ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2016-ENST-0022

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Electronique et Communications »

présentée et soutenue publiquement par

Qianrui LI

le 14 mars 2016

Transmission coopérative et traitement du signal distribué avec feedback et backhaul limité

Directeur de thèse : **David GESBERT**
Co-encadrement de la thèse : **Nicolas GRESSET, Paul DE KERRET**

Jury

M. Angel LOZANO, Professeur, Universitat Pompeu Fabra	Rapporteurs
Mme Mari KOBAYASHI, Maître de conference HDR, Supélec	Rapporteurs
M. Philippe CIBLAT, Professeur, Télécom ParisTech	Président du jury
M. Luca SANGUINETTI, Maître de conference, University of Pisa	Examinateurs
M. Nicolas GRESSET, Dr., MERCE	Examinateurs
M. David GESBERT, Professeur, EURECOM	Directeur de thèse
M. Paul DE KERRET, Dr., EURECOM	Invite

T H È S E

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

Qianrui Li

Distributed Transmitter Cooperation and Signal Processing with Limited Feedback and Backhaul

Defense scheduled the 14th of March 2016 before a committee composed of:

President of the Jury		
Professor	Philippe Ciblat	Telecom ParisTech
Reviewers		
Professor	Angel Lozano	Universitat Pompeu Fabra
Maître de conference HDR	Mari Kobayashi	Supélec
Examiners		
Maître de conference	Luca Sanguinetti	University of Pisa
Dr.	Nicolas Gresset	MERCE
Thesis Supervisor		
Professor	David Gesbert	EURECOM
Guest		
Dr.	Paul de Kerret	EURECOM

THÈSE

présentée pour l'obtention du grade de

Docteur de TELECOM ParisTech

Spécialité: Communication et Electronique

Qianrui Li

Transmission coopérative et traitement du signal distribué avec feedback et backhaul limité

Soutenance de thèse prévue le 14 Mars 2016 devant le jury composé de :

Président du jury		
Professeur	Philippe Ciblat	Telecom ParisTech
Rapporteurs		
Professeur	Angel Lozano	Universitat Pompeu Fabra
Maître de conference HDR	Mari Kobayashi	Supélec
Examinateurs		
Maître de conference	Luca Sanguinetti	University of Pisa
Dr.	Nicolas Gresset	MERCE
Directeur de thèse		
Professeur	David Gesbert	EURECOM
Invite		
Dr.	Paul de Kerret	EURECOM

Abstract

Transmitter cooperation is considered a promising tool for dealing with interference in wireless networks with an aggressive reuse policy of spectral resources. Cooperation is meant here as the joint optimization of certain transmission parameters, where such optimization can be carried out over several independent domains by using different techniques such as power control, user selection in time/frequency, antenna selection or beam/precoder design. Although transmitter cooperation comes in many flavors, a recurrent assumption behind proposed methods lies in the need for cooperating devices to (i) acquire, share information pertaining to the propagation channel toward the multiple receivers and (ii) perform cooperation based on the disseminated information in the previous step. This holds true for instance for coordinated beamforming methods and, to an even greater extent, for network-MIMO (Joint Processing coordinated multi-point (JP CoMP) in the long term evolution (LTE) terminology). As feedback and exchange of channel state information (CSI) come at a price in terms of signaling overhead, there arise two important questions: (i) What information should be fed back or exchanged such that the CSI acquired at each transmitter is most informative to perform cooperation? (ii) Which techniques can reap the benefits of cooperation while living with an imperfect channel representation that varies from transmitter to transmitter?

In this thesis, we address both aforementioned questions. We consider first each transmitter acquires an initial imperfect CSI based on limited receivers feedback. The transmitters can then exchange their initial CSI through the limited backhaul, which leads to an imperfect and non-identical final CSI at each transmitter. We optimize first the use of limited backhaul such that the channel estimation at each transmitter is the most accurate. We also consider the problem of optimally allocating the backhaul resource so as to get more accurate CSI estimates or balancing the accuracy and the consistency of the CSI estimate at each transmitter.

For the design of efficient cooperation techniques that copes with the

imperfect and non-identical CSI configuration at each transmitter, we investigate specifically a regularized zero forcing (RZF) precoder design in large system scenario. With random matrix theory tools, we find the RZF precoder which maximizes the system ergodic sum rate and is robust to the effect of limited feedback and backhaul. Finally, interesting and challenging research directions and open problems are discussed.

Abrégé

La coopération des émetteurs est considérée comme une approche prometteuse pour limiter les interférences dans les réseaux sans fil ayant une réutilisation des ressources spectrales très agressive. La coopération des émetteurs permet l'optimisation conjointe de certains paramètres de transmission. Cette optimisation peut concerner plusieurs domaines indépendants en utilisant des techniques différentes telles que le contrôle de la puissance, la sélection des utilisateurs en temps/fréquence, la sélection des antennes, ou la conception des matrices de précodage. Bien que la coopération des émetteurs existe sous différentes formes, une hypothèse commune est le besoin pour les émetteurs entrant en coopération (i) d'acquérir et de partager des informations concernant le canal de propagation ainsi que (ii) d'effectuer une coopération fondée sur les informations diffusées à l'étape précédente. La conception coordonnée des matrices de précodage et, d'une manière encore plus marquée, la transmission conjointe à différents émetteurs (JP CoMP) sont des exemples importants de méthodes de coopérations présentant ces propriétés. L'acquisition et l'échange de l'information de canal (CSI) étant strictement limités, il se pose deux questions importantes: (i) Quelle information doit être renvoyée ou échangée de manière à permettre la coopération la plus efficace? (ii) Quelles méthodes permettent de réaliser les gains de la coopération dans ce contexte de partage limité et imparfait d'information?

Dans cette thèse, nous abordons les deux questions précédentes. Dans un premier temps, nous considérons que chaque émetteur acquiert une estimée de canal imparfaite. Ces émetteurs peuvent alors échanger leur information de canal initiale par un canal ayant une capacité limité, conduisant à une estimée finale à chaque émetteur. Nous optimisons d'abord l'utilisation des liens limités afin que l'estimation de canal à chaque émetteur soit la plus précise possible. Nous considérons également le problème de la répartition optimale des ressources de partage afin d'obtenir les estimations de la meilleure qualité possible, ou bien d'atteindre un compromis entre la précision et la cohérence des estimées aux émetteurs.

Dans un second temps, nous étudions la conception de techniques de coopération efficaces dans une configuration d'information de canal imparfaitement partagée entre les émetteurs. En particulier, nous nous intéressons à la méthode de précodage par inversion régularisée de canal (RZF). La considération d'un grand nombre d'antennes aux émetteurs permet alors l'utilisation d'outils de la théorie des matrices aléatoires. Ces outils permettent d'obtenir le précodeur RZF qui maximise la capacité ergodique du système lorsque confronté à des informations de canal imparfaitement partagées par les émetteurs. Enfin, les futures directions de recherche découlant de ces travaux sont présentées et discutées.

Acknowledgements

It's my great honor and pleasure to have spent 3 years pursuing my Ph.D. degree in Eurecom under an industrial collaboration program in Mitsubishi Electric R&D Centre Europe (MERCE). During this wonderful time, I have learnt from and worked with a lot of talented minds. Foremost, I would like to thank my thesis advisor Professor David Gesbert for his guidance and support. I have benefit tremendously from his unique blend of wisdom, vision and insights in this field. Apart from his excellence in academic research, his impressive personal charisma truly creates a pleasant atmosphere to conduct research in the group. I would also like to thank my industrial supervisor Dr. Nicolas Gresset for being a nice encadrant and a friend. We have lots of fruitful discussions and I am deeply impressed by his profound thinking, technical insight and sensibility in industrial practice. His generosity and integrity have inspired me as a role model for future career and I truly take him as not only a supervisor in research but also a mentor for life.

I thank many of my colleagues with whom I have the prestige to collaborate and conduct research. It's the joint work with Dr. Paul de Kerret that has led to chapter 5 of this thesis. We have many stimulating discussions together and I really have learned a lot from him. I thank assistant Professor Laura Cottatellucci for the discussions on the subject of random matrix theory. I owe her a great deal for her insight and help. I thank Xinping Yi, Haifan Yin, Miltiades C. Filippou, Rajeev Gangula and many others in the group, it's really a dynamic group to work with and I appreciate a lot for all of your help.

I would also like to thank Loïc Brunel, David Mottier, Mourad Khanfouci, Herve Bonneville, Cristina Ioana Ciochina, Jonathan Letessier, Akl Charaf, Victor Exposito, Jean Christophe Sibel and many others in wireless communication division in MERCE. MERCE is a fantastic company with amazing circumstance for working. I thank Magali Branchereau, Marie Plantard, Lucie Herve and Sophie Pautonnier for your help and support in

administrative works and my travel organizations.

I would like to thank all my friends Jingjing, Shengyun, Nan, Yang, Juan, Yulu, Mintao, I'm lucky to have you guys to share my happiness and sorrows. Finally, I would like to express my gratitude to my parents Xueqing and Xiaowu, thank you for your fully support and indulgence. I cannot accomplish this work without your encourage and love!

Contents

List of Figures

List of Tables

Acronyms

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

AMSE	Average Mean Square Error.
BS	Base Station.
CCP	Composite Concave Program.
CCSI	Centralized Channel State Information.
CoMP	Coordinated Multi-Point.
CSI	Channel State Information.
CSIT	Channel State Information at the Transmitter.
CSIR	Channel State Information at the Receiver.
DCSI	Distributed Channel State Information.
DCSIT	Distributed Channel State Information at the Transmitter.
DeNB	Donor E-UTRAN/Evolved Node B.
DoF	Degrees-of-Freedom.
eNB	E-UTRAN/Evolved Node B.
FDD	Frequency Division Duplexing.
i.i.d.	independent and identically distributed.
HeNB	Home E-UTRAN/Evolved Node B.
HeNB-GW	Home E-UTRAN/Evolved Node B Gate Way.
MC	Mont-Carlo.
LTE	Long Term Evolution.
LTE-A	Long Term Evolution Advanced.
MIMO	Multiple Input Multiple Output.
MME	Mobility Management Entity.
MMSE	Minimum Mean Square Error.
MSE	Mean Square Error.
M-TX	Master Transmitter.
RN	Relay Node.

RRH	Remote Radio Head.
RVQ	Random Vector Quantization.
RX	Receiver(s).
RZF	Regularized Zero Forcing.
S-GW	Serving Gate Way.
SINR	Signal to Noise and Interference Ratio.
SNR	Signal-to-Noise Ratio.
S-TX	Slave Transmitters.
s.t.	subject to.
TDD	Time Division Duplex.
TX	Transmitter(s).
ULA	Uniform Linear Array.
VQ	Vector Quantization.
ZF	Zero Forcing.

Notations

$\|.\|$	Absolute value of a number or the cardinality of a set.
$\|\boldsymbol{a}\|$	Euclidean norm of vector $\boldsymbol{a}$.
$\|\mathbf{A}\|_F$	Frobenius norm of Matrix $\mathbf{A}$.
$\mathbf{A}^*$	Complex conjugate of matrix $\mathbf{A}$.
$\mathbf{A}^{\mathrm{T}}$	Transpose of matrix $\mathbf{A}$.
$\mathbf{A}^{\mathrm{H}}$	Complex conjugate transpose (Hermitian) of matrix $\mathbf{A}$.
$\mathrm{tr}(\mathbf{A})$	Trace of matrix $\mathbf{A}$.
$\det(\mathbf{A})$	Determinant of matrix $\mathbf{A}$.
$\mathbf{0}_{a\times b}$	Zero matrix of size $a \times b$.
$\mathbf{I}_N$	Identity matrix of size $N \times N$.
$[\mathbf{A}]_{i,j}$	The (i,j)th entry of matrix $\mathbf{A}$.
$[\boldsymbol{b}]_i$	The ith entry of vector $\boldsymbol{b}$.
$\mathbf{D} = \mathrm{diag}(\boldsymbol{a})$	Diagonal matrix $\mathbf{D}$ with entries of $\boldsymbol{a}$ along its diagonal.
$\mathbb{R}^{n\times m}$	Real space of dimension $n \times m$.
$\mathbb{C}^{n\times m}$	Complex space of dimension $n \times m$.
$\mathrm{E}\left[.\right]$	Expectation.
$\mathrm{E}_A(f(a))$	Expectation of $f(a)$ when a belongs to A.
$\mathrm{var}\left[.\right]$	Variance.
$\mathrm{cov}\left[.\right]$	Covariance matrix.
sup	Supremum.
inf	Infimum.
lim sup	Limit superior.
lim inf	Limit inferior.
argmax	The feasible point that attains the maximum value.
argmin	The feasible point that attains the minimum value.
$1_{a=b}$	A function that returns 1 when $a = b$ and 0 otherwise.
$\mathfrak{i}$	Imaginary unit that satisfies $\mathfrak{i}^2 = -1$.
$\mathcal{N}_{\mathbb{C}}$	Complex Gaussian distribution.

$\xrightarrow{a.s.}$	almost surely converge to.
$x \asymp y$	x converges to y almost surely as K and M_{TX} go to infinity.

Chapter 1

Résumé [Français]

1.1 État de l'art de la coopération entre émetteurs

La coopération entre émetteurs est un sujet de recherche long terme à fort potentiel dans le domaine des télécommunications, exploitant éfficacement les interférences dans les réseaux cellulaires de prochaine génération [1]. Cependant, elle a aussi ses limites [2]. La version actuelle du 3GPP LTE-Advanced (LTE-A) a déjà intégré les transmissions coordonnées multi-points (CoMP, également connue sous le nom de réseau MIMO) pour lutter contre l'interférence inter-cellulaires. Dans cette version, de multiples stations de base (BS) voisines peuvent être regroupées et former un ensemble CoMP coopérant pour un traitement multi-cellulaires.

1.1.1 Techniques de traitement multi-cellulaires

Il existe différentes stratégies de coopération multi-cellulaires pour des stations de base dans un ensemble de CoMP. Ces stratégies comprennent les transmissions multiple-input-multiple-output (MIMO) et/ou l'ordonnancement coopératif multi-utilisateurs. Dans cette thèse, nous nous concentrons sur les techniques MIMO. Elles peuvent être cataloguées en fonction du niveau du partage de données multi-utilisateurs et du retour d'information sur le canal (CSI). Ceci est représenté en Fig. 1.1.

On nomme une coopération entre émetteurs sans échange d'informations coordonées lorsqu'il n'y a ni échange de données d'utilisateurs, ni échange explicite de données CSI sur les infrastructures. Dans ce cas, on peut par exemple utiliser des techniques MIMO multi-utilisateurs pour cellules individuelles [3] ou de stechniques de beamforming 3D [4, 5].

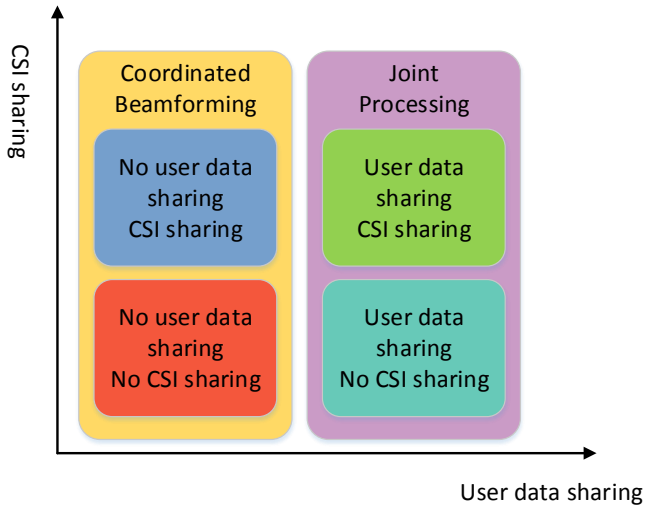


Figure 1.1: Catalogue des techniques de traitement multi-cellulaires selon les niveaux du partage des données utilisateurs et données CSI.

Si chaque utilisateur est uniquement servi par une station de base dans l'ensemble coopérant, c.a.d que les données pour cet utilisateur ne sont disponibles qu'à une seule station de base, alors on peut utiliser des techniques de **beamforming coordonné**. De nombreux travaux ont porté sur la conception de beamforming coordonné fondé sur différents niveaux de partage de CSI en utilisant diverses mesures de performance. Le beamforming coordonné sans partage explicite de CSI est étudié dans [6, 7]. Dans ces travaux, le beamforming coordonné entre les émetteurs est basé sur le CSI local instantané. Dans [6], la maximisation du SINR virtuel est utilisée comme mesure de performance. Dans [7], plusieurs mesures de performance fonctions de l'erreur quadratique moyenne (MSE) ont été examinées. Beaucoup d'autres travaux considèrent également le beamforming coordonné avec les données CSI partagées entre les émetteurs [8–12]. Dans ces travaux, différents niveaux de CSI partagé sont considérés: (i) le partage de CSI avec *coopération lente*, qui se réfère uniquement à l'échange des statistiques de CSI, et (ii) le partage de CSI avec *coopération rapide*, qui indique que les CSI instantanés sont échangés entre les émetteurs. Un exemple de conception de beamforming coordonné avec la coopération lente est décrite dans [8]. Pour une coopération rapide, nous pouvons continuer à différencier les scénarios selon que chaque émetteur puisse acquérir le CSI instantané de quelques liens (par exemple, les liens d'interférences [10]) ou le CSI global entre tous

les émetteurs et récepteurs [9,11]. La conception de beamforming coordonné qui minimise la puissance d'émission totale est présentée dans [9, 11]. Un beamformer qui maximise le débit et minimise les interférences résiduelles est proposé dans [10]. Une conception d'alignement d'interférence est considérée dans [12], où les signaux des émetteurs brouilleurs sont alignés dans le sous-espace orthogonzl au sous-espace engendré par l'espace de signal du récepteur.

On appelle **précodage conjoint** lorsque chaque utilisateur est desservi par toutes les stations de base dans l'ensemble coopérant, c.a.d lorsque les données d'utilisateurs sont partagées entre toutes les stations de base. Les instituts de normalisation débattent des mesures selon lesquelles les systèmes de communications mobiles de prochaine génération utiliseront des techniques de précodage conjoint selong des techniques de *MIMO réseau* afind d'atténuer les interférences inter-cellulaireset de maintenir l'efficacité spectrale et l'équité entre utilisateurs [1, 13]. Dans [14], le précodage coopératif multi-cellulaires a été examiné avec partage des données d'utilisateurs et pas de partage de CSI instantané entre les émetteurs. Le CSI instantané des autres émetteurs n'étant pas partagé, le gain du signal provenant d'autres émetteurs peut être estimé en se basant sur les statistiques des canaux. Beaucoup d'autres travaux analysent également le précodage conjoint avec les données d'utilisateurs et les données de CSI instantanées partagées [15–17].

1.1.2 Transmission centralisée et distribuée

Un autre problème important de la coopération entre émetteurs est de savoir si la stratégie de coopération est conçue dans un mode **centralisé** ou dans un mode **distribué**. la coopération entre émetteurs centralisé indique qu'il y a une *unité centrale* dans le système qui est connecté à tous les émetteurs dans l'ensemble coopérant. Dans la conception centralisée, cette unité centrale recueillera des informations de tous les émetteurs et créera un précodeur pour chacun émetteur. De nombreux travaux antérieurs [13, 18–22] envisagent un précodage conjoint avec une unité centrale. Cependant, la présence d'unité centrale dans le système n'est pas toujours acquise, et des limites sur l'infrastructure du réseau peuvent exister. En outre, cette architecture souffre du problème du point individuel de défaillance, ce qui rend le système moins robuste. Par conséquent, la coopération distribuée est une solution intéressante. La coopération entre émetteurs distribués suggère qu'il n'y a pas d'unité centrale dans le système. Le calcul de précodeur est mis en œuvre au niveau de chaque émetteur. Selon les objectifs,

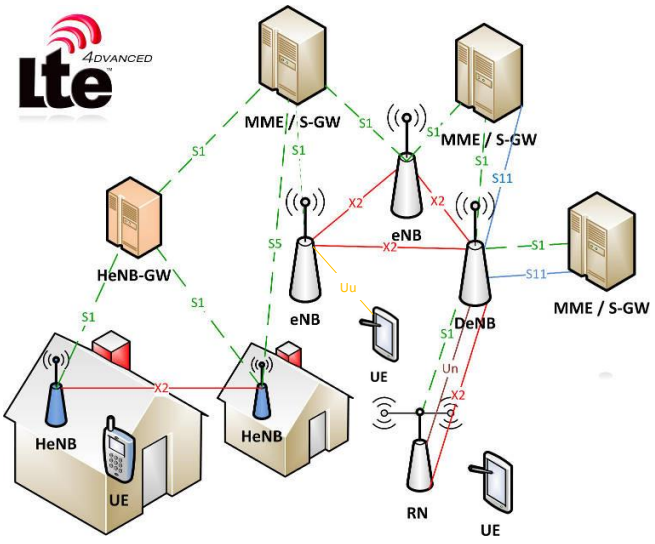


Figure 1.2: L'architecture LTE-A avec l'interface entre les noeuds différents.

les techniques telles que l'optimisation de consensus, la théorie des jeux et de l'optimisation distribuée pourraient être utilisées. De nombreux travaux visent l'optimisation distribuée dans les réseaux multi-agents [23–26]. Pour la transmission coopérative, la conception de beamforming coordonnée distribuée a été introduite dans [9–11, 27, 28] et la conception du précodage conjoint distribué a été étudié par exemple dans [16, 29].

1.2 Architecture des transmissions coopératives

Lors de la discussion en cours sur la 5G, la transmission coopérative pourrait avoir une influence sur l'architecture du système puisque les nouvelles stratégies de coopération conduiront à une redéfinition des fonctions sur chaque nœud [30]. Pour le système LTE-A actuel, l'architecture est décrite dans la Fig. 1.2 où sont illustrées les différentes interfaces entre les nœuds. Les nœuds représentés sont: E-UTRAN/ Evolved Node B (eNBs), Home eNBs (HeNBs), Home eNB Gate Way(HeNB-GW), Mobility Management Entity (MME)/Serving Gate Way (S-GW), Donor eNBs (DeNBs) et Relay Nodes (RNs). La transmission coopérative est considérée entre eNBs, HeNBs, DeNBs et RNs.

Dans l'architecture LTE-A, on observe une interface X2 qui permet l'échange d'information entre deux émetteurs coopérants. Dans le système

3GPP LTE Release 10, l'interface X2 gère deux fonctions: la fonction liée au handover et la fonction liée à l'interférence ou à la répartition de charge. Pour la fonction liée au handover, le trafic de données utilisateur est transmis à travers l'interface X2 en utilisant le protocole X2-U. Par conséquent, le tunnel pour le trafic de données utilisateurs est établi entre les nœuds coopérants lors de la procédure de handover et il n'existe que pendant une période de temps limitée. Pour la fonction liée à la charge et l'interférence, interface X2 supporte soit l'échange des informations dans le domaine fréquence/temporel, soit les informations de charge pour optimiser la gestion des ressources radio. Cet échange est construit sur le protocole X2-AP et est dédié à la au traffic de contrôle.

Dans cette thèse, nous considérons la transmission coopérative à travers une infrastructure du réseau qui est similaire à l'interface X2. Nous supposons que cette infrastructure est *dédié à la signalisation*: nous ne pouvons qu'échanger des informations de contrôle à travers cette interface. Comme nous pouvons le voir, l'interface X2 mentionnée ci-dessus est seulement une interface logique qui relie deux émetteurs coopérants, physiquement l'interface X2 peut être soit fibre, soit câble en cuivre, soit supportée par des liens radio sans fil. Par conséquent, l'infrastructure de coopération entre émetteurs est considérée en général à capacité limitée.

1.3 L'impact de feedback et backhaul limité sur le traitement multi-cellulaires

Une acquisition de CSIT précise est important pour beamforming coordonné et précodage conjoint. Dans un cas extrême de précodage conjoint avec CSI entre tous émetteurs et récepteurs parfaitement connue par chaque émetteur, émetteurs situés aux endoits différents peut être vu comme un réseau de multiples antennes virtuellement colocalisé desservant tous les récepteurs avec interférence étant complètement annulée. Par conséquent, le canal dégénère à un canal de diffusion et les algorithmes de précodage bien connus peuvent être utilisés [31]. Cependant, il faut faire attention que dans le système réel, le feedback par l'interface sans fil et le partage à travers les backhauls coordonnés sont limités.

1.3.1 L'imperfection et/ou delai à cause du feedback et partage de canal

Avec feedback limité entre le récepteur (RX) et l'émetteur (TX), chaque émetteur obtient une CSI imparfaite et/ou retardée. De nombreux travaux ont étudié beamforming coordonné ou l'alignement d'interférence en cas de CSIT imparfaite [7, 10, 32–37]. Certains travaux [33, 34] ont étudié la répartition des ressources CSI feedback tels que certains degrés de liberté (DoF) peuvent être atteints. d'autre travaux [7, 10, 36, 37] ont mis l'accent sur la conception de précodeur robuste avec CSI imparfaite. Certains ont également examiné l'effet de feedback limité et/ou une CSI retardée dans le cas du précodage conjoint à canal de diffusion [38–40].

Comme il est mentionné dans la section précédente, plusieurs algorithmes beamforming coordonné et precodage conjoint demandent différents niveaux de CSI partage entre les émetteurs. Ce partage de données CSI est principalement basé sur le backhaul. Le backhaul limité a introduit une imperfection et/ou un delai CSI en plus au-dessus du feedback limité. L'effet de backhaul limité a été pris en considération pour beamforming coordonné et precodage conjoint dans nombreux travaux. Dans [41], backhaul limité pour le partage CSI est pris en compte dans la conception d'alignement d'interférence. Dans [18–20], le backhaul avec capacité limité est considéré et l'analyse théoriques d'information de la performance du système de precodage conjoint est fourni. Dans [21, 22, 42–44], ils considèrent la conception de précodeur optimal pour le precodage conjoint avec backhaul limité.

Malgré l'erreur et le delai CSI introduit par backhaul limité, les émetteurs peuvent bénéficier beaucoup par partage CSI explicite à travers le backhaul. La procédure d'information partage est proactive, ce qui indique que TX peut choisir de partager les informations que d'autres TXs ont vraiment besoin ou des informations qui facilitent la mesure de la coopération entre émetteurs. Cependant, de nombreux travaux sur la coopération des émetteurs avec backhaul limité se concentrent plus sur l'aspect de la limitation du backhaul capacité, sans tenir compte de la possibilité que TXs peuvent décider les informations à partager. typiquement, cettes conceptions exploitent le backhaul avec quantification vectorielle aléatoire ou quantification scalaire, quelles que soitent les *propriétés statistiques* d'information locale déjà existant à l'émetteur, et sans tenir compte des avantages potentiels de *corrélation des estimations de canal initiales* disponible sur les émetteurs. Par conséquent, les algorithmes d'estimation de canal décentralisé avec backhaul limité qui exploitent de manière optimale le backhaul de coordination devront être envisagées.

1.3.2 Précodage sous information de canal distribuée

Hormis l'imperfection et/ou delai du canal, un autre problème important introduit par feedback et backhaul limité est que, après la procédure de CSI partage, chaque émetteur peut avoir CSI *imparfaite* et *non-identique*. C'est mentionné comme **CSIT distribuée** dans cette thèse et ce concept est introduit par exemple dans [45]. Il convient de noter que pour le traitement multi-cellulaires, de nombreux travaux (voir [13, 18–20, 22] entre les autres) supposent une conception centralisée basée sur une CSI *unique* et *imparfaite* à l'unité centrale. Cette hypothèse, qui est porté dans cette thèse comme **CSIT centralisée**, est défiée par scénarios avec feedback et backhaul limité.

Au meilleur de nos connaissances, nombreuses questions liées à CSIT distribuée sont encore ouverts. Par exemple, la région de la capacité du canal de diffusion avec CSIT distribuée est inconnue. Dans [45], une caractérisation des débit dans le régime de SNR élevé est effectuée en utilisant l'analyse DoF pour le scénario de deux émetteurs. Cette étude met en évidence la peine sévère causée par l'absence d'un CSI cohérente partagée entre les TXs coopérants au point de vue DoF lors de un précodeur conventionnel zéro forceur (ZF) est mis en place. Il est également démontré que les précodeurs robustes régularisés [46] ne restaure pas le DoF. Bien qu'une nouvelle stratégie de précodage décentralisé qui peut restaurer le DoF a été présenté dans [45] pour le cas de deux TXs, le résultat général de plus de 2 TXs reste ouverte. En outre, au régime de SNR limité, le problème de la conception précodeurs robustes décentralisés avec CSIT distribuée est complètement ouvert. L'utilisation de précodeurs conventionnels linéaires qui ne sont pas conscients de la structure de la CSIT distribuée va introduire une perte significative par rapport à un cadre de CSIT centralisée.

Pour la coopération décentralisée de l'émetteur, nous considérons particulièrement la conception du système grande en utilisant les outils de la théorie des matrices aléatoires. L'analyse du système grande est beaucoup utilisé dans la conception de système sans fil (Voir [28, 39, 47–51] parmi d'autres). Cependant, son rôle en aidant à analyser la coopération décentralisée avec émetteur CSIT distribué a reçu peu d'attention avant.

1.4 Organisation de thèse

Dans la première partie de cette thèse, nous étudions comment faire l'échange de l'information le plus efficace en utilisant le backhaul limité. nous considérons que chaque émetteur est doté d'une CSI initiale imparfaite et les émetteurs peuvent échanger leur CSI initiale avec backhaul limité. Nous op-

timisons d'abord le livre-code de quantification lorsque la capacité de backhaul est prédéfinie et chaque émetteur peut obtenir une estimation d'erreur moyenne quadratique minimale (MMSE) de canal après l'échange CSI. Nous étudions également le problème de l'allocation optimale des ressources backhaul pour chaque lien de telle sorte que la somme des MSE à chaque émetteur est réduite ou l'exactitude et la cohérence de la CSI estime à tous les émetteurs sont équilibrés lorsque le total de ressources backhaul est prédéfini.

Dans la deuxième partie de cette thèse, nous considérons l'algorithme distribué de lien descendante d'un système multi-cellulaires avec precodage conjoint qui est robuste à la configuration de la CSIT distribuée à chaque émetteur. Nous étudions un précodeur zéro forceur régularisé en système grande et optimisons le coefficient de régularisation, ainsi que le contrôle de puissance par émetteur telle que le débit total ergodique du système est maximisée.

1.5 Contributions et publications

Les publications suivantes sont le résultat des travaux présentés dans cette thèse.

- Qianrui Li, David Gesbert and Nicolas Gresset, "Joint Precoding over a Master-Slave Coordination Link", in proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2014
- Qianrui Li, David Gesbert and Nicolas Gresset, "A Cooperative Channel Estimation Approach for Coordinated Multipoint Transmission Networks", in proc. IEEE International Conference on Communications (ICC) Workshop on Small Cell and 5G Network, 2015
- Qianrui Li, David Gesbert and Nicolas Gresset, "Cooperative Channel Estimation for Coordinated Multipoint Transmission with Limited Backhauling", submitted to IEEE Trans. Wireless Commun., 2016
- Qianrui Li, Paul de Kerret, David Gesbert and Nicolas Gresset, "Robust Regularized ZF in Decentralized Broadcast Channel with Correlated CSI Noise", in proc. Allerton Conference on Communication (Allerton), 2015
- Qianrui Li, Paul de Kerret, David Gesbert and Nicolas Gresset, "Robust Regularized ZF in Cooperative Broadcast Channel under Distributed CSIT", submitted to IEEE Trans. Inf. Theory, 2016

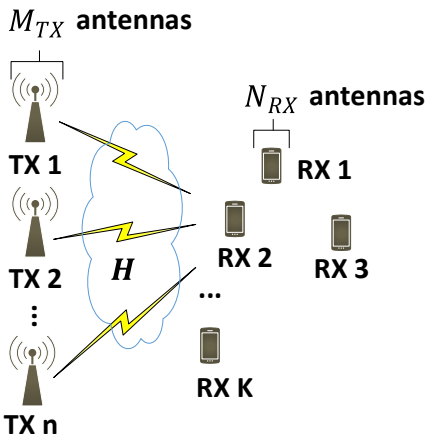


Figure 1.3: Un ystème multi-émetteurs multi-recepteurs aux liens descendants.

1.6 Coopérations des transmetteurs aux liens descendants

Dans cette thèse, nous nous concentrons sur la transmission coopérative avec feedback et backhaul limité aux lien descendantes. De nombreux travaux (voir [18, 52, 53] entre les autres) considèrent également problème très intéressant de transmission coopérative aux liens montants, mais transmission coopérative aux lien descendantes est plus difficile au point de vue de CSI limitée.

1.6.1 Modèle de transmission

Nous considérons un système de transmission multi-émetteurs sur la Fig. 1.3 avec n TXs servant au total K RXs. Chaque TX est équipée de M_{TX} antennes d'émission et chaque RX est équipé de N_{RX} antennes de réception. Le nombre total d'antennes TX est notée par

$$M = nM_{TX} \tag{1.1}$$

et le nombre total d'antennes RX est notée par

$$N = KN_{RX}. \tag{1.2}$$

Le signal $\mathbf{y}_k$ reçu par le $k$ème RX est égal à

$$\mathbf{y}_k = \mathbf{H}_k^{\mathrm{H}}\mathbf{x} + \boldsymbol{n}_k, \tag{1.3}$$

où $\mathbf{y}_k \in \mathbb{C}^{N_{RX}\times 1}$. $\mathbf{H}_k \in \mathbb{C}^{M\times N_{RX}}$ est le canal entre tous TXs et le $k$ème RX. $\boldsymbol{n}_k \in \mathbb{C}^{N_{RX}\times 1}$ est le bruit reçu par RX k.

Le signal reçu pour tous les RXs est présenté par

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{n}, \tag{1.4}$$

où $\mathbf{y} = \begin{bmatrix}\mathbf{y}_1^{\mathrm{T}} & \dots & \mathbf{y}_K^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{N\times 1}$, $\mathbf{H} = \begin{bmatrix}\mathbf{H}_1 & \dots & \mathbf{H}_K\end{bmatrix}^{\mathrm{H}} \in \mathbb{C}^{N\times M}$ est le canal entre tous TXs et RXs. $\mathbf{x} \in \mathbb{C}^{M\times 1}$ est le signal émis multi-TX et $\boldsymbol{n} = \begin{bmatrix}\boldsymbol{n}_1^{\mathrm{T}} & \dots & \boldsymbol{n}_K^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{N\times 1}$ est le bruit reçu multi-RX. Les éléments de $\boldsymbol{n}$ est sont indépendants et identiquement distribués (i.i.d) comme $\mathcal{N}_{\mathbb{C}}(0,1)$.

Le signal émis multi-TX $\mathbf{x} \in \mathbb{C}^{M\times 1}$ est obtenu à partir des données utilisateurs à transmettre $\boldsymbol{s} = \begin{bmatrix}\boldsymbol{s}_1^{\mathrm{T}} & \dots & \boldsymbol{s}_K^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{N\times 1}$ de telle manière que

$$\mathbf{x} = \mathbf{T}\boldsymbol{s} = \sum_{k=1}^{K} \mathbf{T}_k \boldsymbol{s}_k, \tag{1.5}$$

avec $\mathbf{T} = \begin{bmatrix}\mathbf{T}_1 & \dots & \mathbf{T}_K\end{bmatrix} \in \mathbb{C}^{M\times N}$ est le précodeur *multi-utilisateurs*, $\mathbf{T}_k \in \mathbb{C}^{M\times N_{RX}}$ est la matrice de précodage pour les données $\boldsymbol{s}_k$ à RX k. Nous considérons les données utilisateurs sont gaussien et les éléments de $\boldsymbol{s}$ sont i.i.d distribués comme $\mathcal{N}_{\mathbb{C}}(0,1)$.

Le précodeur multi-utilisateurs $\mathbf{T}$ peut être présenté sous une autre forme

$$\mathbf{T} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_n \end{bmatrix}, \tag{1.6}$$

où $\mathbf{W}_i \in \mathbb{C}^{M_{TX}\times N}$ est le précodeur implementé à TX i.

1.6.2 Beamforming coordonée aux liens descendants

Beamforming coordonée aux liens descendants est présenté dans Fig. 1.4, supposons que chaque TX dessert sulement un RX, $n = K$.

La matrice selective $\mathbf{E}_i^{\mathrm{H}} \in \mathbb{C}^{M_{\mathrm{TX}}\times M}$ dans Fig. 1.4 est notée par

$$\mathbf{E}_i^{\mathrm{H}} = \begin{bmatrix}\mathbf{0}_{M_{\mathrm{TX}}\times(i-1)M_{\mathrm{TX}}} & \mathbf{I}_{M_{\mathrm{TX}}} & \mathbf{0}_{M_{\mathrm{TX}}\times(n-i)M_{\mathrm{TX}}}\end{bmatrix}. \tag{1.7}$$

Étant donné que les données de utilisateurs ne sont pas partagées entre les TXs, hormis les contraintes de puissance, il y a d'autre contraintes posées sur la structure de précodeur multi-utilisateurs $\mathbf{T}$ puisque seulement $\boldsymbol{s}_i$ est

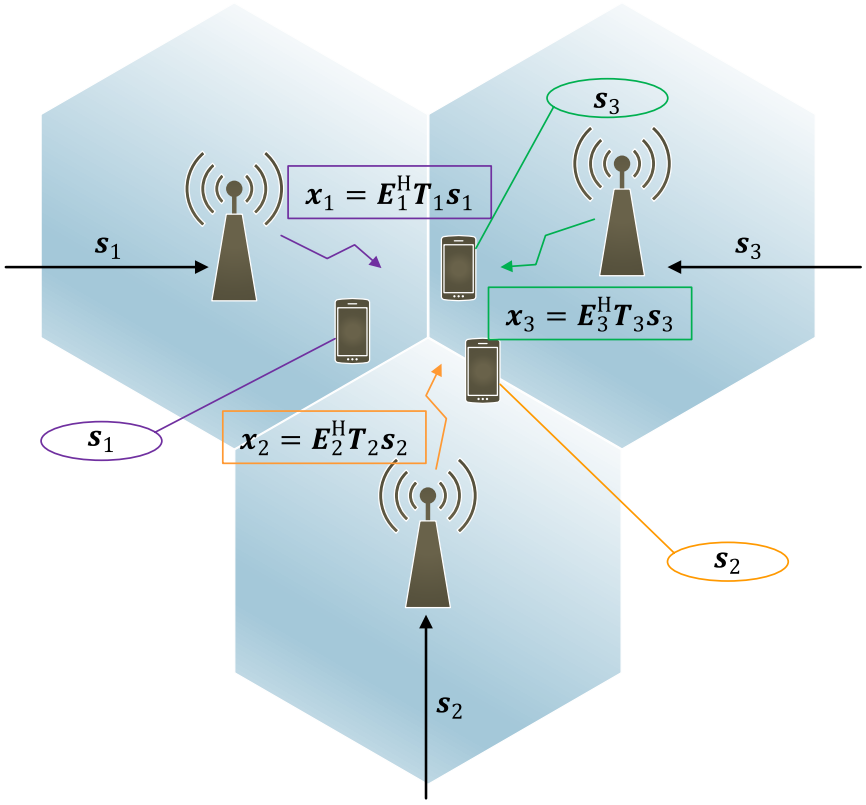


Figure 1.4: Coopérations des transmetteurs aux liens descendants: Beamforming coordonée.

disponible à TX i. Cela veut dire que le précodeur $\mathbf{W}_i$ implementé à TX i est

$$\mathbf{W}_i = \begin{bmatrix} \mathbf{W}_{i1} & \dots & \mathbf{W}_{iK} \end{bmatrix} \tag{1.8}$$

$$\mathbf{W}_{ik} = \mathbf{0}, \qquad \forall k \neq i, \tag{1.9}$$

où $\mathbf{W}_{ik} \in \mathbb{C}^{M_{TX} \times N_{RX}}$ contiens les colonnes de précodeur $\mathbf{W}_i$ dédié au précodage pour $\boldsymbol{s}_k$.

Avec tous les contraintes, on peut garantir que il n'y a que $\boldsymbol{s}_i$ disponible à TX i

$$\mathbf{x}_i = \mathbf{W}_i \boldsymbol{s} = \sum_{k=1}^{K} \mathbf{W}_{ik} \boldsymbol{s}_k = \mathbf{W}_{ii} \boldsymbol{s}_i = \mathbf{E}_i^{\mathrm{H}} \mathbf{T}_i \boldsymbol{s}_i. \tag{1.10}$$

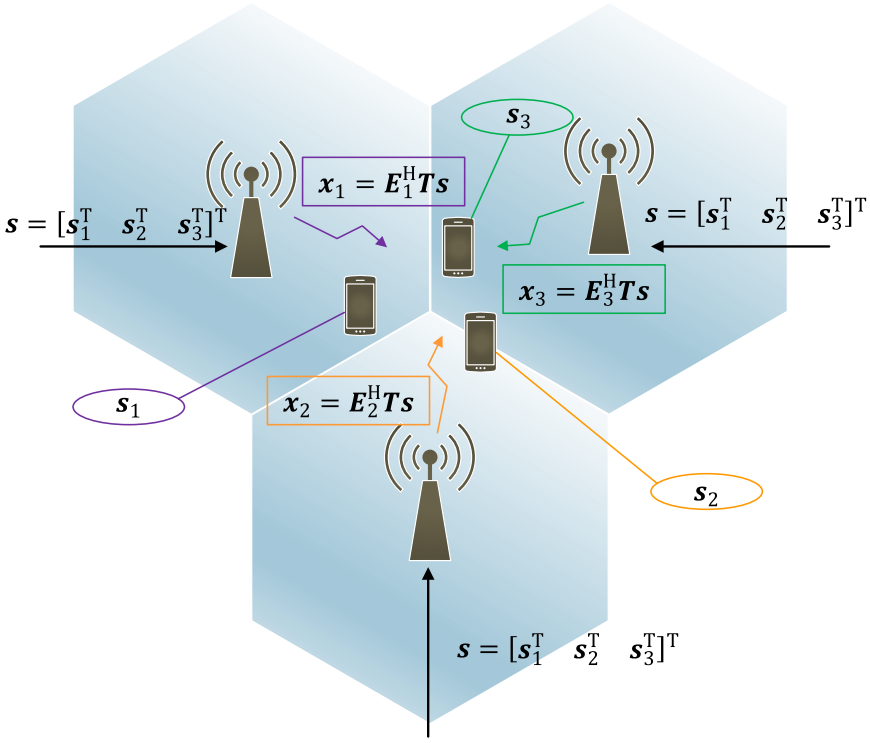


Figure 1.5: Downlink Transmitter Cooperation: Joint Processing.

1.6.3 Precodage conjoint aux liens descendants

Le precodage conjoint aux liens descendants est présenté dans Fig. 1.5. Les données multi-utilisateurs $\boldsymbol{s}$ sont partagées entre tous TXs. Chaque RX est desservi simultanément par tous les TXs dans l'ensemble coopérant. Le signal émis à TX i est défini comme

$$\mathbf{x}_i = \mathbf{W}_i \boldsymbol{s} = \mathbf{E}_i^{\mathrm{H}} \mathbf{T} \boldsymbol{s}. \tag{1.11}$$

Il n'y a aucune constrainte sur la structure de la matrice précodeur multi-utilisateurs $\mathbf{T}$ sauf les contraintes de puissance. On va discuter les contraintes de puissance en détail à Chapitre 6.4.

1.7 Modèle de canal pour feedback et backhaul limité

1.7.1 CSIT distribuée

Dans cette thèse, nous supposons que le backhaul limité est utilisé pour l'échange de l'information *instantanée*. Pour chaque information instantanée qui doit être partagée, le backhaul peut être utilisé une *seule* fois. Cependant, l'informations statistiques (par exemple, les statistiques CSI) qui fluctuent lentement sont partagée *parfaitement* entre les émetteurs dans l'ensemble coopérant.

Dans toute cette thèse, nous supposons également que les informations d'état de canal au récepteur (CSIR) sont parfaites. Nous faisons cette hypothèse afin de se concentrer sur l'effet de feedback et backhaul limité, sachant qu'ils sont responsable de la CSIT imparfaite [54]. Avec feedback et backhaul limité, le CSI obtenu à différents TXs sont imparfaites et non identiques. Ceci est présenté dans cette thèse comme CSIT distribué.

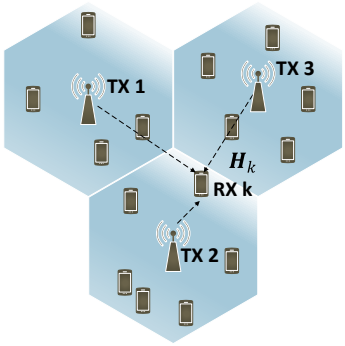
Chaque RX peut porter CSI à un ou plusieurs TXs basé sur feedback stratégies différentes. La CSI que TX i *initialement* obtenu à partir des RXs est notée par $\mathbf{H}^{(i)} \in \mathbb{C}^{N\times M}$.

Avec le backhaul, chaque TX peut partager CSIT avec ses TXs coopérants. Après cette procédure de CSI partage, la CSI que TX i emphfinalement obtenu est notée par $\hat{\boldsymbol{h}}^{(i)} \in \mathbb{C}^{N\times M}$.

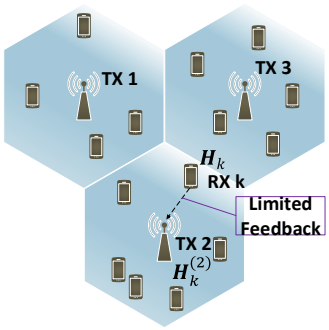
Comme une remarque, le modèle de CSIT distribuée est tout à fait générale et inclut de nombreuses combinaisons de stratégies différentes pour feedback et CSI partage.

Example 1 (LTE duplex à répartition en fréquence (DRF) scénario)**.** *Imaginez un scénario d'estimation de canal aux liens descendants LTE FDD, chaque TX diffusera sa séquence pilote, chaque RX recevra les pilotes et nous supposons qu'ils peuvent obtenir une CSIR parfaite. Ceci est illustré dans la Fig. 1.6a. Chaque RX porta CSI au TX qu'il est associé, comme représenté dans la Fig. 1.6b. Dans l'étape suivante, TXs peut échanger leur CSI à travers le backhaul, ce qui conduit à une CSI finale* $\hat{\mathbf{H}}^{(i)}$ *au TX* i*. Il est décrit dans la Fig. 1.6c.*

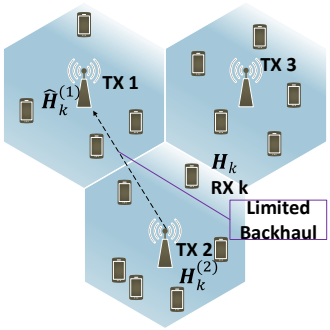
Il faut noter que nous avons légèrement abusé la notation $bH^{(i)}$ *parce que le CSI réele que chaque TX a reçu de RX n'est pas pour le système total, mais il a reçu seulement la CSI venue de RXs qui sont dans le cellulaire. Cependant, nous pouvons facilement résoudre ce problème en imposant que les coefficients de canal correspondant aux RXs hors-cellulaire sont zéro, ce*



(a) LTE DRF estimation de canal aux liens descendants: CSIR parfaite.



(b) LTE DRF estimation de canal aux liens descendants: CSI feedback limité.

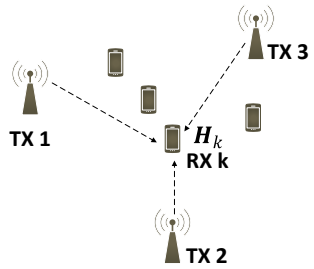


(c) LTE DRF estimation de canal aux liens descendants: backhaul CSI partage limité.

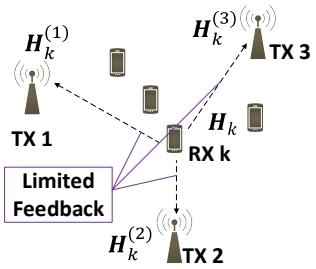
Figure 1.6: LTE DRF estimation de canal et CSI acquisition aux liens descendants avec feedback et backhaul limité.

qui indique qu'il n'y a pas de CSI pour ces RXs initialement disponible au TX.

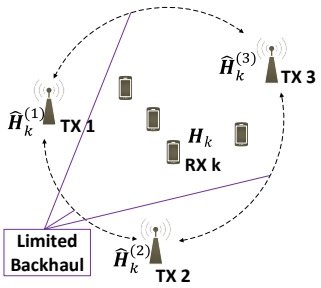
Example 2 (Broadcast feedback scénario)**.** *Dans ce scénario, chaque TX diffuse sa séquence pilote, chaque RX recevra les pilotes et obtiendra CSIR parfaite. Ceci est illustré dans la Fig. 1.7a. Par la suite, chaque RX effectuera un broadcast de CSI à tous les TXs, comme représenté dans la Fig. 1.7b. Après, TXs peut échanger leur CSI à travers le backhaul, ce qui conduit à une CSI finale* $\hat{\mathbf{H}}^{(i)}$ *au TX* i*. Il est décrit dans la Fig. 1.7c. Il faut remarquer que la CSI initiale à chaque TX après le broadcast par RX constitue déjà une structure de CSIT distribués puisque la condition de canal*



(a) Broadcast feedback scénario: CSIR parfaite.



(b) Broadcast feedback scénario: CSI feedback limité.



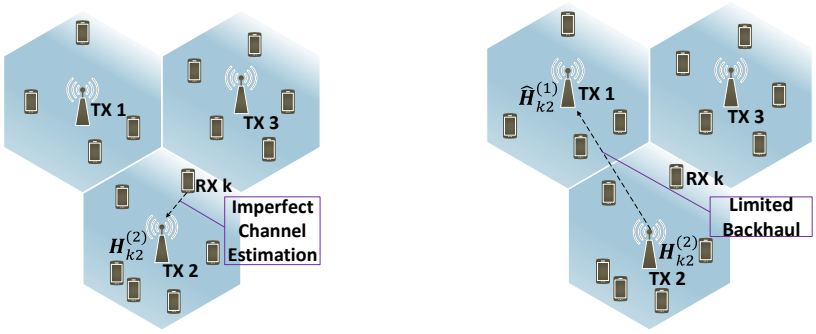
(c) Broadcast feedback scénario: backhaul CSI partage limité.

Figure 1.7: Broadcast feedback scénario avec feedback et backhaul limité.

pour chaque binôme TX-RX est différente.

Example 3 (LTE duplex à répartition dans le temps (DRT) scénario)**.** *Dans ce scénario, la CSIT initiale à chaque TX est acquis à partir d'une procédure d'estimation de canal aux liens montants et la CSI aux liens descendants est obtenue grâce à la réciprocité de canal. Au cours de l'estimation du canal aux liens montants, chaque RX envoie sa séquence pilote à la station de base associée. Comme on peut le voir dans la Fig. 1.8a, chaque BS ne peut que acquérir la CSI pour le canal entre ce BS et RXs dans le cellulaire. À l'étape suivante, les stations de base peuvent partager CSI à travers le backhaul limitée, ce qui conduit à une CSI finale* $\hat{\mathbf{H}}^{(i)}$ *au TX* i*. Ceci est décrit dans la Fig. 1.8b.*

Comme une remarque, nous abusons légèrement la notation $\mathbf{H}^{(i)}$ *au TX* i *puisque le canal acquis ici est que le canal entre TX* i *et ses RXs desservi. Nous pouvons garder la notation* $\mathbf{H}^{(i)}$ *et mettre toutes les autres éléments de canal hors-cellulaire à zéro.*



(a) LTE DRT scénario: estimation de canal imparfaite au TX.

(b) LTE DRT scénario: backhaul CSI partage limité.

Figure 1.8: LTE DRT scénario avec estimation de canal imparfaite et backhaul limité.

1.8 Facteurs de mérite du système

Dans cette section, nous parlons des facteurs de mérite du système dans cette thèse. Section 1.8.1, 1.8.2, 1.8.3 introduire les facteurs de mérite qui sont utilisés pour l'échange efficace d'informations à travers le backhaul. Section 1.8.4 montre le facteur de mérite pour la coopération distribuée à l'émetteur.

1.8.1 L'erreur quadratique moyenne à chaque émetteur

Une métrique intuitive que nous considérons est de laisser chaque TX obtenir une CSI finale plus précise après la procédure de backhauling. Par conséquent, le facteur de mérite est la MSE de CSI finale à chaque TX

$$D^{(i)} = \mathrm{E}\left[\|\mathbf{H} - \hat{\mathbf{H}}^{(i)}\|_F^2\right]. \tag{1.12}$$

Avec ce facteur de mérite, chaque TX peut optimiser de manière distribuée sa fonction de reconstruction pour la CSI finale et les livres-code pour la quantification sur backhaul.

1.8.2 La moyenne d'erreur quadratique moyenne entre tous les émetteurs

Un autre facteur de mérite que nous considérons est de minimiser la moyenne d'erreur quadratique moyenne (AMSE) entre tous les émetteurs

$$D^{av} = \frac{1}{n}\sum_{i=1}^{n} D^{(i)}. \tag{1.13}$$

Ce facteur de mérite est particulièrement utilisé dans le problème de l'allocation des ressources de backhaul pour chaque lien de coopération tels que la moyenne d'erreur quadratique moyenne entre tous les émetteurs est minimisée.

1.8.3 Équilibrer la précision et la cohérence entre les émetteurs

On peut observer que, pour le scénario de CSIT distribuée, la CSI finale à chaque TX n'est pas seulement imparfaite, mais aussi différente d'un TX à l'autre. Par conséquent, en dehors de l'analyse de la performance du système avec MSE ou AMSE, qui indique la précision de la CSI à chaque TX, nous devons aussi tenir compte de la cohérence de CSI à TXs différents puisque de nombreuses stratégies de coopération ont besoin d'une unité centrale pour traiter et CSI plus cohérente à différentes TXs sont plus favorable. L'équilibre entre la précision et la cohérence est basée sur un paramètre ρ

$$D^{bal} = \frac{\rho}{n}\sum_{i=1}^{n} D^{(i)} + \frac{1-\rho}{4n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}\{\|\hat{\mathbf{H}}^{(i)} - \hat{\mathbf{H}}^{(j)}\|_F^2\}. \tag{1.14}$$

Avec $\rho = 1$, ce facteur de mérite revient à l'AMSE entre tous les émetteurs. Avec $\rho = 0$, le facteur de mérite devient la somme des CSI différences carrés entre toutes les paires TXs. En réglant ce paramètre d'équilibrage, nous pouvons contrôler le niveau de centralisation et de la précision pour CSIT distribuée. Ce facteur de mérite est discuté en détail dans le chapitre 5.6.

1.8.4 Débit total ergodique du système

Le facteur de mérite que nous considérons dans la coopération de l'émetteur est le débit total ergodique du système

$$\mathrm{R}_{sum} = \mathrm{E}\left[\sum_{k=1}^{K} R_k\right], \tag{1.15}$$

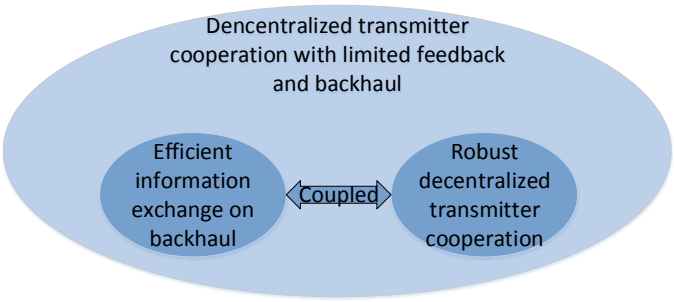


Figure 1.9: Transmission copérative distribuée avec feedback et backhaul limité.

où le débit pour RX k est noté par [55]

$$R_k = \log \frac{|\mathbf{I} + \sum_{j=1}^{K} \mathbf{H}_k^{\mathrm{H}} \mathbf{T}_j \mathbf{T}_j^{\mathrm{H}} \mathbf{H}_k|}{|\mathbf{I} + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{H}_k^{\mathrm{H}} \mathbf{T}_\ell \mathbf{T}_\ell^{\mathrm{H}} \mathbf{H}_k|}. \tag{1.16}$$

1.9 L'énoncé du problème

Comme il est mentionné dans le résumé, la transmission copérative avec feedback et backhaul limité implique deux étapes. La première étape est l'acquisition et partage de CSI entre les TXs, ce qui conduit à une structure d'information CSIT distribuée à chaque TX. La deuxième étape consiste à décider la stratégie à coopérer sur la base du CSIT distribuée disponible à chaque TX.

Comme on le voit dans la Fig. 1.9, l'objectif ultime est de résoudre ensemble le problème de l'exploitation du feedback et backhaul limité pour l'échange d'informations de la manière la plus efficace et concevoir une stratégie de coopération distribuée qui est robuste est performante fondé sur la CSI obtenue après le feedback et partage. Toutefois, cet objectif est trop ambitieux pour atteindre parce que chaque sous-problème est difficile lorsqu'il est abordé séparément, et pour l'optimisation conjointe, il s'agit d'une solution couplée. Par conséquent, nous avons appliqué une stratégie de diviser pour régner et examiné les deux sous-problèmes séparément.

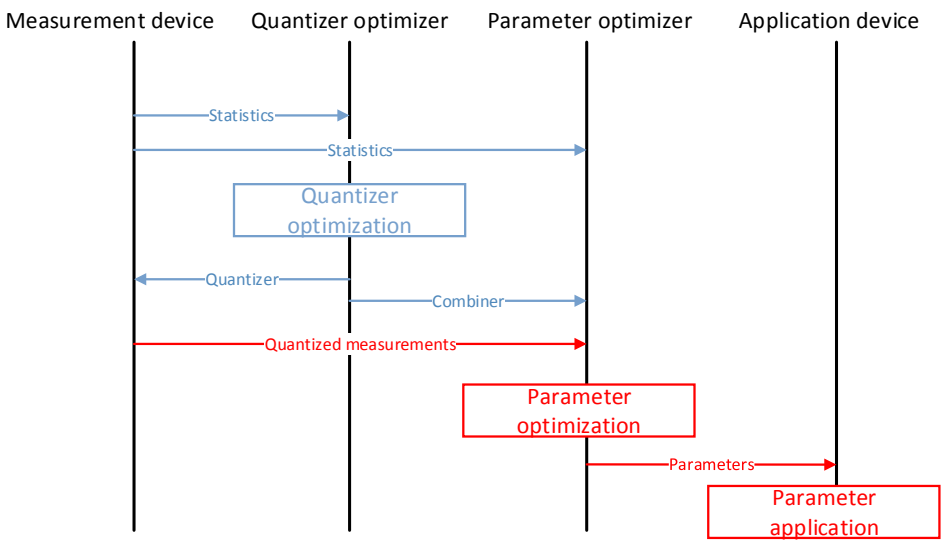


Figure 1.10: Graphique de séquence de temps pour une conception de l'échange efficace et une transmission coopérative distribuée.

1.9.1 Le découplage du problème

Nous avons divisé le problème d'optimisation conjointe en deux sous-problèmes:

- Une conception de l'échange d'informations efficaces, qui vise à *l'optimisation de quantificateurs* pour l'échange d'informations sur backhaul limité et la *fonction de combinaison* à chaque TX qui génère la CSIT finale basée sur CSIT initiale et des informations échangé.

- Une conception de coopération robuste et distribuée, qui effectue une optimisation des paramètres sur la base d'un facteur de mérite proposé et des informations quantifiées après l'échange.

Un graphique de séquence de temps dans Fig. 1.10 représente les procédures pour résoudre les deux problèmes. Selon la séparation de fonction à chaque TX, nous avons introduit les blocs fonctionnels suivants:

- Bloc de mesure, qui obtient des observations à chaque nœud et fournit au bloc quantification optimisée. Il peut fournir à la fois des statistiques de canal et CSI instantanée.

- Bloc quantification optimisée, qui recueille les statistiques et conçoit un quantificateur optimale pour la coordination avec backhaul limitée

et une fonction de combinaison optimale pour générer la CSI finale à chaque TX basée sur certain facteur de mérite.

- Bloc paramètre optimisé, qui conçoit une stratégie robuste de coopération distribuée à l'émetteur.
- Bloc d'application, qui met en œuvre la conception de précodeur coopération distribuée à l'émetteur.

Au cours de la phase de conception l'échange d'informations efficace, le bloc de mesure d'abord fournir au bloc quantification optimisée les statistiques. Le bloc quantification optimisée recueille les statistiques et conçoit un quantificateur optimale pour la coordination backhaul limitée basée sur certaine facteur de mérite. Par la suite, le bloc quantification optimisée informe le bloc de mesure de la quantification optimale qui doit être utilisé sur chaque lien de coordination. Le bloc quantification optimisée informe également le bloc paramètre optimisé comment les informations reçues d'autres nœuds doivent être combinées pour obtenir l'information finale instantanée à chaque TX.

Comme une remarque, après la phase de conception l'échange d'informations efficace, le bloc de mesure sait comment quantifier la CSI partagée sur backhaul et le bloc paramètre optimisé peut construire la CSI finale instantanée basée sur CSI initiale, CSI partagée et la fonction de combinasion.

Pendant la phase de conception d'une coopération robuste distribuée à TX, avec la CSI finale à chaque TX, les statistiques et le facteur de mérite, le bloc paramètre optimisé va concevoir une stratégie robuste de coopération distribuée à l'émetteur. Cette stratégie est informé au bloc d'application pour la mise en œuvre.

1.9.2 La Conception d'échange d'informations efficaces: le codage Wyner-Ziv

Le premier problème nous étudions est de trouver la façon la plus efficace d'utiliser le feedback et backhaul limité. Bien noté que ce problème est profondément enracinée dans le cadre du codage de source avec perte en utilisant des informations collatérales (le codage Wyner-Ziv [56]) et compression multi-sources avec des informations collatérales [57].

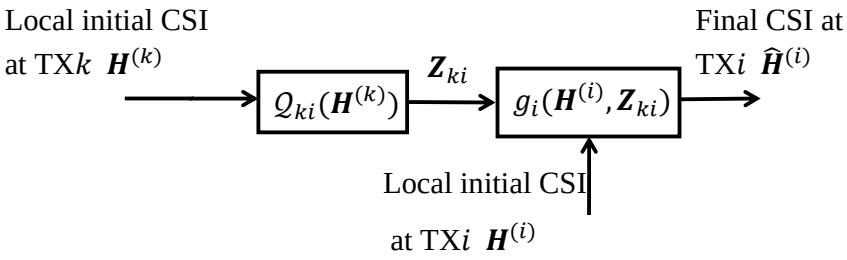


Figure 1.11: Le codage Wyner-Ziv: codage de source avec perte en utilisant des informations collatérales.

Figue. 1.11 révèle la nature du problème. Au TX k, il y a une première CSI $\mathbf{H}^{(k)}$ obtenue à partir de RX feedback. TX k partage sa CSI avec TX i avec le backhaul limité entre les émetteurs, ce processus est modélisé comme une procédure de quantification de débit fixe. Au TX i, il reconstruit une CSI finale $\hat{\mathbf{H}}^{(i)}$ fondé sur la CSI partagée quantifiée $\mathbf{Z}_{ki}$ de TX k et la CSI initiale $\mathbf{H}^{(i)}$ (appelé dans le codage Wyner-Ziv informations collatérales). Il faut noter que s'il y a plusieurs TXs qui partagent CSI avec TX i, on le nomme comme compression multi-sources avec des informations collatérales.

1.9.3 La conception de transmission coopérative: le problème de décision en l'équipe

Dans la deuxième partie de la thèse, nous considérons le problème de la conception de coopération tratégie robuste et distribuée avec CSIT distribuée. Il faut noter que ce problème appartient au problème de décision en l'équipe [58] qui est bien formulé. Dans un problème de décision en l'équipe, plusieurs agents du réseau souhaitent coopérer pour maximiser une utilité commune. Chaque agent a son *propre version limitée* de l'état du système, tous les agents doivent fournir une action / décision en fonction de leur propre point de vue limité. Ces actions doivent être cohérentes vers l'utilité commune.

Dans le cadre de la coopération TX, chaque TX a une CSI individuelle et conçoit le précodeur basé sur cette CSIT. Chaque TX n'a aucune idée sur le précodeur destiné à d'autres TXs. Cependant, chacun d'eux doit prendre la décision appropriée telle qu'une utilité commune est maximisée.

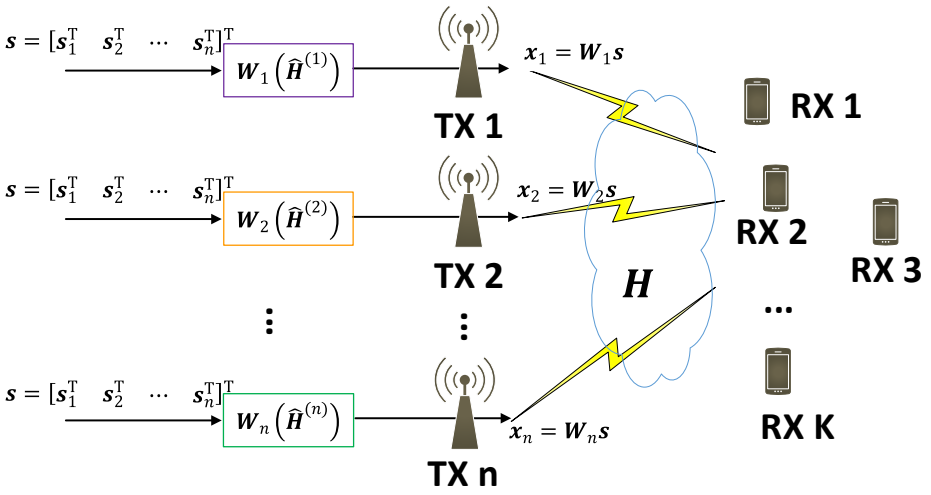


Figure 1.12: Le problème de la décision en l'équipe: Precodage conjoint avec n TXs et K RXs

Ce problème de la décision en l'équipe est formulé comme

$$\max_{\mathbf{W}_i(\hat{\mathbf{H}}^{(i)}),\forall i=1,\ldots,n} \mathrm{E}\left(\sum_i^n u_i\left(\mathbf{W}_1(\hat{\mathbf{H}}^{(1)}),\ldots,\mathbf{W}_n(\hat{\mathbf{H}}^{(n)}),\mathbf{H}\right)\right), \tag{1.17}$$

où $\mathbf{W}_i(\hat{\mathbf{H}}^{(i)})$ est la décision prise à TX i basé sur $\hat{\mathbf{H}}^{(i)}$ et $\mathrm{E}\{\sum_i^n u_i(\cdot)\}$ est l'utilité commune qui dépend de toutes les décisions TXS et l'état du système. Fig. 1.12 illustre le problème de la décision en l'équipe pour un système coopérative avec n TXs et K RXs.

1.10 Conclusion et nouveaux problèmes

Dans la première partie de la thèse, nous étudions d'abord plusieurs types d'information peuvent être transmis à travers le backhaul limité et comment des informations différentes pourrait affecter la conception de la coopération à l'émetteur. Après, nous nous concentrons donc sur l'échange de CSI et analysons le problème de l'échange CSI efficace avec backhaul limité. L'échange d'informations à travers le backhaul limité sont considérées comme des procédures de quantification. Nous étudions comment la CSI à chaque TX devrait être quantifiée et ce que chaque TX devrait faire avec la CSI initiale et l'informations partagées obtenue par le backhaul.

Nous rejetons ces problèmes dans un cadre systématique de l'estimation du canal coopérative avec les informations collatérales. Nous proposons des algorithmes qui permettent d'optimiser la conception du livre-code pour la quantification sur backhaul limité et résoudre le problème de l'allocation des ressources backhaul lorsque le nombre total de ressources backhaul sont limitées. Comme le feedback et backhaul limité conduire à une CSIT structure distribuée, nous considérons aussi à équilibrer la précision et la cohérence de la CSI à différents émetteurs afin d'améliorer la performance de la conception distribuée de transmission coopérative.

Dans la deuxième partie de la thèse, nous étudions la performance du système grande avec conception de précodeur RZF distribué dans la configuration de CSIT distribuée. Nous vise à dériver une expression déterministe équivalente pour le débit total ergodique du système. Nous voulons optimiser le coefficient de régularisation et la puissance d'émission à chaque TX de telle sorte que le débit total du système ergodique est maximisé.

Dans cette thèse, nous avons travaillé uniquement sur la coopération robuste distribuée pour l'émetteur avec RZF précodeur. Une direction intéressante pour les futurs travaux pourrait être l'analyse de coopération robuste distribuée de l'émetteur au-delà du précodeur RZF, e.g, pour d'autres stratégies de précodage possibles. Il est également un sujet très prometteur pour effectuer l'analyse des informations théoriques telles que l'analyse de la région de capacité pour la coopération distribuée à l'émetteur. Par ailleurs, pour la transmission cooperative et traitement du signal distribué avec feedback et backhaul limité, nous avons délibérément divisé en problèmes séparés comme la conception de l'échange d'informations efficace et la conception de la coopération distribuée à l'émetteur. Une autre direction intéressante mais difficile est le but ultime de l'optimisation conjointe. Pour finir, il est très prometteuse de analyser et déterminer quel degré de centralisation est la plus efficace dans le système avenir sans fil. Comme 5G et plus sont supposés être des réseaux très hétérogènes, une meilleure compréhension et décision sur le niveau de centralisation appliquée pour la coopération de l'émetteur pourrait même affecter l'architecture des réseaux futurs.

Part I

Motivations and Models

Chapter 2

Introduction

2.1 State of Art for Transmitter Cooperation

Transmitter cooperation is a long-lasting topic in communication due to its potential benefit in interference harnessing in next generation cellular networks [1]. However, it also has its shortcomings [2]. The current release of 3GPP LTE-Advanced (LTE-A) has already incorporated coordinated multi-point transmission (CoMP, also known as network MIMO) feature to fight against the inter-cell interference. In this release, multiple neighboring base stations (BS) can be aggregated together and form a CoMP cooperating set for multi-cell processing.

2.1.1 Techniques for multi-cell processing

There are various strategies for base stations in the CoMP cooperating set to perform multi-cell processing. These strategies include multiple-input-multiple-output (MIMO) transmission design and/or cooperative multi-user scheduling. In this thesis, we focus on the MIMO techniques. They can be cataloged based on their different levels of user data and CSI data sharing shown in Fig. 2.1.

If there is neither user data nor explicit backhaul-based CSI exchange between base stations, this is the transmitter cooperation without any coordination information exchange. Techniques such as single cell Multi-user MIMO [3] and 3D beamforming [4, 5] can be applied.

If each user is solely served by one base station in the cooperating set, i.e., the user data for this user is only available at one base station, **coordinated beamforming** can be used. Many works have focused on the coordinated beamforming design based on different levels of CSI sharing using a variety

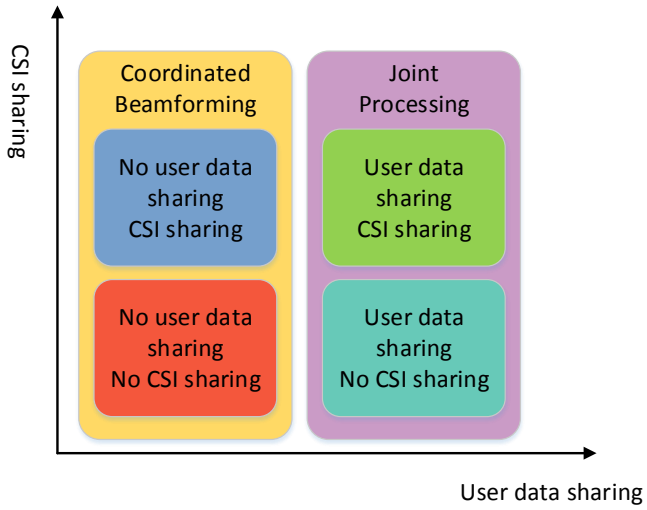


Figure 2.1: Catalog of Multi-cell processing techniques according to the levels of user data and CSI data sharing.

of performance measures. Coordinated beamforming without explicit CSI sharing is considered in [6, 7]. In these works, coordinated beamforming at each transmitter is based on its local instantaneous CSI. In [6], they have chosen virtual SINR maximization as performance measure. In [7], various performance measures that are functions of mean square error (MSE) have been considered. Many other works also consider the coordinated beamforming with the CSI data shared between transmitters [8–12]. In these works, they consider different levels of CSI sharing: (i) the CSI sharing can lead to *slow cooperation*, which refers to the exchange of only CSI statistics, and (ii) *fast cooperation*, which indicates that the instantaneous CSI is exchanged between transmitters. An example of coordinated beamforming design with slow cooperation can be found in [8]. For fast cooperation, we can further differentiate the scenarios based on whether each transmitter can acquire the instantaneous CSI of a few links (e.g, the interference links [10]) or even the global CSI between all transmitters and receivers [9, 11]. The coordinated beamforming designs that minimize the total transmit power are discussed in [9, 11]. A rate maximization beamformer that minimizes the interference leakage is proposed in [10]. An interference alignment design is considered in [12], where the signals of the interfering transmitters are aligned into the null space of the subspace spanned by the receiver's signal space.

If each user is served by all base stations in the cooperating set, i.e.,

the user data for this user is shared among all base stations, this is called **joint processing**. It is currently debated in standardization forums that to what extent next generation mobile communication systems will employ joint processing techniques denoted as *network MIMO* to mitigate inter-cell interference as well as to maintain spectral efficiency and system fairness [1, 13]. In [14], cooperative multi-cell precoding has been considered with user data shared and instantaneous CSI not shared between transmitters. Since the instantaneous CSI of the other transmitters is not shared, they have proposed a way to estimate the signal gain from other transmitters based on statistics of these channels. Many other works also focus on joint processing with both user data and instantaneous CSI data shared [15–17].

2.1.2 Decentralized and centralized transmitter cooperation

Another important issue in transmitter cooperation is whether the cooperation strategy is designed in a **centralized** fashion or in a **decentralized** fashion. Centralized transmitter cooperation indicates that there is a *central unit* in the system that is connected to all the transmitter in the cooperation set. In centralized design, this central unit will gather information from all the transmitters and design precoder for each of them. Many previous works [13, 18–22] consider joint processing with a central unit. However, a central unit in the system is not an assumption that is always valid due to the backhaul limitation and network topology. Also, this architecture suffers from the problem of single point of failure, which makes the system irresilient. Therefore, decentralized transmitter cooperation is introduced. Decentralized transmitter cooperation suggests that there is no central unit available in the system, the computation of the precoder is implemented at each transmitter. Based on different objectives, techniques such as consensus optimization, game theory and distributed optimization could be used. Many works target on the decentralized optimization in multi-agents networks [23–26]. For transmitter cooperation, decentralized coordinated beamforming design has been introduced in [9–11, 27, 28] and decentralized joint precoder design has been studied for example in [16, 29].

2.2 Architecture for Transmitter Cooperation

In the ongoing discussion about 5G, transmitter cooperation might have an influence on the system architecture since new cooperation strategies will lead to a redefinition of the functions on each node [30]. For the current

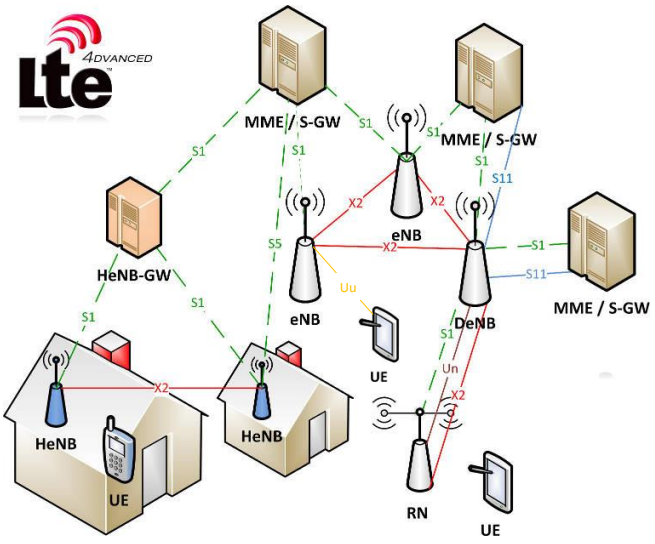


Figure 2.2: LTE-A architecture with interface between different nodes.

LTE-A system, the architecture is described in Fig. 2.2 with interface between different nodes. The nodes here include E-UTRAN/Evolved Node B (eNBs), home eNBs (HeNBs), home eNB gate way (HeNB-GW), mobility management entity (MME)/serving gate way (S-GW), donor eNBs (DeNBs) and relay nodes (RNs). Transmitter cooperation are considered between eNBs, HeNBs, DeNBs and RNs.

It can be found that in LTE-A architecture, there is X2 interface which allows the information exchange between two cooperating transmitters. In 3GPP Release 10, the X2 interface manages two functions: the handover related function and the load or interference related function. For the handover related function, the user plane traffic are conveyed via the X2 interface based on X2-U protocol. Consequently, the user plane tunnel is established between the cooperating nodes during the handover procedure and it only lives for a limited period of time. For the load and interference related function, X2 interface serves as exchanging either frequency/time domain information or load information to optimize radio resource management. This exchange is built upon X2-AP protocol and is dedicated to control plane communication.

In this thesis, we consider the transmitter cooperation through backhaul that is similar to the X2 interface. We assume that this backhaul is *dedicated to signaling*: we can only exchange control plane information through

this interface. As can be seen, the aforementioned X2 interface is just a logical interface that connects two cooperating transmitters, the physical X2 interface can be either wired fiber/copper lines or wireless radio links. Therefore, this backhaul between transmitters is considered in general to be rate limited.

2.3 Impact of Limited Feedback and Backhaul on Multi-cell Processing

An accurate CSIT acquisition is crucial for both coordinated beamforming and joint processing. In an extreme case of joint processing with perfect network-wise CSI at each transmitter, transmitters at different location can be seen as a collocated virtual multiple antennas array serving all receivers with interference being canceled completely. Therefore, the channel degenerates to a broadcast channel and well-known precoding algorithms from the literature can be used [31]. However, in real system, both the feedback through the wireless interface and the sharing through the coordination backhaul links are limited.

2.3.1 Channel feedback and sharing imperfection and/or delay

With limited feedback between receiver (RX) and transmitter (TX), each transmitter obtains imperfect and/or delayed CSI. Many works have considered how coordinated beamforming or interference alignment can cope with imperfect CSIT [7,10,32–37]. Some of these works [33,34] have studied the CSI feedback resource allocation such that certain degrees-of-freedom (DoF) can be achieved. Some of the works [7, 10, 36, 37] have focused on the robust precoder design under the imperfect CSI. A number of research efforts have also considered the effect of limited feedback and/or delay in joint processing broadcast channel setting [38–40].

As is mentioned in the previous section, many coordinated beamforming and joint processing algorithms requires different levels of CSI data sharing between transmitters. This CSI data sharing is mainly based on the coordination backhaul. The limited backhaul has introduced further CSI imperfection and/or delay on top of limited feedback. The effect of limited backhaul has been taken into consideration for both coordinated beamforming and joint processing in many works. In [41], limited backhaul for CSI sharing is considered in interference alignment design. In [18–20], the ca-

pacity limited backhaul is considered and information theoretic analysis of system performance for joint processing is provided. In [21, 22, 42–44], they consider optimal precoder design for the joint processing with limited backhaul.

Despite the error and delay for CSI introduced by limited backhaul, transmitters can benefit a lot from the explicit CSI sharing over the coordination backhaul. The information sharing procedure is proactive, which indicates that TX can choose to share the information that other TXs really desire or information that facilitates to the greatest extent the transmitter cooperation. However, many works on transmitter cooperation with limited backhaul focus more on the aspect of backhaul capacity limitation, without considering the possibility that TXs can decide the information to be shared. Typically, such designs exploit the backhaul using random vector quantization or scalar quantization, with no regard for the *statistical properties* of the local information already existing at the transmitter, and ignoring the potential benefits of *correlated initial channel estimates* available at the transmitters. Therefore, algorithm for decentralized channel estimation with limited backhaul that optimally exploit the coordination backhaul should be considered.

2.3.2 Precoding under distributed channel state information setting

Apart from the channel imperfection and/or delay, another important problem introduced by limited feedback and backhaul is that after the CSI sharing procedure, transmitters can have *imperfect* and *non-identical* CSI that varies from transmitter to transmitter. This is mentioned as **distributed CSIT** setting in this thesis and this concept is introduced for instance in [45]. It should be noticed that for multi-cell processing, many works (see [13, 18–20, 22] among others) assume a centralized design based on a *single* and *imperfect* CSI at central unit. This assumption, which is referred in this thesis as **centralized CSIT**, is challenged by limited feedback and backhaul scenario.

To the best of our knowledge, many questions related to the distributed CSIT setting are still open. For example, the capacity region of the broadcast channel under distributed CSIT setting is unknown. In [45], a rate characterization at high SNR is carried out using DoF analysis for the two transmitters scenario. This study highlights the severe penalty caused by the lack of a consistent CSI shared by the cooperating TXs from a DoF point of view when using a conventional zero forcing (ZF) precoder. It is also shown

that regularized robust precoders [46] do not restore the DoF. Although a new DoF-restoring decentralized precoding strategy was presented in [45] for the two TXs case, the general case of more than 2 TXs remains open. Furthermore, at finite SNR, the problem of designing decentralized robust precoders in distributed CSIT setting is fully open. The use of conventional linear precoders that are unaware of the distributed CSIT structure will introduce a significant loss with respect to a centralized CSIT setting.

For decentralized transmitter cooperation, we consider particularly the large system design using random matrix theory tools. Large system analysis is widely used in wireless system design (See [28, 39, 47–51] among others). However, its role in helping to analyze decentralized transmitter cooperation with distributed CSIT has received little attention before.

2.4 Organization of This Thesis

In the first part of this thesis, we study how to make the most efficient information exchange using the limited backhaul. we consider that each transmitter is endowed with an initial imperfect CSI and transmitters can exchange their initial CSI using limited backhaul. We first optimize the quantizer codebook design when the backhaul capacity is predefined such that each transmitter can obtain a minimum mean square error (MMSE) channel estimate after the CSI exchange. We also study the problem of optimal backhaul resource allocation for each backhaul link such that the sum of MSE at each transmitter is minimized or the accuracy and the consistency of the CSI estimates at all transmitters are balanced when the total number of backhaul resource is predefined.

In the second part of this thesis, we consider the decentralized downlink multi-cell joint processing algorithm that is robust to the distributed CSIT configuration at each transmitter. We investigate a regularized zero forcing precoder design in large system scenario and optimize the regularization coefficient as well as the per transmitter power control such that the system ergodic sum rate is maximized.

2.5 Contributions and Publications

The contributions presented in this thesis are listed below.

- **Information exchange and cooperation design with limited master-slave backhaul:** When the cooperating transmitters are in

master-slave scenario where one master transmitter is endowed with perfect CSIT wile other K slave transmitters have zero prior CSI, we compare the strategy to communicate quantized CSI or quantized precoder design over the limited backhaul between master and slaves. It is shown that these two backhaul exploitation strategies are equivalent. We also propose a sub-optimal and low complexity algorithm for the joint precoding design. These results have been published in

Qianrui Li, David Gesbert and Nicolas Gresset, "Joint Precoding over a Master-Slave Coordination Link", in proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2014

- **Cooperative channel estimation:** Obtaining accurate global CSI at multiple transmitter devices is critical to the performance of many coordinated transmission schemes. Practical CSI local feedback often leads to noisy and partial CSI estimates at each transmitter. With rate-limited bi-directional backhaul, transmitters have the opportunity to exchange a few CSI-related bits towards establishing global CSIT. This work investigates the possible strategies towards this goal. We propose a novel decentralized algorithm that produces MMSE-optimal global channel estimates at each device from combining local feedback and backhaul-exchanged information. The method adapts to arbitrary initial information topologies and feedback noise statistics. We also propose an algorithm to allocated the backhaul resources in order to minimize the average mean square error or balancing the accuracy and consistency of CSI estimates between transmitters. These results have been published in

  Qianrui Li, David Gesbert and Nicolas Gresset, "A Cooperative Channel Estimation Approach for Coordinated Multipoint Transmission Networks", in proc. IEEE International Conference on Communications (ICC) Workshop on Small Cell and 5G Network, 2015

  Qianrui Li, David Gesbert and Nicolas Gresset, "Cooperative Channel Estimation for Coordinated Multipoint Transmission with Limited Backhauling", submitted to IEEE Trans. Wireless Commun., 2016

- **Robust decentralized precoder design for transmitter cooperation under distributed CSIT:** We consider the sum rate performance of decentralized RZF in large joint processing coordinated

multi-point transmission network with a distributed channel state information (DCSI) setting. In DCSI setting, the CSI at various TXs are imperfect due to the limited feedback and the CSI estimates at different TXs are correlated due to the limited backhaul. In decentralized RZF precoder design, each TX computes elements of the precoder based on their individual CSI estimates of the global multiuser channel. Our analysis assumes that there is a finite number n of cooperation TXs, and that the number M_{TX} of transmit antennas and the number K of single antennas RXs are large. The deterministic equivalent of the signal-to-interference-plus-noise ratio (SINR) at the receiver as $M_{TX}, K \to \infty$ is derived. Building upon the deterministic equivalent, we optimize the power allocation and the optimal regularization coefficient at each TX such that the system ergodic sum rate is maximized. Numerical simulations confirm the improved robustness with respect to CSI inconsistency at different TXs, even with moderate number of antennas and receivers. These results have been published in

Qianrui Li, Paul de Kerret, David Gesbert and Nicolas Gresset, "Robust Regularized ZF in Decentralized Broadcast Channel with Correlated CSI Noise", in proc. Allerton Conference on Communication (Allerton), 2015

Qianrui Li, Paul de Kerret, David Gesbert and Nicolas Gresset, "Robust Regularized ZF in Cooperative Broadcast Channel under Distributed CSIT", submitted to IEEE Trans. Inf. Theory, 2016

Chapter 3

System Model and Problem Statement

3.1 Downlink Transmitter Cooperation

In this thesis, we focus on the downlink transmission cooperation with limited feedback and backhaul. Many works (see [18, 52, 53] among others) also consider very interesting problem of uplink transmitter cooperation but from the limited CSI point of view, downlink transmitter cooperation is more challenging.

3.1.1 Transmission model

We consider a multi-transmitters transmission system in Fig. 3.1 with n TXs serving in total K RXs. Each TX is equipped with M_{TX} transmit antennas and each RX is equipped with N_{RX} receive antennas. The total number of TX antennas is denoted by

$$M = nM_{TX} \tag{3.1}$$

and the total number of RX antennas is denoted by

$$N = KN_{RX}. \tag{3.2}$$

The signal $\mathbf{y}_k$ received by the kth RX reads

$$\mathbf{y}_k = \mathbf{H}_k^{\mathrm{H}}\mathbf{x} + \boldsymbol{n}_k, \tag{3.3}$$

where $\mathbf{y}_k \in \mathbb{C}^{N_{RX}\times 1}$. $\mathbf{H}_k \in \mathbb{C}^{M\times N_{RX}}$ is the channel from all TXs to the kth RX. $\boldsymbol{n}_k \in \mathbb{C}^{N_{RX}\times 1}$ is the receive noise for RX k.

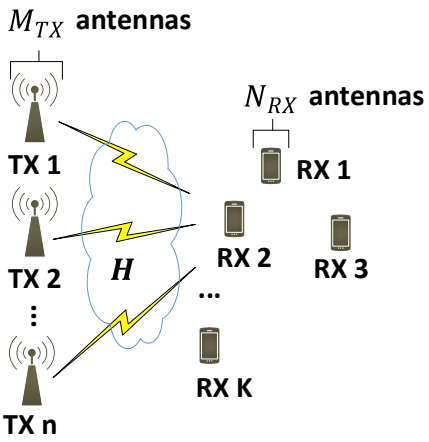


Figure 3.1: Multi-transmitters multi-receivers downlink transmission system.

The overall receiving signal for all the RXs is described as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \boldsymbol{n}, \tag{3.4}$$

where $\mathbf{y} = \begin{bmatrix}\mathbf{y}_1^{\mathrm{T}} & \dots & \mathbf{y}_K^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{N\times 1}$, $\mathbf{H} = \begin{bmatrix}\mathbf{H}_1 & \dots & \mathbf{H}_K\end{bmatrix}^{\mathrm{H}} \in \mathbb{C}^{N\times M}$ is the channel from all TXs to all RXs. $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is the multi-TX transmit signal and $\boldsymbol{n} = \begin{bmatrix}\boldsymbol{n}_1^{\mathrm{T}} & \dots & \boldsymbol{n}_K^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{N\times 1}$ is the multi-RX receive noise. The multi-RX receive noise $\boldsymbol{n}$ is considered to have independent identically distributed (i.i.d) entry distributed as $\mathcal{N}_{\mathbb{C}}(0,1)$.

The multi-TX transmit signal $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is obtained from the symbol vector $\boldsymbol{s} = \begin{bmatrix}\boldsymbol{s}_1^{\mathrm{T}} & \dots & \boldsymbol{s}_K^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{N\times 1}$ and takes the form

$$\mathbf{x} = \mathbf{T}\boldsymbol{s} = \sum_{k=1}^{K} \mathbf{T}_k \boldsymbol{s}_k, \tag{3.5}$$

with $\mathbf{T} = \begin{bmatrix}\mathbf{T}_1 & \dots & \mathbf{T}_K\end{bmatrix} \in \mathbb{C}^{M\times N}$ being the *multi-user* precoder, $\mathbf{T}_k \in \mathbb{C}^{M\times N_{RX}}$ being the precoding matrix for the data string $\boldsymbol{s}_k$ of RX k. Gaussian signaling is considered and the entries of multi-user symbol vector $\boldsymbol{s}$ are i.i.d $\mathcal{N}_{\mathbb{C}}(0,1)$.

The multi-user precoder $\mathbf{T}$ can be denoted in another form as

$$\mathbf{T} = \begin{bmatrix}\mathbf{W}_1 \\ \vdots \\ \mathbf{W}_n\end{bmatrix}, \tag{3.6}$$

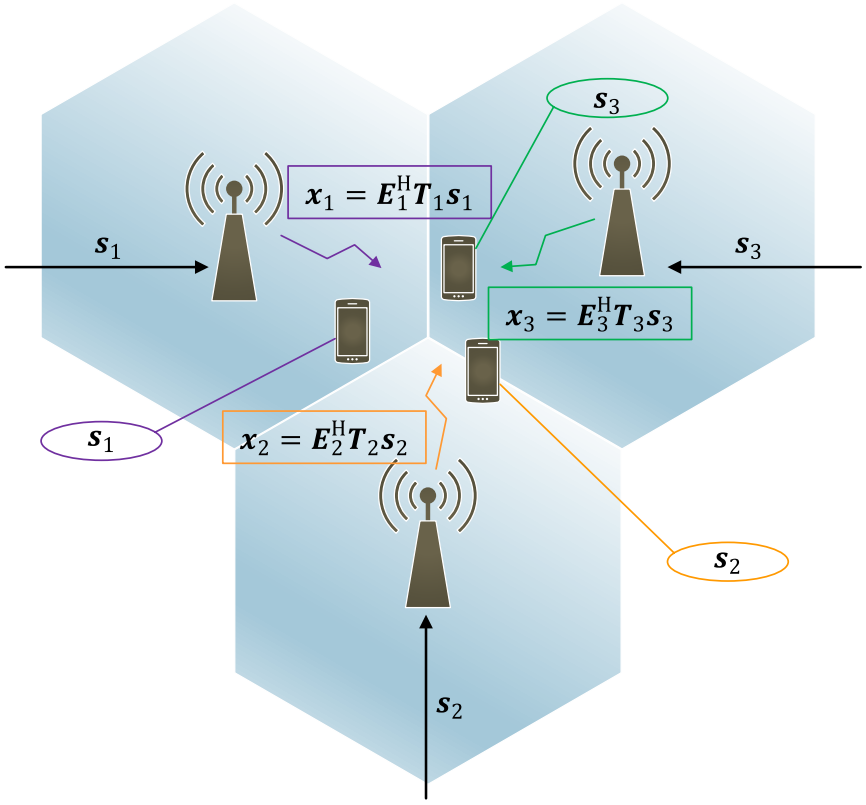


Figure 3.2: Downlink Transmitter Cooperation: Coordinated Beamforming.

where $\mathbf{W}_i \in \mathbb{C}^{M_{TX} \times N}$ is the precoder implemented at TX i.

3.1.2 Downlink coordinated beamforming

Downlink coordinated beamforming is shown in Fig. 3.2, assuming that each TX serves only one RX, $n = K$.

The selection matrix $\mathbf{E}_i^{\mathrm{H}} \in \mathbb{C}^{M_{\mathrm{TX}} \times M}$ in Fig. 3.2 is defined as

$$\mathbf{E}_i^{\mathrm{H}} = \left[\mathbf{0}_{M_{\mathrm{TX}} \times (i-1)M_{\mathrm{TX}}} \quad \mathbf{I}_{M_{\mathrm{TX}}} \quad \mathbf{0}_{M_{\mathrm{TX}} \times (n-i)M_{\mathrm{TX}}}\right]. \tag{3.7}$$

Since the user data is not shared between TXs, apart from the power constraints, there are some constraints on the structure of the multi-user precoder $\mathbf{T}$ since at TX i, only the data string $\boldsymbol{s}_i$ is available. This indicates

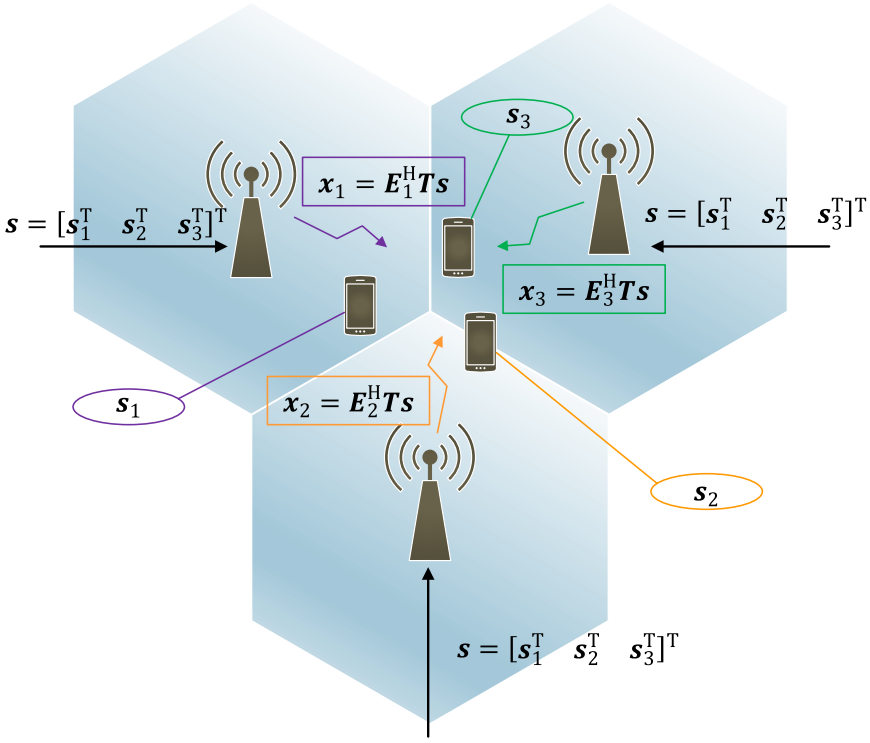


Figure 3.3: Downlink Transmitter Cooperation: Joint Processing.

that the precoder $\mathbf{W}_i$ implemented at TX i reads

$$\mathbf{W}_i = \begin{bmatrix} \mathbf{W}_{i1} & \dots & \mathbf{W}_{iK} \end{bmatrix} \tag{3.8}$$

$$\mathbf{W}_{ik} = \mathbf{0}, \qquad \forall k \neq i, \tag{3.9}$$

where $\mathbf{W}_{ik} \in \mathbb{C}^{M_{TX} \times N_{RX}}$ contains the columns of precoder $\mathbf{W}_i$ dedicated to precode $\boldsymbol{s}_k$.

With these constraints, we can guarantee that only $\boldsymbol{s}_i$ is available at TX i

$$\mathbf{x}_i = \mathbf{W}_i \boldsymbol{s} = \sum_{k=1}^{K} \mathbf{W}_{ik} \boldsymbol{s}_k = \mathbf{W}_{ii} \boldsymbol{s}_i = \mathbf{E}_i^H \mathbf{T}_i \boldsymbol{s}_i. \tag{3.10}$$

3.1.3 Downlink joint processing

Downlink joint processing is depicted in Fig. 3.3. The multi-user data string

$\boldsymbol{s}$ is shared between all TXs. Each RX is jointly served by all the TXs in the cooperating set. The signal transmit at TX i reads

$$\mathbf{x}_i = \mathbf{W}_i \boldsymbol{s} = \mathbf{E}_i^{\mathrm{H}} \mathbf{T} \boldsymbol{s}. \tag{3.11}$$

There is no constraint on the structure of the multi-user precoder matrix $\mathbf{T}$ except the power constraints. The power constraints will be analyzed in detail in Chapter 6.4.

3.2 Channel Model for Limited Feedback and Backhaul

3.2.1 Distributed CSIT

In this thesis, we assume that the limited backhaul is used for *instantaneous* information exchange. For each instantaneous information that needs to be shared, the backhaul can be used *only once*. However, the slowly fluctuated statistic information (e.g, the CSI statistics) are *perfectly* shared among transmitters in the cooperating set.

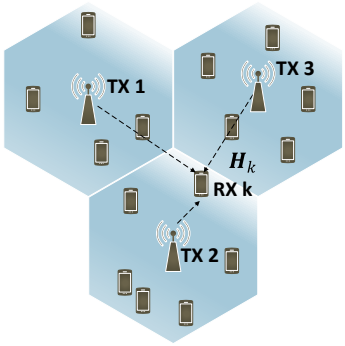
Throughout this thesis, we also assume that the channel state information at receiver (CSIR) is perfect. We make this assumption in order to focus more on the effect of limited feedback and backhaul, knowing that they are the main cause of imperfect CSIT [54]. With limited feedback and backhaul, the CSI obtained at different TXs are imperfect and non-identical. This is introduced in section 2.3.2 as distributed CSIT.

Each RX can feedback CSI to one or several TXs based on different feedback strategies. The CSIT that TX i *initially obtained* from RX feedback is denoted as $\mathbf{H}^{(i)} \in \mathbb{C}^{N \times M}$.

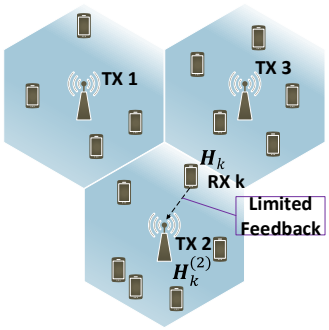
With the backhaul, each TX can share CSIT with its cooperating TXs. After this CSI sharing procedure, the CSI that TX i *finally obtained* is denoted as $\hat{\mathbf{H}}^{(i)} \in \mathbb{C}^{N \times M}$.

As a remark, the distributed CSIT model is quite general and encompasses many different combinations of feedback and CSI sharing strategies.

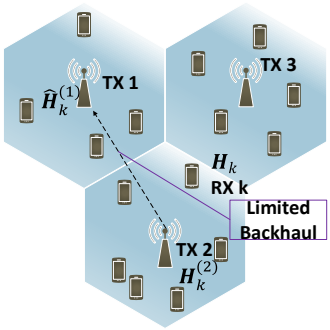
Example 4 (LTE frequency division duplexing (FDD) scenario)**.** *Consider a LTE FDD downlink channel estimation scenario, each TX will broadcast its pilot sequence, each RX will receive the pilots and we assume that they can obtain perfect CSIR. This is shown in Fig. 3.4a. Each RX will then feedback the CSI only to the TX that it is associated with, as is depicted in Fig. 3.4b. In the next step, TXs can exchange their CSI through the*



(a) LTE FDD downlink channel estimation: perfect CSIR.



(b) LTE FDD downlink channel estimation: limited CSI feedback.



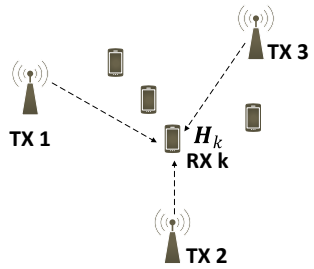
(c) LTE FDD downlink channel estimation: limited backhaul CSI sharing.

Figure 3.4: LTE FDD downlink channel estimation and CSI acquisition with limited feedback and backhaul.

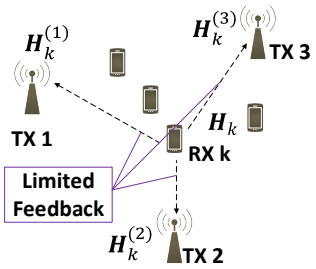
backhaul, which leads to a final CSI $\hat{\mathbf{H}}^{(i)}$ *at TX* i*. This is described in Fig. 3.4c.*

It should be noticed that we have slightly abused the notation $\mathbf{H}^{(i)}$ *because the actual CSI that each TX has received from RX feedback is not for the overall system but only CSI for RXs that are in the cell. However, we can easily solve this problem by imposing the channel coefficients corresponding to the out-of-cell RXs to zero, indicating that there is no CSI for these RXs initially available at this TX.*

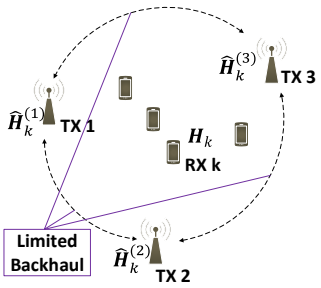
Example 5 (Broadcast feedback scenario)**.** *In this scenario, each TX broadcasts its pilot sequence, each RX receives the pilots and obtains perfect CSIR. This is shown in Fig. 3.5a. Afterward, each RX will perform a broadcast*



(a) Broadcast feedback scenario: perfect CSIR.



(b) Broadcast feedback scenario: limited CSI feedback.



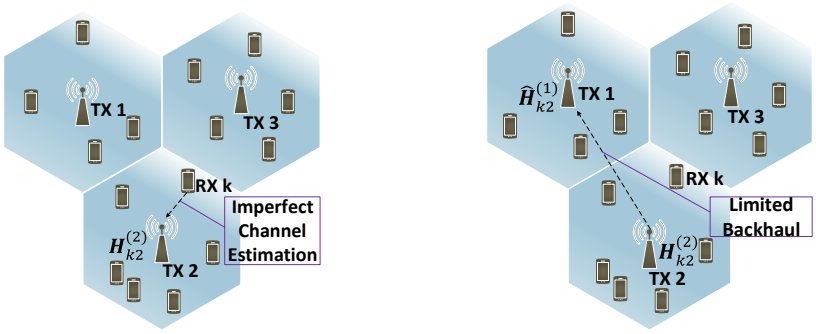
(c) Broadcast feedback scenario: limited backhaul CSI sharing.

Figure 3.5: Broadcast feedback scenario with limited feedback and backhaul.

feedback of CSI to all the TXs, as is depicted in Fig. 3.5b. TXs can then exchange their CSI through the backhaul, which leads to a final CSI $\hat{\mathbf{H}}^{(i)}$ *at TX* i*. This is described in Fig. 3.5c. It should be noticed that the initial CSI at each TX after RX's broadcast feedback already forms a distributed CSIT structure since the channel condition for different TX-RX pairs are different.*

Example 6 (LTE time division duplexing (TDD) scenario)**.** *In this scenario, the initial CSIT at each TX is acquired based on an uplink channel estimation procedure and the downlink CSI is obtained due to channel reciprocity. During the uplink channel estimation, each RX will send its pilot sequence to the associated BS. As is shown in Fig. 3.6a, each BS can only acquire the CSI for the channel between this BS and RXs in the cell. In the next step, BSs can share CSI through the limited coordination backhaul, which leads to a final CSI* $\hat{\mathbf{H}}^{(i)}$ *at TX* i*. This is described in Fig. 3.6b.*

As a remark, we slightly abuse the notation $\mathbf{H}^{(i)}$ *at TX* i *since the channel*



(a) LTE TDD scenario: imperfect channel estimation at TX.

(b) LTE TDD scenario: limited backhaul sharing.

Figure 3.6: LTE TDD scenario with imperfect channel estimation and backhaul.

acquired here is only the channel between TX i and its serving RXs. Again, we can keep the notation $\mathbf{H}^{(i)}$ *and set all the other channel entries to zero.*

3.3 System Figures of Merit

In this section, we talk about the possible system figures of merit in this thesis. Section 3.3.1, 3.3.2, 3.3.3 introduce the figures of merit that are used for efficient information exchange through the coordination backhaul. Section 3.3.4 demonstrates the figure of merit for decentralized transmitter cooperation.

3.3.1 Mean square error at each transmitter

An intuitive metric we consider is to let each TX obtain a more accurate final CSI after the backhauling procedure. Therefore, the figure of metric is the MSE of the final CSI at each TX

$$D^{(i)} = \mathrm{E}\left[\|\mathbf{H} - \hat{\mathbf{H}}^{(i)}\|_F^2\right]. \tag{3.12}$$

With this figure of merit, each TX can optimize in decentralized manner its reconstruction function for the final CSI and the codebooks for the quantization on backhaul.

3.3.2 Average mean square error at all transmitters

Another figure of merit we consider is to minimize the average mean square error (AMSE) at all the transmitters

$$D^{av} = \frac{1}{n}\sum_{i=1}^{n} D^{(i)}. \tag{3.13}$$

This figure of merit is particularly used in the backhaul resource allocation problem to optimally allocate the backhaul resource for each cooperation link such that the average mean square error at all the transmitters is minimized.

3.3.3 Balancing accuracy and consistency between transmitters

It can be observed that for the distributed CSIT scenario, the final CSI at each TX is not only imperfect, but also different from each other. Therefore, apart from analyzing the system performance with MSE or AMSE, which indicates the accuracy of the CSI at each TX, we should also consider the consistency of the CSI at different TXs since many cooperation strategies require a central processing unit and thus more consistent CSI at different TXs are more favorable. The balancing of the accuracy and consistency is based on a balancing factor ρ

$$D^{bal} = \frac{\rho}{n}\sum_{i=1}^{n} D^{(i)} + \frac{1-\rho}{4n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}\{\|\hat{\mathbf{H}}^{(i)} - \hat{\mathbf{H}}^{(j)}\|_F^2\}. \tag{3.14}$$

With $\rho = 1$, this figure of merit falls back to the AMSE at all the transmitters. With $\rho = 0$, the figure of merit becomes the sum of squared CSI discrepancy between all TX pairs. Thus, by tuning this balancing parameter, we can control the level of centralization and accuracy for the distributed CSIT. This figure of merit is discussed in details in Chapter 5.6.

3.3.4 System ergodic sum rate

The figure of merit we consider in the transmitter cooperation is system ergodic sum rate

$$\mathrm{R}_{sum} = \mathrm{E}\left[\sum_{k=1}^{K} R_k\right], \tag{3.15}$$

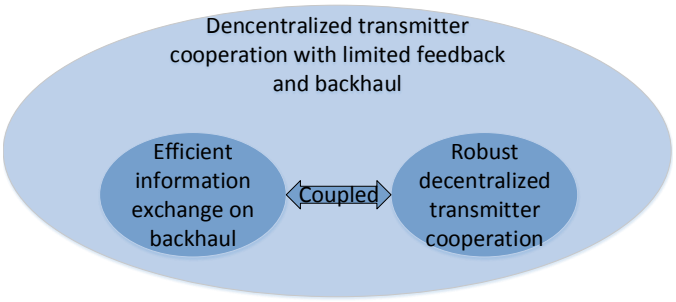


Figure 3.7: Decentralized transmitter cooperation with limited feedback and backhaul.

where the rate for RX k is denoted as [55]

$$R_k = \log \frac{|\mathbf{I} + \sum_{j=1}^{K} \mathbf{H}_k^{\mathrm{H}} \mathbf{T}_j \mathbf{T}_j^{\mathrm{H}} \mathbf{H}_k|}{|\mathbf{I} + \sum_{\substack{\ell=1 \\ \ell \neq k}}^{K} \mathbf{H}_k^{\mathrm{H}} \mathbf{T}_\ell \mathbf{T}_\ell^{\mathrm{H}} \mathbf{H}_k|}. \tag{3.16}$$

3.4 Problem Statement

As is mentioned in the abstract, transmitter cooperation with limited feedback and backhaul involves two steps. The first step is to acquire and share CSI between TXs, which leads to a distributed CSIT information structure at each TX. The second step is to decide the cooperating strategy based on the distributed CSIT available at each TX.

As is shown in Fig. 3.7, the ultimate goal is to jointly solve the problem of exploiting the limited feedback and backhaul for information exchange in the most efficient manner and design a robust and high performance decentralized transmitter cooperation strategy based on the CSI obtained after feedback and sharing. However, this goal is too ambitious to achieve since either sub-problem is difficult when taken alone and the joint optimization makes the two problems coupling together. Therefore, we have applied a divide-and-conquer strategy and considered the two sub-problems separately.

3.4.1 Decoupling the problem

We have split the joint optimization problem into two sub-problems:

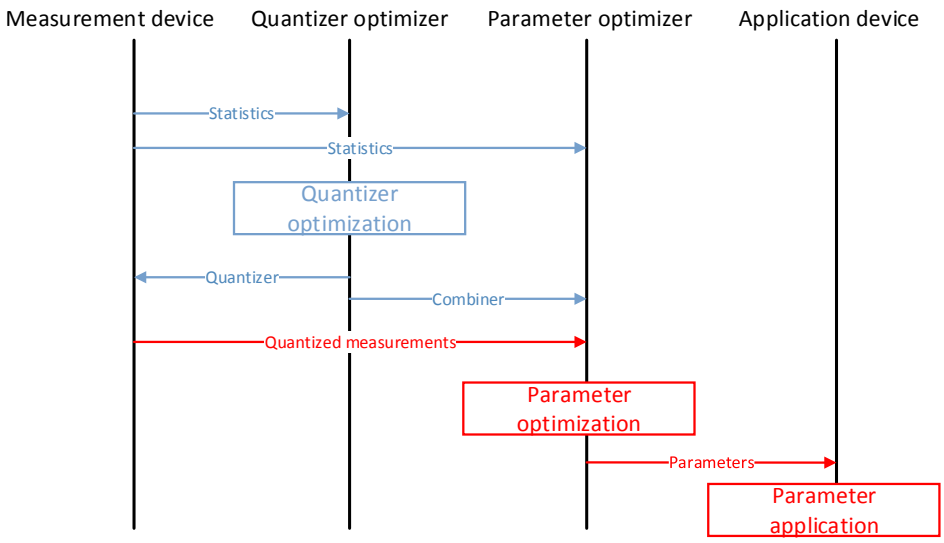


Figure 3.8: Time sequence graph for efficient information exchange design and robust decentralized TX cooperation design.

- An efficient information exchange design, which aims at *optimizing the quantizers* for the information exchange on the limited backhaul and the *combining function* at each TX that generates the final CSIT based on local initial CSIT and exchanged information.

- A robust decentralized transmitter cooperation design, which performs some parameter optimizations based on the a given figure of merit and the quantized information from the exchange.

A time sequence graph in Fig. 3.8 depicts the procedures for solving these two problems. According to the functional split at each TX, we have introduced the following functional blocks:

- Measurement device block, which obtains observations at the node and provides it to the quantizer optimizer block. It can provide both channel statistics and instantaneous CSI.

- Quantizer optimizer block, which collects the statistics and design the optimal quantizer for limited coordination backhaul link and the optimal combiner function for final CSI generating at each TX based on certain figure of merit.

- Parameter optimizer block, which designs a robust decentralized transmitter cooperation strategy.
- Application device block, which implements the decentralized transmitter cooperation precoder design.

During the efficient information exchange design phase, the measurement device will first provide the quantizer optimizer the statistics. The quantizer optimizer will collect the statistics and design the optimal quantizer for limited coordination backhaul link based on certain figure of merit. Thereafter, the quantizer optimizer will inform the measurement device about the optimal quantizer that should be used on each coordination link. The quantizer optimizer will also inform the parameter optimizer how the information received from other nodes should be combined to obtain the final local instantaneous information.

As a remark, after the efficient information exchange design phase, the measurement device knows how to quantize the CSI shared on the backhaul and the parameter optimizer can construct the instantaneous final CSI based on local initial CSI, shared CSI and the combiner function.

During the robust decentralized TX cooperation design phase, with the final CSI at each TX, the statistics and the figure of merit, the parameter optimizer will design a robust decentralized transmitter cooperation strategy. This strategy is then informed to the application device for implementation.

3.4.2 Efficient information exchange design: Wyner-Ziv coding problem

The first problem we investigate is to find the most efficient way to use the limited feedback and backhaul. Noticing that this problem is deeply rooted in the framework of lossy source coding with side information (Wyner-Ziv coding [56]) and multiple-source compressing with side information [57].

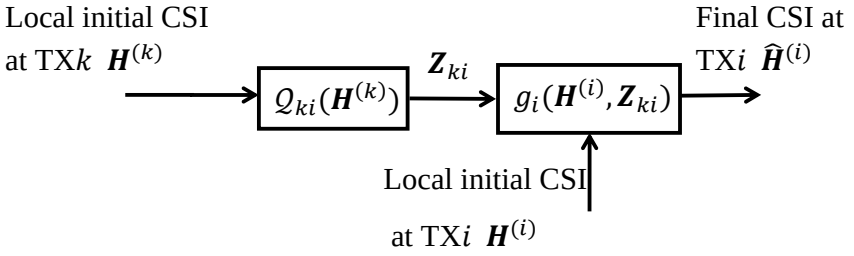


Figure 3.9: Wyner-Ziv coding: lossy source coding with side information.

Fig. 3.9 reveals the nature of the problem. At TX k, there is an initial CSI $\mathbf{H}^{(k)}$ obtained from RX feedback. TX k shares its CSI with TX i through the limited backhaul in between the transmitters, which is modeled as a fix rate quantization procedure. At TX i, it reconstructs a final CSI $\hat{\mathbf{H}}^{(i)}$ based on the quantized sharing CSI $\mathbf{Z}_{ki}$ from TX k and the local initial CSI $\mathbf{H}^{(i)}$ (referred in the Wyner-Ziv coding as side information). Noticing that if there are multiple TXs sharing CSI with TX i, this is referred as multiple-source compressing with side information.

3.4.3 Robust decentralized transmitter cooperation design: team decision problem

In the second part of the thesis, we consider the problem of robust distributed design for the transmitter cooperation strategy under distributed CSIT configuration. It should be noticed that this problem belongs to the well formulated framework of team decision problem [58]. In a team decision problem, several network agents wish to cooperate towards maximizing a common utility. Each agent has its *own limited view* of the system state, all agents have to come up with an action/decision based on their own limited view. These actions should be consistent toward the common utility.

In the context of TX cooperation, each TX has an individual CSI and will design the precoder based on this CSIT. Each TX has no idea on the precoder designed at other TXs. However, each of them should make the proper decision such that a common utility is maximized. This team decision problem is formulated as follows

$$\max_{\mathbf{W}_i(\hat{\mathbf{H}}^{(i)}),\forall i=1,\ldots,n} \mathrm{E}\left(\sum_{i}^{n} u_i\left(\mathbf{W}_1(\hat{\mathbf{H}}^{(1)}),\ldots,\mathbf{W}_n(\hat{\mathbf{H}}^{(n)}),\mathbf{H}\right)\right), \tag{3.17}$$

where $\mathbf{W}_i(\hat{\mathbf{H}}^{(i)})$ is the decision made at TX i based on $\hat{\mathbf{H}}^{(i)}$ and $\mathrm{E}\{\sum_i^n u_i(\cdot)\}$ is the common utility which is dependent on all TXs decisions and the system state. Fig. 3.10 demonstrates the team decision problem for a system of n TXs cooperation with joint processing serving K RXs.

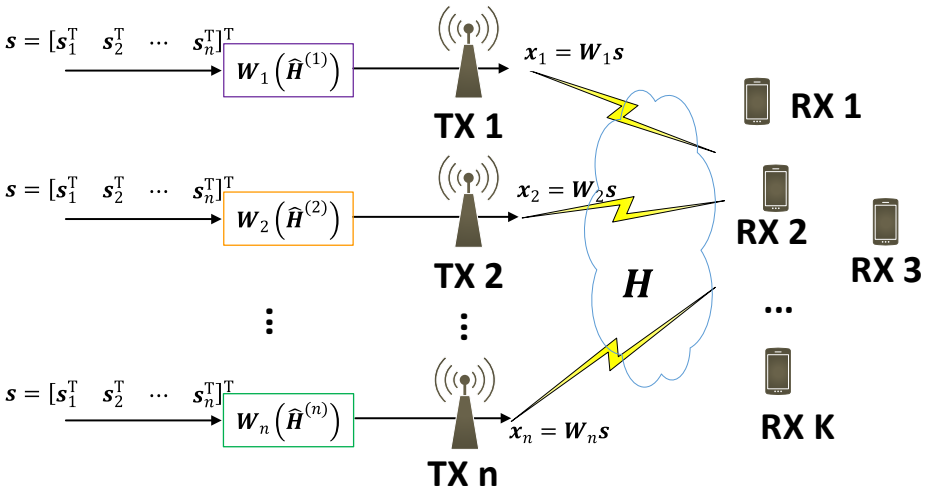


Figure 3.10: Team decision problem: n TXs cooperation with joint processing serving K RXs.

3.5 Summary of the Goals

In the first part of the thesis, we study first what kind of information can be conveyed through the limited backhaul and how different information could affect the transmitter cooperation design. We then focus on exchanging only CSI and analyze the problem of efficient CSI exchange for limited backhaul. Information exchange through the limited backhaul are considered as quantization procedures. We study how the CSI at each TX should be quantized and what each TX should do with the initial CSI and the shared information obtained from the backhaul. We cast these problems into a systematic framework of cooperative channel estimation with side information. We propose algorithms that can optimize the codebook design for the quantization on limited backhaul and solve the problem of backhaul resource allocation when the total number of backhaul resources are limited. As the limited feedback and backhaul will introduce a distributed CSIT structure, we also consider to balance the accuracy and consistency of CSI at different transmitters so as to enhance the performance of distributed transmitter cooperation design.

In the second part of thesis, we investigate the large system performance of decentralized RZF precoder design under the distributed CSIT configuration. We aims at deriving a deterministic equivalent expression for the system ergodic sum rate. We want to optimize the regularization coefficient

and the power control at each TX such that the system ergodic sum rate is maximized.

Part II

Efficient Information Exchange Design

Chapter 4

Design of Information Exchange and Cooperation on Limited Master-Slave Backhaul

When we begin to study the problem of efficient exploitation of the limited backhaul, the immediate question that comes into mind is:

"What is the nature of information that should be conveyed over the limited backhaul?"

Many different types of information can be exchange through the limited backhaul. We can share quantized CSIT, quantized precoder matrices or any quantized parameters that are useful in cooperation strategy design. For all types of information that can be possibly shared through backhaul, what is the difference and how will they affect the design of transmitter cooperation strategy?

We begin with a simple example by analyzing a specific Master-Slave setting [59] and compare different ways of exploiting the limited backhaul and provide accordingly the joint precoding strategy.

4.1 Master-Slave Model

We consider transmitter cooperation in the form of a n TXs network MIMO setup as is described in Section 3.1.1, where TX 1 acts as the master transmitter (M-TX) and TX 2 to TX n are the slave transmitters (S-TX). The M-TX is the TX that can obtain *perfect* CSI for all the serving RXs. The S-TXs are TXs that have *zero* initial CSI. It should be noticed that the master-slave model is a very special case for the distributed CSIT scenario. In LTE system, the M-TX and the S-TXs are similar to eNB and the remote radio heads (RRH) that are connected to the eNB respectively.

The propagation channel between RX j and TX i is denoted as $\mathbf{H}_{ji} \sim \mathbb{C}^{M_{TX} \times N_{RX}}$. Assume $\mathbf{h}_{ji} = vec(\mathbf{H}_{ji}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{C}_{\mathbf{h}_{ji}})$ and $\mathbf{n}_j \in \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I})$. The covariance matrix $\mathbf{C}_{\mathbf{h}_{ji}} = \mathrm{E}\{\mathbf{h}_{ji}\mathbf{h}_{ji}^H\}$ is positive semi-definite. TX i is subjected to an individual power constraint of P_i, that is, $\mathrm{E}\left\{\|\mathbf{x}_i\|^2\right\} = \|\mathbf{W}_i\|_F^2 \leq P_i$. Assume that each RX j is equipped with a linear MMSE receive filter $\mathbf{U}_j^{\mathrm{H}}$. This receive filter is calculated independently at each RX based on perfect CSI. The MMSE receive filter is denoted as [60]

$$\mathbf{U}_j^{\mathrm{H}} = \boldsymbol{\delta}_j^{\mathrm{H}} \mathbf{W}^{\mathrm{H}} \mathbf{H}_j (\mathbf{H}_j^{\mathrm{H}} \mathbf{W} \mathbf{W}^{\mathrm{H}} \mathbf{H}_j + \mathbf{I})^{-1}. \tag{4.1}$$

The mean square error (MSE) matrix for RX j reads

$$\mathbf{G}_j = \boldsymbol{\delta}_j^H (\mathbf{W}^{\mathrm{H}} \mathbf{H}_j \mathbf{H}_j^{\mathrm{H}} \mathbf{W} + \mathbf{I})^{-1} \boldsymbol{\delta}_j, \tag{4.2}$$

where $\boldsymbol{\delta}_j^{\mathrm{H}} = \left[\mathbf{0}_{N_{RX} \times (j-1)N_{RX}}, \mathbf{I}_{N_{RX}}, \mathbf{0}_{N_{RX} \times (K-j)N_{RX}}\right] \in \mathbb{C}^{N_{RX} \times KN_{RX}}$, is a selection matrix with $\mathbf{I}_{N_{RX}}$ at the jth block position and zero matrices elsewhere.

4.1.1 Master-slave coordination

We consider the existence of a coordination link between M-TX and S-TX ℓ with a rate limited to R_ℓ bits. This setting corresponds to a special case in Example 4 described in Section 3.2.1. We assume only a single use of the coordination link is allowed, by which M-TX will inform S-TX about the precoding strategy it should adopt. This setup is illustrated in Fig.4.1 with $K = 3$. We are interested in finding a linear precoding matrix at each TX such that the system sum rate will be maximized to the extent of what the limited coordination rate allows. Note that the sum rate can be conveniently written based on MSE matrices by [60]:

$$f_{SR}(\mathbf{H}, \mathbf{W}) = f_{SR}(\mathbf{H}, \mathbf{W}_1, \ldots, \mathbf{W}_n) = \sum_{j=1}^{K} \log\det(\mathbf{G}_j^{-1}). \tag{4.3}$$

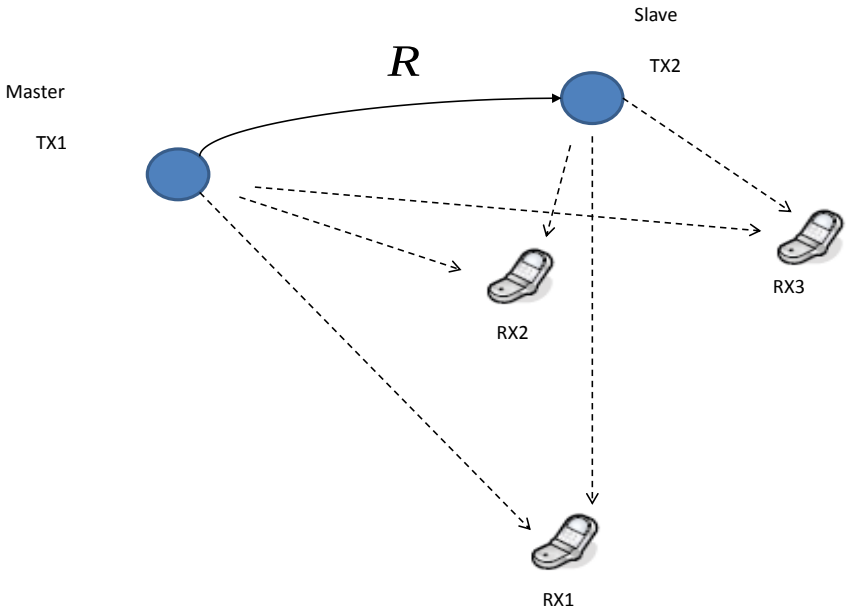


Figure 4.1: Master-slave coordination with a M-TX and a S-TX jointly serving 3 RXs.

4.2 Coordination Strategy in Master-Slave Model

We consider two options to exploit the coordination link: (i) share quantized CSI, or (ii) share quantized precoders.

4.2.1 M-TX sends precoder data

Let $\mathcal{C}_\ell^{prec}$ denote a codebook for the precoder decision sent from M-TX to S-TX ℓ. Since the signaling rate is limited to R_ℓ bits, the cardinality for $\mathcal{C}_\ell^{prec}$ is $\left|\mathcal{C}_\ell^{prec}\right| = 2^{R_\ell}$. Note that the codebook design could in principle be optimized based on the sum rate maximizing precoder distribution.

Optimally quantized precoder

Under the finite coordination rate constraint, the sum rate optimal linear precoding strategy at each TX is obtained from the following hybrid

continuous-discrete optimization problem

$$\begin{cases} \max\limits_{\mathbf{W}_1,\ldots,\mathbf{W}_n} & f_{SR}(\mathbf{H},\mathbf{W}_1,\ldots,\mathbf{W}_n) \\ & \mathbf{W}_1 \in \mathbb{C}^{M_{TX}\times KN_{RX}} \\ st: & \mathbf{W}_\ell \in \mathcal{C}_\ell^{prec}, \left|\mathcal{C}_\ell^{prec}\right| = 2^{R_\ell}, \qquad \forall \ell = 2,\ldots,n \\ & \|\mathbf{W}_i\|_F^2 \leq P_i, i = 1,\ldots,n \end{cases}, \tag{4.4}$$

where the M-TX proceeds with sending to S-TX ℓ the index of the optimal codeword for $\mathbf{W}_\ell$ in (4.4).

Naive quantized precoder

An intuitive yet naive algorithm (referred to later as naive quantized precoder) is to let M-TX compute in continuous domain the sum rate optimal precoders and send the quantized version of the obtained $\mathbf{W}_\ell$ to S-TX ℓ.

4.2.2 M-TX sends CSI data

For this coordination strategy, let $\mathcal{C}_\ell^{CSI} = \{q_1^{(\ell)},\ldots,q_{2^{R_\ell}}^{(\ell)}\}$ denote the codebook for normalized channel matrix quantization from M-TX to S-TX ℓ with $\left|\mathcal{C}_\ell^{CSI}\right| = 2^{R_\ell}$.

Optimal quantized CSI

Under the finite coordination rate constraint, the sum rate optimal linear precoding strategy at each TX can be obtained from the following optimization problem

$$\begin{cases} \max\limits_{\mathbf{W}_1,\ldots,\mathbf{W}_n} & f_{SR}(\mathbf{H},\mathbf{W}_1,\ldots,\mathbf{W}_n) \\ & \mathbf{W}_1 \in \mathbb{C}^{M\times NL} \\ & \mathbf{W}_\ell \in \mathcal{D}_\ell = \{p_1,\ldots,p_{2^{R_\ell}}\}, \qquad \forall \ell = 2,\ldots,n \\ st: & p_k = \mathbf{E}_\ell^{\mathrm{H}} \arg\max\limits_{\mathbf{W}} f_{SR}(q_k^{(\ell)},\mathbf{W}), \qquad \forall k = 1,\ldots,2^{R_\ell} \\ & \|\mathbf{W}_i\|_F^2 \leq P_i, i = 1,\ldots,n \end{cases}. \tag{4.5}$$

We recall that $\mathbf{E}_\ell^{\mathrm{H}} = \begin{bmatrix}\mathbf{0}_{M_{\mathrm{TX}}\times(i-1)M_{\mathrm{TX}}} & \mathbf{I}_{M_{\mathrm{TX}}} & \mathbf{0}_{M_{\mathrm{TX}}\times(n-i)M_{\mathrm{TX}}}\end{bmatrix}$ is defined in Section 3.1.1.

Naive quantized CSI

A common practice yet suboptimal algorithm (referred to later as naive quantized CSI) consists in quantizing $\mathbf{H}$ at the M-TX and sending the codeword index to S-TX. Just like the naive quantized precoder, the problem of this approach is that it ignores the asymmetry of CSI knowledge at the M-TX and S-TX.

4.2.3 An equivalence result

An interesting question is whether there is a fundamental advantage in signaling precoder-based vs. CSI-based data. The following provides an insight into the problem:

Proposition 1. *Given the codebooks $\mathcal{C}_\ell^{CSI}, \forall \ell = 2, \ldots, n$, there always exists codebooks $\mathcal{C}_\ell^{prec}, \forall \ell = 2, \ldots, n$ such that the optimization problem (4.5) and (4.4) attain the same optimum.*

Proof. This can be obtained by selecting a codebook $\mathcal{C}_\ell^{prec}$ that satisfies $\mathcal{C}_\ell^{prec} = \mathcal{D}_\ell$, where $\mathcal{D}_\ell$ is defined in problem (4.5). □

Hence when the codebook is properly designed, the optimal coordination strategies involving a communication of quantized precoders or CSI have equivalent performance.

4.3 Precoding for Master-Slave Coordination

According to proposition 1, without loss of generality, we will focus on a coordination strategy where M-TX sends the quantized precoder data, which is described by problem (4.4). This problem is difficult because it's a non-convex optimization problem over a non-convex set. Additionally, the complexity is prohibitive when R_ℓ increases.
We propose an alternating maximization algorithm for problem (4.4). The optimization is decomposed into 2 phases. In phase 1, M-TX solves $\mathbf{W}_\ell, \forall \ell = 2, \ldots, n$, based on a given $\mathbf{W}_1'$

$$\begin{cases} \max\limits_{\mathbf{W}_2, \ldots, \mathbf{W}_n} & f_{SR}(\mathbf{H}, \mathbf{W}_1', \mathbf{W}_2, \ldots, \mathbf{W}_n) \\ st: & \mathbf{W}_\ell \in \mathcal{C}_\ell^{prec}, \left|\mathcal{C}_\ell^{prec}\right| = 2^{R_\ell} \\ & \|\mathbf{W}_\ell\|_F^2 \leq P_\ell, \qquad \forall \ell = 2, \ldots, n \end{cases} . \tag{4.6}$$

In phase 2, M-TX solves for $\mathbf{W}_1$ in continuous space with given $\mathbf{W}'_\ell, \forall \ell = 2, \ldots, n$

$$\begin{cases} \max\limits_{\mathbf{W}_1} & f_{SR}(\mathbf{H}, \mathbf{W}_1, \mathbf{W}'_2, \ldots, \mathbf{W}'_n) \\ st: & \mathbf{W}_1 \in \mathbb{C}^{M_{TX} \times KN_{RX}} \\ & \|\mathbf{W}_1\|_F^2 \leq P_1 \end{cases} . \tag{4.7}$$

According to [60], the optimal precoder $\mathbf{W}$ maximizing the sum rate is the same as the precoder $\mathbf{W}$ derived by a weighted sum MMSE minimization problem. Here we extend the result to the case of decentralized precoders:

Proposition 2. *The optimization problem (4.7) has the same KKT point as the optimization problem*

$$\begin{cases} \min\limits_{\boldsymbol{W}_1} & \sum\limits_{j=1}^{K} tr(\boldsymbol{M}_j \mathbf{G}_j) \\ st: & \| \boldsymbol{W}_1 \|_F^2 \leq P_1 \end{cases} , \tag{4.8}$$

while $\boldsymbol{W}_\ell$ is fixed as $\boldsymbol{W}'_\ell, \forall \ell = 2, \ldots, n$ and the weight matrix for TXj satisfies

$$\boldsymbol{M}_j = (\mathbf{G}_j)^{-1}, \tag{4.9}$$

which is calculated with the optimal $\boldsymbol{W}_1$.

Proof. The lagrangian dual function for optimization problem (4.7) is

$$f(\mathbf{W}_1) = \sum_{j=1}^{K} - \log\det(\mathbf{G}_j) + \lambda(tr(\mathbf{W}_1 \mathbf{W}_1^H) - P_1). \tag{4.10}$$

Since

$$-\nabla_{[\mathbf{W}_1]_{nm}} logdet\left(\mathbf{G}_j\right) = -tr\left(\mathbf{G}_j^{-1} \frac{\partial \mathbf{G}_j}{\partial \left[\mathbf{W}_1^*\right]_{nm}}\right), \tag{4.11}$$

therefore, we can find

$$\left[\nabla_{\mathbf{W}_1} f(\mathbf{W}_1)\right]_{nm} = \nabla_{[\mathbf{W}_1]_{nm}} f(\mathbf{W}_1) \tag{4.12}$$

$$= \sum_{j=1}^{K} tr\left(\mathbf{G}_j^{-1} \frac{\partial \mathbf{G}_j}{\partial \left[\mathbf{W}_1^*\right]_{nm}}\right) + \lambda tr\left(\mathbf{W}_1 \mathbf{J}_{nm}\right), \tag{4.13}$$

where $\mathbf{J}_{nm}$ is a single entry matrix with 1 at (n, m) and 0 elsewhere.
The lagrangian dual function for optimization problem (4.8) when all slave precoders are fixed is

$$g(\mathbf{W}_1) = \sum_{j=1}^{K} tr(\mathbf{M}_j \mathbf{G}_j) + \mu \left(tr\left(\mathbf{W}_1 \mathbf{W}_1^H\right) - P_1 \right). \tag{4.14}$$

With a constant weight matrix $\mathbf{M}_j$ for RX$j, j = 1, \ldots, K$, we have

$$\left[\nabla_{\mathbf{W}_1} tr(\mathbf{M}_j \mathbf{G}_j)\right]_{nm} = tr\left[\mathbf{M}_j \frac{\partial \mathbf{G}_j}{\partial \left[\mathbf{W}_1^*\right]_{nm}}\right]. \tag{4.15}$$

Hence,

$$\left[\nabla_{\mathbf{W}_1} g(\mathbf{W}_1)\right]_{nm} = \nabla_{[\mathbf{W}_1]_{nm}} g(\mathbf{W}_1) \tag{4.16}$$

$$= \sum_{j=1}^{K} tr\left[\mathbf{M}_j \frac{\partial \mathbf{G}_j}{\partial \left[\mathbf{W}_1^*\right]_{nm}}\right] + \mu tr\left(\mathbf{W}_1 \mathbf{J}_{nm}\right). \tag{4.17}$$

Comparing $\left[\nabla_{\mathbf{W}_1} f(\mathbf{W}_1)\right]_{nm}$ and $\left[\nabla_{\mathbf{W}_1} g(\mathbf{W}_1)\right]_{nm}$, it is clear that the two problem have the same KKT point if $\mathbf{M}_j = \mathbf{G}_j^{-1}$. □

In the single M-TX and single S-TX case, we can simply use a complete search in the discrete set to solve problem (4.6).

4.4 Simulations

In this section the sum rate performance is evaluated for different settings using Mont-Carlo simulations. In all simulations $n = 3$ and $K = 3$, $M_{TX} = N_{RX} = 1$. R_2, R_3 are the rates for link between M-TX 1 and S-TX 2, S-TX 3. Each entry of $\mathbf{H}_{ji}$ is generated independently by a complex Gaussian random variable $u \in \mathcal{N}_{\mathbb{C}}(0, 1)$ multiplying a path loss component $v_{ji} = \alpha d_{ji}^{-\varepsilon}$, where d_{ji} is the distance between RXj and TXi, α is the cell edge SNR, $\varepsilon = 2$ is the decay factor. The TXs and RXs positions are generated uniformly in a circle cell with radius equal to 1km. The codebook are generated using random vector quantization (RVQ). The per-TX power constrain is $P_i = 1$Watt, $i = 1, \ldots, 3$.
Fig. 4.2 compares the sum rate for different strategies averaged over 100 realizations. The greedy quantized precoder algorithm well outperforms the naive algorithms, as these naive algorithms are unable to cope with the asymmetric nature of the CSIT. Our greedy algorithm perform close to the

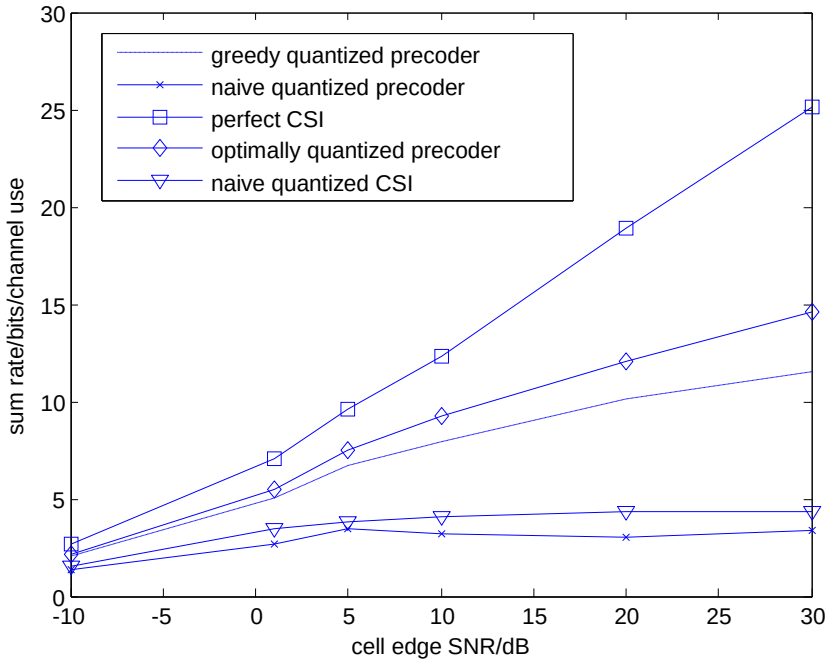


Figure 4.2: Sum rate performance of 1 M-TX, 2 S-TX, and $R_2 = R_3 = 3$bits.

optimal quantized precoder algorithm, while having very clear advantage in complexity against it. Fig. 4.3 confirms the improvement in performance as the signaling rate is increased.

4.5 Conclusion

We have studied the information exchange and the joint processing transmitter cooperation under a master-slave CSI configuration. We have compared signaling strategies based on exchange of CSI vs. precoder decisions through the limited backhaul. It is shown that with properly designed codebooks, these two signaling strategies are equivalent. We have also proposed optimal and suboptimal precoder design algorithms and exhibited good performance complexity trade-off.

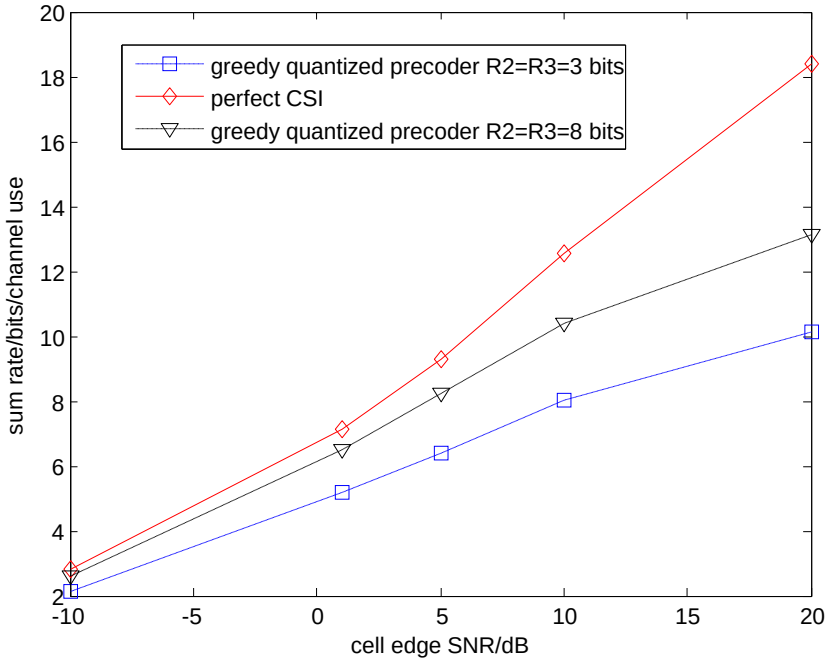


Figure 4.3: Sum rate performance of 1 M-TX, 2 S-TX, various signaling rate.

Chapter 5

Cooperative Channel Estimation and Backhaul Resource Allocation

In the previous chapter, we have compared sharing different information through the coordination backhaul for master-slave scenario. It reveals that for a properly designed codebook for the sharing information, the type of information that conveyed over the backhaul is irrelevant. In this chapter, we consider the scenario that the coordination backhaul is dedicated for *CSI sharing*. Under this assumption, we want to address the following questions:

"Which part of CSI should be shared, particularly, how the CSI shared through the backhaul should be quantized?"

"After receiving the quantized sharing CSI from other TXs, How can each TX constructs a final CSI?"

5.1 System Model and Problem Description

We consider a communication system with n TXs and K RXs as is described in Section 3.1.1. The cooperative TXs could be (small cell) base stations attempting to serve receiving terminals in a cooperative fashion. There exists a possible cooperative transmission strategies, generally requiring the availability of some global CSI at each TX [1,7,9,61]. Although the actual choice of the transmission scheme (joint MIMO precoding, interference alignment,

coordinated scheduling, coordinated resource allocation, etc.) may affect the CSI reconstruction problem at the TX side, such a question is left for further work while this paper focuses instead on the general problem of producing the best possible global CSI at each and every TX in a non discriminatory manner [62]. The impact of our channel estimation framework on the overall system performance is however evaluated in Section 5.7 for a particular example of network-MIMO enabled system.

We assume that the channels are frequency-flat Rayleigh fading, with $\boldsymbol{h} = vec(\mathbf{H}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{Q}_{\boldsymbol{h}})$, where $\mathbf{Q}_{\boldsymbol{h}}$ is an arbitrary multi-user channel covariance matrix.

5.1.1 Distributed CSI model

For the CSI model, we now assume that each TX acquires an initial estimate of the global channel state from an arbitrary pilot-based, digital or analog feedback mechanism. Similar to the CSI model used in [63, 64], the CSI made initially available at TX i is a noisy one. More generally, here the CSI imperfection is TX-dependent, giving rise to a distributed CSI model as initially introduced in [45]. Let

$$\mathbf{H}^{(i)} = \mathbf{H} + \mathbf{E}^{(i)}, \tag{5.1}$$

where $\mathbf{H}^{(i)} \in \mathbb{C}^{KN_{RX} \times nM_{TX}}$ is an initial CSI estimate for $\mathbf{H}$ at TX i. $\mathbf{E}^{(i)} \in \mathbb{C}^{KN_{RX} \times nM_{TX}}$ is the estimation error seen at TX i. Hence, the *estimates* at various TXs can be correlated through $\mathbf{H}$. The channel independent $\mathbf{E}^{(i)}$ satisfies $\mathbf{e}^{(i)} = vec(\mathbf{E}^{(i)}) \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{Q}_i)$. The errors terms seen at different TXs are assumed independent, i.e, $\mathrm{E}\left[\mathbf{e}^{(i)} \cdot \mathbf{e}^{(j)^{\mathrm{H}}}\right] = \mathbf{0}, \forall i \neq j$. Throughout this work, the channel statistics $\mathbf{Q}_{\boldsymbol{h}}$ and all error statistics $\mathbf{Q}_i, \forall i = 1, \ldots, K$ are assumed to be known at every TX by virtue of slow statistical variations.

5.1.2 Limited rate coordination model

Consider that the TX devices are equipped with rate-limited bi-directional communication links over which they can exchange a finite amount of CSI related information. Note that we only allow a *single shot* of coordination which consumes R_{ki} bits of communication from TX k to TX i for all cooperating (k, i) pairs *simultaneously*. The problem is now to optimally exploit this coordination capability so as to acquire the best possible global channel estimate at each TX.

In Fig. 5.1, the cooperation information exchange between two transmitters TX i and TX k is illustrated. TX k sends to TX i a suitably quantized

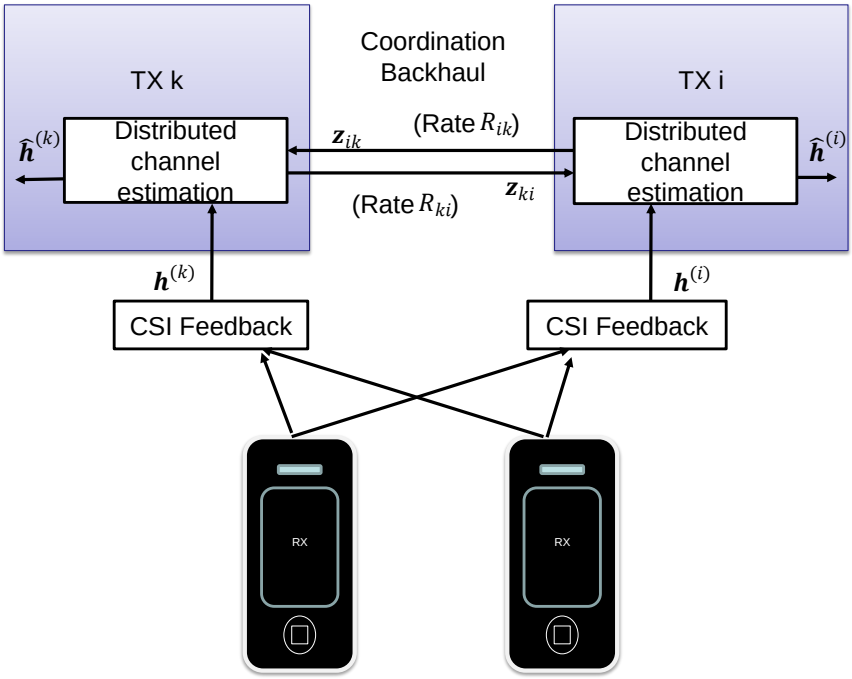


Figure 5.1: Decentralized cooperative channel estimation across two TXs serving two terminals.

version of initial CSI estimate $\boldsymbol{h}^{(k)}$, denoted by $\mathbf{z}_{ki}$. Similar operation is performed by TX i sending $\mathbf{z}_{ik}$ to TX k. The quantization operation associated to the link from TX k to TX i is defined as $\mathcal{Q}_{ki} : \mathbb{C}^{nKN_{RX}M_{TX}\times 1} \mapsto \mathcal{C}_{ki}$, $\mathbf{z}_{ki} \in \mathcal{C}_{ki}$, $\mid \mathcal{C}_{ki} \mid = 2^{R_{ki}}$ where $\mathcal{C}_{ki}$ is the codebook for the quantizer $\mathcal{Q}_{ki}$.

5.1.3 Channel estimation with limited coordination

At TX i, a reconstruction function $g_i(.)$ combines the initial CSI $\boldsymbol{h}^{(i)}$ and the exchanged CSI $\mathbf{z}_{ki}$ to form a final estimate $\hat{\boldsymbol{h}}^{(i)}$.

The MMSE estimation problem at TX i can be formulated as follows:

$$D^{(i)} = \min \mathrm{E}\left[\| \boldsymbol{h} - \hat{\boldsymbol{h}}^{(i)} \|^2\right] \tag{5.2}$$

$$= \min_{g_i, \mathcal{Q}_{ki}} \mathrm{E}\left[\| \boldsymbol{h} - g_i(\boldsymbol{h}^{(i)}, \mathcal{Q}_{ki}(\boldsymbol{h}^{(k)})) \|^2\right], \tag{5.3}$$

where $\hat{\boldsymbol{h}}^{(i)} = g_i(\boldsymbol{h}^{(i)}, \mathcal{Q}_{ki}(\boldsymbol{h}^{(k)}))$.

Note that it is in general a difficult problem since functional optimization is hard and the two functions $g_i(.)$ and $\mathcal{Q}_{ki}$ are intertwined.

The goal of this work is to find (i) a suitable reconstruction function $g_i(.)$ and (ii) the optimal quantizer $\mathcal{Q}_{ki}$ such that at TX i, $D^{(i)}$ is minimized.

5.2 Optimal Vector Quantization Model

We now first introduce a useful model for the optimal vector quantization (VQ) that will be used in the quantizer and reconstruction function design.

Optimal VQ can usually be derived via a Lloyd-Max algorithm as depicted in Fig. 5.2. The quantization result $\mathbf{z}_{ki}$ and quantization error $\mathbf{e}_{\mathcal{Q}_{ki}}$ are uncorrelated but the quantization input $\boldsymbol{h}^{(k)}$ is both dependent on $\mathbf{e}_{\mathcal{Q}_{ki}}$ and $\mathbf{z}_{ki}$. The covariance matrices for $\boldsymbol{h}^{(k)}$, $\mathbf{z}_{ki}$ and $\mathbf{e}_{\mathcal{Q}_{ki}}$ satisfy [65]

$$\mathbf{Q}_{\boldsymbol{h}^{(k)}} = \mathbf{Q}_{\mathbf{z}_{ki}} + \mathbf{Q}_{\mathcal{Q}_{ki}}. \tag{5.4}$$

Since the input of the quantizer $\boldsymbol{h}^{(k)}$ is Gaussian, we obtain an upper bound of the impact of quantization by assuming that the quantization error reads

$$\mathbf{e}_{\mathcal{Q}_{ki}} = \boldsymbol{h}^{(k)} - \mathbf{z}_{ki}, \tag{5.5}$$

which is also Gaussian distributed as $\mathbf{e}_{\mathcal{Q}_{ki}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{Q}_{\mathcal{Q}_{ki}})$ [66]. Similar to [13] and based on the Gaussian assumption for $\mathbf{e}_{\mathcal{Q}_{ki}}$, we can approximate the VQ procedure by a gain-plus-additive-noise model (similar to the scalar quantizer case in [67]) as illustrated in Fig.5.3.

Proposition 1. *Assume the quantization error $\mathbf{e}_{\mathcal{Q}_{ki}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{Q}_{\mathcal{Q}_{ki}})$ and is independent from the quantization result $\mathbf{z}_{ki}$, the optimal vector quantization for $\boldsymbol{h}^{(k)} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_k)$ is given by a gain-plus-additive-noise model:*

$$\mathbf{z}_{ki} = (\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_k - \mathbf{Q}_{\mathcal{Q}_{ki}})(\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_k)^{-1}\boldsymbol{h}^{(k)} + \mathbf{q}_{ki}, \tag{5.6}$$

where $\mathbf{q}_{ki}$ and $\boldsymbol{h}^{(k)}$ are uncorrelated random vectors, $\mathbf{q}_{ki} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, (\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_k - \mathbf{Q}_{\mathcal{Q}_{ki}})(\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_k)^{-1}\mathbf{Q}_{\mathcal{Q}_{ki}})$.

Proof. Since $\mathbf{e}_{\mathcal{Q}_{ki}}$ is assumed to be independent from $\mathbf{z}_{ki}$. Knowing that $\boldsymbol{h}^{(k)} = \mathbf{e}_{\mathcal{Q}_{ki}} + \mathbf{z}_{ki}$, $\mathbf{e}_{\mathcal{Q}_{ki}} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{Q}_{\mathcal{Q}_{ki}})$, $\boldsymbol{h} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{Q}_{\boldsymbol{h}})$, according to the Bayesian estimator [68],

$$\begin{aligned} \mathrm{E}\left[\mathbf{z}_{ki}|\boldsymbol{h}^{(k)}\right] &= (\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_k - \mathbf{Q}_{\mathcal{Q}_{ki}})(\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_k)^{-1}\boldsymbol{h}^{(k)} \\ \mathrm{cov}\left[\mathbf{z}_{ki}|\boldsymbol{h}^{(k)}\right] &= (\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_k - \mathbf{Q}_{\mathcal{Q}_{ki}})(\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_k)^{-1}\mathbf{Q}_{\mathcal{Q}_{ki}}, \end{aligned}$$

which concludes the proof. □

This gain-plus-additive noise model with uncorrelated $\mathbf{z}_{ki}$ and $\mathbf{q}_{ki}$ is helpful in the following derivation. In the reminder of this work, both the design of reconstruction functions and optimal quantizers will be based on (5.6).

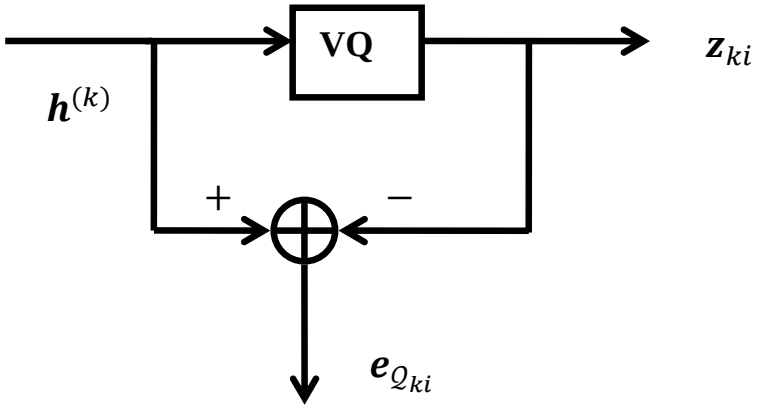


Figure 5.2: Quantizer model for optimal vector quantization.

5.3 Reconstruction Function Design

This section addresses the aforementioned sub-problem of the optimal reconstruction function design in general settings of multi-TXs cooperation ($n \geq 2$). For ease of illustration, we focus on TX i, who is cooperating with TX $k, \forall k \in \mathcal{A}_i$, where the set $\mathcal{A}_i$ contains the indices of TXs that are cooperating with TX i.

We consider hereby the reconstruction function as a weighted linear combination of estimates at TX i, which is suboptimal in a general setting (yet optimal in the particular case of $n = 2$) but leads to a desired closed form optimization.

Hence, the final estimate at TX i is modeled as:

$$\hat{\boldsymbol{h}}^{(i)} = \sum_{k \in \mathcal{A}_i} \mathbf{W}_{ki} \mathbf{z}_{ki} + \mathbf{W}_{ii} \boldsymbol{h}^{(i)}, \tag{5.7}$$

where $\mathbf{W}_{ki}, \mathbf{W}_{ii}$ are weighting matrices. The optimal weight combining matrices are revealed now in the following.

Proposition 2. *Consider a multi-transmitters cooperation described in* (5.7)*, assume the CSI estimate at each TX is distributed according to section 5.1.1 and the limited rate coordination is modeled according to section 5.1.2, let the quantization error covariance matrices be denoted as* $\mathbf{Q}_{\mathcal{Q}_{ki}}, \forall k \in \mathcal{A}_i$*. The optimum per dimensional MSE for the final estimate at TX i is:*

$$D^{(i)opt} = tr\left(\mathbf{Q}_{\boldsymbol{h}}^{-1} + \mathbf{Q}_i^{-1} + \mathbf{\Lambda}_i\right)^{-1}, \tag{5.8}$$

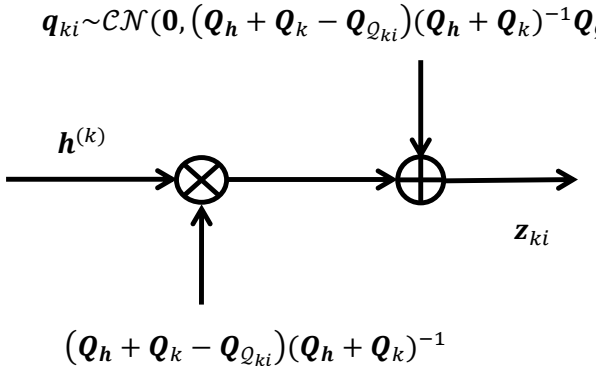


Figure 5.3: Gain-plus-additive-noise model for the optimal vector quantization procedure.

where $\mathbf{\Lambda}_i$ *is defined as:*

$$\mathbf{\Lambda}_i = \sum_{k \in \mathcal{A}_i} \left((\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_k)(\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_k - \mathbf{Q}_{\mathcal{Q}_{ki}})^{-1}(\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_k) - \mathbf{Q}_{\boldsymbol{h}} \right)^{-1}.$$

The optimal weight combing matrices $\{\mathbf{W}_{ki}^{opt}, \mathbf{W}_{ii}^{opt}, \forall k \in \mathcal{A}_i\}$ *are obtained as*

$$\left[\begin{array}{cccc} \mathbf{W}_{ii}^{opt} & \mathbf{W}_{l_1 i}^{opt} & \dots & \mathbf{W}_{l_{C_i} i}^{opt} \end{array} \right] = \mathbf{Q}_{\boldsymbol{h}} \mathbf{\Upsilon}_i \mathbf{\Omega}_i^{-1}, \tag{5.9}$$

where $\mathbf{\Upsilon}_i, \mathbf{\Omega}_i$ *are given below, the set* $\mathcal{A}_i$ *has cardinality* $|\mathcal{A}_i| = C_i$ *and each element in set is denoted by* $\mathcal{A}_i = \{l_1, \dots, l_{C_i}\}$.

$$\begin{array}{rcl} \mathbf{P}_{l_j i} & = & \mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_{l_j} - \mathbf{Q}_{\mathcal{Q}_{l_j i}} \\ \mathbf{A}_{l_j i} & = & \mathbf{P}_{l_j i}(\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_{l_j})^{-1}, \qquad \forall j = 1 \dots C_i \\ \mathbf{\Upsilon}_i & = & \left[\begin{array}{cccc} \mathbf{I} & \mathbf{A}_{l_1 i}^{\mathrm{H}} & \dots & \mathbf{A}_{l_{C_i} i}^{\mathrm{H}} \end{array} \right] \\ \mathbf{\Omega}_i & = & \left[\begin{array}{cccc} \mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_i & \mathbf{Q}_{\boldsymbol{h}} \mathbf{A}_{l_1 i}^{\mathrm{H}} & \dots & \mathbf{Q}_{\boldsymbol{h}} \mathbf{A}_{l_{C_i} i}^{\mathrm{H}} \\ \mathbf{A}_{l_1 i} \mathbf{Q}_{\boldsymbol{h}} & \mathbf{P}_{l_1 i} & \dots & \mathbf{A}_{l_1 i} \mathbf{Q}_{\boldsymbol{h}} \mathbf{A}_{l_{C_i} i}^{\mathrm{H}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_{l_{C_i} i} \mathbf{Q}_{\boldsymbol{h}} & \mathbf{A}_{l_{C_i} i} \mathbf{Q}_{\boldsymbol{h}} \mathbf{A}_{l_1 i}^{\mathrm{H}} & \dots & \mathbf{P}_{l_{C_i} i} \end{array} \right] \end{array}$$

Proof. Adopt the notation in Proposition 2, according to (5.2), (5.6) and

(5.7), the per dimensional MSE can be expressed as

$$\begin{aligned}
D^{(i)} &= \mathrm{E}\left[\| \boldsymbol{h} - \hat{\boldsymbol{h}}^{(i)} \|^2\right] \\
&= \mathrm{tr}\left((\sum_{k\in\mathcal{A}_i} \mathbf{W}_{ki}\mathbf{A}_{ki} + \mathbf{W}_{ii} - \mathbf{I}) \mathbf{Q}_{\boldsymbol{h}} (\sum_{k\in\mathcal{A}_i} \mathbf{W}_{ki}\mathbf{A}_{ki} + \mathbf{W}_{ii} - \mathbf{I})^{\mathrm{H}} \right) \\
&+ \mathrm{tr}\left(\sum_{k\in\mathcal{A}_i} \mathbf{W}_{ki} \left(\mathbf{A}_{ki}\mathbf{Q}_k\mathbf{A}_{ki}^{\mathrm{H}} + \mathbf{A}_{ki}\mathbf{Q}_{\mathcal{Q}_{ki}} \right) \mathbf{W}_{ki}^{\mathrm{H}} + \mathbf{W}_{ii}\mathbf{Q}_i\mathbf{W}_{ii}^{\mathrm{H}} \right).
\end{aligned}$$

Take the partial derivatives and set them to zero:

$$\begin{aligned}
&\frac{\partial D^{(i)}}{\partial \mathbf{W}_{ii}^*} = 0 \\
&\frac{\partial D^{(i)}}{\partial \mathbf{W}_{ki}^*} = 0, \qquad \forall k \in \mathcal{A}_i,
\end{aligned}$$

which leads to:

$$\mathbf{W}_{ii} \left(\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_i \right) = \left(\mathbf{I} - \sum_{k\in\mathcal{A}_i} \mathbf{W}_{ki}\mathbf{A}_{ki} \right) \mathbf{Q}_{\boldsymbol{h}}$$

$$\mathbf{W}_{ki} \left(\mathbf{A}_{ki}\mathbf{Q}_{\boldsymbol{h}}\mathbf{A}_{ki}^{\mathrm{H}} + \mathbf{A}_{ki}\mathbf{Q}_k\mathbf{A}_{ki}^{\mathrm{H}} + \mathbf{A}_{ki}\mathbf{Q}_{\mathcal{Q}_{ki}} \right) = \left(\mathbf{I} - \sum_{\substack{t\in\mathcal{A}_i \\ t\neq k}} \mathbf{W}_{ti}\mathbf{A}_{ti} - \mathbf{W}_{ii} \right) \mathbf{Q}_{\boldsymbol{h}}\mathbf{A}_{ki}^{\mathrm{H}}.$$

Solve the above equation system, the optimal weight combing matrices $\{\mathbf{W}_{ki}^{opt}, \mathbf{W}_{ii}^{opt}, k \in \mathcal{A}_i\}$ can be derived as:

$$\left[\begin{array}{cccc} \mathbf{W}_{ii}^{opt} & \mathbf{W}_{l_1 i}^{opt} & \dots & \mathbf{W}_{l_{C_i} i}^{opt} \end{array} \right] = \mathbf{Q}_{\boldsymbol{h}} \boldsymbol{\Upsilon}_i \boldsymbol{\Omega}_i^{-1}.$$

Let

$$\mathbf{W} = \left[\begin{array}{cccc} \mathbf{W}_{ii}^{opt} & \mathbf{W}_{l_1 i}^{opt} & \dots & \mathbf{W}_{l_{C_i} i}^{opt} \end{array} \right],$$

$$\boldsymbol{\Theta}_i = \left[\begin{array}{cccc} \mathbf{Q}_i & 0 & \dots & 0 \\ 0 & \mathbf{P}_{l_1 i} - \mathbf{A}_{l_1 i}\mathbf{Q}_{\boldsymbol{h}}\mathbf{A}_{l_1 i}^{\mathrm{H}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{P}_{l_{C_i} i} - \mathbf{A}_{l_{C_i} i}\mathbf{Q}_{\boldsymbol{h}}\mathbf{A}_{l_{C_i} i}^{\mathrm{H}} \end{array} \right],$$

then

$$\mathbf{\Omega}_i = \mathbf{\Theta}_i + \mathbf{\Upsilon}_i^{\mathrm{H}} \mathbf{Q}_{\boldsymbol{h}} \mathbf{\Upsilon}_i.$$

Since

$$\mathbf{W} = \mathbf{Q}_{\boldsymbol{h}} \mathbf{\Upsilon}_i \mathbf{\Omega}_i^{-1},$$

the optimum per dimensional MSE satisfies

$$\begin{aligned} D^{(i)opt} &= \mathrm{tr}\big((\mathbf{W}\mathbf{\Upsilon}_i^{\mathrm{H}} - \mathbf{I})\mathbf{Q}_{\boldsymbol{h}}(\mathbf{W}\mathbf{\Upsilon}_i^{\mathrm{H}} - \mathbf{I})^{\mathrm{H}} + \mathbf{W}\mathbf{\Theta}_i\mathbf{W}^{\mathrm{H}}\big) \\ &= \mathrm{tr}\big(\mathbf{Q}_{\boldsymbol{h}} - \mathbf{Q}_{\boldsymbol{h}}\mathbf{\Upsilon}_i\mathbf{\Omega}_i^{-1}\mathbf{\Upsilon}_i^{\mathrm{H}}\mathbf{Q}_{\boldsymbol{h}}\big) \\ &\overset{(a)}{=} \mathrm{tr}\big(\mathbf{Q}_{\boldsymbol{h}}^{-1} + \mathbf{\Upsilon}_i\mathbf{\Theta}_i^{-1}\mathbf{\Upsilon}_i^{\mathrm{H}}\big)^{-1} \\ &= \mathrm{tr}\left(\mathbf{Q}_{\boldsymbol{h}}^{-1} + \sum_{k\in\mathcal{A}_i} \mathbf{A}_{ki}^{\mathrm{H}}(\mathbf{P}_{ki} - \mathbf{A}_{ki}\mathbf{Q}_{\boldsymbol{h}}\mathbf{A}_{ki}^{\mathrm{H}})^{-1}\mathbf{A}_{ki} + \mathbf{Q}_i^{-1}\right)^{-1} \\ &= \mathrm{tr}\big(\mathbf{Q}_{\boldsymbol{h}}^{-1} + \mathbf{Q}_i^{-1} + \mathbf{\Lambda}_i\big)^{-1}, \end{aligned}$$

where (a) follows from the Woodbury identity

$$(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{C}^{\mathrm{H}})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}(\mathbf{B}^{-1} + \mathbf{C}^{\mathrm{H}}\mathbf{A}^{-1}\mathbf{C})^{-1}\mathbf{C}^{\mathrm{H}}\mathbf{A}^{-1}.$$

This concludes the proof. □

Remark 1. $\{\mathbf{W}_{ki}^{opt}, \mathbf{W}_{ii}^{opt}, \forall k \in \mathcal{A}_i\}$ and $D^{(i)opt}$ are merely functions of statistics $\mathbf{Q}_{\boldsymbol{h}}, \mathbf{Q}_i, \mathbf{Q}_k, \mathbf{Q}_{\mathcal{Q}_{ki}}, \forall k \in \mathcal{A}_i$. □

Remark 2. Consider a motivation example of two TXs cooperation, at TX 1, the final estimate is

$$\hat{\boldsymbol{h}}^{(1)} = \mathbf{W}_{21}\mathbf{z}_{21} + \mathbf{W}_{11}\boldsymbol{h}^{(1)},$$

where

$$\begin{aligned} \mathbf{P}_{21} &= \mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_2 - \mathbf{Q}_{\mathcal{Q}_{21}} \\ \mathbf{A}_{21} &= \mathbf{P}_{21}(\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_2)^{-1} \\ [\mathbf{W}_{11}, \mathbf{W}_{21}] &= \mathbf{Q}_{\boldsymbol{h}} \begin{bmatrix} \mathbf{I} & \mathbf{A}_{21}^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_1 & \mathbf{Q}_{\boldsymbol{h}}\mathbf{A}_{21}^{\mathrm{H}} \\ \mathbf{A}_{21}\mathbf{Q}_{\boldsymbol{h}} & \mathbf{P}_{21} \end{bmatrix}^{-1}. \end{aligned}$$

The optimal per dimensional MSE is

$$\begin{aligned} D^{(1)opt} &= \mathrm{tr}\left(\mathbf{Q}_{\boldsymbol{h}}^{-1} + \mathbf{Q}_1^{-1} + \mathbf{\Lambda}_1\right)^{-1} \\ \mathbf{\Lambda}_1 &= \left((\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_2)\mathbf{P}_{21}^{-1}(\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_2) - \mathbf{Q}_{\boldsymbol{h}}\right)^{-1}. \end{aligned}$$

□

Remark 3. The optimal per dimensional MSE and the error covariance matrix for the final estimate is related to 3 covariance terms: $\mathbf{Q}_{\boldsymbol{h}}$ indicates the intrinsic (true) channel statistics, $\mathbf{Q}_i$ refers to the initial estimation error covariance and $\mathbf{\Lambda}_i$ is related to the initial estimation error and the quantization error covariance at all TXs that cooperate with TX i. The covariance of the final estimate is formulated as the inverse of the sum of the 3 terms' individual inverse. □

5.4 Quantizer Design

For the optimal quantizer design, it should be noticed that a conventional optimal VQ (optimal VQ with MSE distortion) implemented by Lloyd-Max algorithm is far from being optimal because rather than minimizing the per dimensional MSE $D^{(i)}$, it only guarantees that the quantization distortion will be minimized.

Therefore, the quantizer should be properly shaped such that the quantization procedure ensures not only the minimization of quantization distortion, but also guarantees that the quantization result, after weighted combination with other estimates, will have the minimal per dimensional MSE $D^{(i)}$. A useful interpretation of this approach is as follows. As $\mathbf{Q}_i, \forall i = 1 \dots K$ reflect the spatial distribution of accuracy of the initial CSI, the quantizer $\mathcal{Q}_{ki}$ should allocate the quantization resource where more bits are needed, i.e. in channel elements or directions that are well known by TX k and least known by TX i. To this end, we use the weighted square error distortion:

$$d_{\mathcal{Q}_{ki}}(\mathbf{x}, \mathbf{y}) = (\mathbf{x} - \mathbf{y})^{\mathrm{H}} \mathbf{B}_{ki} (\mathbf{x} - \mathbf{y}), \tag{5.10}$$

as the distortion measure of the quantizer $\mathcal{Q}_{ki}$ with the positive definite shaping matrix $\mathbf{B}_{ki}$ to be optimized.

As an important intermediate step, we can calculate $\mathbf{Q}_{\mathcal{Q}_{ki}}$ in the asymptotic case as a function of $\mathbf{B}_{ki}$ when the given coordination link rate R_{ki} is sufficiently large, i.e. in the high resolution regime.

Proposition 3. *Consider quantization on a ℓ dimensional complex random vector source $\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{\Gamma})$, in the high resolution regime where the number of quantization level S is large, the optimal VQ using weighted MSE distortion with shaping matrix $\mathbf{B}$ will have a quantization error covariance matrix $\mathbf{Q}_{\mathbf{x}}$ as:*

$$\mathbf{Q}_{\mathbf{x}} = \mathbf{Q}_0^{(S)}(\mathbf{\Gamma}) \det(\mathbf{B})^{\frac{1}{\ell}} \mathbf{B}^{-1},$$

where

$$\mathbf{Q}_0^{(S)}(\mathbf{\Gamma}) = S^{-\frac{1}{\ell}} M_{2\ell} 2\pi \left(\frac{\ell+1}{\ell}\right)^{\ell+1} \det(\mathbf{\Gamma})^{\frac{1}{\ell}} \mathbf{I}_\ell$$

is the quantization error covariance matrix for $\mathbf{x}$ *in the high resolution regime when conventional optimal VQ is applied [69].* $M_{2\ell}$ *is a constant related to* 2ℓ*.*

Proof. Thus, according to Lemma 1,

$$\begin{aligned}\mathbf{Q}_{\mathbf{x}} &= \mathrm{E}\left[(\mathbf{x}-\mathcal{Q}(\mathbf{x}))(\mathbf{x}-\mathcal{Q}(\mathbf{x}))^{\mathrm{H}}\right] \\ &= \mathbf{B}^{-\frac{1}{2}} \mathrm{E}\left[(\mathbf{y}-\mathcal{Q}(\mathbf{y}))(\mathbf{y}-\mathcal{Q}(\mathbf{y}))^{\mathrm{H}}\right] \mathbf{B}^{-\frac{H}{2}} \\ &= \mathbf{B}^{-\frac{1}{2}} \mathbf{Q}_{\mathbf{y}} \mathbf{B}^{-\frac{H}{2}},\end{aligned}$$

where $\mathbf{y} = \mathbf{B}^{\frac{1}{2}}\mathbf{x} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{B}^{\frac{1}{2}}\mathbf{\Gamma}\mathbf{B}^{\frac{H}{2}})$. Let $\mathbf{t} = [\Re(\mathbf{y})^{\mathrm{T}} \Im(\mathbf{y})^{\mathrm{T}}]^{\mathrm{T}}$, then $\mathbf{t} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Phi})$ and

$$\mathbf{\Phi} = \frac{1}{2}\begin{bmatrix} \Re(\mathbf{B}^{\frac{1}{2}}\mathbf{\Gamma}\mathbf{B}^{\frac{H}{2}}) & \Im(\mathbf{B}^{\frac{1}{2}}\mathbf{\Gamma}\mathbf{B}^{\frac{H}{2}}) \\ \Im(\mathbf{B}^{\frac{1}{2}}\mathbf{\Gamma}\mathbf{B}^{\frac{H}{2}}) & \Re(\mathbf{B}^{\frac{1}{2}}\mathbf{\Gamma}\mathbf{B}^{\frac{H}{2}}) \end{bmatrix}.$$

Furthermore, it is proven in [70] that

$$\mathbf{Q}_{\mathbf{t}} = \mathrm{E}\left[(\mathbf{t}-\mathcal{Q}(\mathbf{t}))(\mathbf{t}-\mathcal{Q}(\mathbf{t}))^{\mathrm{T}}\right] = D_{\mathbf{t}}\mathbf{I}_\ell,$$

where the average distortion $D_{\mathbf{t}} = \frac{1}{\ell}\mathrm{tr}(\mathbf{Q}_{\mathbf{t}})$ is obtained from [**?**] for large S as

$$\begin{aligned} D_{\mathbf{t}} &= S^{-\frac{2}{\ell}} M_\ell \left(\int f_{\mathbf{t}}(\mathbf{t})^{\frac{\ell}{\ell+2}} d\mathbf{t}\right)^{\frac{\ell+2}{\ell}} \\ &= S^{-\frac{2}{\ell}} M_\ell 2\pi \left(\frac{\ell+2}{\ell}\right)^{\frac{\ell}{2}+1} \det(\mathbf{\Phi})^{\frac{1}{\ell}}, \end{aligned}$$

where $f_{\mathbf{t}}(.)$ is the probability density function (p.d.f) of $\mathbf{t}$. Finally, from the expression of $\mathbf{Q}_{\mathbf{t}}$ and by a real-to-complex conversion, we get

$$\mathbf{Q}_{\mathbf{y}} = 2S^{-\frac{1}{\ell}} M_{2\ell} 2\pi \left(\frac{\ell+1}{\ell}\right)^{\ell+1} \det(\mathbf{\Phi})^{\frac{1}{2\ell}} \mathbf{I}_\ell,$$

which leads to

$$\begin{aligned}\mathbf{Q}_{\mathbf{x}} &= 2S^{-\frac{1}{\ell}} M_{2\ell} 2\pi \left(\frac{\ell+1}{\ell}\right)^{\ell+1} \det(\mathbf{\Phi})^{\frac{1}{2\ell}} \mathbf{B}^{-1} \\ &= S^{-\frac{1}{\ell}} M_{2\ell} 2\pi \left(\frac{\ell+1}{\ell}\right)^{\ell+1} \det(\mathbf{B})^{\frac{1}{\ell}} \det(\mathbf{\Gamma})^{\frac{1}{\ell}} \mathbf{B}^{-1} \\ &= \mathbf{Q}_0^{(S)}(\mathbf{\Gamma}) \det(\mathbf{B})^{\frac{1}{\ell}} \mathbf{B}_{ki}^{-1},\end{aligned}$$

where

$$\mathbf{Q}_0^{(S)}(\mathbf{\Gamma}) = S^{-\frac{1}{\ell}} M_{2\ell} 2\pi \left(\frac{\ell+1}{\ell}\right)^{\ell+1} \det(\mathbf{\Gamma})^{\frac{1}{\ell}} \mathbf{I}_\ell.$$

□

Remark 4. The aforementioned quantization error covariance matrix expression encompasses the quantization error covariance matrix for conventional optimal VQ by taking $\mathbf{B} = \mathbf{I}$. It can be easily verified that by imposing a constraint that $\det(\mathbf{B}) = 1$, for all value of matrix $\mathbf{B}$, the corresponding quantizers will have the same quantization distortion. □

Remark 5. For the constant $M_{2\ell}$, a look-up table for $2\ell = 1, \ldots, 10$ in [71] can be used. when ℓ is larger, we can approximate $M_{2\ell} = \frac{1}{2\pi e}$. □

We now exploit Proposition 3 in order to derive $\mathbf{Q}_{\mathcal{Q}_{ki}}$:

$$\mathbf{Q}_{\mathcal{Q}_{ki}} = \mathbf{Q}_0^{(S)}(\mathbf{\Gamma}) \det(\mathbf{B}_{ki})^{\frac{1}{\ell}} \mathbf{B}_{ki}^{-1}, \tag{5.11}$$

where

$$\begin{aligned} S &= 2^{R_{ki}} \\ \mathbf{\Gamma} &= \mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_k. \end{aligned}$$

5.4.1 Shaping matrix optimization

Based on the reconstruction function in Section 5.3 and using equations (5.8), (5.11), we can now proceed with the task of jointly optimizing the reconstruction function and the quantizer by solely optimizing the value of $\mathbf{B}_{ki}$:

$$\begin{array}{ll} \min\limits_{\mathbf{B}_{ki}, k \in \mathcal{A}_i} & D^{(i)opt} \\ s.t. & \det(\mathbf{B}_{ki}) = 1, \mathbf{B}_{ki} \succeq 0 \\ & D^{(i)opt} \text{ defined in (5.8)} \end{array} \quad . \tag{5.12}$$

As is mentioned in Remark 4, the constraints on $\mathbf{B}_{ki}$ matrices ensure that all feasible quantizers have the same quantization distortion, the optimization will find $\mathbf{B}_{ki}, \forall k \in \mathcal{A}_i$ that minimize the per dimensional MSE for the final estimate.

We demonstrate that the objective function for Problem (5.12) is in fact a convex function for $\mathbf{B}_{ki}$. However, solving the problem is still difficult due to the complex formulation for the objective function. Therefore, we simplify the objective function based on the high resolution assumption and introduce a convex optimization problem (5.13), which can be easily transformed into a semi-definite quadratic linear programming.

Proposition 4. The objective function in problem (5.12) is convex. In high resolution regime, problem (5.12) can be approximated by the following convex optimization problem:

$$\begin{array}{ll} \min\limits_{\mathbf{B}_{ki}, k\in\mathcal{A}_i} & \mathrm{tr}\left(\sum\limits_{k\in\mathcal{A}_i}(\mathbf{Q}_k^{-1}-\mathbf{Q}_k^{-1}\mathbf{Q}_{\mathcal{Q}_{ki}}\mathbf{Q}_k^{-1})+\mathbf{Q}_i^{-1}+\mathbf{Q}_{\boldsymbol{h}}^{-1}\right)^{-1} \\ s.t. & \det(\mathbf{B}_{ki})\geq 1, \mathbf{B}_{ki}\succeq 0 \\ & \mathbf{Q}_{\mathcal{Q}_{ki}} \text{ defined in (5.11)} \end{array}. \tag{5.13}$$

Proof. Adopt the notations in Proposition 2, we will prove the convexity of problem (5.12). Since

$$\det(\mathbf{B}_{ki}) = 1,$$

according to (5.11),

$$\mathbf{Q}_{\mathcal{Q}_{ki}} = \mathbf{Q}_0^{(S)}(\boldsymbol{\Gamma})\mathbf{B}_{ki}^{-1}.$$

Let

$$f = \mathbf{Q}_{\boldsymbol{h}}^{-1} + \mathbf{Q}_i^{-1} + \boldsymbol{\Lambda}_i,$$

with $\boldsymbol{\Lambda}_i$ defined in Corollary 2. Consider the convexity for composite function: if function h is concave and matrix monotone increasing, and function g is concave, the composite function $h \circ g$ is concave. According to Lemma 2,

$$h = -\mathbf{X}^{-1}$$

is matrix concave and matrix monotone increasing for $\mathbf{X}$,

$$g = \mathbf{A} - \mathbf{B}_{ki}^{-1}$$

with a constant matrix $\mathbf{A}$ is matrix concave for $\mathbf{B}_{ki}$. Therefore the function

$$\boldsymbol{\Lambda}_i = h \circ g$$

is concave for $\mathbf{B}_{ki}$ and $f = \mathbf{Q}_{\boldsymbol{h}}^{-1} + \mathbf{Q}_i^{-1} + \boldsymbol{\Lambda}_i$ is also concave for $\mathbf{B}_{ki}$.

Consider the convexity for composite function: if function h is convex and matrix monotone decreasing, and function g is concave, the composite function $h \circ g$ is convex. According to Lemma 3,

$$h = \mathrm{tr}(\mathbf{X})^{-1}$$

is convex and matrix monotone decreasing for $\mathbf{X}$, since

$$g = \mathbf{Q}_{\boldsymbol{h}}^{-1} + \mathbf{Q}_i^{-1} + \mathbf{\Lambda}_i$$

is concave for $\mathbf{B}_{ki}$, thus

$$D^{(i)opt} = h \circ g$$

is convex for $\mathbf{B}_{ki}$.

When the coordination link rate R_{ki} is sufficiently large, use twice matrix inverse approximation:

$$\begin{aligned} D^{(i)opt} &= \mathrm{tr}\left(\mathbf{Q}_{\boldsymbol{h}}^{-1}+\mathbf{Q}_i^{-1}+\sum_{k\in\mathcal{A}_i}\left((\mathbf{Q}_{\boldsymbol{h}}+\mathbf{Q}_k)(\mathbf{Q}_{\boldsymbol{h}}+\mathbf{Q}_k-\mathbf{Q}_{\mathcal{Q}_{ki}})^{-1}(\mathbf{Q}_{\boldsymbol{h}}+\mathbf{Q}_k)-\mathbf{Q}_{\boldsymbol{h}}\right)^{-1}\right)^{-1} \\ &\overset{(a)}{\simeq} \mathrm{tr}\left(\sum_{k\in\mathcal{A}_i}(\mathbf{Q}_k+\mathbf{Q}_{\mathcal{Q}_{ki}})^{-1}+\mathbf{Q}_i^{-1}+\mathbf{Q}_{\boldsymbol{h}}^{-1}\right)^{-1} \\ &\overset{(b)}{\simeq} \mathrm{tr}\left(\sum_{k\in\mathcal{A}_i}(\mathbf{Q}_k^{-1}-\mathbf{Q}_k^{-1}\mathbf{Q}_{\mathcal{Q}_{ki}}\mathbf{Q}_k^{-1})+\mathbf{Q}_i^{-1}+\mathbf{Q}_{\boldsymbol{h}}^{-1}\right)^{-1}, \end{aligned}$$

where $(a), (b)$ follows from the matrix inverse first order approximation: if $\mathbf{X}^m \to 0$, when $m \to \infty$, then

$$(\mathbf{A}+\mathbf{X})^{-1} \simeq \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{X}\mathbf{A}^{-1}.$$

According to Lemma 3, Lemma 2 and convexity for composite matrix function, the expression above is convex.

If we relax the constraint $\det(\mathbf{B}_{ki}) = 1$ to $\det(\mathbf{B}_{ki}) \geq 1$, the feasible set for $\mathbf{B}_{ki}$ is therefore a convex set and the optimization for this relaxed problem is minimizing a convex function over a convex set, which is a convex optimization problem and the optimal $\mathbf{B}_{ki}$ should attain at the boundary $\det(\mathbf{B}_{ki}) = 1$. This concludes the proof. □

Problem (5.13) can be solved efficiently by optimization toolbox such as CVX. Note that once the optimal weight matrix $\mathbf{B}_{ki}^*$ is obtained, the codebook for optimal quantizer $\mathcal{Q}_{ki}^*$ can be calculated based on Lloyd algorithm and a training set. The optimal weight matrices for estimation combine can be calculated according to (5.9). Noticing that this optimization is semi-static, the weight matrices for the estimates combination and the shaping

matrices for the quantizers will be updated only when the channel statistics or the backhaul resources have been changed.

Interestingly, the asymptotic performance of the proposed algorithm can be characterized in relation to known information theoretic bound.

Proposition 5. For a two TXs cooperation scenario of TX 1 and TX 2 as described in Section 5.1.3, the proposed coordination shaping algorithm can achieve asymptotically in high resolution regime the Wyner-Ziv bound given by

$$D_{\infty}^{(1)opt} = \operatorname{tr}\left(\mathbf{Q}_1^{-1} + \mathbf{Q}_2^{-1} + \mathbf{Q}_{\boldsymbol{h}}^{-1}\right)^{-1} \tag{5.14}$$

at TX 1.

Proof. The asymptotical MSE for proposed algorithm is:

$$\begin{aligned} D_{\infty}^{(1)opt} &= \lim_{R_{21}\to\infty} D^{(1)opt} \\ &= \operatorname{tr}\left(\mathbf{Q}_1^{-1} + \mathbf{Q}_2^{-1} + \mathbf{Q}_{\boldsymbol{h}}^{-1}\right)^{-1}. \end{aligned}$$

It is well known that the information theoretic bound of the per dimension MSE for two TXs cooperation can be achieved using a Wyner-Ziv quantizer and the asymptotic distortion is [72]:

$$D_{\infty}^{(1)NWZ} = \operatorname{tr}(\mathrm{E}_{YZ} \operatorname{var}[X|Y,Z]),$$

where X, Y, Z correspond respectively to the source data, side information and noisy source (i.e, perfect CSI $\boldsymbol{h}$, initial CSI $\boldsymbol{h}^{(1)}$ and initial CSI at it's cooperation TX $\boldsymbol{h}^{(2)}$ as in our case). Since Gaussianity is assumed for the perfect CSI and initial CSI, $D_{\infty}^{(1)NWZ}$ can be calculated as:

$$\begin{aligned} D_{\infty}^{(1)NWZ} =& \operatorname{tr}(\mathbf{Q}_{\boldsymbol{h}}) - \operatorname{tr}\left(\begin{bmatrix}\mathbf{Q}_{\boldsymbol{h}} & \mathbf{Q}_{\boldsymbol{h}}\end{bmatrix}\begin{bmatrix}\mathbf{Q}_{\boldsymbol{h}}+\mathbf{Q}_{\mathbf{1}} & \mathbf{Q}_{\boldsymbol{h}} \\ \mathbf{Q}_{\boldsymbol{h}} & \mathbf{Q}_{\boldsymbol{h}}+\mathbf{Q}_{\mathbf{2}}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{Q}_{\boldsymbol{h}} \\ \mathbf{Q}_{\boldsymbol{h}}\end{bmatrix}\right) \\ =& \operatorname{tr}\left(\mathbf{T}_1 - \mathbf{T}_1(\mathbf{T}_1+\mathbf{Q}_2)^{-1}\mathbf{T}_1\right) \\ =& \operatorname{tr}\left(\mathbf{T}_1^{-1}+\mathbf{Q}_2^{-1}\right)^{-1} \\ =& \operatorname{tr}\left(\mathbf{Q}_1^{-1}+\mathbf{Q}_2^{-1}+\mathbf{Q}_{\boldsymbol{h}}^{-1}\right)^{-1} \\ =& D_{\infty}^{(1)opt}, \end{aligned}$$

where

$$\mathbf{T}_1 = \mathbf{Q}_{\boldsymbol{h}}(\mathbf{Q}_{\boldsymbol{h}} + \mathbf{Q}_1)^{-1}\mathbf{Q}_1.$$

□

Algorithm 1 Iterative algorithm for problem (5.15)

1: Initialize $\mathbf{B}_{ki}, \forall k \in \mathcal{A}_i, \forall i = 1, \ldots, n$
2: **while** not converge **do**
3: Optimize R_{ki} for problem (5.15) with $\mathbf{B}_{ki}$ fixed
4: Optimize $\mathbf{B}_{ki}$ for problem (5.15) with R_{ki} fixed
5: **end while**

Thus, this result reveals that in two TXs cooperation case, the proposed coordination shaping algorithm is asymptotically optimal as backhaul increases.

5.5 Coordination Backhaul Resource Allocation

An interesting consequence of the above analysis is the optimization of coordination where multiple TXs can exchange simultaneously CSI-related information to each other under a global constraint on the coordination bits. The global optimization problem over all coordination links now becomes:

$$\begin{array}{cl} \min\limits_{\substack{\mathbf{B}_{ki}, R_{ki} \\ k\in\mathcal{A}_i, i=1,\ldots,n}} & \frac{1}{n}\sum\limits_{i=1}^{n} D^{(i)opt} \\ s.t. & \det(\mathbf{B}_{ki}) = 1, \mathbf{B}_{ki} \succeq 0 \\ & \sum\limits_{i=1}^{n}\sum\limits_{k\in\mathcal{A}_i} R_{ki} = R_{tot}, R_{ki} \in \mathbb{N}^{+} \\ & D^{(i)opt} \text{ defined in (5.8)} \end{array} . \tag{5.15}$$

Due to the integer constraints on R_{ki}, this problem becomes a non-convex optimization. However, conventional alternating algorithms can be applied to perform the optimization in a two-step iterative approach described in Algorithm 1.

The optimization in the first step is an integer programming problem [73] and the optimization in the second step is a convex optimization problem. Hence, many conventional algorithms can be applied in both steps [74]. It should be noted that the alternating algorithm does not guarantee the global optimum. Based on the initial point, it might converge to a local optimal point as well.

5.6 Balance the Accuracy and Consistency of Estimates

The final CSI estimate at each TX will be used for TX cooperation design. Therefore, estimate accuracy achieved by minimizing the estimation error at each TX is not the only important criteria. As is known, the CoMP system will end up with a distributed CSIT configuration after limited information exchange. Since many TX cooperation designs require a central processing unit which indicates that all the cooperating TX shares a unique imperfect CSI estimate, we should also maintain the consistency of the final CSI estimates at different TXs. Therefore, we introduce the following figure of merit which balances the accuracy and consistency of estimates with a balancing factor ρ defined as:

$$D^{bal} = \frac{\rho}{n}\sum_{i=1}^{n} D^{(i)} + \frac{1-\rho}{4n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}\left[\|\hat{\boldsymbol{h}}^{(i)} - \hat{\boldsymbol{h}}^{(j)}\|^2\right]. \tag{5.16}$$

It should be noticed that D^{bal} is always smaller than $D^{av} = \frac{1}{n}\sum_{i=1}^{n} D^{(i)}$ since

$$\begin{aligned} D^{bal} =& \frac{\rho}{n}\sum_{i=1}^{n} D^{(i)} + \frac{1-\rho}{4n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}\left[\|\hat{\boldsymbol{h}}^{(i)} - \hat{\boldsymbol{h}}^{(j)}\|^2\right] && (5.17)\\ \overset{(a)}{\leq}& \frac{\rho}{n}\sum_{i=1}^{n} D^{(i)} + \frac{1-\rho}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathrm{E}\left[\|\hat{\boldsymbol{h}}^{(i)} - \boldsymbol{h}\|^2 + \|\boldsymbol{h} - \hat{\boldsymbol{h}}^{(j)}\|^2\right] && (5.18)\\ =& \frac{1}{n}\sum_{i=1}^{n} D^{(i)} = D^{av}, && (5.19) \end{aligned}$$

where (a) is based on the inequality

$$\|\boldsymbol{a}-\boldsymbol{b}\|^2 \leq (\|\boldsymbol{a}-\boldsymbol{c}\| + \|\boldsymbol{c}-\boldsymbol{b}\|)^2 \leq 2\left(\|\boldsymbol{a}-\boldsymbol{c}\|^2 + \|\boldsymbol{c}-\boldsymbol{b}\|^2\right) \tag{5.20}$$

for any vector $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$. We can perform similar optimization for the quantizer and the reconstruct function based on D^{bal}.

5.7 Numerical Performance Analysis

In this section, a network MIMO transmission [1] setup is considered. Unless otherwise indicated, the default simulation settings are $n = 2$ and $K = 2$, $M_{TX} = N_{RX} = 1$. The channel $\boldsymbol{h} \in \mathbb{C}^{4\times 1} \sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}, \mathbf{I}_4)$ and the rates on

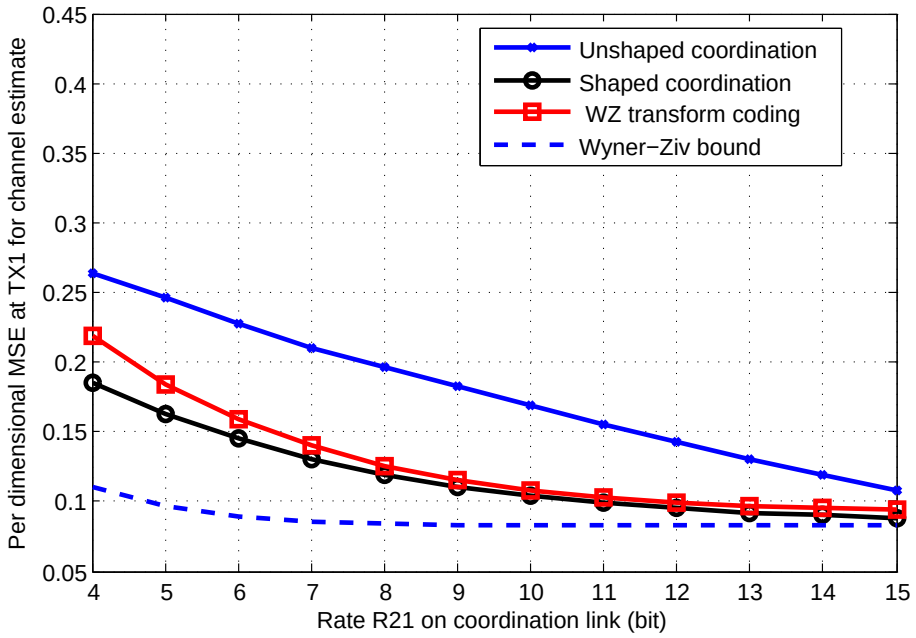


Figure 5.4: Per dimensional MSE for the final channel estimation at TX 1 vs coordination link rate R_{21}.

coordination link from TX 2 to TX 1 and from TX 1 to TX 2 are denoted R_{21} and R_{12}, respectively. Each TX constructs a ZF precoder based on its final channel estimate. The power control at each TX is 20dB. The per dimensional MSE for decentralized channel estimation is evaluated for different settings using Monte-Carlo simulations over 10^5 channel realizations. Since $\ell = nKM_{TX}N_{RX} = 4$, the parameter $M_{2\ell}$ is chosen to be 929/12960 which is related to the $E8$ lattice [71]. In Fig. 5.4, Fig. 5.6 and Fig. 5.8, the CSI information structure is characterized by

$$\mathbf{Q}_1 = \begin{pmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0.9 & 0 \\ 0 & 0 & 0 & 0.9 \end{pmatrix}, \mathbf{Q}_2 = \begin{pmatrix} 0.9 & 0 & 0 & 0 \\ 0 & 0.9 & 0 & 0 \\ 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.1 \end{pmatrix},$$

which corresponds to an example where TX 1 has more accurate CSI about RX1 and less accurate CSI about RX2, while TX 2 has more accurate CSI about RX2 and less accurate CSI about RX1. The $diag(.)$ operator represents a diagonal matrix with diagonal elements in the parenthesis.

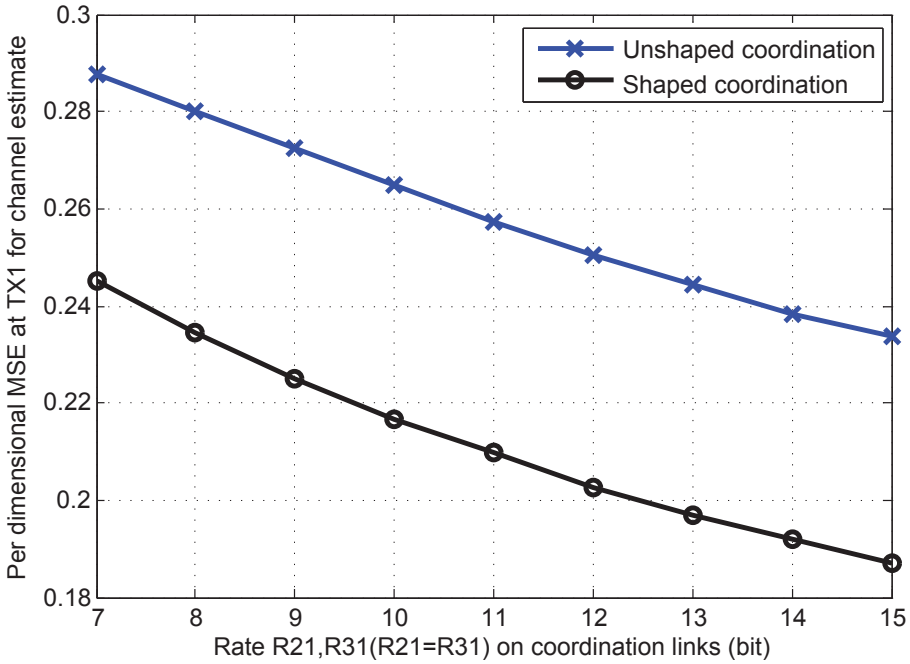


Figure 5.5: 3 TX cooperation: per dimensional MSE for the final channel estimation at TX 1 vs coordination link rate $R_{21}(= R_{31})$.

Fig.5.4 shows the per dimensional MSE for the final estimation at TX 1. The Wyner-Ziv bound is the information theoretic bound. The shaped coordination curve applies the proposed algorithm. The unshaped coordination implements the traditional optimal VQ and finds $\mathbf{W}_{21}, \mathbf{W}_{11}$ accordingly using (5.9). From the figure we can conclude that the shaped coordination algorithm outperforms the unshaped coordination algorithm, which not surprisingly shows the benefit of taking the priori statistic information into account. The WZ transform coding curve refers to the asymptotic optimal Wyner-Ziv transform coding for noisy source in [72]. It reveals that our algorithm outperforms the Wyner-Ziv transform coding algorithm in low coordination rate region and converges asymptotically to the Wyner-Ziv bound D_∞ as expected.

In Fig.5.5, the per dimensional MSE for the final estimation at TX 1 is plotted for a 3 TX cooperation network. The rate constraint on coordination link from TX 2 to TX 1 and TX 3 to TX 1 are R_{21} and R_{31} respectively. In

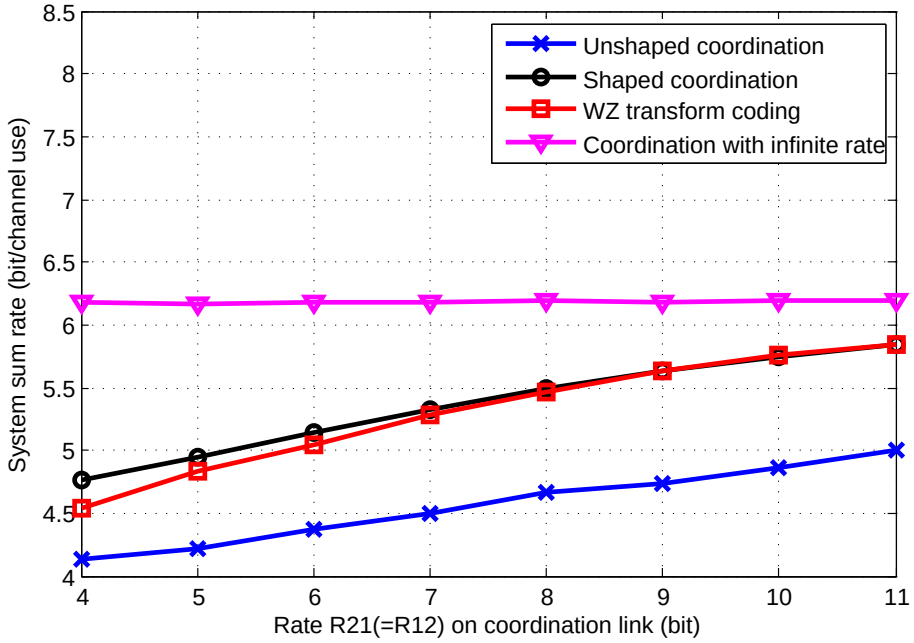


Figure 5.6: Sum rate for the 2 TX cooperation system vs coordination link rate $R_{21}(= R_{12})$, each TX implements a ZF precoder.

this simulation, the parameters are denoted below:

$$
\begin{aligned}
n =& K = 3, M_{TX} = N_{RX} = 1, \ell = nKM_{TX}N_{RX} = 9, M_{2\ell} = \frac{1}{2\pi e} \\
Q_1 =& \text{diag}(0.81, 0.91, 0.13, 0.91, 0.63, 0.10, 0.28, 0.55, 0.96) \\
Q_2 =& \text{diag}(0.96, 0.16, 0.97, 0.96, 0.49, 0.80, 0.14, 0.42, 0.92) \\
Q_3 =& \text{diag}(0.79, 0.96, 0.66, 0.04, 0.85, 0.93, 0.68, 0.76, 0.74).
\end{aligned}
$$

The simulation clearly shows the performance enhancement of the coordination shaping algorithm over the unshaped coordination algorithm in a multiple TX cooperation scenario.

Fig. 5.6 exhibits the sum rate for a 2 TX cooperation system. Each TX implements a ZF precoder based on its final CSI estimate, the power constraint is $P = 20$dB per TX. The rate on the coordination links satisfies $R_{21} = R_{12}$. We also provide the sum rate for the case when coordination links have infinite bandwidth. The figure shows that the proposed shaped coordination algorithm will improve the system sum rate beyond the unshaped coordination algorithm and WZ transform coding algorithm when a simple

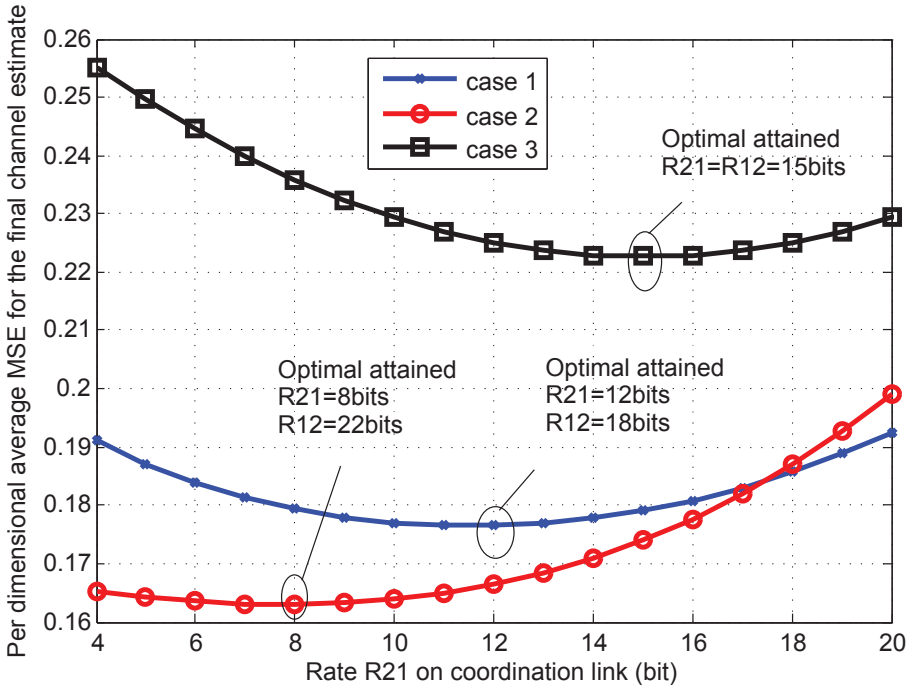


Figure 5.7: Per dimensional AMSE for the final channel estimation at TX 1, TX 2 for different backhaul resource allocation.

ZF precoder is implemented. As the rate on coordination link increases, the sum rate for all algorithms will converge to the infinite coordination rate case.

Fig.5.7 considers the coordination backhaul resource allocation problem for a 2 TX cooperation system. The total amount of bits for coordination link is $R_{12} + R_{21} = 30$bits. Three cases for the CSI estimate matrices are considered.

cases	$\mathbf{Q}_1$	$\mathbf{Q}_2$
case 1	$\mathrm{diag}(0.1, 0.1, 0.9, 0.9)$	$\mathrm{diag}(0.5, 0.5, 0.5, 0.5)$
case 2	$\mathrm{diag}(0.4, 0.2, 0.3, 0.1)$	$\mathrm{diag}(0.7, 0.8, 0.6, 0.9)$
case 3	$\mathrm{diag}(0.5, 0.5, 0.5, 0.5)$	$\mathrm{diag}(0.5, 0.5, 0.5, 0.5)$

Table 5.1: CSI estimate matrices at TX 1 and TX 2.

The figure reveals that the cooperation information exchange is not nec-

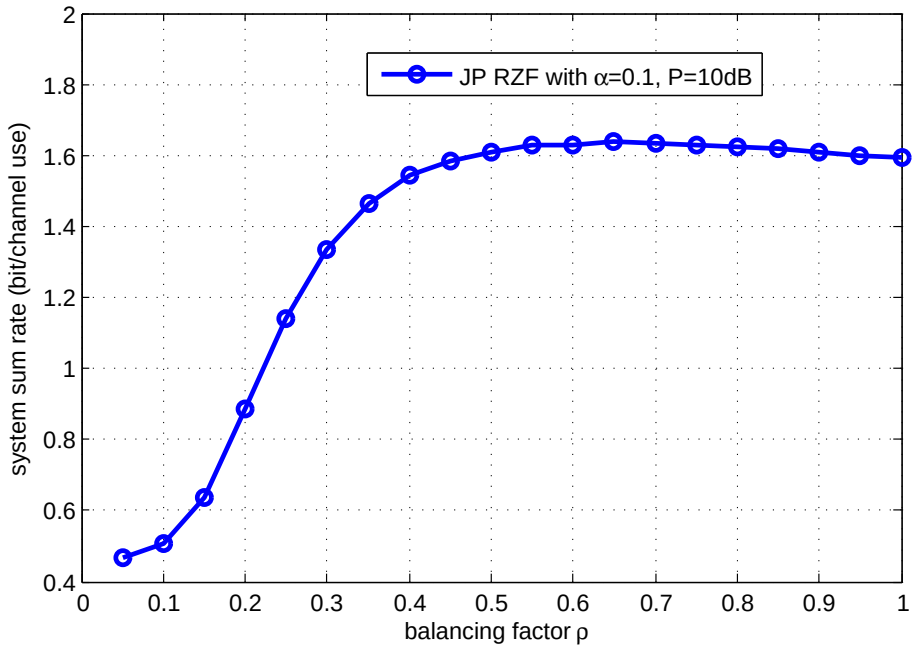


Figure 5.8: Sum rate for the 2 TX cooperation system vs the balancing factor ρ, each TX implements a RZF precoder.

essarily symmetric. In case 3, the two TXs exchange information with equal rate $R_{12} = R_{21} = 15$bits because the accuracy of CSI at both end is the same. However, in case 2, the optimal coordination link bit allocation strategy is to let TX 1 share the cooperation information to TX 2 through a $R_{21} = 8$bits coordination link and vice versa through a $R_{12} = 22$bits coordination link. It's intuitive because for every channel coefficient, TX 1 has a more accurate initial CSI than TX 2. It reveals that if one TX has a better CSI, it is more encouraged to share his information through a higher rate coordination link.

Fig. 5.8 reveals that balancing the accuracy and consistency of estimates will affect the system sum rate performance when a RZF precoder (see Chapter 6.1.3 for details) is implemented at each TX based on its final channel estimation. The regularization coefficient $\alpha = 0.1$, the power constraint is $P = 10$dB per TX, the coordination backhaul is $R_{21} = R_{12} = 9$bits. It demonstrates that if the decentralized transmitter strategy is fixed, we can balancing the accuracy and the consistency for the CSI estimates at different TXs in order to achieve higher system sum rate performance. We can

conclude that there is an optimal balancing factor. This balancing factor depends on the chosen TX cooperation strategy and the optimal value is not necessarily 1, which falls back to the per TX MMSE figure of merit.

5.8 Conclusion

We study the decentralized cooperative channel estimation for coordinated transmission with limited backhauling. We derive a low-complexity algorithm which exploits the finite-capacity backhaul near-optimally and is robust to arbitrary feedback noise statistics. We exhibit clear advantages over CSI acquisition and exploitation methods used in conventional CoMP systems.

Part III

Robust Decentralized Precoder Design under Limited Feedback and Backhaul

Chapter 6

Robust Regularized ZF in Cooperative Broadcast Channel under Distributed CSIT

In the previous part, we have studied the problem of efficient exchange design for the limited coordination backhaul. Thanks to the cooperative channel estimation algorithm proposed in Chapter 5, each TX can acquire a more accurate final CSI. In this part, we will analyze the problem of decentralized joint processing transmitter cooperation design based on the final CSI obtained at each TX. We target specifically on the large system analysis of a fixed family of joint processing techniques: the RZF precoder. Particularly, we focus on the question below:

"What is the robust decentralized RZF precoder design for the joint processing CoMP under distributed CSIT?"

6.1 System Model

6.1.1 Transmission model

We consider the system described in Section 3.1.1 with $N_{RX} = 1$. We assume that the ratio of transmit antennas with respect to the number of users is

fixed and given by

$$\beta = \frac{M}{K} \geq 1. \tag{6.1}$$

Assume the transmission noise has i.i.d entry $n_k \sim \mathcal{N}_{\mathbb{C}}(0,1), \forall k = 1, \ldots, K$. In this particular single RX antenna case, the notations introduced in Section 3.1.1 simplify as follows. The transmitted multi-user signal $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is obtained from the symbol vector $\boldsymbol{s} = [s_1, \ldots, s_K]^{\mathrm{T}} \in \mathbb{C}^{K\times 1}$:

$$\mathbf{x} = \mathbf{T}\boldsymbol{s} = \sum_{k=1}^{K} \boldsymbol{t}_k s_k \tag{6.2}$$

with $\mathbf{T} = \left[\boldsymbol{t}_1, \ldots, \boldsymbol{t}_K\right] \in \mathbb{C}^{M\times K}$ being the *multi-user* precoder, $\boldsymbol{t}_k \in \mathbb{C}^{M\times 1}$ being the beamforming vector for RX k. We consider an average sum power constraint

$$\mathrm{E}\left[\mathrm{tr}\left(\mathbf{T}\mathbf{T}^{\mathrm{H}}\right)\right] = P, \tag{6.3}$$

where P is the total transmit power for all TXs.

With the assumption of Gaussian signaling $s_k \sim \mathcal{N}_{\mathbb{C}}(0,1), \forall k$ and each user decoding with perfect CSIR, the signal-to-interference and noise ratio (SINR) at RX k is given by [75]

$$\mathrm{SINR}_k = \frac{\left|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{t}_k\right|^2}{1 + \sum\limits_{\ell=1,\ell\neq k}^{K} \left|\boldsymbol{h}_k^{\mathrm{H}}\boldsymbol{t}_\ell\right|^2}. \tag{6.4}$$

The ergodic sum rate for the CoMP network is defined as:

$$\mathrm{R}_{sum} = \sum_{k=1}^{K} \mathrm{E}\left[\log_2\left(1 + \mathrm{SINR}_k\right)\right], \tag{6.5}$$

where the expectation is taken over the random channel realizations.

6.1.2 Distributed CSI channel model

RX k's channel $\boldsymbol{h}_k$ is modeled as

$$\boldsymbol{h}_k = \sqrt{M}\boldsymbol{\Theta}_k^{\frac{1}{2}}\boldsymbol{z}_k, \tag{6.6}$$

where $\boldsymbol{\Theta}_k \in \mathbb{C}^{M\times M}$ is the channel correlation matrix of RX k and $\boldsymbol{z}_k$ has i.i.d complex entries of zero mean, variance $\frac{1}{M}$ and eighth order moment of order $O(\frac{1}{M^4})$. The channel correlation matrices $\boldsymbol{\Theta}_k, \forall k = 1, \ldots, K$ are

assumed to be slowly varying compared to the channel coherence time and therefore to be perfectly known by *all* TXs.

In this distributed CSIT model, each TX receives its own CSI estimate for the CoMP channel. Specifically, TX j receives the multi-user channel estimate $\hat{\mathbf{H}}^{(j)} = \left[\begin{array}{ccc} \hat{\boldsymbol{h}}_1^{(j)} & \dots & \hat{\boldsymbol{h}}_K^{(j)} \end{array} \right]^{\mathrm{H}} \in \mathbb{C}^{K \times M}$ and designs its transmit coefficients solely as a function of $\hat{\mathbf{H}}^{(j)}$. We model the imperfect channel estimate $\hat{\boldsymbol{h}}_k^{(j)}$ for RX k at TX j as

$$\hat{\boldsymbol{h}}_k^{(j)} = \sqrt{M}\boldsymbol{\Theta}_k^{\frac{1}{2}} \left(\sqrt{1-(\sigma_k^{(j)})^2}\boldsymbol{z}_k + \sigma_k^{(j)}\boldsymbol{q}_k^{(j)} \right) = \sqrt{1-(\sigma_k^{(j)})^2}\boldsymbol{h}_k + \sigma_k^{(j)}\boldsymbol{\delta}_k^{(j)}. \tag{6.7}$$

The estimation error $\boldsymbol{\delta}_k^{(j)} = \sqrt{M}\boldsymbol{\Theta}_k^{\frac{1}{2}}\boldsymbol{q}_k^{(j)} \in \mathbb{C}^{M\times 1}$, where $\boldsymbol{q}_k^{(j)}$ has i.i.d complex entries of zero mean, variance $\frac{1}{M}$, eighth order moment of order $O(\frac{1}{M^4})$ and are independent of $\boldsymbol{z}_k$ and n_k. $\sigma_k^{(j)} \in [0,1]$ is a parameter indicating the accuracy or quality of the channel for RX k seen at TX j, i.e., $\sigma_k^{(j)} = 0$ correspond to perfect CSIT, whereas $\sigma_k^{(j)} = 1$ the CSIT is completely uncorrelated to the true channel. This model is widely used in the literature and is particularly well suited to model a channel quantization [38, 39].

Further, we assume that the estimation errors at TX j and TX j' satisfies

$$\boldsymbol{q}_k^{(j)} = \rho_k^{(j,j')}\boldsymbol{q}_k^{(j')} + \sqrt{1-(\rho_k^{(j,j')})^2}\boldsymbol{e}_k^{(j,j')}, \qquad \forall j, j', k, \tag{6.8}$$

where $\rho_k^{(j,j')} \in [0,1]$ is the correlation between $\boldsymbol{q}_k^{(j)}$ and $\boldsymbol{q}_k^{(j')}$. $\boldsymbol{e}_k^{(j,j')}$ has i.i.d complex entries of zero mean, variance $\frac{1}{M}$, eighth order moment of order $O(\frac{1}{M^4})$ and are independent of $\boldsymbol{q}_k^{(j')}$. Hence, the CSI estimate error are correlated as

$$\mathbb{E}\left[\boldsymbol{\delta}_k^{(j)}(\boldsymbol{\delta}_k^{(j')})^{\mathrm{H}}\right] = \boldsymbol{\Theta}_k^{\frac{1}{2}}\mathbb{E}\left[\boldsymbol{q}_k^{(j)}(\boldsymbol{q}_k^{(j')})^{\mathrm{H}}\right]\boldsymbol{\Theta}_k^{\frac{H}{2}} = \rho_k^{(j,j')}\boldsymbol{\Theta}_k. \tag{6.9}$$

Note that $\rho_k^{(j,j)} = 1, \forall j, k$.

This distributed CSI model which allows for correlation between the estimate errors at different TXs is very general. It is particularly adapted to model *imperfect* CSI backhaul between TXs where delay and/or imperfections are introduced.

Example 7. *Consider a particular CoMP network setting described in Fig. 6.1.*

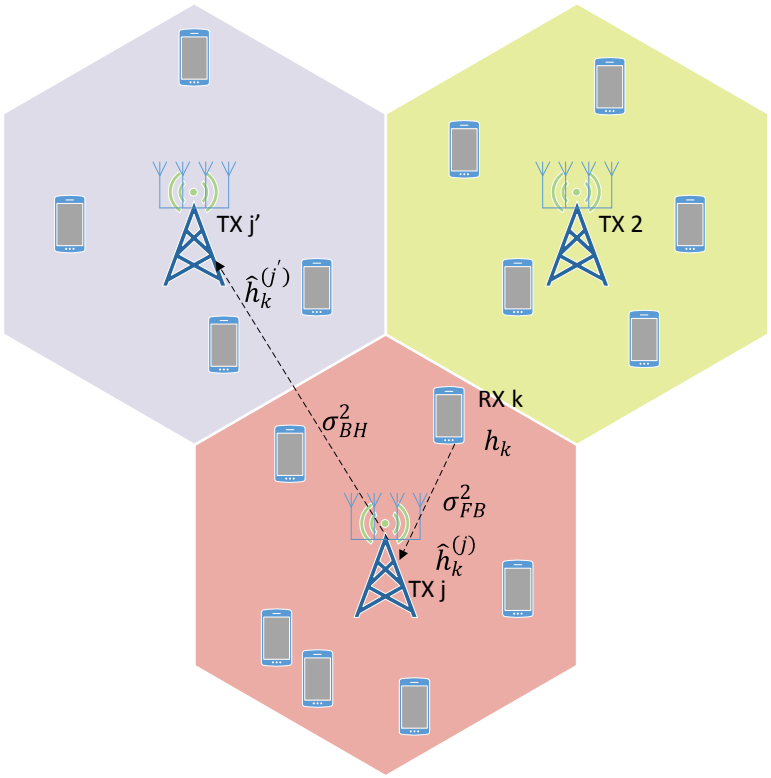


Figure 6.1: CoMP transmission network with limited CSI feedback and limited CSI sharing.

In a LTE FDD downlink channel estimation scenario, each base station sends pilots to all the served users. The RX k only feedback its downlink CSI to its associated base station, the TX j. Assume RX k has perfect CSIR and the channel degradation caused by limited feedback from RX k to TX j is σ^2_{FB}*, the CSIT seen at TX j for RX k is*

$$\hat{\boldsymbol{h}}_k^{(j)} = \sqrt{1-\sigma^2_{FB}}\boldsymbol{h}_k + \sigma_{FB}\boldsymbol{\delta}_k^{(j)},$$

where $\boldsymbol{\delta}_k^{(j)}$ *is the channel independent feedback noise.*

With the backhaul between TX j and TX j'*, assume the channel degradation caused by limited backhaul from TX j to TX* j' *is* σ^2_{BH}*, TX* j' *can*

obtain a quantized version of RX k channel

$$\begin{aligned}\hat{\boldsymbol{h}}_k^{(j')} &= \sqrt{1-\sigma_{BH}^2}\hat{\boldsymbol{h}}_k^{(j)} + \sigma_{BH}\boldsymbol{\epsilon}_k^{(j,j')} \\ &= \sqrt{\left(1-\sigma_{BH}^2\right)\left(1-\sigma_{FB}^2\right)}\boldsymbol{h}_k + \sigma_{FB}\sqrt{1-\sigma_{BH}^2}\boldsymbol{\delta}_k^{(j)} + \sigma_{BH}\boldsymbol{\epsilon}_k^{(j,j')},\end{aligned}$$

where $\boldsymbol{\epsilon}_k^{(j,j')}$ is the sharing noise independent from $\boldsymbol{h}_k, \boldsymbol{\delta}_k^{(j)}$.

The above scenario is well presented by the distributed CSI model with following setting

$$\begin{aligned}\sigma_k^{(j)} &= \sigma_{FB} \\ \sigma_k^{(j')} &= \sqrt{1-\left(1-\sigma_{BH}^2\right)\left(1-\sigma_{FB}^2\right)} \\ \rho_k^{(j,j')} &= \frac{\sigma_{FB}\sqrt{1-\sigma_{BH}^2}}{\sqrt{1-\left(1-\sigma_{BH}^2\right)\left(1-\sigma_{FB}^2\right)}}.\end{aligned} \tag{6.10}$$

Example 8. *The distributed CSI model allows to bridge the gap between the two extreme configuration: centralized CSI (CCSI) and distributed CSI with independent estimate errors. Indeed, the setting*

$$\sigma_k^{(j)} = \sigma_k^{(j')},\ \rho_k^{(j,j')} = 1, \qquad \forall j, j' = 1, \ldots, n,\ k = 1, \ldots, K \tag{6.11}$$

corresponds to the CCSI configuration [38, 39].

The setting

$$\rho_k^{(j,j')} = 0, \qquad \forall j, j' = 1, \ldots, n,\ j \neq j',\ k = 1, \ldots, K \tag{6.12}$$

simplifies to the distributed CSI with independent estimate errors configuration as previously studied in the literature [45].

6.1.3 Regularized zero forcing with distributed CSI

We restrain the work to the analysis of precoders falling into the catalog of *regularized ZF* precoder [46, 76], when faced with DCSI in the large system regime. Hence, the precoder designed at TX j is assumed to take the form

$$\mathbf{T}_{\mathrm{rZF}}^{(j)} = \left((\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)} + M\alpha^{(j)}\mathbf{I}_M\right)^{-1}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\frac{\sqrt{P}}{\sqrt{\Psi^{(j)}}}, \tag{6.13}$$

with regularization factor $\alpha^{(j)} > 0$. We also define

$$\mathbf{C}^{(j)} = \frac{(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}}{M} + \alpha^{(j)}\mathbf{I}_M \tag{6.14}$$

such that the precoder at TX j can be rewritten as

$$\mathbf{T}_{\mathrm{rZF}}^{(j)} = \frac{1}{M}(\mathbf{C}^{(j)})^{-1}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\frac{\sqrt{P}}{\sqrt{\Psi^{(j)}}}. \tag{6.15}$$

The scalar $\Psi^{(j)}$ corresponds to the power normalization at TX j. Hence, it holds that

$$\Psi^{(j)} = \|\Big((\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)} + M\alpha^{(j)}\mathbf{I}_M\Big)^{-1}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\|_{\mathrm{F}}^2. \tag{6.16}$$

Upon concatenation of all TX's precoding vectors, the effective global precoder denoted by $\mathbf{T}_{\mathrm{rZF}}^{\mathrm{DCSI}}$, is equal to

$$\mathbf{T}_{\mathrm{rZF}}^{\mathrm{DCSI}} = \begin{bmatrix} \mu_1\mathbf{E}_1^{\mathrm{H}}\mathbf{T}_{\mathrm{rZF}}^{(1)} \\ \mu_2\mathbf{E}_2^{\mathrm{H}}\mathbf{T}_{\mathrm{rZF}}^{(2)} \\ \vdots \\ \mu_n\mathbf{E}_n^{\mathrm{H}}\mathbf{T}_{\mathrm{rZF}}^{(n)} \end{bmatrix}, \tag{6.17}$$

where $\mathbf{E}_j^{\mathrm{H}} \in \mathbb{C}^{M_{\mathrm{TX}} \times M}$ is defined as

$$\mathbf{E}_j^{\mathrm{H}} = \begin{bmatrix} \mathbf{0}_{M_{\mathrm{TX}}\times(j-1)M_{\mathrm{TX}}} & \mathbf{I}_{M_{\mathrm{TX}}} & \mathbf{0}_{M_{\mathrm{TX}}\times(n-j)M_{\mathrm{TX}}} \end{bmatrix}. \tag{6.18}$$

The scalar $\mu_j > 0$ is the transmit power scaling at TX j. The average transmit power allocated at TX j is

$$P_{TX_j} = \mu_j^2\mathrm{E}\left[\mathrm{tr}\left(\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}\mathbf{T}_{\mathrm{rZF}}^{(j)}(\mathbf{T}_{\mathrm{rZF}}^{(j)})^{\mathrm{H}}\right)\right]. \tag{6.19}$$

According to the sum power constraint,

$$\sum_{j=1}^{n} P_{TX_j} = \sum_{j=1}^{n} \mu_j^2\mathrm{E}\left[\mathrm{tr}\left(\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}\mathbf{T}_{\mathrm{rZF}}^{(j)}(\mathbf{T}_{\mathrm{rZF}}^{(j)})^{\mathrm{H}}\right)\right] = P. \tag{6.20}$$

Although the finite SNR rate analysis under the precoding structure (6.17) and the DCSI model in (6.7) is challenging in the general case because of the dependency of one user performance on all channel estimates, some useful insights can be obtained in the large antenna regime as shown below.

6.2 Main Theoretical Result: Deterministic Equivalent of the SINR

In this section, the large system analysis of deterministic equivalent of the SINR under the RZF precoding for various assumptions is presented.

In order to derive a deterministic equivalent, we make the following assumption on the correlation matrices $\mathbf{\Theta}_k$ and the Gram matrix $\frac{1}{M}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}$.

Assumption 1. *All correlation matrices $\mathbf{\Theta}_k, \forall k = 1, \ldots, K$ have uniformly bounded spectral norm on M, i.e.,*

$$\limsup_{M,K\to\infty} \sup_{1\leq k\leq K} \| \mathbf{\Theta}_k \| < \infty. \tag{6.21}$$

Assumption 2. *The random matrices $\frac{1}{M}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}, \forall j = 1, \ldots, n$ have uniformly bounded spectral norm on M with probability one, i.e.,*

$$\limsup_{M,K\to\infty} \| \frac{1}{M}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)} \| < \infty \tag{6.22}$$

with probability one.

Our approach will be based on the following fundamental result based on the Stieltjes transform in the analysis of wireless networks [39, 49].

Theorem 1. *[49, 77] Let the matrix $\mathbf{U}$ be any matrix with bounded spectral norm and the ith column $\boldsymbol{h}_i$ of $\mathbf{H}^{\mathrm{H}}$ be $\boldsymbol{h}_i = \sqrt{M}\mathbf{\Theta}_i^{\frac{1}{2}}\boldsymbol{z}_i$, where the entries of $\boldsymbol{z}_i$ are i.i.d of zero mean, variance $\frac{1}{M}$ and have eighth moment of order $O(\frac{1}{M^4})$. Let Assumption 1 holds true. Consider the resolvent matrix $\mathbf{Q} = \left(\frac{\mathbf{H}^{\mathrm{H}}\mathbf{H}}{M} + \alpha\mathbf{I}_M\right)^{-1}$ with regularization coefficient $\alpha > 0$. Let*

$$\mathbf{Q}_o = \left(\frac{1}{M}\sum_{k=1}^{K}\frac{\mathbf{\Theta}_k}{1+m_k} + \alpha\mathbf{I}_M\right)^{-1}, \tag{6.23}$$

where m_k satisfies:

$$m_k = \frac{1}{M}\operatorname{tr}\left(\mathbf{\Theta}_k\left(\frac{1}{M}\sum_{\ell=1}^{K}\frac{\mathbf{\Theta}_\ell}{1+m_\ell} + \alpha\mathbf{I}_M\right)^{-1}\right). \tag{6.24}$$

Then,

$$\frac{1}{M}\operatorname{tr}(\mathbf{UQ}) - \frac{1}{M}\operatorname{tr}(\mathbf{UQ}_o) \xrightarrow[K,M\to\infty]{a.s.} 0. \tag{6.25}$$

The fixed point m_k can easily be obtained by an iterative algorithm given in [39, 50] and recalled in Appendix D for the sake of completeness.

Adopt the shorthand notation used in [39], we introduce

$$c_{0,k}^{(j)} = 1 - (\sigma_k^{(j)})^2, \quad c_{1,k}^{(j)} = (\sigma_k^{(j)})^2, \quad c_{2,k}^{(j)} = \sigma_k^{(j)}\sqrt{1-(\sigma_k^{(j)})^2}. \tag{6.26}$$

We can further define the term $\mathbf{Q}_o^{(j)}$ and $m_k^{(j)}$ respectively as $\mathbf{Q}_o$ and m_k in Theorem 1 using instead the local CSI estimate $\hat{\mathbf{H}}^{(j)}$ and regularization coefficient $\alpha^{(j)}$ at TX j. A deterministic equivalent of the SINR under RZF precoding is therefore provided in the following theorem.

Theorem 2. *Let the Assumptions 1 and 2 hold true, then the SINR of RX k under RZF precoding satisfies*

$$\mathrm{SINR}_k - \mathrm{SINR}_k^o \xrightarrow[K,M_{TX}\to\infty]{a.s.} 0. \tag{6.27}$$

SINR_k^o *is defined as*

$$\mathrm{SINR}_k^o = \frac{P\left(\sum_{j=1}^n \mu_j \sqrt{\frac{c_{0,k}^{(j)}}{\Gamma_{j,j}^o(\mathbf{I}_M)}} \frac{\Phi_{j,k}^o}{1+m_k^{(j)}}\right)^2}{1+I_k^o}. \tag{6.28}$$

$I_k^o \in \mathbb{R}$ *is given by*

$$I_k^o = P\sum_{j=1}^n\sum_{j'=1}^n \frac{\mu_j\mu_{j'}}{\sqrt{\Gamma_{j,j}^o(\mathbf{I}_M)\Gamma_{j',j'}^o(\mathbf{I}_M)}}\left(\Gamma_{j,j'}^o(\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\mathbf{\Theta}_k\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}})\right.$$
$$\left. -2\Gamma_{j,j'}^o(\mathbf{\Theta}_k\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}})\frac{c_{0,k}^{(j')}\Phi_{j',k}^o}{1+m_k^{(j')}} + \Phi_{j',k}^o\Phi_{j,k}^o\Gamma_{j,j'}^o(\mathbf{\Theta}_k)\frac{c_{0,k}^{(j)}c_{0,k}^{(j')}+\rho_k^{(j,j')}c_{2,k}^{(j)}c_{2,k}^{(j')}}{(1+m_k^{(j)})(1+m_k^{(j')})}\right), \tag{6.29}$$

where $\Phi_{j,k}^o \in \mathbb{R}$ *is defined as*

$$\Phi_{j,k}^o = \frac{\mathrm{tr}\left(\mathbf{\Theta}_k\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}\mathbf{Q}_o^{(j)}\right)}{M} \tag{6.30}$$

and the function $\Gamma_{j,j'}^o(\mathbf{X}) : \mathbb{C}^{M\times M} \mapsto \mathbb{C}$ *is defined in Lemma 12. The transmit power scaling* μ_j *for TX j satisfies*

$$\sum_{j=1}^n \mu_j^2 \frac{\Gamma_{j,j}^o(\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}})}{\Gamma_{j,j}^o(\mathbf{I}_M)} = 1. \tag{6.31}$$

Proof. The proof of Theorem 2 is given in Appendix B. □

The theorem demonstrates that in the large system setting, the SINR expression for each RX can be derived as a given function of (i) n, M_{TX}, K that indicate the system dimensions, (ii) $\sigma_k^{(j)}, \rho_k^{(j,j')}, \mathbf{\Theta}_k$ which reflect the statistic of the channel and of CSI estimates at each TX, and (iii) the precoder parameters regularization coefficient $\alpha^{(j)}$ and power scaling $\mu^{(j)}$.

This result is very general and encompasses several important results from the literature.

6.2.1 Regularized ZF precoding for centralized CSI isotropic channel

For a CCSI scenario described in [39], $\sigma_k^{(j)} = \sigma_k^{(j')} = \sigma_k$, $\alpha^{(j)} = \alpha^{(j')} = \alpha$, $\rho_k^{(j,j')} = 1, \forall j, j' = 1, \ldots, n,\ k = 1, \ldots, K$. Further assume that $\mathbf{\Theta}_k = \mathbf{I}_M$, this will simplify the calculation for $m_k^{(j)}$. Indeed, $m_k^{(j)}$ can be obtained in closed form

$$m_k^{(j)} = m^o = \frac{\beta - 1 - \alpha\beta + \sqrt{(\alpha\beta - \beta + 1)^2 + 4\alpha\beta^2}}{2\alpha\beta}. \tag{6.32}$$

In this setting, the total power constraint (6.31) simplifies to

$$\frac{1}{n}\sum_{j=1}^{n} \mu_j^2 = 1. \tag{6.33}$$

Since

$$\Gamma_{j,j}^o(\mathbf{I}_M) = \frac{(m^o)^2}{\beta(1+m^o)^2 - (m^o)^2} \tag{6.34}$$

$$\Gamma_{j,j}^o(\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}) = \frac{1}{n}\Gamma_{j,j}^o(\mathbf{I}_M), \tag{6.35}$$

assume $\mu_j = 1, \forall j = 1, \ldots, n$, the transmit power p_{TX_j} at TX j denotes

$$p_{TX_j} = \mu_j^2 P \frac{\Gamma_{j,j}^o(\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}})}{\Gamma_{j,j}^o(\mathbf{I}_M)} = \frac{P}{n}. \tag{6.36}$$

This indicates an equal power allocation per TX. Since $\mathbf{\Theta}_k = \mathbf{I}_M$ the channel is isotropic, the above setting also indicates that the signal power for RX k satisfies

$$p_k = \frac{P}{K}, \tag{6.37}$$

which is an equal power per RX.

With a few simplification, we can obtain the deterministic equivalent of SINR in (6.28)

$$\mathrm{SINR}_k^o = \frac{(1-\sigma_k^2)(\beta\,(1+m^o)^2-(m^o)^2)}{\left(1-\sigma_k^2+(1+m^o)^2\sigma_k^2+\frac{(1+m^o)^2}{P}\right)}. \tag{6.38}$$

Note that this coincides with [39, Corollary 2].

6.2.2 Regularized ZF precoding for distributed CSI with independent estimate error isotropic channel

In this settings, $\rho_k^{(j,j')}=0, \forall j\neq j',\ j,j'=1,\ldots,n,\ k=1,\ldots,K$. Assume same regularization coefficient at each TX $\alpha^{(j)}=\alpha^{(j')}=\alpha$, $\mathbf{\Theta}_k=\mathbf{I}_M$ and $\mu_j=1$ indicating equal per TX power allocation.

Adopt the notation for m^o in Section 6.2.1, with a few simplification, the deterministic SINR in (6.28) becomes

$$\mathrm{SINR}_k^o = \frac{P\left(\frac{1}{n}\sum_{j=1}^n\sqrt{c_{0,k}^{(j)}}\right)^2\frac{\beta(1+m^o)^2-(m^o)^2}{(1+m^o)^2}}{I_k^o+1} \tag{6.39}$$

with

$$I_k^o = P - P\sum_{j=1}^n\sum_{j'=1}^n \frac{\left(\beta\,(1+m^o)^2-(m^o)^2\right)\Gamma_{j,j'}^o}{n^2(1+m^o)^2m^o}\cdot\left[2c_{0,k}^{(j)}+m^o\left(2c_{0,k}^{(j)}-c_{0,k}^{(j)}c_{0,k}^{(j')}\right)\right] \tag{6.40}$$

$$\Gamma_{j,j'}^o = \frac{\frac{1}{M}\sum_{\ell=1}^K\sqrt{c_{0,\ell}^{(j)}c_{0,\ell}^{(j')}}}{\frac{(1+m^o)^2}{(m^o)^2}-\frac{1}{M}\sum_{\ell=1}^K c_{0,\ell}^{(j)}c_{0,\ell}^{(j')}} \tag{6.41}$$

Note that this result coincides with [78].

6.2.3 Regularized ZF precoding for distributed CSI isotropic channel

Assume that $\mathbf{\Theta}_k=\mathbf{I}_M, \forall k=1,\ldots,K$ and $\mu_j=1, \forall j=1,\ldots,n$ indicating equal per TX power allocation. In this specific setting, the terms $m_k^{(j)}$ can be obtained in closed form as

$$m_k^{(j)} = m^{(j)} = \frac{\beta-1-\alpha^{(j)}\beta+\sqrt{(\alpha^{(j)}\beta-\beta+1)^2+4\alpha^{(j)}\beta^2}}{2\alpha^{(j)}\beta} \tag{6.42}$$

With a few simplification, the deterministic SINR in (6.28) becomes

$$\mathrm{SINR}_k^o = \frac{P\left(\frac{1}{n}\sum_{j=1}^{n}\sqrt{\frac{1-(\sigma_k^{(j)})^2}{\Gamma_{j,j}^o}}\frac{m^{(j)}}{1+m^{(j)}}\right)^2}{1+I_k^o} \tag{6.43}$$

with $I_k^o \in \mathbb{R}$ defined as

$$I_k^o = P - P\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\Gamma_{j,j'}^o m^{(j)}}{\sqrt{\Gamma_{j,j}^o\Gamma_{j',j'}^o}}\left[\frac{2c_{0,k}^{(j)} - \left(\rho_k^{(j,j')}c_{2,k}^{(j)}c_{2,k}^{(j')} + c_{0,k}^{(j)}c_{0,k}^{(j')} - 2c_{0,k}^{(j)}\right)m^{(j')}}{n^2\left(1+m^{(j)}\right)\left(1+m^{(j')}\right)}\right], \tag{6.44}$$

where $\Gamma_{j,j'}^o \in \mathbb{R}$ is defined as

$$\Gamma_{j,j'}^o = \frac{\frac{1}{M}\sum_{\ell=1}^{K}\sqrt{c_{0,\ell}^{(j)}c_{0,\ell}^{(j')}} + \sqrt{c_{1,\ell}^{(j)}c_{1,\ell}^{(j')}}\rho_\ell^{(j,j')}}{\frac{1+m^{(j)}}{m^{(j)}}\frac{1+m^{(j')}}{m^{(j')}} - \frac{\sum_{\ell=1}^{K}\left(\sqrt{c_{0,\ell}^{(j)}c_{0,\ell}^{(j')}} + \sqrt{c_{1,\ell}^{(j)}c_{1,\ell}^{(j')}}\rho_\ell^{(j,j')}\right)^2}{M}}. \tag{6.45}$$

Note that this result coincides with [79].

6.3 Application of the Theorem

The deterministic equivalent of the SINR expression is helpful in the robust distributed RZF precoder design for coordinated multi-point transmission network. It can be applied to various of optimization problems discussed in the sequel.

6.3.1 Naive regularized ZF

When each TX is not aware of the CSI imperfection at the other TXs, it chooses its regularization parameter on the basis of its own CSI quality, which yields a naive (suboptimal) precoding scheme. Specifically, assume equal average power is allocated to each TX and each TX j optimize its regularization coefficient $\alpha^{(j)}$ based on $\hat{\mathbf{H}}^{(j)}$ considering that $\hat{\mathbf{H}}^{(j)}$ is the centralized CSIT shared among all TXs

$$\alpha_{naive}^{(j)} = \underset{\alpha^{(j)}}{\mathrm{argmax}}\,\mathrm{R}_{sum}\left(\hat{\mathbf{H}}^{(j)},\ldots,\hat{\mathbf{H}}^{(j)}\right), \qquad \forall j = 1,\ldots,n. \tag{6.46}$$

Algorithm 2 Robust Regular ZF Design

1: Discretize the search space for $\boldsymbol{\alpha}$
2: **for** each feasible $\boldsymbol{\alpha}$ point **do**
3: Optimize $\boldsymbol{\mu}$
4: **end for**
5: Find the optimal tuple $(\boldsymbol{\alpha}^\star, \boldsymbol{\mu}^\star)$ that maximizes the ergodic sum rate

In general, this can be done using a linear search and the large system approximation for the CCSI case is given in [39].

In the particular case with same quality across all links for a fixed TX, i.e., $\sigma_k^{(j)} = \sigma^{(j)}, \forall k = 1, \ldots, K$, and if the channel is isotropic, i.e., $\boldsymbol{\Theta}_k = \mathbf{I}_M$ and each TX has equal average power allocation, i.e., $\mu_j = 1$, a closed form solution exists [39]

$$\alpha^{(j),\mathrm{CCSI}} = \frac{1 + (\sigma^{(j)})^2 P}{1 - (\sigma^{(j)})^2} \frac{1}{\beta P}. \tag{6.47}$$

Naive RZF will be our benchmark precoding scheme as it corresponds to the configuration that neglects the DCSI structure.

6.3.2 Robust regularized ZF

Consider the DCSI setting, the optimal tuple of regularization coefficients and power scaling factors $(\boldsymbol{\alpha}^\star, \boldsymbol{\mu}^\star)$ that maximizing the system ergodic sum rate is defined as

$$(\boldsymbol{\alpha}^\star, \boldsymbol{\mu}^\star) = \underset{\boldsymbol{\alpha},\boldsymbol{\mu}}{\operatorname{argmax}} \sum_{k=1}^{K} \log\left(1 + \mathrm{SINR}_k^o\right), s.t. \sum_{j=1}^{n} \mu_j^2 \frac{\Gamma_{j,j}^o(\mathbf{E}_j \mathbf{E}_j^{\mathrm{H}})}{\Gamma_{j,j}^o(\mathbf{I}_M)} = 1. \tag{6.48}$$

In the general case, $\boldsymbol{\alpha}$ and $\boldsymbol{\mu}$ are coupled through (6.31). Since optimization problem is non-convex in $\boldsymbol{\alpha}$, if the power scaling $\boldsymbol{\mu}$ is fixed, we resort in this work to a grid search to find the optimal value of $\boldsymbol{\alpha}^*$.

$$\boldsymbol{\alpha}^\star = \underset{\boldsymbol{\alpha}}{\operatorname{argmax}} \sum_{k=1}^{K} \log\left(1 + \mathrm{SINR}_k^o\right). \tag{6.49}$$

Regarding to the power scaling $\boldsymbol{\mu}$, when $\boldsymbol{\alpha}$ is fixed, the optimal $\boldsymbol{\mu}^\star$ can be solved in a rather elegant manner described latter. This lead to Algorithm 2 for robust Regularized ZF design.

6.3.3 Robust regularized ZF with equal regularization

If the power scaling tuple is fixed, in order to reduce the complexity and ensure global optimality, we introduce the single parameter optimization imposing that the regularization coefficients at different TXs are identical

$$\alpha^{\star}_{same} = \underset{\alpha_{same}}{\operatorname{argmax}} \sum_{k=1}^{K} \log\left(1 + \mathrm{SINR}_k^o\right). \tag{6.50}$$

In this case, the global optimal regularization can be find with line search.

6.4 Optimal Power Control

Normally an equal power allocation per TX is considered and the corresponding power scaling takes the form

$$\mu_j^{eq} = \sqrt{\frac{P}{n \cdot \mathrm{E}\left[\mathrm{tr}\left(\mathbf{E}_j \mathbf{E}_j^{\mathrm{H}} \mathbf{T}_{\mathrm{rZF}}^{(j)} (\mathbf{T}_{\mathrm{rZF}}^{(j)})^{\mathrm{H}}\right)\right]}}\ . \tag{6.51}$$

However, in large system settings, according to Theorem 2, the power constraint (6.20) becomes (6.31) with deterministic $\Gamma_{j,j}^o(\mathbf{I}_M), \Gamma_{j,j}^o(\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}})$. Hence, each TX can optimize the power scaling for all the TXs in the absence of instantaneous CSI at other TXs.As this power control depends only on the long term statistics, it can be carried out in the same way at each TX since statistical information are perfectly shared among all TXs. In this way, average power control is enforced in the system.

Assume that the regularization coefficients $\alpha^{(j)}, \forall j$ are all fixed. The optimal power scaling tuple $\boldsymbol{\mu}^{\star} = [\mu_1^{\star}, \ldots, \mu_n^{\star}]^{\mathrm{T}}$ that maximizes the sum rate is obtained by solving the optimization problem:

$$\boldsymbol{\mu}^{\star} = \mathrm{argmax}_{\boldsymbol{\mu}} \textstyle\sum_{k=1}^{K} \log\left(1 + \mathrm{SINR}_k^o\right), s.t. \sum_{j=1}^{n} \mu_j^2 \frac{\Gamma_{j,j}^o(\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}})}{\Gamma_{j,j}^o(\mathbf{I}_M)} = 1\ , \tag{6.52}$$

where SINR^o is obtained from Equation (6.28).With a few simplifications, Problem (6.52) can be reformulated as:

$$\boldsymbol{\mu}^{\star} = \underset{\boldsymbol{\mu}}{\operatorname{argmin}} \prod_{k=1}^{K} \frac{\boldsymbol{\mu}^{\mathrm{T}} \mathbf{B}_k \boldsymbol{\mu}}{\frac{1}{P} + \boldsymbol{\mu}^{\mathrm{T}} (\mathbf{A}_k + \mathbf{B}_k) \boldsymbol{\mu}}, s.t. \parallel \mathbf{C}\boldsymbol{\mu} \parallel_F^2 = 1, \boldsymbol{\mu} \in \mathbb{R}^n\ , \tag{6.53}$$

where $\mathbf{A}_k, \mathbf{B}_k, \mathbf{C}, \forall k$ are constant matrices defined as

$$[\mathbf{A}_k]_{j,j'} = \sqrt{\frac{c_{0,k}^{(j)} c_{0,k}^{(j')}}{\Gamma_{j,j}^o(\mathbf{I}_M)\Gamma_{j',j'}^o(\mathbf{I}_M)}} \frac{\Phi_{j,k}^o \Phi_{j',k}^o}{\left(1+m_k^{(j)}\right)\left(1+m_k^{(j')}\right)} \tag{6.54}$$

$$[\mathbf{B}_k]_{j,j'} = \frac{1}{\sqrt{\Gamma_{j,j}^o(\mathbf{I}_M)\Gamma_{j',j'}^o(\mathbf{I}_M)}} \left(\Gamma_{j,j'}^o(\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\mathbf{\Theta}_k\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}) \right.$$
$$\left. -2\Gamma_{j,j'}^o\,(\mathbf{\Theta}_k\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}})\frac{c_{0,k}^{(j')}\Phi_{j',k}^o}{1+m_k^{(j')}} + \Phi_{j',k}^o\Phi_{j,k}^o\Gamma_{j,j'}^o(\mathbf{\Theta}_k)\frac{c_{0,k}^{(j)}c_{0,k}^{(j')} + \rho_k^{(j,j')}c_{2,k}^{(j)}c_{2,k}^{(j')}}{(1+m_k^{(j)})(1+m_k^{(j')})} \right) \tag{6.55}$$

$$\mathbf{C} = \mathrm{diag}\left(\sqrt{\frac{\Gamma_{1,1}^o(\mathbf{E}_1\mathbf{E}_1^{\mathrm{H}})}{\Gamma_{1,1}^o(\mathbf{I}_M)}}, \ldots, \sqrt{\frac{\Gamma_{n,n}^o(\mathbf{E}_n\mathbf{E}_n^{\mathrm{H}})}{\Gamma_{n,n}^o(\mathbf{I}_M)}} \right). \tag{6.56}$$

Optimization Problem (6.53) then falls in the category of composite concave program (CCP) [80].We will therefore exploit a result provided in [80] which shows that a CCP attains the same optimum as an associated parametric optimization. Hence, Problem (6.53) can then be solved via Algorithm 3.

Algorithm 3 Iterative algorithm for CCP Problem (6.53)

1: Initialize $\boldsymbol{\mu}^{[0]}$
2: $t = 0$
3: **while** not converge **do**
4: $\quad y_k(\boldsymbol{\mu}^{[t]}) = \frac{(\boldsymbol{\mu}^{[t]})^{\mathrm{T}}\mathbf{B}_k\boldsymbol{\mu}^{[t]}}{\frac{1}{P}+(\boldsymbol{\mu}^{[t]})^{\mathrm{T}}(\mathbf{A}_k+\mathbf{B}_k)\boldsymbol{\mu}^{[t]}}$
5: $\quad$ Solve Problem (6.57)
6: $\quad t = t+1$
7: **end while**

Problem (6.57) in the iteration is defined as

$$\begin{aligned} \boldsymbol{\mu}^{[t+1]} \quad &= \underset{\mathbf{x}}{\mathrm{argmin}} \textstyle\prod_{\ell=1}^K y_\ell(\boldsymbol{\mu}^{[t]}) \cdot \sum_{k=1}^K \frac{y_k(\mathbf{x})}{y_k(\boldsymbol{\mu}^{[t]})} \\ &\quad \| C\mathbf{x} \|^2 \leq 1 \end{aligned}, \tag{6.57}$$

which can be rewritten as

$$\begin{aligned} &\min_{\mathbf{x}} \textstyle\sum_{k=1}^K \frac{1}{y_k(\boldsymbol{\mu}^{[t]})} \cdot \frac{\mathbf{x}^{\mathrm{T}}\mathbf{B}_k\mathbf{x}}{\frac{1}{P}+\mathbf{x}^{\mathrm{T}}(\mathbf{A}_k+\mathbf{B}_k)\mathbf{x}} \\ &\| C\mathbf{x} \|^2 \leq 1 \end{aligned}. \tag{6.58}$$

Problem (6.58) is a nonlinear sum-of-ratio optimization and the global optimum can be found iteratively as in [81]. A more detailed discussion of the properties of these optimization problems is provided for the interested reader in Appendix C.

The complete algorithm with the inside optimization problem solved iteratively is given in Algorithm 4 and it then fulfils the following global optimality result.

Theorem 3. *The global optimal solution of power scaling optimization problem* (6.53) *can be solved using a two-level iterative algorithm described in Algorithm 4.*

Proof. The proof of Theorem 3 is given in Appendix C. □

Algorithm 4 Iterative algorithm for optimal power scaling

Initialize $\boldsymbol{\mu}^{[0]}$
$t = 0$
while not converge **do**
  $y_k(\boldsymbol{\mu}^{[t]}) = \frac{\frac{1}{P}+(\boldsymbol{\mu}^{[t]})^{\mathrm{T}}\mathbf{B}_k\boldsymbol{\mu}^{[t]}}{\frac{1}{P}+(\boldsymbol{\mu}^{[t]})^{\mathrm{T}}(\mathbf{A}_k+\mathbf{B}_k)\boldsymbol{\mu}^{[t]}}, \forall k$
  initialize $\mathbf{x}^{[0]}$
  $i = 0$
  while not converge **do**
    $\beta_k^{[i+1]} = \frac{\frac{1}{y_k(\boldsymbol{\mu}^{[t]})}\left(\frac{1}{P}+(\mathbf{x}^{[i]})^{\mathrm{T}}\mathbf{B}_k\mathbf{x}^{[i]}\right)}{\frac{1}{P}+(\mathbf{x}^{[i]})^{\mathrm{T}}(\mathbf{A}_k+\mathbf{B}_k)\mathbf{x}^{[i]}}$
    $u_k^{[i+1]} = \frac{1}{\frac{1}{P}+(\mathbf{x}^{[i]})^{\mathrm{T}}(\mathbf{A}_k+\mathbf{B}_k)\mathbf{x}^{[i]}}$
    $\mathbf{x}^{[i+1]} = \arg\min_{\mathbf{x}} \sum_{k=1}^{K} u_k^{[i+1]} \cdot \left(\frac{\frac{1}{P}+\mathbf{x}^{\mathrm{T}}\mathbf{B}_k\mathbf{x}}{y_k(\boldsymbol{\mu}^{[t]})} - \beta_k^{[i+1]} \cdot \left(\frac{1}{P}+\mathbf{x}^{\mathrm{T}}(\mathbf{A}_k+\mathbf{B}_k)\mathbf{x}\right)\right)$
    $s.t. \quad \| \mathbf{Cx} \|^2 \leq 1$
    $i = i + 1$
  end while
  $\boldsymbol{\mu}^{[t+1]} = \mathbf{x}^{[i]}$
  $t = t + 1$
end while

6.4.1 Particular case: isotropic channel

When the channel is isotropic, (i.e., $\boldsymbol{\Theta}_k = \mathbf{I}_M, \forall k = 1, \ldots, K$) the optimization for the regularization coefficient and the power scaling factors can be

largely simplified. In this case, $m_k^{(j)}$ can be derived in closed form as [79]

$$m_k^{(j)} = \frac{\beta - 1 - \alpha^{(j)}\beta + \sqrt{(\alpha^{(j)}\beta - \beta + 1)^2 + 4\alpha^{(j)}\beta^2}}{2\alpha^{(j)}\beta}$$

and

$$\Gamma_{j,j}^{o}(\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}) = \frac{1}{n}\Gamma_{j,j}^{o}(\mathbf{I}_M).$$

In this case, the power constraint reads

$$\sum_{j=1}^{n} \mu_j^2 = n \tag{6.59}$$

which is no longer intertwined with $\alpha^{(j)}$. This leads to an intuitive alternating optimization shown in Algorithm 5.

Algorithm 5 Joint optimization for power scaling and regularization coefficient

1: Initialize power scaling tuple $\boldsymbol{\mu}$
2: **while** not converge **do**
3: Optimize $\alpha^{(j)}, \forall j$ for fixed $\boldsymbol{\mu}$
4: Optimize $\boldsymbol{\mu}$ as described in Section 6.4
5: **end while**

Consider step (3) in Algorithm 5, the regularization coefficients optimization for given power scaling tuple $\boldsymbol{\mu}$ can be either a vector optimization described in Section 6.3.2 or a single parameter optimization mentioned in Section 6.3.3.

6.5 Simulation Results

This section is divided into two parts. In the first part, we will first verify the deterministic equivalent result. Then, under isotropic channel case, the effect of different parameters on the deterministic equivalent of the system sum rate will be analyzed. In the second part, a simulation in a realistic cellular network settings is provided and the gain of optimal precoder design over the naive precoder design is exhibited.

6.5.1 The effect of different parameters on the deterministic equivalent

In this subsection, unless otherwise specified, we adopt the simulation settings listed in Table 6.1, Table 6.2 and Table 6.3.

M	K	n	β	$\mathbf{\Theta}_k, \forall k$	P
30	30	3	1	$\mathbf{I}_M$	20dB

Table 6.1: Simulation parameters for isotropic channel setting.

Correlation / Scenario	$\rho_k^{(j,j')}, \forall k, \forall j \neq j'$
DCSI uncorrelated error	0
DCSI weakly correlated error	0.01
DCSI strongly correlated error	0.81
CCSI	1

Table 6.2: Different settings for correlation parameter $\rho_k^{(j,j')}$.

Setting / Accuracy	asymmetric setting	symmetric setting
$(\sigma_k^{(1)})^2, \forall k$	0.01	0.1
$(\sigma_k^{(2)})^2, \forall k$	0.16	0.1
$(\sigma_k^{(3)})^2, \forall k$	0.49	0.1

Table 6.3: Different settings for channel accuracy parameter $(\sigma_k^{(1)})^2$.

We verify using Monte-Carlo (MC) simulations the accuracy of the asymptotic expression derived in Theorem 2.

Fig.6.2 depicts the absolute error of the deterministic equivalent R_{sum}^0 compared to the ergodic sum rate R_{sum} as a function of K (or M equivalently). The ergodic sum rate is averaged over 1000 independent channel realizations. For ease of illustration, we choose the symmetric channel accuracy setting, the power allocation is assumed to be equal power per TX and the regularization coefficient at each TX j is chosen as $\alpha^{(j)} = \frac{1}{\beta P}$.

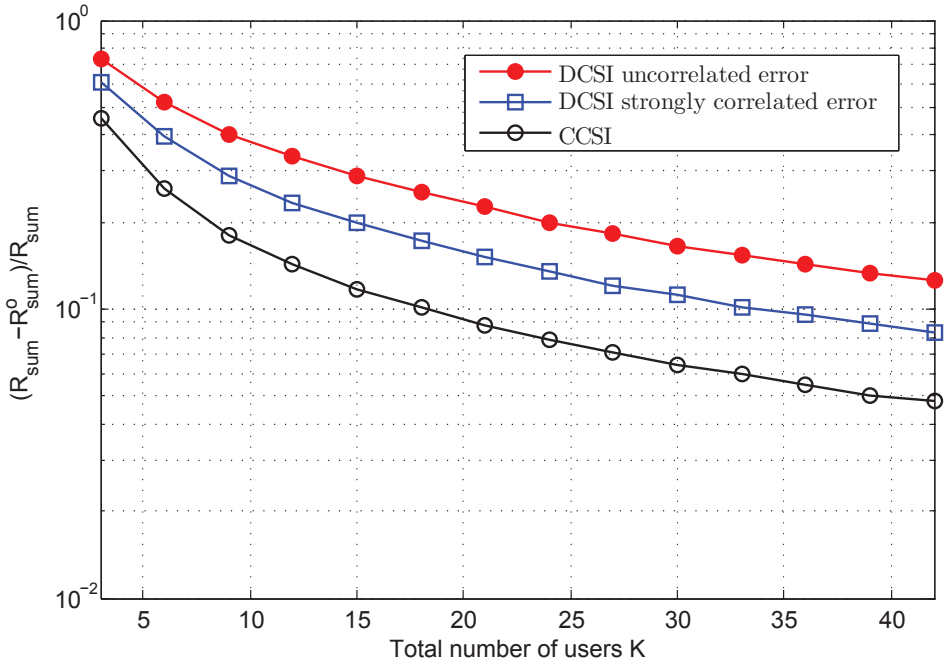


Figure 6.2: RZF, $(R_{sum} - R^0_{sum})/R_{sum}$ vs K for different configurations.

In Fig. 6.2, it confirms that Monte-Carlo simulation converges to the deterministic equivalent as the system becomes large. It also reveals that when the DCSI configuration becomes more "centralized", the deterministic equivalent will be more accurate.

Joint optimization of regularization coefficient and power control

We consider hereby the optimization of regularization coefficient and power scaling. The performances of three algorithms have been compared.

- $(\boldsymbol{\alpha}^\star, \boldsymbol{\mu}^\star)$: Apply iterative algorithm 5, with regularization optimized according to (6.49) and power scaling optimized according to (6.53).
- $(\alpha^\star_{same}, \boldsymbol{\mu}^\star)$: Apply iterative algorithm 5, with regularization optimized according to (6.50) and power scaling optimized according to (6.53).
- $(\boldsymbol{\alpha}_{naive}, \boldsymbol{\mu}_{eq})$: Regularization coefficient is calculated according to (6.47) and equal power allocation per TX according to (6.51).

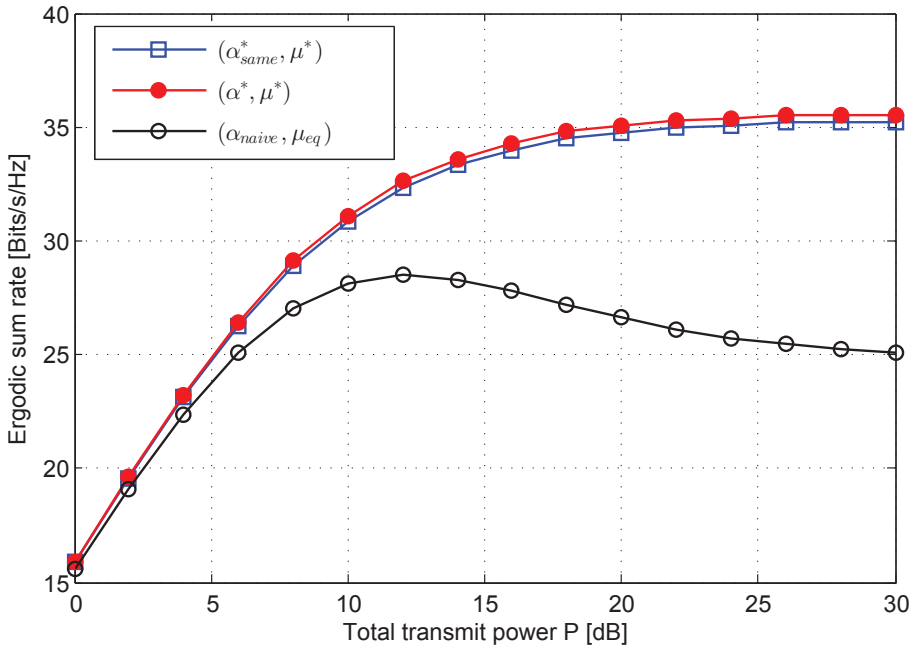


Figure 6.3: RZF, ergodic sum rate vs total transmit power for power allocation and regularization coefficient optimization.

Choose the DCSI weakly correlated error and asymmetric channel accuracy setting, plot the ergodic sum rate as a function of the total transmit power P varies from 0 dB to 30dB.

In Fig. 6.3, we can observe that there is a large performance increasing when the power control and the regularization coefficient are jointly optimized. The joint optimization is robust to the DCSI structure. We can also conclude that only a negligible performance degradation is introduced when imposing identical regularization coefficient at each TX.

The impact of CSIT distributiveness

As is mention in Section 6.1.2, the CSI estimate noise correlation parameter $\rho_k^{(j,j')}$ reflects the distributiveness of this CoMP network.

Adopt the symmetric channel accuracy setting, let the CSI estimate noise correlation be $\rho_k^{(j,j')} = \rho, \forall k, \forall j \neq j'$, we plot the ergodic sum rate as a function of the CSI estimate noise correlation ρ varies from 0 to 1, namely, the CSI structure varies from DCSI with uncorrelated error to CCSI.

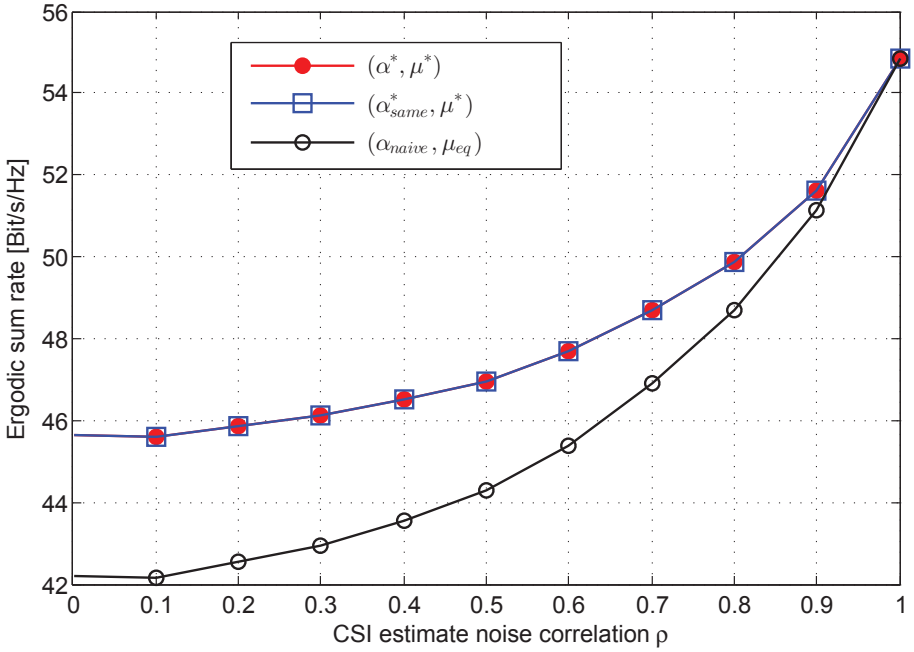


Figure 6.4: RZF, sum rate vs CSI estimate noise correlation ρ.

From Fig. 6.4, it reveals that the proposed algorithms outperforms the naive one in the DCSI scenarios. We can also conclude the distributed CSI at different TXs introduce a non-vanishing performance degradation from the centralized CSI case.

6.5.2 A cellular configuration simulation

In this subsection, we simulate a realistic cellular setting.

The correlation matrix $\mathbf{\Theta}_k$ of the kth RX channel is a block diagonal matrix that reads $\mathbf{\Theta}_k = \text{blockdiag}(\mathbf{\Theta}_{k,1}, \ldots, \mathbf{\Theta}_{k,n})$. The correlation matrix $\mathbf{\Theta}_{k,j}$ between the kth RX and the jth TX denotes [39]

$$[\mathbf{\Theta}_{k,j}]_{\ell,m} = \gamma d_{k,j}^{-\epsilon} \cdot \frac{1}{\theta_{k,\max} - \theta_{k,\min}} \int_{\theta_{k,\min}}^{\theta_{k,\max}} e^{\mathrm{i}\frac{2\pi}{\lambda}\cdot(\ell-m)d_{as}\cdot\cos\theta} d\theta. \qquad (6.60)$$

The $\gamma d_{k,j}^{-\epsilon}$ part indicates the pathloss, with ϵ being the pathloss component, $d_{k,j}$ the distance between RX k and TX j and γ a coefficient to further adjust the model. The rest part models the Uniform Linear Array (ULA) assuming a diffuse two-dimensional field of isotropic scatters around the receivers. The

waves impinge the receiver k uniformly at an azimuth angle θ ranging from $\theta_{k,\min}$ to $\theta_{k,\max}$. The angle spread is φ. The antennas spacing $d_{as} = \frac{\lambda}{2}$, λ is the signal wavelength. To ensure that $\| \mathbf{\Theta}_k \|$ is bounded as M grows large, we assume that the distance $d_{k,j}$ is larger than some threshold distance d_0 and antennas spacing is independent of M.

The parameters are listed in the Table 6.4.

M	30
K	30
n	2
β	1
hexagon cell radius	$1km$
γ	10^9
ϵ	3
d_0	$0.1km$
f	2GHz
d_{as}	$\frac{\lambda}{2}$
φ	$\frac{\pi}{6}$
P	20dB
σ^2_{FB}	$\frac{1}{10}$
σ^2_{BH}	$\frac{7}{9}$

Table 6.4: Simulation parameters for a cellular setting.

According to (6.10), this cellular setting indicates that the accuracy parameters for RX ℓ in the cell is $(\sigma_\ell^{(n)})^2 = 0.1$ and for RX m outside cell is $(\sigma_m^{(n)})^2 = 0.8$ $\forall \ell, m = 1, \ldots, \frac{K}{n}, \forall n$. The corresponding correlation $\rho_k^{(j,j')} = \frac{1}{6}, \forall k = 1, \ldots, K, \forall j \neq j'$.

Fig. 6.5 exhibits the ergodic sum rate as a function of total power P. The performances of two algorithms: (i)$(\boldsymbol{\alpha}^\star, \boldsymbol{\mu}^\star)$ and (ii) $(\boldsymbol{\alpha}_{naive}, \boldsymbol{\mu}_{eq})$ we can conclude that the proposed joint optimization of regularization and power scaling outperforms the naive algorithm. It is also robust to the DCSI structure for cellular setting.

6.6 Conclusion

In this chapter, We have studied regularized ZF joint precoding in a ditributed CSI configuration. We extend the conventional centralized CSI to dis-

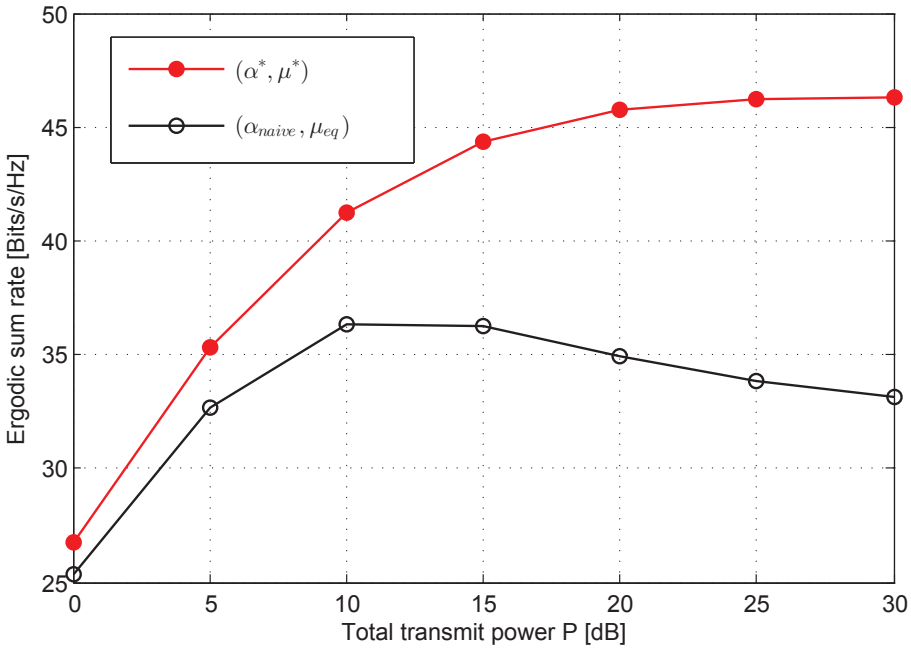


Figure 6.5: Cellular setting RZF, ergodic sum rate vs total transmit power for power allocation and regularization coefficient optimization.

tributed CSI scenario by allowing the CSI errors at the different TXs to be arbitrarily correlated. Using random matrix theory tools, an analytical expression is derived to approximate the average rate per user in the large system limit. This deterministic equivalent expression is then used to optimize the regularization coefficients as well as the power allocation at the different TXs in order to reduce the impact of the DCSI configuration.

Chapter 7

Conclusion and Future Works

In this thesis, we consider the problem of decentralized transmitter cooperation and signal processing with limited feedback and backhaul. we have decoupled the problem into two sub-problems: finding an efficient strategy for information exchange over the limited coordination backhaul and design robust decentralized transmitter cooperation algorithm.

In the first part of the thesis, we first analyze an simple transmitter cooperation scenario named master-slave scenario. In this configuration where the master transmitter is endowed with perfect network-wise CSIT and the slave transmitters have zero initial CSIT, we have proved that with carefully designed codebook, exchanging CSI based data and precoder based data over the limited coordination backhaul are equivalent. We have also proposed suboptimal but low complexity algorithm for the precoder design. We then focus on using the coordination backhaul links only for CSI exchange and target the problem of optimal design for CSI sharing. We have cast this problem into a framework of cooperative channel estimation with side information, which is rooted in the Wyner-Ziv coding and multi-source compressing. We have introduced an approach to jointly optimize the quantizer on the coordination backhaul and the combing function that each TX should apply to obtain the final CSI estimate. We have also proposed algorithms to solve the problem of backhaul resource allocation when the total backhaul resource is limited.

The focus of the second part of the thesis is on the design of robust decentralized transmitter cooperation algorithm. It should be noticed that after the CSI sharing over the coordination backhaul, the CSI at different

TXs are imperfect and non-identical. This CSI configuration is referred as distributed CSIT configuration and we mainly focus on the joint processing transmitter cooperation with RZF precoder under the CSI configuration. Since the decentralized transmitter cooperation design belongs to the well-known and difficult team decision problem, we have made the analysis under large system assumption. With the help of random matrix theory tools, we can find the deterministic equivalent expression for the system ergodic sum rate and optimize the regularization coefficient and the power control for the robust decentralized RZF precoder design.

In this thesis, we have only worked on the robust decentralized transmitter cooperation for RZF precoder. An interesting direction for the future works could be the analysis on robust decentralized transmitter cooperation beyond the RZF precoder, e.g, for other possible precoding strategies. It is also very promising topic to perform information theoretic analysis such as the capacity region analysis for the decentralized transmitter cooperation. Besides, for the decentralized transmitter cooperation and signal processing with limited feedback and backhaul, we have deliberately divided it into separate problems of efficient information exchange design and decentralized transmitter cooperation design. Another interesting yet challenging direction is the ultimate goal of joint optimization of the two. Last but not least, it is very promising to analysis and determine what degree of centralization is most efficient in future wireless system. As 5G and beyond are assumed to be very heterogeneous networks, a better understanding and decision on what level of centralization should be applied for transmitter cooperation could even affect the architecture of future networks.

Appendices

A Useful Lemmas

Lemma 1. *for a random vector* $\mathbf{x}$*, if an optimal* S *levels Euclidean distance distortion based quantizer applied on a random vector* $\mathbf{y} = \mathbf{B}^{\frac{1}{2}}\mathbf{x}$ *has the codebook* $\{\mathbf{y}_1, \ldots, \mathbf{y}_S\}$ *and associated partition* $\{\mathcal{P}_1, \ldots, \mathcal{P}_S\}$*, then the optimal* S *level weighted square error distortion based quantizer applied on the random vector* $\mathbf{x}$ *with shaping matrix* $\mathbf{B}$ *will have the codebook* $\{\mathbf{B}^{-\frac{1}{2}}\mathbf{y}_1, \ldots, \mathbf{B}^{-\frac{1}{2}}\mathbf{y}_S\}$ *and associated partition* $\{\mathbf{B}^{-\frac{1}{2}}[\mathcal{P}_1], \ldots, \mathbf{B}^{-\frac{1}{2}}[\mathcal{P}_S]\}$*,where the* $\mathbf{B}^{-\frac{1}{2}}[\mathcal{P}_i]$ *is defined as* $\mathbf{B}^{-\frac{1}{2}}[\mathcal{P}_i] = \{\mathbf{x} : \exists \mathbf{y} \in \mathcal{P}_i \text{ s.t. } \mathbf{x} = \mathbf{B}^{-\frac{1}{2}}\mathbf{y}\}$.

Proof. Consider the distortion associated with the optimal codebook and partition:

$$
\begin{aligned}
D_{\mathbf{y}} &= \sum_{j=1}^{S} f_{\mathbf{y}}(\mathbf{y}_j) \int\limits_{\mathbf{y}\in\mathcal{P}_j} (\mathbf{y}-\mathbf{y}_j)^H(\mathbf{y}-\mathbf{y}_j)d\mathbf{y}. \\
D_{\mathbf{x}} &= \sum_{j=1}^{S} f_{\mathbf{x}}(\mathbf{x}_j) \int\limits_{\mathbf{x}\in\mathcal{M}_j} (\mathbf{x}-\mathbf{x}_j)^H B(\mathbf{x}-\mathbf{x}_j)d\mathbf{x} \\
&= \det(\mathbf{B}^{-1}) \sum_{j=1}^{S} f_{\mathbf{x}}(\mathbf{x}_j) \int\limits_{\mathbf{u}\in\mathbf{B}^{\frac{1}{2}}[\mathcal{M}_j]} (\mathbf{u}-\mathbf{B}^{\frac{1}{2}}\mathbf{x}_j)^H(\mathbf{u}-\mathbf{B}^{\frac{1}{2}}\mathbf{x}_j)d\mathbf{u}.
\end{aligned}
$$

Where $\{\mathbf{x}_1, \ldots, \mathbf{x}_S\}$ and $\{\mathcal{M}_1, \ldots, \mathcal{M}_S\}$ denote the optimal codebook and partition for the quantizing $\mathbf{x}$. Let $\mathbf{B}^{\frac{1}{2}}\mathbf{x}_j = \mathbf{y}_j$ and $\mathbf{B}^{\frac{1}{2}}[\mathcal{M}_j] = \mathcal{P}_j$, since

$$
f_{\mathbf{y}}(\mathbf{y}_j) = f_{\mathbf{x}}(\mathbf{x}_j)\det(\mathbf{B}^{-1}),
$$

by change of variable in the integral,we can get $D_{\mathbf{y}} = D_{\mathbf{x}}$. □

Lemma 2. *[82, Lemma 2.5] The matrix function* $h : \mathbf{X} \mapsto -\mathbf{X}^{-1}$ *is both a strictly matrix concave function and a matrix monotone increasing function.*

Lemma 3. *If* $\mathbf{X}$ *is positive semi-definite hermitian matrix, then matrix function* $h : \mathbf{X} \mapsto tr(\mathbf{X}^{-1})$ *is convex and matrix monotone decreasing.*

Proof. Consider $g(t) = h(\mathbf{Z} + t\mathbf{V})$, where $\mathbf{Z}$ is positive semi-definite hermitian matrix, $\mathbf{V}$ is hermitian matrix, since

$$
\frac{d^2}{dt^2}g(t)|_{t=0} = \operatorname{tr}\left(2(\mathbf{Z}^{-1}\mathbf{V})\mathbf{Z}^{-1}(\mathbf{Z}^{-1}\mathbf{V})^{-1}\right) \geq 0,
$$

therefore, $f(\mathbf{X})$ is convex.
For $\mathbf{X} \succeq \mathbf{Y} \succeq \mathbf{0}$, according to Lemma 2,

$$\mathbf{Y}^{-1} \succeq \mathbf{X}^{-1} \succeq \mathbf{0}$$
$$h(\mathbf{X}) - h(\mathbf{Y}) = \mathrm{tr}(\mathbf{X}^{-1} - \mathbf{Y}^{-1}) \leq 0,$$

therefore, the function is matrix monotone decreasing. □

B Proof of Theorem 2

The proof is built upon results from both [39] and [49] and novel lemmas Lemma 12 and Lemma 13. We also make extensive use of classical RMT lemmas recalled in Appendix D. In particular, Lemma 12 extends [49, Lemma 15] and is an interesting result in itself.

B.1 Deterministic equivalent for $\Psi^{(j)}$

We start by finding a deterministic equivalent for $\Psi^{(j)}$. Apply Lemma 12 with $\hat{\mathbf{H}}^{(j')} = \hat{\mathbf{H}}^{(j)},\ \mathbf{A} = \mathbf{I}_M$, which gives

$$\begin{aligned}\Psi^{(j)} &\asymp \Gamma^o_{j,j}(\mathbf{I}_M) \\ &= \left(\frac{1}{M} \sum_{\ell=1}^{K} \frac{\frac{1}{M}\,\mathrm{tr}\left(\mathbf{\Theta}_\ell \mathbf{Q}_o^{(j)} \mathbf{Q}_o^{(j)}\right)}{(1+m_\ell^{(j)})^2} + \frac{1}{M} \sum_{\ell=1}^{K} \frac{\Gamma^o_{j,j}(\mathbf{\Theta}_\ell)\frac{1}{M}\,\mathrm{tr}\left(\mathbf{\Theta}_\ell \mathbf{Q}_o^{(j)} \mathbf{Q}_o^{(j)}\right)}{(1+m_\ell^{(j)})^2} \right).\end{aligned}$$

From (1), it can be noted that, as expected, this deterministic equivalent does not depend on $\sigma_\ell^{(j)}$. The total power constraint for large scale system reads

$$\begin{aligned}&\| \mathbf{T}_{\mathrm{rZF}}^{\mathrm{DCSI}} \|_{\mathrm{F}}^2 \\ &= \sum_{j=1}^{n} \mu_j^2 \mathrm{tr}\left(\mathbf{E}_j^{\mathrm{H}} \mathbf{T}_{\mathrm{rZF}}^{(j)} (\mathbf{T}_{\mathrm{rZF}}^{(j)})^{\mathrm{H}} \mathbf{E}_j\right) \\ &\overset{(a)}{\asymp} \sum_{j=1}^{n} \mu_j^2 \frac{P}{\Gamma^o_{j,j}(\mathbf{I}_M)} \Gamma^o_{j,j}(\mathbf{E}_j \mathbf{E}_j^{\mathrm{H}}) \\ &= P,\end{aligned}$$

where (a) follows from Lemma 12. Therefore, there is a constraint for the power scaling factors μ_j:

$$\sum_{j=1}^{n} \mu_j^2 \frac{\Gamma^o_{j,j}(\mathbf{E}_j \mathbf{E}_j^{\mathrm{H}})}{\Gamma^o_{j,j}(\mathbf{I}_M)} = 1.$$

B.2 Deterministic equivalent for $\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{t}_{\mathrm{rZF},k}^{\mathrm{DCSI}}$

Turning to the desired signal at RX k, we can write

$$\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{t}_{\mathrm{rZF},k}^{\mathrm{DCSI}} = \sum_{j=1}^{n} \frac{1}{M} \frac{\mu_j \sqrt{P}}{\sqrt{\Psi^{(j)}}} \boldsymbol{h}_k^{\mathrm{H}} \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}} (\mathbf{C}^{(j)})^{-1} \hat{\boldsymbol{h}}_k^{(j)} \tag{1}$$

$$\overset{(a)}{\asymp} \sqrt{P} \sum_{j=1}^{n} \mu_j \sqrt{\frac{1}{\Gamma_{j,j}^{o}(\mathbf{I}_M)}} \frac{\frac{1}{M} \boldsymbol{h}_k^{\mathrm{H}} \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}} (\mathbf{C}_{[k]}^{(j)})^{-1} \hat{\boldsymbol{h}}_k^{(j)}}{1 + \frac{1}{M} (\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}} (\mathbf{C}_{[k]}^{(j)})^{-1} \hat{\boldsymbol{h}}_k^{(j)}} \tag{2}$$

$$\overset{(b)}{\asymp} \sqrt{P} \sum_{j=1}^{n} \mu_j \sqrt{\frac{1 - (\sigma_k^{(j)})^2}{\Gamma_{j,j}^{o}(\mathbf{I}_M)}} \frac{\frac{1}{M} \boldsymbol{h}_k^{\mathrm{H}} \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}} (\mathbf{C}_{[k]}^{(j)})^{-1} \boldsymbol{h}_k}{1 + \frac{1}{M} (\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}} (\mathbf{C}_{[k]}^{(j)})^{-1} \hat{\boldsymbol{h}}_k^{(j)}} \tag{3}$$

$$\overset{(c)}{\asymp} \sqrt{P} \sum_{j=1}^{n} \mu_j \sqrt{\frac{1 - (\sigma_k^{(j)})^2}{\Gamma_{j,j}^{o}(\mathbf{I}_M)}} \frac{\frac{1}{M} \operatorname{tr}\left(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}} (\mathbf{C}_{[k]}^{(j)})^{-1} \right)}{1 + \frac{1}{M} \operatorname{tr}\left(\boldsymbol{\Theta}_k (\mathbf{C}_{[k]}^{(j)})^{-1} \right)} \tag{4}$$

$$\overset{(d)}{\asymp} \sqrt{P} \sum_{j=1}^{n} \mu_j \sqrt{\frac{1 - (\sigma_k^{(j)})^2}{\Gamma_{j,j}^{o}(\mathbf{I}_M)}} \frac{\frac{1}{M} \operatorname{tr}\left(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}} (\mathbf{C}^{(j)})^{-1} \right)}{1 + \frac{1}{M} \operatorname{tr}\left(\boldsymbol{\Theta}_k (\mathbf{C}^{(j)})^{-1} \right)} \tag{5}$$

$$\overset{(e)}{\asymp} \sqrt{P} \sum_{j=1}^{n} \mu_j \sqrt{\frac{1 - (\sigma_k^{(j)})^2}{\Gamma_{j,j}^{o}(\mathbf{I}_M)}} \frac{\frac{1}{M} \operatorname{tr}\left(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}} \mathbf{Q}_o^{(j)} \right)}{1 + \frac{1}{M} \operatorname{tr}\left(\boldsymbol{\Theta}_k \mathbf{Q}_o^{(j)} \right)}, \tag{6}$$

where we have defined

$$\mathbf{C}_{[k]}^{(j)} = \frac{\hat{\mathbf{H}}_{[k]}^{(j)} (\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}}{M} + \alpha^{(j)} \mathbf{I}_M, \qquad \forall j \tag{7}$$

with

$$(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}} = \left[\hat{\boldsymbol{h}}_1^{(j)} \quad \dots \quad \hat{\boldsymbol{h}}_{k-1}^{(j)} \quad \hat{\boldsymbol{h}}_{k+1}^{(j)} \quad \dots \quad \hat{\boldsymbol{h}}_K^{(j)} \right], \qquad \forall j. \tag{8}$$

Equality (a) follows then from Lemma 7 and the use of the deterministic equivalent derived for $\Psi^{(j)}$, (b) from Lemma 9, (c) from Lemma 8, (d) from Lemma 10 and (e) from the fundamental Theorem 1.

It follows then directly that

$$\left| \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{t}_{\mathrm{rZF},k}^{\mathrm{DCSI}} \right|^2 \asymp P \left(\sum_{j=1}^{n} \mu_j \sqrt{\frac{1 - (\sigma_k^{(j)})^2}{\Gamma_{j,j}^{o}(\mathbf{I}_M)}} \frac{\frac{1}{M} \operatorname{tr}\left(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}} \mathbf{Q}_o^{(j)} \right)}{1 + \frac{1}{M} \operatorname{tr}\left(\boldsymbol{\Theta}_k \mathbf{Q}_o^{(j)} \right)} \right)^2. \tag{9}$$

B.3 Deterministic Equivalent for $\mathcal{I}_k$

Our first step is to write explicitly the interference term using the definition of $\mathbf{T}^{\mathrm{DCSI}}$ and replace $\Psi^{(j)}$ by its deterministic equivalent.

$$\mathcal{I}_k = \sum_{\ell=1,\ell\neq k}^{K} |\boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{t}_{\mathrm{rZF},\ell}^{\mathrm{DCSI}}|^2 \tag{10}$$

$$= \boldsymbol{h}_k^{\mathrm{H}} \mathbf{T}_{\mathrm{rZF}}^{\mathrm{DCSI}} (\mathbf{T}_{\mathrm{rZF}}^{\mathrm{DCSI}})^{\mathrm{H}} \boldsymbol{h}_k - \boldsymbol{h}_k^{\mathrm{H}} \boldsymbol{t}_{\mathrm{rZF},k}^{\mathrm{DCSI}} (\boldsymbol{t}_{\mathrm{rZF},k}^{\mathrm{DCSI}})^{\mathrm{H}} \boldsymbol{h}_k \tag{11}$$

$$= \frac{1}{M^2} \sum_{j=1}^{n} \sum_{j'=1}^{n} \frac{\mu_j \mu_{j'} P}{\sqrt{\Psi^{(j)} \Psi^{(j')}}} \boldsymbol{h}_k^{\mathrm{H}} \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}} (\mathbf{C}^{(j)})^{-1} (\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}} \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}^{(j')})^{-1} \mathbf{E}_{j'} \mathbf{E}_{j'}^{\mathrm{H}} \boldsymbol{h}_k \tag{12}$$

$$\asymp \frac{P}{M^2} \sum_{j=1}^{n} \sum_{j'=1}^{n} \frac{\mu_j \mu_{j'}}{\sqrt{\Gamma_{j,j}^{o}(\mathbf{I}_M) \Gamma_{j',j'}^{o}(\mathbf{I}_M)}} \boldsymbol{h}_k^{\mathrm{H}} \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}} (\mathbf{C}_{[k]}^{(j)})^{-1} \tag{13}$$

$$\cdot (\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}} \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}^{(j')})^{-1} \mathbf{E}_{j'} \mathbf{E}_{j'}^{\mathrm{H}} \boldsymbol{h}_k \tag{14}$$

$$+ \frac{P}{M^2} \sum_{j=1}^{n} \sum_{j'=1}^{n} \frac{\mu_j \mu_{j'}}{\sqrt{\Gamma_{j,j}^{o}(\mathbf{I}_M) \Gamma_{j',j'}^{o}(\mathbf{I}_M)}} \boldsymbol{h}_k^{\mathrm{H}} \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}} \left((\mathbf{C}^{(j)})^{-1} - (\mathbf{C}_{[k]}^{(j)})^{-1} \right) \tag{15}$$

$$\cdot (\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}} \hat{\mathbf{H}}_{[k]}^{(j')} (\mathbf{C}^{(j')})^{-1} \mathbf{E}_{j'} \mathbf{E}_{j'}^{\mathrm{H}} \boldsymbol{h}_k. \tag{16}$$

To obtain a deterministic equivalent for the second summation in (16) we use the following relation

$$(\mathbf{C}^{(j)})^{-1} - (\mathbf{C}_{[k]}^{(j)})^{-1}$$

$$= (\mathbf{C}^{(j)})^{-1} \left(\mathbf{C}_{[k]}^{(j)} - \mathbf{C}^{(j)} \right) (\mathbf{C}_{[k]}^{(j)})^{-1} \tag{17}$$

$$= -\frac{(\mathbf{C}^{(j)})^{-1}}{M} \left(c_{0,k}^{(j)} \boldsymbol{h}_k \boldsymbol{h}_k^{\mathrm{H}} + c_{1,k}^{(j)} \boldsymbol{\delta}_k^{(j)} (\boldsymbol{\delta}_k^{(j)})^{\mathrm{H}} + c_{2,k}^{(j)} \boldsymbol{\delta}_k^{(j)} \boldsymbol{h}_k^{\mathrm{H}} + c_{2,k}^{(j)} \boldsymbol{h}_k (\boldsymbol{\delta}_k^{(j)})^{\mathrm{H}} \right) (\mathbf{C}_{[k]}^{(j)})^{-1}, \tag{18}$$

where $c_{0,k}^{(j)}, c_{1,k}^{(j)}, c_{2,k}^{(j)}$ is defined in (6.26). It is important to note that

$$c_{0,k}^{(j)} c_{1,k}^{(j)} = (c_{2,k}^{(j)})^2, c_{0,k}^{(j)} + c_{1,k}^{(j)} = 1,$$

as these relations will be used several times through the proof.

Inserting (18) into (16), the interference term can be denoted as

$$
\begin{aligned}
\tilde{\mathcal{I}}_k \asymp & \frac{P}{M^2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\mu_j\mu_{j'}}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M)\Gamma^o_{j',j'}(\mathbf{I}_M)}}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)}_{[k]})^{-1} \\
& \qquad \cdot (\hat{\mathbf{H}}^{(j)}_{[k]})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}_{[k]}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k \\
& - \frac{P}{M^3}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\mu_j\mu_{j'}}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M)\Gamma^o_{j',j'}(\mathbf{I}_M)}}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)})^{-1} \\
& \qquad \cdot \left[\boldsymbol{h}_k c^{(j)}_{0,k}\boldsymbol{h}_k^{\mathrm{H}}\right](\mathbf{C}^{(j)}_{[k]})^{-1}(\hat{\mathbf{H}}^{(j)}_{[k]})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}_{[k]}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k \\
& - \frac{P}{M^3}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\mu_j\mu_{j'}}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M)\Gamma^o_{j',j'}(\mathbf{I}_M)}}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)})^{-1} \\
& \qquad \cdot \left[\boldsymbol{\delta}^{(j)}_k c^{(j)}_{1,k}(\boldsymbol{\delta}^{(j)}_k)^{\mathrm{H}}\right](\mathbf{C}^{(j)}_{[k]})^{-1}(\hat{\mathbf{H}}^{(j)}_{[k]})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}_{[k]}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k \\
& - \frac{P}{M^3}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\mu_j\mu_{j'}}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M)\Gamma^o_{j',j'}(\mathbf{I}_M)}}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)})^{-1} \\
& \qquad \cdot \left[\boldsymbol{\delta}^{(j)}_k c^{(j)}_{2,k}\boldsymbol{h}_k^{\mathrm{H}}\right](\mathbf{C}^{(j)}_{[k]})^{-1}(\hat{\mathbf{H}}^{(j)}_{[k]})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}_{[k]}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k \\
& - \frac{P}{M^3}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\mu_j\mu_{j'}}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M)\Gamma^o_{j',j'}(\mathbf{I}_M)}}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)})^{-1} \\
& \qquad \cdot \left[\boldsymbol{h}_k c^{(j)}_{2,k}(\boldsymbol{\delta}^{(j)}_k)^{\mathrm{H}}\right](\mathbf{C}^{(j)}_{[k]})^{-1}(\hat{\mathbf{H}}^{(j)}_{[k]})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}_{[k]}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k \\
& = A + B + C + D + E. \qquad (19)
\end{aligned}
$$

We proceed by calculating terms A to E in (19) successively, using Lemma 13. For the sake of simplicity, we only proceed the calculation of term A and the rest terms can be calculated in similar manner.

$$
\begin{aligned}
A = & \frac{P}{M^2}\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\mu_j\mu_{j'}}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M)\Gamma^o_{j',j'}(\mathbf{I}_M)}}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)}_{[k]})^{-1} \\
& \qquad \cdot(\hat{\mathbf{H}}^{(j)}_{[k]})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}_{[k]}(\mathbf{C}^{(j')})^{-1}\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{h}_k \\
\asymp & P\sum_{j=1}^{n}\sum_{j'=1}^{n}\frac{\mu_j\mu_{j'}}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M)\Gamma^o_{j',j'}(\mathbf{I}_M)}} \\
& \qquad \cdot\Bigg[\frac{\mathrm{tr}\left(\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{\Theta}_k\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)}_{[k]})^{-1}(\hat{\mathbf{H}}^{(j)}_{[k]})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}_{[k]}(\mathbf{C}^{(j')}_{[k]})^{-1}\right)}{M^2} \\
& \qquad\quad - c^{(j')}_{0,k}\frac{\mathrm{tr}\left(\boldsymbol{\Theta}_k\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)}_{[k]})^{-1}(\hat{\mathbf{H}}^{(j)}_{[k]})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}_{[k]}(\mathbf{C}^{(j')}_{[k]})^{-1}\right)}{M^2} \\
& \qquad\qquad \cdot\frac{\mathrm{tr}\left(\boldsymbol{\Theta}_k\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}(\mathbf{C}^{(j')}_{[k]})^{-1}\right)}{M}\frac{1+c^{(j')}_{1,k}\frac{\mathrm{tr}\left(\boldsymbol{\Theta}_k(\mathbf{C}^{(j')}_{[k]})^{-1}\right)}{M}}{1+\frac{\mathrm{tr}\left(\boldsymbol{\Theta}_k(\mathbf{C}^{(j')}_{[k]})^{-1}\right)}{M}} \\
& \qquad\quad +(c^{(j')}_{2,k})^2\frac{\mathrm{tr}\left(\boldsymbol{\Theta}_k\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)}_{[k]})^{-1}(\hat{\mathbf{H}}^{(j)}_{[k]})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}_{[k]}(\mathbf{C}^{(j')}_{[k]})^{-1}\right)}{M^2} \\
& \qquad\qquad \cdot\frac{\mathrm{tr}\left(\boldsymbol{\Theta}_k\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}(\mathbf{C}^{(j')}_{[k]})^{-1}\right)}{M}\frac{\frac{\mathrm{tr}\left(\boldsymbol{\Theta}_k(\mathbf{C}^{(j')}_{[k]})^{-1}\right)}{M}}{1+\frac{\mathrm{tr}\left(\boldsymbol{\Theta}_k(\mathbf{C}^{(j')}_{[k]})^{-1}\right)}{M}}\Bigg]. \qquad (20)
\end{aligned}
$$

According to Lemma 10 and Lemma 12, we can have:

$$
\begin{aligned}
&\frac{1}{M^2}\mathrm{tr}\left(\boldsymbol{\Theta}_k\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)}_{[k]})^{-1}(\hat{\mathbf{H}}^{(j)}_{[k]})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}_{[k]}(\mathbf{C}^{(j')}_{[k]})^{-1}\right)\asymp\Gamma^o_{j,j'}(\boldsymbol{\Theta}_k\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}) \\
&\frac{1}{M^2}\mathrm{tr}\left(\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{\Theta}_k\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}(\mathbf{C}^{(j)}_{[k]})^{-1}(\hat{\mathbf{H}}^{(j)}_{[k]})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}_{[k]}(\mathbf{C}^{(j')}_{[k]})^{-1}\right)\asymp\Gamma^o_{j,j'}(\mathbf{E}_{j'}\mathbf{E}_{j'}^{\mathrm{H}}\boldsymbol{\Theta}_k\mathbf{E}_j\mathbf{E}_j^{\mathrm{H}}). \qquad (21)
\end{aligned}
$$

Inserting (21) in (20) and using the fundamental Theorem 1 yields

$$
\begin{aligned}
A \asymp P \sum_{j=1}^{n} \sum_{j'=1}^{n} \mu_j \mu_{j'} & \frac{\Gamma^o_{j,j'}(\mathbf{E}_{j'} \mathbf{E}_{j'}^{\mathrm{H}} \boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}})}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M) \Gamma^o_{j',j'}(\mathbf{I}_M)}} \\
& - \mu_j \mu_{j'} c_{0,k}^{(j')} \frac{\Gamma^o_{j,j'}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}})}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M) \Gamma^o_{j',j'}(\mathbf{I}_M)}} \frac{\operatorname{tr}\left(\boldsymbol{\Theta}_k \mathbf{E}_{j'} \mathbf{E}_{j'}^{\mathrm{H}} \mathbf{Q}_o^{(j')} \right)}{M} \frac{1 + c_{1,k}^{(j')} m_k^{(j')}}{1 + m_k^{(j')}} \\
& + \mu_j \mu_{j'} (c_{2,k}^{(j')})^2 \frac{\Gamma^o_{j,j'}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}})}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M) \Gamma^o_{j',j'}(\mathbf{I}_M)}} \frac{\operatorname{tr}\left(\boldsymbol{\Theta}_k \mathbf{E}_{j'} \mathbf{E}_{j'}^{\mathrm{H}} \mathbf{Q}_o^{(j')} \right)}{M} \frac{m_k^{(j')}}{1 + m_k^{(j')}} \\
\overset{(a)}{\asymp} P \sum_{j=1}^{n} \sum_{j'=1}^{n} \mu_j \mu_{j'} & \frac{\Gamma^o_{j,j'}(\mathbf{E}_{j'} \mathbf{E}_{j'}^{\mathrm{H}} \boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}})}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M) \Gamma^o_{j',j'}(\mathbf{I}_M)}} \\
& - \mu_j \mu_{j'} c_{0,k}^{(j')} \frac{\Gamma^o_{j,j'}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}})}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M) \Gamma^o_{j',j'}(\mathbf{I}_M)}} \frac{\operatorname{tr}\left(\boldsymbol{\Theta}_k \mathbf{E}_{j'} \mathbf{E}_{j'}^{\mathrm{H}} \mathbf{Q}_o^{(j')} \right)}{M} \frac{1}{1 + m_k^{(j')}},
\end{aligned}
\tag{22}
$$

where equality (a) follows from $c_{0,k}^{(j')} c_{1,k}^{(j')} = (c_{2,k}^{(j')})^2$. Proceed similarly for the remaining 4 terms and add term A, B, C, D and E together, we can get

$$
\begin{aligned}
\mathcal{I}_k &= A + B + C + D + E \qquad (23) \\
&= P \sum_{j=1}^{n} \sum_{j'=1}^{n} \frac{\mu_j \mu_{j'} \Gamma^o_{j,j'}(\mathbf{E}_{j'} \mathbf{E}_{j'}^{\mathrm{H}} \boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}})}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M) \Gamma^o_{j',j'}(\mathbf{I}_M)}} \\
&\quad - 2P \sum_{j=1}^{n} \sum_{j'=1}^{n} \frac{\mu_j \mu_{j'} \Gamma^o_{j,j'}(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}})}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M) \Gamma^o_{j',j'}(\mathbf{I}_M)}} \frac{\operatorname{tr}\left(\boldsymbol{\Theta}_k \mathbf{E}_{j'} \mathbf{E}_{j'}^{\mathrm{H}} \mathbf{Q}_o^{(j')} \right)}{M} \frac{c_{0,k}^{(j')}}{1 + m_k^{(j')}} \\
&\quad + P \sum_{j=1}^{n} \sum_{j'=1}^{n} \mu_j \mu_{j'} \frac{\frac{\operatorname{tr}\left(\boldsymbol{\Theta}_k \mathbf{E}_{j'} \mathbf{E}_{j'}^{\mathrm{H}} \mathbf{Q}_o^{(j')} \right)}{M} \frac{\operatorname{tr}\left(\boldsymbol{\Theta}_k \mathbf{E}_j \mathbf{E}_j^{\mathrm{H}} \mathbf{Q}_o^{(j)} \right)}{M} \Gamma^o_{j,j'}(\boldsymbol{\Theta}_k)}{\sqrt{\Gamma^o_{j,j}(\mathbf{I}_M) \Gamma^o_{j',j'}(\mathbf{I}_M)}} \\
&\qquad \cdot \frac{c_{0,k}^{(j)} c_{0,k}^{(j')} + \rho_k^{(j,j')} c_{2,k}^{(j)} c_{2,k}^{(j')}}{(1 + m_k^{(j)})(1 + m_k^{(j')})}. \qquad (24)
\end{aligned}
$$

This concludes the proof.

C Proof of Theorem 3

The proof is organized as follows: we first prove that the optimal solution of a CCP Problem (6.53) can be obtained by solving an associated parametric problem. Then we prove that this parametric problem, which is the sum of quadratic fraction, can be solved by an iterative optimization algorithm (6.57). According to [81], the optimal solution of sum of quadratic ratio optimization can be computed by an iterative algorithm with two tuples of auxiliary variables as described in (27). Therefore, the last step is to prove that the optimal solution of (27) can be attained.

Define the Composite Concave Programing (CCP):

$$v^* = \min_{x \in X} \Phi(u(x)),$$

where X is any arbitrary set, u is a function on X with values in $\mathbb{R}^K$ and Φ is a differentiable real valued function on some open convex set U containing the set $U(X) = \{u(x), x \in X\}$. Let X^* be the set of global optimal solution of this problem.

Define the parametric problem of CCP (CCP(λ)):

$$v^*(\lambda) = \min_{x \in X} v(x; \lambda) = \sum_{i=1}^{K} \lambda_i u_i(x),$$

where $\lambda \in \Lambda \subset \mathbb{R}^K$ is viewed as parameter. Let $X^*(\lambda)$ denote the set of global optimal solutions to problem CCP(λ).

Consider first the following Lemmas.

Lemma 4 (Chapter 4, Theorem 2.1, [80])**.** *Let the function Φ be a differentiable real valued function on some open convex set U containing the set $U(X) = \{u(x) : x \in X\}$. If $\mathbf{\Phi}$ is pseudoconcave with respect to u, then:*

$$x \in X^* \Rightarrow x \in X^*(\nabla\Phi(u(x))),$$
$$x \in X^* \Rightarrow X^*(\nabla\Phi(u(x))) \subset X^*.$$

Lemma 5. *Function $\Phi : \mathbb{R}^K \mapsto \mathbb{R}, \Phi(\mathbf{y}) = \prod_{i=1}^{K} y_i$ is differentiable and pseudoconcave on the open convex set $(0, \infty)^K$.*

Proof. According to the definition of pseudoconcave function, if $\forall \boldsymbol{a}, \boldsymbol{b} \in (0, \infty)^K$, we have

$$-\nabla\Phi(\boldsymbol{a}) \cdot (\boldsymbol{b} - \boldsymbol{a}) \geq 0,$$

which yields

$$K \geq \sum_{k=1}^{K} \frac{b_k}{a_k}.$$

According to the AM-GM inequality,

$$\frac{1}{K}\sum_{k=1}^{K} \frac{b_k}{a_k} \geq \sqrt[K]{\frac{\prod_{k=1}^{K} b_k}{\prod_{k=1}^{K} a_k}}.$$

Therefore,

$$-\Phi(\boldsymbol{b}) = -\prod_{k=1}^{K} b_k \geq -\prod_{k=1}^{K} a_k = -\Phi(\boldsymbol{a}),$$

this indicates that the function is pseudoconcave. □

Now consider the optimization problem (6.53), according to Lemma 4, 5, the optimal solution to the following problem

$$\begin{aligned} \min_{\boldsymbol{\mu}} \quad & \sum_{k=1}^{K} \lambda_k \frac{\boldsymbol{\mu}^{\mathrm{T}} \mathbf{B}_k \boldsymbol{\mu}}{\frac{1}{P} + \boldsymbol{\mu}^{\mathrm{T}} (\mathbf{A}_k + \mathbf{B}_k) \boldsymbol{\mu}} \\ s.t. \quad & \lambda_k = \prod_{\ell=1, \ell \neq k}^{K} \frac{\boldsymbol{\mu}^{\mathrm{T}} \mathbf{B}_\ell \boldsymbol{\mu}}{\frac{1}{P} + \boldsymbol{\mu}^{\mathrm{T}} (\mathbf{A}_\ell + \mathbf{B}_\ell) \boldsymbol{\mu}} \\ & \| \mathbf{C}\boldsymbol{\mu} \|_F^2 = 1, \boldsymbol{\mu} \in \mathbb{R}^{n \times 1} \end{aligned} \tag{25}$$

is the optimal solution for the optimization problem (6.53).

Now we will prove that the optimum of Problem (25) can be achieved via the proposed iterative algorithm (6.57). Denote that

$$y_k(\boldsymbol{\mu}^{[t]}) = \frac{(\boldsymbol{\mu}^{[t]})^{\mathrm{T}} \mathbf{B}_k \boldsymbol{\mu}^{[t]}}{\frac{1}{P} + (\boldsymbol{\mu}^{[t]})^{\mathrm{T}} (\mathbf{A}_k + \mathbf{B}_k) \boldsymbol{\mu}^{[t]}},$$

problem (6.57) can be further expressed as

$$\begin{aligned} \boldsymbol{\mu}^{[t+1]} \quad & = \arg\min_{\mathbf{x}} \prod_{\ell=1}^{K} y_\ell(\boldsymbol{\mu}^{[t]}) \cdot \sum_{k=1}^{K} \frac{y_k(\mathbf{x})}{y_k(\boldsymbol{\mu}^{[t]})} \\ & \| C\mathbf{x} \|^2 \leq 1 \end{aligned} . \tag{26}$$

From (26) we can have

$$\prod_{\ell=1}^{K} y_\ell(\boldsymbol{\mu}^{[t]}) \cdot \sum_{k=1}^{K} \frac{y_k(\boldsymbol{\mu}^{[t+1]})}{y_k(\boldsymbol{\mu}^{[t]})} \leq \prod_{\ell=1}^{K} y_\ell(\boldsymbol{\mu}^{[t]}) \cdot \sum_{k=1}^{K} \frac{y_k(\boldsymbol{\mu}^{[t]})}{y_k(\boldsymbol{\mu}^{[t]})} = K \cdot \prod_{\ell=1}^{K} y_\ell(\boldsymbol{\mu}^{[t]}),$$

together with the AM-GM inequality, it indicates

$$K \sqrt[K]{\frac{\prod_{k=1}^{K} y_k(\boldsymbol{\mu}^{[t+1]})}{\prod_{k=1}^{K} y_k(\boldsymbol{\mu}^{[t]})}} \leq \sum_{k=1}^{K} \frac{y_k(\boldsymbol{\mu}^{[t+1]})}{y_k(\boldsymbol{\mu}^{[t]})} \leq K.$$

This suggests that in each iteration, the value $\prod_{k=1}^{K} y_k(\mathbf{x})$ decreases, which coincides with the objective function of problem (25). This concludes that by solving the parametric problem iteratively, it will converge to the optimal solution of problem (6.53).

According to [81], the optimal solution of problem (6.58) is computed as follows:

Algorithm 6 Iterative algorithm for sum of ratios Problem (6.58)

1: Initialize $\mathbf{x}^{[0]}$
2: $i = 0$
3: **while** not converge **do**
4: $\quad \beta_k^{[i+1]} = \frac{\frac{1}{y_k(\boldsymbol{\mu}^{[t]})}\left(\frac{1}{P}+(\mathbf{x}^{[i]})^{\mathrm{T}}\mathbf{B}_k\mathbf{x}^{[i]}\right)}{\frac{1}{P}+(\mathbf{x}^{[i]})^{\mathrm{T}}(\mathbf{A}_k+\mathbf{B}_k)\mathbf{x}^{[i]}}, \forall k$
$\quad\quad u_k^{[i+1]} = \frac{1}{\frac{1}{P}+(\mathbf{x}^{[i]})^{\mathrm{T}}(\mathbf{A}_k+\mathbf{B}_k)\mathbf{x}^{[i]}}, \forall k$
5: $\quad$ Solve Problem (27)
6: $\quad i = i + 1$
7: **end while**

where the Problem (27) is defined as

$$\begin{aligned} \mathbf{x}^{[i+1]} &= \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{k=1}^{K} u_k^{[i+1]} \cdot \left(\frac{\frac{1}{P}+\mathbf{x}^{\mathrm{T}}\mathbf{B}_k\mathbf{x}}{y_k(\boldsymbol{\mu}^{[t]})} - \beta_k^{[i+1]} \cdot \left(\frac{1}{P} + \mathbf{x}^{\mathrm{T}}\left(\mathbf{A}_k + \mathbf{B}_k\right)\mathbf{x}\right)\right) \\ &\quad \| \mathbf{Cx} \|^2 \leq 1 \end{aligned} . \tag{27}$$

Now we prove that Problem (27) can be efficiently solved by semi-definite relaxation.

Define $\mathbf{X} = \mathbf{x}\mathbf{x}^{\mathrm{T}}$, (27) has an semi-definite programming (SDP) relaxation

$$\begin{aligned} \mathbf{x}^{[i+1]} &= \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{k=1}^{K} u_k^{[i+1]} \cdot \left(\frac{\frac{1}{P}+\mathrm{tr}(\mathbf{B}_k\mathbf{X})}{y_k(\boldsymbol{\mu}^{[t]})} - \beta_k^{[i+1]} \cdot \left(\frac{1}{P} + \mathrm{tr}\left(\left(\mathbf{A}_k + \mathbf{B}_k\right)\mathbf{X}\right)\right)\right), \\ s.t. \quad & \mathrm{tr}\left(\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{X}\right) \leq 1, \mathbf{X} \in \mathbb{R}^{n\times n}, \mathbf{X} \succeq 0 \end{aligned} \tag{28}$$

with the rank($\mathbf{X}$) = 1 constrained been dropped. According to the Shapiro-Barvinok-Pataki result [83], when there is no more than 2 quadratic constraints, the rank 1 SDP relaxation is tight and there is always an optimal $\mathbf{X}^*$ with rank($\mathbf{X}^*$) = 1.

Therefore, the optimal solution of problem (6.53) can be obtained by solving a two-level iterative algorithm. This concludes the proof.

D Classical Lemmas from the Literature

Lemma 6 (Adapted from [39,50]). *Let $\alpha^{(j)} > 0, j = 1, \ldots, n$ and $m_k^{(j)[t]}, t \geq 0$ be the sequence defined as*

$$\begin{cases} m_k^{(j)[0]} = \frac{1}{\alpha^{(j)}} & \forall k = 1, \ldots, K \\ m_k^{(j)[t]} = \frac{1}{M} \operatorname{tr}\left(\boldsymbol{\Theta}_k \left(\frac{1}{M} \sum_{\ell=1}^{K} \frac{\boldsymbol{\Theta}_\ell}{1+m_\ell^{(j)[t-1]}} + \alpha^{(j)} \mathbf{I}_M \right)^{-1} \right) & \text{for } t \geq 1 \end{cases}. \tag{29}$$

Then $m_k^{(j)[t]} \xrightarrow{t\to\infty} m_k^{(j)}$, with $m_k^{(j)}$ being by construction an iterative algorithm of (29).

Lemma 7 (Resolvent Identities [49, 50]). *Given any matrix $\mathbf{H} \in \mathbb{C}^{K\times M}$, let $\boldsymbol{h}_k^{\mathrm{H}}$ denote its kth row and $\mathbf{H}_{[k]} \in \mathbb{C}^{(K-1)\times M}$ denote the matrix obtained after removing the kth row from $\mathbf{H}$. The resolvent matrices of $\mathbf{H}$ and $\mathbf{H}_{[k]}$ are denoted by $\mathbf{Q} = \left(\mathbf{H}^{\mathrm{H}}\mathbf{H} + \alpha\mathbf{I}_M\right)^{-1}$ and $\mathbf{Q}_{[k]} = \left(\mathbf{H}_{[k]}^{\mathrm{H}}\mathbf{H}_{[k]} + \alpha\mathbf{I}_M\right)^{-1}$, with $\alpha > 0$, respectively. It then holds that*

$$\mathbf{Q} = \mathbf{Q}_{[k]} - \frac{1}{M} \frac{\mathbf{Q}_{[k]}\boldsymbol{h}_k\boldsymbol{h}_k^{\mathrm{H}}\mathbf{Q}_{[k]}}{1 + \frac{1}{M}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{Q}_{[k]}\boldsymbol{h}_k} \tag{30}$$

and

$$\boldsymbol{h}_k^{\mathrm{H}}\mathbf{Q} = \frac{\boldsymbol{h}_k^{\mathrm{H}}\mathbf{Q}_{[k]}}{1 + \frac{1}{M}\boldsymbol{h}_k^{\mathrm{H}}\mathbf{Q}_{[k]}\boldsymbol{h}_k}. \tag{31}$$

Lemma 8 ([49,50]). *Let $(\mathbf{A}_N)_{N\geq 1}, \mathbf{A}_N \in \mathbb{C}^{N\times N}$ be a sequence of matrices such that $\limsup \|\mathbf{A}_N\| < \infty$, and $(\mathbf{x}_N)_{N\geq 1}, \mathbf{x}_N \in \mathbb{C}^{N\times 1}$ be a sequence of random vectors of i.i.d. entries of zero mean, unit variance, and finite eighth order moment independent of $\mathbf{A}_N$. Then,*

$$\frac{1}{N}\mathbf{x}_N^{\mathrm{H}}\mathbf{A}_N\mathbf{x}_N - \frac{1}{N}\operatorname{tr}\left(\mathbf{A}_N\right) \xrightarrow[N\to\infty]{a.s.} 0. \tag{32}$$

Lemma 9 ([49,50]). *Let* $(\mathbf{A}_N)_{N\geq 1}, \mathbf{A}_N \in \mathbb{C}^{N\times N}$ *be a sequence of matrices such that* $\limsup \|\mathbf{A}_N\| < \infty$*, and* $\mathbf{x}_N, \mathbf{y}_N$ *be random, mutually independent with i.i.d. entries of zero mean, unit variance, finite eighth order moment, and independent of* $\mathbf{A}_N$*. Then,*

$$\frac{1}{N}\mathbf{x}_N^{\mathrm{H}}\mathbf{A}_N\mathbf{y}_N \xrightarrow[N\to\infty]{a.s.} 0. \tag{33}$$

Lemma 10 ([39,50]). *Let* $\mathbf{Q}$ *and* $\mathbf{Q}_{[k]}$ *be as given in Lemma 7. Then, for any matrix* $\mathbf{A}$*, we have*

$$\operatorname{tr}\left(\mathbf{A}\left(\mathbf{Q}-\mathbf{Q}_{[k]}\right)\right) \leq \|\mathbf{A}\|_2. \tag{34}$$

Lemma 11 ([39,50]). *Let* $\mathbf{U}, \mathbf{V}, \mathbf{\Theta}$ *be of uniformly bounded spectral norm with respect to* N *and let* $\mathbf{V}$ *be invertible. Further, define* $\mathbf{x} = \mathbf{\Theta}^{\frac{1}{2}}\boldsymbol{z}$ *and* $\mathbf{y} = \mathbf{\Theta}^{\frac{1}{2}}\boldsymbol{q}$ *where* $\boldsymbol{z}, \boldsymbol{q} \in \mathbb{C}^N$ *have i.i.d. complex entries of zero mean, variance* $1/N$ *and finite* 8*th order moment and be mutually independent as well as independent of* $\mathbf{U}, \mathbf{V}$*. Define* $c_0, c_1, c_2 \in \mathbb{R}^+$ *such that* $c_0c_1 - c_2^2 \geq 0$*, and let* $u = \frac{1}{N}\operatorname{tr}\left(\mathbf{\Theta}\mathbf{V}^{-1}\right)$ *and* $u' = \frac{1}{N}\operatorname{tr}\left(\mathbf{\Theta}\mathbf{U}\mathbf{V}^{-1}\right)$*. Then we have:*

$$\mathbf{x}^{\mathrm{H}}\mathbf{U}\left(\mathbf{V} + c_0\mathbf{x}\mathbf{x}^{\mathrm{H}} + c_1\mathbf{y}\mathbf{y}^{\mathrm{H}} + c_2\mathbf{x}\mathbf{y}^{\mathrm{H}} + c_2\mathbf{y}\mathbf{x}^{\mathrm{H}}\right)^{-1}\mathbf{x} \tag{35}$$

$$- \frac{u'\left(1+c_1u\right)}{(c_0c_1-c_2^2)u^2 + (c_0+c_1)u + 1} \to 0 \tag{36}$$

as well as

$$\mathbf{x}^{\mathrm{H}}\mathbf{U}\left(\mathbf{V} + c_0\mathbf{x}\mathbf{x}^{\mathrm{H}} + c_1\mathbf{y}\mathbf{y}^{\mathrm{H}} + c_2\mathbf{x}\mathbf{y}^{\mathrm{H}} + c_2\mathbf{y}\mathbf{x}^{\mathrm{H}}\right)^{-1}\mathbf{y} \tag{37}$$

$$- \frac{-c_2uu'}{(c_0c_1-c_2^2)u^2 + (c_0+c_1)u + 1} \to 0. \tag{38}$$

E New Lemmas

Lemma 12. *Consider the channel matrices* $\hat{\mathbf{H}}^{(j)}, \hat{\mathbf{H}}^{(j')}$ *are distributed according to the DCSI model in Section 6.1.2. Let*

$$\mathbf{Q}^{(j)} = \left(\frac{(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}}{M} + \alpha^{(j)}\mathbf{I}_M\right)^{-1} \tag{39}$$

$$\mathbf{Q}^{(j')} = \left(\frac{(\hat{\mathbf{H}}^{(j')})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}}{M} + \alpha^{(j')}\mathbf{I}_M\right)^{-1} \tag{40}$$

with $\alpha^{(j)}, \alpha^{(j')} > 0$. *Let* $\mathbf{X} \in \mathbb{C}^{M\times M}$ *be of uniformly bounded spectral norm with respect to* M. *Then,*

$$\frac{1}{M^2}\operatorname{tr}\left(\mathbf{X}\mathbf{Q}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}^{(j')}\right) - \Gamma^o_{j,j'}(\mathbf{X}) \xrightarrow{a.s.} 0 \tag{41}$$

where the function $\Gamma^o_{j,j'}(\mathbf{X}) : \mathbb{C}^{M\times M} \mapsto \mathbb{C}$ *is defined as*

$$\begin{aligned}\Gamma^o_{j,j'}(\mathbf{X}) = &\frac{1}{M}\sum_{k=1}^{K}\frac{\left(\sqrt{c^{(j)}_{0,k}c^{(j')}_{0,k}} + \sqrt{c^{(j)}_{1,k}c^{(j')}_{1,k}}\rho^{(j,j')}_k\right)\frac{1}{M}\operatorname{tr}\left(\mathbf{\Theta}_k\mathbf{Q}^{(j')}_o\mathbf{X}\mathbf{Q}^{(j)}_o\right)}{(1+m^{(j)}_k)(1+m^{(j')}_k)}\\ &+\frac{1}{M}\sum_{k=1}^{K}\frac{\left(\sqrt{c^{(j)}_{0,k}c^{(j')}_{0,k}} + \sqrt{c^{(j)}_{1,k}c^{(j')}_{1,k}}\rho^{(j,j')}_k\right)^2\Gamma^o_{j,j'}(\mathbf{\Theta}_k)\frac{1}{M}\operatorname{tr}\left(\mathbf{\Theta}_k\mathbf{Q}^{(j')}_o\mathbf{X}\mathbf{Q}^{(j)}_o\right)}{(1+m^{(j)}_k)(1+m^{(j')}_k)},\end{aligned} \tag{42}$$

with

$$c^{(j)}_{0,k} = 1-(\sigma^{(j)}_k)^2, \quad c^{(j)}_{1,k} = (\sigma^{(j)}_k)^2, \quad c^{(j)}_{2,k} = \sigma^{(j)}_k\sqrt{1-(\sigma^{(j)}_k)^2}. \tag{43}$$

$m^{(j)}_k, \mathbf{Q}^{(j)}_o, m^{(j')}_k, \mathbf{Q}^{(j')}_o$ *are defined in Theorem 1 using* $\hat{\mathbf{H}}^{(j)}, \alpha^{(j)}, \hat{\mathbf{H}}^{(j')}, \alpha^{(j')}$ *respectively.* $\Gamma^o_{j,j'}(\mathbf{\Theta}_k),\ k = 1,\ldots,K$ *is the* k*th entry of vector* $\boldsymbol{\gamma} \in \mathbb{C}^{K\times 1}$. *Vector* $\boldsymbol{\gamma}$ *is the solution for equation system*

$$\mathbf{A}\boldsymbol{\gamma} = \boldsymbol{b}. \tag{44}$$

$\mathbf{A} \in \mathbb{C}^{K\times K}$ *with*

$$[\mathbf{A}]_{\ell,t} = 1_{\ell=t} - \frac{\left(\sqrt{c^{(j)}_{0,t}c^{(j')}_{0,t}} + \sqrt{c^{(j)}_{1,t}c^{(j')}_{1,t}}\rho^{(j,j')}_t\right)^2}{M(1+m^{(j)}_t)(1+m^{(j')}_t)}\frac{\operatorname{tr}\left(\mathbf{\Theta}_t\mathbf{Q}^{(j')}_o\mathbf{\Theta}_\ell\mathbf{Q}^{(j)}_o\right)}{M}. \tag{45}$$

$\boldsymbol{b} \in \mathbb{C}^{K\times 1}$ *with*

$$[\boldsymbol{b}]_\ell = \frac{1}{M}\sum_{k=1}^{K}\frac{\sqrt{c^{(j)}_{0,k}c^{(j')}_{0,k}} + \sqrt{c^{(j)}_{1,k}c^{(j')}_{1,k}}\rho^{(j,j')}_k}{(1+m^{(j)}_k)(1+m^{(j')}_k)}\frac{\operatorname{tr}\left(\mathbf{\Theta}_k\mathbf{Q}^{(j')}_o\mathbf{\Theta}_\ell\mathbf{Q}^{(j)}_o\right)}{M}. \tag{46}$$

Proof. We start by introducing

$$\mathbf{Q}^{(j)}_{[k]} = \left(\frac{(\hat{\mathbf{H}}^{(j)}_{[k]})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}_{[k]}}{M} + \alpha^{(j)}\mathbf{I}_M\right)^{-1} \tag{47}$$

with

$$\hat{\mathbf{H}}_{[k]}^{(j)} = \left[\hat{\boldsymbol{h}}_1^{(j)} \quad \dots \quad \hat{\boldsymbol{h}}_{k-1}^{(j)} \quad \hat{\boldsymbol{h}}_{k+1}^{(j)} \quad \dots \quad \hat{\boldsymbol{h}}_K^{(j)}\right]^{\mathrm{H}}. \tag{48}$$

$\mathbf{Q}_{[k]}^{(j')}$ and $\hat{\mathbf{H}}_{[k]}^{(j')}$ are defined respectively in similar manner as $\mathbf{Q}_{[k]}^{(j)}$ and $\hat{\mathbf{H}}_{[k]}^{(j)}$. Let us start by writing the simple equality

$$\begin{aligned}
&\mathbf{Q}^{(j)} - \mathbf{Q}_o^{(j)} \\
=&\mathbf{Q}_o^{(j)} \left((\mathbf{Q}_o^{(j)})^{-1} - (\mathbf{Q}^{(j)})^{-1}\right) \mathbf{Q}^{(j)} \\
=&\mathbf{Q}_o^{(j)} \left(\frac{1}{M}\sum_{k=1}^{K} \frac{\boldsymbol{\Theta}_k}{1+m_k^{(j)}} - \frac{(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\mathbf{H}^{(j)}}{M}\right) \mathbf{Q}^{(j)}.
\end{aligned} \tag{49}$$

We can then replace $\mathbf{Q}^{(j)}$ using (49) to obtain

$$\begin{aligned}
&\frac{1}{M^2}\operatorname{tr}\left(\mathbf{X}\mathbf{Q}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}^{(j')}\right) \\
=&\frac{1}{M^2}\operatorname{tr}\left(\mathbf{X}\mathbf{Q}_o^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}^{(j')}\right) \\
&+\sum_{k=1}^{K}\frac{\operatorname{tr}\left(\mathbf{X}\mathbf{Q}_o^{(j)}\boldsymbol{\Theta}_k\mathbf{Q}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}^{(j')}\right)}{M^3\left(1+m_k^{(j)}\right)} \\
&-\frac{1}{M^3}\operatorname{tr}\left(\mathbf{X}\mathbf{Q}_o^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}\mathbf{Q}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}^{(j')}\right) \\
&= Z_1 + Z_2 + Z_3.
\end{aligned} \tag{50}$$

We will now calculate separately each of the term Z_i. Starting with Z_1 gives

$$Z_1 = \frac{1}{M^2} \operatorname{tr}\left(\mathbf{X}\mathbf{Q}_o^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}^{(j')}\right) \tag{51}$$

$$= \frac{1}{M}\sum_{k=1}^{K}\frac{1}{M}(\hat{\boldsymbol{h}}_k^{(j')})^{\mathrm{H}}\mathbf{Q}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\hat{\boldsymbol{h}}_k^{(j)} \tag{52}$$

$$\overset{(a)}{=} \frac{1}{M}\sum_{k=1}^{K}\frac{1}{M}\frac{(\hat{\boldsymbol{h}}_k^{(j')})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\hat{\boldsymbol{h}}_k^{(j)}}{1+\frac{1}{M}(\hat{\boldsymbol{h}}_k^{(j')})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j')}\hat{\boldsymbol{h}}_k^{(j')}} \tag{53}$$

$$\overset{(b)}{\asymp} \frac{1}{M}\sum_{k=1}^{K}\frac{\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right)\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_{[k]}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{1+\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_{[k]}^{(j')}\right)} \tag{54}$$

$$\overset{(c)}{\asymp} \frac{1}{M}\sum_{k=1}^{K}\frac{\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right)\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{1+\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}^{(j')}\right)} \tag{55}$$

$$\overset{(d)}{\asymp} \frac{1}{M}\sum_{k=1}^{K}\frac{\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right)\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_0^{(j')}\mathbf{X}\mathbf{Q}_0^{(j)}\right)}{1+m_k^{(j')}}, \tag{56}$$

where equality (a) follows from Lemma 7, equality (b) from Lemma 8, equality (c) from Lemma 10, and equality (d) from the fundamental Theorem 1. The following calculations are very similar and the same lemmas are used.

Turning to Z_3 gives

$$Z_3 = -\frac{1}{M^3}\operatorname{tr}\left(\mathbf{X}\mathbf{Q}_o^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}\mathbf{Q}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}^{(j')}\right) \tag{57}$$

$$= -\frac{1}{M^3}\sum_{k=1}^{K}\operatorname{tr}\left((\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\hat{\boldsymbol{h}}_k^{(j)}\right) \tag{58}$$

$$= -\frac{1}{M^3}\sum_{k=1}^{K}\frac{\operatorname{tr}\left((\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\hat{\boldsymbol{h}}_k^{(j)}\right)}{1+\frac{1}{M}(\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j)}\hat{\boldsymbol{h}}_k^{(j)}} \tag{59}$$

$$\overset{(e)}{=} -\frac{1}{M^3}\sum_{k=1}^{K}\frac{\operatorname{tr}\left((\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}_{[k]}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\hat{\boldsymbol{h}}_k^{(j)}\right)}{1+\frac{1}{M}(\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j)}\hat{\boldsymbol{h}}_k^{(j)}}$$
$$+\frac{1}{M^4}\sum_{k=1}^{K}\frac{\operatorname{tr}\left((\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}_{[k]}^{(j')}\hat{\boldsymbol{h}}_k^{(j')}(\hat{\boldsymbol{h}}_k^{(j')})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\hat{\boldsymbol{h}}_k^{(j)}\right)}{\left(1+\frac{1}{M}(\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j)}\hat{\boldsymbol{h}}_k^{(j)}\right)\left(1+\frac{1}{M}(\hat{\boldsymbol{h}}_k^{(j')})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j')}\hat{\boldsymbol{h}}_k^{(j')}\right)} \tag{60}$$
$$= Z_4 + Z_5, \tag{61}$$

with equality (e) obtained using Lemma 7 for $\mathbf{Q}^{(j')}$. We also split the calculation in two and start by calculating Z_4 as follows.

$$\begin{aligned}
Z_4 &= -\frac{1}{M^3}\sum_{k=1}^{K}\frac{\operatorname{tr}\left((\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j)}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{[k]}^{(j')}\mathbf{Q}_{[k]}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\hat{\boldsymbol{h}}_k^{(j)}\right)}{1+\frac{1}{M}(\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j)}\hat{\boldsymbol{h}}_k^{(j)}} \\
&\quad -\frac{1}{M^3}\sum_{k=1}^{K}\frac{\operatorname{tr}\left((\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j)}\hat{\boldsymbol{h}}_k^{(j)}(\hat{\boldsymbol{h}}_k^{(j')})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\hat{\boldsymbol{h}}_k^{(j)}\right)}{1+\frac{1}{M}(\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j)}\hat{\boldsymbol{h}}_k^{(j)}} \qquad (62) \\
&\asymp -\frac{1}{M^3}\sum_{k=1}^{K}\frac{\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_{[k]}^{(j)}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{[k]}^{(j')}\mathbf{Q}_{[k]}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{1+\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_{[k]}^{(j)}\right)} \\
&\quad -\frac{1}{M}\sum_{k=1}^{K}\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right) \\
&\qquad\qquad \cdot\frac{\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_{[k]}^{(j)}\right)\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_{[k]}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{1+\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}'_{[k]}\right)} \qquad (63) \\
&\overset{(f)}{\asymp} -\frac{1}{M}\sum_{k=1}^{K}\frac{\frac{1}{M^2}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{1+m_k^{(j)}} \\
&\quad -\frac{1}{M}\sum_{k=1}^{K}\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right)\frac{m_k^{(j)}\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_0^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{1+m_k^{(j)}} \qquad (64) \\
&\asymp -Z_2 - \frac{1}{M}\sum_{k=1}^{K}\frac{\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right)\frac{m_k^{(j)}}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_0^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{1+m_k^{(j)}}, \qquad (65)
\end{aligned}$$

where (f) applies multiple times Lemma 10. Finally, Z_5 is calculated as

$$Z_5 \asymp \frac{1}{M^4}\sum_{k=1}^{K}\frac{\operatorname{tr}\left((\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}_{[k]}^{(j')}\hat{\boldsymbol{h}}_k^{(j')}(\hat{\boldsymbol{h}}_k^{(j')})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\hat{\boldsymbol{h}}_k^{(j)}\right)}{\left(1+m_k^{(j)}\right)\left(1+m_k^{(j')}\right)} \tag{66}$$

$$= \frac{1}{M^4}\sum_{k=1}^{K}\frac{\operatorname{tr}\left((\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j)}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{[k]}^{(j')}\mathbf{Q}_{[k]}^{(j')}\hat{\boldsymbol{h}}_k^{(j')}(\hat{\boldsymbol{h}}_k^{(j')})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\hat{\boldsymbol{h}}_k^{(j)}\right)}{\left(1+m_k^{(j)}\right)\left(1+m_k^{(j')}\right)}$$
$$+ \frac{1}{M^4}\sum_{k=1}^{K}\frac{\operatorname{tr}\left((\hat{\boldsymbol{h}}_k^{(j)})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j)}\hat{\boldsymbol{h}}_k^{(j)}(\hat{\boldsymbol{h}}_k^{(j')})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j')}\hat{\boldsymbol{h}}_k^{(j')}(\hat{\boldsymbol{h}}_k^{(j')})^{\mathrm{H}}\mathbf{Q}_{[k]}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\hat{\boldsymbol{h}}_k^{(j)}\right)}{\left(1+m_k^{(j)}\right)\left(1+m_k^{(j')}\right)} \tag{67}$$

$$\asymp \frac{1}{M}\sum_{k=1}^{K}\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right)^2$$
$$\cdot\frac{\frac{1}{M^2}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_{[k]}^{(j)}(\hat{\mathbf{H}}_{[k]}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_{[k]}^{(j')}\mathbf{Q}_{[k]}^{(j')}\right)\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_{[k]}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{\left(1+m_k^{(j)}\right)\left(1+m_k^{(j')}\right)}$$
$$+\frac{1}{M}\sum_{k=1}^{K}\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right)$$
$$\cdot\frac{\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_{[k]}^{(j)}\right)\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_{[k]}^{(j')}\right)\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_{[k]}^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{\left(1+m_k^{(j)}\right)\left(1+m_k^{(j')}\right)} \tag{68}$$

$$\asymp \frac{1}{M}\sum_{k=1}^{K}\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right)^2$$
$$\cdot\frac{\frac{1}{M^2}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}^{(j')}\right)\frac{1}{M}\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_o^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{\left(1+m_k^{(j)}\right)\left(1+m_k^{(j')}\right)}$$
$$+\frac{1}{M}\sum_{k=1}^{K}\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right)\frac{m_k^{(j)}m_k^{(j')}\frac{\operatorname{tr}\left(\boldsymbol{\Theta}_k\mathbf{Q}_o^{(j)}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{M}}{\left(1+m_k^{(j)}\right)\left(1+m_k^{(j')}\right)}. \tag{69}$$

Add all the Z_i gives

$$\frac{1}{M^2}\operatorname{tr}\left(\mathbf{X}\mathbf{Q}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}^{(j')}\right) \tag{70}$$

$$\asymp \frac{1}{M}\sum_{k=1}^{K}\frac{\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right)\frac{1}{M}\operatorname{tr}\left(\mathbf{\Theta}_k\mathbf{Q}_o^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{(1+m_k^{(j)})(1+m_k^{(j')})}$$

$$+\frac{1}{M}\sum_{k=1}^{K}\frac{1}{M^2}\operatorname{tr}\left(\mathbf{\Theta}_k\mathbf{Q}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j')}\mathbf{Q}^{(j')}\right)$$

$$\cdot\frac{\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right)^2\frac{1}{M}\operatorname{tr}\left(\mathbf{\Theta}_k\mathbf{Q}_o^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{(1+m_k^{(j)})(1+m_k^{(j')})}, \tag{71}$$

which yields

$$\Gamma_{j,j'}^{o}(\mathbf{X})=\frac{1}{M}\sum_{k=1}^{K}\frac{\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right)\frac{1}{M}\operatorname{tr}\left(\mathbf{\Theta}_k\mathbf{Q}_o^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{(1+m_k^{(j)})(1+m_k^{(j')})}$$

$$+\frac{1}{M}\sum_{k=1}^{K}\frac{\left(\sqrt{c_{0,k}^{(j)}c_{0,k}^{(j')}}+\sqrt{c_{1,k}^{(j)}c_{1,k}^{(j')}}\rho_k^{(j,j')}\right)^2\Gamma_{j,j'}^{o}(\mathbf{\Theta}_k)\frac{1}{M}\operatorname{tr}\left(\mathbf{\Theta}_k\mathbf{Q}_o^{(j')}\mathbf{X}\mathbf{Q}_o^{(j)}\right)}{(1+m_k^{(j)})(1+m_k^{(j')})}. \tag{72}$$

It remains then to calculate $\Gamma_{j,j'}^{o}(\mathbf{\Theta}_k)$ to conclude the calculation. Indeed, it is the solution of equation system when asserting $\mathbf{X}=\mathbf{\Theta}_k, \forall k=1,\ldots,K$ into (72). □

Lemma 13. *Let $\mathbf{L},\mathbf{R},\bar{\mathbf{A}},\mathbf{\Theta}\in\mathbb{C}^{M\times M}$ be of uniformly bounded spectral norm with respect to M and let $\bar{\mathbf{A}}$ be invertible. Further define $\mathbf{x}=\mathbf{\Theta}^{\frac{1}{2}}\mathbf{z}$, $\mathbf{x}'=\mathbf{\Theta}^{\frac{1}{2}}\mathbf{z}'$ and $\mathbf{y}=\mathbf{\Theta}^{\frac{1}{2}}\mathbf{q}$. $\mathbf{z},\mathbf{z}'$ satisfies $\mathbf{z}=\rho\mathbf{z}'+\sqrt{1-\rho^2}\mathbf{w}$. $\mathbf{z},\mathbf{q}$ and $\mathbf{z}',\mathbf{q},\mathbf{w}$ are mutually independent as well as independent of $\mathbf{L},\mathbf{R},\bar{\mathbf{A}}$. $\mathbf{z},\mathbf{z}',\mathbf{q},\mathbf{w}$ have i.i.d. complex entries of zero mean, variance $1/M$ and finite 8th order moment. Let us define*

$$\mathbf{A}=\bar{\mathbf{A}}+c_0\mathbf{x}\mathbf{x}^{\mathrm{H}}+c_1\mathbf{y}\mathbf{y}^{\mathrm{H}}+c_2\mathbf{x}\mathbf{y}^{\mathrm{H}}+c_2\mathbf{y}\mathbf{x}^{\mathrm{H}}$$
$$\mathbf{A}'=\bar{\mathbf{A}}+c_0\mathbf{x}'\mathbf{x}'^{\mathrm{H}}+c_1\mathbf{y}\mathbf{y}^{\mathrm{H}}+c_2\mathbf{x}'\mathbf{y}^{\mathrm{H}}+c_2\mathbf{y}\mathbf{x}'^{\mathrm{H}},$$

let $c_0, c_1, c_2 \in \mathbb{R}^+$ *with* $c_0 + c_1 = 1$ *and* $c_0 c_1 - c_2^2 = 0$, *and*

$$u = \frac{\operatorname{tr}(\boldsymbol{\Theta}\bar{\mathbf{A}}^{-1})}{M}, \qquad u_{\mathrm{L}} = \frac{\operatorname{tr}(\boldsymbol{\Theta}\mathbf{L}\bar{\mathbf{A}}^{-1})}{M},$$
$$u_{\mathrm{R}} = \frac{\operatorname{tr}(\boldsymbol{\Theta}\bar{\mathbf{A}}^{-1}\mathbf{R})}{M}, \qquad u_{\mathrm{LR}} = \frac{\operatorname{tr}(\boldsymbol{\Theta}\mathbf{L}\bar{\mathbf{A}}^{-1}\mathbf{R})}{M}.$$

Then we have:

$$\begin{aligned}
\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\mathbf{R}\mathbf{x} &\asymp u_{\mathrm{LR}} - \frac{c_0 u_{\mathrm{L}} u_{\mathrm{R}}}{1+u} \\
\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\mathbf{R}\mathbf{y} &\asymp -\frac{c_2 u_{\mathrm{L}} u_{\mathrm{R}}}{1+u} \\
\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}'^{-1}\mathbf{R}\mathbf{y} &\asymp -\rho\frac{c_2 u_{\mathrm{L}} u_{\mathrm{R}}}{1+u}.
\end{aligned}$$

Proof. Focusing first on the first equality gives

$$\begin{aligned}
&\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\mathbf{R}\mathbf{x} - \mathbf{x}^{\mathrm{H}}\mathbf{L}\bar{\mathbf{A}}^{-1}\mathbf{R}\mathbf{x} \\
&= \mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\left(\bar{\mathbf{A}} - \mathbf{A}\right)\bar{\mathbf{A}}^{-1}\mathbf{R}\mathbf{x} \\
&= -\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\left(c_0\mathbf{x}\mathbf{x}^{\mathrm{H}} + c_1\mathbf{y}\mathbf{y}^{\mathrm{H}} + c_2\mathbf{y}\mathbf{x}^{\mathrm{H}} + c_2\mathbf{x}\mathbf{y}^{\mathrm{H}}\right)\bar{\mathbf{A}}^{-1}\mathbf{R}\mathbf{x} \\
&\overset{(a)}{\asymp} -\left(c_0\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\mathbf{x} + c_2\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\mathbf{y}\right)\operatorname{tr}\left(\boldsymbol{\Theta}\bar{\mathbf{A}}^{-1}\mathbf{R}\right) \\
&\overset{(b)}{\asymp} -c_0\frac{\operatorname{tr}\left(\boldsymbol{\Theta}\mathbf{L}\bar{\mathbf{A}}^{-1}\right)}{M}\frac{\operatorname{tr}\left(\boldsymbol{\Theta}\bar{\mathbf{A}}^{-1}\mathbf{R}\right)}{M}\frac{1 + c_1\frac{\operatorname{tr}\left(\boldsymbol{\Theta}\bar{\mathbf{A}}^{-1}\right)}{M}}{1 + \frac{\operatorname{tr}\left(\boldsymbol{\Theta}\bar{\mathbf{A}}^{-1}\right)}{M}} \\
&\quad + c_2^2\frac{\operatorname{tr}\left(\boldsymbol{\Theta}\mathbf{L}\bar{\mathbf{A}}^{-1}\right)}{M}\frac{\operatorname{tr}\left(\boldsymbol{\Theta}\bar{\mathbf{A}}^{-1}\mathbf{R}\right)}{M}\frac{\frac{\operatorname{tr}\left(\boldsymbol{\Theta}\bar{\mathbf{A}}^{-1}\right)}{M}}{1 + \frac{\operatorname{tr}\left(\boldsymbol{\Theta}\bar{\mathbf{A}}^{-1}\right)}{M}},
\end{aligned}$$

where equality (a) is obtained from using Lemma 9 and Lemma 8 and equality (b) follows from Lemma 11. Similarly, we turn to the second equality to

write

$$
\begin{aligned}
&\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\mathbf{R}\mathbf{y}-\mathbf{x}^{\mathrm{H}}\mathbf{L}\bar{\mathbf{A}}^{-1}\mathbf{R}\mathbf{y}\\
=&\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\left(\bar{\mathbf{A}}-\mathbf{A}\right)\bar{\mathbf{A}}^{-1}\mathbf{R}\mathbf{y}\\
=&-\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\left(c_0\mathbf{x}\mathbf{x}^{\mathrm{H}}+c_1\mathbf{y}\mathbf{y}^{\mathrm{H}}+c_2\mathbf{y}\mathbf{x}^{\mathrm{H}}+c_2\mathbf{x}\mathbf{y}^{\mathrm{H}}\right)\bar{\mathbf{A}}^{-1}\mathbf{R}\mathbf{y}\\
\overset{(c)}{\asymp}&-\left(c_1\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\mathbf{y}+c_2\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}^{-1}\mathbf{x}\right)\frac{\mathrm{tr}\left(\mathbf{\Theta}\bar{\mathbf{A}}^{-1}\mathbf{R}\right)}{M}\\
\overset{(d)}{\asymp}&c_1c_2\frac{\mathrm{tr}\left(\mathbf{\Theta}\mathbf{L}\bar{\mathbf{A}}^{-1}\right)}{M}\frac{\mathrm{tr}\left(\mathbf{\Theta}\bar{\mathbf{A}}^{-1}\mathbf{R}\right)}{M}\frac{\frac{\mathrm{tr}\left(\mathbf{\Theta}\bar{\mathbf{A}}^{-1}\right)}{M}}{1+\frac{\mathrm{tr}\left(\mathbf{\Theta}\bar{\mathbf{A}}^{-1}\right)}{M}}\\
&-c_2\frac{\mathrm{tr}\left(\mathbf{\Theta}\mathbf{L}\bar{\mathbf{A}}^{-1}\right)}{M}\frac{\mathrm{tr}\left(\mathbf{\Theta}\bar{\mathbf{A}}^{-1}\mathbf{R}\right)}{M}\frac{1+c_1\frac{\mathrm{tr}\left(\mathbf{\Theta}\bar{\mathbf{A}}^{-1}\right)}{M}}{1+\frac{\mathrm{tr}\left(\mathbf{\Theta}\bar{\mathbf{A}}^{-1}\right)}{M}},
\end{aligned}
$$

where equality (c) is obtained from using Lemma 9 and Lemma 8 and equality (d) follows from Lemma 11. For the third equality,

$$
\begin{aligned}
&\mathbf{x}^{\mathrm{H}}\mathbf{L}\mathbf{A}'^{-1}\mathbf{R}\mathbf{y}\\
=&\rho\mathbf{x}'\mathbf{L}\mathbf{A}'^{-1}\mathbf{R}\mathbf{y}+\sqrt{1-\rho^2}\mathbf{\Theta}^{\frac{1}{2}}\mathbf{w}\mathbf{L}\mathbf{A}'^{-1}\mathbf{R}\mathbf{y}\\
\overset{(e)}{\asymp}&\rho\mathbf{x}'\mathbf{L}\mathbf{A}'^{-1}\mathbf{R}\mathbf{y}\\
\asymp&-\rho\frac{c_2u_{\mathrm{L}}u_{\mathrm{R}}}{1+u},
\end{aligned}
$$

□

where equality (e) is obtained from using Lemma 9.

Bibliography

[1] D. Gesbert, S. Hanly, H. Huang, S. Shamai (Shitz), O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: a new look at interference," *IEEE J. Sel. Areas Commun.*, 2010.

[2] 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.

[3] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, 2010.

[4] J. Koppenborg, H. Halbauer, S. Saur, and C. Hoek, "3D beamforming trials with an active antenna array," in *Proc. International ITG Workshop on Smart Antennas (WSA)*, 2012.

[5] Y.-S. Cheng and C.-H. Chen, "A novel 3D beamforming scheme for LTE-Advanced system," in *Proc. Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 2014.

[6] R. Zakhour and D. Gesbert, "Distributed multicell-MISO precoding using the layered virtual SINR framework," *IEEE Trans. Wireless Commun.*, 2010.

[7] E. Björnson, G. Zheng, M. Bengtsson, and B. Ottersten, "Robust monotonic optimization framework for multicell MISO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2508–2523, 2012.

[8] M. C. Filippou, D. Gesbert, and G. A. Ropokis, "Optimal combining of instantaneous and statistical CSI in the SIMO interference channel," in *Proc. IEEE Vehicular Technology Conference (VTC Spring)*, 2013.

[9] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, 2010.

[10] W. W. L. Ho, T. Q. S. Quek, S. Sun, and R. W. Heath, "Decentralized precoding for multicell MIMO downlink," *IEEE Trans. Wireless Commun.*, 2011.

[11] A. Tolli, H. Pennanen, and P. Komulainen, "Decentralized minimum power multi-cell beamforming with limited backhaul signaling," *IEEE Trans. Wireless Commun.*, 2011.

[12] S. Peters and R. Heath, "Cooperative algorithms for MIMO interference channels," *IEEE Trans. Veh. Technol.*, 2011.

[13] P. Marsch and G. Fettweis, "On downlink network mimo under a constrained backhaul and imperfect channel knowledge," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2009.

[14] E. Björnson, R. Zakhour, D. Gesbert, and B. Ottersten, "Cooperative multicell precoding: Rate region characterization and distributed strategies with instantaneous and statistical CSI," *IEEE Trans. Signal Process.*, 2010.

[15] S. Shamai and B. M. Zaidel, "Enhancing the cellular downlink capacity via co-processing at the transmitting end," in *Proc. IEEE Vehicular Technology Conference (VTC Spring)*, 2001.

[16] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, 2008.

[17] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, "Networked MIMO with clustered linear precoding," *IEEE Trans. on Wireless Commun.*, 2009.

[18] A. Sanderovich, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Uplink macro diversity of limited backhaul cellular network," *IEEE Trans. Inf. Theory*, vol. 55, no. 8, pp. 3457–3478, 2009.

[19] O. Simeone, O. Somekh, H. V. Poor, and S. Shamai (Shitz), "Downlink multicell processing with limited-backhaul capacity," *EURASIP Journal on Advances in Signal Processing*, 2009.

[20] R. Zakhour and D. Gesbert, "Optimized data sharing in multicell MIMO with finite backhaul capacity," *IEEE Trans. Signal Process.*, 2011.

[21] J. Zhao, T. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," *IEEE Trans. Wireless Commun.*, 2013.

[22] T. Lakshmana, A. Tölli, R. Devassy, and T. Svensson, "Precoder design with limited feedback and backhauling for joint transmission," *arXiv preprint arXiv:1503.07590*, 2015.

[23] F. Iutzeler, P. Bianchi, P. Ciblat, and W. Hachem, "Explicit convergence rate of a distributed alternating direction method of multipliers," *arXiv preprint arXiv:1312.1085*, 2013.

[24] A. Nedic, A. Ozdaglar, and P. Parrilo, "Constrained consensus and optimization in multi-agent networks," *IEEE Trans. Automat. Contr.*, vol. 55, 2010.

[25] I. Schizas, A. Ribeiro, and G. Giannakis, "Consensus in Ad Hoc WSNs with noisy links-Part I: distributed estimation of deterministic signals," *IEEE Trans. Signal Process.*, vol. 56, no. 1, 2008.

[26] F. Iutzeler, P. Bianchi, P. Ciblat, and W. Hachem, "Asynchronous distributed optimization using a randomized alternating direction method of multipliers," in *Proc. IEEE Annual Conference onDecision and Control (CDC)*, 2013.

[27] P. Komulainen, A. Tolli, and M. Juntti, "Effective CSI signaling and decentralized beam coordination in TDD multi-cell MIMO systems," *IEEE Trans. Signal Process.*, vol. 61, no. 9, pp. 2204–2218, 2013.

[28] H. Asgharimoghaddam, A. Tölli, L. Sanguinetti, and M. Debbah, "Decentralized multi-cell beamforming with QoS guarantees via large system analysis," in *proc. IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, 2015.

[29] S. Kaviani, O. Simeone, W. A. Krzymien, and S. Shamai (Shitz), "Linear precoding and equalization for network MIMO with partial cooperation," *IEEE Trans. Veh. Technol.*, 2012.

[30] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, "Five disruptive technology directions for 5G," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 74–80, 2014.

[31] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Communications*, 2006.

[32] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "A distributed numerical approach to interference alignment and applications to wireless interference networks," *IEEE Trans. Inf. Theo.*, 2011.

[33] H. Bolcskei and I. J. Thukral, "Interference alignment with limited feedback," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2009.

[34] R. Krishnamachari and M. Varanasi, "Interference alignment under limited feedback for MIMO interference channels," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2010.

[35] O. E. Ayach and R. W. Heath, "Interference alignment with analog channel state feedback," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 626–636, Feb. 2012.

[36] M. Rezaee and M. Guillaud, "Limited feedback for interference alignment in the K-user MIMO interference channel," in *Proc. IEEE Information Theory Workshop (ITW)*, 2012.

[37] X. Rao, L. Ruan, and V. K. N. Lau, "CSI feedback reduction for MIMO interference alignment," *IEEE Trans. Signal Process.*, 2013.

[38] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, 2006.

[39] 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*, 2012.

[40] M. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Trans. Inf. Theory*, 2012.

[41] M. Rezaee, M. Guillaud, and F. Lindqvist, "CSIT sharing over finite capacity backhaul for spatial interference alignment," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2013.

[42] P. Marsch and G. Fettweis, "On multicell cooperative transmission in backhaul-constrained cellular systems," *Annals of Telecommunications*, vol. 63, no. 5, 2008.

[43] T. R. Lakshmana, C. Botella, and T. Svensson, "Partial joint processing with efficient backhauling using particle swarm optimization," *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, pp. 1–18, 2012.

[44] R. Apelfrojd and M. Sternad, "Robust linear precoder for coordinated multipoint joint transmission under limited backhaul with imperfect CSI," in *Proc. IEEE International Symposium on Wireless Communication Systems (ISWCS)*, 2014.

[45] P. de Kerret and D. Gesbert, "Degrees of freedom of the network MIMO channel with distributed CSI," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6806–6824, Nov. 2012.

[46] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication-part I: channel inversion and regularization," *IEEE Trans. on Commun.*, vol. 53, no. 1, pp. 195–202, 2005.

[47] B. Hochwald and S. Vishwanath, "Space-time multiple access: Linear growth in the sum rate," in *Proc. Allerton Conference on Communication, Control, and Computing (Allerton)*, 2002.

[48] A. Tulino and S. Verdu, *Random matrix theory and wireless communications*. Now Publisher Inc., 2004.

[49] A. Müller, A. Kammoun, E. Björnson, and M. Debbah, "Linear precoding based on polynomial expansion: reducing complexity in massive MIMO," 2013. [Online]. Available: http://arxiv.org/abs/1310.1806

[50] R. Couillet and M. Debbah, *Random matrix methods for wireless Communications*. Cambridge University Press, 2011.

[51] L. Sanguinetti, R. Couillet, and M. Debbah, "Base station cooperation for power minimization in the downlink: Large system analysis," in *IEEE Global Communications Conference (GLOBECOM)*, 2015.

[52] J. Hoydis, M. Kobayashi, and M. Debbah, "Optimal channel training in uplink network MIMO systems," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2824–2833, 2011.

[53] P. Marsch and G. Fettweis, "Uplink comp under a constrained backhaul and imperfect channel knowledge," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1730–1742, 2011.

[54] M. Kobayashi, N. Jindal, and G. Caire, "Training and Feedback Optimization for Multiuser MIMO Downlink," *IEEE Trans. Communications*, 2011.

[55] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the MIMO broadcast channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1473–1482, 2008.

[56] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[57] A. Del Coso and S. Simoens, "Distributed compression for the uplink of a backhaul-constrained coordinated cellular network," *arXiv preprint arXiv:0802.0776*, 2008.

[58] Y.-C. Ho and K.-C. Chu, "Team decision theory and information structures in optimal control problems–Part I," *IEEE Trans. Automat. Contr.*, 1972.

[59] Q. Li, D. Gesbert, and N. Gresset, "Joint precoding over a master-slave coordination link," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014.

[60] S. S. Christensen, R. Agarwal, E. Carvalho, and J. M. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, 2008.

[61] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the-user interference channel," *Information Theory, IEEE Transactions on*, vol. 54, no. 8, pp. 3425–3441, 2008.

[62] Q. Li, D. Gesbert, and N. Gresset, "A cooperative channel estimation approach for coordinated multipoint transmission networks," in *Proc.IEEE International Conference on Communication Workshop (ICCW)*, 2015.

[63] R. Fritzsche and G. Fettweis, "Distributed robust sum rate maximization in cooperative cellular networks," in *Proc. IEEE Workshop on Cooperative and Cognitive Mobile Networks (CoCoNet)*, 2013.

[64] S. Wagner, R. Couillet, M. Debbah, and D. T. Slock, "Large system analysis of linear precoding in correlated miso broadcast channels under limited feedback," *Information Theory, IEEE Transactions on*, vol. 58, no. 7, pp. 4509–4537, 2012.

[65] I. Jee and R. Haddaad, "Optimum design of vector-quantized subband codecs," *IEEE Trans. Signal Process.*, 1998.

[66] T. M. Cover and J. A. Thomas, *Elements of information theory.* John Wiley & Sons, 2012.

[67] P. H. Westerink, J. Biemond, and D. E. Boekee, "Scalar quantization error analysis for image subband coding using qmfs," *IEEE Trans. Signal Process.*, vol. 40, no. 2, pp. 421–428, 1992.

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

[69] P. Zador, "Asymptotic quantization error of continuous signals and the quantization dimension," *IEEE Trans. Inf. Theory*, 1982.

[70] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, 1996.

[71] J. Conway and N. Sloane, "Voronoi regions of lattices, second moments of polytopes, and quantization," *IEEE Trans.Info. Theory*, 1982.

[72] D. Rebollo-Monedero, S. Rane, and B. Girod, "Wyner-Ziv quantization and transform coding of noisy sources at high rates," in *Proc. Asilomar Conference*, 2004.

[73] R. S. Garfinkel and G. L. Nemhauser, *Integer programming.* Wiley New York, 1972.

[74] S. Boyd and L. Vandenberghe, *Convex optimization.* Cambridge university press, 2004.

[75] T. Cover and A. Thomas, *Elements of information theory.* Wiley-Interscience, Jul. 2006.

[76] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO Channels," *IEEE Trans. Signal Process.*, 2004.

[77] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for mutual information analysis of larger dimensional multi-antenna channels," *IEEE Trans. Inf. Theory*, 2008.

[78] P. de Kerret, D. Gesbert, and U. Salim, "Large system analysis of joint regularized Zero Forcing precoding with distributed CSIT," in *Proc. IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2015.

[79] Q. Li, P. de Kerret, D. Gesbert, and N. Gresset, "Robust Regularized ZF in Decentralized Broadcast Channel with Correlated CSI Noise," in *Proc. Allerton Conference on Communication, Control, and Computing (Allerton)*, 2015.

[80] A. M. Rubinov and B. M. Glover, *Optimization and Related Topics*. Springer Science & Business Media, 2013, vol. 47.

[81] Y. Jong, "Practical global optimization algorithm for the sum-of-ratios problem," *arXiv preprint arXiv:1207.1153v3*, 2012.

[82] J. Brinkhuis, Z.-Q. Luo, and S. Zhang, "Matrix convex functions with applications to weighted centers for semidefinite programming," Tech. Rep., 2005.

[83] G. Pataki, "On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues," *Mathematics of Operations Research*, vol. 23, no. 2, pp. 339–358, 1998.