l'exposant du canal. Le canal entre l'utilisateur de la ℓ-ième cellule et la station de base cible est donc dénoté par $\mathbf{h}_\ell$.

Lors de la transmission des pilotes, le signal $M \times \tau$ reçu à la station de base cible est donc

$$\mathbf{Y} = \sum_{l=1}^{L} \mathbf{h}_l \mathbf{s}_l^T + \mathbf{N}, \tag{1.10}$$

où $\mathbf{N} \in \mathbb{C}^{M\times\tau}$ est le bruit blanc additif Gaussien, centré et de variance σ_n^2.

1.4.2 Estimation bayésienne du canal

Par la suite, nous développons des estimateurs de canal se basant sur les statistiques de second ordre des canaux. Deux estimateurs bayésiens de canal peuvent ainsi être construits. Dans un premier temps, tous les canaux sont estimés à la station de base cible (y compris les interférents). Dans le second temps, seulement $\mathbf{h}_1$ est estimé.

En représentation vectorielle, notre modèle (5.3) peut être représenté comme

$$\mathbf{y} = \widetilde{\mathbf{S}}\mathbf{h} + \mathbf{n}, \tag{1.11}$$

où $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\mathbf{n} = \text{vec}(\mathbf{N})$, and $\mathbf{h} \in \mathbb{C}^{LM\times 1}$ est obtenu en écrivant successivement les L canaux. La matrice des pilotes, dénotée par $\tilde{\mathbf{S}}$, est ainsi définie par

$$\widetilde{\mathbf{S}} \triangleq \left[\begin{array}{ccc} \mathbf{s}_1 \otimes \mathbf{I}_M & \cdots & \mathbf{s}_L \otimes \mathbf{I}_M \end{array} \right]. \tag{1.12}$$

Utilisant la méthode du maximum a posteriori (MAP), l'estimation bayésienne est donnée par :

$$\widehat{\mathbf{h}} = (\sigma_n^2 \mathbf{I}_{LM} + \mathbf{R}\widetilde{\mathbf{S}}^H\widetilde{\mathbf{S}})^{-1}\mathbf{R}\widetilde{\mathbf{S}}^H\mathbf{y}. \tag{1.13}$$

Il est remarquable que l'estimation bayésienne dans (4.13) coïncide avec l'erreur quadratique moyenne minimum (MMSE) lorsque le canal $\mathbf{h}$ suit la distribution gaussienne complexe.

Considerons le scénario pessimiste où une seule séquence de pilotes est utilisée dans les L cellules et représentée par

$$\mathbf{s} = \left[\begin{array}{cccc} s_1 & s_2 & \cdots & s_\tau \end{array} \right]^T. \tag{1.14}$$

Nous définissons ensuite la matrice d'apprentissage $\bar{\mathbf{S}} \triangleq \mathbf{s} \otimes \mathbf{I}_M$. Il s'ensuite alors que $\bar{\mathbf{S}}^H\bar{\mathbf{S}} = \tau\mathbf{I}_M$. Le signal vectorisé d'apprentissage reçue à la station de base cible peut donc être exprimé comme

$$\mathbf{y} = \bar{\mathbf{S}}\sum_{l=1}^{L}\mathbf{h}_l + \mathbf{n}. \tag{1.15}$$

L'estimateur bayésien (équivalent à l'estimateur MMSE) pour le canal spécifique $\mathbf{h}_1$ est ainsi représenté par :

$$\widehat{\mathbf{h}}_1 = \mathbf{R}_1\bar{\mathbf{S}}^H\left(\bar{\mathbf{S}}\left(\sum_{l=1}^{L}\mathbf{R}_l\right)\bar{\mathbf{S}}^H + \sigma_n^2\mathbf{I}_{\tau M}\right)^{-1}\mathbf{y} \tag{1.16}$$

$$= \mathbf{R}_1\left(\sigma_n^2\mathbf{I}_M + \tau\sum_{l=1}^{L}\mathbf{R}_l\right)^{-1}\bar{\mathbf{S}}^H\mathbf{y}. \tag{1.17}$$

Dans la section ci-dessous, nous examinons la dégradation causée par la contamination de pilote sur la qualité de l'estimation.

Analyse à grande échelle

Nous cherchons à analyser la performance des estimateurs ci-dessus dans le régime d'un grand nombre d'antennes M. Afin d'obtenir des resultats analytiques, notre analyse est basée sur l'hypothèse d'ULA.

Théorème 3. *Supposons que l'angle arrivée θ des trajets multiples du canal $\mathbf{h}_j, j = 1, \ldots, L$, est distribué selon une densité arbitraire $p_j(\theta)$ ayant un support borné, i.e., $p_j(\theta) = 0$, $\theta \notin [\theta_j^{\min}, \theta_j^{\max}]$, pour $\theta_j^{\min} \leqslant \theta_j^{\max} \in [0, \pi]$. Si les $L-1$ intervalles $[\theta_i^{\min}, \theta_i^{\max}]$, $i = 2, \ldots, L$ sont strictement distincts du support de l'angle d'arrivé du canal direct*[1] *$[\theta_1^{\min}, \theta_1^{\max}]$, on a*

$$\lim_{M \to \infty} \widehat{\mathbf{h}}_1 = \widehat{\mathbf{h}}_1^{no\ int}. \tag{1.18}$$

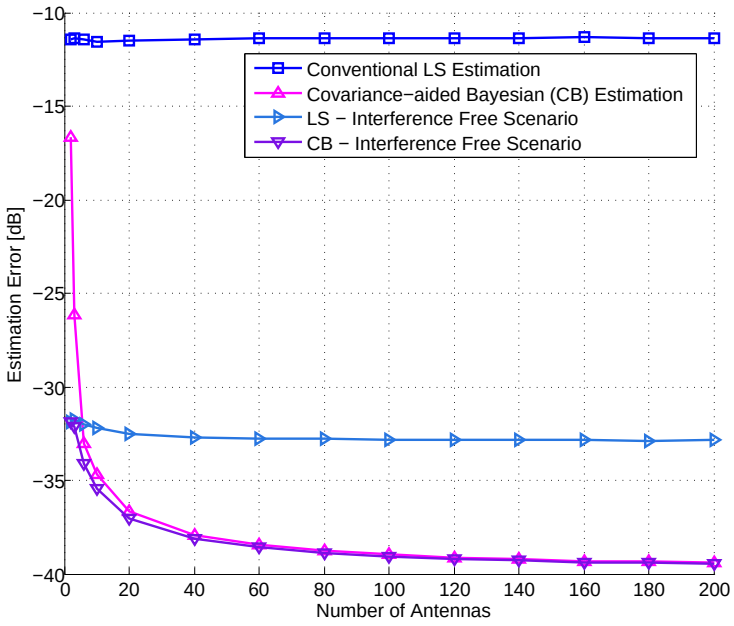

FIGURE 1.6 – Estimation MSE vs. nombre BS antennes , 2-cellule network, deux utilisateurs en position fixée, AOA uniformément distribuées avec $\theta_\Delta = 20$ degrés, nans chevauchement trajets multiples.

1. Cette condition correspond simplement à un scénario pratique conduisant à des sous-espaces de signaux distincts entre la matrice de covariance souhaitée et la matrice de covariance des interférences. Cependant, des scénarios plus généraux pourraient être utilisés.

Les résultats présentés dans le Theorème 3 sont vérifiés numériquement dans Le graphique4.1 dans le cas d'un réseau composé de deux cellules et où les positions de deux utilisateurs sont fixées. Les angles d'arrivées du canal direct sont uniformément distribuées avec une moyenne égale à 90 degrés et avec un étalement angulaire égale à 20 degrés, ce qui donne pas de chevauchement entre les multi-trajets souhaités et celui de l'interférence. Comme on le voit, l'erreur d'estimation de canal converge vers le cas sans interférence, ce qui indique que la contamination des pilotes est éliminée avec le nombre d'antennes en croissant.

Theorème 3 donne la condition suffisante pour atteindre la suppression totale de l'interférence lorsque M est grand :

$$\bigcup_{l=2}^{L} \text{span}\left\{\mathbf{R}_l\right\} \subset \text{null}\left\{\mathbf{R}_1\right\} \tag{1.19}$$

où la condition ci-dessus nécessite que la matrice de covariance du canal désiré aie un noyau non-vide et que les sous-espaces engendrés par les matrices de covariance de toutes les interferences tombent dans ce noyau.

Le Theorème 3 est limité au cas où le support de l'angle d'arrivé du canal direct est un unique interval. Nous allons donner ci-dessous les résultats étendus dans des conditions moins restrictives.

Cas de plusieurs groupe d'angles d'arrivé

Nous considérons maintenant le modèle général de propagation par trajets multiples lorsque les angles d'arrivé du canal d'un certain utilisateur sont limitées, mais proviennent de plusieurs groupes disjoints, comme décrit dans la Section 1.3.2. Soit $\overline{\Phi}_d$ l'union des angles d'arrivés possibles du canal direct et $\overline{\Phi}_i$ l'union des angles d'arrivé de l'interférence. Nous avons le résultat suivant pour le cas d'un réseau d'antenne uniforme massif :

Corollaire 2. *Soit $D \leq \lambda/2$ and $\overline{\Phi}_d \cap \overline{\Phi}_i = \emptyset$. L'estimation MMSE pour (1.17) satisfait :*

$$\lim_{M\to\infty} \widehat{\mathbf{h}}_1 = \widehat{\mathbf{h}}_1^{no\ int}. \tag{1.20}$$

1.4.3 Affectation coordonnée des pilotes

Dans cette section, nous concevons un protocole de coordination permettant de distribuer les séquences de pilotes aux utilisateurs. Le rôle de la coordination est d'optimiser l'utilisation des matrices de covariance dans le but de satisfaire la condition de non chevauchement des supports des angles

d'arrivé. Nous utilisons l'estimation de MMSE comme métrique de performance afin de trouver le meilleur groupe d'utilisateur. Le principe de la distribution des pilotes coordonnée consiste à exploiter la connaissance des matrices de covariance.

Nous proposons une approche gloutonne/gourmande classique appelé Affectation Coordonnée des Pilotes (CPA). La coordination peut être interprétée comme suit : Pour minimiser l'erreur d'estimation, une station de base tend à affecter un pilote donné à l'utilisateur dont la caractéristique spatiale diffère fortement avec les utilisateurs interférents utilisant le même pilote.

Deux types de distributions de l'angle d'arrivé sont considérées ici, un non-borné (Gaussien) et un borné (uniforme) :

Distribution Gaussienne

Pour certain canal $\mathbf{h}_{\ell k}^{(j)}$, les angles d'arrivé pour tous les trajets P sont i.i.d. Gaussiennes de moyenne $\bar{\theta}_{\ell k}^{(j)}$ et de variance σ^2.

Notez que les distributions Gaussienne des angles d'arrivé ne peuvent pas satifaire les conditions de non chevauchement des supports des angles d'arrivé du Théorème 3, néanmoins l'utilisation de la méthode proposée dans ce contexte donne également des gains substantiels lorsque σ^2 diminue.

Distribution uniforme

Pour les canaux $\mathbf{h}_{\ell k}^{(j)}$, les angles d'arrivé sont uniformément réparties sur $[\bar{\theta}_{lk}^{(j)} - \theta_\Delta, \bar{\theta}_{lk}^{(j)} + \theta_\Delta]$, où $\bar{\theta}_{\ell k}^{(j)}$ est la moyenne des angles d'arrivé.

Dans ces figures, "LS" dénote l'estimation classique de canal en utilisant la méthode des moindres carrés. "CB" dénote l'estimation Bayésienne reposant sur la connaissance de la matrice de covariance (sans affectation coordonnée des pilotes), et "CPA" est l'estimation Bayésienne basée sur Affectation Pilote Coordonnée proposée dans cette thèse.

Dans le graphique 4.3 et le graphique 4.4, l'estimation MSE par rapport au nombre d'antennes est illustrée. Quand les angles d'arrivés ont des distributions uniformes avec $\theta_\Delta = 10$ degrés, comme dans Le graphique4.3, la performance de l'estimateur CPA est améloirée si M varie de 2 à 10. Le graphique4.4 est obtenue avec une distribution Gaussienne des angles d'arrivé. Nous pouvons aussi observer un écart entre la CPA et l'estimateur sans interference, en raison de la distribution Gaussienne qui ne verifie pas la condition sur les bornes. Néanmoins, les gains sur l'estimateur classique restent importants.

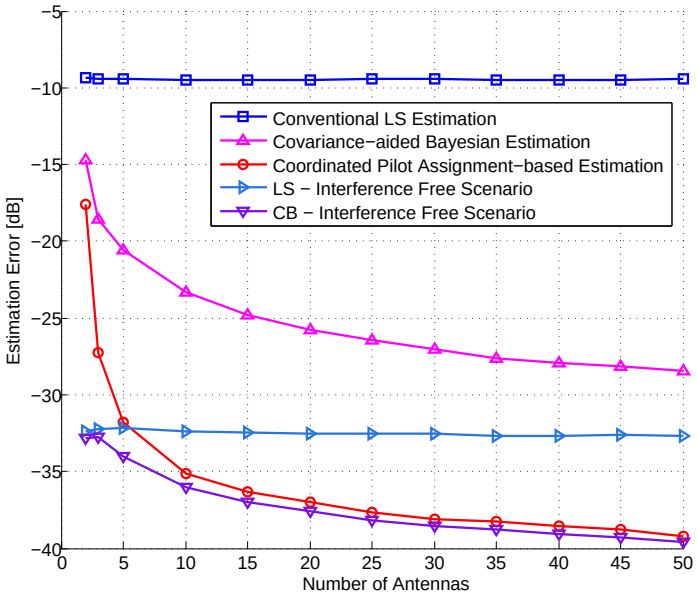

FIGURE 1.7 – Estimation MSE vs. nombre de BS antennes, AOA uniformément distribuées avec $\theta_\Delta = 10$ degrés, 2-cellules network.

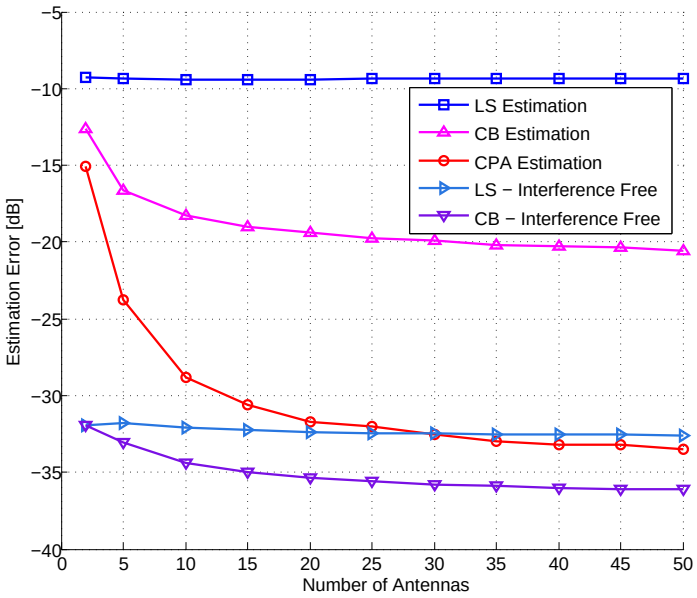

FIGURE 1.8 – Estimation MSE vs. number of BS antennas, AOA distribués gaussienne $\sigma = 10$ degrees, 2-cell network.

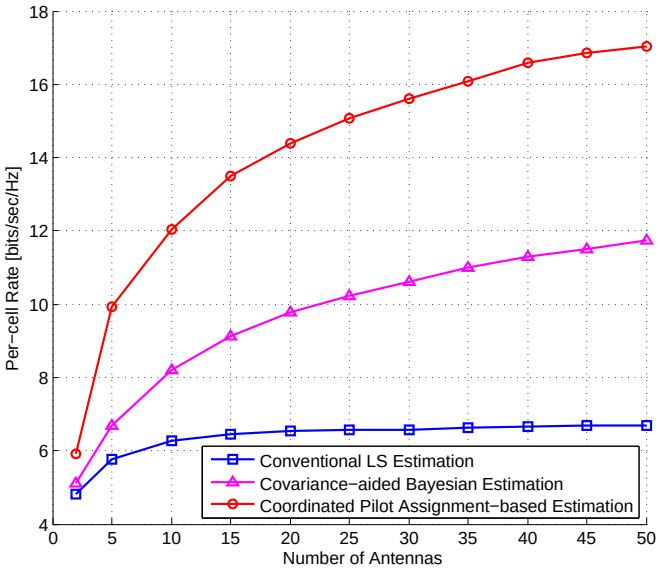

FIGURE 1.9 – Taux par cellule vs. nombre de BS antennes, 2-cellules network, AOA distribués gaussienne avec $\sigma = 10$ degrés.

Le graphique4.6 décrit le débit de la connection descendante pour une cellule, obtenu par la stratégie de precodage adapté et comfirme les gains obtenus lorsque l'estimation Bayésienne est utilisé en conjonction avec la stratégie d'affectation coordonnée des pilotes (CPA) proposée et des gains intermédiaires quand il est utilisé tout seul.

1.4.4 Conclusions

Cette partie propose une métode d'estimation du canal se basant sur la connaissance de la matrice de covariance dans le contexte de systèmes multi-cellulaires et d'antennes multiples dans le cas d'un régime limité par les interférences. Nous développons des estimateurs Bayésiens et démontrons analytiquement l'efficacité d'une telle approche pour les systèmes ayant un grand nombre d'antennes, conduisant à une élimination complète des effets de la contamination de pilote dans le cas de matrices de covariance satisfaisant une certaine condition de non-chevauchement des supports des angles d'arrivé. Nous proposons une stratégie d'affectation coordonnée des pilotes (CPA) qui aide à former des matrices de covariance verifiant la condition nécessaire et mettons en évidence que des performances proches de l'estimation sans interférence peuvent être atteintes.

1.5 Décontamination basée sur la puissance et les angles

Dans la section précédente, nous avons montré plusieurs techniques d'atténuation des interférences se basant sur la connaissance statistique des canaux des utilisateurs. Plus particulièrement, nous avons démontré que dans le cas où le support des angles d'arrivé du canal direct ne se chevauche pas avec le support des angles d'arrivés des utilisateurs interférents, une méthode basée sur l'estimation MMSE atteint l'élimination totale des interférences. Cependant, quand une telle condition de non-chevauchement n'est pas satisfaite, une interférence résiduelle existe lors de l'estimation MMSE, ce qui limite la performance des système MIMO massif. Dans cette section, nous cherchons à résoudre ce type de problème en exploitant les statistiques à court terme du canal.

1.5.1 Transmission de données

Soit un réseau multi-cellules multi-utilisateur ayant K utilisateurs par cellule, et chacun desservi par sa propre station de base. Le canal multi-utilisateur MIMO entre les K utilisateurs dans la cellule ℓ et la station de base j est :

$$\mathbf{H}_{\ell}^{(j)} \triangleq \begin{bmatrix} \mathbf{h}_{\ell 1}^{(j)} & \mathbf{h}_{\ell 2}^{(j)} & \cdots & \mathbf{h}_{\ell K}^{(j)} \end{bmatrix}, \tag{1.21}$$

et la matrice des pilote se composant de toutes les séquences de pilotes utilisés par ces K utilisateurs est :

$$\mathbf{S} \triangleq \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \cdots & \mathbf{s}_K \end{bmatrix}^T. \tag{1.22}$$

Pendant la phase d'apprentissage, le signal reçu à la station de base j est

$$\mathbf{Y}^{(j)} = \sum_{\ell=1}^{L} \mathbf{H}_{\ell}^{(j)} \mathbf{S} + \mathbf{N}^{(j)}, \tag{1.23}$$

où $\mathbf{N}^{(j)} \in \mathbb{C}^{M \times \tau}$ est le spatialement et temporellement blanc bruit gaussien additif (AWGN) avec moyenne nulle et variance élément par élément σ_n^2.

Ensuite, pendant la phase de transmission de données de liaison montante, chaque utilisateur transmet C symboles de données. Le signal de données reçu à la station de base j est donné par :

$$\mathbf{W}^{(j)} = \sum_{\ell=1}^{L} \mathbf{H}_{\ell}^{(j)} \mathbf{X}_{\ell} + \mathbf{Z}^{(j)}, \tag{1.24}$$

où $\mathbf{X}_\ell \in \mathbb{C}^{K\times C}$ est la matrice de symboles émis par tous les utilisateurs dans la ℓ-ième cellule. Les symboles sont i.i.d. avec une moyenne nulle et une variance unitaire. La matrice $\mathbf{Z}^{(j)} \in \mathbb{C}^{M\times C}$ a ses éléments i.i.d Gaussien avec moyenne nulle et variance σ_n^2.

1.5.2 Méthode basée sur la projection dans le domaine de l'amplitude

Nous proposons ci-dessous de nouvelles méthodes d'estimation permettant une estimation plus robust dans un scénario cellulaire réaliste.

Un seul utilisateur par cellule

Par soucis de clareté, nous considérons d'abord un scénario simple où chaque cellule a un seul utilisateur, à savoir $K = 1$. Les utilisateurs dans des cellules différentes partagent la même séquence de pilote $\mathbf{s}$.

On introduit un filtre $\mathbf{\Xi}_j$, qui est basé sur la connaissance des matrices de covariance de canal d'une manière similaire à celle utilisée par le filtre MMSE.

$$\mathbf{\Xi}_j = \left(\sum_{l=1}^{L} \mathbf{R}_l^{(j)} + \sigma_n^2 \mathbf{I}_M\right)^{-1} \mathbf{R}_j^{(j)}. \tag{1.25}$$

Le filtre spatial est appliqué au signal reçu à la station de base j comme

$$\widetilde{\mathbf{W}}_j \triangleq \mathbf{\Xi}_j \mathbf{W}^{(j)}. \tag{1.26}$$

La méthode basée sur l'amplitude peut maintenant être appliqué sur les données reçues filtrées pour se débarrasser de l'interférence résiduelle. Prendre le vecteur propre correspondant à la plus grande valeur propre de la matrice $\widetilde{\mathbf{W}}_j \widetilde{\mathbf{W}}_j^H / C$:

$$\widetilde{\mathbf{u}}_{j1} = \mathbf{e}_1\{\frac{1}{C}\widetilde{\mathbf{W}}_j \widetilde{\mathbf{W}}_j^H\}. \tag{1.27}$$

Donc $\widetilde{\mathbf{u}}_{j1}$ peut être considérée comme une estimation de la direction du vecteur $\mathbf{\Xi}_j \mathbf{h}_j^{(j)}$.

Nous annulons alors l'effet de la multiplication par la matrice $\mathbf{\Xi}_j$ en utilisant

$$\mathbf{\Xi}_j{}' \triangleq \mathbf{R}_j^{(j)\dagger} \left(\sum_{l=1}^{L} \mathbf{R}_l^{(j)} + \sigma_n^2 \mathbf{I}_M\right), \tag{1.28}$$

et on obtient une estimation de la direction du vecteur de canal $\mathbf{h}_j^{(j)}$ comme la suite :

$$\overline{\mathbf{u}}_{j1} = \frac{\boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1}}{\left\| \boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1} \right\|_2}. \tag{1.29}$$

Enfin, les ambiguïtés de phase et d'amplitude du canal peuvent être résolus par la projection de l'estimation selon la méthode des moindres carrés sur le sous-espace engendré par $\overline{\mathbf{u}}_{j1}$:

$$\widehat{\mathbf{h}}_j^{(j)\mathrm{CA}} = \frac{1}{\tau} \overline{\mathbf{u}}_{j1} \overline{\mathbf{u}}_{j1}^H \mathbf{Y}^{(j)} \mathbf{s}^*, \tag{1.30}$$

où "CA" désigne la projection dans le domaine d'amplitude en covariance.

L'algorithme est résumé ci-dessous :

Algorithm 1 Covariance-aided Amplitude based Projection

1: Prenez le premier vecteur propre de $\widetilde{\mathbf{W}}_j \widetilde{\mathbf{W}}_j^H / C$ comme dans (1.27), avec $\widetilde{\mathbf{W}}_j$ étant le signal de données filtré.
2: Inverser l'effet du filtre spatial utilisant(1.29).
3: Résoudre les ambiguïtés de phase et d'amplitude par(1.30).

Les performances asymptotique de l'estimateur ci-dessus sont analysées théoriquement. Afin de faciliter l'analyse, nous introduisons la condition suivante :

Condition C1 : La norme spectrale de $\mathbf{R}_l^{(j)}$ est uniformément bornée :

$$\forall M \in \mathbb{Z}^+ \text{ et} \forall l \in \{1, \ldots, L\}, \exists \zeta, \text{s.t.} \left\| \mathbf{R}_l^{(j)} \right\|_2 < \zeta, \tag{1.31}$$

où $\mathbb{Z}^+$ est l'ensemble des nombres entiers positifs, et ζ est une constante.

Performance asymptotique de l'estimateur CA

Soit

$$\alpha_l^{(j)} \triangleq \lim_{M \to \infty} \frac{1}{M} \operatorname{tr}\{\boldsymbol{\Xi}_j \mathbf{R}_l^{(j)} \boldsymbol{\Xi}_j^H\}, \forall l = 1, \ldots, L. \tag{1.32}$$

Théorème 4. *Compte tenu de la condition C1, si l'inégalité suivante est vraie :*

$$\alpha_j^{(j)} > \alpha_l^{(j)}, \forall l \neq j, \tag{1.33}$$

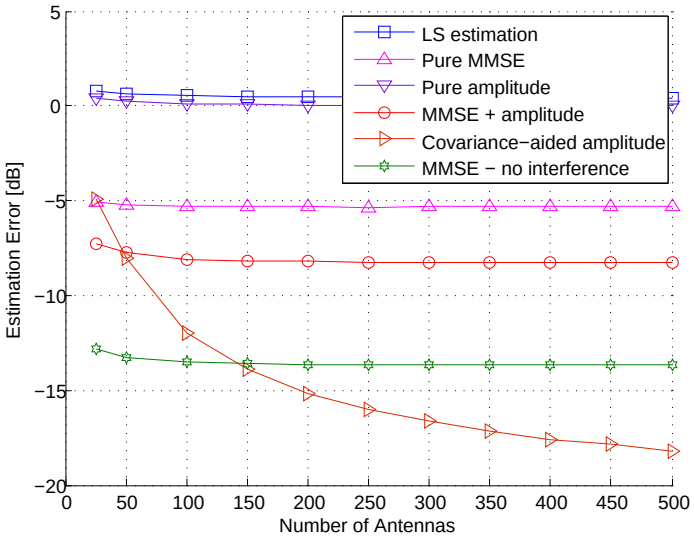

FIGURE 1.10 – Estimation des performance versus M pour un réseau à deux-cellules avec un utilisateur par cellule, un exposant d'atténuation $\gamma = 0$ et un support angulaire recouvrant partiellement les angles d'arrivés à hauteur de 60 degrés, $C = 500$, SNR $= 0$ dB.

alors, l'erreur d'estimation de (1.30) disparaît :

$$\lim_{M,C\to\infty} \frac{\left\|\widehat{\mathbf{h}}_j^{(j)CA} - \mathbf{h}_j^{(j)}\right\|_2^2}{\left\|\mathbf{h}_j^{(j)}\right\|_2^2} = 0. \tag{1.34}$$

De plus, la condition (1.33) du Theorème 4 peut alors être remplacé par

$$\left\|\mathbf{\Xi}_j \mathbf{h}_j^{(j)}\right\|_2 > \left\|\mathbf{\Xi}_j \mathbf{h}_l^{(j)}\right\|_2, \forall l \neq j. \tag{1.35}$$

Nous illustrons maintenant le Theorème 4 dans le graphique 5.1. Supposons qu'on aie un réseau composé de deux cellules et que chaque cellule aie deux utilisateurs. L'exposant d'attenuation est $\gamma = 0$, i.e., la puissance de l'interférence de canal est la même que celle du canal direct. Les supports angulaires à trajets multiples de l'interférence et le canal désiré se superposent donc à hauteur de 50 Dans cette courbe, "Pure amplitude" dénote la méthode de projection basée seulement sur l'amplitude. "MMSE + amplitude" représente l'estimateur à faible complexité proposé dans [4]. "Covariance-aided amplitude" dénote la méthode de projection utilisant l'amplitude et

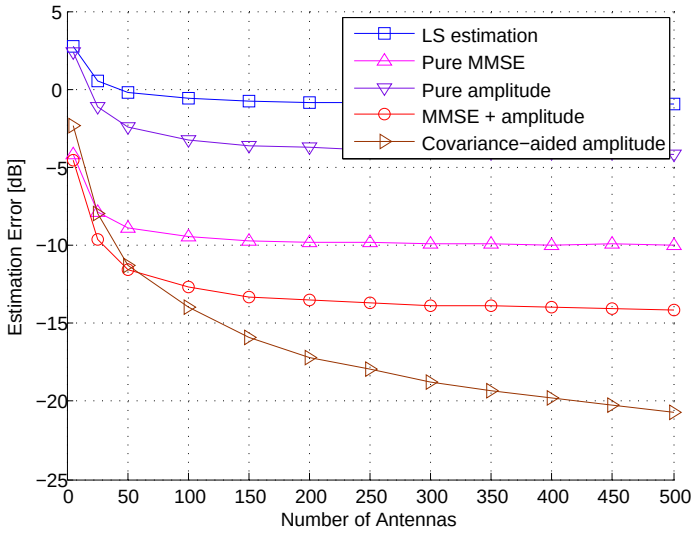

FIGURE 1.11 – Estimation performance vs. M, 7-cellules network, 1 utilisateurs par cellule, AoA réparties 60 degrés, exposant de perte de trajet $\gamma = 2$, cellule-bord SNR = 0 dB.

la connaisance de la matrice de covariance (1.30). La courbe "MMSE - no interference" montre l'erreur d'estimation de l'estimateur MMSE dans un scénario sans interférence et sert ainsi de réference. Comme on peut le voir sur la courbe 5.1, la méthode de projection utlisant l'amplitude et la covariance surpasse la méthode d'estimation MMSE sans interférences.

Dans la courbe 5.2, nous considérons un réseau composé de 7 cellules, avec un seul utilisateur par cellule. Les utilisateurs sont supposés être distribués de façon aléatoire et uniforme au sein de leurs propres cellules. La propagation angulaire du canal d'utilisateur est de 30 degrés. L'exposant d'attenuation est fixé à $\gamma = 2$. Il est possible d'observer comment la méthode de projection exploitant à la fois la connaissance de l'amplitude et de la matrice de covariance est plus performante que les autres méthodes.

Généralisation à plusieurs utilisateurs par cellule

Nous appliquons maintenant la méthode de projection basé sur l'amplitude et la connaissance de la matrice de covariance dans le cas général, c'est à dire lorsque K utilisateurs sont servis simultanément dans chaque cellule. Nous considérons l'estimation du canal d'utilisateur $\mathbf{h}_{jk}^{(j)}$ dans la suite de cette section.

Pour simplifier les notations, nous introduisons la notation $\mathbf{H}_{j\backslash k}^{(j)}$ pour representer $\mathbf{H}_{j}^{(j)}$ après avoir enlevé sa k-ième colonne :

$$\mathbf{H}_{j\backslash k}^{(j)} \triangleq \left[\mathbf{h}_{j1}^{(j)} \quad \cdots \quad \mathbf{h}_{j(k-1)}^{(j)} \quad \mathbf{h}_{j(k+1)}^{(j)} \quad \cdots \quad \mathbf{h}_{jK}^{(j)}\right]. \tag{1.36}$$

L'estimation correspondant à (1.36), dénotée par $\widehat{\mathbf{H}}_{j\backslash k}^{(j)}$, est obtenue en éliminant la k-ième colonne de $\widehat{\mathbf{H}}_{j}^{(j)}$, qui est une estimation selon la méthode des moindre carrés de $\mathbf{H}_{j}^{(j)}$.

Nous neutralisons tout d'abord l'interférence intra-cellulaire avec un filtre dit de Zero-Forçage (ZF) et dénoté par $\mathbf{T}_{jk}$ en utilisant l'estimée $\widehat{\mathbf{H}}_{j\backslash k}^{(j)}$. Ensuite, le filtre spatial $\mathbf{\Xi}_{jk}$ est appliquée. A l'issue de l'application de ces 2 filtres, le signal obtanu est :

$$\widetilde{\mathbf{W}}_{jk} \triangleq \mathbf{\Xi}_{jk}\mathbf{T}_{jk}\mathbf{W}^{(j)}, \tag{1.37}$$

où

$$\mathbf{T}_{jk} \triangleq \mathbf{I}_M - \widehat{\mathbf{H}}_{j\backslash k}^{(j)}(\widehat{\mathbf{H}}_{j\backslash k}^{(j)H}\widehat{\mathbf{H}}_{j\backslash k}^{(j)})^{-1}\widehat{\mathbf{H}}_{j\backslash k}^{(j)H}, \tag{1.38}$$

et

$$\mathbf{\Xi}_{jk} \triangleq \left(\sum_{l=1}^{L}\mathbf{R}_{lk}^{(j)} + \sigma_n^2\mathbf{I}_M\right)^{-1}\mathbf{R}_{jk}^{(j)}. \tag{1.39}$$

Hormis ces modifications, la méthode décrite dans le cas d'un seul utilisateur peut être appliquée.

Soit le vecteur propre correspondant à la plus grande valeur propre de la matrice $\widetilde{\mathbf{W}}_{jk}\widetilde{\mathbf{W}}_{jk}^{H}/C$:

$$\widetilde{\mathbf{u}}_{jk1} = \mathbf{e}_1\{\frac{1}{C}\widetilde{\mathbf{W}}_{jk}\widetilde{\mathbf{W}}_{jk}^{H}\}. \tag{1.40}$$

L'estimation de la direction de $\mathbf{h}_{jk}^{(j)}$ est obtenue par :

$$\overline{\mathbf{u}}_{jk1} = \frac{\mathbf{\Xi}'_{jk}\widetilde{\mathbf{u}}_{jk1}}{\left\|\mathbf{\Xi}'_{jk}\widetilde{\mathbf{u}}_{jk1}\right\|_2}, \tag{1.41}$$

où

$$\mathbf{\Xi}'_{jk} \triangleq \mathbf{R}_{jk}^{(j)\dagger}\left(\sum_{l=1}^{L}\mathbf{R}_{lk}^{(j)} + \sigma_n^2\mathbf{I}_M\right). \tag{1.42}$$

Enfin, les ambiguïtés de phase et d'amplitude sont résolu durant la séquence d'apprentissage pour obtenir l'estimation de $\mathbf{h}_{jk}^{(j)}$:

$$\widehat{\mathbf{h}}_{jk}^{(j)\mathrm{CA}} = \frac{1}{\tau}\overline{\mathbf{u}}_{jk1}\overline{\mathbf{u}}_{jk1}^{H}\mathbf{Y}^{(j)}\mathbf{S}^{H}. \tag{1.43}$$

Notez que dans cette méthode, nous construisons un filtre de type ZF $\mathbf{T}_{jk}$ en se basant sur une estimation des moindres carréess. D'autres améliorations peuvent être obtenus avec une estimation de plus grande qualité au prix d'une complexité plus élevé.

1.5.3 Conclusions

Dans cette section, nous avons proposé plusieurs algorithmes robustes d'estimation du canal en exploitant la diversité des trajets dans les domaines d'angulaires et d'amplitude. La méthode de projection utilisant l'amplitude et la covariance est robuste même dans le cas où le canal désiré et les canaux d'interférence ont des angles d'arrivée ayant des supports se supperposant, et ne sont pas séparables seulement en termes de puissance. L'analyse asymptotique montre les conditions dans lesquelles l'erreur d'estimation de canal converge vers zéro.

1.6 Feedback coopératif pour le FDD

Dans la section précédente, nous avons abordé l'un des problèmes fondamentaux de système MIMO massif : la contamination de pilote. Dans cette section, nous traitons un autre défi - l'acquisition de l'information pour les systèmes MIMO massifs dans le cas d'une transmission dans le mode FDD. De manière à réduire la quantité d'information à transmettre dans les système non-réciproques, nous proposons une methode exploitant une coopération entre les utilisateurs. Plus précisemment, nous introduisons une phase supplémentaire à la tranmsmission classique des informations de canal durant laquelle les utilisateurs échangent entre eux l'information de canal obtenue locallement. Cette information est ensuite transmisme coopérativement afin de réduire la quantité d'information à transmettre.

1.6.1 Signal et modèles de canal pour FDD

Nous considérons une station de base MIMO massif desservant un groupe de K utilisateurs ayant chacun une seule antenne. La station de base a M antennes et fonctionne en mode FDD. Le canal de liaison descendante entre

la station de base et l'utilisateur k est noté $\mathbf{h}_k^H \in \mathbb{C}^{1\times M}$. Le canal de liaison descendante entier peut donc être représenté par

$$\mathbf{H}^H = \begin{bmatrix} \mathbf{h}_1^H \\ \vdots \\ \mathbf{h}_K^H \end{bmatrix}_{\in \mathbb{C}^{K\times M}}. \tag{1.44}$$

La transmission de liaison descendante est modélisée par :

$$\mathbf{y} = \mathbf{H}^H \mathbf{B}\mathbf{s} + \mathbf{n}, \tag{1.45}$$

où $\mathbf{y} \in \mathbb{C}^{K\times 1}$ est le signal reçu par tous les utilisateurs, $\mathbf{s} \in \mathbb{C}^{K\times 1}$ représente le vecteur de signaux Gaussiens i.i.d., et $\mathbf{n}$ représente le bruit blanc Gaussien, centré, et une variance égale à σ_n^2. $\mathbf{B}$ est la formation de faisceau de liaison descendante ayant la puissance totale P.

Nous considérons un scénario dans lequel K utilisateurs sont situés à proximité les uns des autres de telle sorte que leurs canaux présentent les même caractistiques statistiques, i.e., $\forall k, \mathbb{E}\{\mathbf{h}_k \mathbf{h}_k^H\} = \mathbf{R}$. Le rang de $\mathbf{R}$ est supposé être égale à d. Soit la décomposition en valeur propres (EVD) de $\mathbf{R}$:

$$\mathbf{R} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H \tag{1.46}$$

En supposant que les valeurs propres de $\mathbf{\Sigma}$ sont triées par ordre décroissant, les d premières colonnes de $\mathbf{U}$ sont extraites pour former une sous-matrice $\mathbf{U}_1 \in \mathbb{C}^{M\times d}$. Le vecteur de canal $\mathbf{h}_k$ est donc dans l'espace engendré par $\mathbf{U}_1$, c'est-à-dire $\forall k, \mathbf{h}_k$ est une combinaison linéaire des colonnes de $\mathbf{U}_1$.

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_k \end{bmatrix} = \mathbf{U}_1 \mathbf{A}, \tag{1.47}$$

où $\mathbf{A} \in \mathbb{C}^{d\times K}$ est défini comme :

$$\mathbf{A} \triangleq \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_K \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ a_{d1} & a_{d2} & \cdots & a_{dK} \end{bmatrix}. \tag{1.48}$$

1.6.2 Conception de l'acquisition de l'information de canal sans D2D

La stratégie traditionnelle de feedback pour les système multi-utilisateurs est de laisser chaque utilisateur quantifier son vecteur de canal de liaison descendante puis envoyer l'information quantifiée à la station de base [5]. À

noter que dans la configuration MIMO massif, seulement une part de l'information de canal (notamment K coefficients par utilisateur) est nécessaire pour atteindre l'orthogonalité entre les signaux de l'utilisateur.

En extrayant N lignes avec $K \leq N \leq d$ de la matrice $\mathbf{A}$, on peux obtenir une matrice $\mathbf{A}_s \in \mathbb{C}^{N \times K}$; En prenant N colonnes correspondantes de $\mathbf{U}_1$, on obtaient ainsi la matrice $\mathbf{U}_s \in \mathbb{C}^{M \times N}$. Nous pouvons alors reconstruire partiellement la matrice de canal de la manière suivante :

$$\widetilde{\mathbf{H}} \triangleq \mathbf{U}_s \mathbf{A}_s. \tag{1.49}$$

Un filtre ZF utilisant cette information de canal peut alors être écrit comme

$$\mathbf{B} = \frac{\sqrt{P}\widetilde{\mathbf{H}}^{\dagger}}{||\widetilde{\mathbf{H}}^{\dagger}||_{\mathrm{F}}}. \tag{1.50}$$

Il est décrit dans [6] comment former la matrice $\mathbf{A}_s$ en extrayant linéairement *arbitrairement* une sous-matrice de taille $K \times d$ de $\mathbf{A}$. Si la station de base connait obtient cette information de canal incomplète, elle garantit que les utilisateurs ne reçoivent pas d'interférences après application du filtrage ZF (1.50). Lorsque l'échange d'information de canal entre les utilisateurs n'est pas possible, il est raisonnable de choisir les vecteurs propres des premiers K lignes depuis $\mathbf{U}_1$ dans la mesure où ils représentent les K vecteurs propres les plus forts en terme de statistiques.

1.6.3 Acquisistion coopérative de l'information de canal avec D2D

Lorsque l'échange de l'information de canal est rendue possible entre les utilisateurs, ceux-ci peuvent prendre une décision conjointe et déterminer conjointement quel ensemble de vecteurs propres de $\mathbf{U}_1$, ou également quels lignes extraire de $\mathbf{A}$ pour former la matrice $\mathbf{A}_s$. Par exemple, nous pouvons considérer le SNR à côté de l'utilisateur comme critère :

$$\mathrm{SNR} = \frac{P}{||\widetilde{\mathbf{H}}^{\dagger}||_{\mathrm{F}}^2} = \frac{P}{\mathrm{tr}\{(\mathbf{A}_s^H \mathbf{A}_s)^{-1}\}}. \tag{1.51}$$

Nous choisissons alors K parmi les d vecteurs propres de $\mathbf{U}_1$ pour que le SNR soit maximisée.

$$\mathcal{G}^{(2)} = \underset{N=K}{\arg} \min \mathrm{tr}\{(\mathbf{A}_s^H \mathbf{A}_s)^{-1}\}. \tag{1.52}$$

L'optimalité de la décision de sélection des vecteurs propres est atteinte par un algorithme de sélection décrémentielle qui est expliqué dans [7] [8].

Nous commençons par considére le canal effectif $\mathbf{A}$ dans sa totalité. Sous la condition que le SNR est minimum, les lignes de $\mathbf{A}$ sont retirés un par un, jusqu'à il reste K lignes.

Lorsque la décision conjointe est prise, i.e., pour $\mathcal{G}^{(2)}$, les utilisateurs transmettent alors à la station de base la matrice $\mathbf{A}_s$ (quantifié) correspondante, ainsi que les indices de $\mathcal{G}^{(2)}$.

Trois régimes différents d'acquisition de l'information de canal sont évalués et les performances des débits totaux sont donnés dans la Figure 6.1.

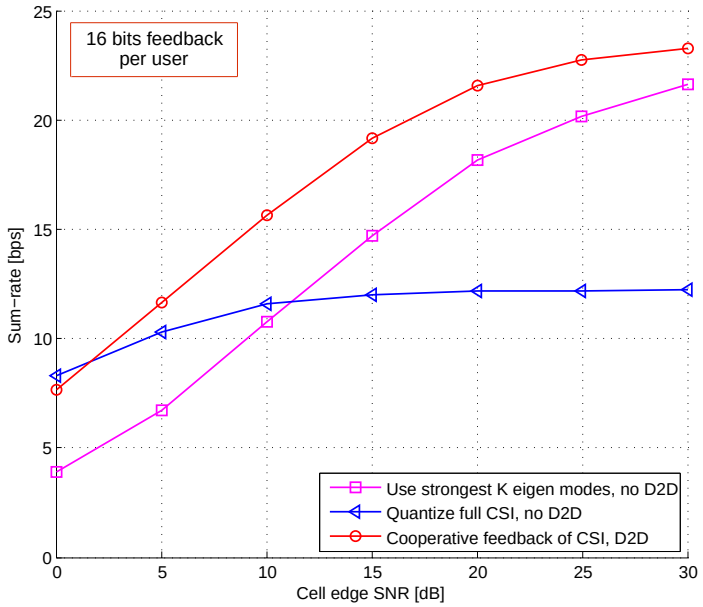

FIGURE 1.12 – DL sum-rates with/without feedback cooperation, feedback overhead : 16 bits per user, $K = 3$, $M = 50$.

Pour comparison, on suppose que la même quantité de bits de quantification est disponible dans chacun des trois régimes. Chaque utilisateur transmet alors 16 bits d'information à la station de base. La courbe"Quantize full CSI, no D2D" désigne la performance lorsque la matrice de canal effective $\mathbf{A}$ est transmie dans sa totalité. La courbe "Use strongest K eigen modes, no D2D" indique la performance de chaque utilisateur pour K vecteurs propres dominants de $\mathbf{R}$. La courbe "Cooperative feedback of CSI, D2D" fait référence à la nouvelle méthode proposée. Les utilisateurs transmettent alors les projections quantifiées et les indices des trois vecteurs propres qui sont choisis. Les simulations numériques mettent clairement en évidence le gain

de la méthode proposée.

1.6.4 Acquisition coopérative de l'indice du précodeur avec D2D

Nous considérons maintenant une autre approche dans laquelle l'information de canal n'est pas explicitement transmise mais l'indice du précodeur à utiliser est transmis. L'échange de l'information de canal via D2D communications entre les utilisateurs permet à ceux-ci de choisir conjointement une matrice de précodage. Après l'échange de cette information, les utilisateurs sélectionnent conjointement la meilleure matrice de précodage selon un certain critère, et transmettent l'indice de la matrice de précodage sélectionnée. Le critère de sélection peut varier en fonction de l'exigence de la complexité du système. Un choix intuitive est la maximisation du débit total $\mathcal{C}$.

Pour clarifier cette description, les principales étapes sont rappelées dans la suite :

(1) Les utilisateurs envoient leur information de canal à un certain utilisateur "maître". Cet utilisateur maître a maintenant la matrice effective de canal $\mathbf{A}$.

(2) L'utilisateur maître recherche dans le dictionnaire l'indice du précodage qui maximise le débit total :

(3) L'utisateur maître transmet cet indice à la station de base.

(4) La station de base effectue un filtrage en utilisant la matrice du dictionnaire désignée par l'indice reçu.

Les performances atteintes en utilisant cette méthodes sont illustrées numériquement dans le graphique 6.2. Comme on peut le voir, la méthode proposée apporte de significatifs gains de performance.

1.6.5 Conclusions

Dans cette section, nous proposons nue nouvelle méthode permettant l'acquisition de l'information de canal lorsque les transmission s'effectuent en mode FDD. Cette méthode répose sur l'échange d'information de canal entre les utilisateurs de manière à optimiser la transmission vers la station de base. Deux approches sont proposées pour l'optimisation de la transmission vers la station de base après l'étape de partage de l'information. Dans la première, l'information de canal est expicitement transmise alors que dans la deuxième c'est l'indice du filtre à utiliser qui est transmis. Ces méthodes aident à réduire le coût de l'acquisition de l'information de canal dans des systèmes FDD MIMO massif.

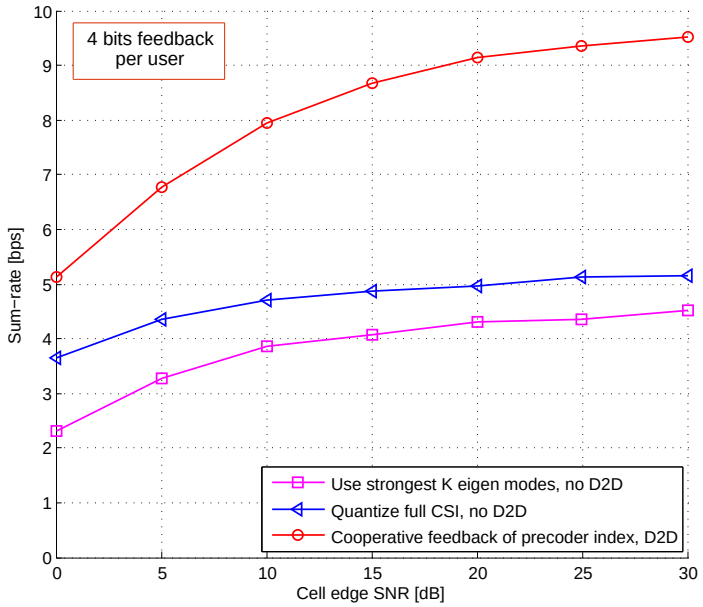

FIGURE 1.13 – DL Sum-taux de sélection coopérative de pré-codage et non-coopérative CSI feedback, feedback overhead : 4 bits par utilisateur.

1.7 Publications

Les publications suivantes sont le résultat des travaux réalisés au cours du doctorat.

1.7.1 Conférences

1. Haifan Yin, David Gesbert, Miltiades Filippou, and Yingzhuang Liu "Decontaminating pilots in MIMO massifs systems", International Conference on Communications (ICC 2013), Jun. 9-13, 2013, Budapest, Hungary.

2. Miltiades Filippou, David Gesbert, and Haifan Yin, "Decontaminating pilots in cognitive MIMO massifs networks," 9th IEEE International Symposium on Wireless Communications Systems (ISWCS 2012), Aug. 28-31, 2012, Paris, France (**invited**).

3. Haifan Yin, David Gesbert, and Laura Cottatellucci "A statistical

approach to interference reduction in distributed large-scale antenna systems", International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014), May, 2014, Florence, Italy.

4. Haifan Yin, Laura Cottatellucci, and David Gesbert, "Enabling MIMO massifs systems in the FDD mode thanks to D2D communications", Asilomar Conference on Signals, Systems, and Computers (Asilomar 2014), Nov. 2-5, 2014, Pacific Grove, CA, USA (**invited**),

5. Haifan Yin, Laura Cottatellucci, David Gesbert, Ralf R. Müller, and Gaoning He, "Pilot decontamination using combined angular and amplitude based projections in MIMO massifs systems", IEEE 16th Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2015), Jun. 28 - Jul. 1, 2015, Stockholm, Sweden. (**invited**).

6. Junting Chen, Haifan Yin, Laura Cottatellucci, and David Gesbert, "Precoder feedback versus channel feedback in MIMO massifs under user cooperation," 49th Asilomar Conference on Signals, Systems, and Computers (Asilomar 2015), Nov. 8-11, 2015, Pacific Grove, CA, USA (**invited**).

7. Haifan Yin, Laura Cottatellucci, David Gesbert, Ralf R. Müller, and Gaoning He, "Robust pilot decontamination : A joint angle and power domain approach", International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016), Mar. 20 - Mar. 25, 2016, Shanghai, China.

1.7.2 Journaux

1. Haifan Yin, David Gesbert, Miltiades Filippou, and Yingzhuang Liu "A coordinated approach to channel estimation in large-scale multiple-antenna systems", IEEE Journal on Selected Areas in Communications, special issue on large-scale antenna systems, Vol. 31, No. 2, pp. 264-273, Feb. 2013.

2. Haifan Yin, David Gesbert, and Laura 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, Oct. 2014.

3. Haifan Yin, Laura Cottatellucci, David Gesbert, Ralf R. Müller, and Gaoning He, "Robust pilot decontamination based on joint angle and power domain discrimination", accepted for publication in IEEE Transactions on Signal Processing. Sept. 2015. [Online]. Available : http ://arxiv.org/abs/1509.06024

1.7.3 Brevets

1. Laura Cottatellucci, Haifan Yin, David Gesbert, Gaoning He, and Georg M. Kreuz, "Closed-loop CSI feedback with co-operative feedback design for use in MIMO/MISO systems", European Patent, PCT application number : PCT/EP2014/073501, Oct. 31, 2014,

Chapter 2

Introduction

Full reuse of the frequency across neighboring cells leads to severe interference, which in turn limits the quality of service offered to cellular users, especially those located at the cell edge. As service providers seek some solutions to restore performance in low-SINR cell locations, several approaches aimed at mitigating inter-cell interference have emerged in the last few years. Among these, the solutions which exploit the additional degrees of freedom made available by the use of multiple antennas seem the most promising, particularly so at the base station side where such arrays are more affordable.

In the cooperation approach, the so-called network MIMO (or CoMP in the 3GPP terminology) schemes mimic the transmission over a virtual MIMO array encompassing the *spatially distributed* base station antennas. It goes at the expense of fast signaling links over the backhaul, a need for tight synchronization, and seemingly multi-user detection schemes that are computationally demanding in practice.

In an effort to solve this problem while limiting the requirements for user data sharing over the backhaul network, coordinated beamforming approaches have been proposed in which 1) multiple-antenna processing is exploited at each base station, and 2) the optimization of the beamforming vectors at all cooperating base stations is performed jointly. Coordinated beamforming does not require the exchange of user message information (e.g., in network MIMO). Yet it still demands the exchange of channel state information (CSI) across the transmitters on a fast time scale and low-latency basis, making almost as challenging to implement in practice as the above mentioned network MIMO schemes. Additionally, a major hurdle preventing from realizing the full gains of MIMO multi-cell cooperation lies in the cost of acquiring and sharing channel estimates using orthogonal training

sequences over large clusters of cells [9].

Fortunately a path towards solving some of the essential practical problems related to beamforming-based interference avoidance was suggested in [10]. In this work, it was pointed out that the need for exchanging Channel State Information at Transmitter (CSIT) between base stations could be alleviated by simply increasing the number of antennas, M, at each transmitter (so-called massive MIMO). The added cost of hardware is compensated by the fact that simple distributed beamforming schemes that require little inter-cell cooperation can efficiently mitigate interference [10–13]. This result is rooted in the law of large numbers, which predicts that, as the number of antennas increases, the vector channel for a desired terminal will tend to become more orthogonal to the vector channel of a randomly selected interfering user. This makes it possible to reject interference at the base station side by simply aligning the beamforming vector with the desired channel ("Maximum Ratio Combining" or spatial matched filter), providing that local channel information is known at the base station. Hence in theory, a simple fully distributed per-cell beamforming scheme can offer performance scaling (with M) similar to a more complex centralized optimization.

Despite its promising potential, there are several challenges and limiting factors that restrain the performance of massive MIMO in practical system. A brief review of the challenges (non-comprehensive) of massive MIMO is given below.

- Pilot contamination problem. In reality, channel information is acquired on the basis of finite-length pilot sequences, and crucially, in the presence of inter-cell interference. Therefore, the pilot sequences from neighboring cells would contaminate each other, giving rise to the so-called pilot contamination (PC) problem [14], [15], [16]. It was pointed out in [10] that pilot contamination constitutes a bottleneck for performance.
- CSI acquisition in FDD mode. In FDD system, the channel reciprocity does not hold. Thus CSIT has to be estimated and fed back to the base station by user terminals. The required overhead of training and feedback grows with the number of base station antennas. As a result, the FDD deployment of massive MIMO is hampered by the practical issue of large training and feedback overhead [17] [18].
- Reciprocity calibration. In TDD system, the wireless uplink and downlink channels are in general reciprocal. However, when it comes to the hardware of base station and user terminal, the channel reciprocity may not hold anymore. This introduces an unknown amplitude scaling and phase shift between the uplink and downlink channels. In order

to have an accurate CSIT via uplink training, reciprocity calibration is needed in practice [19].

– Hardware impairments. Some initial studies in [20] show that hardware impairments give rise to channel estimation error and finite ceilings on the channel estimation error and on the capacity. Another study in [21] shows that phase noise causes the estimated channel outdated and thus limits the length of data sequences and the number of scheduled users.

In this thesis, we look into the first two challenges and provide solutions to these problems. More precisely, in chapter 3 we introduce some useful properties of massive MIMO channels. In chapter 4 and chapter 5 we address the CSI acquisition of TDD systems, particularly the pilot contamination problem. In order to enhance the estimation quality, several novel channel estimation methods are proposed based on the newly identified channel properties in chapter 3. In chapter 6 we consider the CSI acquisition of massive MIMO in FDD mode. We show novel approaches which help reduce the CSI feedback overhead in FDD system.

2.1 Motivations

The realization of the high energy efficiency and throughput of massive MIMO, especially in the downlink transmission, relies on the accurate knowledge of the CSI. The system performance is very sensitive to the accuracy of the channel information. Thus CSI acquisition is crucial, in both TDD and FDD deployments. However, the acquisition of CSI is confronted by several challenges in massive MIMO systems. This thesis is mainly motivated by these challenges in both TDD and FDD duplexing schemes.

In practical wireless system, CSI is acquired based on training sequences sent by user terminals. Due to limited time and frequency resources, non-orthogonal pilot sequences are typically used by user terminals in neighboring cells, resulting in residual channel estimation error. This effect, called *pilot contamination* [14], [16], has a detrimental impact on the actual achievable spectral and energy efficiencies in real systems. In fact, the residual error in channel estimation due to the unavoidable reuse of identical training sequences by user terminals in different cells was identified as a limiting factor to cancel interference in massive MIMO networks. It has been shown in [10], [15], [14], [16] that pilot contamination effects determine a quick saturation of the interference rejection performance with the number of antennas, thereby undermining the value of massive MIMO systems in cellular networks. This problem will be addressed in chapter 4 and chapter 5 and

efficient interference mitigation methods will be given.

In today's cellular system, the large majority of currently deployed cellular networks is operating in FDD mode. In this mode, the transmission on uplink and downlink are in different frequency bands, thus, unlike in TDD mode, channel reciprocity does not hold. The downlink channel information is first obtained at user side via training signals, then fed back to the base station. In massive MIMO systems with large number of BS antennas, this feedback overhead in principle scales linearly with the number of BS antennas, which renders the deployment of massive MIMO in FDD mode impractical. Thus, new feedback techniques have to be developed towards relieving the feedback overhead before FDD mode is made more applicable in massive MIMO systems. To this end, we propose in chapter 6 novel feedback schemes which exploits the synergies between device-to-device communications and massive MIMO. We demonstrate that with a fixed amount of feedback bits available, our proposed methods have significant performance gains in terms of sum-rate.

2.2 CSI Acquisition: Challenges and Avenues

The performance gains of massive MIMO largely rely on the channel information, which enables beamforming and interference canceling. In massive MIMO systems, CSI acquisition faces challenges in both TDD mode and FDD mode, due to pilot contamination and training and feedback overhead respectively. Before presenting our works, we briefly review the state-of-art of the related research topics in this section.

2.2.1 CSI Acquisition in TDD Massive MIMO

Pilot contamination has fueled extensive research efforts in recent years. This problem was first identified in [14] and analyzed in [22] [23] [16] for the scenario with a large number of antennas at base station. It was pointed out in [10] that this problem constitutes an important practical challenge for massive MIMO. In [16], a multi-cell precoder optimization problem is formulated in order to reduce interference. In [24] and [25] the authors studied time-shifted pilots scheme in neighboring cells. A multi-cell joint precoding method is proposed in [26], based on the assumptions that the data signals for all users in all cells are available at each BS and that large-scale fading coefficients are known globally. In this case, multi-cell joint precoding is performed.

Some blind or semi-blind channel estimation methods are also proposed [27–29] to reduce pilot contamination. These estimation schemes are based on the assumption of asymptotic pairwise orthogonality between channel vectors of different users. This assumption enables the identification of channel vectors based on the received data signal. In particular, [27] proposed an eigenvalue decomposition (EVD)-based estimation scheme, which estimated the channel subspace blindly from the received data signal. The remaining ambiguity is resolved using training sequences. An enhanced iterative method is also proposed in [27] in order to improve the performance. In [28,29], a blind pilot decontamination scheme is proposed. Analysis shows that pilot contamination can be eliminated asymptotically if the signal of interest is stronger than interference, a condition that may hold in a network with power control and power-controlled handoff. In [30] [31], the authors improve the EVD-based channel estimation with maximum a-posteriori (MAP) criterion. A non-convex optimization problem is formulated and a local optimum can be achieved using gradient ascent method.

A pilot sequence hopping scheme is proposed in [32], which enables the users to change the pilot sequences in each transmission time slot. By doing so, a randomization of the pilot contamination can be achieved. Instead of estimating the channels in each transmission time slot, the authors propose to perform channel estimation once for several time slots. A Kalman filter is developed in order to exploit this randomization effect, leading to pilot contamination reduction. Another smart pilot reuse scheme is proposed in [33]. Inspired by fractional frequency reuse, the authors propose the idea of fractional pilot reuse, where a fraction of the users which are close to their own base stations reuse the same pilot pool in all the cells, while the other users are assigned orthogonal pilots according to a reuse factor. In [34], the authors propose a semi-blind estimation method which is similar to Robust Independent Component Analysis (RobustICA) [35]. An optimization algorithm is developed based on asynchronous pilot protocol.

Coordinated pilot allocation methods can be found in [36, 37]. The authors of [37] formulate a combinatorial network utility maximization problem with respect to the pilot assignment strategy. Exhaustive searching and low-complexity suboptimal solutions are given.

2.2.2 CSI Acquisition in FDD Massive MIMO

Some preliminary effort to address the problem of massive MIMO in FDD deployment is [17]. In this paper, the authors propose joint spatial division and multiplexing (JSDM) which exploits the property of the covariance ma-

trices of channel vectors in order to reduce the dimension of instantaneous CSI. A two-stage precoding strategy is proposed, with the prebeamforming matrix depending on the second-order statistics of user channels only and an instantaneous beamformer based on the reduced-dimensional instantaneous CSI. This work identifies a low-rankness property of the channel covariance matrix, independently of our work [36]. This property is utilized to reduce training overhead and CSI feedback overhead, thus opening the way to the feasibility of massive MIMO in FDD mode. A similar idea of two-stage precoder is given in [38], where a subspace-tracking algorithm is formulated for the computation of the statistical precoder.

In [39], a non-coherent trellis-coded CSI quantization method is proposed, whose encoding complexity scales linearly with the number of BS antennas, in contrast to traditional CSI quantization method where for a fixed feedback rate per antenna, the size of the codebook for quantizing the channel grows exponentially with the number of antennas. With such efficient CSI quantization and feedback method, massive MIMO system operating in FDD mode is made more feasible.

An antenna grouping based CSI feedback reduction method is proposed in [40] [41]. The proposed algorithm maps multiple correlated antenna elements to a single representative value using pre-designed patterns. A header is introduced in the feedback protocol in order to select a group pattern. Then the reduced-dimensional channel vector is quantized and fed back to the base station.

There are also numerous papers [42–49] discussing the training and/or CSI overhead reduction methods using compressed sensing [50–54], based on the (hidden) sparsity of massive MIMO channels in the domains of time, frequency, and space.

2.2.3 Massive MIMO and D2D

Some initial concepts exploiting device-to-device communication was proposed as early as [55] for the purpose of multihop relays in cellular networks. Later, it is shown in that D2D can be exploited to increase the spectral efficiency [56–59], as well as facilitate multicasting [60, 61], video dissemination [62–64], peer-to-peer communications [65], etc. Despite all these research efforts on D2D, relatively few literatures investigate the benefits of combining D2D technique with massive MIMO. Among them, [66] studies the spectral efficiencies of cellular user and D2D user - which coexist in the network - under both perfect CSI and imperfect CSI conditions. It is found that under perfect CSI, the received SINRs of cellular users increase without

limit, and that the effects of noise, fast fading, and the interference from the other co-channel cellular users and D2D users vanish completely. On the contrary, under more realistic CSI condition, i.e., with MMSE-based channel estimate, the received SINR is further degraded due to interference from D2D users, in addition to pilot contamination. The works [67,68] investigate the average sum-rate and energy efficiency of a network with underlaid D2D communications. The authors show that underlaid D2D users and cellular users can co-exist well without significant performance degradation when the density of D2D users is small. This limitation is due to the co-channel interference between D2D users and cellular users, a similar conclusion of [66].

2.3 Contributions and Publications

This thesis mainly concerns the signal processing aspect of massive MIMO. We address the problem of CSI acquisition in both TDD and FDD modes. Interference mitigation in channel estimation phase and CSI feedback overhead reduction are the focus of this thesis. We propose novel schemes exploiting the channel properties that are specific to massive MIMO. These properties are modeled theoretically and verified by Monte Carlo experiments. The performance gains of the new signal processing methods are confirmed by simulations.

The contributions of this PhD thesis are listed below:

2.3.1 MMSE-based pilot decontamination

We reveal a fundamental property of massive MIMO channel – the reduced-rankness of channel covariance matrix. We derive closed-form expression of the rank of the covariance matrix for uniform linear array (ULA). The rank is a function of the angular support of multipath AoA. This is typically verified in realistic scenarios due to the limited angle spread followed by incoming paths originating from street-level users [69]. This low-rankness property is proved to be helpful in massive MIMO interference mitigation, including pilot contamination reduction. Therefore we exploit the fact that the desired user signals and interfering user signals are received at the base station with (at least approximately) low-rank. We show that the performance of the LMMSE channel estimator depends on the level of overlapping between the desired signal subspace and the interference subspace. Importantly, we are the first to prove that pilot contamination can be eliminated completely under certain condition on the AoA distributions. This condition

is identified as the non-overlapping condition: if the AoA supports of interference channels do not overlap with the AoA support of desired channel, then pilot contamination vanishes asymptotically with the number of BS antennas. Moreover, we propose a strategy for the assignment of pilot sequences. This strategy helps assign carefully selected groups of users to identical pilot sequences so as to minimize the channel estimation error. Performance close to the interference-free channel estimation scenario is obtained for moderate numbers of antennas and users.

These results were published in

– Haifan Yin, David Gesbert, Miltiades Filippou, and Yingzhuang Liu "A coordinated approach to channel estimation in large-scale multiple-antenna systems", IEEE Journal on Selected Areas in Communications, special issue on large-scale antenna systems, Vol. 31, No. 2, pp. 264-273, Feb. 2013.
– Haifan Yin, David Gesbert, Miltiades Filippou, and Yingzhuang Liu "Decontaminating pilots in massive MIMO systems", International Conference on Communications (ICC 2013), Jun. 9-13, 2013, Budapest, Hungary.

2.3.2 Generalized low-rankness of channel covariance and its applications

We consider a uniform linear massive array system yet with *several clusters* of AoA supports from each user. In this case we establish a low-rank model for the channel's covariance that directly extends that of our previous work [36], where the rank is shown to be a function of the incoming/departing angular spread of multipath. We prove that the reduce-rank property of the channel covariance matrix revealed in [36] also holds in this more general setting. We then show that a similar low-rank result holds for a linear array with random placement of antenna elements. More surprisingly, we show that the low-rankness property extends to two dimensional distributed large scale arrays. In particular we build our analysis under the classical one-ring model, which shows the dependence of the signal subspace's richness on the scattering radius around the user terminal. Closed-form expression of the upper bound of the channel covariance's rank is given. The applications of the low-rankness covariance property to channel estimation's denoising and low-complexity interference filtering are highlighted.

These results are published in

– Haifan Yin, David Gesbert, and Laura Cottatellucci "Dealing with interference in distributed large-scale MIMO systems: A statistical ap-

proach", IEEE Journal of Selected Topics in Signal Processing, Vol. 8, No. 5, Oct. 2014.

– Haifan Yin, David Gesbert, and Laura Cottatellucci "A statistical approach to interference reduction in distributed large-scale antenna systems", International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014), May, 2014, Florence, Italy.

2.3.3 Robust Angle/Power based Pilot Decontamination

In an another effort to cope with pilot contamination, we propose a family of robust algorithms which provide significant improvement over known methods. Specifically, our contributions are as follows: We propose a spatial filter which helps bring down the power of interference while preserving the signal of interest. With this spatial filter, we present a novel channel estimation scheme called "covariance-aided amplitude based projection". It combines the merits of linear MMSE estimator and the amplitude based projection method proposed in [28, 29], yet can be shown to have significant gains over these known schemes.

Then we give asymptotic analysis on this proposed method and provide weaker condition compared to the previous methods where the estimation error of the proposed method goes to zero asymptotically in the limit of large number of antennas and data symbols.

In addition, as the uniformly boundedness of the largest eigenvalue of channel covariance was reported to be useful in previous work (such as [70]) but not formally analyzed, we identify in the case of ULA a sufficient propagation condition under which the uniformly bounded spectral norm of channel covariance is satisfied exactly.

Finally we propose two low-complexity alternatives of the first method. An asymptotic performance characterization is also given.

These results are published in

– Haifan Yin, Laura Cottatellucci, David Gesbert, Ralf R. Müller, and Gaoning He, "Robust pilot decontamination based on joint angle and power domain discrimination", accepted for publication in IEEE Transactions on Signal Processing. Sept. 2015. [Online]. Available: http://arxiv.org/abs/1509.06024

– Haifan Yin, Laura Cottatellucci, David Gesbert, Ralf R. Müller, and Gaoning He, "Pilot decontamination using combined angular and amplitude based projections in massive MIMO systems", IEEE 16th Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2015), Jun. 28 - Jul. 1, 2015, Stockholm, Sweden. (**invited**).

– Haifan Yin, Laura Cottatellucci, David Gesbert, Ralf R. Müller, and Gaoning He, "Robust pilot decontamination: A joint angle and power domain approach", International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016), Mar. 20 - Mar. 25, 2016, Shanghai, China.

2.3.4 Cooperative Feedback for FDD Massive MIMO

We address the problem of CSI acquisition for massive MIMO in FDD deployments. In this setting, CSI is obtained by user terminal and then fed back to the base station. Due to the large number of BS antennas, CSI acquisition is in general challenging in FDD mode. To this end, we propose a three-phase *cooperative* closed-loop feedback as an alternative to the traditional per-user feedback loop, in order to reduce the amount of feedback for FDD systems. This novel feedback mechanism is enabled by device to device communications. Specifically, the UTs estimate the channel parameters in the channel sounding phase, then exchange the acquired local CSI in a cluster of users. After this the users perform a joint optimized design of the feedback before transmitting it to the base station. Two methods of the joint feedback are proposed. The first method performs an optimal selection of the reduced effective subspace based on the shared CSI. The second approach benefits from the knowledge of the shared CSI by selecting the best precoder from a predefined codebook.

These results are published in the following paper:

– Haifan Yin, Laura Cottatellucci, and David Gesbert, "Enabling massive MIMO systems in the FDD mode thanks to D2D communications", Asilomar Conference on Signals, Systems, and Computers (Asilomar 2014), Nov. 2-5, 2014, Pacific Grove, CA, USA (**invited**),

and the following patent:

– Laura Cottatellucci, Haifan Yin, David Gesbert, Gaoning He, and Georg M. Kreuz, "Closed-loop CSI feedback with co-operative feedback design for use in MIMO/MISO systems", European Patent, PCT application number: PCT/EP2014/073501, Oct. 31, 2014,

Chapter 3

Properties of Massive MIMO Channels

A promising approach to improve CSI acquisition method lies in exploiting the specific features of the channel itself. In massive MIMO, some particular properties of the wireless channel emerge, that can be exploited towards improved channel estimation and interference mitigation. In this chapter, we will review these specific properties of massive MIMO channels.

3.1 Network Model

We consider a network of L time-synchronized cells, with full spectrum reuse. Each base station is equipped with M antennas. There are K single-antenna users in each cell simultaneously served by their base station. In order to ease the notation, we consider a narrow band system, e.g., one carrier of orthogonal frequency-division multiplexing (OFDM) system. The uplink channel estimation is obtained at the base station (BS) by uplink training signal and possibly data signal sent by user terminals. When the cellular network operates in TDD mode, due to channel reciprocity, this uplink channel estimate can directly be utilized in the downlink precoding and data transmission. Denote the $M \times 1$ channel vector between the k-th user located in the l-th cell and the j-th base station by $\mathbf{h}_{lk}^{(j)}$. In this chapter, we investigate the special properties of massive MIMO channels. In particular, we characterize the low-dimensional property of the channel covariance, which will be shown very helpful in interference mitigation techniques in chapter 4, 5, and 6.

3.2 Low-Rank Model in ULA

3.2.1 Channel Model

In this section, we will study the wireless channels of the widely deployed ULA. The following classical multipath channel model [1] is adopted in this thesis:

$$\mathbf{h}_{lk}^{(j)} = \frac{\beta_{lk}^{(j)}}{\sqrt{P}} \sum_{p=1}^{P} \mathbf{a}(\theta_{lkp}^{(j)}) e^{i\varphi_{lkp}^{(j)}}, \tag{3.1}$$

where P is the arbitrary large number of i.i.d. paths, and $e^{i\varphi_{lkp}^{(j)}}$ is the i.i.d. random phase, which is independent over channel indices l, k, j, and path index p. $\mathbf{a}(\theta)$ is the steering (or phase response) vector by the array to a path originating from the angle of arrival θ:

$$\mathbf{a}(\theta) \triangleq \begin{bmatrix} 1 \\ e^{-j2\pi \frac{D}{\lambda} \cos(\theta)} \\ \vdots \\ e^{-j2\pi \frac{(M-1)D}{\lambda} \cos(\theta)} \end{bmatrix}, \tag{3.2}$$

where λ is the signal wavelength and D is the antenna spacing which is assumed fixed. Note that we can limit θ to $\theta \in [0, \pi]$ because any $\theta \in [-\pi, 0)$ can be replaced by $-\theta$ giving the same steering vector. $\beta_{lk}^{(j)}$ is the path-loss coefficient

$$\beta_{lk}^{(j)} = \sqrt{\frac{\alpha}{{d_{lk}^{(j)}}^{\gamma}}}, \tag{3.3}$$

in which γ is the path-loss exponent, $d_{lk}^{(j)}$ is the geographical distance between the user and the j-th base station, and α is a constant.

3.2.2 Low-rankness property of ULA

Without loss of generality, we consider a certain user (either desired user or interference user). The angle support of this user is $\Phi = [\theta_{\min}, \theta_{\max}]$, which means the probability density function (PDF) $p(\theta)$ of the AoA of this user channel $\mathbf{h}$ satisfies $p(\theta) > 0$ if $\theta \in \Phi$ and $p(\theta) = 0$ if $\theta \notin \Phi$. We have the following claim on the rank of the covariance matrix $\mathbf{R}$ of this user:

Theorem 1. *The rank of the channel covariance matrix* $\mathbf{R}$ *satisfies:*

$$\frac{\text{rank}(\mathbf{R})}{M} \leqslant d, \text{ when } M \text{ is sufficiently large },$$

where d is defined as

$$d \triangleq \big(\cos(\theta^{min}) - \cos(\theta^{max})\big)\frac{D}{\lambda}.$$

Proof: The proof relies on an intermediate lemma given below:

Lemma 1. *Define* $\boldsymbol{\alpha}(x) \triangleq [\ 1 \quad e^{-j\pi x} \quad \cdots \quad e^{-j\pi(M-1)x}\ ]^T$ *and* $\mathcal{A} \triangleq \text{span}\{\boldsymbol{\alpha}(x), x \in [-1,1]\}$. *Given* $b_1, b_2 \in [-1,1]$ *and* $b_1 < b_2$, *define* $\mathcal{B} \triangleq \text{span}\{\boldsymbol{\alpha}(x), x \in [b_1, b_2]\}$, *then*

- $\dim\{\mathcal{A}\} = M$
- $\dim\{\mathcal{B}\} \sim (b_2 - b_1)M/2$ *when* M *grows large.*

Proof: See Appendix .1. □

Lemma 1 characterizes the number of dimensions a linear space has, which is spanned by $\boldsymbol{\alpha}$(x), in which x plays the role of spatial frequency.

The rest of the proof of Theorem 1 is shown in Appendix .2. □

Theorem 1 indicates that for large M, there exists a null space null($\mathbf{R}$) of dimension $(1-d)M$, which can be exploited for the purpose of interference rejection. Interestingly, related low-rank properties of the covariance matrices were independently derived in [17] for the purpose of reducing the overhead of downlink channel estimation and CSI feedback in massive MIMO for FDD systems.

Theorem 1 states the low-rankness property of a channel covariance when the multipath AoAs span *one* continuous interval. A more general result is now given. Consider a general multipath model when the AoAs corresponding to a certain user's channel are still bounded, but come from several disjoint clusters. Let Q denote the number of clusters. Let $[\theta_q^{\min}, \theta_q^{\max}]$ denote the interval of AoAs for the q-th cluster of the paths in the $[0, \pi]$ interval. See an illustration in Fig. 3.1 for $Q = 2$. Define the total set of AoAs of the channel of interest as

$$\overline{\Phi} \triangleq \cup_{q=1}^{Q}[\theta_q^{\min}, \theta_q^{\max}], \tag{3.4}$$

so that the PDF $p(\theta)$ of the AoA satisfies $p(\theta) > 0$ if $\theta \in \overline{\Phi}$ and $p(\theta) = 0$ if $\theta \notin \overline{\Phi}$. we have the following corollary for this general setting:

Corollary 1. *The rank of channel covariance matrix* $\mathbf{R}$ *satisfies:*

$$\frac{\text{rank}(\mathbf{R})}{M} \leqslant d, \ \textit{when } M \textit{ is sufficiently large},$$

where d is defined as

$$d \triangleq \min(1, \sum_{q=1}^{Q}\big(\cos(\theta_q^{min}) - \cos(\theta_q^{max})\big)\frac{D}{\lambda}).$$

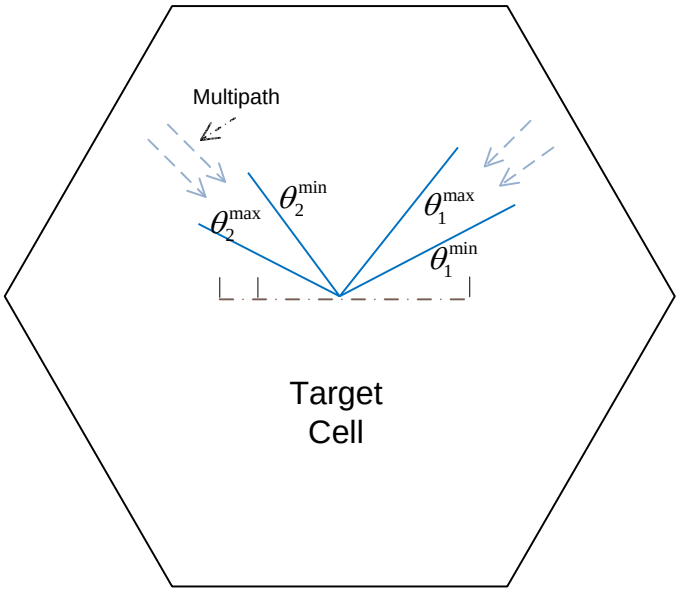

Figure 3.1 – Desired channel composed of $Q = 2$ clusters of multipath.

Proof: The channel can be seen as the sum of elementary channels each of which corresponds to one separate clusters. Then $\mathbf{R}$ can be decomposed into a sum of covariances over these clusters. Since the clusters are separated, the signal subspaces of the corresponding covariances are orthogonal and therefore their dimensions add up. Then based on Theorem 1 and Lemma 1, the proof of Corollary 1 can be readily obtained. □

The low-rankness property shown in Theorem 1 and Corollary 1 reveal a fundamental feature of massive MIMO channel. This feature can be exploited in application scenarios like spatial multiplexing [17], pilot contamination reduction, and interference mitigation.

3.3 Low-Rank Model in Random Linear Arrays

3.3.1 Channel Model

Tightly calibrated arrays with uniform spacing are hard to realize in practice. An interesting question is whether the low-rankness results shown in section 3.2.2 carry on to the setting of linear arrays with random antenna placement. To study this case, we consider a set of antennas randomly located over a line, and spanning a total aperture of $\mathcal{D}$ meters. We investigate

the extended array and $\mathcal{D}$ is allowed to grow with M.

In this case, an elementary path coming from an angle θ can be represented via the corresponding array response vector as:

$$\mathbf{a}(\theta) \triangleq \begin{bmatrix} e^{-j2\pi \frac{d_1}{\lambda}\cos(\theta)} \\ \vdots \\ e^{-j2\pi \frac{d_M}{\lambda}\cos(\theta)} \end{bmatrix}, \tag{3.5}$$

where the position of the m-th antenna[1] $(1 \leq m \leq M)$, d_m, follows a uniform distribution, i.e., $d_m \sim \mathcal{U}(0, \mathcal{D})$.

3.3.2 Low-rankness Property of Random Linear Array

We now study the rank of channel covariance for a random linear array. Define the average antenna spacing $\overline{D} \triangleq \mathcal{D}/M$. Assuming the aperture of antenna array $\mathcal{D}$ is increasing linearly with M, i.e., $\overline{D}$ is constant, we now have the extended results on the low-dimensional property:

Proposition 3. *Define*

$$\begin{aligned} \boldsymbol{\alpha}(x) &\triangleq \left[e^{-j2\pi \frac{d_1}{\lambda}x}, \cdots, e^{-j2\pi \frac{d_M}{\lambda}x}\right]^T \\ \mathcal{B} &\triangleq \text{span}\{\boldsymbol{\alpha}(x), x \in [b_1, b_2]\} \\ \mathcal{C} &\triangleq \text{span}\{\boldsymbol{\alpha}(x), x \in \bar{b}\}, \end{aligned}$$

where $b_1, b_2 \in [-1, 1]$, $\bar{b} \triangleq \cup_{q=1}^{Q}[b_q^{min}, b_q^{max}]$, *and* b_q^{min}, b_q^{max} *are values such that*

$$-1 \leqslant b_1^{min} < b_1^{max} < \cdots < b_q^{min} < b_q^{max} < \cdots < b_Q^{min} < b_Q^{max} \leqslant 1$$

then we have

- $\dim\{\mathcal{B}\} \leq (b_2 - b_1)M\overline{D}/\lambda + o(M)$
- $\dim\{\mathcal{C}\} \leq \sum_{q=1}^{Q}\left(b_q^{max} - b_q^{min}\right) M\overline{D}/\lambda + o(M)$

Proof: See Appendix .4. □

Proposition 3 indicates the dimensions spanned in massive MIMO regime by elementary paths for (i) single cluster of AoA, and (ii) multiple disjoint clusters of AoA, respectively. The following result now directly generalizes Corollary 1 to random arrays.

1. Note that antenna ordering has no impact on our results.

Proposition 4. *With a bounded support of AoAs* $\overline{\Phi}$ *as in (3.4), the rank of channel covariance matrix* $\mathbf{R}$ *satisfies:*

$$\text{rank}(\mathbf{R}) \leq \sum_{q=1}^{Q} \left(\cos(\theta_q^{min}) - \cos(\theta_q^{max})\right) \frac{M\overline{D}}{\lambda} + o(M), \tag{3.6}$$

Proof: We can readily obtain this result by replacing x with $\cos(\theta)$ in Proposition 3. □

This result above suggests that the low-dimensional feature of signal subspaces in massive MIMO is not critically linked to the Fourier structure of the steering vectors. Furthermore, it should be noted that the above upper bound is actually very tight for large M, as witnessed from the simulation in Fig. 3.2, where we take $Q = 1, \overline{D} = \lambda/2$ for example. The AoA spread is 40 degrees, and the closed form model refers to

$$f(M) \triangleq \sum_{q=1}^{Q} \left(\cos(\theta_q^{\min}) - \cos(\theta_q^{\max})\right) \frac{M\overline{D}}{\lambda}.$$

We can observe that $\text{rank}(\mathbf{R})$ is well approximated by $f(M)$.

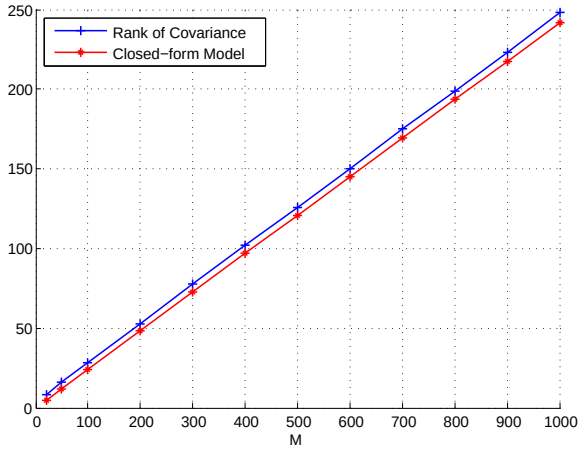

Figure 3.2 – Closed-form rank model for the channel covariance vs. actual rank.

Proposition 4 and Fig. 3.2 suggest that a property of rank additivity holds for multiple disjoint clusters of AoAs in the massive MIMO regime, i.e., for $M \to +\infty$.

3.4 Low-Rank Model in DAS

3.4.1 Channel Model

We now turn to another popular form of large scale antenna regime, often referred to in the literature as distributed antenna systems. In such a setting, a virtual base station is deployed having its M antennas scattered throughout the cell. [2] We consider again the uplink in which joint combining across all BS antennas is assumed possible. The M base station antennas are assumed uniformly and randomly located in a fixed size network, serving single-antenna users. M is allowed to grow large giving rise to a so-called dense network. Our model assumes a disk-shaped cell of radius R_c, although simulation and intuition confirm that the actual shape of the cell's boundary is irrelevant to the main result. In order to facilitate the analysis, we adopt the one-ring model [2, 3] where users are surrounded by a ring of P local scatterers (see Fig. 3.3) located r meters away from the user. The positions of the scatterers are considered to follow a uniform distribution on the ring. In the one-ring model, the propagation from user to base is assumed to follow P paths (hereafter referred to as scattering paths), where each path p bounces once on the p-th scatterer before reaching all M destinations. [3]

Hence, the path length from user k to the m-th antenna via the p-th path is $r + d_{kpm}$, where d_{kpm} is the distance between the p-th scatterer of the k-th user and the m-th BS antenna. The path loss of the p-th scattering path is modeled by:

$$\beta_{kpm} = \frac{\alpha}{(d_{kpm} + r)^{\gamma}}, \tag{3.7}$$

where α is a constant that can be computed based on desired cell-edge SNR, and γ is the path loss exponent. We scale the amplitude of each path by $\sqrt{P}$. The channel between user k and all BS antennas is given by:

$$\mathbf{h}_k \triangleq \frac{1}{\sqrt{P}} \sum_{p=1}^{P} \mathbf{h}_{kp}, \tag{3.8}$$

where $\mathbf{h}_{kp}$ is the p-th scattering path vector channel between user k and all

2. For ease of exposition we temporarily consider a single cell setting in this section, i.e., $L = 1$. However simulation is also done later in a multi-cell scenario.

3. Note that this model assumes the BS antennas are high enough above clutter so that there is no local scattering around the BS antennas.

base stations:

$$\mathbf{h}_{kp} \triangleq \begin{bmatrix} \sqrt{\beta_{kp1}} e^{-j2\pi \frac{d_{kp1}+r}{\lambda}} \\ \vdots \\ \sqrt{\beta_{kpM}} e^{-j2\pi \frac{d_{kpM}+r}{\lambda}} \end{bmatrix} e^{j\varphi_{kp}}, \tag{3.9}$$

where $e^{j\varphi_{kp}}$ denotes the random common phase of that scattering path vector due to possible random perturbations of the user location around the ring center or the phase shift due to the reflection on the scatterer. φ_{kp} is assumed i.i.d. and uniformly distributed between 0 to 2π.

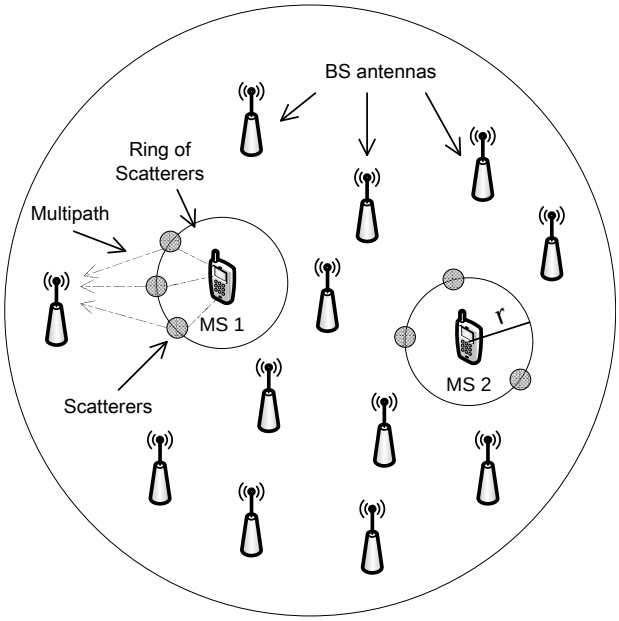

Figure 3.3 – The distributed large-scale antenna setting with a one-ring model.

3.4.2 Low-rankness Property of DAS

We have shown in section 3.2.2 and 3.3.2 the low-dimension property for co-located linear antenna array systems. In attacking this problem it is important to distinguish the rank reduction effect due to path loss from the intrinsic finite-rank behavior of the large antenna channel covariance in an equal path loss regime. In fact, in an extended network (i.e. where some base station antennas can be arbitrarily far from some users), the signal

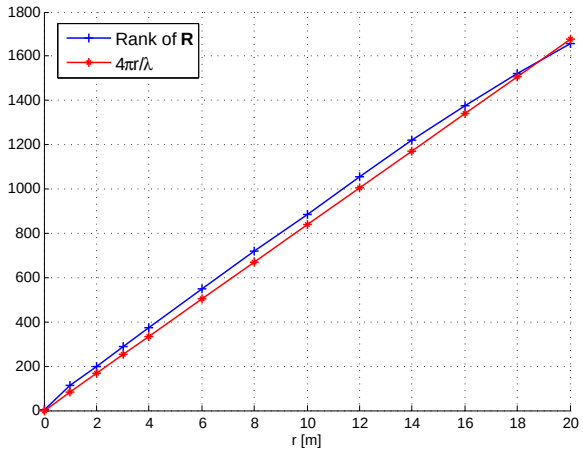

Figure 3.4 – Rank vs. r, $M = 2000$, $\lambda = 0.15$m, $R_c = 500$m.

of any given user will be received over only a limited number of antennas in its vicinity, thereby effectively limiting the channel rank to the size of this neighborhood. To circumvent this problem, we consider below a (dense) network where the path loss terms are set artificially to be all equal (to one) and study the finite-rankness under such conditions. In this model, the channel covariance is defined as $\mathbf{R} \triangleq \mathbb{E}\{\mathbf{h}\mathbf{h}^H\}$ where the expectation is taken over the random positions of the scatterers on the ring. Note that our analysis indicates that a randomization over the user's location inside the scattering's disk would produce an identical upper bound on the rank.

Theorem 2. *The rank of the channel covariance matrix for a distributed antenna system satisfies:*

$$rank(\mathbf{R}) \leq \frac{4\pi r}{\lambda} + o(r). \tag{3.10}$$

Proof: See Appendix .5. □

In reality we show below that the right hand side of (3.10) is a very close approximation of the actual rank, which is defined as the number of eigenvalues of $\mathbf{R}$ which are greater than a prescribed threshold (in our simulations it is taken to be 10e-5). Theorem 2 shows a linear dependency of the rank on the size of the scattering ring. When r increases, the richer scattering environment expands the dimension of signal space. Fig. 3.4 shows the behavior of the covariance rank with respect to the scattering radius r. We can see the rank scales linearly with the slope $4\pi/\lambda$. However because of the finite

number of antennas the rank will finally saturate towards M when r keeps increasing.

3.5 Uniformly Boundedness of Channel Covariance in ULA

Now we characterize the property of the uniformly boundedness of the spectral norm of the channel covariance matrix in ULA. This property will be shown useful in the analysis of chapter 5. In addition, many papers (such as [70]) using random matrix method build the analysis on the assumption of the uniformly bounded spectral norm of covariance matrix. Yet, the validity of this assumption has not been proved before. In this section, we identify in the case of ULA a sufficient propagation condition under which the uniformly bounded spectral norm of channel covariance is satisfied exactly.

The spectral norm of $\mathbf{R}$ denotes its largest eigenvalue. As shown in section 3.2.2, the rank of a channel covariance scales linearly with the number of antennas, which gives us an intuition that the channel energy spreads over a subspace whose dimension grows with M, and that the energy is not concentrated on one or several eigenvectors of the channel covariance. This intuition is proved in the following preposition:

Proposition 5. *Let Φ be the AoA support of a certain user. Let $p(\theta)$ be the probability density function of AoA of that user. If $p(\theta)$ is uniformly bounded, i.e., $p(\theta) < +\infty, \forall \theta \in \Phi$, and Φ lies in a closed interval that does not include the parallel directions with respect to the array , i.e., $0, \pi \notin \Phi$, then, the spectral norm of the user's covariance $\mathbf{R}$ is uniformly bounded:*

$$\forall M, \|\mathbf{R}\|_2 < +\infty. \tag{3.11}$$

Proof: See Appendix .7. □

Note that this result is hinted upon [17] by resorting to approximation of $\mathbf{R}$ by a circulant matrix. Our Proposition 5 here gives a formal proof of the previous approximated result.

3.6 Conclusions

In this chapter we point out several fundamental properties of massive MIMO users' channels. We investigate the low-dimensional properties of covariance signal subspaces in general topologies of massive arrays, including

uniform linear array, random linear array, and 2D distributed array. Another revealed property is the uniformly boundedness of the spectral norm of channel covariance for ULA. These properties will be shown helpful in later chapters when dealing with interference and reducing the amount of CSI feedback.

Chapter 4

Covariance based Channel Estimation

4.1 Introduction

As shown in chapter 3, the channel covariance matrix of massive MIMO exhibits a useful low-rankness property. In this chapter, we exploit this property towards improving CSI acquisition of massive MIMO in TDD deployments. This chapter focuses on the problem of channel estimation and interference reduction in the presence of multi-cell interference generated from pilot contamination. We propose an estimation method which provides a substantial improvement in performance. It relies on the exploitation of dormant side-information lying in the second-order statistics of the user channels, both for desired and interfering users. In particular, we demonstrate a powerful result indicating that the exploitation of covariance information under certain subspace conditions on the covariance matrices can lead to a complete removal of pilot contamination effects in the large M limit. We then turn to a practical algorithm design where this concept is exploited. The key idea behind the new algorithm is the use of a covariance-aware pilot assignment strategy within the channel estimation phase itself. While diversity-based scheduling methods have been popularized for maximizing various throughput-fairness performance criteria [71], [72], [73], [74], the potential benefit of user-to-pilot assignment in the context of interference-prone channel estimation has received very little attention so far. Finally, as another application of the previous identified low-rankness of the channel covariance matrix in chapter 3, a simple subspace-based interference mitigation scheme is put forward, which exploits the statistical information of the interference

channels.

4.2 UL training

Consider the uplink training of a massive MIMO system. We assume the pilots, of length τ, used by single-antenna users in the same cell are mutually orthogonal. As a result, intra-cell interference is negligible in the channel estimation phase. However, non-orthogonal (possibly identical) pilots are reused from cell to cell, resulting in pilot contamination from $L-1$ interfering cells. For ease of exposition, in this chapter we consider the case of single user per cell, unless otherwise notified. The pilot sequence used in the l-th cell is denoted by:

$$\mathbf{s}_l = [\ s_{l1} \quad s_{l2} \quad \cdots \quad s_{l\tau}\ ]^T. \tag{4.1}$$

The powers of pilot sequences are assumed equal such that $|s_{l1}|^2 + \cdots + |s_{l\tau}|^2 = \tau, l = 1, 2, \ldots, L$. Without loss of generality, we assume the first cell is the target cell so that we can drop the BS index on the superscript of channel vector. The channel vector between the l-th cell user and the target base station is $\mathbf{h}_l$. Thus, $\mathbf{h}_1$ is the desired channel while $\mathbf{h}_l$, $l > 1$ are interference channels. All channel vectors are assumed to be $M \times 1$ complex Gaussian, undergoing correlation due to the finite multipath angle spread at the base station side [75]:

$$\mathbf{h}_l = \mathbf{R}_l^{1/2}\mathbf{h}_{\mathrm{W}l}, l = 1, 2, \ldots, L, \tag{4.2}$$

where $\mathbf{h}_{\mathrm{W}l} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is the spatially white $M \times 1$ SIMO channel, and $\mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ denotes zero-mean complex Gaussian distribution with covariance matrix $\mathbf{I}_M$. In this chapter, we make the assumption that covariance matrix $\mathbf{R}_l \triangleq \mathbb{E}\{\mathbf{h}_l\mathbf{h}_l^H\}$ can be obtained separately from the desired and interference channels (see section 4.8 for how this could be done in practice).

During the pilot phase, the $M \times \tau$ signal received at the target base station is

$$\mathbf{Y} = \sum_{l=1}^{L} \mathbf{h}_l\mathbf{s}_l^T + \mathbf{N}, \tag{4.3}$$

where $\mathbf{N} \in \mathbb{C}^{M\times\tau}$ is the spatially and temporally white additive Gaussian noise (AWGN) with zero-mean and element-wise variance σ_n^2.

4.3 Pilot Contamination Problem

Conventional channel estimation relies on correlating the received signal with the known pilot sequence (referred here as Least Squares (LS) estimate for example). Hence, using the model in (5.3), an LS estimator for the desired channel $\mathbf{h}_1$ is

$$\widehat{\mathbf{h}}_1^{\mathrm{LS}} = \mathbf{Y}\mathbf{s_1}^*(\mathbf{s_1}^T\mathbf{s_1}^*)^{-1}. \tag{4.4}$$

The conventional estimator suffers from a lack of orthogonality between the desired and interfering pilots, an effect known as pilot contamination [14], [22], [23]. In particular, when the *same* pilot sequence is reused in all L cells, i.e., $\mathbf{s}_1 = \cdots = \mathbf{s}_L = \mathbf{s}$, the estimator can be written as

$$\widehat{\mathbf{h}}_1^{\mathrm{LS}} = \mathbf{h}_1 + \sum_{l\neq 1}^{L} \mathbf{h}_l + \mathbf{N}\mathbf{s}^*/\tau \ . \tag{4.5}$$

As it appears in (4.5), the interfering channels leak directly into the desired channel estimate. The estimation performance is then limited by the signal to interfering ratio at the base station, which in turns limits the ability to design an effective interference-avoiding beamforming solution.

4.4 Covariance-aided Channel Estimation

We hereby propose an improved channel estimator with the aim of reducing the pilot contamination effect, and taking advantage of the multiple antenna dimensions. We suggest to do so by exploiting side information lying in the second order statistics of the channel vectors. The role of covariance matrices is to capture structure information related to the distribution (mainly mean and spread) of the multipath angles of arrival at the base station. Due to the typically elevated position of the base station, rays impinge on the antennas with a finite AoA spread and a user location-dependent mean angle. Note that covariance-aided channel estimation itself is not a novel idea, e.g., in [76]. In [77], the authors focus on optimal design of pilot sequences and they exploit the covariance matrices of desired channels and colored interference. The optimal training sequences were developed with adaptation to the statistics of disturbance. In our work, however, the pilot design is shown not having an impact on interference reduction, since fully aligned pilots are transmitted. Instead, we focus on i) studying the limiting behavior of covariance-based estimates in the presence of interference and large-scale antenna arrays, and ii) how to *shape* covariance information for the full benefit of channel estimation quality.

4.4.1 Bayesian Estimation

In this section we develop channel estimators based on the second-order statistics of the channels. Two Bayesian channel estimators can be formed. In the first, all channels are estimated at the target base station (including interfering ones). In the second, only $\mathbf{h_1}$ is estimated.

By vectorizing the received signal and noise, our model (5.3) can be represented as

$$\mathbf{y} = \widetilde{\mathbf{S}}\mathbf{h} + \mathbf{n}, \tag{4.6}$$

where $\mathbf{y} = \mathrm{vec}(\mathbf{Y})$, $\mathbf{n} = \mathrm{vec}(\mathbf{N})$, and $\mathbf{h} \in \mathbb{C}^{LM\times 1}$ is obtained by stacking all L channels into a vector. The pilot matrix $\tilde{\mathbf{S}}$ is defined as

$$\widetilde{\mathbf{S}} \triangleq \left[\begin{array}{ccc} \mathbf{s}_1 \otimes \mathbf{I}_M & \cdots & \mathbf{s}_L \otimes \mathbf{I}_M \end{array} \right]. \tag{4.7}$$

Applying Bayes' rule, the conditional distribution of the channels $\mathbf{h}$ given the received training signal $\mathbf{y}$ is

$$p(\mathbf{h}|\mathbf{y}) = \frac{p(\mathbf{h})p(\mathbf{y}|\mathbf{h})}{p(\mathbf{y})}. \tag{4.8}$$

We use the multivariate Gaussian probability density function (PDF) of the random vector $\mathbf{h}$ and assume its rows $\mathbf{h}_1, \cdots, \mathbf{h}_L$ are mutually independent, giving the joint PDF:

$$p(\mathbf{h}) = \frac{\exp\left(-\sum_{l=1}^{L} \mathbf{h}_l^H \mathbf{R}_l^{-1} \mathbf{h}_l\right)}{\pi^{LM}(\det \mathbf{R}_1 \cdots \det \mathbf{R}_L)^M}. \tag{4.9}$$

Note that we derive this Bayesian estimator under the standard condition of covariance matrix invertibility, although we show later this hypothesis is actually challenged by reality in the large-number-of-antennas regime. Fortunately, our final expressions for channel estimators completely skip the covariance inversion.

Using (5.5), we may obtain:

$$p(\mathbf{y}|\mathbf{h}) = \frac{\exp\left(-(\mathbf{y} - \widetilde{\mathbf{S}}\mathbf{h})^H (\mathbf{y} - \widetilde{\mathbf{S}}\mathbf{h})/\sigma_n^2\right)}{(\pi\sigma_n^2)^{M\tau}}. \tag{4.10}$$

Combining the equations (4.9) and (4.10), the expression of (4.8) can be rewritten as

$$p(\mathbf{h}|\mathbf{y}) = \frac{\exp\left(-l(\mathbf{h})\right)}{AB}, \tag{4.11}$$

where $A \triangleq p(\mathbf{y})(\pi\sigma_n^2)^{M\tau}$, $B \triangleq \pi^{LM}(\det \mathbf{R})^M = \pi^{LM}(\det \mathbf{R}_1 \cdots \det \mathbf{R}_L)^M$, and

$$l(\mathbf{h}) \triangleq \mathbf{h}^H \bar{\mathbf{R}} \mathbf{h} + (\mathbf{y} - \widetilde{\mathbf{S}}\mathbf{h})^H (\mathbf{y} - \widetilde{\mathbf{S}}\mathbf{h})/\sigma_n^2, \tag{4.12}$$

in which $\mathbf{R} \triangleq \mathrm{diag}(\mathbf{R}_1, \cdots, \mathbf{R}_L)$, $\bar{\mathbf{R}} \triangleq \mathbf{R}^{-1}$.

Using the maximum a posteriori (MAP) decision rule, the Bayesian estimator yields the most probable value given the observation $\mathbf{y}$ [78]:

$$\begin{aligned} \widehat{\mathbf{h}} &= \arg \max_{\mathbf{h} \in \mathbb{C}^{LM \times 1}} p(\mathbf{h}|\mathbf{y}) \\ &= \arg \min_{\mathbf{h} \in \mathbb{C}^{LM \times 1}} l(\mathbf{h}) \\ &= (\sigma_n^2 \mathbf{I}_{LM} + \mathbf{R}\widetilde{\mathbf{S}}^H \widetilde{\mathbf{S}})^{-1} \mathbf{R} \widetilde{\mathbf{S}}^H \mathbf{y}. \end{aligned} \tag{4.13}$$

Interestingly, the Bayesian estimate as shown in (4.13) coincides with the minimum mean square error (MMSE) estimate, which has the form

$$\widehat{\mathbf{h}}^{\mathrm{MMSE}} = \mathbf{R}\widetilde{\mathbf{S}}^H (\widetilde{\mathbf{S}}\mathbf{R}\widetilde{\mathbf{S}}^H + \sigma_n^2 \mathbf{I}_{\tau M})^{-1} \mathbf{y}. \tag{4.14}$$

(4.13) and (4.14) are equivalent thanks to the matrix inversion identity $(\mathbf{I} + \mathbf{AB})^{-1}\mathbf{A} = \mathbf{A}(\mathbf{I} + \mathbf{BA})^{-1}$.

4.4.2 Channel Estimation with Full Pilot Reuse

Previously we have given expressions whereby interfering channels are estimated simultaneously with the desired channel. This could be of use in designing zero-forcing type receivers. Even though it is clear that Zero-Forcing (ZF) type (or other sophisticated) receivers would give better performance at finite M (see [15] for an analysis of this problem), in this chapter, however, we focus on simple matched filters, since such filters are made more relevant by the users of massive MIMO. Matched filters require the knowledge of the desired channel only, so that interference channels can be considered as nuisance parameters. For this case, the single user channel estimation shown below can be used. For ease of exposition, the worst case situation with a unique pilot sequence reused in all L cells is considered:

$$\mathbf{s} = \begin{bmatrix} s_1 & s_2 & \cdots & s_\tau \end{bmatrix}^T. \tag{4.15}$$

Similar to (5.6), we define a training matrix $\bar{\mathbf{S}} \triangleq \mathbf{s} \otimes \mathbf{I}_M$. Note that $\bar{\mathbf{S}}^H \bar{\mathbf{S}} = \tau \mathbf{I}_M$. Then the vectorized received training signal at the target base station can be expressed as

$$\mathbf{y} = \bar{\mathbf{S}} \sum_{l=1}^{L} \mathbf{h}_l + \mathbf{n}. \tag{4.16}$$

Since the Bayesian estimator and the MMSE estimator are identical, we omit the derivation and simply give the expression of this estimator for the desired channel $\mathbf{h}_1$ only:

$$\widehat{\mathbf{h}}_1 = \mathbf{R}_1 \bar{\mathbf{S}}^H \left(\bar{\mathbf{S}} \left(\sum_{l=1}^{L} \mathbf{R}_l \right) \bar{\mathbf{S}}^H + \sigma_n^2 \mathbf{I}_{\tau M} \right)^{-1} \mathbf{y} \tag{4.17}$$

$$= \mathbf{R}_1 \left(\sigma_n^2 \mathbf{I}_M + \tau \sum_{l=1}^{L} \mathbf{R}_l \right)^{-1} \bar{\mathbf{S}}^H \mathbf{y}. \tag{4.18}$$

Note that the MMSE channel estimation in the presence of identical pilots is also undertaken in other works such as [15].

In the section below, we examine the degradation caused by the pilot contamination on the estimation performance. In particular, we point out the role played by the use of covariance matrices in dramatically reducing the pilot contamination effects under certain conditions on the rank structure.

We are interested in the mean squared error (MSE) of the proposed estimators, which can be defined as: $\mathcal{M} \triangleq \mathbb{E}\{\|\widehat{\mathbf{h}} - \mathbf{h}\|_2^2\}$, or for the single user channel estimate $\mathcal{M}_1 \triangleq \mathbb{E}\{\|\widehat{\mathbf{h}}_1 - \mathbf{h}_1\|_2^2\}$.

The estimation MSE of (4.13) is

$$\mathcal{M} = \mathrm{tr} \left\{ \mathbf{R} \left(\mathbf{I}_{LM} + \frac{\widetilde{\mathbf{S}}^H \widetilde{\mathbf{S}}}{\sigma_n^2} \mathbf{R} \right)^{-1} \right\}. \tag{4.19}$$

Specifically, when identical pilots are used in all cells, the MSEs are

$$\mathcal{M} = \mathrm{tr} \left\{ \mathbf{R} \left(\mathbf{I}_{LM} + \frac{\tau \mathbf{J}_{\mathrm{LL}} \otimes \mathbf{I}_M}{\sigma_n^2} \mathbf{R} \right)^{-1} \right\}, \tag{4.20}$$

$$\mathcal{M}_1 = \mathrm{tr} \left\{ \mathbf{R}_1 - \mathbf{R}_1^2 \left(\frac{\sigma_n^2}{\tau} \mathbf{I}_M + \sum_{l=1}^{L} \mathbf{R}_l \right)^{-1} \right\}, \tag{4.21}$$

where $\mathbf{J}_{\mathrm{LL}}$ is an $L \times L$ unit matrix consisting of all 1s. The derivations to obtain $\mathcal{M}$ and $\mathcal{M}_1$ use standard methods and the details are omitted here due to lack of space. However, similar methods can be found in [79]. Of course, it is clear from (4.20) and (4.21) that the MSE is not dependent on the specific design of the pilot sequence, but on the power of it.

We can readily obtain the channel estimate of (4.18) in an interference-free scenario, by setting interference terms to zero:

$$\widehat{\mathbf{h}}_1^{\text{no int}} = \mathbf{R}_1\big(\sigma_n^2\mathbf{I}_M + \tau\mathbf{R}_1\big)^{-1}\bar{\mathbf{S}}^H(\bar{\mathbf{S}}\mathbf{h}_1 + \mathbf{n}), \tag{4.22}$$

where the superscript *no int* refers to the "no interference case", and the corresponding MSE:

$$\mathcal{M}_1^{\text{no int}} = \text{tr}\left\{\mathbf{R}_1\left(\mathbf{I}_M + \frac{\tau}{\sigma_n^2}\mathbf{R}_1\right)^{-1}\right\}. \tag{4.23}$$

4.4.3 Large-scale Analysis

We seek to analyze the performance for the above estimators in the regime of large number of antennas M. For tractability, our analysis is based on the assumption of uniform linear array (ULA) with supercritical antenna spacing (i.e., less than or equal to half wavelength).

Below, we momentarily assume that the selected users exhibit multipath AoAs that do not overlap with the AoAs for the desired user, i.e., the AoA spread and user locations are such that multipath for the desired user are confined to a region of space where interfering paths are very unlikely to exist. Although the asymptotic analysis below makes use of this condition, it will be shown in Section 4.5 how such a structure can be shaped implicitly by the coordinated pilot assignment. Finally, simulations reveal in Section 6.6 the robustness with respect to an overlap between AoA regions of desired and interference channels (for instance in the case of Gaussian AoA distribution).

The main result of this chapter is as follows:

Theorem 3. *Assume the multipath angle of arrival θ yielding channel $\mathbf{h}_j, j = 1, \ldots, L$, in (6.21), is distributed according to an arbitrary density $p_j(\theta)$ with bounded support, i.e., $p_j(\theta) = 0$ for $\theta \notin [\theta_j^{\min}, \theta_j^{\max}]$ for some fixed $\theta_j^{\min} \leqslant \theta_j^{\max} \in [0, \pi]$. If the $L-1$ intervals $[\theta_i^{\min}, \theta_i^{\max}]$, $i = 2, \ldots, L$ are strictly non-overlapping with the desired channel's AoA interval*[1] *$[\theta_1^{\min}, \theta_1^{\max}]$, we have*

$$\lim_{M\to\infty} \widehat{\mathbf{h}}_1 = \widehat{\mathbf{h}}_1^{no\ int}. \tag{4.24}$$

1. This condition is just one example of practical scenario leading to non-overlapping signal subspaces between the desired and the interference covariances, however, more general multipath scenarios could be used.

Proof: *Proof:* From the channel model (6.21), we get

$$\mathbf{R}_i = \frac{\delta_i^2}{P}\sum_{p=1}^{P}\mathbb{E}\{\mathbf{a}(\theta_{ip})\mathbf{a}(\theta_{ip})^H\} = \delta_i^2\mathbb{E}\{\mathbf{a}(\theta_i)\mathbf{a}(\theta_i)^H\},$$

where θ_i has the PDF $p_i(\theta)$ for all $i = 1, \ldots, L$. The proof of Theorem 3 relies on three intermediate results, namely Lemma 1, Theorem 1, and Lemma 2 as given blow. The essential ingredient is to exhibit an asymptotic (at large M) orthogonal vector basis for $\mathbf{R}_i$ constructed from steering vectors at regularly sampled spatial frequencies.

Lemma 2. *The null space* $\text{null}(\mathbf{R}_i)$ *includes a certain set of unit-norm vectors:*

$$\text{null}(\mathbf{R}_i) \supset \text{span}\left\{\frac{\mathbf{a}(\Phi)}{\sqrt{M}}, \forall\Phi \notin [\theta_i^{min}, \theta_i^{max}]\right\}, \text{ as } M \to \infty.$$

Proof: See Appendix .6. □

This lemma indicates that multipath components with AoA outside the AoA region for a given user will tend to fall in the null space of its covariance matrix in the large-number-of-antennas case.

We now return to the proof of Theorem 3. $\mathbf{R}_i$ can be decomposed into

$$\mathbf{R}_i = \mathbf{U}_i\mathbf{\Sigma}_i\mathbf{U}_i^H, \tag{4.25}$$

where $\mathbf{U}_i$ is the signal eigenvector matrix of size $M \times m_i$, in which $m_i \leq d_i M$. $\mathbf{\Sigma}_i$ is an eigenvalue matrix of size $m_i \times m_i$. Due to Lemma 2 and the fact that densities $p_i(\theta)$ and $p_1(\theta)$ have non-overlapping supports, we have

$$\mathbf{U}_i^H\mathbf{U}_1 = 0, \forall i \neq 1, \text{ as } M \to \infty. \tag{4.26}$$

Combining the channel estimate (4.18) and the channel model (4.16), we obtain

$$\widehat{\mathbf{h}}_1 = \mathbf{R}_1\left(\sigma_n^2\mathbf{I}_M + \tau\sum_{l=1}^{L}\mathbf{R}_l\right)^{-1}\bar{\mathbf{S}}^H\left(\bar{\mathbf{S}}\sum_{i=1}^{\mathrm{L}}\mathbf{h}_i + \mathbf{n}\right).$$

According to (4.26), matrices $\mathbf{R}_1$ and $\sum_{l=2}^{L}\mathbf{R}_l$ span orthogonal subspaces in the large M limit. Therefore we place ourselves in the asymptotic regime for M, when $\tau\sum_{l=2}^{L}\mathbf{R}_l$ can be eigen-decomposed according to

$$\tau\sum_{l=2}^{L}\mathbf{R}_l = \mathbf{W}\mathbf{\Sigma}\mathbf{W}^H, \tag{4.27}$$

where $\mathbf{W}$ is the eigenvector matrix such that $\mathbf{W}^H\mathbf{W} = \mathbf{I}$ and span $\{\mathbf{W}\}$ is included in the orthogonal complement of span $\{\mathbf{U}_1\}$. Now denote $\mathbf{V}$ the unitary matrix corresponding to the orthogonal complement of both span $\{\mathbf{W}\}$ and span $\{\mathbf{U}_1\}$, so that the $M \times M$ identity matrix can now be decomposed into:

$$\mathbf{I}_M = \mathbf{U}_1\mathbf{U}_1^H + \mathbf{W}\mathbf{W}^H + \mathbf{V}\mathbf{V}^H. \tag{4.28}$$

Thus, for large M,

$$\begin{aligned}\widehat{\mathbf{h}}_1 \sim \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{U}_1^H \big(&\sigma_n^2\mathbf{U}_1\mathbf{U}_1^H + \sigma_n^2\mathbf{V}\mathbf{V}^H + \sigma_n^2\mathbf{W}\mathbf{W}^H \\ &+\tau\mathbf{U}_1\mathbf{\Sigma}_1\mathbf{U}_1^H + \mathbf{W}\mathbf{\Sigma}\mathbf{W}^H\big)^{-1}\left(\tau\sum_{i=1}^{\mathrm{L}}\mathbf{h}_i + \bar{\mathbf{S}}^H\mathbf{n}\right).\end{aligned}$$

Due to asymptotic orthogonality between $\mathbf{U}_1$, $\mathbf{W}$ and $\mathbf{V}$,

$$\begin{aligned}\widehat{\mathbf{h}}_1 &\sim \mathbf{U}_1\mathbf{\Sigma}_1(\sigma^2\mathbf{I}_{m_1} + \tau\mathbf{\Sigma}_1)^{-1}\mathbf{U}_1^H(\tau\sum_{i=1}^{\mathrm{L}}\mathbf{h}_i + \bar{\mathbf{S}}^H\mathbf{n}) \\ &\sim \mathbf{U}_1\mathbf{\Sigma}_1(\sigma^2\mathbf{I}_{m_1} + \tau\mathbf{\Sigma}_1)^{-1}\tau(\mathbf{U}_1^H\mathbf{h}_1 + \sum_{i=2}^{\mathrm{L}}\mathbf{U}_1^H\mathbf{h}_i + \frac{\bar{\mathbf{S}}^H\mathbf{n}_1}{\tau}).\end{aligned}$$

However, since $\mathbf{h}_i \subset \text{span}\left\{\mathbf{a}(\theta), \forall\theta \in [\theta_i^{\min}, \theta_i^{\max}]\right\}$, we have from Lemma 2 that $\frac{\left\|\mathbf{U}_1^H\mathbf{h}_i\right\|_2}{\left\|\mathbf{U}_1^H\mathbf{h}_1\right\|_2} \to 0$, for $i \neq 1$ when $M \to \infty$. Therefore

$$\lim_{M\to\infty}\widehat{\mathbf{h}}_1 = \tau\mathbf{U}_1\mathbf{\Sigma}_1\big(\sigma_n^2\mathbf{I}_{m_1} + \tau\mathbf{\Sigma}_1\big)^{-1}\left(\mathbf{U}_1^H\mathbf{h}_1 + \frac{\bar{\mathbf{S}}^H\mathbf{n}}{\tau}\right),$$

which is identical to $\widehat{\mathbf{h}}_1^{\text{no int}}$ if we apply the EVD decomposition (4.25) for $\mathbf{R}_1$ in (4.22). This proves Theorem 3. □

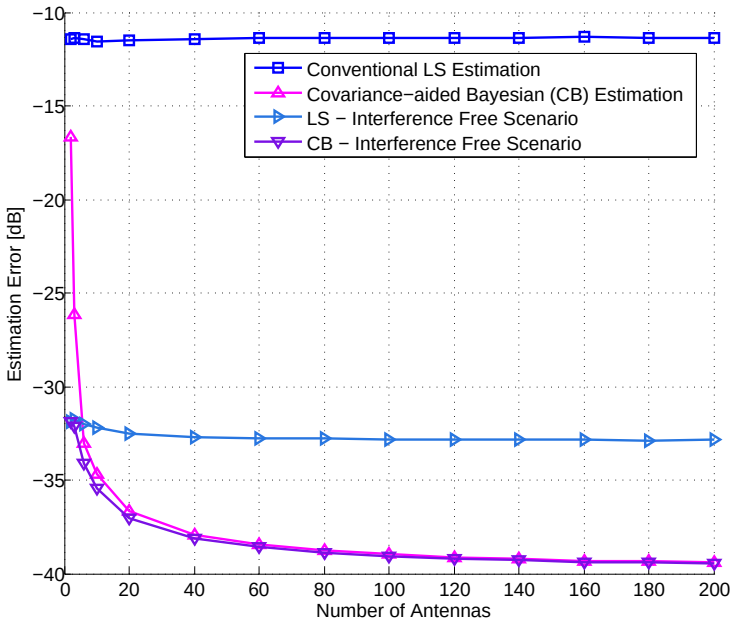

Figure 4.1 – Estimation MSE vs. number of BS antennas , 2-cell network, fixed positions of two users, uniformly distributed AoAs with $\theta_\Delta = 20$ degrees, non-overlapping multipath.

We validate Theorem 3 in Fig. 4.1 with a 2-cell network, where the two users' positions are fixed. AoAs of desired channels are uniformly distributed with a mean of 90 degrees, and the angle spreads of all channels are 20 degrees, yielding no overlap between desired and interfering multipaths. As can be seen, the channel estimation error converges to interference-free case, indicating that the pilot contamination is quickly eliminated with growing number of antennas. More details of simulation parameters can be found in section 4.7.1.

Theorem 3 reveals the following sufficient condition for achieving total interference suppression in the large M regime:

$$\bigcup_{l=2}^{L} \text{span}\left\{\mathbf{R}_l\right\} \subset \text{null}\left\{\mathbf{R}_1\right\} \tag{4.29}$$

where the above condition requires the target channel covariance to exhibit a non-empty null space (aka low-dimensional subspace) *and* for all other interference covariances' signal subspaces to fall within this null space. In practice, the inclusion condition in (4.29) can be realized by a user grouping algorithm (an example will be presented in 4.5), as long as the rank of each

covariance is small enough in relation to M. However Theorem 3 is limited to the case of single AoA support interval and tightly calibrated antenna array. We will give below extended results under less restrictive condition.

Multiple Clusters of AoA Support

We now consider the general multipath model when the AoAs of a certain user's channel are bounded, but come from several disjoint clusters, as described in Section 3.2.2. Denote by $\overline{\Phi}_d$ as the union of possible AoAs of desired channel and by $\overline{\Phi}_i$ the union of all possible interference AoAs. We have the following result for the massive uniform array:

Corollary 2. *if $D \leq \lambda/2$ and $\overline{\Phi}_d \cap \overline{\Phi}_i = \emptyset$, then the MMSE estimate of (4.18) satisfies:*

$$\lim_{M\to\infty} \widehat{\mathbf{h}}_1 = \widehat{\mathbf{h}}_1^{no\ int}. \tag{4.30}$$

Proof: It can be shown from Lemma 2 that , condition (4.29) will be fulfilled as long as interfering AoAs do not overlap with *any* of the clusters for the desired channel, in which case if we analyze the received signal using eigen-value decomposition, we can find the interference disappears asymptotically because of its orthogonality with the signal space of desired channel covariance. (4.30) is obtained in the same way as in 4.4.3. As a result we omit the detailed proof here. □

Rank Additivity Condition

In the following proposition we extend the results of Corollary 2 to the weaker assumption of rank additivity for the covariance matrices of the desired and interference channels.

Proposition 6. *Let $\mathbf{R}_d$ be the covariance matrix of desired channel and $\mathbf{R}_i$ be the covariance of the sum of all interference channels. If $\mathbf{R}_d$ and $\mathbf{R}_i$ satisfy the following rank additivity property*

$$\operatorname{rank}(\mathbf{R}_d + \mathbf{R}_i) = \operatorname{rank}(\mathbf{R}_d) + \operatorname{rank}(\mathbf{R}_i),$$

then in the high SNR regime, the linear MMSE estimate of the desired channel is error free, or, in other words, its error covariance matrix $\mathbf{C}_e$ vanishes.

Proof: In the case of absence of white Gaussian noise, i.e., $\sigma_n^2 = 0$, and rank deficient signal and interference covariance matrices, the error covariance matrix of linear MMSE estimator [79] can be generalized as

$$\mathbf{C}_e = \mathbf{R}_d - \mathbf{R}_d(\mathbf{R}_d + \mathbf{R}_i)^{\dagger}\mathbf{R}_d \tag{4.31}$$

where $(\cdot)^\dagger$ denotes the Moore-Penrose generalized inverse of the matrix argument. Let us denote by $\mathbf{R}_x = \mathbf{U}_x\mathbf{\Sigma}_x\mathbf{U}_x^H$, with $x \in \{d, e\}$, $\mathbf{U}_x$ unitary matrix, and $\mathbf{\Sigma}_x$ diagonal matrix, the eigenvalue decomposition of the Hermitian matrix $\mathbf{R}_x$. Then, $\mathbf{R}_x^\dagger = \mathbf{U}_x\mathbf{\Sigma}_x^\dagger\mathbf{U}_x^H$, where the elements i, j of the matrix $\mathbf{\Sigma}_x^\dagger$ are given by

$$\mathbf{\Sigma}_{x,ij}^\dagger = \begin{cases} \mathbf{\Sigma}_{x,ii}^{-1}, & \text{if } i = j \text{ and } \mathbf{\Sigma}_{x,ii} \neq 0; \\ 0, & \text{otherwise.} \end{cases}$$

Additionally, $\widetilde{\mathbf{U}}_x$ denotes the column space of $\mathbf{R}_x$ and $\widetilde{\mathbf{\Sigma}}_x$ the corresponding nonzero eigenvalues such that $\mathbf{R}_x = \widetilde{\mathbf{U}}_x\widetilde{\mathbf{\Sigma}}_x\widetilde{\mathbf{U}}_x^H$. Then, under the assumption of rank additivity of the covariance matrices $\mathbf{R}_d$ and $\mathbf{R}_i$, the theorem on the Moore-Penrose generalized inverse for sum of matrices in [80] yields

$$(\mathbf{R}_d + \mathbf{R}_i)^\dagger = (\mathbf{I} - \mathbf{S}^\dagger)\mathbf{R}_d^\dagger(\mathbf{I} - \mathbf{T}^\dagger) + \mathbf{S}^\dagger\mathbf{R}_i^\dagger\mathbf{T}^\dagger, \tag{4.32}$$

where $\mathbf{S} = \widetilde{\mathbf{U}}_i\widetilde{\mathbf{U}}_i^H(\mathbf{I} - \widetilde{\mathbf{U}}_d\widetilde{\mathbf{U}}_d^H)$ and $\mathbf{T} = (\mathbf{I} - \widetilde{\mathbf{U}}_d\widetilde{\mathbf{U}}_d^H)\widetilde{\mathbf{U}}_i\widetilde{\mathbf{U}}_i^H$.

Let us observe that

$$\mathbf{T}^\dagger\mathbf{R}_d = \mathbf{0} \quad \text{and} \quad \mathbf{R}_d\mathbf{S}^\dagger = \mathbf{0}. \tag{4.33}$$

We focus on the first equality. The proof of the second equation follows along the same line. By appealing to the mixed type reverse order laws of the $r \times s$ matrix $\mathbf{A}$ and the $s \times t$ matrix $\mathbf{B}$ in [81]

$$(\mathbf{AB})^\dagger = \mathbf{B}^H(\mathbf{A}^H\mathbf{ABB}^H)^\dagger\mathbf{A}^H,$$

$\mathbf{T}^\dagger$ can be rewritten as

$$\begin{aligned} \mathbf{T}^\dagger &= \widetilde{\mathbf{U}}_i\widetilde{\mathbf{U}}_i^H\left[(\mathbf{I} - \widetilde{\mathbf{U}}_d\widetilde{\mathbf{U}}_d^H)\widetilde{\mathbf{U}}_i\widetilde{\mathbf{U}}_i^H\right]^\dagger(\mathbf{I} - \widetilde{\mathbf{U}}_d\widetilde{\mathbf{U}}_d^H) \\ &= \widetilde{\mathbf{U}}_i\widetilde{\mathbf{U}}_i^H\mathbf{T}^\dagger(\mathbf{I} - \widetilde{\mathbf{U}}_d\widetilde{\mathbf{U}}_d^H). \end{aligned}$$

The first equality is obtained utilizing the fact that the matrices $\widetilde{\mathbf{U}}_i\widetilde{\mathbf{U}}_i^H$ and $(\mathbf{I} - \widetilde{\mathbf{U}}_d\widetilde{\mathbf{U}}_d^H)$ are orthogonal projectors and thus idempotent. Then,

$$\mathbf{T}^\dagger\mathbf{R}_d = \widetilde{\mathbf{U}}_i\widetilde{\mathbf{U}}_i^H\mathbf{T}^\dagger(\mathbf{I} - \widetilde{\mathbf{U}}_d\widetilde{\mathbf{U}}_d^H)\widetilde{\mathbf{U}}_d\widetilde{\mathbf{\Sigma}}_d\widetilde{\mathbf{U}}_d^H = \mathbf{0}.$$

Finally, substituting (4.32) into (4.31) and accounting for orthogonality in (4.33)

$$\begin{aligned} \mathbf{C}_e &= \mathbf{R}_d - \mathbf{R}_d\left[(\mathbf{I} - \mathbf{S}^\dagger)\mathbf{R}_d^\dagger(\mathbf{I} - \mathbf{T}^\dagger) + \mathbf{S}^\dagger\mathbf{R}_i^\dagger\mathbf{T}^\dagger\right]\mathbf{R}_d \\ &= \mathbf{R}_d - \mathbf{R}_d\mathbf{R}_d^\dagger\mathbf{R}_d = \mathbf{0}. \end{aligned}$$

In the last equality we use one of the fundamental relations defining the Moore-Penrose generalized inverse. □

According to Proposition 4, the rank additivity condition is in general satisfied even for random linear array when the AoA support of desired channel and that of interference channels span disjoint region of spaces, i.e., $\overline{\Phi}_d \cap \overline{\Phi}_i = \emptyset$. This property can be exploited in pilot decontamination or interference rejection. As an example, we show in Fig. 4.2 the channel estimation performance for random linear array setting. In this figure, we

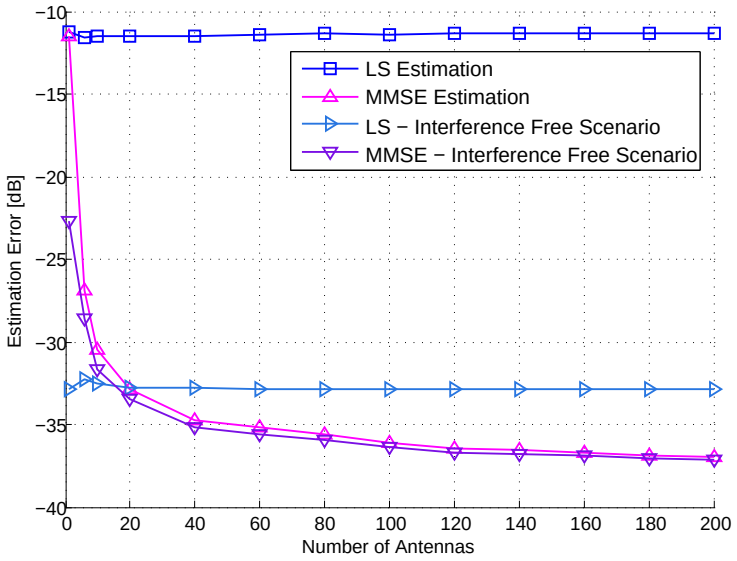

Figure 4.2 – Channel Estimation MSE vs. M, $\overline{D} = \lambda/2$, 2-cell network, angle spread 30 degrees, $\overline{\Phi}_d \cap \overline{\Phi}_i = \emptyset$, cell-edge SNR is 20dB. We compare the standard LS to MMSE estimators, in interference and interference-free scenarios.

consider a 2-cell network. Each cell has one single-antenna user who uses identical pilot sequence. The mean squared error of uplink channel estimation is shown. The simulation suggests that the MMSE channel estimator is able to rid itself from pilot contamination effects as the number of antennas is (even moderately) large, which verifies Proposition 6.

4.5 Coordinated Pilot Assignment

We have seen from above that the performance of the covariance-aided channel estimation is particularly sensitive to the degree with which the signal subspaces of covariance matrices for the desired and the interference

channels overlap with each other. In the ideal case where the desired and the interference covariances span distinct subspaces, we have demonstrated that the pilot contamination effect tends to vanish in the large-antenna-array case. In this section, we make use of this property by designing a suitable coordination protocol for assigning pilot sequences to users in the L cells. The role of the coordination is to optimize the use of covariance matrices in an effort to try and satisfy the non-overlapping AoA constraint of Theorem 1. We assume that in all L cells, the considered pilot sequence will be assigned to one (out of K) user in each of the L cells. Let $\mathcal{G} \triangleq \{1, \ldots, K\}$, then $\mathcal{K}_l \in \mathcal{G}$ denotes the index of the user in the l-th cell who is assigned the pilot sequence $\mathbf{s}$. The set of selected users is denoted by $\mathcal{U}$ in what follows.

We use the estimation MSE (4.21) as a performance metric to be minimized in order to find the best user set. (4.20) is an alternative if we take the estimates of interfering channels into consideration. For a given user set $\mathcal{U}$, we define a network utility function

$$\mathrm{F}(\mathcal{U}) \triangleq \sum_{j=1}^{|\mathcal{U}|} \frac{\mathcal{M}_j(\mathcal{U})}{\mathrm{tr}\left\{\mathbf{R}_j^{(j)}(\mathcal{U})\right\}}, \tag{4.34}$$

where $|\mathcal{U}|$ is the cardinal number of the set $\mathcal{U}$. $\mathcal{M}_j(\mathcal{U})$ is the estimation MSE for the desired channel at the j-th base station, with a notation readily extended from $\mathcal{M}_1$ in (4.21), where this time cell j is the target cell when computing $\mathcal{M}_j$. $\mathbf{R}_j^{(j)}(\mathcal{U})$ is the covariance matrix of the desired channel at the j-th cell.

The principle of the coordinated pilot assignment consists in exploiting covariance information at all cells (a total of KL^2 covariance matrices) in order to minimize the sum MSE metric. Hence, L users are assigned an identical pilot sequence when the corresponding L^2 covariance matrices exhibit the most orthogonal signal subspaces. Note that the MSE-based criterion (4.34) implicitly exploits the property of subspace orthogonality, e.g., at high SNRs, the proposed MSE-based criterion will be minimized by choices of users with covariance matrices showing maximum signal subspace orthogonality, thereby implicitly satisfying the conditions behind Theorem 3. In view of minimizing the search complexity, a classical greedy approach is proposed:

1) Initialize $\mathcal{U} = \emptyset$
2) For $l = 1, \ldots, L$ do:
   $\mathcal{K}_l = \arg\min_{k \in \mathcal{G}} \mathrm{F}(\mathcal{U} \cup \{k\})$
   $\mathcal{U} \leftarrow \mathcal{U} \cup \{\mathcal{K}_l\}$

End

The coordination can be interpreted as follows: To minimize the estimation error, a base station tends to assign a given pilot to the user whose spatial feature has most differences with the interfering users assigned the same pilot. Clearly, the performance will improve with the number of users, as it becomes more likely to find users with discriminable second-order statistics.

4.6 Interference filtering via subspace projection

In this section, we propose a simple beamforming strategy building on the low-dimensionality of the signal subspace, which does not require an accurate channel estimation. Consider a network where K' users share the same pilot sequence $\mathbf{s}$. These users can either be in different cells or in the same cell (e.g., in a distributed antenna system without cell boundary). Assume the first user is the target user and all other $K'-1$ users are interference users. Denote the sum of interference covariances as $\mathbf{R}_I = \mathbf{R}_2 + \mathbf{R}_3 + \cdots + \mathbf{R}_{K'}$. The eigenvalue decomposition of $\mathbf{R}_I$ is $\mathbf{R}_I = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H$, where $\mathbf{\Sigma}$ is a $M \times M$ diagonal matrix with the eigenvalues of $\mathbf{R}_I$ on its main diagonal. Suppose the eigenvalues are in descending order and the first m eigenvalues are non-negligible while the others can be neglected. We construct the spatial filter at the BS side for user 1 as:

$$\mathbf{W}_1 = \begin{bmatrix} \mathbf{u}_{m+1} & \mathbf{u}_{m+2} & \dots & \mathbf{u}_M \end{bmatrix}^H, \tag{4.35}$$

where $\mathbf{u}_m$ is the m-th column of $\mathbf{U}$. We can assume approximately that:

$$\mathbf{W}_1\mathbf{h}_k \approx \mathbf{0}, \forall k \neq 1,$$

$$\mathbf{W}_1\mathbf{Y} \approx \mathbf{W}_1\mathbf{h}_1\mathbf{s}^T + \mathbf{W}_1\mathbf{N}, \tag{4.36}$$

where $\mathbf{N} \in \mathbb{C}^{M\times\tau}$ is the spatially and temporally white additive Gaussian noise, $\mathbf{Y} \in \mathbb{C}^{M\times\tau}$ is the received training signal, and $\mathbf{s} \in \mathbb{C}^{\tau\times 1}$ is the shared pilot sequence. Define the effective channel $\underline{\mathbf{h}}_1 \triangleq \mathbf{W}_1\mathbf{h}_1$. Note that $\underline{\mathbf{h}}_1$ has a reduced size, which is $(M-m)\times 1$. An LS estimate of $\underline{\mathbf{h}}_1$ is:

$$\widehat{\underline{\mathbf{h}}}_1 = \mathbf{W}_1\mathbf{Y}\mathbf{s}^*(\mathbf{s}^T\mathbf{s}^*)^{-1}, \tag{4.37}$$

The key idea is that $\widehat{\underline{\mathbf{h}}}_1$ is coarse if considered as a channel estimate, yet it can be used as a modified MRC beamformer as it lies in a subspace orthogonal to the interference and is also aligned with the signal subspace of $\mathbf{h}_1$. During

uplink data transmission phase:

$$\mathbf{y} = \mathbf{h}_1 \mathbf{s}_1^T + \sum_{k=2}^{K'} \mathbf{h}_k \mathbf{s}_k^T + \mathbf{n}, \tag{4.38}$$

where $\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_K \in \mathbb{C}^{\tau_u \times 1}$ are the transmitted signal sequence. $\mathbf{y}, \mathbf{n} \in \mathbb{C}^{M \times \tau_u}$ are the received signal and noise respectively. The subspace-based MRC beamformer is $\widehat{\underline{\mathbf{h}}}_1^H \mathbf{W}_1$:

$$\widehat{\underline{\mathbf{h}}}_1^H \mathbf{W}_1 \mathbf{y} = \widehat{\underline{\mathbf{h}}}_1^H \underline{\mathbf{h}}_1 \mathbf{s}_1^T + \underbrace{\widehat{\underline{\mathbf{h}}}_1^H \mathbf{W}_1 \sum_{k=2}^{K'} \mathbf{h}_k \mathbf{s}_k^T}_{\approx \mathbf{0}} + \widehat{\underline{\mathbf{h}}}_1^H \mathbf{W}_1 \mathbf{n}. \tag{4.39}$$

In case there is no null space for $\mathbf{R}_I$, e.g., the number of users K' is large or the interference users have rich scattering environments, the subspace-based method can still avoid the strong eigen modes of interference and therefore reject a good amount of interference.

Note that the subspace projection method has a certain similarity with [28] which also uses eigen-value decomposition in order to perform blind channel estimation. However there are two main differences: 1) They address only the case of classical massive arrays, not distributed antenna arrays; 2) They use the received power levels domain to separate desired channel and interfering channels. In our approach, the discrimination against interference is related to the phases with which the interference and desired signals arrive at the array. In fact, the two techniques could in principle be combined.

4.7 Numerical Results

In this section we evaluate the performance of our proposed channel estimation and interference filtering schemes in simulations. In the first part of this section, we consider traditional cellular network where BS is equipped with co-located antenna array. The second part contains the numerical results for a distributed antenna system.

Two performance metrics are used to evaluate the proposed schemes. The first one is a normalized channel estimation error

$$\mathrm{MSE}_{jk} \triangleq 10\log_{10}\left(\frac{\left\|\widehat{\mathbf{h}}_{jk}^{(j)} - \mathbf{h}_{jk}^{(j)}\right\|_2^2}{\left\|\mathbf{h}_{jk}^{(j)}\right\|_2^2}\right), \tag{4.40}$$

where $\mathbf{h}_{jk}^{(j)}$ and $\widehat{\mathbf{h}}_{jk}^{(j)}$ are the (desired) channel of user k at cell j and its estimate respectively. In the simulation we average the channel estimation MSE over all users in order to obtain the MSE curve.

The second performance metric is the per-cell rate of the uplink/downlink obtained assuming MRC beamformer based on the channel estimates. We define the per-cell rate as follows:

$$\mathcal{C} \triangleq \frac{\sum_{j=1}^{L}\sum_{k=1}^{K}\log_2\left(1+\mathrm{SINR}_{jk}\right)}{L},$$

where SINR_{jk} is the signal-to-noise-plus-interference ratio (SINR) of the k-th scheduled user in the j-th cell.

4.7.1 Co-located Antenna Array

In order to preserve fairness between users and avoid having high-SNR users being systematically assigned the considered pilot, we consider a symmetric multicell network where the users are all distributed on the cell edge and have the same distance from their base stations. In practice, users with greater average SNR levels (but equal across cells) can be assigned together on a separate pilot pool. We adopt the model of a cluster of synchronized and hexagonally shaped cells. Some basic simulation parameters are given in Table 4.1. We keep these parameters in the following simulations, unless otherwise stated.

Table 4.1 – Basic simulation parameters

Cell radius	1 km
Cell edge SNR	20 dB
Number of users per-cell	10
Distance from a user to its BS	800 m
Path loss exponent	3
Carrier frequency	2 GHz
Antenna spacing	$\lambda/2$
Number of paths	50
Pilot length	10

Two types of AoA distributions are considered here, a non-bounded one (Gaussian) and a bounded one (uniform):

Gaussian distribution

For a certain channel vector $\mathbf{h}_{lk}^{(j)}$, the AoAs of all P paths are i.i.d. Gaussian random variables with mean $\bar{\theta}_{lk}^{(j)}$ and standard deviation σ. Here we suppose all the desired channels and interference channels have the same standard deviation of AoA. Note that Gaussian AoA distributions cannot fulfill the conditions of non-overlapping AoA support domains in Theorem 3, nevertheless the use of the proposed method in this context also gives substantial gains as σ^2 decreases.

Uniform distribution

For the channel $\mathbf{h}_{lk}^{(j)}$, the AoAs are uniformly distributed over $[\bar{\theta}_{lk}^{(j)} - \theta_\Delta, \bar{\theta}_{lk}^{(j)} + \theta_\Delta]$, where $\bar{\theta}_{lk}^{(j)}$ is the mean AoA.

Note that in this section, we suppose the number of users simultaneously served by each base station is one. In reality, however, a base station can serve more than one users when these users have mutually orthogonal pilot sequences.

Numerical results of the proposed channel estimation scheme are now shown. In the figures, "LS" stands for conventional LS channel estimation. "CB" denotes the Covariance-aided Bayesian estimation (without coordinated pilot assignment), and "CPA" is the proposed Coordinated Pilot Assignment-based Bayesian estimation.

In Fig. 4.3 and Fig. 4.4, the estimation MSEs versus the BS number of antennas are illustrated. When the AoAs have uniform distributions with $\theta_\Delta = 10$ degrees, as shown in Fig. 4.3, the performance of CPA estimator improves quickly with M from 2 to 10. In the 2-cell network, CPA has the ability of avoiding the overlap between AoAs for the desired and interference channels. For comparison, Fig. 4.4 is obtained with Gaussian AoA distribution. We can observe a gap remains between the CPA and the interference-free one, due to the non-boundedness of the Gaussian PDF. Nevertheless, the gains over the classical estimator remain substantial.

We then examine the impact of standard deviation σ of Gaussian AoAs on the estimation. Fig. 4.5 shows that the estimation error is a monotonically increasing function of σ. In contrast, an angle spread tending toward zero will cause the channel direction to collapse into a deterministic quantity, yielding large gains for covariance-based channel estimation.

Figs. 4.6 depicts the downlink per-cell rate achieved by the MRC beamforming strategy and suggests large gains when the Bayesian estimation is used in conjunction with the proposed coordinated pilot assignment strategy

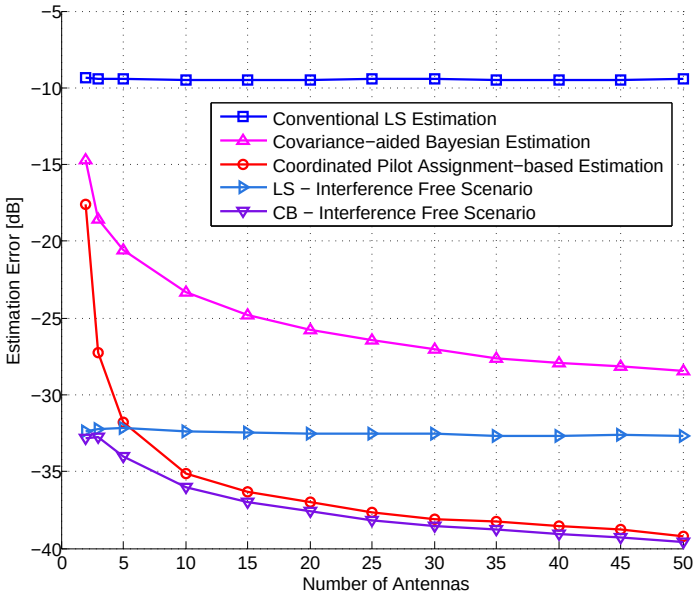

Figure 4.3 – Estimation MSE vs. number of BS antennas, uniformly distributed AoAs with $\theta_\Delta = 10$ degrees, 2-cell network.

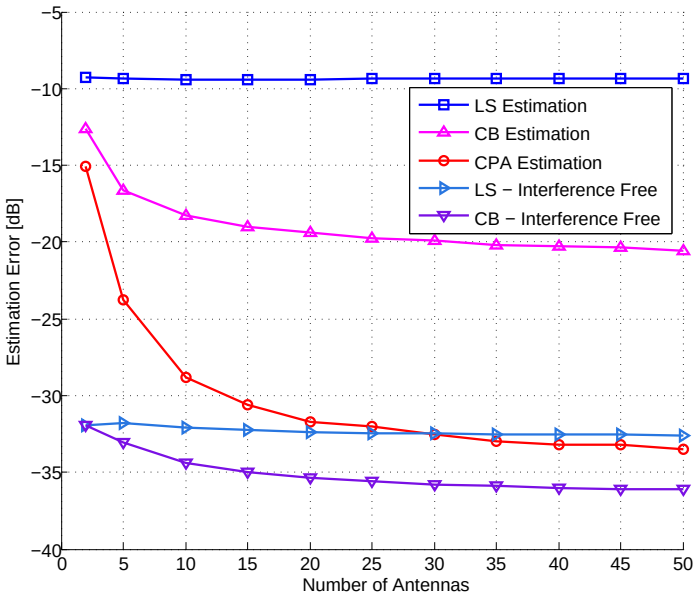

Figure 4.4 – Estimation MSE vs. number of BS antennas, Gaussian distributed AoAs with $\sigma = 10$ degrees, 2-cell network.

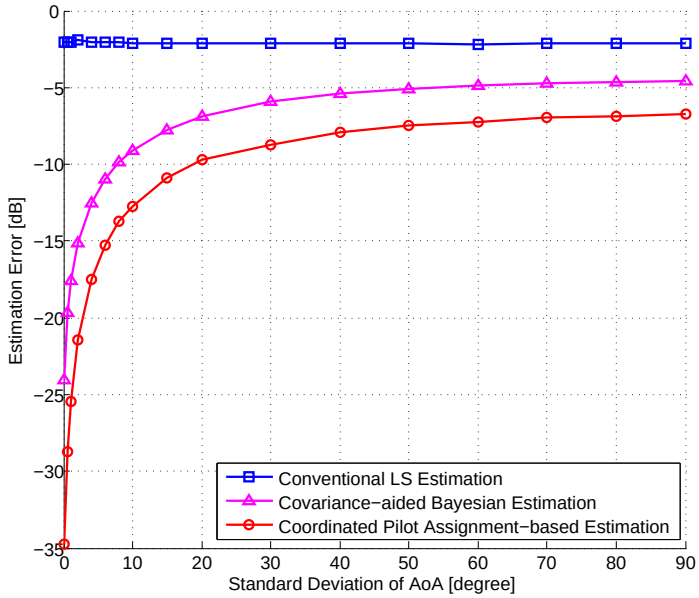

Figure 4.5 – Estimation MSE vs. standard deviation of Gaussian distributed AoAs with $M = 10$, 7-cell network.

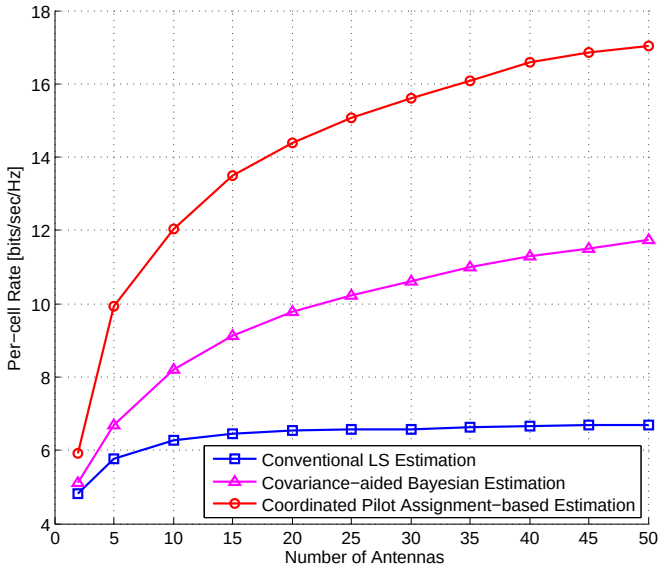

Figure 4.6 – Per-cell rate vs. number of BS antennas, 2-cell network, Gaussian distributed AoAs with $\sigma = 10$ degrees.

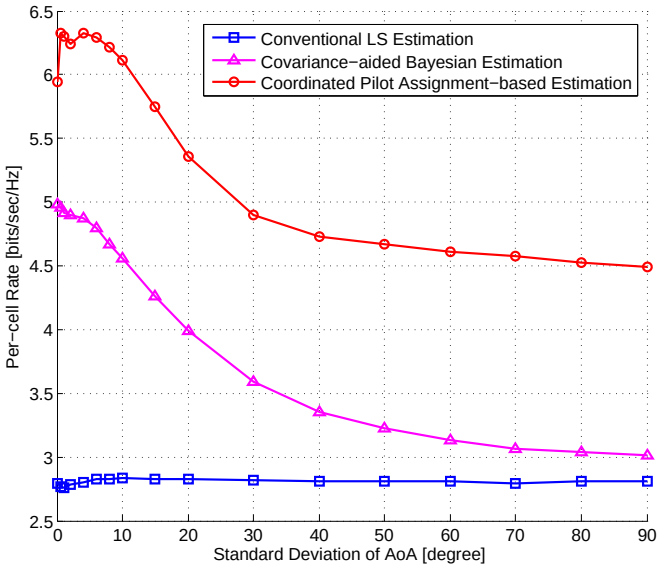

Figure 4.7 – Per-cell rate vs. standard deviation of AoA (Gaussian distribution) with $M = 10$, 7-cell network.

and intermediate gains when it is used alone. Obviously the rate performance almost saturates with M in the classical LS case (due to pilot contamination) while it increases quickly with M for the proposed estimators, indicating the full benefits of massive MIMO systems are exploited.

Finally, we show the per-cell rate performance with a small number of antennas ($M = 10$) in the setting of unbounded Gaussian distributed AoAs 4.7. As can be seen, even in traditional MIMO system with small number of BS antennas, our proposed channel estimation scheme has significant performance gains.

4.7.2 Distributed Antenna Array

We now consider the channel estimation quality in a distributed antenna system based random network with radius $R_c = 500$ meters. The path loss exponent $\gamma = 2.5$. The scattering radius is $r = 15$ meters. $P = 50$ scatterers are randomly distributed in the scattering ring, which is centered at the user.

In Fig. 4.8, we assume the target user is located at the origin while an interfering user (they share the same pilot sequence) is moving over the horizontal axis at increasing distances from user 1. As we can observe,

when the MMSE estimator (4.18) is used, the channel estimation error is a monotonous decreasing function of the distance between the desired user and the interference user. One may also notice the constant performance gap between LS and MMSE estimator in interference-free scenario, which indicates that covariance information is still helpful even in a highly distributed antenna system. As shown in the blue curve on the top, an LS estimator is unable to separate the desired channel and the interference channel. In contrast, an MMSE estimator has much better performance as its MSE is decreasing almost linearly with inter-user spacing, hence confirming our claims.

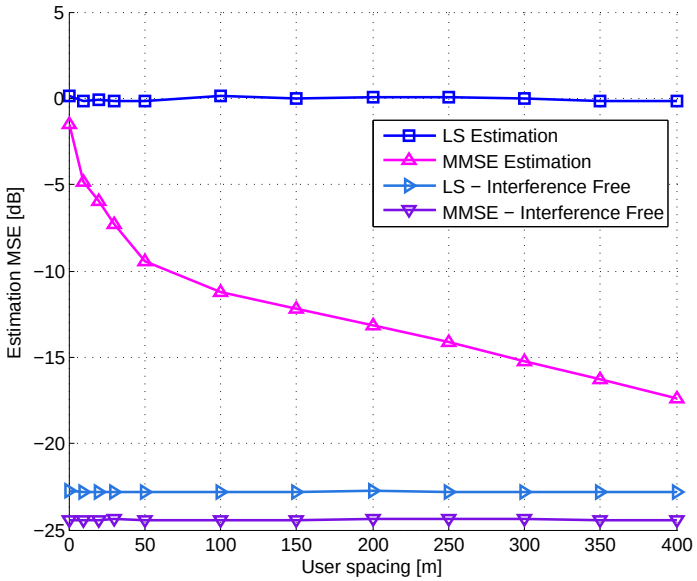

Figure 4.8 – Estimation performance vs. distance between two users, $M = 2000$, $r = 15$m, single-cell network.

We then examine the performance of four MRC beamformers in terms of uplink sum-rate in a single-cell setting (Fig. 4.9) and per-cell rate in a multi-cell setting (Fig. 4.10).

In Fig. 4.9, we show the performance of subspace-based MRC beamforming in a single-cell network where two users share the same pilot. The total number of distributed antennas is 500. In the figure "LS + MRC" denotes the sum-rate performance of MRC beamforming using the LS channel estimate, while the curve "MMSE + MRC" is the performance of MRC beamforming using the MMSE estimate (4.18). "MMSE + MMSE" denotes

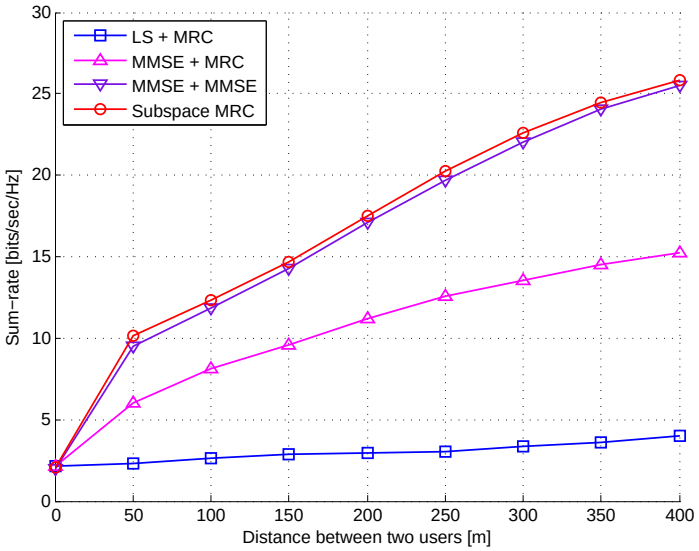

Figure 4.9 – Uplink sum-rate vs. distance between 2 users, $M = 500$, $r =$ 15m, cell-edge SNR 20dB, single-cell network.

the performance curve of MMSE beamforming using MMSE channel estimate when channel covariances (including the interference covariances) are assumed known during both channel estimation and signal detection. The simulation shows the simple subspace-based method has a very good performance. Due to pilot contamination, the MRC beamformer using MMSE channel estimate is not as good as subspace-based method. The reason is that $\mathbf{R}_1$ and $\mathbf{R}_2$ generally have overlapping signal subspaces here. We may also notice that the subspace-based MRC beamformer has some slight performance gains over the MMSE beamformer.

Fig. 4.10 depicts the uplink per-cell rate achieved by the above-mentioned MRC beamformers as a function of scattering radius r. In the simulation we have 7 hexagonal cells with one center cell and 6 surrounding cells. Each cell has one user. All the users share the same pilot sequence. The per-cell rate is defined as the sum-rate (6.19) divided by the number of cells. As can be seen, the subspace-based beamforming shows performance gains over other traditional MRC methods especially when the radius of scattering ring is smaller. It also shows more robustness than MMSE beamformer when the radius of the scattering ring is larger.

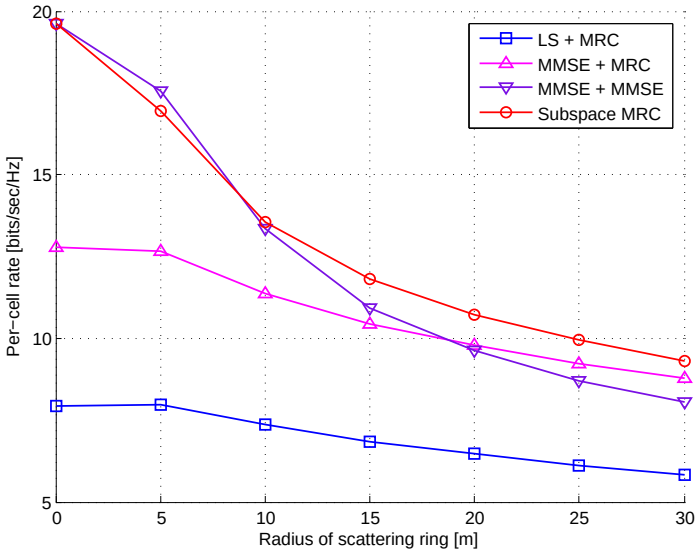

Figure 4.10 – Uplink per-cell rate vs. r, cell-edge SNR 20dB, 7-cell network, each cell has $M = 500$ distributed antennas.

4.8 Discussions

In this chapter, we assumed the individual covariance matrices can be estimated separately. This could be done in practice by exploiting resource blocks where the desired user and interference users are known to be assigned at different times. In future networks, one may imagine a specific training design for learning second-order statistics. Since covariance information varies much slower than fast fading, such training may not consume a substantial amount of resources.

4.9 Conclusions

This chapter proposes a covariance-aided channel estimation framework in the context of interference-limited multi-cell multiple antenna systems. We develop Bayesian estimators and demonstrate analytically the efficiency of such an approach for large-scale antenna systems, leading to a complete removal of pilot contamination effects in the case covariance matrices satisfy a certain non-overlapping condition on their dominant subspaces. We propose a coordinated pilot assignment strategy that helps shape covariance matrices toward satisfying the needed condition and show channel estima-

tion performance close to interference-free scenarios.

We also put forward a modified MRC beamformer relaying on channel covariance's low-rankness property. Simulation in a distributed antenna system shows that such a low-complexity method approaches and even outperforms the MMSE beamformer which has higher complexity.

Chapter 5

Joint Angle/Power based Decontamination

5.1 Introduction

In chapter 4 we have shown several interference mitigation techniques based on the long-term statistics of user channels. Asymptotic performance analysis indicates that interference can in principle be discriminated on the basis of the distributions of multipath AoAs. In particular, we demonstrated that in the case when the AoA support of the desired user does not overlap with the AoA supports of the interference users, an MMSE-based method achieves full interference elimination. However, when such a non-overlapping condition is not satisfied, interference persists in MMSE estimation, which limits the ultimate performance of massive MIMO. In this chapter, we seek to solve the remaining interference by exploiting the short-term statistics of channel information.

Two key features of massive MIMO channels that have been previously reported are of particular interest here: 1) channels of different users tend to be pairwise orthogonal when the number of antennas increases, thus leading to a specific subspace structure for the received data vectors that depend on these channels [28] and 2) the channel covariance matrix exhibits a low-rankness property whenever the multipath impinging on the MIMO array spans a finite angular spread [17,36,82]. The blind signal subspace estimation in [28] capitalizes on the first property. The second property has been utilized in [6,17,36,82,83] assuming the complete knowledge of the long-term channel covariance matrices. While the exploitation of the two properties *individually* has given rise to a set of distinct original decontamination approaches, in

this work we will exploit these two key features in a *combined* manner. Doing so we can propose a novel approach towards mitigating pilot contamination that exhibits much higher levels of robustness.

More specifically, in [28, 29], the pairwise channel orthogonality property allows to blindly estimate the user-of-interest channel subspace and discriminate between user-of-interest signals and interference based on the channel powers. In practice, decontamination occurs via a projection driven by the channel amplitudes. This approach works well within the constraint that the interference channel is received with a power level sufficiently lower than that of the desired channel, a condition which is hard to guarantee for some edge-of-cell terminals.

In a way completely different from [28,29], another approach based on a MMSE estimator is adopted in [36] to estimate the channel of interest via projection of the received signals onto the user-of-interest subspace. This subspace, identified by a channel covariance matrix (a long-term one, as opposed to the instantaneous signal correlation matrix of [28,29]), is related to the angular spread of the signal of interest [36] and enables to annihilate the interference from users with non-overlapping domains of multipath angles-of-arrival (AoA). Interestingly, this latter approach makes no assumption on received signal amplitudes and can also discriminate against interfering users that are received with similar or even higher powers. Yet, the approach fails to decontaminate pilots when propagation scattering creates large angle spread, causing spatial overlapping among desired and interference channels.

In this chapter, we point out that the strengths of these two previously unrelated estimation methods are strongly complementary, offering a unique opportunity for developing robust massive MIMO channel estimation schemes. Thus, we aim to properly merge the two projections in complementary domains while keeping the individual benefits. In fact, we propose a family of algorithms striking various performance/complexity trade-offs.

We start by presenting a first scheme named "covariance-aided amplitude based projection" that effectively combines projections in the angular and amplitude domains and exhibits robustness to interference power/angles overlapping conditions. We present an asymptotic analysis which reveals the conditions under which the channel estimation error due to pilot contamination and noise can be made to vanish. An intuitive physical interpretation of this condition for a ULA is given in the form of the residual interference channel energy contained in the multipath components that overlap in angle with those of the desired channel. Although the physical explanation is given for the ULA example, the general principle apply to other antenna placement topologies.

The obtained condition for decontamination is in general less restrictive than the condition required by previous MMSE and the amplitude projection-based methods taken separately to achieve complete removal of pilot contamination.

Following this, we then propose two low-complexity alternative schemes called "subspace and amplitude based projection" and "MMSE + amplitude based projection" respectively. Such schemes achieve different complexity-performance trade-off at moderate number of antennas. Specifically, the "subspace and amplitude based projection" can be shown to reach asymptotic (in the number of antennas) decontamination result under the same channel topology conditions as the first scheme.

5.2 UL Training/Data Transmission

Consider a multi-user multi-cell network with K users in each cell served by their own BS. Recall from chapter 3 the multi-user MIMO channel between all K users in cell l and BS j:

$$\mathbf{H}_l^{(j)} \triangleq \left[\mathbf{h}_{l1}^{(j)} \quad \mathbf{h}_{l2}^{(j)} \quad \cdots \quad \mathbf{h}_{lK}^{(j)}\right], \tag{5.1}$$

and the pilot matrix consists of all pilot sequences used by these K users:

$$\mathbf{S} \triangleq \left[\mathbf{s}_1 \quad \mathbf{s}_2 \quad \cdots \quad \mathbf{s}_K\right]^T. \tag{5.2}$$

During the training phase, the received signal at the base station j is

$$\mathbf{Y}^{(j)} = \sum_{l=1}^{L} \mathbf{H}_l^{(j)}\mathbf{S} + \mathbf{N}^{(j)}, \tag{5.3}$$

where $\mathbf{N}^{(j)} \in \mathbb{C}^{M\times\tau}$ is the spatially and temporally white additive Gaussian noise (AWGN) with zero-mean and element-wise variance σ_n^2. Then, during the uplink data transmission phase, each user transmits C data symbols. The received data signal at base station j is given by:

$$\mathbf{W}^{(j)} = \sum_{l=1}^{L} \mathbf{H}_l^{(j)}\mathbf{X}_l + \mathbf{Z}^{(j)}, \tag{5.4}$$

where $\mathbf{X}_l \in \mathbb{C}^{K\times C}$ is the matrix of transmitted symbols of all users in the l-th cell. The symbols are i.i.d. with zero-mean and unit average element-wise variance. $\mathbf{Z}^{(j)} \in \mathbb{C}^{M\times C}$ is the AWGN noise with zero-mean and element-wise variance σ_n^2. Note that the block fading channel is constant during the transmission for the τ pilot symbols and the C data symbols.

5.3 A review of LMMSE estimation

We briefly recall the MMSE channel estimator in a multi-cell single-user per cell setting. Without loss of generality, we assume cell j is the target cell, and $\mathbf{h}_j^{(j)} \in \mathbb{C}^{M\times 1}$ is the desired channel, while $\mathbf{h}_l^{(j)} \in \mathbb{C}^{M\times 1}, \forall l \neq j$ are the interference channels. We rewrite (5.3) in a vectorized form,

$$\mathbf{y}^{(j)} = \overline{\mathbf{S}} \sum_{l=1}^{L} \mathbf{h}_l^{(j)} + \mathbf{n}^{(j)}, \tag{5.5}$$

where $\mathbf{y}^{(j)} = \mathrm{vec}(\mathbf{Y}^{(j)})$, $\mathbf{n}^{(j)} = \mathrm{vec}(\mathbf{N}^{(j)})$. A pilot sequence $\mathbf{s}$ is shared by all users. The pilot matrix $\overline{\mathbf{S}}$ is given by

$$\overline{\mathbf{S}} \triangleq \mathbf{s} \otimes \mathbf{I}_M. \tag{5.6}$$

We define the channel covariance matrices

$$\mathbf{R}_l^{(j)} \triangleq \mathbb{E}\{\mathbf{h}_l^{(j)} \mathbf{h}_l^{(j)H}\} \in \mathbb{C}^{M\times M}, l = 1, \ldots, L. \tag{5.7}$$

A linear MMSE estimator for $\mathbf{h}_j^{(j)}$ is given by

$$\widehat{\mathbf{h}}_j^{(j)\mathrm{MMSE}} = \mathbf{R}_j^{(j)} \left(\tau(\sum_{l=1}^{L} \mathbf{R}_l^{(j)}) + \sigma_n^2 \mathbf{I}_M \right)^{-1} \overline{\mathbf{S}}^H \mathbf{y}^{(j)}. \tag{5.8}$$

5.3.1 Asymptotic performance of MMSE

As shown in chapter 4, for a base station equipped with a ULA, the above MMSE estimator can fully eliminate the effects of interfering channels when $M \to \infty$, under a specific "non overlap" condition on the distributions of multipath AoAs for the desired and interference channels. This condition is formalized as follows: Assume the user in cell j is our target (desired user). Denote the angular support of the desired channel as Φ_d, (i.e., the probability density function (PDF) $p_d(\theta)$ of the AoA of desired channel $\mathbf{h}_j^{(j)}$ satisfies $p_d(\theta) > 0$ if $\theta \in \Phi_d$ and $p_d(\theta) = 0$ if $\theta \notin \Phi_d$) and similarly the union of the angular supports of all interference channels $\mathbf{h}_l^{(j)} (l \neq j)$ as Φ_i. If $\Phi_d \cap \Phi_i = \emptyset$, then, as $M \to \infty$, (5.8) converges to an interference-free estimate. In practice the "non overlap" condition is hard to guarantee and the finite-M performance of the MMSE scheme will depend on angular spreading and user location, although the latter can be shaped via the use of so-called coordinated pilot assignment [36].

5.4 A review of power domain discrimination

Interestingly, angle is not the only domain where interference can be discriminated upon, as revealed from a completely different approach to pilot decontamination [28,29]. In that approach the empirical instantaneous covariance matrix built from the received data (5.4) is exploited, in contrast with the use of long-term covariance matrices in (5.8). Assume cell j is our target cell and each cell has K users. The eigenvalue decomposition (EVD) of $\mathbf{W}^{(j)}\mathbf{W}^{(j)H}/C$ is written as

$$\frac{1}{C}\mathbf{W}^{(j)}\mathbf{W}^{(j)H} = \mathbf{U}^{(j)}\mathbf{\Lambda}^{(j)}\mathbf{U}^{(j)H}, \tag{5.9}$$

where $\mathbf{U}^{(j)} \in \mathbb{C}^{M\times M} = [\mathbf{u}_1^{(j)}|\mathbf{u}_2^{(j)}|\cdots|\mathbf{u}_M^{(j)}]$ is a unitary matrix and $\mathbf{\Lambda}^{(j)} = \text{diag}\{\lambda_1^{(j)},\cdots,\lambda_M^{(j)}\}$ with its diagonal entries sorted in a non-increasing order. By extracting the first K columns of $\mathbf{U}^{(j)}$, i.e., the eigenvectors corresponding to the strongest K eigenvalues, we obtain an orthogonal basis

$$\mathbf{E}^{(j)} \triangleq \begin{bmatrix}\mathbf{u}_1^{(j)} & \mathbf{u}_2^{(j)} & \cdots & \mathbf{u}_K^{(j)}\end{bmatrix} \in \mathbb{C}^{M\times K}. \tag{5.10}$$

The basic idea in [28,29] is to use the orthogonal basis $\mathbf{E}^{(j)}$ as an estimate for the span of $\mathbf{H}_j^{(j)}$, which includes all desired user channels in cell j. Then, by projecting the received signal onto the subspace spanned by $\mathbf{E}^{(j)}$, most of the signal of interest is preserved. In contrast, the interference signal is canceled out thanks to the asymptotic property that the user channels are pairwise orthogonal as the number of antennas tends to infinity. Thus after the above mentioned projection, the estimate of the multi-user channel $\mathbf{H}_j^{(j)}$ is given by:

$$\widehat{\mathbf{H}}_j^{(j)\text{AM}} = \frac{1}{\tau}\mathbf{E}^{(j)}\mathbf{E}^{(j)H}\mathbf{Y}^{(j)}\mathbf{S}^H. \tag{5.11}$$

Note here that interference and desired channel directions are discriminated on the basis of channel amplitudes and not AoA, hence the estimate is labeled "AM" for "Amplitude". As a way to guarantee an asymptotic separation between the signal of interest and the interference in terms of power, it has been suggested to introduce power control in the network [28,29].

5.4.1 Generalized amplitude projection

As shown in [28, 29], the above method works well when the desired channels and interference channels are separable in power domain, i.e., the

instantaneous powers of any desired channels are higher than that of any interference channels. In practice however, this assumption is not always guaranteed. For a finite number of antennas, the short-term fading realization can cause the interference subspace to spill over the desired one. An enhanced version can somewhat mitigate this problem by considering a generalized amplitude based projection. This consists in selecting a possibly larger number ($\kappa^{(j)}$) of dominant eigenvectors to form $\mathbf{E}^{(j)}$, where $\kappa^{(j)}$ is the number of eigenvalues in $\mathbf{\Lambda}^{(j)}$ that are greater than $\mu\lambda_K^{(j)}$. μ is a design parameter that satisfies $0 \leq \mu < 1$. See section 5.7 for details on the choice of μ.

5.5 Covariance-aided amplitude based projection

Note that both previous methods, while being able to tackle pilot contamination in quite different ways, perform well only in some restricted user/channel topologies. For a ULA base station, the MMSE method leads to interference free channel estimates under the strict requirement that the desired and interference channel do not overlap in their AoA regions. While the amplitude based projection requires that no interference channel power exceeds that of a desired channel to achieve a similar result. Unfortunately, due to the random user location and scattering effects, it is quite unlikely to achieve these conditions at all times. As a result, by combining the useful properties of both the MMSE and the amplitude projection method, we propose below novel estimation methods that will lead to enhanced robustness in a realistic cellular scenario.

5.5.1 Single user per cell

For ease of exposition we first consider a simplified scenario where intra-cell interference is ignored by assuming that each cell has only one user, i.e., $K = 1$. The users in different cells share the same pilot sequence $\mathbf{s}$. Then with proper modifications we will generalize this method to the setting of multiple users per cell in section 5.5.3.

The objective is to combine long-term statistics which include spatial distribution information together with short-term empirical covariance which contains instantaneous amplitude and direction channel information. Hence, a spatial distribution filter can be associated to an instantaneous projection operator to help discriminate against any interference terms whose spatial directions live in a subspace orthogonal to that of the desired channel. The

intuition is that such a spatial filter may bring the residual interference to a level that is acceptable to the instantaneous projection-based channel estimator.

In order to carry out the above intuition, we introduce a long-term statistical filter $\mathbf{\Xi}_j$, which is based on channel covariance matrices in a way similar to that used by the MMSE filter in (5.8).

$$\mathbf{\Xi}_j = \left(\sum_{l=1}^{L} \mathbf{R}_l^{(j)} + \sigma_n^2 \mathbf{I}_M \right)^{-1} \mathbf{R}_j^{(j)}. \tag{5.12}$$

Note that the linear filter $\mathbf{\Xi}_j$ allows to discriminate against the interference in angular domain by projecting away from multipath AoAs that are occupied by interference. Note also that the choice of spatial filter $\mathbf{\Xi}_j$ is justified from the fact that the full information of desired channel $\mathbf{h}_j^{(j)}$ is preserved, as $\mathbf{h}_j^{(j)}$ lies in the signal space of $\mathbf{R}_j^{(j)}$. In fact, the desired channel is recoverable using another linear transformation $\mathbf{\Xi}_j{}'$:

$$\mathbf{\Xi}_j{}' \triangleq \mathbf{R}_j^{(j)^\dagger} \left(\sum_{l=1}^{L} \mathbf{R}_l^{(j)} + \sigma_n^2 \mathbf{I}_M \right), \tag{5.13}$$

as can be seen from the following equality

$$\mathbf{\Xi}_j' \mathbf{\Xi}_j \mathbf{h}_j^{(j)} = \mathbf{R}_j^{(j)\dagger} \mathbf{R}_j^{(j)} \mathbf{h}_j^{(j)} = \mathbf{V}_j \mathbf{V}_j^H \mathbf{h}_j^{(j)} = \mathbf{h}_j^{(j)}, \tag{5.14}$$

where the columns of $\mathbf{V}_j$ are the eigenvectors of $\mathbf{R}_j^{(j)}$ corresponding to non-zero eigenvalues.

The spatial filter is applied to the received data signal at base station j as

$$\widetilde{\mathbf{W}}_j \triangleq \mathbf{\Xi}_j \mathbf{W}^{(j)}. \tag{5.15}$$

The amplitude-based method as shown in section 5.4 can now be applied on the filtered received data to get rid of the residual interference. Take the eigenvector corresponding to the largest eigenvalue of the matrix $\widetilde{\mathbf{W}}_j \widetilde{\mathbf{W}}_j^H / C$:

$$\widetilde{\mathbf{u}}_{j1} = \mathbf{e}_1 \{ \frac{1}{C} \widetilde{\mathbf{W}}_j \widetilde{\mathbf{W}}_j^H \}. \tag{5.16}$$

Hence $\widetilde{\mathbf{u}}_{j1}$ can be considered as an estimate of the direction of the vector $\mathbf{\Xi}_j \mathbf{h}_j^{(j)}$.

We then cancel the effect of the pre-multiplicative matrix $\mathbf{\Xi}_j$ using $\mathbf{\Xi}'_j$ in (5.13), and we obtain an estimate of the direction of the channel vector $\mathbf{h}_j^{(j)}$ as follows:

$$\overline{\mathbf{u}}_{j1} = \frac{\mathbf{\Xi}'_j \widetilde{\mathbf{u}}_{j1}}{\left\| \mathbf{\Xi}'_j \widetilde{\mathbf{u}}_{j1} \right\|_2}. \tag{5.17}$$

Finally, the phase and amplitude ambiguities of the desire channel can be resolved by projecting the LS estimate onto the subspace spanned by $\overline{\mathbf{u}}_{j1}$:

$$\widehat{\mathbf{h}}_j^{(j)\mathrm{CA}} = \frac{1}{\tau} \overline{\mathbf{u}}_{j1} \overline{\mathbf{u}}_{j1}^H \mathbf{Y}^{(j)} \mathbf{s}^*, \tag{5.18}$$

where the superscript "CA" denotes the covariance-aided amplitude domain projection.

The algorithm is summarized below:

Algorithm 2 Covariance-aided Amplitude based Projection

1: Take the first eigenvector of $\widetilde{\mathbf{W}}_j \widetilde{\mathbf{W}}_j^H / C$ as in (5.16), with $\widetilde{\mathbf{W}}_j$ being the filtered data signal.
2: Reverse the effect of the spatial filter using (5.17).
3: Resolve the phase and amplitude ambiguities by (5.18).

The complexity of this proposed estimation scheme is briefly evaluated.

We note that the computation of the matrix inversions in (5.12) has a complexity order of $O(M^{2.37})$. However, these computations are performed in a preamble phase and their cost is negligible under the underlying assumption of channel stationarity implicitly made in this article. In practical systems, the matrix inversion in (5.12) is performed when the channel statistics are updated. Since the channel statistics are typically updated in a time scale much larger than the channel coherence time, i.e., the time scale for the applicability of Algorithm 2, then their computational cost is negligible. Therefore, we can focus on the complexity of Algorithm 2 only.

In step 1, the spatial filtering of the data signals in (5.15) and the computation of the covariance matrix $\widetilde{\mathbf{W}}_j \widetilde{\mathbf{W}}_j^H$ is performed along with the computation of the dominant eigenvector of an $M \times M$ matrix as in (5.16). The former computation has a complexity order $O(CM^2)$ while, by applying the classical power method, the computation of the dominant vector has a complexity order $O(M^2)$. Both step 2 and step 3 require multiplications of matrices by M-dimensional vectors and thus both have a complexity order

$O(M^2)$. Then, the global complexity of the algorithm is dominated by the complexity of step 1, which is $O(CM^2)$.

The ability for the above estimator to combine the advantages of the previously known angle and amplitude projection based estimators is now analyzed theoretically. In particular we are interested in the conditions under which full pilot decontamination can be achieved asymptotically in the limit of the number of antennas M and data symbols C. In order to facilitate the analysis, we introduce the following condition:

Condition C1: The spectral norm of $\mathbf{R}_l^{(j)}$ is uniformly bounded:

$$\forall M \in \mathbb{Z}^+ \text{ and } \forall l \in \{1, \ldots, L\}, \exists \zeta, \text{s.t.} \left\| \mathbf{R}_l^{(j)} \right\|_2 < \zeta, \tag{5.19}$$

where $\mathbb{Z}^+$ is the set of positive integers, and ζ is a constant.

Condition C1 can be interpreted as describing all the scenarios in which the channel energy is spread over a subspace whose dimension grows with M. Note that the same assumption can be found in some other papers, e.g., [70]. The corresponding physical condition for a ULA has been analyzed in section 3.5.

As another interpretation of the condition, it is worth noting that when this condition is not satisfied, there is no guarantee that the asymptotic pairwise orthogonality of different users' channels holds. In other words, the quantity $\mathbf{h}_j^{(j)H}\mathbf{h}_l^{(j)}/M, l \neq j$ may not converge to zero, which is an adverse condition for all massive MIMO methods. However our proposed methods still have significant performance gains under this adverse circumstance. Moreover, C1 is a sufficient condition and we believe it can be weakened.

5.5.2 Asymptotic performance of the proposed CA estimator

We now look into the performance analysis of the proposed estimation scheme. Let us define

$$\alpha_l^{(j)} \triangleq \lim_{M\to\infty} \frac{1}{M} \operatorname{tr}\{\mathbf{\Xi}_j \mathbf{R}_l^{(j)} \mathbf{\Xi}_j^H\}, \forall l = 1, \ldots, L. \tag{5.20}$$

Theorem 4. *Given condition C1, if the following inequality holds true:*

$$\alpha_j^{(j)} > \alpha_l^{(j)}, \forall l \neq j, \tag{5.21}$$

then, the estimation error of (5.18) vanishes:

$$\lim_{M,C\to\infty} \frac{\left\| \widehat{\mathbf{h}}_j^{(j)CA} - \mathbf{h}_j^{(j)} \right\|_2^2}{\left\| \mathbf{h}_j^{(j)} \right\|_2^2} = 0. \tag{5.22}$$

Proof: For the sake of notational convenience, in this proof we assume the user in cell j is the target user and thus drop the superscript (j). The desired channel is denoted by $\mathbf{h}_j = \mathbf{h}_j^{(j)}$ and the interference channels are $\mathbf{h}_l = \mathbf{h}_l^{(j)}, l \neq j$. Since $\mathbf{h}_l, l = 1, \ldots, L$, is considered as $M \times 1$ complex Gaussian with the spatial correlation matrices $\mathbf{R}_l = \mathbb{E}\{\mathbf{h}_l \mathbf{h}_l^H\}$, the channels can be factorized as [75]

$$\mathbf{h}_l = \mathbf{R}_l^{1/2} \mathbf{h}_{\mathrm{W}l}, l = 1, \ldots, L, \tag{5.23}$$

where $\mathbf{h}_{\mathrm{W}l} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$, is i.i.d. $M \times 1$ vector with unit variance. We build the proof of Theorem 4 on the general correlation model (6.6). The proof consists in three parts, corresponding to the three steps in Algorithm 2 respectively. More specifically, Lemma 3 (and the intermediate results towards Lemma 3) is the first part of the proof. It shows that $\widetilde{\mathbf{u}}_{j1}$ aligns asymptotically with the direction of the filtered channel vector $\underline{\mathbf{h}}_j = \mathbf{\Xi}_j \mathbf{h}_j$. The second part of the proof is provided in Lemma 8, which proves that after canceling the effect of the spatial filter using $\mathbf{\Xi}_j'$, we obtain the direction of the true channel $\mathbf{h}_j$ in $\overline{\mathbf{u}}_{j1}$. The final part of the proof shows that by projecting the LS estimate onto the subspace of $\overline{\mathbf{u}}_{j1}$, we resolve the phase and amplitude of the true channel.

Lemma 3. *Given condition C1, if* $\alpha_j^{(j)} > \alpha_l^{(j)}, \forall l \neq j$, *then there exists a unique* $0 \leq \phi < 2\pi$, *such that*

$$\lim_{M,C \to \infty} \left\| \frac{\underline{\mathbf{h}}_j}{\left\| \underline{\mathbf{h}}_j \right\|_2} - \widetilde{\mathbf{u}}_{j1} e^{j\phi} \right\|_2 = 0. \tag{5.24}$$

where $\underline{\mathbf{h}}_l \triangleq \mathbf{\Xi}_j \mathbf{h}_l, l = 1, \ldots, L$.

Proof: The proof of Lemma 3 relies on several intermediate results, namely Lemma 4 - Lemma 7.

Lemma 4. *Under condition C1, the spectral norm of* $\mathbf{\Xi}_j \mathbf{\Xi}_j^H$ *satisfies:*

$$\lim_{M \to \infty} \frac{1}{M} \left\| \mathbf{\Xi}_j \mathbf{\Xi}_j^H \right\|_2 = 0. \tag{5.25}$$

Proof: See Appendix .8. □

Lemma 4 indicates that the spectral norm of the covariance of the noise (after multiplying $\mathbf{\Xi}_j$) is bounded and does not scale with M. This conclusion will be exploited when we prove in Lemma 7 that the impact of noise on the dominant eigenvector/eigenvalue vanishes.

Lemma 5. *[84] Let $\mathbf{A}_M$ be a deterministic $M \times M$ complex matrix with uniformly bounded spectral radius for all M. Let $\mathbf{q} = \frac{1}{\sqrt{(M)}}[q_1, \cdots, q_M]^T$ where $q_i, \forall i = 1, \cdots, M$ is i.i.d. complex random variable with zero mean, unit variance, and finite eighth moment. Let $\mathbf{r}$ be a similar vector independent of $\mathbf{q}$. Then as $M \to \infty$,*

$$\mathbf{q}^H \mathbf{A}_M \mathbf{q} \xrightarrow{a.s.} \frac{1}{M} \operatorname{tr}\{\mathbf{A}_M\}, \tag{5.26}$$

and

$$\mathbf{q}^H \mathbf{A}_M \mathbf{r} \xrightarrow{a.s.} 0, \tag{5.27}$$

where $\xrightarrow{a.s.}$ denotes almost sure convergence.

Note that in this chapter, the condition on the finite eighth moment always holds, as when we apply Lemma 5, the components of the vector of interest are i.i.d. complex Gaussian variables. It is well known that a complex Gaussian variable with zero mean, unit variance has finite eighth moment.

Lemma 6. *Given condition C1,*

$$\lim_{M\to\infty} \frac{1}{M} \underline{\mathbf{h}}_j^H \underline{\mathbf{h}}_l = 0, \forall l \neq j \tag{5.28}$$

$$\lim_{M\to\infty} \frac{1}{M} \underline{\mathbf{h}}_l^H \underline{\mathbf{h}}_l \overset{a.s.}{=} \alpha_l, l = 1, \ldots, L. \tag{5.29}$$

Proof: See Appendix .9. □

Lemma 7. *When condition C1 is satisfied,*

$$\lim_{M,C\to\infty} \left\| \frac{\widetilde{\mathbf{W}}_j \widetilde{\mathbf{W}}_j^H}{MC} \frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2} - \alpha_j \frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2} \right\|_2 = 0, \tag{5.30}$$

Proof: See Appendix .10. □

Lemma 7 proves that as $M, C \to \infty$, α_j is an asymptotic eigenvalue of the random matrix $\widetilde{\mathbf{W}}_j \widetilde{\mathbf{W}}_j^H / MC$, with its corresponding eigenvector converging to $\underline{\mathbf{h}}_j / \left\|\underline{\mathbf{h}}_j\right\|_2$ up to a random phase.

We now return to the proof of Lemma 3. Since $\alpha_j > \alpha_l, \forall l \neq j$, one may readily obtain from Lemma 7 and (5.28):

$$\lim_{M,C\to\infty} \lambda_1 \left\{ \frac{\widetilde{\mathbf{W}}_j \widetilde{\mathbf{W}}_j^H}{MC} \right\} = \alpha_j, \tag{5.31}$$

and that there exists a unique $0 \leq \phi < 2\pi$, such that

$$\lim_{M,C\to\infty} \left\| \frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2} - \mathbf{e}_1 \left\{ \frac{\widetilde{\mathbf{W}}_j \widetilde{\mathbf{W}}_j^H}{MC} \right\} e^{j\phi} \right\|_2 = 0, \tag{5.32}$$

which completes the proof of Lemma 3. □

Now we show the second part of the proof of Theorem 4. Note that in this part we make the implicit assumption that the spectral norm of $\mathbf{\Xi}'_j$ satisfies $\left\|\mathbf{\Xi}'_j\right\|_2 < +\infty$. A sufficient (but not necessary) condition of such an assumption is that the spectral norm of $\mathbf{R}_j^{\dagger}$ is finite.

Lemma 8. *Given (5.24), we have*

$$\lim_{M,C\to\infty} \left\| \frac{\mathbf{h}_j}{\left\|\mathbf{h}_j\right\|_2} - \bar{\mathbf{u}}_{j1} e^{j\phi} \right\|_2 = 0. \tag{5.33}$$

Proof: See Appendix .11. □

The final part of the proof of Theorem 4 can be found in Appendix .12, which corresponds to step 3 of Algorithm 2. The proof shows that projecting the LS estimate onto the subspace of $\bar{\mathbf{u}}_{j1}$ will lead to noise-free estimate asymptotically as $M, C \to \infty$. This concludes the proof of Theorem 4. □

Interestingly, condition (5.21) in Theorem 4 can be replaced with

$$\left\|\mathbf{\Xi}_j \mathbf{h}_j^{(j)}\right\|_2 > \left\|\mathbf{\Xi}_j \mathbf{h}_l^{(j)}\right\|_2, \forall l \neq j, \tag{5.34}$$

which indicates that under suitable conditions on the spectral norm of channel covariance, after multiplying the filter $\mathbf{\Xi}_j$, if the power of the desired channel is higher than that of interference channel, then, pilot contamination disappears asymptotically, along with noise.

Note that we have so far no assumption on antenna placement in the analysis, other than the requirement for uniformly boundedness of the spectral norm of channel covariance. In the sequel we look into a specific model of ULA as an example and seek to further understand the physical meaning of the proposed method.

We still assume $\mathbf{h}_j^{(j)}$ is the channel of interest. Denote its angular support as Φ_d. Decompose the interference channel $\mathbf{h}_l^{(j)}, \forall l \neq j$, as follows:

$$\mathbf{h}_l^{(j)} = \mathbf{h}_{l\mathrm{i}}^{(j)} + \mathbf{h}_{l\mathrm{o}}^{(j)}, \tag{5.35}$$

where

$$\mathbf{h}_{l\mathrm{i}}^{(j)} = \frac{\beta_l^{(j)}}{\sqrt{P}} \sum_{\theta \in \Phi_d} \mathbf{a}(\theta) e^{i\varphi_\theta} \tag{5.36}$$

$$\mathbf{h}_{l\mathrm{o}}^{(j)} = \frac{\beta_l^{(j)}}{\sqrt{P}} \sum_{\theta \notin \Phi_d} \mathbf{a}(\theta) e^{i\varphi_\theta}, \tag{5.37}$$

which means $\mathbf{h}_{l\mathrm{i}}^{(j)}$ is the residual multipath component of the interference channel within the AoA region Φ_d of the desired channel, while $\mathbf{h}_{l\mathrm{o}}^{(j)}$ is the multipath component which is outside Φ_d.

Theorem 5. *For a ULA base station, under condition C1, if the residual multipath component of the interference channel satisfies:*

$$\forall l \neq j, \left\| \mathbf{\Xi}_j \mathbf{h}_{li}^{(j)} \right\|_2 < \left\| \mathbf{\Xi}_j \mathbf{h}_j^{(j)} \right\|_2, \tag{5.38}$$

then, the estimation error of the estimator (5.18) vanishes:

$$\lim_{M,C\to\infty} \frac{\left\| \widehat{\mathbf{h}}_j^{(j)CA} - \mathbf{h}_j^{(j)} \right\|_2^2}{\left\| \mathbf{h}_j^{(j)} \right\|_2^2} = 0. \tag{5.39}$$

Proof: See Appendix .13. □

Theorem 5 further confirms the fact that for a base station equipped with ULA, only the interference multipath components that overlap with those of the desired channel affect the performance of our pilot decontamination method. In other words, the spatial filter $\mathbf{\Xi}_j$ removes the energy located in all interference multipath originating from directions that do not overlap with those of the desired channel. It is then sufficient for the energy of the *residual* interference components to be below that of the desired channel to allow for a full decontamination.

5.5.3 Generalization to multiple users per cell

Now we generalize the covariance-aided amplitude based projection into multi-user setting where K users are served simultaneously in each cell. We consider the estimation of user channel $\mathbf{h}_{jk}^{(j)}$ in the reminder of this section.

Define a matrix $\mathbf{H}_{j\backslash k}^{(j)}$ as a sub-matrix of $\mathbf{H}_j^{(j)}$ after removing its k-th column:

$$\mathbf{H}_{j\backslash k}^{(j)} \triangleq \left[\mathbf{h}_{j1}^{(j)} \quad \cdots \quad \mathbf{h}_{j(k-1)}^{(j)} \quad \mathbf{h}_{j(k+1)}^{(j)} \quad \cdots \quad \mathbf{h}_{jK}^{(j)} \right]. \tag{5.40}$$

A corresponding estimate of (5.40), denoted by $\widehat{\mathbf{H}}_{j\backslash k}^{(j)}$, is obtained by removing the k-th column of $\widehat{\mathbf{H}}_{j}^{(j)}$, which can be an LS estimate, MMSE estimate, or other linear/non-linear estimate of $\mathbf{H}_{j}^{(j)}$. For demonstration purpose only, in this thesis we use the simplest LS estimate, which already shows very good performance.

In order to adapt the method in section 5.5.1 to multi-user scenario, we propose to first neutralize the intra-cell interference with a Zero-Forcing (ZF) filter $\mathbf{T}_{jk}$ based on the LS estimate $\widehat{\mathbf{H}}_{j\backslash k}^{(j)}$, and then apply the spatial filter $\mathbf{\Xi}_{jk}$. After these two filters, the data signal is now:

$$\widetilde{\mathbf{W}}_{jk} \triangleq \mathbf{\Xi}_{jk}\mathbf{T}_{jk}\mathbf{W}^{(j)}, \tag{5.41}$$

where

$$\mathbf{T}_{jk} \triangleq \mathbf{I}_M - \widehat{\mathbf{H}}_{j\backslash k}^{(j)}(\widehat{\mathbf{H}}_{j\backslash k}^{(j)H}\widehat{\mathbf{H}}_{j\backslash k}^{(j)})^{-1}\widehat{\mathbf{H}}_{j\backslash k}^{(j)H}, \tag{5.42}$$

and

$$\mathbf{\Xi}_{jk} \triangleq \left(\sum_{l=1}^{L}\mathbf{R}_{lk}^{(j)} + \sigma_n^2\mathbf{I}_M\right)^{-1}\mathbf{R}_{jk}^{(j)}. \tag{5.43}$$

The rest of this method proceeds as in the single user setting. Take the dominant eigenvector of $\widetilde{\mathbf{W}}_{jk}\widetilde{\mathbf{W}}_{jk}^H/C$:

$$\widetilde{\mathbf{u}}_{jk1} = \mathbf{e}_1\{\frac{1}{C}\widetilde{\mathbf{W}}_{jk}\widetilde{\mathbf{W}}_{jk}^H\}. \tag{5.44}$$

The estimate of the direction of $\mathbf{h}_{jk}^{(j)}$ is obtained by:

$$\overline{\mathbf{u}}_{jk1} = \frac{\mathbf{\Xi}'_{jk}\widetilde{\mathbf{u}}_{jk1}}{\left\|\mathbf{\Xi}'_{jk}\widetilde{\mathbf{u}}_{jk1}\right\|_2}, \tag{5.45}$$

where

$$\mathbf{\Xi}'_{jk} \triangleq \mathbf{R}_{jk}^{(j)\dagger}\left(\sum_{l=1}^{L}\mathbf{R}_{lk}^{(j)} + \sigma_n^2\mathbf{I}_M\right). \tag{5.46}$$

Finally the phase and amplitude ambiguities are resolved by the training sequence, and we have the estimate of $\mathbf{h}_{jk}^{(j)}$:

$$\widehat{\mathbf{h}}_{jk}^{(j)\mathrm{CA}} = \frac{1}{\tau}\overline{\mathbf{u}}_{jk1}\overline{\mathbf{u}}_{jk1}^H\mathbf{Y}^{(j)}\mathbf{S}^H. \tag{5.47}$$

Note that in this method, we build the ZF type filter $\mathbf{T}_{jk}$ based on a rough LS estimate. Further improvements can be attained with higher quality estimate

at the cost of higher complexity. As a simple example, we can reduce the effect of noise on the estimate $\widehat{\mathbf{H}}_{j\backslash k}^{(j)}$ by first applying EVD of $\mathbf{W}^{(j)}\mathbf{W}^{(j)H}/C$, then removing the subspace where the noise lies, and finally performing LS estimation. These extensions are out of the scope of this thesis.

5.6 Low-complexity alternatives

In this section, we propose two alternatives of the method shown in section 5.5, aiming at lower computational complexity at the cost of mild performance loss.

5.6.1 Subspace and amplitude based projection

The low-rankness of channel covariance implies that the uplink received desired signal lives in a reduced subspace. By projecting the received data signal $\mathbf{W}^{(j)}$ onto the signal space of $\mathbf{R}_{jk}^{(j)}$, we are able to preserve the signal from user k in cell j while remove the interference and noise that live in its complementary subspace. In the following, we show a subspace-based signal space projection method that relies on the covariance of desired channel only. For ease of exposition, we simplify the system setup to single user per cell. Let the user in cell j be the target user. The EVD of the covariance of the desired channel is

$$\mathbf{R}_j^{(j)} = \mathbf{V}_j \, \mathbf{\Sigma}_j \mathbf{V}_j^H, \tag{5.48}$$

where the diagonal entries of $\mathbf{\Sigma}_j$ contains the non-negligible eigenvalues of $\mathbf{R}_j^{(j)}$. Then we project the received data signal onto the signal space of $\mathbf{R}_j^{(j)}$, or the column space of $\mathbf{V}_j$:

$$\overline{\mathbf{W}}_j \triangleq \mathbf{V}_j \mathbf{V}_j^H \mathbf{W}^{(j)}. \tag{5.49}$$

The rest of this method follows the same idea as the covariance-aided amplitude based projection scheme. Taking the eigenvector corresponding to the largest eigenvalue of $\overline{\mathbf{W}}_j \overline{\mathbf{W}}_j^H/C$:

$$\overline{\mathbf{u}}_{j1} = \mathbf{e}_1 \left\{ \frac{1}{C} \overline{\mathbf{W}}_j \overline{\mathbf{W}}_j^H \right\}, \tag{5.50}$$

the channel estimate of $\mathbf{h}_j^{(j)}$ is given by

$$\widehat{\mathbf{h}}_j^{(j)\mathrm{SA}} = \frac{1}{\tau} \overline{\mathbf{u}}_{j1} \overline{\mathbf{u}}_{j1}^H \mathbf{Y}^{(j)} \mathbf{s}^*, \tag{5.51}$$

where the superscript "SA" stands for "subspace and amplitude based projection". Note that this method does not require the covariance of interference channels or variance of noise. It explicitly relies on the assumption that the desired covariance matrix has a low-dimensional signal subspace, with some degradations expected when this condition is not realized in practice. In fact, if $\mathbf{R}_j^{(j)}$ has full rank, this method degrades to pure amplitude based projection.

Note that this "SA" estimator has lower complexity than the "CA" estimator (5.18) in the sense that 1) "SA" estimator does not require the statistical knowledge of the interference channels or the variance of the noise, and 2) "SA" estimator skips step 2 in Algorithm 2.

The physical condition under which full decontamination is achieved with this method is shown below in the case of a ULA. We denote the angular support of desired channel $\mathbf{h}_j^{(j)}$ by Φ_d and the multipath components of the interference channel $\mathbf{h}_l^{(j)}$ falling in Φ_d as $\mathbf{h}_{\mathrm{li}}^{(j)}$.

Theorem 6. *For a ULA base station, if the power of interference channel that falls into the angular support Φ_d satisfies*

$$\forall l \neq j, \left\|\mathbf{h}_{li}^{(j)}\right\|_2 < \left\|\mathbf{h}_j^{(j)}\right\|_2, \tag{5.52}$$

and the channel covariance satisfies

$$\forall M \in \mathbb{Z}^+, \forall l \neq j, \left\|\mathbf{R}_j^{(j)\frac{1}{2}} \mathbf{V}_j \mathbf{V}_j^H \mathbf{R}_l^{(j)\frac{1}{2}}\right\|_2 < +\infty, \tag{5.53}$$

then, the estimation error of the estimator (5.51) vanishes:

$$\lim_{M,C\to\infty} \frac{\left\|\widehat{\mathbf{h}}_j^{(j)SA} - \mathbf{h}_j^{(j)}\right\|_2^2}{\left\|\mathbf{h}_j^{(j)}\right\|_2^2} = 0. \tag{5.54}$$

Proof: Due to lack of space, we skip the complete proof and only give two key steps below. By applying the asymptotic orthogonality between two steering vectors which are associated with different AoAs (Lemma 3 in [36]), we may readily obtain:

$$\lim_{M\to\infty} \frac{1}{\sqrt{M}} \mathbf{V}_j \mathbf{V}_j^H \mathbf{h}_j^{(j)} = \frac{1}{\sqrt{M}} \mathbf{h}_j^{(j)} \tag{5.55}$$

$$\lim_{M\to\infty} \frac{1}{\sqrt{M}} \mathbf{V}_j \mathbf{V}_j^H \mathbf{h}_l^{(j)} = \frac{1}{\sqrt{M}} \mathbf{h}_{\mathrm{li}}^{(j)}, \forall l \neq j, \tag{5.56}$$

which means the multipath components of interference that fall outside Φ_d disappear asymptotically after the projection by $\mathbf{V}_j\mathbf{V}_j^H$. Then, equation (5.53) ensures that

$$\lim_{M\to\infty}\frac{1}{M}\underline{\mathbf{h}}_j^{(j)H}\underline{\mathbf{h}}_l^{(j)}=0, l\neq j, \tag{5.57}$$

where

$$\underline{\mathbf{h}}_l^{(j)}\triangleq\mathbf{V}_j\mathbf{V}_j^H\mathbf{h}_l^{(j)}, l=1,\ldots,L. \tag{5.58}$$

□

Note that in Theorem 6 condition (5.53) is less restrictive than the uniformly boundedness of the spectral norm of the channel covariance. In the special case of zero angular spread, the rank of channel covariance becomes one. Denote the deterministic AoA from the user in cell l to base station j as $\overline{\theta}_l^{(j)}$. We can easily see that the channel estimation error of (5.51) vanishes completely as $M, C\to\infty$ as long as

$$\forall l\neq j, \overline{\theta}_l^{(j)}\neq\overline{\theta}_j^{(j)}, \tag{5.59}$$

which occurs with probability one.

When channel covariance is not available, we can still benefit from the subspace projection method by approximating $\mathbf{V}_j$ with a subset of discrete Fourier transform (DFT) basis as shown in [85]. This DFT basis can be chosen based on a small number of channel observations. The generalization to multi-user case can be done by introducing the ZF filter (5.42) as in section 5.5.3. Due to lack of space, we skip the details.

5.6.2 MMSE + amplitude based projection

Another alternative is to directly project the MMSE estimate onto the subspace of $\mathbf{E}^{(j)}$ obtained by EVD of $\mathbf{W}^{(j)}\mathbf{W}^{(j)H}/C$ as in section 5.4. The estimator for the multi-user channel $\mathbf{H}_j^{(j)}$ is given by:

$$\widehat{\mathbf{H}}_j^{(j)\mathrm{MA}}=\overline{\mathbf{E}}^{(j)}\overline{\mathbf{E}}^{(j)H}\overline{\mathbf{R}}_j^{(j)}\left(\tau(\sum_{l=1}^{L}\overline{\mathbf{R}}_l^{(j)})+\sigma_n^2\mathbf{I}_{KM}\right)^{-1}\overline{\mathbf{S}}^H\mathbf{y}^{(j)}, \tag{5.60}$$

where

$$\overline{\mathbf{E}}^{(j)}\triangleq\mathbf{I}_K\otimes\mathbf{E}^{(j)}, \tag{5.61}$$

$$\overline{\mathbf{S}} \triangleq \mathbf{S}^T \otimes \mathbf{I}_M = \left[\begin{array}{ccc} \mathbf{s}_1 \otimes \mathbf{I}_M & \cdots & \mathbf{s}_K \otimes \mathbf{I}_M \end{array} \right], \tag{5.62}$$

and

$$\overline{\mathbf{R}}_l^{(j)} = \text{diag}\{\mathbf{R}_{l1}^{(j)}, ..., \mathbf{R}_{lK}^{(j)}\}, l = 1, \ldots, L. \tag{5.63}$$

The superscript "MA" denotes MMSE + amplitude based projection. It is worth noting that both the amplitude-based projection and angular-based projection require large number of antennas to achieve complete decontamination. In contrast, the MMSE estimator is efficient with very small number of antennas. As M grows, MMSE estimator starts to reduce interference earlier than the previously proposed methods, as will be shown by simulations in section 5.7. However, unlike the previously proposed schemes, this "MA" estimator cannot achieve complete decontamination when the interference channel is overlapping with desired channel in both angular and power domains.

5.7 Numerical Results

This section contains numerical results of our different channel estimation schemes compared with prior methods. In the simulation, we have multiple hexagonally shaped adjacent cells in the network. The radius of each cell is 1000 meters. Each base station has M antennas, which forms a ULA, with half wavelength antenna spacing. The length of pilot sequence is $\tau = 10$.

Two performance metrics are considered. The first is the normalized channel estimation error

$$\epsilon \triangleq \frac{1}{KL} \sum_{j=1}^{L} \sum_{k=1}^{K} \left(\frac{\left\| \widehat{\mathbf{h}}_{jk}^{(j)} - \mathbf{h}_{jk}^{(j)} \right\|_2^2}{\left\| \mathbf{h}_{jk}^{(j)} \right\|_2^2} \right). \tag{5.64}$$

The estimation errors in the plots are obtained in Monte Carlo simulations and finally displayed in dB scale.

The second metric is the uplink per-cell rate when MRC receiver (based on the obtained channel estimate) is used at base station side.

In all simulations presented in this section, we assume that the channel covariance matrix is estimated using 1000 exact channel realizations. The multipath angle of arrival of any channel (including the interference channel) follows a uniform distribution centered at the direction corresponding to line-of-sight (LoS). The number of multipath is $P = 50$. According to the coherence time model in [86], for a mobile user moving at a vehicular speed of 70 km/h in an environment of 2.6 GHz carrier frequency and 5μs high

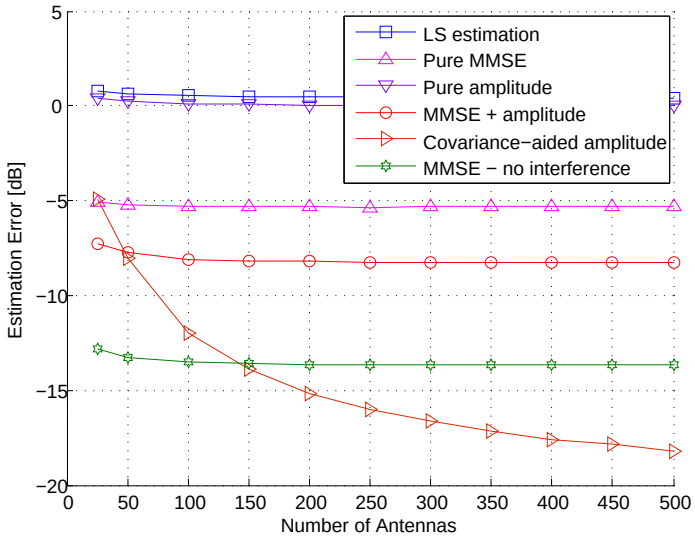

Figure 5.1 – Estimation performance vs. M, 2-cell network, 1 user per cell, path loss exponent $\gamma = 0$, partially overlapping angular support, AoA spread 60 degrees, SNR = 0 dB.

delay spread (corresponding to an excess distance of 1.5 km), the channel can be assumed coherent over 500 transmitted symbols. Thus we will let $C = 500$ in simulations, although larger coherence time can be expected in practice for a user with lower mobility.

Note that in all simulations, the amplitude-based projection and MMSE + amplitude based projection follow the enhanced eigenvector selection strategy shown in section 5.4.1 with the design parameter $\mu = 0.2$.

We first illustrate Theorem 4 in Fig. 5.1. Suppose we have a two-cell network, with each cell having one user. In order to make the interference overlapping in power domain with the desired signal, we set the path loss exponent $\gamma = 0$. The power of the interference channel has equal probability to be higher or lower than the power of the desired channel. The user in each cell is deliberately put in a symmetrical position such that the multi-path angular supports of the interference and the desired channel are half overlapping with each other.

In the figure, "LS estimation" and "Pure MMSE" denote the system performances when an LS estimator and an MMSE estimator (5.8) are used respectively. "Pure amplitude" denotes the case when we apply the generalized amplitude based projection method only. "MMSE + amplitude" represents

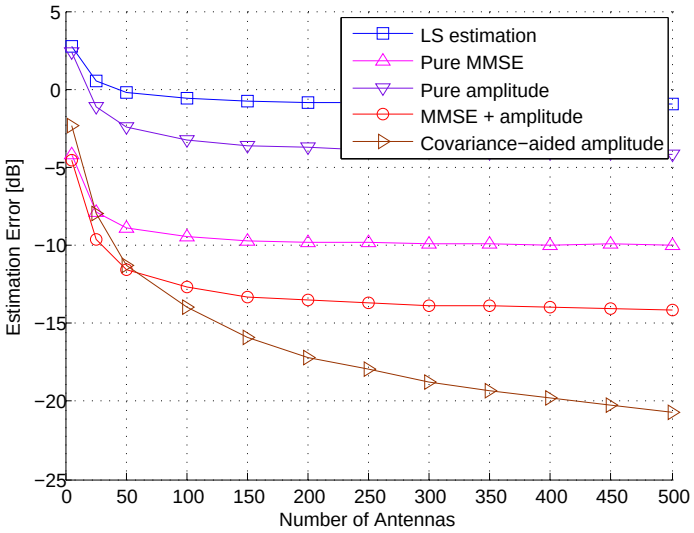

Figure 5.2 – Estimation performance vs. M, 7-cell network, one user per cell, AoA spread 30 degrees, path loss exponent $\gamma = 2$, cell-edge SNR = 0 dB.

the proposed estimator (5.60). "Covariance-aided amplitude" denotes the proposed covariance-aided amplitude based projection method (5.18). The curve "MMSE - no interference" shows the estimation error of an MMSE estimator in an interference-free scenario. As can be seen from Fig. 5.1, due to the overlapping interference in both angle and power domains, the performance of all estimators saturate quickly with the number of antennas, except the proposed covariance-aided amplitude based projection method, which eventually outperforms interference-free MMSE estimation. [1]

In Fig. 5.2 and Fig. 5.3, we show the performance of estimation error and the corresponding uplink per-cell rate for a 7-cell network, with single user per cell. The users are assumed to be distributed randomly and uniformly within their own cells excluding a central disc with radius 100 meters. The angular spread of the user channel (including interference channel) is 30 degrees. The path loss exponent is now $\gamma = 2$. As we may observe, the traditional LS estimator suffers from severe pilot contamination. The pure amplitude based method and the pure MMSE method alleviate the pilot

1. The reason is that the performance of the interference-free MMSE estimation has a non-vanishing lower bound due to white Gaussian noise. On the contrary, our proposed covariance-aided amplitude based projection method eliminates the effects of noise and interference asymptotically.

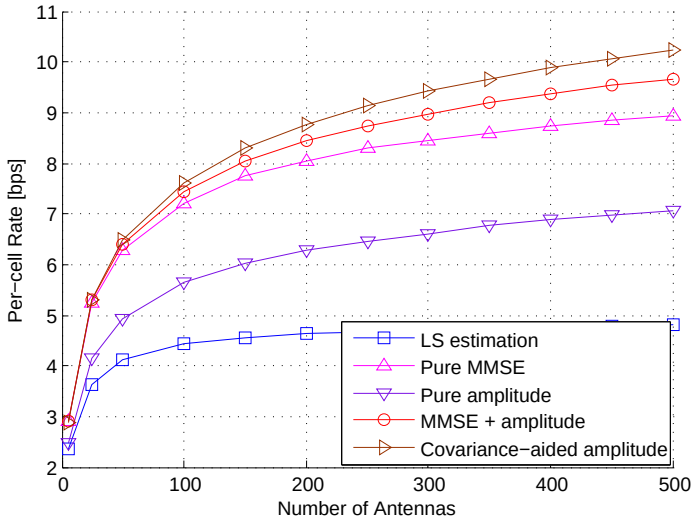

Figure 5.3 – Uplink per-cell rate vs. M, 7-cell network, one user per cell, AoA spread 30 degrees, path loss exponent $\gamma = 2$, cell-edge SNR = 0 dB.

interference, yet saturate with the number of antennas. These saturation effects come from the overlapping of the interference and the desired channels in power and angular domains respectively. The "MMSE + amplitude" approach outperforms these two known methods as it discriminates against interference in both amplitude and angular domains. However this scheme cannot cope with the case of overlapping in both domains. Owing to its robustness, the covariance-aided amplitude projection method outperforms the rest in terms of both estimation error and uplink per-cell rate.

We now turn our attention to multi-cell multi-user scenario. Fig. 5.4 and Fig. 5.5 show the channel estimation performance and the corresponding uplink per-cell rate for a 7-cell network with each cell having 4 users. In these two figures, we add the curve of subspace and amplitude based projection, which is denoted in the figures as "Subspace + amplitude". The other parameters remain unchanged compared with those in Fig. 5.2 and Fig. 5.3. We can notice that in Fig. 5.4 the covariance-aided amplitude projection method has some performance loss with respect to the low-complexity MMSE + amplitude method and the MMSE method when the number of antennas is small. It is due to the following two facts: 1) when M is small, it is well known that MMSE works well, but not the amplitude based methods, and 2) with small M, the asymptotical orthogonality of channels of different users is not fully exhibited, and consequently a small amount of signal of interest

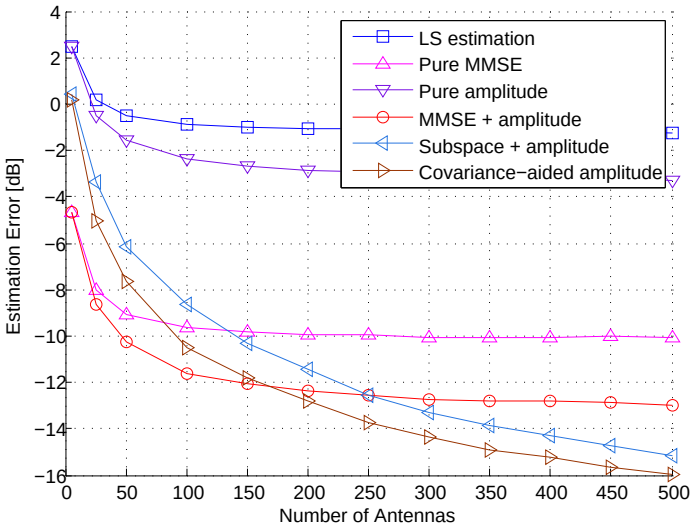

Figure 5.4 – Estimation performance vs. M, 7-cell network, 4 users per cell, AoA spread 30 degrees, path loss exponent $\gamma = 2$, cell-edge SNR = 0 dB.

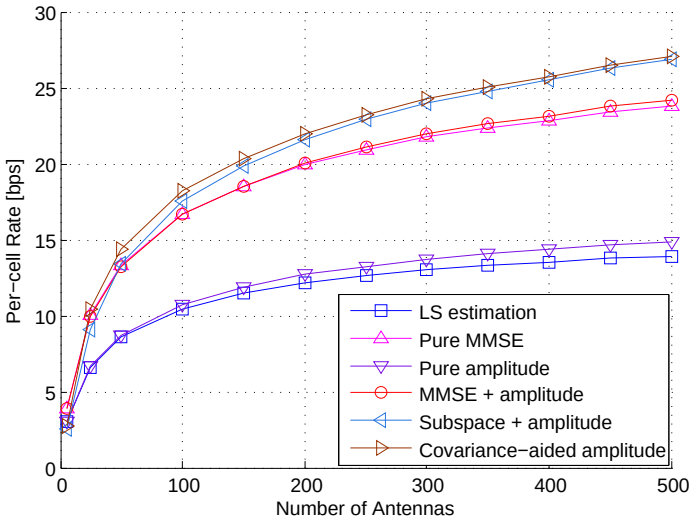

Figure 5.5 – Uplink per-cell rate vs. M, 7-cell network, 4 users per cell, AoA spread 30 degrees, path loss exponent $\gamma = 2$, cell-edge SNR = 0 dB.

is removed by the ZF filter $\mathbf{T}_{kj}$, along with intra-cell interference. However it is not disturbing in the sense that 1) as the number of antennas grows, the covariance-aided amplitude projection method quickly outperforms the other methods; and 2) The per-cell rate of this proposed method is still good even with moderate number of antennas, e.g., $M > 25$. It is also interesting to note that the low-complexity alternative scheme, subspace and amplitude based projection method, has some minor performance loss, yet keeps approximately the same slope as the covariance-aided amplitude projection.

5.8 Conclusions

In this chapter we proposed a series of robust channel estimation algorithms exploiting path diversity in both angle and amplitude domains. The first method called "covariance-aided amplitude based projection" is robust even when the desired channel and the interference channels overlap in multipath AoA and are not separable just in terms of power. Two low-complexity alternative schemes were proposed, namely "subspace and amplitude based projection" and "MMSE + amplitude based projection". Asymptotic analysis shows the condition under which the channel estimation error converges to zero.

Chapter 6

Cooperative Feedback Design in FDD

In previous chapters, we have addressed one of the fundamental problems of massive MIMO - pilot contamination in TDD deployment. In this chapter, we deal with another challenge - CSI acquisition for massive MIMO in FDD setting. In particular, we propose CSI feedback reduction methods based on novel approaches of feedback design in FDD massive MIMO systems. We exploit the synergies between massive MIMO systems and inter-user communications based on D2D. The exchange of local CSI among users, enabled by D2D communications, makes available global CSI at the terminals. Thus, we can construct more *informative* forms of feedback based on this shared knowledge. Two feedback variants are highlighted: 1) cooperative CSI feedback, and 2) cooperative precoder index feedback. For a given feedback overhead, the sum-rate performance is assessed and the gains compared with a conventional massive MIMO setup without D2D are shown.

6.1 Introduction

Very large antenna array or massive MIMO networks have impressive potentials to combat interference [10, 13, 87] based on simple beamforming techniques without requiring complex inter-cell coordination approaches. The challenge of this kind of systems concerns the CSI acquisition at the access point, which is crucial for downlink transmission. In [10], Marzetta limited the applicability of massive MIMO networks to TDD mode. By appealing to the reciprocity principle, TDD mode enables the acquisition of the CSI for downlink by channel estimation in the uplink via an open-loop

feedback scheme that avoids costly feedback. Although the reuse of the same set of pilot sequences in adjacent cells, usually referred to as pilot contamination, seems to have severe detrimental effects on the spectral [10] and power efficiency [88] of the massive MIMO networks compared with ideal CSI knowledge. Nevertheless the promised gains are still of several orders of magnitude [88] and fueled intensive research activities on massive MIMO networks in TDD mode.

In setups where channel reciprocity does not hold such as massive MIMO systems in FDD mode, closed-loop feedback is required which consists of a preliminary phase that we refer to as channel sounding, where the access point transmits training sequences for channel estimation or local CSI acquisition at each user terminal. In a second phase, shortly after CSI feedback, each UT retransmits its local CSI such that the global CSI is available at the access point. Using traditional closed-loop approaches, both the length of training sequences and the necessary feedback for large antenna arrays can become prohibitive. Very recently, building on the reduced rankness of UT channel covariance matrices in massive MIMO systems [17,36,82], promising schemes for FDD mode have been proposed in [17, 83, 89–91]. In [17, 83, 89], the authors refer to the subspace actually spanned by the UT channel as the *effective channel* and they cluster UTs with almost completely overlapping effective channels. A pre-beamforming designed using only second-order channel statistics and projecting signals of each cluster onto the effective channel enables a drastic reduction of the training sequence length. Further reductions are also possible by restricting the projection to *reduced-dimension effective channel*. Then, a reduced amount of feedback is required to design the precoder on this latter subspace. In [90,91], the authors exploit the hidden joint sparsity of the channel for clusters of UTs to reduce both training and feedback by applying compressed sensing techniques.

In [83], a drastic reduction of the required CSI is obtained in the ideal case as the access point is equipped with a large - theoretically infinite - uniform linear antenna array and each cluster consists of UTs located on a ray with origin at the access point. In those ideal conditions the selection of a reduced effective channel with negligible performance loss is strongly simplified and the required training length and CSI is drastically reduced.

The low-rankness of UT channel covariance matrices for more general antenna array settings has been studied in [36, 82]: under more realistic conditions, with UTs of a cluster randomly located in a given sector and arbitrary topology of large antenna array, the dimensions of the effective channel might be still too large.

In order to reduce the amount of feedback in non-reciprocal setups, we

propose a three-phase *cooperative* closed-loop feedback as an alternative to the traditional per-user feedback loop. The new scheme exploits a novel synergy between multiuser networks with access points equipped with multiple antenna arrays and D2D communications. More specifically, we introduce an intermediate phase in the classical closed-loop feedback: once estimated the channel parameters in the channel sounding phase, the UTs in a cluster exchange the acquired local CSI such that the global CSI is available to a master receiver in the case of centralized processing or to all the receivers in the case of distributed processing. The availability of the global CSI enables a *joint optimized design of the feedback*. We propose two methods to design the feedback under a constraint on the total amount of bits available for the feedback. The first method performs an optimal selection of the reduced effective subspace based on instantaneous knowledge of the global CSI. Then, the coefficients of the reduced effective subspace are quantized and retransmitted. The second approach benefits from the knowledge of the global CSI by selecting the best precoder from a predefined codebook. In both cases we adopt as optimality criterion the maximization of the sum-rate. In the first approach a zero-forcing precoder is implemented at the access point. Both schemes show clear performance improvements compared with the reference massive MIMO system without D2D.

6.2 Signal and Channel Models for FDD

We consider a massive MIMO base station serving K single-antenna users in the cell. The base station has M antennas and operates in FDD mode. The downlink channel between the base station and the k-th user is denoted by $\mathbf{h}_k^H \in \mathbb{C}^{1\times M}$. The full downlink channel can therefore be represented by

$$\mathbf{H}^H = \begin{bmatrix} \mathbf{h}_1^H \\ \vdots \\ \mathbf{h}_K^H \end{bmatrix}_{\in \mathbb{C}^{K\times M}} . \tag{6.1}$$

The downlink transmission is modeled by:

$$\mathbf{y} = \mathbf{H}^H \mathbf{B}\mathbf{s} + \mathbf{n}, \tag{6.2}$$

where $\mathbf{y} \in \mathbb{C}^{K\times 1}$ is the received signal at all users, $\mathbf{s} \in \mathbb{C}^{K\times 1}$ denotes the vector of i.i.d. Gaussian signals with zero-mean and unit-variance, and $\mathbf{n}$ represents the spatially and temporally additive white Gaussian noise with zero-mean and element-wise variance σ_n^2. $\mathbf{B}$ is the downlink beamformer which has the total power P.

We investigate a scenario where a group of K users are located close to each other so they can be assumed to share similar spatial statistics of channels. Hence we assume the channel covariances of these users are identical, i.e., $\forall k, \mathbb{E}\{\mathbf{h}_k \mathbf{h}_k^H\} = \mathbf{R}$. Due to limited angle spread followed by incoming paths originating from high-level base station, the channel covariance $\mathbf{R}$ typically exhibits low-rank property [36] [83] [82]. We denote the rank of $\mathbf{R}$ as d. Applying eigen value decomposition (EVD) to $\mathbf{R}$:

$$\mathbf{R} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^H \tag{6.3}$$

Without loss of generality, we assume the eigenvalues in $\mathbf{\Sigma}$ are in descending order, so that the first d eigenvalues are non-negligible while the others can be neglected. We extract the first d columns of $\mathbf{U}$ in order to form a submatrix $\mathbf{U}_1 \in \mathbb{C}^{M \times d}$. The columns in $\mathbf{U}_1$ are ranked in descending order according to their average powers (or their corresponding eigenvalues). Now the channel vector $\mathbf{h}_k$ is in the column space of $\mathbf{U}_1$, e.g., $\forall k, \mathbf{h}_k$ is a linear combination of the columns of $\mathbf{U}_1$. We may write:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \cdots & \mathbf{h}_k \end{bmatrix} = \mathbf{U}_1 \mathbf{A}, \tag{6.4}$$

where $\mathbf{A} \in \mathbb{C}^{d \times K}$ is defined as:

$$\mathbf{A} \triangleq \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_K \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & \cdots & \vdots \\ a_{d1} & a_{d2} & \cdots & a_{dK} \end{bmatrix}. \tag{6.5}$$

The channel vector $\mathbf{h}_k, 1 \leq k \leq K$, is assumed to be $M \times 1$ complex Gaussian, undergoing correlation due to the finite multipath angle spread at the base station side [75]:

$$\mathbf{h}_k = \mathbf{R}^{1/2} \mathbf{h}_{\mathrm{W}k} = \mathbf{U}\mathbf{\Sigma}^{1/2}\mathbf{U}^H \mathbf{h}_{\mathrm{W}k}, k = 1, 2, \ldots, K, \tag{6.6}$$

where $\mathbf{h}_{\mathrm{W}k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is the spatially white $M \times 1$ SIMO channel, $\mathbf{I}_M$ is the $M \times M$ identity matrix, and $\mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ denotes zero-mean complex Gaussian distribution with covariance matrix $\mathbf{I}_M$. From (6.6) and (6.5) we may readily obtain the distribution of $\mathbf{a}_k$ as:

$$\mathbf{a}_k \sim \mathcal{CN}(0, \mathbf{\Sigma}_1), 1 \leq k \leq K, \tag{6.7}$$

where $\mathbf{\Sigma}_1$ is a diagonal matrix with the d greatest eigenvalues of $\mathbf{R}$ on its diagonal in decending order.

6.3 Feedback Design without D2D

A traditional feedback strategy for multiuser system is to let each user quantize its downlink channel vector and then send the quantized information back to the base station [5]. In the subsequent precoding stage the user signals cannot be made perfectly orthogonal to each other because of the quantization errors [92]. Note, in massive MIMO regime only partial channel information (namely K coefficients per user) is needed to achieve orthogonality between user signals. This gives a first step towards increasing the CSI quality with given amount of feedback overhead. This is highlighted by simple Proposition 7 below.

By extracting N ($K \leq N \leq d$) rows from $\mathbf{A}$, we form a matrix $\mathbf{A}_s \in \mathbb{C}^{N\times K}$; and by extracting the corresponding N columns of $\mathbf{U}_1$, we form a matrix $\mathbf{U}_s \in \mathbb{C}^{M\times N}$. We can partially reconstruct the channel matrix as follows:

$$\widetilde{\mathbf{H}} \triangleq \mathbf{U}_s\mathbf{A}_s. \tag{6.8}$$

A zero-forcing (ZF) beamformer based on the incomplete CSI can be written as:

$$\mathbf{B} = \frac{\sqrt{P}\widetilde{\mathbf{H}}^{\dagger}}{||\widetilde{\mathbf{H}}^{\dagger}||_{\mathrm{F}}}, \tag{6.9}$$

where $||\cdot||_{\mathrm{F}}$ denotes the Frobenius norm, and $\widetilde{\mathbf{H}}^{\dagger}$ is the Moore-Penrose pseudoinverse:

$$\widetilde{\mathbf{H}}^{\dagger} = \widetilde{\mathbf{H}}^{H}(\widetilde{\mathbf{H}}\widetilde{\mathbf{H}}^{H})^{-1}.$$

Proposition 7. *The ZF beamformer (6.9) is able to eliminate inter-user interference completely.*

Proof: We may rewrite $\widetilde{\mathbf{H}}^{\dagger}$ as follows:

$$\widetilde{\mathbf{H}}^{\dagger} = \mathbf{U}_s\mathbf{A}_s(\mathbf{A}_s^H\mathbf{A}_s)^{-1}. \tag{6.10}$$

The channel model is now

$$\mathbf{y} = \mathbf{A}^H\mathbf{U}^H\mathbf{U}_s\mathbf{A}_s(\mathbf{A}_s^H\mathbf{A}_s)^{-1}\frac{\sqrt{P}}{||\widetilde{\mathbf{H}}^{\dagger}||_{\mathrm{F}}}\mathbf{s} + \mathbf{n}. \tag{6.11}$$

Since the following equation holds:

$$\mathbf{A}^H\mathbf{U}^H\mathbf{U}_s = \mathbf{A}_s^H, \tag{6.12}$$

the received signal vector is

$$\mathbf{y} = \frac{\sqrt{P}}{||\widetilde{\mathbf{H}}^{\dagger}||_{\mathrm{F}}}\mathbf{s} + \mathbf{n}. \tag{6.13}$$

Hence interference is nulled, proving Proposition 7. □

Proposition 7 indicates that we may form an $\mathbf{A}_s$ by extracting *any* linearly independent K out of d rows of $\mathbf{A}$. If the base station knows this incomplete CSI, it will ensure the users receive zero interference after downlink beamforming (6.9). When CSI exchange between users is not possible, a typical choice of eigenvectors is the first K rows in $\mathbf{U}_1$, as they are the strongest K eigen modes in statistical point of view. Denote the index of the i-th chosen eigenvector as e_i, and define the set of chosen indices as $\mathcal{G} \triangleq \{e_1, \cdots, e_N\}$ so that $\forall i, 1 \leq e_i \leq d$, and that $\forall i \neq j, e_i \neq e_j$. When CSI exchange between users is not allowed, the users select the following set of rows from $\mathbf{A}$:

$$\mathcal{G}^{(1)} = \{1, \cdots, K\}, \tag{6.14}$$

which is known by the base station by default. The users will only send back a quantized version of $\mathbf{A}_s$, i.e., the first K rows of $\mathbf{A}$.

6.4 Cooperative Feedback of CSI with D2D

Once CSI exchange is allowed between users, the users can make a joint decision of which set of eigenvectors in $\mathbf{U}_1$ to choose, or equivalently which set of rows of $\mathbf{A}$ to extract in order to form $\mathbf{A}_s$. As a simple example, we may consider the signal-to-noise ratio (SNR) at user side as a criterion:

$$\mathrm{SNR} = \frac{P}{||\widetilde{\mathbf{H}}^{\dagger}||_{\mathrm{F}}^2} = \frac{P}{\mathrm{tr}\{(\mathbf{A}_s^H\mathbf{A}_s)^{-1}\}}. \tag{6.15}$$

We choose K out of d eigenvectors from $\mathbf{U}_1$ so that the SNR is maximized.

$$\mathcal{G}^{(2)} = \underset{N=K}{\arg}\ \min \mathrm{tr}\{(\mathbf{A}_s^H\mathbf{A}_s)^{-1}\}. \tag{6.16}$$

The optimality of the eigenvector selection decision is achieved by exhaustive search. This problem is loosely reminiscent of the TX antenna selection in conventional MIMO systems. The number of possible candidates is $\binom{d}{K}$, which scales exponentially with K. Despite its optimality, the exhaustive method has a high computational complexity, which necessitates

low-complexity selection algorithms. Gorokhov et al. proposed a decremental selection algorithm in [7] [8] for the purpose of antenna selection. In his approach, the rows of the channel matrix are removed one by one, while minimizing the capacity degradation. This approach can be adapted to the selection of eigenvectors. We start with the full effective channel $\mathbf{A}$. Under the condition that the SNR reduction is minimum, the rows of $\mathbf{A}$ are removed one by one, until we have K rows left. The searching space is reduced from $\binom{d}{K}$ down to an order of d^2.

After the joint decision is made, i.e., $\mathcal{G}^{(2)}$ is obtained, the users will send back the corresponding $\mathbf{A}_s$ (quantized), as well as the indices in $\mathcal{G}^{(2)}$ to the base station.

6.5 Cooperative Feedback of Precoder Index with D2D

We now consider another cooperative feedback design approach based on precoder feedback. Local CSI exchange via D2D communications enables users to jointly choose a precoding matrix, which makes the multi-user system analogous to a point-to-point MIMO system. Some classical precoder selection schemes [93, 94] can directly apply. A precoder codebook is composed of a finite set of precoding matrices predetermined a priori. Such a codebook is used to approximate the precoder, e.g., the normalized Moore-Penrose pseudoinverse of $\mathbf{A}$ for a ZF type precoder. The codewords can be generated according to the distribution of $\mathbf{A}$, which is given in (6.7). The codebook design and optimization are non-trivial and out of the scope of this thesis. Nevertheless a simple random codebook generation method is given in Section 6.6.2. After local CSI exchange, the users jointly select the best precoding matrix according a certain criterion and feed back the index of the selected precoding matrix. The selection criterion may vary depending on the complexity requirement of the system. However an intuitive choice is the downlink sum-rate. Assume the codebook $\mathcal{X}$ is known at both the base station and the master UT. Any codeword $\mathbf{X} \in \mathcal{X}$ is of the size $d \times K$. The codewords are normalized such that $\forall \mathbf{X} \in \mathcal{X}, ||\mathbf{X}||_F = \sqrt{P}$. If $\mathbf{X}$ is chosen, the downlink precoding matrix is $\mathbf{U}_1\mathbf{X}$. Since the users know the instantaneous channel, they are able to compute the downlink sum-rate when a certain precoder $\mathbf{X}$ is selected. We define $\underline{\mathbf{H}} \triangleq \mathbf{A}^H\mathbf{X}$. The downlink data model is now

$$\mathbf{y} = \underline{\mathbf{H}}\mathbf{s} + \mathbf{n}. \tag{6.17}$$

The rate of user k is:

$$r_k \triangleq \log_2 \left(1 + \frac{|\underline{\mathbf{H}}_{k,k}|^2}{\sigma_n^2 + \sum_{l \neq k} |\underline{\mathbf{H}}_{k,l}|^2} \right), \tag{6.18}$$

where $\underline{\mathbf{H}}_{k,l}$ stands for the (k,l)-th element of $\underline{\mathbf{H}}$. The downlink sum-rate is defined as

$$\mathcal{C} \triangleq \sum_{k=1}^{K} r_k. \tag{6.19}$$

A description of the algorithm is as follows:

(1) The slave UTs send their measured downlink CSI to the master UT. The master UT now has the effective channel matrix $\mathbf{A}$.

(2) The master UT searches the codebook and finds the index of the precoder that maximizes the sum-rate:

$$i = \underset{X_i \in \mathcal{X}}{\arg \max} \{\mathcal{C}\}. \tag{6.20}$$

(3) The master UT feeds back the index i to the base station.

(4) The base station performs downlink beamforming (6.17) using the selected precoder.

6.6 Numerical Results

This section contains numerical evaluations of different feedback mechanisms. First we introduce our physical channel model. We assume the base station antennas form a uniform linear array (ULA). It is worthwhile to note that the proposed methods of this thesis and their results also hold for some other different settings of antenna placement, e.g., the random linear array, two dimensional uniform array, or even the two-dimensional distributed array [82]. For ease of exposition we take a ULA for example. The downlink channel between the base station and the k-th user is obtained by the following model [1]:

$$\mathbf{h}_k^H = \sqrt{\beta_k} \sum_{q=1}^{Q} \left(\mathbf{a}(\theta_{kq})\right)^H e^{j\varphi_{kq}}, \tag{6.21}$$

where Q is the number of i.i.d. paths, β_k denotes the path loss for channel $\mathbf{h}_k$, and it is dependent on the prescribed average SNR at cell edge. $e^{j\varphi_{kp}}$

is the i.i.d. random phase, which is independent over channel index k and path index q. $\mathbf{a}(\theta)$ is the signature (or phase response) vector by the array to a path originating from the angle θ, as shown in [2]

$$\mathbf{a}(\theta) \triangleq \begin{bmatrix} 1 \\ e^{-j2\pi \frac{D}{\lambda}\cos(\theta)} \\ \vdots \\ e^{-j2\pi \frac{(M-1)D}{\lambda}\cos(\theta)} \end{bmatrix}, \tag{6.22}$$

where D is the antenna spacing at the base station and λ is the signal wavelength.

In simulations of this chapter, we assume the users share the same scattering environment, giving rise to identical channel covariance matrix for all users. In other words, θ_{kq} has an i.i.d. distribution, and $\forall k, \beta_k = \beta$. This is the worst case scenario due to the resemblance of the channel of all users.

We consider a cluster of 3 single-antenna UTs being served by a base station equipped with 50 antennas. The cell radius is 1000 meters and the users are located 800 meters away from the base station. The angle of departure (AoD) of any user channel follows a uniform distribution from 80 degrees to 100 degrees, i.e., $\forall k, \forall q, \theta_{kq} \sim \mathcal{U}(80, 100)$. Due to limited angle spread (20 degrees), the rank of the channel covariance $\mathbf{R}$ is around $d = 15$ (see [36]). Since channel covariance is assumed known by the users, we consider only the reduced-dimension subspace (effective subspace), where the 50×1 channel vector can be effectively represented by a linear combination of 15 eigenvectors of the channel covariance.

When CSI exchange is not allowed between users, the traditional approach is that each user quantizes its own channel and sends back the quantized CSI to the base station. Then the base station designs a precoder based on the quantized CSI. We introduce this method as a reference system.

6.6.1 Cooperative CSI Feedback

Three different CSI feedback regimes are evaluated and the sum-rate performances are given in Fig. 6.1. For the sake of fairness, we assume the same amount of quantization bits is available in the three regimes. Each user will send back 16 bits of information to the base station. The curve "Quantize full CSI, no D2D" denotes the sum-rate performance of the conventional method when the full effective channel matrix $\mathbf{A}$ is quantized and sent back to base station. In this approach each user quantizes its effective channel vector, i.e., its corresponding column in $\mathbf{A}$, into 16 bits. Upon the reception of

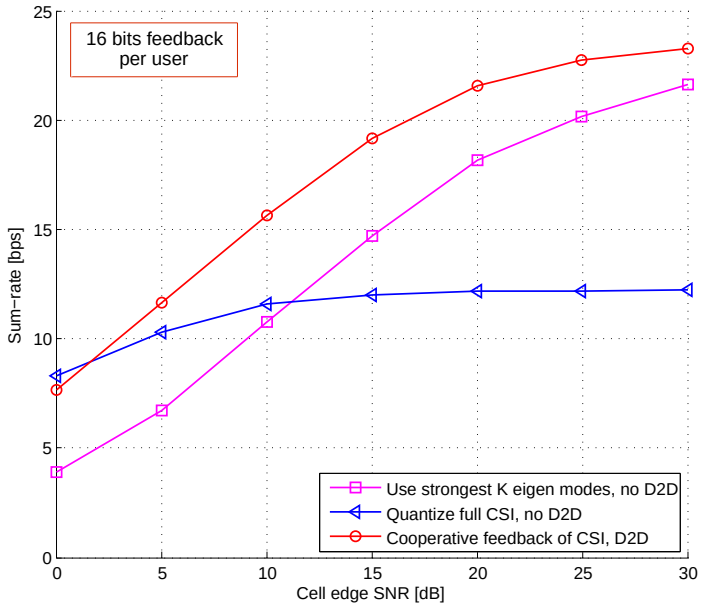

Figure 6.1 – DL sum-rates with/without feedback cooperation, feedback overhead: 16 bits per user.

users' feedback, the base station constructs the effective channel matrix and performs ZF precoding based on it. The curve "Use strongest K eigen modes, no D2D" shows the performance of each user using K dominant eigenvectors of $\mathbf{R}$, as described in the reference system (6.14). The curve "Cooperative feedback of CSI, D2D" refers to the proposed feedback regime where the users exchange CSI and search using the decremental selection method for the best 3 eigen modes. The users need to feed back the quantized projections and the indices of the three eigenvectors that are chosen. We omit the result of exhaustive search due to the fact that it has higher complexity yet negligible performance improvement compared with the low-complexity decremental method. Despite the fact that the D2D method has fewer quantization bits available, which results from the requirements of feeding back the indices, we still observe a clear performance gain of the D2D method.

6.6.2 Cooperative Precoder Index Feedback

In the following we will evaluate the performance of precoder index feedback method. We keep the simulation settings the same as in section 6.6.1 except that the amount of feedback overhead is now 4 bits per user. We still

work on the effective subspace. The curve "Quantize full CSI, no D2D" and "Use strongest K eigen modes, no D2D" in Fig. 6.2 denote the same non-cooperative methods as shown in section 6.6.1. We omit the performance of cooperative CSI feedback here due to the lack of available quantization bits, as the representation of the indices of the chosen eigen modes alone requires $\log_2 \binom{d}{K} \approx 9$ bits.

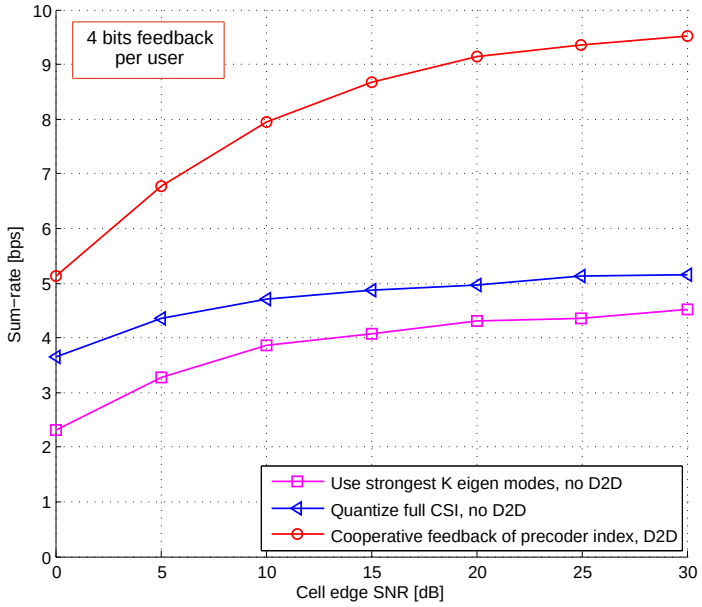

Figure 6.2 – DL sum-rates of cooperative precoder selection and non-cooperative CSI feedback, feedback overhead: 4 bits per user.

Once D2D is enabled, the users can compute the system performance, i.e., the sum-rates, when different precoders in the codebook are used, and finally pick the best precoder. Note that in simulation, we generate a random codebook as follows: 1) generate a fixed number of realizations of $\mathbf{A}$ according to its distribution (6.7); 2) compute the normalized Moore-Penrose pseudoinverse of $\mathbf{A}$ for each realization. For the sake of fairness, the cooperative precoder index feedback scheme, marked with "Cooperative feedback of precoder index, D2D", also has totally 12 bits of feedback overhead available. We can observe significant gain of this method over the traditional ones when D2D is not possible.

Finally, we would like to remark that the cooperative precoder index feedback approach has higher complexity due to exhaustive search. However

it has better performance than cooperative CSI feedback, especially when the feedback overhead is small.

6.7 Conclusions

In this chapter we propose a new cooperative feedback framework for FDD massive MIMO whereby devices rely on local CSI exchange so as to compute a suitable feedback signal. We show two approaches for feedback design, cooperative CSI feedback and cooperative precoder index feedback. These methods help reduce feedback overhead for FDD massive MIMO systems for a given sum-rate performance target, compared with the conventional non-cooperative feedback design.

Chapter 7

Conclusion

This thesis addresses the challenges of CSI acquisition of massive MIMO in both TDD mode and FDD mode.

The first part of this thesis concerns mainly the channel estimation in TDD mode. We tackle the pilot contamination problem, which was believed to constitute a bottleneck of the performance of massive MIMO systems. We show that under certain non-overlapping condition on the multipath AoA distribution for a ULA base station, pilot contamination effect can be made to vanish completely using an MMSE channel estimator. This non-overlapping condition ensures that the interference falls into the null space of the covariance matrix of desired channel asymptotically. This is mainly due to the low-rankness property of channel covariance matrix, which is identified and proved in this thesis. Furthermore, we show that such a low-rank property is not limited to ULA. It can be generalized to non-uniform array, and even to two-dimensional distributed large scale arrays. We then propose coordinated pilot assignment method based on the second-order statistics of user channels. This method is proved to be powerful in discriminating across interfering users with even identical pilot sequences.

Although the above-mentioned MMSE-based estimation schemes can lead to pilot contamination elimination under the strict non-overlapping condition, in reality however, due to the randomness of users' locations, this condition is unlikely to hold at all times. Thus, we propose a series of robust channel estimation schemes that exploit both the long-term statistical information of the channels and the short-term instantaneous amplitudes of the channels. Asymptotic analysis and simulation show that the proposed methods require milder conditions to achieve full pilot decontamination compared to known methods.

In the second part of this thesis, we consider the CSI acquisition of massive MIMO in FDD mode. Novel cooperative feedback mechanisms are proposed based on D2D communications, aiming at reducing the amount of CSI feedback. In these schemes, the users are allowed to exchange CSI with each other so as to jointly design the feedback. We show that for a given amount of feedback overhead, the new methods have significant sum-rate performance gain over the traditional schemes without D2D.

Appendices

.1 Proof of Lemma 1:

Proof:

Define the series

$$x_i \triangleq -1 + \frac{2(i-1)}{M}, i = 1, \ldots, M,$$

and

$$\boldsymbol{\mu}_i \triangleq \frac{\boldsymbol{\alpha}(x_i)}{\sqrt{M}}.$$

Then we have $\boldsymbol{\mu}_i \subset \mathcal{A}, \forall i = 1, \ldots, M$ and

$$\boldsymbol{\mu}_k^H \boldsymbol{\mu}_i = \frac{1 - e^{-j2\pi(i-k)}}{M(1 - e^{-\frac{j2\pi(i-k)}{M}})} = 0, k \neq i.$$

Thus $\{\boldsymbol{\mu}_i \,|i = 1, \ldots, M\}$ forms an orthogonal basis of $\mathcal{A}$, and therefore

$$\dim\{\mathcal{A}\} = M.$$

Define

$$\tilde{B} \triangleq \left\{ \boldsymbol{\mu}_i \middle| i \in \mathbb{Z} \cap \left[\lfloor \frac{M(b_1+1)}{2} + 1 \rfloor + 1, \lceil \frac{M(b_2+1)}{2} + 1 \rceil - 1 \right] \right\},$$

where $\lceil x \rceil$ and $\lfloor x \rfloor$ are rounded-above and rounded-below operators respectively. Then $\tilde{B}$ is part of an orthogonal basis of the space $\mathcal{B}$, which indicates $\dim\{\mathcal{B}\} \geqslant |\tilde{B}|$. By counting vectors in $\tilde{B}$, we have that

$$\begin{aligned} \dim\{\mathcal{B}\} &\geqslant \lceil \frac{M(b_2+1)}{2} + 1 \rceil - \lfloor \frac{M(b_1+1)}{2} + 1 \rfloor - 1 \\ &= \lceil \frac{M(b_2+1)}{2} \rceil - \lfloor \frac{M(b_1+1)}{2} \rfloor - 1. \end{aligned} \tag{1}$$

Now we define

$$\tilde{C} \triangleq \left\{ \boldsymbol{\mu}_i \middle| i \in \mathbb{Z} \text{ and } i \in \left[1, \lfloor \frac{M(b_1+1)}{2} + 1 \rfloor\right] \cup \left[\lceil \frac{M(b_2+1)}{2} + 1 \rceil, M \right] \right\}.$$

Then $\tilde{C}$ is part of an orthogonal basis of $\mathcal{A}$. Furthermore,

$$\begin{aligned} |\tilde{C}| &= \lfloor \frac{M(b_1+1)}{2} + 1 \rfloor + M - \lceil \frac{M(b_2+1)}{2} + 1 \rceil + 1 \\ &= M - \lceil \frac{M(b_2+1)}{2} \rceil + \lfloor \frac{M(b_1+1)}{2} \rfloor + 1. \end{aligned}$$

Consider the equivalent form of $\mathcal{B}$

$$\mathcal{B} = \left\{ \int_{b_1}^{b_2} f(x)\boldsymbol{\alpha}(x)dx \middle| \forall |f(x)| < \infty,\ x \in [b_1, b_2] \right\}.$$

Take any vector $\boldsymbol{\mu}_i \in \tilde{C}$, we have

$$\begin{aligned}\boldsymbol{\mu}_i^H \int_{b_1}^{b_2} f(x)\boldsymbol{\alpha}(x)dx &= \frac{1}{\sqrt{M}} \int_{b_1}^{b_2} f(x)\boldsymbol{\alpha}(x_i)^H \boldsymbol{\alpha}(x)dx \\ &= \frac{1}{\sqrt{M}} \int_{b_1}^{b_2} f(x) \frac{1 - e^{-j\pi M(x-x_i)}}{1 - e^{-j\pi(x-x_i)}} dx.\end{aligned}$$

Since $\boldsymbol{\mu}_i \in \tilde{C}$, we can observe $x_i \notin [b_1, b_2]$, thus

$$\lim_{M\to\infty} \boldsymbol{\mu}_i^H \int_{b_1}^{b_2} f(x)\boldsymbol{\alpha}(x)dx = 0.$$

Therefore $\tilde{C} \subset \mathcal{B}^\perp$ when $M \to \infty$. Hence we have

$$\begin{aligned}&\dim\{\mathcal{B}^\perp\} = M - \dim\{\mathcal{B}\} \geqslant |\tilde{C}| \\ \Rightarrow &\dim\{\mathcal{B}\} \leqslant \lceil \frac{M(b_2+1)}{2} \rceil - \lfloor \frac{M(b_1+1)}{2} \rfloor - 1. \end{aligned} \tag{2}$$

Combining (1) and (2), we can easily obtain

$$\begin{aligned}\dim\{\mathcal{B}\} &\sim \lceil \frac{M(b_2+1)}{2} \rceil - \lfloor \frac{M(b_1+1)}{2} \rfloor - 1 \\ &\sim \frac{M(b_2 - b_1)}{2} + o(M),\end{aligned}$$

and Lemma 1 is proved. □

.2 Proof of Theorem 1:

Proof: We define

$$\mathbf{b}(x) \triangleq \mathbf{a}\left(\cos^{-1}(x\frac{\lambda}{D})\right), x \in [-\frac{D}{\lambda}, \frac{D}{\lambda}]. \tag{3}$$

It is clear from Lemma 1 that $\mathbf{b}(x) = \boldsymbol{\alpha}(2\mathrm{x})$. Hence, for any interval $[x^{\min}, x^{\max}]$ in $[-\frac{1}{2}, \frac{1}{2}]$,

$$\begin{aligned}&\dim\left\{\mathrm{span}\left\{\mathbf{b}(x), \forall x \in [x^{\min}, x^{\max}]\right\}\right\} \\ &\sim (x^{\max} - x^{\min})M \quad \text{when } M \text{ is large.}\end{aligned} \tag{4}$$

Additionally, we have

$$\text{span}\left\{\mathbf{R}\right\} = \text{span}\left\{\int_0^{\pi} \mathbf{a}(\theta)\mathbf{a}(\theta)^H p(\theta)d\theta\right\},$$

Thus, due to the bounded support of $p(\theta)$, we can obtain

$$\begin{aligned}\text{span}\left\{\mathbf{R}\right\} &= \text{span}\left\{\int_{\theta^{\min}}^{\theta^{\max}} \mathbf{a}(\theta)\mathbf{a}(\theta)^H p(\theta)d\theta\right\} \\ &= \text{span}\left\{\int_{\theta^{\min}}^{\theta^{\max}} \mathbf{b}(\frac{D}{\lambda}\cos(\theta))\mathbf{b}^H(\frac{D}{\lambda}\cos(\theta))p(\theta)d\theta\right\}.\end{aligned}$$

Then, by interpreting the integral as a (continuous) sum, we have

$$\text{span}\left\{\mathbf{R}\right\} \subset \text{span}\left\{\mathbf{b}(x), \forall x \in [\frac{D}{\lambda}\cos(\theta^{\max}), \frac{D}{\lambda}\cos(\theta^{\min})]\right\}.$$

From (4), we obtain

$$\text{rank}(\mathbf{R}) \leq \left(\cos(\theta^{\min}) - \cos(\theta^{\max})\right)\frac{D}{\lambda}M,$$

for large M, and Theorem 1 is proved. □

.3 Proof of Proposition 5:

Proof: Denote the associated path loss as β. The covariance $\mathbf{R}$ is a Toeplitz matrix, with its mn-th entry given by

$$\begin{aligned}\mathbf{R}(m,n) &= \beta\int_0^{\pi} p(\theta)e^{j2\pi\frac{D}{\lambda}(n-m)\cos(\theta)}d\theta \qquad (5)\\ &= \beta\int_{-1}^{1} p\left(\arccos(x)\right)e^{j2\pi\frac{D}{\lambda}(n-m)x}\frac{1}{\sqrt{1-x^2}}dx \\ &= \frac{\beta\lambda}{2\pi D}\int_{-\frac{2\pi D}{\lambda}}^{\frac{2\pi D}{\lambda}} \frac{p\left(\arccos(\frac{\lambda x}{2\pi D})\right)}{\sqrt{1-\left(\frac{\lambda x}{2\pi D}\right)^2}}e^{j(n-m)x}dx, \\ &= \frac{1}{2\pi}\int_{-\frac{2\pi D}{\lambda}}^{\frac{2\pi D}{\lambda}} f(x)e^{j(n-m)x}dx, \qquad (6)\end{aligned}$$

where

$$f(x) \triangleq \frac{\beta\lambda}{D}\frac{p\left(\arccos(\frac{\lambda x}{2\pi D})\right)}{\sqrt{1-\left(\frac{\lambda x}{2\pi D}\right)^2}}. \qquad (7)$$

Since $0, \pi \notin \Phi$, or in other words, $p(0) = p(\pi) = 0$, and that $p(\theta) < \infty, \forall \theta \in \Phi$, it follows that $f(x)$ is uniformly bounded:

$$f(x) < +\infty, -\frac{2\pi D}{\lambda} \leq x \leq \frac{2\pi D}{\lambda}. \tag{8}$$

Thus, the Toeplitz matrix $\mathbf{R}$ is related to the real integrable and uniformly bounded generating function $f(x)$, with its entries being Fourier coefficients of $f(x)$. We now resort to the known result on the spectrum of the $n \times n$ Toeplitz matrices $\mathbf{T}_n(f)$ defined by the generating function $f(x)$. Denote by ess inf and ess sup the essential minimum and the essential maximum of f, i.e., the infimum and the supremum of f up to within a set of measure zero. Let $m_f \triangleq \text{ess inf} f$ and $M_f \triangleq \text{ess sup} f$.

Theorem 7. *[95] If $\lambda_0^{(n)} \leq \lambda_1^{(n)} \leq \cdots \leq \lambda_{n-1}^{(n-1)}$ are the eigenvalues of $\mathbf{T}_n(f)$, then, the spectrum of $\mathbf{T}_n(f)$ is contained in (m_f, M_f); moreover $\lim_{n\to\infty} \lambda_0^{(n)} = m_f$ and $\lim_{n\to\infty} \lambda_{n-1}^{(n-1)} = M_f$.*

By invoking Theorem 8, we obtain that $\lim_{M\to\infty} \|\mathbf{R}\|_2 = M_f < \infty$. In addition, for any finite M, the inequality $\|\mathbf{R}\|_2 < \infty$ always holds true. This concludes the proof. □

.4 Proof of Proposition 3:

Proof: We first consider an N-antenna ULA with aperture $\mathcal{D}$ and antenna spacing $D = \mathcal{D}/(N-1)$. Define

$$\begin{aligned}\tilde{\boldsymbol{\alpha}}(x) &\triangleq \left[0, e^{-j2\pi\frac{D}{\lambda}x}, \cdots, e^{-j2\pi\frac{D(N-1)}{\lambda}x}\right]^T \\ &= \left[0, e^{-j2\pi\frac{\mathcal{D}}{\lambda(N-1)}x}, \cdots, e^{-j2\pi\frac{\mathcal{D}(N-1)}{\lambda(N-1)}x}\right]^T.\end{aligned}$$

Now we define $\tilde{\mathcal{A}} \triangleq \text{span}\{\tilde{\boldsymbol{\alpha}}(x), x \in [b_1, b_2]\}$. Recall from section 3.2.2 the following result:

If $\boldsymbol{\beta}(\text{x}) \triangleq [\ 1 \quad e^{-j\pi x} \quad \cdots \quad e^{-j\pi(N-1)x}\ ]^T$. Given $b_1, b_2 \in [-1, 1]$ and $b_1 < b_2$, define $\mathcal{A} \triangleq \text{span}\{\boldsymbol{\beta}(\text{x}), \text{x} \in [b_1, b_2]\}$, when N is large,

$$\dim\{\mathcal{A}\} = \frac{(b_2 - b_1)N}{2} + o(\frac{(b_2 - b_1)N}{2}). \tag{9}$$

The above conclusion can directly apply: when N is large,

$$\begin{aligned}\dim\{\tilde{\mathcal{A}}\} &= \frac{N\mathcal{D}}{(N-1)\lambda}(b_2-b_1)+o\left(\frac{N\mathcal{D}}{(N-1)\lambda}(b_2-b_1)\right)\\ &= \frac{DN}{\lambda}(b_2-b_1)+o\left(\frac{DN}{\lambda}(b_2-b_1)\right)\\ &= \frac{M\overline{D}}{\lambda}(b_2-b_1)+o\left(\frac{M\overline{D}}{\lambda}(b_2-b_1)\right)\\ &= \frac{M\overline{D}}{\lambda}(b_2-b_1)+o(M).\end{aligned}$$

We can observe that $\dim\{\tilde{\mathcal{A}}\}$ has no dependency on N. Imagine for any finite aperture $\mathcal{D}$, we let $N\to\infty$ so that $D\to 0$. In this case, all elements of $\boldsymbol{\alpha}(x)$ can be seen as M (finite) random samples in the vector $\tilde{\boldsymbol{\alpha}}(x)$. Hence

$$\dim\{\mathcal{B}\}\le\dim\{\tilde{\mathcal{A}}\}=\frac{M\overline{D}}{\lambda}(b_2-b_1)+o(M).$$

Now consider the space $\mathcal{C}$. We define $\mathcal{C}_q\triangleq\text{span}\{\boldsymbol{\alpha}(x),x\in[b_q^{\min},b_q^{\max}]\}$. An upper bound of its dimension can be obtained by considering the extreme case when all Q spaces are mutually orthogonal so that their dimensions can add up:

$$\dim\{\mathcal{C}\}\le\sum_{q=1}^{Q}\dim\{\mathcal{C}_q\}=\sum_{q=1}^{Q}\frac{M\overline{D}}{\lambda}(b_q^{\max}-b_q^{\min})+o(M).$$

Thus, Proposition 3 is proven. □

.5 Proof of Theorem 2:

Proof:

For ease of exposition we omit the user index k. Imagine a special case when the scatterers are located in a line which has the length $\tilde{L}$, as shown in Fig. 1. Assume the antennas are far away so that the scatterers are in the same planar wavefront region. We denote the right end of the scattering line as the reference point. The m-th antenna is located $\tilde{d}_m$ meters away from the reference point, at the angle θ_m. The p-th scatter is $\tilde{l}(p)$ meters away from the reference point. $\tilde{l}(p)$ follows a uniform distribution, i.e., $\tilde{l}(p)\sim\mathcal{U}(0,\tilde{L})$. The phase shift between the scatterer p and the reference point is $2\pi\tilde{l}(p)\cos(\theta_m)/\lambda$, $1\le m\le M$.

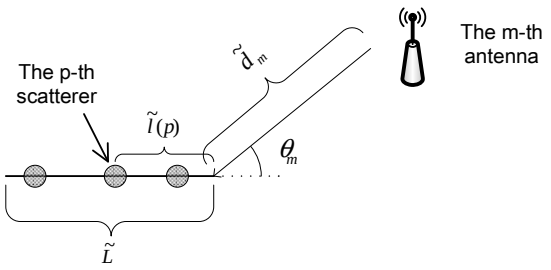

Figure 1 – Illustration of a line of scatterers.

Define a diagonal phase matrix

$$\mathbf{\Xi} \triangleq \text{diag}\{e^{-j2\pi\frac{\tilde{d}_1+r}{\lambda}}, ..., e^{-j2\pi\frac{\tilde{d}_M+r}{\lambda}}\}.$$

The p-th scattering path vector channel is now given by:

$$\mathbf{h}_p \triangleq \begin{bmatrix} e^{-j2\pi\frac{d_{p1}+r}{\lambda}} \\ \vdots \\ e^{-j2\pi\frac{d_{pM}+r}{\lambda}} \end{bmatrix} e^{j\varphi_p} = e^{j\varphi_p}\mathbf{\Xi}\begin{bmatrix} e^{-j2\pi\frac{\tilde{l}(p)\cos(\theta_1)}{\lambda}} \\ \vdots \\ e^{-j2\pi\frac{\tilde{l}(p)\cos(\theta_M)}{\lambda}} \end{bmatrix}$$

If we define $\tilde{\mathbf{h}}_p \triangleq [e^{-j2\pi\frac{\tilde{l}(p)\cos(\theta_1)}{\lambda}}, \cdots, e^{-j2\pi\frac{\tilde{l}(p)\cos(\theta_M)}{\lambda}}]^T$, we may see that $\tilde{\mathbf{h}}_p$ and $\mathbf{h}_p$ are unitarily equivalent, since $e^{j\varphi_p}\mathbf{\Xi}$ forms a unitary matrix. If we vary $\tilde{l}(p)$, we can obtain two linear spaces spanned by $\mathbf{h}_p$ and/or $\tilde{\mathbf{h}}_p$. The dimensions of the two spaces are equal. Therefore in the following we will find the dimension spanned by $\tilde{\mathbf{h}}_p$ instead of $\mathbf{h}_p$. Assume N is a large integer, we define $x_n \triangleq -1 + \frac{2(n-1)}{N}, n = 1, ..., N$, the set $\mathcal{X} \triangleq \{x_n\}$, as well as the vector

$$\boldsymbol{\mu}_p \triangleq \begin{bmatrix} e^{-j2\pi\frac{\tilde{l}(p)x_1}{\lambda}} \\ e^{-j2\pi\frac{\tilde{l}(p)x_2}{\lambda}} \\ \vdots \\ e^{-j2\pi\frac{\tilde{l}(p)x_N}{\lambda}} \end{bmatrix} = e^{j2\pi\tilde{l}(p)}\begin{bmatrix} e^{-j2\pi\frac{\tilde{l}(p)(0)}{\lambda}} \\ e^{-j2\pi\frac{\tilde{l}(p)(\frac{2}{N})}{\lambda}} \\ \vdots \\ e^{-j2\pi\frac{\tilde{l}(p)(\frac{2(N-1)}{N})}{\lambda}} \end{bmatrix}.$$

Since $\tilde{l}(p) \sim \mathcal{U}(0, \tilde{L})$, we reuse the result shown in (9):

$$\begin{aligned}
&\dim\{\text{span}\{\boldsymbol{\mu}_p, \tilde{l}(p) \in [0, \tilde{L}]\}\} \\
&= \frac{4\tilde{L}}{N\lambda}\frac{N}{2} + o\left(\frac{4\tilde{L}}{N\lambda}\frac{N}{2}\right) \\
&= \frac{2\tilde{L}}{\lambda} + o(\tilde{L}).
\end{aligned}$$

Note that the above result holds when N is arbitrarily large. We again observe that when $N \to \infty$, any $\cos(\theta_m)$ will fall into the set $\mathcal{X}$, which indicates

$$\dim\{\text{span}\{\tilde{\mathbf{h}}_p, \tilde{l}(p) \in [0, \tilde{L}]\}\} \leq \frac{2\tilde{L}}{\lambda} + o(\tilde{L}). \tag{10}$$

Recall that the covariance matrix $\mathbf{R} = \mathbb{E}\{\frac{1}{P}\sum_{p=1}^{P}\sum_{q=1}^{P}\mathbf{h}_p\mathbf{h}_q^H\}$. Because of the random and independent phases φ_p and φ_q,

$$\forall p \neq q, \mathbb{E}\{\mathbf{h}_p\mathbf{h}_q^H\} = \mathbf{0}.$$

$$\mathbf{R} = \mathbb{E}\{\frac{1}{P}\sum_{p=1}^{P}\mathbf{h}_p\mathbf{h}_p^H\} = \mathbb{E}\{\mathbf{h}_p\mathbf{h}_p^H\}.$$

We can see that the number of scatterers has no impact on the rank of channel covariance matrix. Hence according to (10), the rank of $\mathbf{R}$ is upper bounded by

$$\text{rank}(\mathbf{R}) \leq \frac{2\tilde{L}}{\lambda} + o(\tilde{L}).$$

Returning to the one-ring model, we can interpret the ring as the sum of lines, with the total length $2\pi r$. An extreme case is when all of the channels corresponding to different pieces of the ring span orthogonal spaces, i.e., the rank of the covariance matrix is the sum of the spatial dimensions corresponding to every pieces of the ring. This is the case when the covariance rank is maximized. Therefore the rank is upper bounded by:

$$\text{rank}(\mathbf{R}) \leq \frac{4\pi r}{\lambda} + o(r).$$

Thus Theorem 2 is proven. □

.6 Proof of Lemma 2:

Proof: Take an angle $\Phi \notin [\theta_i^{\min}, \theta_i^{\max}]$ and define

$$\mathbf{u} \triangleq \frac{\mathbf{a}(\Phi)}{\sqrt{M}}.$$

Then we have

$$\begin{aligned}
\mathbf{u}^H \mathbf{R}_i \mathbf{u} &= \frac{1}{M} \mathbf{a}(\Phi)^H \mathbf{R}_i \mathbf{a}(\Phi) \\
&= \frac{1}{M} \mathbf{a}^H(\Phi) \mathbb{E}\left\{ \mathbf{a}(\theta) \mathbf{a}^H(\theta) \right\} \mathbf{a}(\Phi) \\
&= \frac{1}{M} \mathbb{E}\left\{ \left| \mathbf{a}^H(\Phi) \mathbf{a}(\theta) \right|^2 \right\} \\
&= \frac{1}{M} \mathbb{E}\left\{ \left| \sum_{m=0}^{M-1} e^{2\pi j (m-1) \frac{D}{\lambda} (\cos(\Phi) - \cos(\theta))} \right|^2 \right\} \\
&= \int_{\theta_i^{\min}}^{\theta_i^{\max}} \frac{1}{M} \left| \sum_{m=0}^{M-1} e^{2\pi j (m-1) \frac{D}{\lambda} (\cos(\Phi) - \cos(\theta))} \right|^2 p_i(\theta) d\theta.
\end{aligned}$$

According to the well-known result on the sum of geometric series, we can easily obtain

$$\lim_{M \to \infty} \frac{1}{M} \left| \sum_{m=0}^{M-1} e^{2\pi j (m-1) \frac{D}{\lambda} (\cos(\Phi) - \cos(\theta))} \right|^2 = 0,$$

since $\Phi \neq \theta, \forall \theta \in [\theta_i^{\min}, \theta_i^{\max}]$. Thus

$$\lim_{M \to \infty} \mathbf{u}^H \mathbf{R}_i \mathbf{u} = 0,$$

which proves Lemma 2. □

.7 Proof of Proposition 5:

Proof: Denote the associated path loss as β. The covariance $\mathbf{R}$ is a Toeplitz matrix, with its mn-th entry given by

$$\begin{aligned}\mathbf{R}(m,n) &= \beta \int_0^{\pi} p(\theta) e^{j2\pi \frac{D}{\lambda}(n-m)\cos(\theta)} d\theta \qquad (11)\\ &= \beta \int_{-1}^{1} p\left(\arccos(x)\right) e^{j2\pi \frac{D}{\lambda}(n-m)x} \frac{1}{\sqrt{1-x^2}} dx \\ &= \frac{\beta\lambda}{2\pi D} \int_{-\frac{2\pi D}{\lambda}}^{\frac{2\pi D}{\lambda}} \frac{p\left(\arccos(\frac{\lambda x}{2\pi D})\right)}{\sqrt{1-\left(\frac{\lambda x}{2\pi D}\right)^2}} e^{j(n-m)x} dx, \\ &= \frac{1}{2\pi} \int_{-\frac{2\pi D}{\lambda}}^{\frac{2\pi D}{\lambda}} f(x) e^{j(n-m)x} dx, \qquad (12)\end{aligned}$$

where

$$f(x) \triangleq \frac{\beta\lambda}{D} \frac{p\left(\arccos(\frac{\lambda x}{2\pi D})\right)}{\sqrt{1-\left(\frac{\lambda x}{2\pi D}\right)^2}}. \qquad (13)$$

Since $0, \pi \notin \Phi$, or in other words, $p(0) = p(\pi) = 0$, and that $p(\theta) < \infty, \forall \theta \in \Phi$, it follows that $f(x)$ is uniformly bounded:

$$f(x) < +\infty, -\frac{2\pi D}{\lambda} \leq x \leq \frac{2\pi D}{\lambda}. \qquad (14)$$

Thus, the Toeplitz matrix $\mathbf{R}$ is related to the real integrable and uniformly bounded generating function $f(x)$, with its entries being Fourier coefficients of $f(x)$. We now resort to the known result on the spectrum of the $n \times n$ Toeplitz matrices $\mathbf{T}_n(f)$ defined by the generating function $f(x)$. Denote by ess inf and ess sup the essential minimum and the essential maximum of f, i.e., the infimum and the supremum of f up to within a set of measure zero. Let $m_f \triangleq \text{ess inf} f$ and $M_f \triangleq \text{ess sup} f$.

Theorem 8. *[95] If $\lambda_0^{(n)} \leq \lambda_1^{(n)} \leq \cdots \leq \lambda_{n-1}^{(n-1)}$ are the eigenvalues of $\mathbf{T}_n(f)$, then, the spectrum of $\mathbf{T}_n(f)$ is contained in (m_f, M_f); moreover $\lim_{n\to\infty} \lambda_0^{(n)} = m_f$ and $\lim_{n\to\infty} \lambda_{n-1}^{(n-1)} = M_f$.*

By invoking Theorem 8, we obtain that $\lim_{M\to\infty} \|\mathbf{R}\|_2 = M_f < \infty$. In addition, for any finite M, the inequality $\|\mathbf{R}\|_2 < \infty$ always holds true. This concludes the proof. □

.8 Proof of Lemma 4:

Proof: Since $\mathbf{R}_j$ and $\left(\sum_{l=1}^{L}\mathbf{R}_l+\sigma_n^2\mathbf{I}_M\right)^{-1}$ are both positive semi-definite (PSD) Hermitian matrices, we can directly apply the inequalities of [96] on the eigenvalues of the product of two PSD Hermitian matrices:

$$\|\mathbf{\Xi}_j\|_2 \leq \left\|\left(\sum_{l=1}^{L}\mathbf{R}_l+\sigma_n^2\mathbf{I}_M\right)^{-1}\right\|_2 \|\mathbf{R}_j\|_2 < \frac{\zeta}{\sigma_n^2}. \tag{15}$$

It is straightforward to show that

$$\left\|\mathbf{\Xi}_j\mathbf{\Xi}_j^H\right\|_2 = \|\mathbf{\Xi}_j\|_2^2 < \frac{\zeta^2}{\sigma_n^4}, \tag{16}$$

which indicates that the spectral norm of $\mathbf{\Xi}_j\mathbf{\Xi}_j^H$ is also uniformly bounded. This proves Lemma 4. □

.9 Proof of Lemma 6:

Proof: Using the spatial correlation model (6.6), we may write

$$\frac{1}{M}\underline{\mathbf{h}}_j^H\underline{\mathbf{h}}_l = \frac{1}{M}\mathbf{h}_{\mathrm{W}j}^H\mathbf{R}_j^{\frac{1}{2}}\mathbf{\Xi}_j^H\mathbf{\Xi}_j\mathbf{R}_l^{\frac{1}{2}}\mathbf{h}_{\mathrm{W}l}. \tag{17}$$

By an abuse of notation, we now use the operator $\lambda_1\{\cdot\}$ to represent the largest singular value of a matrix. Appealing to the singular value inequalities in [97], we can show that the maximum singular value of $\mathbf{R}_j^{\frac{1}{2}}\mathbf{\Xi}_j^H\mathbf{\Xi}_j\mathbf{R}_l^{\frac{1}{2}}$ yields:

$$\lambda_1\{\mathbf{R}_j^{\frac{1}{2}}\mathbf{\Xi}_j^H\mathbf{\Xi}_j\mathbf{R}_l^{\frac{1}{2}}\} \leq \lambda_1\{\mathbf{R}_j^{\frac{1}{2}}\}\lambda_1\{\mathbf{\Xi}_j^H\mathbf{\Xi}_j\mathbf{R}_l^{\frac{1}{2}}\} \tag{18}$$

$$< \zeta^{\frac{1}{2}}\lambda_1\{\mathbf{\Xi}_j^H\mathbf{\Xi}_j\}\lambda_1\{\mathbf{R}_l^{\frac{1}{2}}\} \tag{19}$$

$$< \frac{\zeta^3}{\sigma_n^4}, \tag{20}$$

which means the spectral radius of the complex matrix $\mathbf{R}_j^{\frac{1}{2}}\mathbf{\Xi}_j^H\mathbf{\Xi}_j\mathbf{R}_l^{\frac{1}{2}}$ is uniformly bounded for any M. Thus, according to Lemma 5, $\frac{1}{M}\underline{\mathbf{h}}_j^H\underline{\mathbf{h}}_l, \forall l \neq j$, converges almost surely to zero. Thus (5.28) holds true. In a similar way, we can prove (5.29). This concludes the proof of Lemma 6. □

.10 Proof of Lemma 7:

Proof: Define

$$\Gamma \triangleq \lim_{C\to\infty}\left(\frac{1}{C}\widetilde{\mathbf{W}}_j\widetilde{\mathbf{W}}_j^H\right) \tag{21}$$

$$= \underline{\mathbf{h}}_j\underline{\mathbf{h}}_j^H + \sum_{l\neq j}\underline{\mathbf{h}}_l\underline{\mathbf{h}}_l^H + \sigma_n^2\boldsymbol{\Xi}_j\boldsymbol{\Xi}_j^H. \tag{22}$$

In this proof, we first consider the noise free scenario and let

$$\Gamma_{\text{nf}} = \underline{\mathbf{h}}_j\underline{\mathbf{h}}_j^H + \sum_{l\neq j}\underline{\mathbf{h}}_l\underline{\mathbf{h}}_l^H, \tag{23}$$

where the subscript "nf" denotes noise free. We can then write

$$\begin{aligned}
&\lim_{M\to\infty}\left\|\frac{\Gamma_{\text{nf}}}{M}\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2} - \alpha_j\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2}\right\|_2^2 \qquad (24)\\
&= \lim_{M\to\infty}\left(\frac{\Gamma_{\text{nf}}}{M}\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2} - \alpha_j\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2}\right)^H\left(\frac{\Gamma_{\text{nf}}}{M}\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2} - \alpha_j\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2}\right)\\
&= \lim_{M\to\infty}\frac{1}{M^2}\left\|\underline{\mathbf{h}}_j\right\|_2^2 - \lim_{M\to\infty}\frac{2\alpha_j}{M}\left\|\underline{\mathbf{h}}_j\right\|_2^2 + \alpha_j^2\\
&= \alpha_j^2 - 2\alpha_j^2 + \alpha_j^2\\
&= 0,
\end{aligned}$$

Which proves that when $M\to\infty$, an eigenvalue of the random matrix Γ_{nf}/M converges to α_j, with its corresponding eigenvector converging to $\underline{\mathbf{h}}_j/\left\|\underline{\mathbf{h}}_j\right\|_2$ up to a random phase.

Then we consider the Hermitian matrix $\sigma_n^2\boldsymbol{\Xi}_j\boldsymbol{\Xi}_j^H$ as a perturbation on Γ_{nf}/M. Due to the Bauer-Fike Theorem [98] on the perturbation of eigenvalues of Hermitian matrices, together with Lemma 4, we have for $1\leq i\leq L$:

$$\lim_{M\to\infty}\left|\lambda_i\left\{\frac{\Gamma}{M}\right\} - \lambda_i\left\{\frac{\Gamma_{\text{nf}}}{M}\right\}\right| \tag{25}$$

$$\leq \lim_{M\to\infty}\frac{\sigma_n^2}{M}\left\|\boldsymbol{\Xi}_j\boldsymbol{\Xi}_j^H\right\|_2 \tag{26}$$

$$= 0. \tag{27}$$

The above result shows that the impact of the perturbation on the eigenvalues of Γ_{nf}/M vanishes as $M\to\infty$. In other words, α_j is again an asymptotic

eigenvalue of Γ/M. Now we verify that despite the perturbation, the eigenvector of Γ/M corresponding to the asymptotic eigenvalue α_j also converges to $\underline{\mathbf{h}}_j/\left\|\underline{\mathbf{h}}_j\right\|_2$ up to a random phase. To prove this, it is sufficient to show that

$$\begin{aligned}
&\lim_{M\to\infty}\left\|\frac{\Gamma}{M}\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2}-\alpha_j\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2}\right\|_2 \\
&\leq \lim_{M\to\infty}\left\|\frac{\Gamma_{\mathrm{nf}}}{M}\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2}-\alpha_j\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2}\right\|_2+\left\|\frac{\sigma_n^2\boldsymbol{\Xi}_j\boldsymbol{\Xi}_j^H}{M}\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2}\right\|_2 \\
&\overset{(a)}{=} 0,
\end{aligned} \tag{28}$$

where (a) is due to the definition of the spectral norm:

$$\lim_{M\to\infty}\left\|\frac{\sigma_n^2\boldsymbol{\Xi}_j\boldsymbol{\Xi}_j^H}{M}\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2}\right\|_2=0. \tag{29}$$

It follows that

$$\lim_{M,C\to\infty}\left\|\frac{\widetilde{\mathbf{W}}_j\widetilde{\mathbf{W}}_j^H}{MC}\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2}-\alpha_j\frac{\underline{\mathbf{h}}_j}{\left\|\underline{\mathbf{h}}_j\right\|_2}\right\|_2=0, \tag{30}$$

which concludes the proof of Lemma 7. □

.11 Proof of Lemma 8:

Proof: We can derive:

$$\begin{aligned}
&\lim_{M,C\to\infty}\left\|\frac{\boldsymbol{\Xi}_j'\underline{\mathbf{h}}_j}{\left\|\boldsymbol{\Xi}_j'\underline{\mathbf{h}}_j\right\|_2}-\frac{\boldsymbol{\Xi}_j'\widetilde{\mathbf{u}}_{j1}e^{j\phi}}{\left\|\boldsymbol{\Xi}_j'\widetilde{\mathbf{u}}_{j1}\right\|_2}\right\|_2^2 \\
&=\lim_{M,C\to\infty}\left(\frac{\boldsymbol{\Xi}_j'\underline{\mathbf{h}}_j}{\left\|\boldsymbol{\Xi}_j'\underline{\mathbf{h}}_j\right\|_2}-\frac{\boldsymbol{\Xi}_j'\widetilde{\mathbf{u}}_{j1}e^{j\phi}}{\left\|\boldsymbol{\Xi}_j'\widetilde{\mathbf{u}}_{j1}\right\|_2}\right)^H\cdot \\
&\qquad\left(\frac{\boldsymbol{\Xi}_j'\underline{\mathbf{h}}_j}{\left\|\boldsymbol{\Xi}_j'\underline{\mathbf{h}}_j\right\|_2}-\frac{\boldsymbol{\Xi}_j'\widetilde{\mathbf{u}}_{j1}e^{j\phi}}{\left\|\boldsymbol{\Xi}_j'\widetilde{\mathbf{u}}_{j1}\right\|_2}\right) \\
&=2-\lim_{M,C\to\infty}\left(\frac{\underline{\mathbf{h}}_j^H\boldsymbol{\Xi}_j'^H\boldsymbol{\Xi}_j'\widetilde{\mathbf{u}}_{j1}e^{j\phi}}{\left\|\boldsymbol{\Xi}_j'\underline{\mathbf{h}}_j\right\|_2\left\|\boldsymbol{\Xi}_j'\widetilde{\mathbf{u}}_{j1}\right\|_2}+\frac{e^{-j\phi}\widetilde{\mathbf{u}}_{j1}^H\boldsymbol{\Xi}_j'^H\boldsymbol{\Xi}_j'\underline{\mathbf{h}}_j}{\left\|\boldsymbol{\Xi}_j'\underline{\mathbf{h}}_j\right\|_2\left\|\boldsymbol{\Xi}_j'\widetilde{\mathbf{u}}_{j1}\right\|_2}\right)
\end{aligned} \tag{31}$$

We treat the following quantity separately:

$$\lim_{M,C\to\infty} \frac{\underline{\mathbf{h}}_j^H \underline{\boldsymbol{\Xi}}_j'^H \boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1} e^{j\phi}}{\left\|\boldsymbol{\Xi}_j' \underline{\mathbf{h}}_j\right\|_2 \left\|\boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1}\right\|_2} \tag{32}$$

$$= \lim_{M,C\to\infty} \frac{\underline{\mathbf{h}}_j^H \underline{\boldsymbol{\Xi}}_j'^H \boldsymbol{\Xi}_j' \left(\frac{\underline{\mathbf{h}}_j}{\|\underline{\mathbf{h}}_j\|_2} + \widetilde{\mathbf{u}}_{j1} e^{j\phi} - \frac{\underline{\mathbf{h}}_j}{\|\underline{\mathbf{h}}_j\|_2} \right)}{\left\|\boldsymbol{\Xi}_j' \underline{\mathbf{h}}_j\right\|_2 \left\|\boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1}\right\|_2}$$

$$= \lim_{M,C\to\infty} \frac{\left\|\boldsymbol{\Xi}_j' \frac{\underline{\mathbf{h}}_j}{\|\underline{\mathbf{h}}_j\|_2}\right\|_2}{\left\|\boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1}\right\|_2}$$

$$= \lim_{M,C\to\infty} \frac{\left\|\boldsymbol{\Xi}_j' \left(\frac{\underline{\mathbf{h}}_j}{\|\underline{\mathbf{h}}_j\|_2} - \widetilde{\mathbf{u}}_{j1} e^{j\phi} + \widetilde{\mathbf{u}}_{j1} e^{j\phi} \right)\right\|_2}{\left\|\boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1}\right\|_2}$$

$$\leq \lim_{M,C\to\infty} \frac{\left\|\boldsymbol{\Xi}_j' (\frac{\underline{\mathbf{h}}_j}{\|\underline{\mathbf{h}}_j\|_2} - \widetilde{\mathbf{u}}_{j1} e^{j\phi})\right\|_2}{\left\|\boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1}\right\|_2} + \lim_{M,C\to\infty} \frac{\left\|\boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1} e^{j\phi}\right\|_2}{\left\|\boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1}\right\|_2}$$

$$= 1 \tag{33}$$

In a similar way, we can prove that

$$\lim_{M,C\to\infty} \frac{\left\|\boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1}\right\|_2}{\left\|\boldsymbol{\Xi}_j' \frac{\underline{\mathbf{h}}_j}{\|\underline{\mathbf{h}}_j\|_2}\right\|_2} \leq 1. \tag{34}$$

Combining (33) and (34), we obtain:

$$\lim_{M,C\to\infty} \frac{\underline{\mathbf{h}}_j^H \underline{\boldsymbol{\Xi}}_j'^H \boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1} e^{j\phi}}{\left\|\boldsymbol{\Xi}_j' \underline{\mathbf{h}}_j\right\|_2 \left\|\boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1}\right\|_2} = 1. \tag{35}$$

With analogous derivation, we can prove

$$\lim_{M,C\to\infty} \frac{e^{-j\phi} \widetilde{\mathbf{u}}_{j1}^H \underline{\boldsymbol{\Xi}}_j'^H \boldsymbol{\Xi}_j' \underline{\mathbf{h}}_j}{\left\|\boldsymbol{\Xi}_j' \underline{\mathbf{h}}_j\right\|_2 \left\|\boldsymbol{\Xi}_j' \widetilde{\mathbf{u}}_{j1}\right\|_2} = 1. \tag{36}$$

Applying (35) and (36) to (31) gives:

$$\lim_{M,C\to\infty}\left\|\frac{\boldsymbol{\Xi}_j'\underline{\mathbf{h}}_j}{\left\|\boldsymbol{\Xi}_j'\underline{\mathbf{h}}_j\right\|_2}-\frac{\boldsymbol{\Xi}_j'\widetilde{\mathbf{u}}_{j1}e^{j\phi}}{\left\|\boldsymbol{\Xi}_j'\widetilde{\mathbf{u}}_{j1}\right\|_2}\right\|_2^2=0. \tag{37}$$

The following equality holds:

$$\boldsymbol{\Xi}_j'\underline{\mathbf{h}}_j=\boldsymbol{\Xi}_j'\boldsymbol{\Xi}_j\mathbf{h}_j=\mathbf{R}_j^{\dagger}\mathbf{R}_j\mathbf{h}_j=\mathbf{h}_j, \tag{38}$$

proving that

$$\lim_{M,C\to\infty}\left\|\frac{\mathbf{h}_j}{\|\mathbf{h}_j\|_2}-\bar{\mathbf{u}}_{j1}e^{j\phi}\right\|_2=0, \tag{39}$$

which completes the proof of Lemma 8. □

.12 Proof of Theorem 4:

Proof: From (5.33) we readily obtain

$$\lim_{M,C\to\infty}\frac{\mathbf{h}_j^H\bar{\mathbf{u}}_{j1}}{\|\mathbf{h}_j\|_2}=1. \tag{40}$$

Recall from the uplink training (5.3), we have

$$\widehat{\mathbf{h}}_j^{\mathrm{CA}}=\frac{1}{\tau}\bar{\mathbf{u}}_{j1}\bar{\mathbf{u}}_{j1}^H\left(\mathbf{h}_j\mathbf{s}^T+\sum_{l\neq j}\mathbf{h}_l\mathbf{s}^T+\mathbf{N}\right)\mathbf{s}^*, \tag{41}$$

and hence

$$\lim_{M,C\to\infty} \frac{\left\|\widehat{\mathbf{h}}_j^{\mathrm{CA}} - \mathbf{h}_j\right\|_2^2}{\|\mathbf{h}_j\|_2^2} \tag{42}$$

$$= \lim_{M,C\to\infty} \frac{(\widehat{\mathbf{h}}_j^{\mathrm{CA}} - \mathbf{h}_j)^H(\widehat{\mathbf{h}}_j^{\mathrm{CA}} - \mathbf{h}_j)}{\|\mathbf{h}_j\|_2^2}$$

$$= \lim_{M,C\to\infty} \frac{1}{\|\mathbf{h}_j\|_2^2} \Bigg(\sum_{l\neq j} \mathbf{h}_l \bar{\mathbf{u}}_{j1} \bar{\mathbf{u}}_{j1}^H \sum_{l\neq j} \mathbf{h}_l + \sum_{l\neq j} \mathbf{h}_l \bar{\mathbf{u}}_{j1} \bar{\mathbf{u}}_{j1}^H \mathbf{N} \frac{\mathbf{s}^*}{\tau}$$
$$+ \frac{\mathbf{s}^T}{\tau} \mathbf{N}^H \bar{\mathbf{u}}_{j1} \bar{\mathbf{u}}_{j1}^H \sum_{l\neq j} \mathbf{h}_l + \frac{\mathbf{s}^T}{\tau} \mathbf{N}^H \bar{\mathbf{u}}_{j1} \bar{\mathbf{u}}_{j1}^H \mathbf{N} \frac{\mathbf{s}^*}{\tau}$$
$$- \mathbf{h}_j^H \bar{\mathbf{u}}_{j1} \bar{\mathbf{u}}_{j1}^H \mathbf{h}_j + \mathbf{h}_j^H \mathbf{h}_j \Bigg)$$

$$= \lim_{M,C\to\infty} \frac{1}{\|\mathbf{h}_j\|_2^2} \left(\mathbf{h}_j^H \mathbf{h}_j - \mathbf{h}_j^H \bar{\mathbf{u}}_{j1} \bar{\mathbf{u}}_{j1}^H \mathbf{h}_j \right). \tag{43}$$

Equation (40) ensures that

$$\lim_{M,C\to\infty} \frac{1}{\|\mathbf{h}_j\|_2^2} \mathbf{h}_j^H \bar{\mathbf{u}}_{j1} \bar{\mathbf{u}}_{j1}^H \mathbf{h}_j = \frac{1}{\|\mathbf{h}_j\|_2^2} \mathbf{h}_j^H \mathbf{h}_j = 1, \tag{44}$$

which concludes the proof. □

.13 Proof of Theorem 5:

Proof: This proof follows similar steps towards Theorem 4. Thus we give a sketch of the proof only. Define

$$\Gamma \triangleq \lim_{C\to\infty} \left(\frac{1}{C} \widetilde{\mathbf{W}}_j \widetilde{\mathbf{W}}_j^H \right) \tag{45}$$

$$= \underline{\mathbf{h}}_j^{(j)} \underline{\mathbf{h}}_j^{(j)H} + \sum_{l\neq j} \underline{\mathbf{h}}_l^{(j)} \underline{\mathbf{h}}_l^{(j)H} + \sigma_n^2 \mathbf{\Xi}_j \mathbf{\Xi}_j^H, \tag{46}$$

where $\underline{\mathbf{h}}_l^{(j)} \triangleq \mathbf{\Xi}_j \mathbf{h}_l^{(j)}, l = 1, \ldots, L$. Due to the asymptotic orthogonality between steering vectors in disjoint angular support, i.e., Lemma 3 in [36], we can easily show that in large antenna limit, $\mathbf{h}_{\mathrm{lo}}^{(j)}$ falls into the null space of $\mathbf{R}_j^{(j)}$. Thus

$$\lim_{M\to\infty} \frac{1}{M} \underline{\mathbf{h}}_l^{(j)} \underline{\mathbf{h}}_l^{(j)H} = \lim_{M\to\infty} \frac{1}{M} \underline{\mathbf{h}}_{li}^{(j)} \underline{\mathbf{h}}_{li}^{(j)H}. \tag{47}$$

Then we have

$$\lim_{M\to\infty}\frac{\Gamma}{M}=\frac{1}{M}\left(\underline{\mathbf{h}}_j^{(j)}\underline{\mathbf{h}}_j^{(j)H}+\sum_{l\neq j}\underline{\mathbf{h}}_{li}^{(j)}\underline{\mathbf{h}}_{li}^{(j)H}+\sigma_n^2\boldsymbol{\Xi}_j\boldsymbol{\Xi}_j^H\right).$$

Under condition C1, it is easy to show that

$$\lim_{M\to\infty}\left\|\frac{\Gamma}{M}\frac{\underline{\mathbf{h}}_j^{(j)}}{\left\|\underline{\mathbf{h}}_j^{(j)}\right\|_2}-\frac{\underline{\mathbf{h}}_j^{(j)H}\underline{\mathbf{h}}_j^{(j)}}{M}\frac{\underline{\mathbf{h}}_j^{(j)}}{\left\|\underline{\mathbf{h}}_j^{(j)}\right\|_2}\right\|_2=0. \tag{48}$$

Given the following condition

$$\forall l\neq j,\left\|\boldsymbol{\Xi}_j\mathbf{h}_{li}^{(j)}\right\|_2<\left\|\boldsymbol{\Xi}_j\mathbf{h}_j^{(j)}\right\|_2, \tag{49}$$

it is clear that the dominant eigenvector of Γ/M converges to $\underline{\mathbf{h}}_j^{(j)}/\left\|\underline{\mathbf{h}}_j^{(j)}\right\|_2$ (up to a random phase), with its corresponding eigenvalue converging to $\underline{\mathbf{h}}_j^{(j)H}\underline{\mathbf{h}}_j^{(j)}/M$. Then, using the same technique in the proof of Lemma 8, we obtain:

$$\lim_{M,C\to\infty}\frac{\mathbf{h}_j^{(j)H}\bar{\mathbf{u}}_{j1}}{\left\|\mathbf{h}_j^{(j)}\right\|_2}=1. \tag{50}$$

Finally, we readily obtain (5.39) by analogous derivations in Appendix .12. □

Bibliography

[1] A. F. Molisch, *Wireless communications*. Wiley, 2010.

[2] W. C. Jakes, *Mobile microwave communication*. Wiley, 1974.

[3] D. S. Shiu, G. J. Foschini, M. J. Gans, and J. M. Kahn, "Fading correlation and its effect on the capacity of multielement antenna systems," *IEEE Trans. Commun.*, vol. 48, no. 3, pp. 502–513, Mar. 2000.

[4] H. Yin, L. Cottatellucci, D. Gesbert, R. Müller, and G. He, "Robust pilot decontamination based on joint angle and power domain discrimination," *Submitted to IEEE Trans. Signal Process.*, 2015. [Online]. Available: http://arxiv.org/abs/1509.06024

[5] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, 2006.

[6] H. Yin, L. Cottatellucci, and D. Gesbert, "Enabling massive MIMO systems in the FDD mode thanks to D2D communications," in *Proc. of Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, Nov 2014, pp. 656–660.

[7] A. Gorokhov, "Antenna selection algorithms for MEA transmission systems," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, Orlando, FL, USA, May 2002, pp. III–2857–III–2860.

[8] A. Gorokhov, D. A. Gore, and A. J. Paulraj, "Receive antenna selection for MIMO spatial multiplexing: theory and algorithms," *IEEE Trans Signal Process.*, vol. 51, no. 11, pp. 2796–2807, 2003.

[9] A. Lozano, R. W. Heath, and J. G. Andrews, "Fundamental limits of cooperation," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5213–5226, 2013.

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

[11] 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 Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2013.

[12] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.

[13] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[14] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination problem in multi-cell TDD systems," in *Proc. IEEE International Symposium on Information Theory (ISIT'09)*, Seoul, Korea, Jun. 2009, pp. 2184–2188.

[15] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO: How many antennas do we need?" in *Proc. 2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Sep. 2011, pp. 545 –550.

[16] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.

[17] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing – the large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct 2013.

[18] E. Björnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: 10 myths and one grand question," *To appear in IEEE Communications Magazine*, 2015. [Online]. Available: http://arxiv.org/abs/1503.06854

[19] R. Rogalin, O. Y. Bursalioglu, H. Papadopoulos, G. Caire, A. F. Molisch, A. Michaloliakos, V. Balan, and K. Psounis, "Scalable synchronization and reciprocity calibration for distributed multiuser MIMO," *IEEE Trans. Wireless Commun*, vol. 13, no. 4, pp. 1815–1831, 2014.

[20] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: energy efficiency, estimation, and capacity limits," *Submitted to IEEE Trans. Inf. Theory*, 2013. [Online]. Available: http://arxiv.org/abs/1307.2584

[21] A. Pitarokoilis, S. K. Mohammed, and E. G. Larsson, "Uplink performance of time-reversal MRC in massive MIMO systems subject to

phase noise," *IEEE Trans. Wireless Commun*, vol. 14, no. 2, pp. 711–723, 2015.

[22] H. Q. Ngo, T. L. Marzetta, and E. G. Larsson, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2011)*, Prague, Czech Republic, May 2011, pp. 3464–3467.

[23] B. Gopalakrishnan and N. Jindal, "An analysis of pilot contamination on multi-user MIMO cellular systems with many antennas," in *2011 IEEE 12th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, Jun. 2011, pp. 381–385.

[24] K. Appaiah, A. Ashikhmin, and T. L. Marzetta, "Pilot contamination reduction in multi-user TDD systems," in *2010 IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, May 2010, pp. 1–5.

[25] F. Fernandes, A. Ashikhmin, and T. L. Marzetta, "Inter-cell interference in noncooperative TDD large scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 192–201, 2013.

[26] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *IEEE International Symposium on Information Theory Proceedings (ISIT2012)*. Cambridge, MA, USA: IEEE, 2012, pp. 1137–1141.

[27] H. Q. Ngo and E. G. Larsson, "EVD-based channel estimation in multicell multiuser MIMO systems with very large antenna arrays," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2012, pp. 3249–3252.

[28] R. R. Müller, L. Cottatellucci, and M. Vehkaperä, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[29] L. Cottatellucci, R. R. Müller, and M. Vehkapera, "Analysis of pilot decontamination based on power control," in *Proc. of IEEE Vehicular Technology Conference (VTC)*, Jun 2013, pp. 1–5.

[30] D. Neumann, M. Joham, and W. Utschick, "Suppression of pilot-contamination in massive MIMO systems," in *IEEE International Workshop on Signal Processing Advances in Wireless Communications (SPAWC2014)*. IEEE, Jun. 2014, pp. 11–15.

[31] ——, "Channel estimation in massive MIMO systems," *arXiv preprint*, 2015. [Online]. Available: http://arxiv.org/abs/1503.08691

[32] J. H. Sørensen and E. D. Carvalho, "Pilot decontamination through pilot sequence hopping in massive MIMO systems," in *Proc. of IEEE Global Telecommunications Conference (Globecom)*, Dec. 2014, pp. 1–6.

[33] I. Atzeni, J. Arnau, and M. Debbah, "Fractional pilot reuse in massive MIMO systems." [Online]. Available: http://arxiv.org/abs/1503.07321

[34] D. Hu, L. He, and X. Wang, "Semi-blind pilot decontamination for massive MIMO systems," *IEEE Trans. Wireless Commun.*, no. 99, Sept. 2015.

[35] V. Zarzoso and P. Comon, "Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size," *IEEE Transactions on Neural Networks*, vol. 21, no. 2, pp. 248–261, 2010.

[36] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, Feb. 2013.

[37] D. Neumann, A. Gruendinger, M. Joham, and W. Utschick, "Pilot coordination for large-scale multi-cell TDD systems," in *18th International ITG Workshop on Smart Antennas (WSA)*, Mar. 2014.

[38] J. Chen and V. K. N. Lau, "Two-tier precoding for FDD multi-cell massive MIMO time-varying interference networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1230–1238, 2014.

[39] J. Choi, Z. Chance, D. J. Love, and U. Madhow, "Noncoherent trellis coded quantization: A practical limited feedback technique for massive MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5016–5029, 2013.

[40] B. Lee, J. Choi, J.-Y. Seol, D. J. Love, and B. Shim, "Antenna grouping based feedback reduction for FDD-based massive MIMO systems," in *IEEE International Conference on Communications (ICC)*, Sydney, Australia, Jun. 2014, pp. 4477–4482.

[41] B. Lee, J. Choi, J.-y. Seol, D. J. Love, and B. Shim, "Antenna grouping based feedback compression for FDD-based massive MIMO systems," *To appear in IEEE Trans. Commun.*, 2015. [Online]. Available: http://arxiv.org/abs/1408.6009

[42] S. T. Qaseem and T. Y. Al-Naffouri, "Compressive sensing for reducing feedback in MIMO broadcast channels," in *IEEE International Conference on Communications (ICC)*, Cape Town, South Africa, 2010, pp. 1–5.

[43] C. Qi and L. Wu, "A hybrid compressed sensing algorithm for sparse channel estimation in MIMO OFDM systems," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3488–3491.

[44] P.-H. Kuo, H. Kung, and P.-A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Paris, France, Apr. 2012, pp. 492–497.

[45] S. L. H. Nguyen and A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser mimo systems," in *IEEE Wireless Communications and Networking Conference (WCNC)*, Shanghai, China, Apr. 2013, pp. 2890–2895.

[46] Z. Gao, L. Dai, and Z. Wang, "Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems," *Electronics Letters*, vol. 50, no. 12, pp. 896–898, 2014.

[47] M. E. Eltayeb, T. Al-Naffouri, and H. Bahrami, "Compressive sensing for feedback reduction in MIMO broadcast channels," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3209–3222, Sept. 2014.

[48] M. Masood, L. H. Afify, and T. Y. Al-Naffouri, "Efficient coordinated recovery of sparse channels in massive MIMO," *IEEE Trans Signal Process.*, vol. 63, no. 1, pp. 104–118, 2015.

[49] W. Shen, B. Wang, J. Feng, C. Gao, and J. Ma, "Differential CSIT acquisition based on compressive sensing for FDD massive MIMO systems," in *IEEE Vehicular Technology Conference (VTC Spring)*, Glasgow, U.K., May 2015, pp. 1–4.

[50] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[51] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[52] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on pure and applied mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[53] E. J. Candè and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[54] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[55] Y.-D. Lin and Y.-C. Hsu, "Multihop cellular: A new architecture for wireless communications," in *Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE INFOCOM 2000)*, vol. 3, 2000, pp. 1273–1282.

[56] B. Kaufman and B. Aazhang, "Cellular networks with an overlaid device to device network," in *2008 Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, USA, Oct 2008, pp. 1537–1541.

[57] K. Doppler, M. Rinne, C. Wijting, C. B. Ribeiro, and K. Hugl, "Device-to-device communication as an underlay to LTE-advanced networks," *IEEE Communications Magazine*, vol. 47, no. 12, pp. 42–49, 2009.

[58] K. Doppler, M. P. Rinne, P. Jänis, C. Ribeiro, and K. Hugl, "Device-to-Device communications; functional prospects for LTE-advanced networks," in *IEEE International Conference on Communications Workshops (ICC Workshops 2009)*, Dresden, Germany, Jun. 2009, pp. 1–6.

[59] T. Peng, Q. Lu, H. Wang, S. Xu, and W. Wang, "Interference avoidance mechanisms in the hybrid cellular and device-to-device systems," in *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Jun. 2009, pp. 617–621.

[60] J. Du, W. Zhu, J. Xu, Z. Li, and H. Wang, "A compressed HARQ feedback for device-to-device multicast communications," in *IEEE Vehicular Technology Conference (VTC Fall)*, Quebec City, Canada, Sept. 2012, pp. 1–5.

[61] B. Zhou, H. Hu, S.-Q. Huang, and H.-H. Chen, "Intracluster device-to-device relay algorithm with optimal resource utilization," *IEEE Trans. Veh. Technol.*, vol. 62, no. 5, pp. 2315–2326, 2013.

[62] N. Golrezaei, A. G. Dimakis, and A. F. Molisch, "Device-to-device collaboration through distributed storage," in *IEEE Global Communications Conference (GLOBECOM)*, Anaheim, CA, USA, Dec. 2012, pp. 2397–2402.

[63] J. C. Li, M. Lei, and F. Gao, "Device-to-device (d2d) communication in mu-mimo cellular networks," in *IEEE Global Communications Conference (GLOBECOM)*, Anaheim, CA, USA, Dec. 2012, pp. 3583–3587.

[64] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis, "Base-station assisted device-to-device communications for high-throughput

wireless video networks," *IEEE Trans. Wireless Commun.*, vol. 13, no. 7, pp. 3665–3676, 2014.

[65] L. Lei, Z. Zhong, C. Lin, and X. Shen, "Operator controlled device-to-device communications in LTE-advanced networks," *IEEE Wireless Communications*, vol. 19, no. 3, p. 96, 2012.

[66] X. Lin, R. Heath, and J. Andrews.

[67] S. Shalmashi, E. Björnson, M. Kountouris, K. W. Sung, and M. Debbah, "Energy efficiency and sum rate when massive MIMO meets device-to-device communication," in *IEEE International Conference on Communications (ICC)*, London, UK, Jun. 2015.

[68] ——, "Energy efficiency and sum rate tradeoffs for massive mimo systems with underlaid device-to-device communications," *Submitted*, 2015. [Online]. Available: http://arxiv.org/abs/1506.00598

[69] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Pers. Commun.*, vol. 5, no. 1, pp. 10 –22, Feb. 1998.

[70] S. Wagner, R. Couillet, M. Debbah, and D. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, 2012.

[71] R. Knopp and P. A. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proc. IEEE International Conference on Communications*, vol. 1, Seattle, WA, USA, Jun. 1995, pp. 331–335.

[72] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1277–1294, Jun. 2002.

[73] J. Jang and K. B. Lee, "Transmit power adaptation for multiuser OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 2, pp. 171–178, Feb. 2003.

[74] D. Gesbert, M. Kountouris, R. Heath, C. Chae, and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.

[75] A. J. Paulraj, R. Nabar, and D. Gore, *Introduction to space-time wireless communications*. Cambridge University Press, 2003.

[76] A. Scherb and K. D. Kammeyer, "Bayesian channel estimation for doubly correlated MIMO systems," in *Proc. IEEE Workshop Smart Antennas*, 2007.

[77] E. Bjornson and B. Ottersten, "A framework for training-based estimation in arbitrarily correlated Rician MIMO channels with Rician disturbance," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1807–1820, Mar. 2010.

[78] R. M. Gray and L. D. Davisson, *An introduction to statistical signal processing*. Cambridge University Press, 2004.

[79] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[80] J. A. Fill and D. E. Fishkind, "The Moore–Penrose generalized inverse for sums of matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 2, pp. 629–635, 2000.

[81] M. A. Rakha, "On the Moore–Penrose generalized inverse matrix," *Applied Mathematics and Computation*, vol. 158, no. 1, pp. 185–200, 2004.

[82] H. Yin, D. Gesbert, and L. Cottatellucci, "Dealing with interference in distributed large-scale MIMO systems: A statistical approach," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 942–953, Oct. 2014.

[83] A. Adhikary and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming and user grouping," *arXiv preprint*, 2013. [Online]. Available: http://arxiv.org/abs/1305.7252

[84] J. Evans and D. N. C. Tse, "Large system performance of linear multiuser receivers in multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 6, pp. 2059–2078, 2000.

[85] H. Yin, L. Cottatellucci, D. Gesbert, R. Müller, and G. He, "Pilot decontamination using combined angular and amplitude based projections in massive MIMO systems," in *IEEE 16th Workshop on Signal Processing Advances in Wireless Communications, SPAWC 2015*, Stockholm, Sweden, Jun. 2015, pp. 216–220.

[86] T. S. Rappaport, *Wireless communications: principles and practice*. Prentice Hall PTR New Jersey, 1996, vol. 2.

[87] T. L. Marzetta, "How much training is required for multiuser MIMO?" in *Proc. of the Fortieth Asilomar Conf. on Signals, Systems, & Computers*, Pacific Grove, CA, USA, Oct. 2006.

[88] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[89] J. Nam, J. Ahn, and G. Caire, "Joint spatial division and multiplexing: Realizing massive MIMO gains with limited channel state information," in *46th Annual Conference on Information Sciences and Systems, (CISS 2012)*, Princeton University, NJ, USA, Mar. 2012.

[90] X. Rao, V. K. Lau, and X. Kong, "CSIT estimation and feedback for FDD multi-user massive MIMO systems," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 3157–3161.

[91] X. Rao and V. K. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans Signal Process.*, vol. 62, no. 12, pp. 3261–3271, 2014.

[92] P. Ding, D. J. Love, and M. D. Zoltowski, "Multiple antenna broadcast channels with shape feedback and limited feedback," *IEEE Trans Signal Process.*, vol. 55, no. 7, pp. 3417–3428, 2007.

[93] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, 2005.

[94] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[95] U. Grenander and G. Szegö, *Toeplitz forms and their applications*. 2nd ed. Chelsea, New York, 1984.

[96] A. W. Marshall, I. Olkin, and B. Arnold, *Inequalities: theory of majorization and its applications*. Springer Science & Business Media, 2010.

[97] R. A. Horn and C. R. Johnson, *Topics in matrix analysis*. Cambridge University Press, 1991.

[98] F. L. Bauer and C. T. Fike, "Norms and exclusion theorems," *Numerische Mathematik*, vol. 2, no. 1, pp. 137–141, 1960.