Transmitter cooperation with distributed feedback in wireless networks

Paul de Kerret

ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2013-ENST-0089

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité « Communication et Electronique »

présentée et soutenue publiquement par

Paul de Kerret

le 17 décembre 2013

Transmission coopérative dans les réseaux sans-fil avec feedback distribué

Directeur de thèse : **David Gesbert**

Jury

M. Jean-Claude Belfiore, Professeur, Telecom ParisTech	Président du jury
M. Helmut Bölcskei, Professeur, ETH Zürich	Rapporteur
M. Samson Lasaulce, Directeur de Recherche, CNRS, LSS Supelec	Rapporteur
M. Aydin Sezgin, Professeur, Ruhr-Universität à Bochum	Examinateur
M. Petros Elia, Professeur adjoint, EURECOM	Examinateur

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

THÈSE

DISSERTATION

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

Specialization: Communication and Electronics

Paul de Kerret

Transmitter Cooperation with Distributed Feedback in Wireless Networks

Successfully defended the 17th of December 2013 before a committee composed of:

President of the Jury		
Professor	Jean-Claude Belfiore	Telecom ParisTech
Reviewers		
Professor	Helmut Bölcskei	ETH Zürich
Directeur de recherche	Samson Lasaulce	CNRS, LSS Supelec
Examiners		
Professor	Aydin Sezgin	Ruhr-Universität Bochum
Professor	Petros Elia	EURECOM
Thesis Supervisor		
Professor	David Gesbert	EURECOM

THÈSE

présentée pour l'obtention du grade de

Docteur de TELECOM ParisTech

Spécialité: Communication et Electronique

Paul de Kerret

Transmission coopérative dans les réseaux sans-fil avec feedback distribué

Soutenance de thèse prévue le 17 Décembre 2013 devant le jury composé de :

Président du jury		
Professeur	Jean-Claude Belfiore	Telecom ParisTech
Rapporteurs		
Professeur	Helmut Bölcskei	ETH Zürich
Directeur de recherche	Samson Lasaulce	CNRS, LSS Supelec
Examinateurs		
Professeur	Aydin Sezgin	Ruhr-Universität Bochum
Professeur adjoint	Petros Elia	EURECOM
Directeur de thèse		
Professeur	David Gesbert	EURECOM

Abstract

Multiple-antenna "based" transmitter cooperation has been established as a promising tool towards avoiding, aligning, or shaping the interference resulting from aggressive spectral reuse. The price paid in the form of feedback and exchanging channel state information (CSI) between cooperating devices in most existing methods is however often underestimated. In particular, although the impact of imperfect knowledge of the CSI is often investigated, it is usually assumed that the channel estimates are perfectly shared between all the transmitters. This assumption is meaninguful if the cooperating transmitters are colocated or can cooperate via "perfect" backhaul links (e.g., fiber based). It is however not adapted to many practical cases of transmitter cooperation between distant transmitters. Indeed, in many settings (e.g., current mobile networks), each transmitter acquires one part of the full multi-user estimate which has to be shared to the cooperating transmitters via a limited and imperfect backhaul network. This sharing step is already challenging in conventional networks due to the delay in the sharing step, but it becomes even more critical when considering dense networks with a large number of cooperating transmitters, which can only communicate via low cost, backhaul links (e.g., wireless backhaul).

We focus in this thesis on the network scenario where the transmitters would like to cooperate in their transmission but can only imperfectly exchange on CSI which is acquired locally. This imperfect CSI sharing step gives rise to a CSI configuration, denoted as "distributed CSI", where each transmitter has its own imperfect estimate of the global multi-user channel based on which it determines its transmit parameters. We study first the impact of having distributed CSI over the precoder design. Specifically, we show that conventional precoding schemes are not adapted to the distributed CSI configuration and lead to poor performance. Instead, we advocate the use of new, more robust, precoding paradigms.

A problem which is dual to the problem of designing precoders being robust to unequally shared CSI, is the problem of optimally allocating

CSI across space. This corresponds to the question "Who needs to know what?", when it comes to CSI at cooperating transmitters. In contrast to the resource-hungry solution consisting in providing the same CSI to all transmitters, it is shown how a non-uniform spatial allocation of the CSI to the transmitters can provide strong gains depending on the network's topology. Throughout this thesis, both joint precoding across the transmitters and coordinated beamforming (interference alignment) are investigated as our leading examples of cooperation methods in interference limited wireless networks. Finally, interesting and challenging open problems are discussed.

Abrégé

La coopération des transmetteurs dans les système multi-antennes a été reconnue comme un outil prometteur pour éviter ou aligner les interférences résultant d'une réutilisation agressive de la bande spectrale. Le coùt de cette coopération en terme de transmission d'information de canal aux transmetteurs qui coopérent est cependant souvent sous-estimé. En particulier, bien que les conséquences d'une connaissance imparfaite de l'état du canal aux transmetteurs soient souvent étudiées, il est usuellement supposé que les estimées de canal sont parfaitement partagées entre tous les transmetteurs entrant en coopération. Bien que cette hypothèse apparaisse comme raisonnable lorsque les transmetteurs sont localisés au même endroit ou peuvent coopérer via des connections de très bonne qualités n'introduisant que peu d'imperfections, cette hypothèse n'est pas adaptée à de nombreuses situations où des émetteurs éloignés visent à coopérer. En effet, dans de nombreux réseaux sans fil (comme par exemple les réseaux mobiles actuels), chaque transmetteur recoit seulement une partie de l'information relative au canal qui relie tous les emetteurs à tous les receveurs. Cette estimée doit ensuite être partagée à tous les transmetteurs entrant en coopération de manière à rendre possible la méthode de coopération choisie. Cette étape de partage de l'information de canal représente un réel défi dans la mesure où elle peut conduire à l'introduction de délai et/ou d'imperfections, ce qui a des conséquences très importantes sur les performances de la communication. De plus, les réseaux actuels tendent à se densifier et à utiliser des liens bon marché, et donc de qualité médiocre, entre les transmetteurs.

C'est pourquoi nous étudions le cas de réseaux sans-fil où des transmetteurs émettent d'une manière coopérative bien qu'ils ne puissent échanger que d'une manière imparfaite l'information de canal obtenue localement. Ce partage imparfait de l'information de canal donne lieu à une configuration d'information de canal que nous dénotons comme «distribuée». En résumé, dans cette configuration d'information de canal distribuée, chaque transmetteur reçoit une estimée du canal multi-utilisateur qui lui est propre, à partir

de laquelle il détermine ses paramètres de transmission. Nous étudions tout d'abord les conséquences du partage imparfait de l'information de canal sur le précodage. En particulier, nous mettons en évidence l'inefficacité des méthodes conventionnelles de précodage lorsque confrontées à une configuration à information de canal distribuée. Nous passons ensuite à un autre aspect de ce scenario, qui est la détermination de «qui doit savoir quoi», lorsqu'il s'agit de l'information de canal disponible aux transmetteurs engagés dans la coopération. Il est démontré comment une allocation non-uniforme de l'information de canal aux transmetteurs peut donner lieu à des gains importants, en fonction de la géométrie du réseau considéré.

Acknowledgements

My supervisor David Gesbert has been the one who first introduced me to the “team decision” problem, which represents the core of this theis, and I would like to thank him for having given me the opportunity to work on this very interesting problem. Beyond his many impressive qualities in the academic field, I would like to thank him for his qualities as a person. They have created a very enjoyable and relaxed atmosphere in the group thanks to which it is a pleasure to be in Eurecom.

I think that academic research becomes fun only when you have the chance to work together with talented persones. I have had that chance and each of them has taught me a particular aspect of our work and for that I would like to thank them: Maxime Guillaud, Jakob Hoydis, Georg Böcherer, Rajeev Gangula, and Xinping Yi. I would also like to thank especially Professor Kramer for having welcomed me at the LNT inside the TU München and having spent time to discuss with me some interesting problems of multi-user information theory.

Finally, I would like to thank my friends and my family as they make life, exiting, fun, bright and full of trumpets and because without them there would be no point in working on this thesis.

Contents

List of Figures

Acronyms

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

BC	Broadcast Channel.
BS	Base Station.
CB	Coordinated Beamforming.
CSI	Channel State Information.
CSIT	Channel State Information at the Transmitter.
DPC	Dirty Paper Coding.
FDD	Frequency Division Duplexing.
i.i.d.	independent and identically distributed.
IA	Interference Alignment.
IC	Interference Channel.
LHS	left hand side.
LTE	Long Term Evolution.
MIMO	Multiple Input Multiple Output.
MISO	Multiple Input Single Output.
MSE	Mean Square Error.
RHS.	right hand side.
RX	Receiver(s).
SIMO	Single Input Multiple Output.
SINR	Signal to Noise and Interference Ratio.
SNR	Signal-to-Noise Ratio.
SISO	Single-Input Single-Output.
SVD	Singular Value Decomposition.
TDD	Time Division Duplex.
TX	Transmitter(s).
w.l.o.g.	Without loss of generality.
ZF	Zero Forcing.

Notations

$\imath$	The complex number verifying $\imath^2 = -1$.
$\mathbf{0}_K$	Zero matrix of size $K \times K$.
$\mathbf{I}_K$	Identity matrix of size $K \times K$.
$\boldsymbol{e}_i$	ith column of the identity $\mathbf{I}_K$.
$\lvert a \rvert$	Absolute value of the complex number a.
$\lVert \boldsymbol{a} \rVert$	Euclidean norm of the vector $\mathbf{a}$.
$\lVert \mathbf{A} \rVert$	Frobenius norm of the matrix $\mathbf{A}$.
$\{\boldsymbol{a}\}_i$	ith element of the vector $\boldsymbol{a}$.
$\{\mathbf{A}\}_{ij}$	(i,j)th element of the matrix $\mathbf{A}$.
A_{ij}	(i,j)th element of the matrix $\mathbf{A}$.
$\mathrm{tr}\,(\mathbf{A})$	Trace of the square matrix $\mathbf{A}$.
$\det(\mathbf{A})$	Determinant of the square matrix $\mathbf{A}$.
$\odot$	Element-wise product (Hadamard product).
$\mathbf{A}^*$	Complex conjugate of the matrix $\mathbf{A}$.
$\mathbf{A}^{\mathrm{H}}$	Complex conjugate transpose (Hermitian) of the matrix $\mathbf{A}$.
$\mathbf{A}^{\mathrm{T}}$	Transpose of the matrix $\mathbf{A}$.
$\mathbf{A}^{-1}$	Inverse of the matrix $\mathbf{A}$.
$\mathbf{D} = \mathrm{diag}\,(\boldsymbol{a})$	Diagonal matrix with the entries of the vector $\boldsymbol{a}$ along its diagonal.
$\mathbf{D} = \mathrm{diag}\,(\mathbf{A})$	Diagonal matrix with its diagonal elements equal to the diagonal elements of the square matrix $\mathbf{A}$.
$\mathrm{E}\,[x]$	Expected value of the random variable x.
$\mathrm{E}_{\mathcal{A}}[f(a)]$	Expected value of $f(a)$ when a belongs to $\mathcal{A}$, i.e.,$\mathrm{E}[f(a)\lvert a \in \mathcal{A}]$.
$\mathrm{eig}_{\max}(\mathbf{A})$	Maximum eigenvector of the Hermitian matrix $\mathbf{A}$.
$\lambda_{\max}(\mathbf{A})$	Maximum eigenvalue of the Hermitian matrix $\mathbf{A}$.
$\mathrm{eig}_{\min}(\mathbf{A})$	Minimum eigenvector of the Hermitian matrix $\mathbf{A}$.
$\lambda_{\min}(\mathbf{A})$	Minimum eigenvalue of the Hermitian matrix $\mathbf{A}$.

EVD	Eigenbasis for the Hermitian matrix $\mathbf{A}$.
$\Pi_{\mathbf{A}}(\boldsymbol{v})$	Orthogonal projection of the vector $\boldsymbol{v}$ over the space spanned by the columns of the matrix $\mathbf{A}$.
$\Pi^{\perp}_{\mathbf{A}}(\boldsymbol{v})$	Orthogonal projection of the vector $\boldsymbol{v}$ over the orthogonal complement of the space spanned by the columns of the matrix $\mathbf{A}$.
$\bar{i}$	$\bar{i}$ denotes the complementary indice of i when only two users are considered, i.e., $\bar{i} = i \mod 2 + 1$.
δ_{ij}	Kronecker symbol which is equal to 1 if $j = i$ and to zero otherwise.
$\max, \min$	Maximum, minimum.
$\mathcal{N}(0, \sigma^2)$	The complex circularly symmetric Gaussian distribution with mean 0 and variance σ^2.
$\lceil a \rceil$	Ceiling operator. It rounds the real number a to the nearest integer towards infinity.
$\lfloor a \rfloor$	Floor operator. It rounds the real number a to the nearest integer towards minus infinity.
$[a]^+$	Maximum between the real number a and 0.
$\log_2(a)$	Logarithm base 2 of the positive real number a.
$f(x) = o(g(x))$	Represents the fact that $\lim_{x\to\infty} f(x)/g(x) = 0$.
$f(x) = O(g(x))$	Represents the fact that $\lim_{x\to\infty} \lvert f(x)\rvert/\lvert g(x)\rvert \leq a$, with $a > 0$.
$f(x) \sim g(x)$	Represents the fact that $f(x) = g(x) + o(g(x))$.
$\boldsymbol{a} = \boldsymbol{b} + o(x)$	If $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{C}^n$, $x \in \mathbb{R}$, it represents the fact that $\exists \boldsymbol{\tau} \in \mathbb{C}^n, \boldsymbol{a} = \boldsymbol{b} + \boldsymbol{\tau}$ with $\lim_{x\to 0} \frac{\lVert\boldsymbol{\tau}\rVert}{x} = 0$.
$f(x) \doteq x^b$	Exponential equality as in [1]. Denotes that $\lim_{x\to\infty} \frac{\log(f(x))}{\log(x)} = b$.

Chapter 1

Résumé [Français]

1.1 État de l'art pour la coopérations des transmetteurs

1.1.1 Saturation des bandes de fréquence

Les communications sans fil sont devenues essentielles à nos vies par de nombreux aspects et par une grande diversité de services et d'appareils, du téléphone mobile classique, aux ordinateurs et aux tablettes en passant par les capteurs. Le passage à un trafic dominé par le multimédia crée de nouvelles demandes en terme de débit, de délai, et en général en terme d'efficacité spectrale. Dans les réseaux mobiles, la demande en débit a augmenté exponentiellement durant les dix dernières années: En 2012, le volume de données échangées via les réseaux mobiles était égal à 10 fois le volume total de données échangées par internet en 2000. De plus, il est prévu que cette augmentation exponentielle du volume des informations échangées continue dans les années à venir, de telle sorte que la taille totale des échanges sera encore 10 fois plus importante dans 5 ans [2]. Pour répondre à la saturation des ressources disponibles, il est donc nécessaire de repenser l'architecture des réseaux sans fil ainsi que les méthodes de transmission. Certains éléments sont récemment apparus comme des éléments clés pour atteindre les performances escomptées: Un premier élément clé est la densification du réseau avec un accroissement du nombre d'équipements (cf. les études relatives aux petites cellules [3]). Un second élément clé est l'utilisation plus agressive des ressources ce qui permet d'augmenter les ressources pouvant être allouées. En revanche, ces deux éléments n'apportent des améliorations de performance qu'à la condition qu'ils ne soient pas responsables d'un ac-

croissement du niveau d'interférence. En effet, une mauvaise gestion des interférences réduirait considérablement les gains et pourrait même rendre ces changements contre-productifs. C'est pourquoi il est apparu très clairement à la fois dans la communauté académique et dans le milieu industriel l'importance critique d'une gestion efficace des interférences (Voir [4] et ses références).

1.1.2 Transmission à partir d'un transmetteur vers de multiples récepteurs

Ces dernières années, un nombre impressionnant de travaux a porté sur les transmissions à partir d'un transmetteur (TX) vers plusieurs récepteurs (RXs), ce qui modèle les transmissions vers plusieurs utilisateurs dans une cellule unique. Dans la communauté de la théorie de l'information, ce scénario de transmission est bien connu sous le nom de canal de broadcast [5,6]. De nombreux travaux se sont concentrés sur le canal de broadcast que ce soit dans la communauté de la théorie de l'information ou dans l'industrie de telle sorte que ce type de transmission est maintenant relativement bien compris. La capacité du canal de broadcast Gaussien à entrées et sorties multiples (MIMO) a été trouvé [7, 8] et peut être atteinte par une méthode de transmission non-linéaire appelé «Dirty Paper Coding» (DPC) [9]. De plus, les performances obtenues avec des méthodes linéaires de précodage ont été évaluées [10, 11] et les méthodes linéaires de précodage se sont imposés dans la pratiques grâce à leur plus faible complexité et leur efficacité. De nombreux algorithmes de précodage linéaires ont ainsi été développés pour maximiser différents objectifs [12, 13]. En revanche, ces méthodes ne restent efficaces qu'à la condition qu'une information de canal suffisamment précise soit disponible aux transmetteurs [14, 15].

Pour que ces performances théoriques deviennent une réalité, l'estimation du canal et son feedback aux transmetteurs ont été étudiés dans des scénarios réalistes de transmission et des méthodes rendant possible l'obtention d'information de canal précise aux transmetteurs ont été développées (voir [16] et ses références). De plus, les conséquences d'une réduction de la qualité de l'information de canal ont été évaluées [15]. Il a aussi été montré comment la sélection des RXs pouvait être exploitée pour améliorer les performances et rendre les transmissions plus résistantes aux imperfections de l'information de canal [17, 18]. Une analyse détaillée et complète des transmissions dans les systèmes MIMO avec plusieurs RXs dans le cas du précodage linéaire a été réalisée dans [19].

Cependant, même avec ces nouvelles méthodes, il demeure impossible

d'obtenir une information de canal parfaite aux transmetteurs à cause de la nature changeante du canal. En conséquence, des méthodes de transmission plus résistances à une connaissance imparfaite du canal ont été développées. Deux approches ont été particulièrement fructueuses. La première consiste à modéliser la connaissance imparfaite du canal par une variable aléatoire pour ensuite maximiser les performances moyennées sur ces imperfections [20]. Une deuxième approche consiste à considérer un ensemble de réalisations possibles pour ensuite maximiser les performances minimales vis à vis des réalisations possibles [21, 22].

Parallèlement, d'autres travaux ont étudiés comment exploiter des estimées de canaux «retardées», dans le sens où ces estimées sont reçues au transmetteur après un délai tel que ces estimées sont décorrélées avec l'état actuel du canal. Il a été montré dans [23] que même une estimée décorrélée avec l'état actuel du canal pouvait conduire à une amélioration des performances en comparaison avec un système sans aucune information de canal au transmetteur. Ce résultat surprenant a suscité un intérêt très important dans la communauté académique et de nombreux travaux ont depuis étudiés comment exploiter de la meilleure manière possible des estimées de canal «retardées» (Voir [24–26] parmi d'autres).

Enfin, l'usage de transmetteurs avec un très large nombre d'antennes, appelés «MIMO massif», est récemment apparu comme une possibilité intéressante pour améliorer les performances tout en diminuant les besoins en terme de traitement du signal et d'information de canal [27]. Les systèmes MIMO massifs apparaissent maintenant comme une méthode prometteuse et sont le sujet de nombreuses recherches, que ce soit dans l'industrie ou dans les universités (voir [28–30] et leurs références).

1.1.3 Coopération de plusieurs transmetteurs

Bien que les progrès accomplis pour les transmissions à partir d'un seul transmetteur aient permis une très forte amélioration des performances, les débits atteignables restent profondément limités par les interférences entre transmetteurs. Ainsi, la coopération entre transmetteurs est récemment apparue comme un élément clé pour dépasser ces limites et atteindre les performances désirées [4].

Une méthode traditionnelle pour réduire les interférences entre transmetteurs consiste à optimiser conjointement l'allocation de ressources. En particulier, de nombreuses méthodes d'allocation des fréquences ont été proposées dans le but de s'adapter aux interférences et d'améliorer ainsi l'efficacité des transmissions [31–34].

Plus récemment, il a été montré comment la coopération des transmetteurs aux niveau du précodage pouvait apporter une importante augmentation des débits et en particulier éliminer efficacement les interférences. C'est ce type de scénario et de coopération que nous étudions dans cette thèse. En particulier, nous considérons un canal multi-utilisateur avec une propagation dans un milieu sans fil et avec K TXs et K RXs où TX j est équipé avec M_j antennes et RX i avec N_i antennes. Nous définissons le nombre total d'antennes aux TXs

$$M_{\text{tot}} \triangleq \sum_{i=1}^{K} M_i$$

et le nombre total d'antennes aux RXs

$$N_{\text{tot}} \triangleq \sum_{i=1}^{K} N_i.$$

De plus, nous nous restreignons au cas où un seul flot de données est transmis à chaque utilisateur. L'extension au cas général avec plusieurs flots par utilisateur sera en revanche discutée dans le manuscrit. Les transmissions peuvent alors être représentées mathématiquement selon le modèle discret suivant. Le signal reçu au RX i est représenté par $\mathbf{y}_i$ et est égal à [35]

$$\mathbf{y}_i = \mathbf{H}_{ii}^{\text{H}} \mathbf{x}_i + \sum_{j \neq i} \mathbf{H}_{ij}^{\text{H}} \mathbf{x}_j$$

où $\mathbf{x}_j \in \mathbb{C}^{M_j \times 1}$ est le signal émis par le TX j et $\mathbf{H}_{ij}^{\text{H}} \in \mathbb{C}^{N_i \times M_j}$ est la matrice contenant les éléments du canal entre TX j et RX i. Les signaux transmis $\mathbf{x} = [\mathbf{x}_1, \ldots, \mathbf{x}_K]^{\text{T}} \in \mathbb{C}^{M_{\text{tot}} \times 1}$ sont obtenus à partir des données utilisateur à transmettre $\boldsymbol{s} = [s_1, \ldots, s_K]^{\text{T}} \in \mathbb{C}^{K \times 1}$ après multiplication par un précodeur $\mathbf{T} = [\boldsymbol{t}_1, \ldots, \boldsymbol{t}_K] \in \mathbb{C}^{M_{\text{tot}} \times K}$ de telle manière que

$$\mathbf{x} = \mathbf{T}\boldsymbol{s}.$$

Si les données utilisateurs ne sont pas partagées entre TXs, le précodeur $\mathbf{T}$ doit respecter une structure diagonale par blocs, ce qui n'est pas le cas si les données utilisateurs sont partagées. Dans tous les cas, le précodeur $\mathbf{T}$ doit vérifier la contrainte de puissance $\text{E}[\|\mathbf{T}\|_{\text{F}}^2] = KP$. Le filtre $\boldsymbol{g}_i^{\text{H}} \in \mathbb{C}^{1 \times N_i}$ est ensuite appliqué au signal $\boldsymbol{y}_i$ qui est reçu par le RX i pour obtenir une estimée du symbole transmis.

Nous étudions principalement le débit moyen par utilisateur pour mesurer l'efficacité spectrale des transmissions. On suppose que les données des utilisateurs sont distribuées comme $\mathcal{N}_{\mathbb{C}}(0, 1)$ de telle sorte que le débit moyen

de l'utilisateur i est égal à [6]

$$R_i = \mathrm{E}\left[\log_2\left(1 + \frac{|\boldsymbol{g}_i^{\mathrm{H}}\mathbf{H}_i^{\mathrm{H}}\boldsymbol{t}_i|^2}{1+\sum_{j\neq i}|\boldsymbol{g}_i^{\mathrm{H}}\mathbf{H}_i^{\mathrm{H}}\boldsymbol{t}_j|^2}\right)\right]$$

où nous avons défini $\mathbf{H}_i^{\mathrm{H}} \triangleq [\mathbf{H}_{i1}^{\mathrm{H}}, \ldots, \mathbf{H}_{iK}^{\mathrm{H}}]$ qui représente le canal de tous les TXs vers le RX i. De plus, nous nous intéressons aux transmissions à haut rapport signal-à-bruit (SNR) où le nombre de degrés de liberté (DoF) est une mesure intéressante de l'efficacité spectrale. Le DoF de l'utilisateur i, ou coefficient pré-logarithmique de l'utilisateur i, est défini comme [35]

$$\mathrm{DoF}_i \triangleq \lim_{P\to\infty}\frac{R_i}{\log_2(P)}.$$

Bien qu'une mesure imparfaite de l'efficacité de la transmission, l'analyse du DoF permet d'obtenir des résultats analytiques dans des scénarios de transmission compliqués. Les intuitions obtenus à partir de l'analyse du DoF ont été à l'origine de nombreuses innovations qui ont ensuite connus un succès très important: les transmissions MIMO [36], l'alignement d'interférence [37], l'exploitation de l'information de canal retardée [38], etc...

Nous allons maintenant présenter très brièvement les techniques principales de précodage dans le cas où les données utilisateurs sont partagées entre les TXs et lorsque ces données ne sont pas partagées.

Sans partage des données des utilisateurs: Précodage coordonné

L'utilisation de multiples antennes au TX offre de nouvelles opportunités de coopération: Dans le cas où un utilisateur n'est servi que par un seul transmetteur, il est possible de concevoir ce précodeur de telle sorte que les interférences sont réduites, ce qui est appelé «précodage coordonné» [39]. Avec des RXs ayant une seule antenne, cette coordination peut se réaliser sur la base d'information de canal «locale» au transmetteur, au sens où elle est toujours relative au canal entre ce transmetteur et des récepteurs. Dans ce cas, des méthodes efficaces pour optimiser le précodeur ont été trouvés [40–48].

À l'opposé, lorsque les RXs ont plusieurs antennes, la problématique du précodage change totalement. Il devient alors possible d'aligner les interférences dans un nombre limité de dimensions de manière à ce que le RX puisse ensuite éliminer les interférences par l'application d'un filtre linéaire. Cette méthode de transmission est appelée «alignement d'interférence» dans

la communauté [37,38,49] et a suscité un engouement hors-du-commun. Cela provient du fait que l'alignement d'interférence permet d'atteindre le DoF maximal dans de nombreux scénarios [37, 49]. Nous nous intéressons dans cette thèse à la gestion des interférences et donc aux transmissions à haut SNR pour lesquelles l'alignement d'interférence est une technique primordiale.

Une question particulièrement intéressante avec l'alignement d'interférence est la question de la *faisabilité* de l'alignement d'interférence. On dit qu'un canal d'interférence est «faisable» pour l'alignement d'interférence si la configuration d'antennes (i.e., comment les antennes sont allouées aux TXs et aux RXs) offre assez de variables d'optimisation pour assurer une transmission sans interférence de toutes les données utilisateurs. Cela revient à vérifier que [49]

$$\boldsymbol{g}_i^{\mathrm{H}}\mathbf{H}_{ij}\boldsymbol{t}_j = 0, \qquad \forall i, \forall j \neq i.$$

Dans le cas d'un seul flot de données pour chaque utilisateur, il a alors été montré qu'il est possible de vérifier la faisabilité de l'alignement d'interférence en comparant le nombre d'équations engendrées par la suppression des interférences au nombre de variables d'optimisation disponibles [50]. Nous avons reformulé ces résultats dans cette thèse pour obtenir le théorème suivant.

Théorème 1. Le canal d'interférence $\prod_{k=1}^{K}(N_k, M_k)$ est faisable pour l'alignement d'interférence si et seulement si pour tout sous-ensemble de TXs $\mathcal{S}_{\mathrm{TX}}$ et tout sous-ensemble de RXs $\mathcal{S}_{\mathrm{RX}}$, il est vérifié que

$$\mathcal{N}_{\mathrm{var}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}) \geq \mathcal{N}_{\mathrm{eq}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}), \qquad \forall \mathcal{S}_{\mathrm{TX}}, \mathcal{S}_{\mathrm{RX}} \subseteq \mathcal{K}$$

où $\mathcal{N}_{\mathrm{var}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}})$ et $\mathcal{N}_{\mathrm{eq}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}})$ sont respectivement le nombre de variables et le nombre d'équations provenant de l'ensemble de RXs $\mathcal{S}_{\mathrm{RX}}$ et de l'ensemble de TXs $\mathcal{S}_{\mathrm{TX}}$, et sont donc mathématiquement définis comme

$$\begin{aligned}
\mathcal{N}_{\mathrm{var}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}) &\triangleq \sum_{i \in \mathcal{S}_{\mathrm{RX}}} N_i - 1 + \sum_{i \in \mathcal{S}_{\mathrm{TX}}} M_i - 1, \\
\mathcal{N}_{\mathrm{eq}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}) &\triangleq \sum_{k \in \mathcal{S}_{\mathrm{TX}}} \sum_{j \in \mathcal{S}_{\mathrm{RX}}, j \neq k} 1.
\end{aligned}$$

De nombreux algorithmes d'alignement d'interférence ont été développés dans la littérature (Voir [51–53] parmi d'autres) et de nombreux travaux ont analysé les performances qui peuvent être atteintes par l'alignement d'interférence (Voir [49,50,54] et les références à l'intérieur). Avec ces méthodes, les interférences sont supprimées en exploitant à la fois les possibilités de précodage des TXs et les possibilités de filtrage des RXs, ce qui amène des

améliorations considérables des performances. En revanche, cela demande un degré plus important de coordination entre les TXs puisque les TXs doivent s'accorder sur les dimensions où les interférences sont alignées. En conséquence, les algorithmes d'alignement d'interférence présentés dans la littérature demandent que chaque TX dispose de l'information de l'intégralité du canal multi-utilisateur. Ainsi, lorsque les données des utilisateurs ne sont pas partagées, les méthodes d'alignement d'interférence ont le plus fort potentiel mais aussi le besoin le plus important en information de canal aux TXs. C'est pourquoi, les méthodes d'alignement d'interférences seront étudiées en détail dans ce manuscrit.

Avec partage des données des utilisateurs: Précodage conjoint

Quand les données des utilisateurs peuvent être partagées entre les TXs, il est alors possible pour un RX d'être servi conjointement par plusieurs TXs. Ce scénario de transmission où des TXs éloignés partagent les données des utilisateurs de telle sorte qu'il est possible d'appliquer un «précodeur conjoint» est parfois appelé «MIMO virtuel», et est actuellement un candidat pour la future génération de réseaux sans fil [4, 55, 56]. Avec le partage des données des utilisateurs et une connaissance parfaite du canal à tous les TXs, les TXs peuvent alors être vus comme formant un «TX virtuel» servant tous les RXs, comme dans le cas du canal de broadcast. Il devient alors possible d'utiliser un précodeur conjoint $\mathbf{T} \in \mathbb{C}^{M_{\text{tot}} \times K}$.

Nous étudions les performances à haut SNR de telle sorte que les précodeurs qui éliminent complètement les interférences sont particulièrement intéressants. Ces précodeurs sont qualifiés de «zéro-forceur (ZF)». Il en existe de nombreux types, en particulier en fonction du contrôle de puissance aux TXs [57, 58]. Nous considérons dans cette partie le précodeur ZF $\mathbf{T}^\star \triangleq [\boldsymbol{t}_1^\star, \ldots, \boldsymbol{t}_K^\star] \in \mathbb{C}^{K \times K}$ avec

$$\boldsymbol{t}_i^\star \triangleq \sqrt{P} \left(\mathbf{I}_K - \mathbf{H}_i \left(\mathbf{H}_i^{\mathrm{H}} \mathbf{H}_i \right)^{-1} \mathbf{H}_i^{\mathrm{H}} \right) \boldsymbol{h}_i, \qquad \forall i \in \{1, \ldots, K\}$$

où

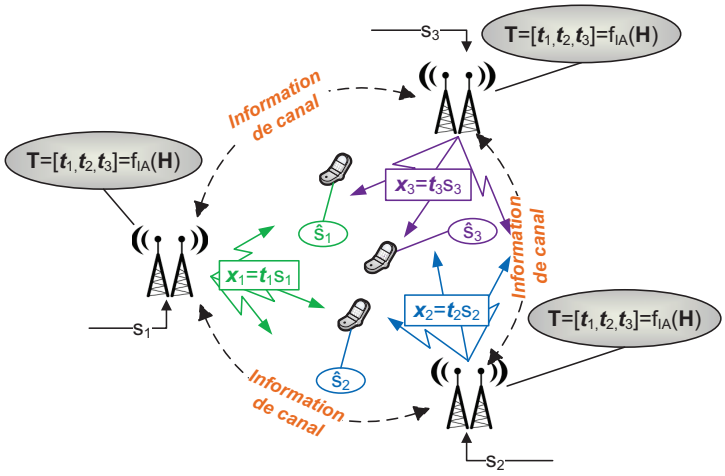
$$\mathbf{H}_i \triangleq \left[\boldsymbol{h}_1 \ \ \ldots \ \ \boldsymbol{h}_{i-1} \ \ \boldsymbol{h}_{i+1} \ \ \ldots \ \ \boldsymbol{h}_K \right], \qquad \forall i \in \{1, \ldots, K\}.$$

et où P correspond à la puissance disponible par utilisateur. Nous utilisons dans ce manuscrit le super-script $\star$ pour représenter le précodeur obtenu à partir d'une connaissance parfaite du canal.

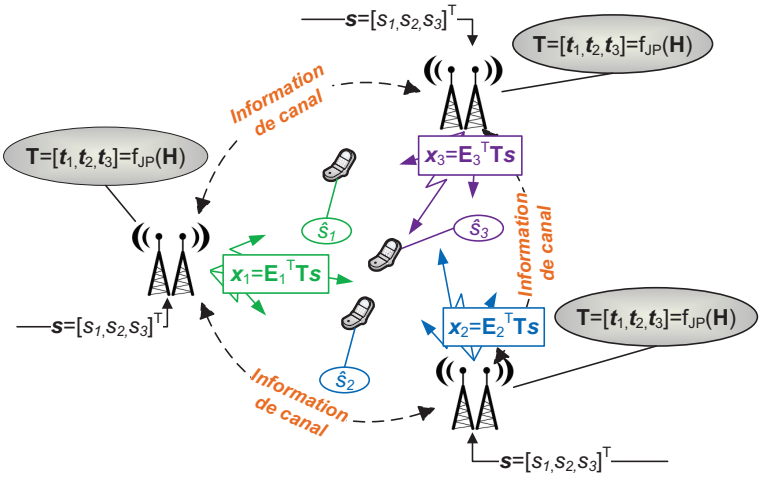
Il est évident que cette méthode de coopération est celle qui promet en théorie les meilleures performances puisque c'est la méthode qui demande le plus d'échange d'information entre les TXs. L'avantage de ce type de

coopération comparée aux approches reposant sur une maximisation égoïste des performances à chaque TX, repose sur le plus faible nombre d'antennes qui est nécessaire aux RXs et aux TXs pour supprimer totalement les interférences. Ce gain est déjà présent pour l'alignement d'interférence et devient encore plus important dans le cas du précodage conjoint. Si nous considérons par exemple un réseau où trois TXs avec deux antennes chacun s'interfèrent. Il est nécessaire que chaque RX aie trois antennes dans le cas où il n'y a pas de coopération. Avec l'alignement d'interférence, seulement deux antennes par RX sont nécessaires alors que le précodage conjoint permet de supprimer les interférences dans le cas de RXs ayant une seule antenne.

Les deux scénarios de transmission (avec et sans partage des données utilisateurs sont représentés dans la Figure 1.1.



(a) Alignement d'interférence



(b) Précodage conjoint

Figure 1.1: Le fonctionnement d'un algorithme de précodage à alignement d'interférence est décrit schématiquement dans la Figure a. La Figure b montre le précodage distribué dans le cas de précodage conjoint avec partage des données utilisateurs. La matrice $\mathbf{E}_i^{\mathrm{T}}$ est une matrice qui sélectionne les lignes du précodeur multi-utilisateur $\mathbf{T}$ qui correspondent aux antennes localisées au TX i.

1.2 Les défis de l'obtention de l'information de canal aux transmetteurs

1.2.1 Information de canal imparfaite aux transmetteurs

De la même manière que pour le cas d'un transmetteur unique, les bénéfices de la coopération des TXs ne peuvent être atteinte qu'au prix de l'obtention d'une information de canal précise aux TXs. En effet, il n'est possible de coordonner les actions des TXs que si ces TXs connaissent l'état du canal sans fil. L'obtention de cette information de canal est encore plus délicate dans le cas de la coopération des TXs (alignement d'interférence et précodage conjoint) car il est alors nécessaire d'obtenir les informations relatives aux canaux entre tous les TXs et tous les RXs. Ainsi la quantité d'information à partager augmente très vite avec la taille du réseau.

En conséquence, la réduction des besoins en information de canal aux TXs des méthodes d'alignement d'interférence est devenu un domaine actif de recherche à part entière [51, 52, 59–61]. Une synthèse des difficultés techniques et pratiques liées à l'alignement d'interférence est réalisée dans [62]. Une autre approche consiste à étudier le nombre minimal de bits utilisés pour la quantification de l'information de canal de manière à atteindre le nombre maximal de degrés de liberté [59, 60]. Il a aussi été étudié comment la conception des filtres aux RXs et aux TXs peut être modifiée pour réduire encore la quantité d'information de canal nécessaire aux TXs [63, 64]. Enfin, une autre approche appelée «gestion topologique des interférences» a récemment été découverte et présentée dans [65]. Au lieu d'avoir comme point de départ une information parfaite aux TXs et d'essayer ensuite de réduire les besoins en terme d'information de canal, cette approche consiste à considérer que très peu d'information de canal est disponible aux TXs et d'essayer ensuite d'améliorer les performances lorsque plus d'information est disponible aux TXs. En particulier, cette approche vise à exploiter seulement une connaissance topologique du réseau sans-fil par les TXs (i.e., la connectivité).

Pour ce qui est du précodage conjoint avec des TXs éloignés, les algorithmes conçus pour le canal de broadcast avec un seul TX servant plusieurs RXs peuvent être utilisés dans ce scénario de coopération entre TXs. Des approches utilisant des mises-à-jour itératives entre les TXs et les RXs ont aussi été développées dans le but d'éviter le feedback explicite de l'information de canal aux TXs [66]. En revanche, ces itérations demandent plusieurs échanges consécutifs d'information entre les TXs. Ces échanges introduisent des délais importants dans de nombreux scénarios ce qui réduit considérable-

ment la qualité de la suppression des interférences. En conséquence, le précodage conjoint est habituellement limité à de petits groupes de coopération au sein desquels les TXs échangent leurs estimées de canal et transmettent coopérativement. La manière optimale de former ces groupes a aussi été étudiée [67–71] mais l'utilisation de petits groupes de coopération présente certaines limitations fondamentales. Tout d'abord, les performances aux limites des groupes sont limitées par les interférences entre groupes de coopération. Ensuite, cela demande aux TXs de recevoir l'information de canal relative à la totalité du groupe de coopération. Le volume d'information à partager augmente très vite avec la taille du groupe de coopération ce qui limite l'élargissement des groupes de coopération. Plusieurs équipes de recherche ont travaillé sur la détermination de la taille optimale des groupes de coopération et l'évaluation du coût de l'estimation de l'information de canal [72, 73]. Ces travaux suggèrent que la coopération des TXs ne permet pas la gestion efficace des interférences, même lorsqu'il est possible de former de grands groupes de coopération.

1.2.2 Configuration à information de canal distribuée

Nous avons vu dans le paragraphe précédent que de nombreux travaux ont étudié le problème de la coopération des transmetteurs lorsque l'information de canal obtenue aux TXs est imparfaite. En revanche, il est considéré dans ces travaux que l'estimée, bien que imparfaite, est la même à tous les TXs impliqués dans le précodage conjoint. En pratique, cela signifie soit que le précodage est réalisé centralement, soit que les estimées de canal sont parfaitement partagées entre les TXs. Nous qualifions par la suite ce scénario à information de canal comme étant «centralisé». Cette hypothèse est bien justifiée dans certains cas où les liens entre TXs sont de très bonne qualité et en particulier lorsque les TXs sont situés au même endroit.

En revanche, cette hypothèse est peu réaliste dans de nombreux scénarios où les TXs sont éloignés et où le nombre de TXs souhaitant entrer en coopération est important. L'information de canal est acquise localement aux RXs et doit être transmise aux entités responsables du précodage. Une possibilité est l'utilisation d'un noeud central, mais celle-ci ne peut être envisagée que dans certains scénarios et requiert une architecture centralisée qui n'est pas adaptée aux réseaux de grande taille. La solution alternative est le précodage distribué mais cela requiert que chaque TX reçoive à priori l'information de canal relative à tout le canal multi-utilisateur. Dans le cas du précodage distribué, deux scénarios sont actuellement envisagés pour l'acquisition de l'information de canal aux TXs.

Dans le premier scénario, l'information de canal disponible aux RXs est émise directement vers tous les TXs comme dans un canal de broadcast. Cette approche présente l'avantage de ne nécessiter aucun partage d'information entre les TXs. En revanche, elle n'est pour l'instant pas privilégiée par les organismes de standardisation pour les réseaux mobiles [74].

Le scénario alternatif consiste à une transmission du RX vers un unique TX. Cette information de canal est ensuite partagée vers les autres TXs. Le partage de cette information sans délai et sans dégradation demande l'utilisation de liens très coûteux qui ne peuvent pas être utilisés pour connecter tous les TXs entre eux. Ainsi, cette étape de partage de l'information de canal introduit dans de nombreux cas des délais supplémentaires.

Ces deux scénarios sont représentés dans la Figure 1.2. Dans les deux cas, les informations de canal disponibles aux différents TXs ne sont pas exactement les mêmes, en particulier lorsque le nombre de TXs entrant en coopération augmente. En effet, chaque TX doit recevoir l'information de canal relative à l'intégralité du canal multi-utilisateur ce qui implique que la quantité d'information à échanger augmente très vite avec la taille du réseau.

Pour modéliser ces scénarios où différents TXs obtiennent des estimées de canal étant imparfaites et imparfaitement partagées entre les TXs, nous introduisons la notion de configuration à «information de canal distribuée». Dans ce modèle, le TX j reçoit sa propre estimée du canal multi-utilisateur que nous dénotons par $(\mathbf{H}^{(j)})^{\mathrm{H}} \in \mathbb{C}^{N_{\mathrm{tot}} \times M_{\mathrm{tot}}}$. Cette estimée est reçue avant que la transmission n'est lieu par un protocole de coopération entre les TXs. Chaque TX effectue ensuite le précodage en utilisant seulement cette information et sans autre forme de communication avec les autres TXs.

En dépit de son importance pratique, très peu de travaux ont étudié cette configuration. Dans [75], un algorithme de précodage adapté à cette configuration est proposé pour le précodage conjoint avec deux TXs. Cependant, cet algorithme a une forte complexité et ne peut pas être facilement généralisé. Dans [76], la capacité du canal d'interférence avec deux utilisateurs est étudiée lorsque chaque TX connaît seulement une partie de l'information de canal. Néanmoins, il ne prend pas en considération le cas d'information de canal imparfaitement connue et imparfaitement partagée entre TXs. Postérieurs à nos publications, certains travaux ont poursuivi l'étude de l'impacte de la configuration à information de canal distribuée sur les méthodes de transmissions [77–79].

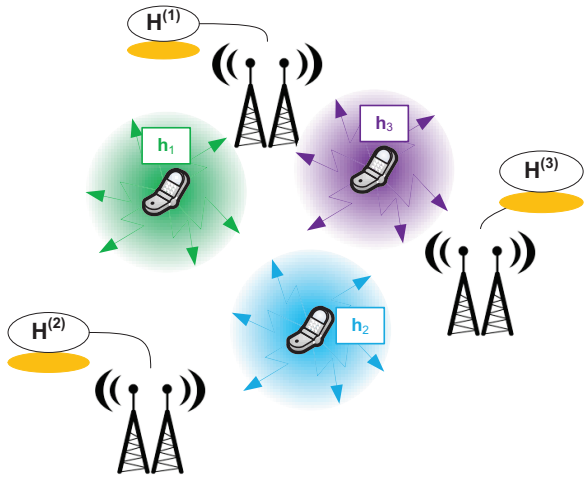
La considération de ce nouveau modèle pose de nombreuses questions. Bien que ces questions soient bien souvent difficiles à répondre et restent pour la plupart non-résolues, nous présentons de nouveaux résultats et une

nouvelle vision de certains aspects de ce problème dans cette thèse. En particulier, nous montrons comment d'importants gains peuvent être réaliser en optimisant le partage de l'information de canal entre les TXs.

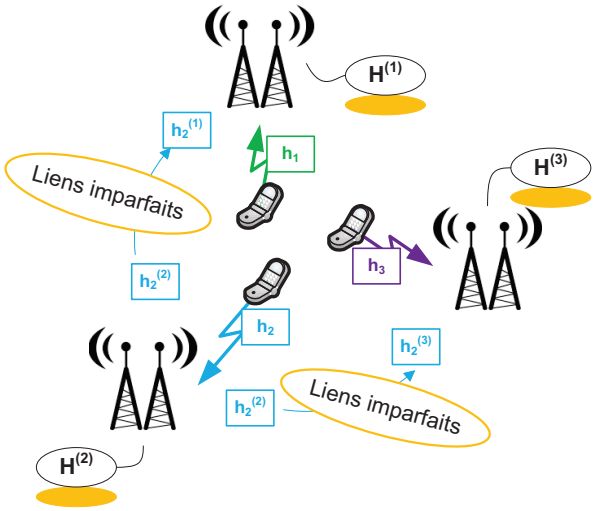
Dans la première partie de cette thèse, nous étudions le précodage à partir d'information de canal distribuée. Considérant tout d'abord le précodage conjoint, nous évaluons les performances des méthodes conçues pour le cas d'information de canal centralisée lorsque confrontées à une information de canal distribuée. Nous mettons alors en évidence la très forte perte d'efficacité de ces méthodes et nous proposons des techniques de transmission plus robustes. Nous étudions aussi plus en détail les besoins en terme d'information de canal et la conception des canaux de feedback. Enfin, nous étudions la performances de l'alignement d'interférence dans le cas d'information de canal distribuée et nous mettons aussi en évidence dans ce cas la nécessité de concevoir de nouvelles approches plus robustes à l'information de canal distribuée.

Dans la seconde partie de cette thèse, nous discutons un autre aspect du problème: Considérant la méthode de précodage comme fixée, nous étudions l'allocation des ressources de feedback disponibles. En particulier, nous étudions quels sont les besoins en terme de feedback et de liens entre TXs pour s'assurer d'une coopération efficace des TXs. En optimisant directement l'allocation des ressources de feedback, nous faisons un pas vers l'allocation à chaque TX de l'information qui lui est vraiment nécessaire. Nous montrons comment l'adaptation de l'allocation des ressources en fonction de l'atténuation du canal ou de la configuration des antennes permet d'importantes réductions de la quantité d'information nécessaire aux TXs sans réduction significative des performances. Ces approches sont ainsi prometteuses pour permettre l'utilisation des méthodes de coopération de TXs dans la pratique.

Nous présentons maintenant un court résumé de certains des principaux résultats de la thèse. Ce résumé est non-exhaustif mais donne une idée des résultats obtenus et permet de comprendre les éléments clés du modèle étudié ainsi que l'approche suivie.



(a) Broadcast de l'information de canal



(b) Echange de l'information de canal

Figure 1.2: Dans la Figure a, chaque TX transmet directement l'estimée disponible vers tous les TXs. Dans la figure b, l'estimée de canal est transmise à un unique TX qui la partage ensuite aux autres TXs. Le canal de tous les TXs vers RX i est dénoté par $\boldsymbol{h}_i^{\mathrm{H}}$ et on représente le fait que l'estimée corresponde à l'information disponible au TX j par l'utilisation du super-script (j).

1.3 Publications

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

1.3.1 Conférences

1. Paul de Kerret and David Gesbert, "The multiplexing gain of a two-cell MIMO channel with unequal CSI", in proc. IEEE International Symposium on Information Theory (ISIT), 2011.

2. Paul de Kerret, Maxime Guillaud, and David Gesbert, "Degrees of freedom of certain interference alignment schemes with distributed CSIT", in Proc. IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC), 2013.

3. Paul de Kerret, Jakob Hoydis, and David Gesbert, "Rate Loss Analysis of Transmitter Cooperation with Distributed CSIT", in Proc. IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC), 2013.

4. Paul de Kerret and David Gesbert, "CSI feedback allocation in multi-cell MIMO channels", in Proc. International Conference on Communications (ICC), 2012.

5. Paul de Kerret and David Gesbert, "MIMO interference alignment algorithms with hierarchical CSIT", in Proc. IEEE International Symposium on Wireless Communication Systems (IWCS), 2012.

   Les publications suivantes ont été réalisées au cours du doctorat mais leurs résultats ne sont pas présentés dans ce manuscript par soucis de brièveté.

6. Paul de Kerret, and David Gesbert, "The asymptotic limits of interference in multicell networks with channel aware scheduling", in Proc. IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC), 2011.

7. Rajeev Gangula, Paul de Kerret, and David Gesbert, "Optimized data symbol allocation in multicell MIMO channels", in proc. IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC), 2011.

8. Paul de Kerret, and David Gesbert, "Sparse precoding in multicell MIMO systems", in proc. IEEE Wireless Communications and Networking Conference (WCNC), 2012.

9. Paul de Kerret, and David Gesbert, "A practical precoding scheme for multicell MIMO channels with partial user's data sharing", in proc. IEEE Wireless Communications and Networking Conference Workshops (WCNCW), 2012.

10. Xinping Yi, Paul de Kerret, and David Gesbert, "The DoF of network MIMO with backhaul delays", in Proc. IEEE International Conference on Communications (ICC), 2013.

11. Paul de Kerret, Xinping Yi, and David Gesbert, "On the degrees of freedom of the K-user time correlated broadcast channel with delayed CSIT", in proc. IEEE International Symposium on Information Theory (ISIT), 2013.

1.3.2 Journaux

1. Paul de Kerret and David Gesbert, "Degrees of freedom of the network MIMO with distributed CSI", in IEEE Trans. Inf. Theory, vol. 58, no. 11, pp. 6806-6824, Nov. 2012

2. Paul de Kerret and David Gesbert, "CSI sharing strategies for transmitter cooperation in wireless networks", in IEEE Wireless Commun. Mag., vol. 20, no. 1, pp. 43-49, Feb. 2013.

3. Paul de Kerret and David Gesbert, "MIMO interference alignment with incomplete CSIT", accepted for publication in IEEE Trans. on Wireless Commun., Jan. 2014.

4. Paul de Kerret and David Gesbert, "Spatial CSIT allocations policies in network MIMO", submitted to IEEE Trans. Inf. Theory, June 2013.

5. Paul de Kerret, Jakob Hoydis, and David Gesbert, "Rate loss analysis of transmitter cooperation with distributed CSIT", to be submitted, 2014.

1.4 Précodage conjoint avec information de canal distribuée

1.4.1 Précodage ZF avec information de canal distribuée

Nous considérons maintenant le cas de précodage conjoint et nous supposons par soucis de clarté que les RXs et les TXs n'ont qu'une seule antenne chacun. Nous dénotons alors le canal de tous les TXs vers le RX i par $\boldsymbol{h}_i^{\mathrm{H}} \in \mathbb{C}^{1\times K}$ (cela correspond à la i-ème ligne de la matrice de canal $\mathbf{H}^{\mathrm{H}} \in \mathbb{C}^{K\times K}$). L'estimé du canal $\boldsymbol{h}_i^{\mathrm{H}}$ au TX j est dénotée par $(\boldsymbol{h}_i^{(j)})^{\mathrm{H}}$. Suite à l'analyse des méthodes de quantification et de feedback [15, 16, 80], cette estimée est définie par

$$\hat{\boldsymbol{h}}_i^{(j)} = \sqrt{1-2^{-\frac{B_i^{(j)}}{K-1}}}\,\boldsymbol{h}_i + 2^{-\frac{B_i^{(j)}}{K-1}}\boldsymbol{\delta}_i^{(j)}, \qquad \forall i,j \in \{1,\ldots,K\}$$

où $B_i^{(j)}$ représente le nombre de bits utilisés pour obtenir $\hat{\boldsymbol{h}}_i^{(j)}$ au TX j. Le vecteur $\boldsymbol{\delta}_i^{(j)} \in \mathbb{C}^{K\times 1}$ a ses éléments distribués comme $\mathcal{CN}(0,1)$ et est indépendant du canal $\boldsymbol{h}_i$.

Nous considérons que les TXs utilisent le précodeur ZF décrit auparavant dans le cas où les TXs ont une connaissance parfaite du canal. Cela signifie que le précodeur ZF obtenu au TX j est $\mathbf{T}^{(j)} = [\boldsymbol{t}_1^{(j)},\ldots,\boldsymbol{t}_K^{(j)}] \in \mathbb{C}^{K\times K}$ avec

$$\boldsymbol{t}_i^{(j)} \triangleq \left(\mathbf{I}_K - \hat{\mathbf{H}}_i^{(j)}\left((\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_i^{(j)}\right)^{-1}(\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}\right)\hat{\boldsymbol{h}}_i^{(j)}, \qquad \forall i \in \{1,\ldots,K\}$$

et avec la matrice $(\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}$ contenant les canaux vers tous les RXs à l'exception de RX i telle que

$$(\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}} \triangleq \begin{bmatrix} (\hat{\boldsymbol{h}}_1^{(j)})^{\mathrm{H}} \\ \vdots \\ (\hat{\boldsymbol{h}}_{i-1}^{(j)})^{\mathrm{H}} \\ (\hat{\boldsymbol{h}}_{i+1}^{(j)})^{\mathrm{H}} \\ \vdots \\ (\hat{\boldsymbol{h}}_K^{(j)})^{\mathrm{H}} \end{bmatrix}, \qquad \forall i \in \{1,\ldots,K\}.$$

Il faut remarquer que TX j obtient le précodeur multi-utilisateur $\mathbf{T}^{(j)}$ mais seulement la j-ième ligne de ce précodeur est utilisée en pratique. Cela provient du précodage qui est distribué parmi tous les TXs. En effet, TX j

transmet seulement $x_j = \boldsymbol{e}_j^{\mathrm{H}} \mathbf{T}^{(j)} \boldsymbol{s}$ (où la multiplication par $\boldsymbol{e}_j^{\mathrm{H}}$ revient à sélectionner la j-ème ligne), de telle sorte que le précodeur obtenu au total est écrit comme $\mathbf{T}^{\mathrm{DCSI}} = [\boldsymbol{t}_1^{\mathrm{DCSI}}, \ldots, \boldsymbol{t}_K^{\mathrm{DCSI}}] \in \mathbb{C}^{K \times K}$ où

$$\boldsymbol{t}_i^{\mathrm{DCSI}} \triangleq \begin{bmatrix} \boldsymbol{e}_1^{\mathrm{H}} \boldsymbol{t}_i^{(1)} \\ \vdots \\ \boldsymbol{e}_K^{\mathrm{H}} \boldsymbol{t}_i^{(K)} \end{bmatrix}, \qquad \forall i \in \{1, \ldots, K\}.$$

1.4.2 Analyse du nombre de degrés de liberté (DoF)

Le précodage ZF est très largement utilisé en pratique et il est connu que le précodage ZF atteint le DoF maximal dans le cas du canal de broadcast lorsque le TX dispose d'une connaissance parfaite du canal [15,16,80]. Il est donc intéressant de déterminer quelles sont les performances du précodage ZF dans le cas de l'information de canal distribuée.

Le scénario de référence est le cas centralisé et nous allons donc rappeler les résultats de la littérature relatifs au cas centralisé pour pouvoir comprendre l'effet de l'information de canal distribuée. Dans le cas centralisé, il n'y a qu'une seule estimée du canal $\boldsymbol{h}_i$ que l'on dénote par $\hat{\boldsymbol{h}}_i$ et qui est obtenue à partir de l'expression précédente avec B_i bits pour la quantification.

Théorème 2 ([15]). Dans le canal de broadcast avec information de canal centralisée, si l'estimée de canal $\hat{\boldsymbol{h}}_i$ qui est disponible centralement est obtenue en utilisant $B_i = A_i(M-1)\log_2(P)$ bits de quantification (avec $A_i \in (0,1]$), le DoF atteint au RX i avec le précodage ZF est égal à

$$\mathrm{DoF}_i^{\mathrm{CCSI}} = A_i.$$

On peut ainsi observer que le DoF atteint au RX i est déterminé seulement par A_i, et donc seulement par B_i. Nous allons maintenant montrer comment l'information de canal distribuée change fondamentalement cette propriété.

Théorème 3. Dans le canal de broadcast avec information de canal distribuée, si l'estimée de canal $\hat{\boldsymbol{h}}_i^{(j)}$ qui est disponible au TX j est obtenue en utilisant $B_i^{(j)} = A_i^{(j)}(M-1)\log_2(P)$ bits de quantification (avec $A_i^{(j)} \in (0,1]$), le DoF atteint au RX i avec le précodage ZF est égal à

$$\mathrm{DoF}_i^{\mathrm{ZF}} = \min_{i,j \in \{1,\ldots,K\}} A_i^{(j)}.$$

Le DoF au RX i est limité par l'estimée la moins précise parmi les estimées de tous les TXs. Ce résultat contraste fortement avec le cas centralisé et montre l'influence critique de l'information de canal distribuée. C'est aussi un résultat très pessimiste car une seule mauvaise estimée suffit à réduire les performances de tous les utilisateurs. Il est donc important de développer des précodeurs étant plus robustes à ce genre d'imperfections.

Des résultats préliminaires dans [81] montrent qu'il est possible dans certains scénarios d'augmenter dramatiquement le DoF atteint. Par exemple, pour la coopération de deux TXs, nous avons développée une méthode de précodage appelée *actif-passif (AP) ZF* qui améliore le DoF comparé au précodage ZF conventionnel.

Théorème 4. Dans le canal de broadcast avec information de canal distribuée, si l'estimée de canal $\hat{\boldsymbol{h}}_i^{(j)}$ qui est disponible à TX j est obtenue en utilisant $B_i^{(j)} = A_i^{(j)}(M-1)\log_2(P)$ bits de quantification (avec $A_i^{(j)} \in (0,1]$), le DoF atteint à RX i avec le précodage AP-ZF est égal à

$$\mathrm{DoF}_i^{\mathrm{APZF}} = \max_{j\in[1,2]} A_i^{(j)}.$$

Ainsi, le précodage AP-ZF permet d'atteindre le DoF qui serait obtenu avec la meilleure des estimées disponibles parmi les deux TXs. AP-ZF consiste à fixer arbitrairement le coefficient de précodage du TX ayant l'estimée la moins précise pour laisser le TX avec la meilleure estimée éliminer les interférences. Les débits moyens atteints avec les précodeurs ZF et AP-ZF sont comparés dans la Figure 1.3 dans le cas où:

$$A_1^{(1)} = 1,\ A_1^{(2)} = 0.5,\ A_2^{(1)} = 0,\ A_2^{(2)} = 0.7.$$

On peut observer que le débit moyen atteint avec le précodage ZF «conventionnel» sature lorsque le SNR augmente (i.e., le DoF est zéro) alors que le précodeur AP ZF est plus robuste et permet d'obtenir un DoF positif (plus exactement, la somme des deux DoFs est égale à 1.3).

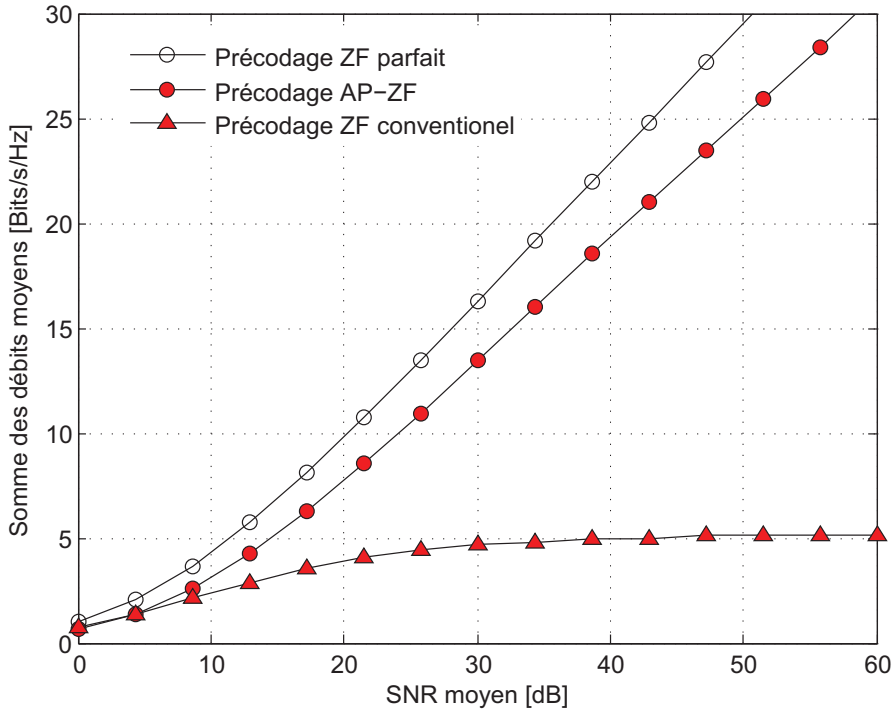


Figure 1.3: Le débit total $R_1 + R_2$ est présenté en fonction du rapport signal-à-bruit pour $A_1^{(1)} = 1, A_1^{(2)} = 0.5, A_2^{(1)} = 0, A_2^{(2)} = 0.7$.

1.4.3 Analyse du précodage et des canaux de feedback

Dans la partie précédente, nous avons évalué l'impacte de l'information de canal distribuée sur le DoF. L'analyse du DoF permet d'obtenir une intuition du fonctionnement et de discuter la qualité de la gestion des interférences. En revanche, cette analyse est difficilement exploitable pour concevoir des systèmes pratiques dans la mesure où le DoF n'est atteint que pour un SNR asymptotiquement grand. C'est pourquoi nous étudions maintenant les dégradations dues à l'information de canal imparfaite en terme de débit.

Une première étape consiste à évaluer la qualité du précodage en fonction de la qualité de l'information de canal disponible.

Proposition 1. Dans le canal de broadcast avec information de canal distribuée, il est vérifié avec une probabilité 1, que

$$\boldsymbol{u}_i^{(j)} = \boldsymbol{u}_i^{\star} + \boldsymbol{a}_i^{(j)} + o(\max_q 2^{-\frac{B_q^{(j)}}{K-1}}), \qquad \forall i,j \in \{1,\ldots,K\}$$

avec

$$\mathrm{E}[|\boldsymbol{e}_p^{\mathrm{H}} \boldsymbol{a}_i^{(j)}|^2] = \frac{2\sum_{k=1,k\neq i}^{K} (2^{-\frac{B_k^{(j)}}{K-1}})^2 + (2^{-\frac{B_i^{(j)}}{K-1}}}{K}, \qquad \forall i,j,p.$$

Cette proposition est le premier résultat qui relie la précision du précodage (au sens de la différence quadratique avec le précodeur obtenu à partir d'une information parfaite de canal) à la qualité de l'information de canal. En effet, dans le cas centralisé, ce sont directement les interférences $|\boldsymbol{h}_i^{\mathrm{H}} \boldsymbol{u}_j|^2$ qui sont analysées. Dans le cas distribué, ces interférences ont une distribution statistique complexe (et différente de la distribution des interférences dans le cas centralisé) due au précodage à partir d'information de canal distribuée qui rend leur analyse difficile. De plus, le précodeur ZF obtenu avec l'information de canal distribuée est aussi particulier par le fait qu'il n'est orthogonal à aucune estimée de canal connue aux TXs, à cause du précodage distribué.

À partir de cette proposition, le débit moyen atteint avec l'information de canal distribuée est minoré dans le théorème suivant.

Théorème 5. Dans le canal de broadcast avec information de canal distribuée, le débit moyen $\mathrm{R}_i^{\mathrm{DCSI}}$ atteint par l'utilisateur i peut être minoré à haut SNR par

$$\mathrm{R}_i^{\mathrm{DCSI}} \geq \mathrm{R}_i^{\star} - \Delta_{\mathrm{R},i}^{\mathrm{DCSI}}$$

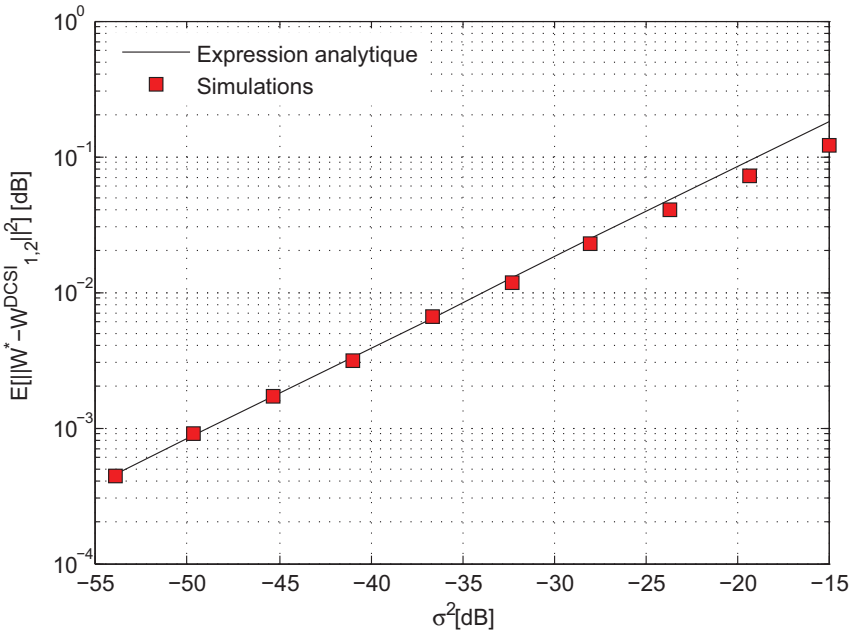


Figure 1.4: Différence moyenne $\mathrm{E}[|\boldsymbol{e}_1^{\mathrm{T}}(\boldsymbol{u}_2^{(1)} - \boldsymbol{u}_2^{\star})|^2]$ en fonction de la qualité de l'information de canal $2^{-\frac{B_i^{(1)}}{K-1}}$.

où $\mathrm{R}_i^{\star}$ correspond au débit atteint en utilisant le précodage ZF avec une connaissance parfaite du canal à chaque TX et

$$\Delta_{\mathrm{R},i}^{\mathrm{DCSI}} \leq \mu(K) + \log_2\left(1 + P\sum_{j=1}^{K}\left((2K-3)\sum_{k=1,k\neq i}^{K} 2^{-\frac{B_k^{(j)}}{K-1}} + 2(K-1)2^{-\frac{B_i^{(j)}}{K-1}}\right)\right) + o(1)$$

avec la fonction $\mu(x)$ définie pour $x > 0$ par

$$\mu(x) \triangleq \log_2\left(3 + 2\log(x)\right).$$

Nous considérons maintenant un exemple d'information de canal pour visualiser les résultats théoriques précédents. Soit un canal d'interférence avec $K = 5$ utilisateurs où

$$B_i^{(1)} = 4\log_2(P^{\frac{2}{3}}), \qquad \forall i \tag{1.1}$$

$$B_i^{(j)} = 4\log_2(P), \qquad \forall i, j \ \ , j \neq 1. \tag{1.2}$$

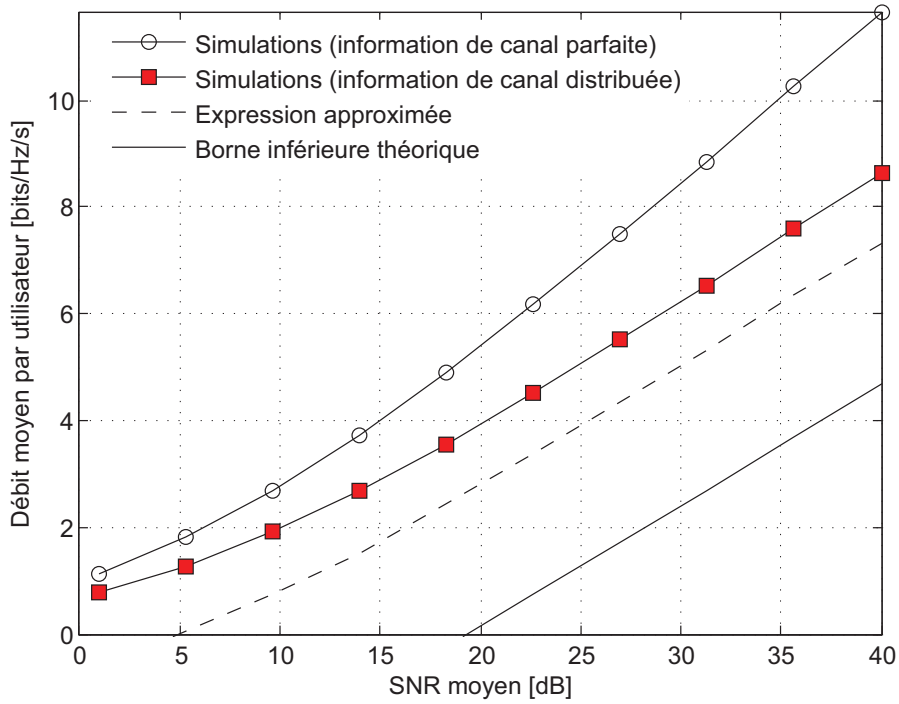


Figure 1.5: Débit moyen par utilisateur en fonction du SNR moyen P avec une information de canal distribuée donnée par $B_i^{(1)} = 4\log_2(P^{\frac{2}{3}}),\ \forall i, B_i^{(j)} = 4\log_2(P),\ \forall i, j, j \neq 1.$.

Nous montrons dans la Figure 1.4 la différence moyenne en norme quadratique $\mathrm{E}[|\boldsymbol{e}_1^{\mathrm{T}}(\boldsymbol{u}_2^{\mathrm{DCSI}} - \boldsymbol{u}_2^{\star})|^2] = \mathrm{E}[|\boldsymbol{e}_1^{\mathrm{T}}(\boldsymbol{u}_2^{(1)} - \boldsymbol{u}_2^{\star})|^2]$. Nous pouvons vérifier que les résultats des simulations concordent avec les résultats analytiques.

Enfin, nous montrons dans la Figure 1.5 le débit moyen atteint et nous le comparons à la borne obtenue analytiquement. Nous présentons aussi une expression analytique qui correspond à la borne du Théorème 5 mais sans la fonction $\mu(K)$. Nous conjecturons en effet que ce terme est seulement du à la difficulté de borner précisément les performances dans le cas distribué. En effet, la distribution statistique des interférences est complexe et la dérivation de la borne inférieure requiert de nombreuses approximations. Bien que la borne obtenue apparaisse relativement éloignée, elle prend efficacement en compte la précision de l'information de canal et peut être utilisée dans la conception des canaux de feedback. À partir du Théorème 5, nous pouvons

ainsi obtenir le critère suivant pour la conception des canaux de feedback.

Théorème 6. Dans le canal de broadcast avec information de canal distribuée, le débit moyen $\mathrm{R}_i^{\mathrm{DCSI}}$ atteint par l'utilisateur i peut être minoré à haut SNR par $R_i^\star - \log_2(1+b) + o(1)$ bits si $b > 2 + 2\log(K)$ et $B_i \geq B^{\mathrm{DCSI}}, \forall i$ avec

$$B^{\mathrm{DCSI}} = (K-1)\log_2\left(\frac{(2K-1)(K-1)P}{b}\right) + (K-1)\log_2\left(\frac{b(3+2\log(K))}{b-2-2\log(K)}\right).$$

Ce dernier résultat donne un critère pour concevoir les canaux de feedback dans le cas d'information de canal distribuée. Cela permet de déterminer le nombre de bits nécessaires pour la quantification de l'information de canal de manière à atteindre les performances désirées. En comparant cette expression aux résultats obtenus dans le cas centralisé, on remarque que l'expression obtenue dans le cas centralisé n'est pas adaptée au cas distribué, et peut mener à d'importantes pertes de performance.

Nous présentons dans la Figure 1.6 le débit moyen par utilisateur atteint en utilisant le nombre de bits de quantification B^{DCSI} avec $b = 1$. Nous le comparons au débit moyen obtenu en utilisant $B_i = B^{\mathrm{CCSI}}, \forall i$ où $B^{\mathrm{CCSI}} = (K-1)\log_2((K-1)P)$ correspond au nombre de bits suffisant pour assurer une perte maximale de $\log_2(1+1) = 1$ bit dans le cas centralisé. On peut observer que l'utilisation de B^{DCSI} induit bien une perte plus petite que 1 bit comparé au débit moyen obtenu à partir d'une information de canal parfaite, ce qui n'est pas le cas en utilisant B^{CCSI}.

Nous avons ainsi pu montrer dans ce manuscript que les règles de conception des canaux de feedback pour le cas centralisé ne sont pas adaptées au cas distribué et nous avons développé de nouveaux critères qui permettent de garantir une transmission efficace à partir d'information de canal distribuée.

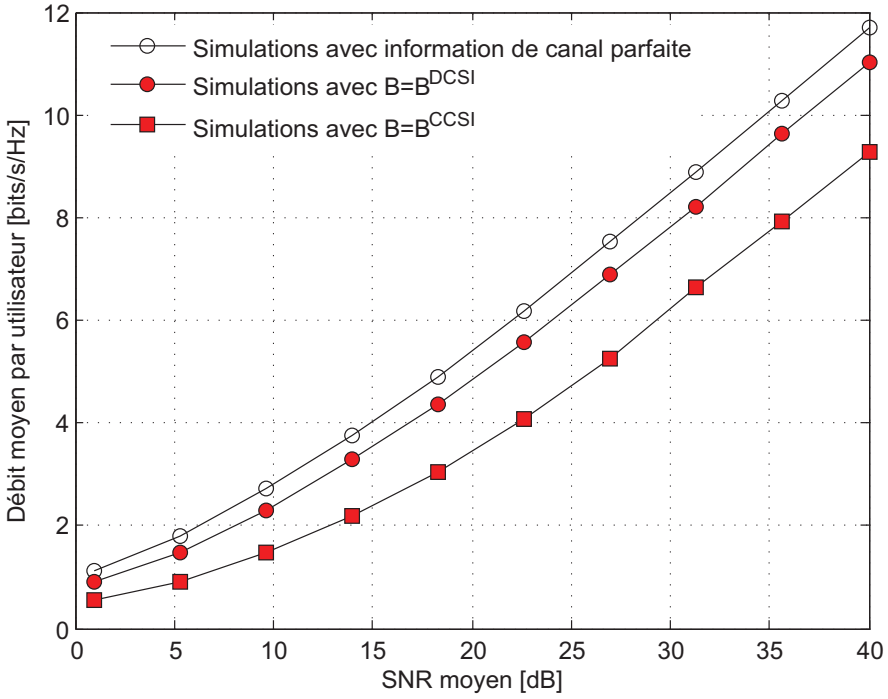


Figure 1.6: Débit moyen en fonction du SNR moyen P dans une configuration de canal distribuée. Le nombre de bits B^{DCSI} est donné par le Théorème 6 avec $b = 1$ et $B^{\mathrm{CCSI}} = (K-1)\log_2\left((K-1)P\right)$ correspond au nombre de bits suffisant pour assurer une perte maximale de $\log_2(1+1) = 1$ bit dans le cas centralisé.

1.5 Alignement d'interférence avec information de canal incomplète

Nous considérons dans cette section un canal d'interférence, i.e., que les données ne sont pas partagées entre les TXs. Comme déjà mentionné auparavant, de nombreux travaux traitent de la faisabilité de l'alignement d'interférence en fonction du nombre d'antennes disponibles aux TXs et aux RXs. En revanche, ces travaux supposent toujours que chaque TX dispose d'une connaissance parfaite de tous les coefficients de canal. Cependant, il devient très vite clair qu'il n'est pas toujours nécessaire de fournir l'intégralité de l'information de canal à chaque TX. Supposons par exemple que chaque RX aie à sa disposition un nombre suffisant d'antennes. Il n'est alors pas nécessaire d'effectuer de précodage puisque les interférences peuvent être supprimées par simple filtrage aux RXs. Cela a pour conséquence qu'il est possible dans ce cas «d'aligner les interférences» sans aucune information de canal aux TXs. Ce simple exemple suggère l'existence d'un compromis entre le nombre d'antennes disponibles et les besoins en terme d'information de canal. Nous allons en fait aller plus loin et montrer qu'il est possible dans certains cas de réduire les besoins en information de canal sans nécessiter aucune antenne additionnelle.

Plus généralement, nous formulons le problème de la minimisation de l'information de canal transmise aux TXs à la condition que la faisabilité de l'alignement d'interférence soit préservée. La minimalité fait référence à la taille de l'allocation de feedback aux TXs que l'on définit comme le nombre total de nombres complexes transmis aux TXs par l'intermédiaire d'un canal de feedback multi-utilisateur.

1.5.1 Modèle d'information de canal incomplète et introduction du problème considéré

Nous considérons maintenant une structure particulière d'information de canal distribuée que nous qualifions d'information de canal «incomplète». Dans ce modèle, un TX obtient soit une connaissance parfaite d'un coefficient de la matrice de canal, soit aucune information sur ce coefficient. Cette structure d'information de canal est représentée par l'intermédiaire des matrices $\mathbf{F}^{(j)} \in \{0,1\}^{N_{\text{tot}} \times M_{\text{tot}}}, j = 1, \ldots, K$. Ces matrices sont définies de telle sorte que si le coefficient $\{\mathbf{H}^{\text{H}}\}_{ik}$ est connu au TX j, alors $\{\mathbf{F}^{(j)}\}_{ik} = 1$, et sinon $\{\mathbf{F}^{(j)}\}_{ik} = 0$. Ainsi l'estimée du canal multi-utilisateur $\hat{\mathbf{H}}^{(j)}$ qui est

disponible à TX j est donnée par

$$(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}} = \mathbf{F}^{(j)} \odot \mathbf{H}^{\mathrm{H}}$$

où $\odot$ représente le produit de Hadamard. Nous définissons à partir de ces matrices l'allocation d'information de canal $\mathcal{F}$:

$$\mathcal{F} = \{\mathbf{F}^{(j)} | \mathbf{F}^{(j)} \in \{0,1\}^{N_{\mathrm{tot}} \times M_{\mathrm{tot}}}, j = 1, \ldots, K\}$$

et nous définissons aussi l'espace $\mathbb{F}$ qui contient les allocations d'information de canal possibles. Enfin, pour représenter la consommation des ressources de feedback, nous définissons la taille d'une allocation d'information de canal comme suit.

Définition 1. La taille d'une allocation d'information de canal $\mathcal{F}$, dénotée par $\mathrm{s}(\mathcal{F})$, est égale au nombre total de coefficients complexes transmis aux TXs:

$$\mathrm{s}(\mathcal{F}) \triangleq \sum_{j=1}^{K} \|\mathbf{F}^{(j)}\|_{\mathrm{F}}^2.$$

Nous nous intéressons aux allocations d'information de canal qui préservent la faisabilité de l'alignement d'interférence de telle sorte que nous considérons uniquement l'ensemble $\mathbb{F}_{\mathrm{feas}}$ contenant précisément les allocations d'information de canal préservant la faisabilité de l'alignement d'interférence. Il a été démontré comment vérifier facilement la faisabilité d'une configuration d'antenne [49,50] de telle sorte que déterminer $\mathbb{F}_{\mathrm{feas}}$ peut être facilement réalisé.

Le problème qui nous intéresse est alors le suivant: *Quelle est l'allocation d'information de canal ayant la plus petite taille tout en préservant la faisabilité de l'alignement d'interférence?* Cette question peut alors être reformulée mathématiquement comme:

$$\mathcal{F}_{\min} = \underset{\mathcal{F} \in \mathbb{F}_{\mathrm{feas}}}{\operatorname{argmin}} \mathrm{s}(\mathcal{F}).$$

Il est nécessaire de remarquer que nous nous intéressons à l'influence de l'information de canal et non au problème de la faisabilité en lui-même, de telle sorte que nous considérons que toutes les configurations d'antennes considérées sont faisables si tous les TXs reçoivent une information complète du canal multi-utilisateur.

Nous introduisons maintenant la notion de *étroitement faisable* et de *largement faisable*, qui se révélera fondamentale dans l'analyse de la faisabilité de l'alignement d'interférence avec information incomplète.

Définition 2. Un canal d'interférence est dit *étroitement faisable* si ce canal d'interférence est faisable pour l'alignement d'interférence dans cette configuration d'antennes et si aucune antenne ne peut être enlevée sans rendre l'alignement d'interférence infaisable. Un canal d'interférence est *étroitement faisable* si et seulement il est faisable et

$$\sum_{i=1}^{K} N_i + M_i = K(K+1).$$

Un canal d'interférence qui est faisable mais non *étroitement faisable* est dit *largement faisable*.

1.5.2 Analyse des scénarios étroitement faisables

Théorème 7. Dans un canal d'interférence $\prod_{k=1}^{K}(N_k, M_k)$ étroitement faisable, supposons qu'il existe un sous-ensemble de TXs et un sous-ensemble de RXs formant un sous-canal d'interférence (i.e., un canal d'interférence inclus dans le premier) étroitement faisable, i.e.,

$$\mathcal{N}_{\text{var}}(\mathcal{S}_{\text{RX}}, \mathcal{S}_{\text{TX}}) = \mathcal{N}_{\text{eq}}(\mathcal{S}_{\text{RX}}, \mathcal{S}_{\text{TX}})$$

et définissons l'allocation incomplète d'information de canal $\mathcal{F} = \{\mathbf{F}^{(j)} | j = 1, \ldots, K\}$ telle que

$$\begin{aligned} \mathbf{F}^{(j)} &= \mathbf{F}_{\mathcal{S}_{\text{RX}}, \mathcal{S}_{\text{TX}}}, & \forall j \in \mathcal{S}_{\text{TX}} \\ \mathbf{F}^{(j)} &= \mathbf{F}_{\mathcal{K},\mathcal{K}} = \mathbf{1}_{N_{\text{tot}} \times M_{\text{tot}}} & \forall j \notin \mathcal{S}_{\text{TX}} \end{aligned}$$

où $\mathbf{F}_{\mathcal{S}_{\text{RX}}, \mathcal{S}_{\text{TX}}}$ représente l'allocation d'information de canal contenant les canaux reliant les RXs dans $\mathcal{S}_{\text{RX}}$ et les TXs dans $\mathcal{S}_{\text{TX}}$.

Alors l'allocation d'information de canal $\mathcal{F}$ préserve la faisabilité de l'alignement d'interférence, i.e., $\mathcal{F} \in \mathbb{F}_{\text{feas}}$.

L'intuition derrière ce résultat est l'importance de reconnaître les sous-ensembles de TXs et de RXs qui forment un canal d'interférence étroitement faisable. Ainsi, il est possible d'aligner les interférences en fournissant à chaque TX seulement l'information relative au plus petit canal d'interférence étant étroitement faisable. A partir de ce résultat, il apparaît facilement comment obtenir une approche itérative d'allocation de l'information de canal. L'algorithme consiste à d'abord former les précodeurs des TXs appartenant au canal d'interférence étroitement faisable le plus petit. Une fois ces précodeurs fixés, les autres précodeurs sont obtenus itérativement de la même manière. Ainsi chaque TX nécessite seulement l'information de

canal relative au sous canal d'interférence étroitement faisable le plus petit auquel il appartient. Pour la description détaillé de l'algorithme et la preuve que cette approche permet bien d'aligner les interférences, nous renvoyons le lecteur à la partie principale du manuscript.

Nous verrons dans les simulations que la réduction atteinte de la taille de l'information de canal est importante. En fait, cette réduction exploite l'hétérogénéité de la configuration d'antennes: Le plus hétérogène est la configuration, le plus grand sera le gain réalisé par notre approche. Cette propriété est particulièrement intéressante dans la mesure où les réseaux du future contiendront des RXs et des TXs de différentes générations et de différentes natures ce qui favorise l'apparition de configurations hétérogènes.

1.5.3 Analyse des scénarios largement faisables

Dans les configurations largement faisables, chaque antenne additionnelle peut être exploitée pour réduire la taille minimale de l'information de canal. Cependant, comment exploiter ces antennes additionnelles représente un problème difficile. Nous avons donc proposé une approche simple qui consiste à utiliser le fait qu'il est possible d'obtenir des configurations étroitement faisables en ignorant certaines antennes à certains RXs et/ou certains TXs dans une configuration largement faisable. Il y a en général plusieurs configurations étroitement faisables qui peuvent ainsi être obtenues à partir d'un scénario largement faisable avec à priori des besoins différents en information de canal.

En conséquence, nous considérons à la place du problème de minimisation initial, le problème d'optimisation suivant :

$$\mathcal{F} = \underset{\mathcal{F}\in\mathbb{F}_{\text{feas}}}{\operatorname{argmin}} \min_{\prod_{k=1}^{K}(N_k',M_k')} \mathrm{s}(\mathcal{F}) \quad \text{s.t.} \sum_{i=1}^{K} M_i' + N_i' = (K+1)K \quad \text{s.t. } 1 \leq M_i' \leq M_i \text{ and } 1 \leq N_i' \leq N_i.$$

Ainsi la minimisation de l'allocation d'information de canal a été réduite au problème consistant à trouver le scénario étroitement faisable (contenant tous les utilisateurs) inclus dans le canal d'interférence initial et nécessitant le moins d'information de canal. Étant donné que nous avons déjà développé un algorithme pour les cas étroitement faisables, il reste seulement à déterminer quels RXs ou quels TXs ne devraient pas exploiter toutes leur antennes pour éliminer les interférences, i.e., où des antennes sont «enlevées» en terme de faisabilité d'alignement d'interférence. Un algorithme heuristique est proposé dans le manuscript et nous verrons qu'il atteint des réduc-

tions importantes avec une faible complexité. Cet algorithme est divisé en deux phases: la phase 1) consiste en l'identification des TXs et des RXs où il est impossible d'enlever une antenne sans rendre l'alignement d'interférence infaisable, ce qui revient à trouver les TXs et les RXs qui appartiennent à un sous-canal d'interférence étroitement faisable. La phase 2) consiste en l'application de l'heuristique qui détermine à quel TX ou quel RX retirer l'antenne parmi les TXs et les RXs où cela est possible sans rendre infaisable l'alignement d'interférence.

Il faut remarquer que l'expression «enlever des antennes» ne signifie pas que les antennes sont physiquement enlevées ou même éteintes. Cela signifie simplement que des dimensions de précodage ne sont pas utilisées pour réduire les interférences mais dans un autre but, comme par exemple augmenter la puissance émise vers un utilisateur où améliorer la diversité de la transmission.

Par soucis de clarté, nous omettons la description de l'algorithme et montrons simplement sur un exemple simple comment l'algorithme fonctionne de manière à obtenir une intuition du principe général de notre approche.

Exemple 1. Nous étudions ici le canal d'interférence $[(2,2).(3,2).(2,3)]$. On peut facilement vérifier que cette configuration d'antennes est faisable pour l'alignement d'interférence. De plus, elle est largement faisable avec deux antennes additionnelles puisque $\sum_{i=1}^{K} N_i + M_i - K(K+1) = 2$. Nous allons maintenant présenter les étapes de notre algorithme d'allocation de l'information de canal pour les canaux d'interférences largement faisables.

- $n = 1$: Durant la phase 1), aucun sous-canal d'interférence étroitement faisable n'est trouvé. Cela signifie qu'il est possible d'enlever une antenne à n'importe quel TX ou n'importe quel RX sans remettre la faisabilité de l'alignement d'interférence en question. Durant la phase 2), une antenne est alors enlevée au TX 1.

- $n = 2$: On obtient donc le canal d'interférence $[(2,1).(3,2).(2,3)]$. Durant la phase 1), l'ensemble des TXs appartenant à un sous-canal d'interférence étroitement faisable est $\mathcal{S}_{\mathrm{TX}}^{\mathrm{Tight}}(n) = \{1,2\}$ et l'ensemble de RXs est $\mathcal{S}_{\mathrm{RX}}^{\mathrm{Tight}}(n) = \{1,3\}$. On peut alors enlever une antenne aux TXs et aux RXs n'appartenant pas à ces ensembles. Durant la phase 2), une antenne est enlevée au TX 3.

Avec cet algorithme itératif, nous avons donc obtenu le canal d'interférence $[(2,1).(3,2).(2,2)]$. Cette configuration d'antennes étant étroitement faisable, nous pouvons utiliser l'algorithme développé précédemment pour les

canaux d'interférence étroitement faisables. Cet algorithme retourne l'allocation d'information de canal

$$\mathcal{F} = \{\mathbf{F}^{(1)} = \mathbf{F}_{\emptyset,\emptyset}, \mathbf{F}^{(2)} = \mathbf{F}_{\{3\},\{1,2\}}, \mathbf{A}^{(F)} = \mathbf{A}_{\{1,3\},\{1,2,3\}}\}.$$

La taille de l'information de canal obtenue avec notre algorithme est de 20 alors que la taille complète est de 99. Les 2 antennes additionnelles ont donc été exploitées pour réduire par un facteur de pratiquement 4 la taille de l'allocation d'information de canal.

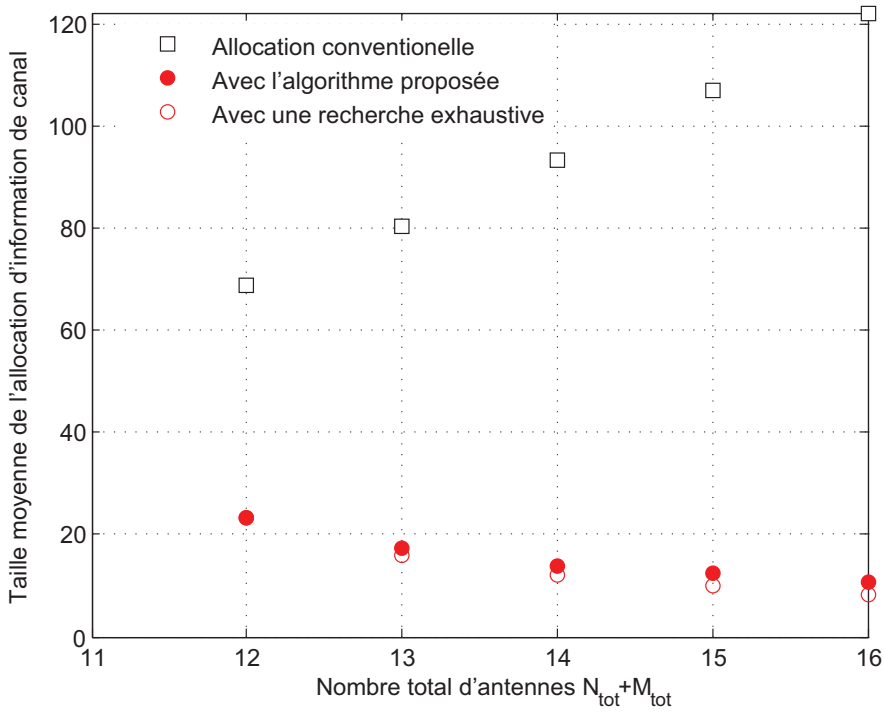


Figure 1.7: Taille moyenne de l'allocation d'information de canal en fonction du nombre d'antennes distribuées de manière aléatoire et uniformément aux TXs et aux RXs pour $K = 3$ paires de TX/RX.

La réduction moyenne de l'information de canal qui est obtenue par l'utilisation de notre algorithme est représentée dans la Figure 1.7 dans le cas d'un canal d'interférence avec 3 paires de TX/RX. Les résultats sont moyennés sur 1000 configurations d'antennes. Ces configurations sont

obtenues en allouant aléatoirement selon une distribution uniforme les antennes aux TXs et aux RXs. Si le canal d'interférence contient 12 antennes, le canal d'interférence est étroitement faisable et l'algorithme pour les scénarios étroitement faisables est utilisé. Avec plus de 12 antennes disponibles, chaque antenne additionnelle est exploitée par notre algorithme pour réduire la taille de l'information de canal nécessaire. On peut observer que notre approche réduit fortement la taille de l'information de canal, et cela même lorsque la configuration est étroitement faisable.

1.6 Conclusion et nouveaux problèmes

Nous avons mis en évidence dans cette thèse la nécessité de considérer le cas d'information de canal distribuée dans le mesure où cela correspond à une réalité pratique. De plus, l'évaluation des performances atteintes par les méthodes de précodage conventionnelles montre l'importance de considérer ce modèle dans la conception du précodage, au prix d'une forte dégradation des performances. Nous avons aussi mis en évidence que l'allocation uniforme des ressources de feedback vers tous les TXs n'utilise pas efficacement les ressources de feedback disponibles: La dissémination de l'information de canal doit être conçue de manière à fournir à chaque TX seulement l'information qui lui est réellement utile. L'optimisation du partage de l'information de canal réduit considérablement les besoins en information de canal aux TXs, rendant ainsi la coopération des TXs plus adaptée aux contraintes pratiques des réseaux et ouvrant la voie vers un usage plus important de la coopération des TXs dans la gestion des interférences.

Au delà des nouvelles méthodes présentées, la considération d'une information de canal distribuée aux TXs mène à de nombreuses nouvelles questions mais aussi à de nombreuses opportunités. Nous avons souvent étudié dans ce manuscript le nombre de degrés de liberté comme métrique et plus généralement les performances à haut SNR. L'étude des performances à des SNRs plus faibles et dans des modèles de canal plus proches de la réalité permettra de transposer les gains obtenus théoriquement en des améliorations de l'efficacité spectrale dans la pratique. De plus, le problème du précodage avec information de canal distribuée est un problème difficile qui reste dans sa majeure partie non résolu. L'obtention de nouvelles méthodes de précodage plus robustes à l'information de canal distribuée représente ainsi une direction de recherche prometteuse. Enfin, nous avons montré dans certains scénarios comment l'optimisation du partage de l'information de canal permet d'utiliser plus efficacement les ressources de feedback disponibles. Cette

approche peut potentiellement être étendue à de nombreux autres scénarios et contribuer à rendre les réseaux sans fil plus efficaces.

Part I

Motivations and Models

Chapter 2

Introduction

2.1 State of the Art for Transmitter Cooperation

2.1.1 Saturation of the Wireless Medium

Wireless communication has become essential to our lives in many ways, through a variety of services and devices ranging from pocket phones to laptops, tablets, sensors and controllers. The advent of multimedia dominated traffic poses extra-ordinary constraints on data rates, latency and above all spectral efficiency. In mobile networks, the demand for data rate has increased exponentially in the past decade. In 2012, the global mobile traffic was equal to 10 times the size of the entire global internet traffic in 2000 and it is expected that the global mobile traffic will continue its exponential increase to reach a tenfolds increase within the next 5 years [2]. In order to deal with the expected saturation of available resources in currently used bands, the architecture of the wireless networks and their transmission schemes have to be rethought. Key enablers for the strong performance of new wireless systems will be a *i*) greater densification of infrastructure equipments (small cells [3]), and *ii*) a very aggressive spatial frequency reuse, which in turn results in severe *interference* conditions for cell-edge terminals. It has then become increasingly clear that the bottleneck of the future wireless networks will be the management of interference [4].

2.1.2 Downlink Multi-user Single-cell Transmission

In the last decade, an impressive number of works have been focused on the downlink transmission where one single transmitter (TX) serves multiple receivers (RXs). In the information-theoretic community, this trans-

mission scenario is well known as the *broadcast channel (BC)* [5, 6]. This scenario has been heavily investigated both in the information theoretic society and in the industry, and is now relatively well understood. The capacity of the Gaussian multiple-input multiple-out (MIMO) channel has been obtained [7, 8] and shown to be achieved by a non-linear scheme called dirty-paper coding (DPC) in which the interference are subtracted on the TX side [9]. In addition, the performance of linear precoding has been evaluated [10, 11] and it has become clear that linear precoding is a practically interesting transmission scheme with lower complexity than DPC but good performance. Efficient algorithms have also been developed in order to maximize the performance with regards to different figures-of-merit while having a low complexity [12, 13]. However, the performance improvement can only be obtained at the cost of an accurate knowledge of the channel state at both the TX and the RXs [14, 15].

To translate these theoretic gains into practical performance, it has then been investigated how to estimate and feedback the channel state in realistic scenarios. Methods to obtain accurate feedback at the TX at low cost have been developed [16] while the impact of having imperfect channel state information (CSI) at the TX has been evaluated [15]. It has also be shown how channel dependent scheduling could help improve the performance and make the transmission more robust to imperfect CSIT [17, 18]. A comprehensive study of multiuser-MIMO transmissions with linear precoding is provided in [19].

Even with the novel developed schemes, obtaining perfect CSIT remains unrealistic due to the changing nature of the channel. Therefore, transmission methods being more robust to imperfect CSIT have been provided, optimizing either the average performance over the CSIT errors [20] or optimizing the worst case behavior [21, 22]. Another line of work aiming at exploiting *delayed* CSIT has been triggered by the work [23] where it was shown that even completely outdated CSIT (not correlated with the current channel state) could help improve the performance over a setting without CSIT. Since then, many works have study how to exploit delayed CSIT (See [24–26] among others).

Finally, using TXs with a very large number of antennas, so-called *massive MIMO*, has been recently advocated in [27] as a solution to improve further the performance while easing the requirements in terms of signal processing and CSI. It is now considered a promising method and is the focus of the research of an increasingly large community. It is investigated both by companies developing prototypes of such TXs and by the academic world (see [28–30], among others).

2.1.3 Multi-cell Processing

Although the progresses and innovations done regarding the single-cell transmission have lead to great performance improvements, they remain fundamentally limited by the inter-cell interference. Thus, TX cooperation has appeared recently as the key to further performance improvements [4]. One conventional method to reduce inter-cell interference is by coordinating resource allocation via flexible and coordinated scheduling. Different frequency allocations schemes have been proposed with the goal to adapt to the interference generated in order to improve the transmission efficiency [31–34].

Without user's data sharing: Coordinated Beamforming (CB)

The use of multiple-antennas at the TX offers additional opportunities for the TX cooperation. If each user is served only by one TX via linear precoding, it is possible to design the precoder (also called beamformer) so as to emit little inter-cell interference, so-called *coordinated beamforming* [39]. With a single-antenna at each RX, the coordination of the TXs can be done based on mostly local CSIT and efficients methods have been found to optimize the precoder design or the feedback [40–48].

In contrast, with multiple-antennas available at the RXs, the transmission paradigm changes completely. It becomes then possible to *align* interference at a restricted number of dimensions such that a RX can then suppress the remaining interference via RX zero forcing (ZF). This precoding scheme has been called *interference alignment (IA)* [37, 38, 49] and has lead to an impressive number of new algorithms (See [51–53] among others) and analysis (See [49, 50, 54], among others). In IA, the interference are ZF jointly at the TXs and the RXs which can lead to a strong improvement of performance. However, it requires a higher degree of coordination among the TXs since all the TXs have to agree on the RX dimensions in which all the interference are restricted. Hence, all the algorithms cited above require global CSI at *every* TX. Without data sharing between the TXs (i.e., among the coordinated beamforming techniques), IA represents the transmission scheme with the strongest potential and the higher feedback requirements. Hence, in the absence of user's data symbol sharing, we will always consider that the RXs have multiple-antenna such that IA can be applied.

With user's data sharing: Joint Precoding (JP)

When the user's data symbols can be shared to several TXs via for example a backhaul network, it is then possible for one user to be served jointly by several TXs. This scenario whereby multiple interfering TXs share user messages and allow for joint precoding, denoted as "Network MIMO" or "Multi-cell MIMO", is currently considered for next generation wireless networks [4, 55, 56]. With perfect message and CSI sharing, the different TXs can be seen as a unique virtual multiple-antenna array serving all RXs, in a multiple-antenna BC fashion. It is in fact clear that the cooperation through JP allows theoretically for the largest improvement as it requires more exchange of information between the TXs than the other alternatives.

A distinct advantage of TX cooperation over conventional approaches relying on egoistic interference rejection at the RXs, lies in the reduced number of antennas needed at each RX to ZF residual interference. This gain is further amplified when user data messages exchange among TXs is made possible. For instance, in the case of three interfering two-antenna TXs, relying on RX based interference rejection alone requires three antennas at each RX to ZF the interference, while just two are needed when coordination is enabled via IA. Further, if the three user messages are exchanged among the TXs, thus enabling JP, then just one antenna per TX and RX is sufficient to preserve interference-free transmission.

2.2 The Challenges of Obtaining CSIT

2.2.1 Imperfect CSIT

Similarly to the single-cell case, the benefits of multiple antenna transmit cooperation go at the expense of requiring CSI at the TXs (CSIT). Indeed, it becomes possible to coordinate the actions of the TXs only if they share the knowledge relative to the state of the wireless channel. Obtaining this CSIT is even more challenging in multi-cell cooperation because in most of the schemes (IA and JP), *global* multi-user CSI is necessary at *every TX*.

Consequently, the study of how CSIT requirements for IA methods can somehow be alleviated has become an active research topic in its own right [51, 52, 59–61]. An overview of the practical challenges of IA and of some possible solutions is provided in [62]. Another line of work consists in studying the minimal number of CSI quantization bits that should be conveyed to the TXs to achieve some given number of degrees-of-freedom (DoF) using IA [59, 60]. It has then been investigated how to reduce the

requirements by optimizing the design of the RX and TX filters [63, 64]. Another approach denoted as *topological interference management* has been recently introduced in [65]. Instead of starting from full CSIT and reducing the requirements, it considers the problem of exploiting only topological information of the network (i.e., connectivity) to obtain very robust schemes in terms of CSI requirements.

Regarding JP across distant TXs, the algorithms designed for the conventional BC can then be applied in the multi-cell setting. Distributed schemes based on iterative updates of the transmit coefficients were also designed to avoid the requirements of explicit CSI at the TXs, both for IA [51] and for joint precoding [66]. However, the computation of precoders typically relies on iterative techniques where each iteration involves the acquisition of local feedback. As local feedback is updated over the iterations, this approach implicitly allows each TX to collect information about the precoders and channels of other TXs, hence amounting to an iterative global CSI acquisition at all TXs.

As a result, JP is usually limited to small *cooperation clusters* inside which the TXs exchange their CSI and cooperate. The optimal way of forming these clusters has recently become an active research topic [67–71]. Still, clustering leads to some fundamental limitations. Firstly, there is inevitably inter-cluster interference on the boundaries of the cluster and secondly, it requires the obtaining at all the TXs inside the cluster of the CSI relative to the entire cluster. This means that the amount of CSI feedback required quickly increases with the number of TXs inside the cluster. Several works have focused on determining the optimal size of the clusters when taking into account the cost of estimating the channel elements, e.g., [72,73]. They suggest that TX cooperation cannot efficiently manage interference, even if the backhaul links are strong enough to form large clusters.

2.2.2 The Distributed Channel State Information Setting

As it was mentioned in the previous section, a large body of literature has been focused on the problem of TX cooperation with imperfect CSIT. However, it is usually assumed that the channel estimate, although imperfect, is the *same* at all the TXs involved in the joint processing. This means that either the precoding is done in a central node or that the precoding is distributed across the TXs with the channel estimate being *perfectly shared* between the TXs. This CSIT scenario will be called hereafter the *centralized* CSIT case. This assumption comes partly from the legacy of previous works where all the transmit antennas were colocated so that this assumption was

justified, and partly because it makes the model simple and intuitive.

However, this assumption is likely to be breached in many scenarios where the TXs are not colocated. Indeed, precoding in a centralized node can be considered only in some scenarios and requires a centralized architecture which does not scale well with the number of cooperating TXs. With distributed precoding, the CSI acquisition is inherently acquired at each TX through a different feedback channel. Two scenarios are actually considered for the acquisition of the CSIT in wireless networks and both scenarios are illustrated in Figure 2.1.

The first one consists in direct broadcast of the local estimates from each RX to all the listening TXs. This scenario is interesting as it does not require any CSIT sharing through the backhaul network. It is however not possible in the current 3GPP LTE-A standards [74].

The alternative is an over-the-air feedback from the UE to the *home* base station alone, followed by an exchange of the local estimates over the backhaul, as it is currently advocated by 3GPP LTE-A standards [74]. Sharing the channel estimates without delay and without quantization requires expensive fiber-based backhaul links or dedicated wireless links which will not be available everywhere or will be too costly. In many settings, the CSI sharing will therefore not be possible without further quantization loss and without a certain delay due both to scheduling and to protocol latency.

Either case, the channel estimates available at the various TXs will not be exactly the same. This becomes particularly clear as the number of cooperating TXs increases. Indeed, in both cases, the practical difficulties of acquiring the CSI at the TXs in a timely manner become more challenging as the number of TXs increases. In addition, the amount of CSI to provide to each TX increases very quickly with the size of the network: every TX has to receive the whole multi-user channel matrix which is of size $K \times K$ for a setting with K TX/RX pairs and only single-antenna nodes.

This means that a form of CSI discrepancy between the channel estimates at the different TXs will arise in many scenarios. In order to capture multiple-antenna precoding scenarios whereby different TXs obtain an imperfect *and* imperfectly shared estimate of the overall multi-user channel, we introduce the framework of *distributed CSIT*. In this CSIT configuration, TX j is assumed to have received its *own* estimate of the multi-user channel, denoted by $\mathbf{H}^{(j)}$, before the transmission occurs. It then designs its transmit parameters solely based on this channel estimate and without additional communications with the other TXs.

In spite of its practical relevance, very few works have considered this CSIT configuration, although the analysis of distributed cooperation is gain-

ing in momentum. In [75], a robust algorithm was designed for joint precoding across two TXs having distributed CSIT. However, the algorithm provided is computationally demanding and does not provide any insight. In [76], the capacity of the two-user interference channel (IC) is studied when each TX knows perfectly a subset of the channel coefficients which are fixed. This work discusses whether it is possible to transmit reliably without knowing some channel coefficients. It does not study however the transmission with imperfect and imperfectly shared CSIT. Recently, distributed scheduling with local state information has also been investigated in [82] and a heuristic algorithm to efficiently exploit the local information available has been developed. Posterior to our publications, some works have pursued analyzing the impact of distributed CSIT over both JP across the TXs [77] and IA [78, 79].

Several fundamental questions follow from this distributed CSIT configuration and have not yet been answered. Although these questions prove to be difficult and to a large extent remain open, we shed some light on this problem in this thesis. Specifically, we show that significant gains can be realized from the consideration of the distributed CSIT framework.

In the first part of this thesis, we study the design of precoders based on distributed CSIT. Considering first JP, we evaluate the performance of conventional precoders designed for the case of *centralized* CSIT when confronted to distributed CSIT. We show the extremely deleterious impact of the CSIT discrepancies over the performance and we discuss the design of robust precoders. We also study how distributed CSIT impact the CSI feedback requirements. Finally, we discuss the impact of distributed CSIT over the IA algorithms.

In the second part of this thesis, we consider the other face of the problem and we assume this time the transmission scheme to be fixed and we study the spatial allocation of feedback resources. We study what are the requirements for an efficient TX cooperation in terms of feedback and backhaul architecture. In particular, we study how complete and accurate should the estimate $\mathbf{H}^{(j)}$ be for each j given some performance requirements. By optimizing directly the spatial allocation of the CSIT, we go closer towards providing each TX solely with the information which it really needs for an efficient transmission. We show how this approach leads to strong reductions of the CSIT requirements at *virtually* no cost, and is therefore promising to make TX cooperation more practical under the constraints of realistic networks.

2.3 Contributions and Publications

2.3.1 Contributions Presented in this Thesis

The contributions presented in this thesis are as follows.

- **DoF of JP with distributed CSIT:** When the TXs apply joint precoding with distributed CSIT, we derive the DoF, or prelog factor (which will be defined and discussed in Chapter 3), as a function of the quality of the CSIT available at each TX when conventional ZF is applied at every TX. It is shown that the DoF at *every* user is limited by the worst CSIT quality across all TXs and across all RXs. This is in strong contrast with the results with imperfect *centralized* CSIT where the quality of the feedback of user i impacts only the DoF at user i [15]. We also provide some simple schemes improving the achieved DoF. These results were published in

  Paul de Kerret and David Gesbert, "The multiplexing gain of a two-cell MIMO channel with unequal CSI", in proc. IEEE International Symposium on Information Theory (ISIT), 2011

  Paul de Kerret and David Gesbert, "Degrees of freedom of the network MIMO with distributed CSI", in IEEE Trans. Inf. Theory, vol. 58, no. 11, pp. 6806-6824, Nov. 2012

- **DoF of IA with distributed CSIT:** Considering some particular IA schemes with distributed CSIT, we lower bound the DoF achieved. The results suggest a similar behavior as when JP is used in the sense that the quality of one channel estimate at one TX impacts all the other users. These results were published in

  Paul de Kerret, Maxime Guillaud, and David Gesbert, "Degrees of freedom of certain interference alignment schemes with distributed CSIT", in Proc. IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC), 2013.

- **Rate loss of JP with distributed CSIT:** Considering JP with distributed CSIT, we go beyond the DoF analysis to discuss the rate loss incurred by the CSI inconsistencies across the TXs. We provide design

guidelines for the feedback schemes to ensure that the performance remains within some given bounds of the performance achieved when perfect CSIT is available at all TXs. This work has been published in

Paul de Kerret, Jakob Hoydis, and David Gesbert, "Rate Loss Analysis of Transmitter Cooperation with Distributed CSIT", in Proc. IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC), 2013.

Paul de Kerret, Jakob Hoydis, and David Gesbert, "Rate Loss Analysis of Transmitter Cooperation with Distributed CSIT", to be submitted, 2013

- **Distance-dependent CSI spatial allocation for JP:** Turning to the optimization of the spatial CSIT allocation and considering first JP, we study how the spatial allocation should be made dependent on the pathloss geometry of the wireless networks. We study an extension of the DoF analysis called *generalized* DoF (which will be defined and discussed in Chapter 3), and we show that the maximal generalized DoF can be achieved in an arbitrarily large network with each TX having only access to the CSI relative to a finite neighbourhood surrounding him. This result provides a theoretical foundation for the well known intuition that the quality with which a channel coefficient is known at a TX should decrease as the distance between the TX and the RX concerned by the channel coefficient increases. This work shows that it is possible to overcome the fundamental limitations of clustering by optimizing directly the spatial CSIT allocation, hence allowing for global interference management with only cooperation on a local scale. These results have can be found in

  Paul de Kerret and David Gesbert, "CSI feedback allocation in multicell MIMO channels", in Proc. International Conference on Communications (ICC), 2012.

  Paul de Kerret and David Gesbert, "Spatial CSIT allocations policies in network MIMO", submitted to IEEE Trans. Inf. Theory, June 2013.

- **IA with incomplete CSIT sharing:** Turning to IA, we provide the

first IA algorithm achieving *perfect* IA with only incomplete CSIT, where the incompleteness refers to some TXs having only the knowledge of a subset of channel coefficients. More generally, we exploit the distributed (here incomplete) CSIT framework to develop an IA algorithm achieving IA with significant reductions of the CSIT requirements. Particularly striking it the fact that these savings come at no cost since perfect IA is achieved. These results can be found in

Paul de Kerret and David Gesbert, "MIMO interference alignment algorithms with hierarchical CSIT", in Proc. IEEE International Symposium on Wireless Communication Systems (IWCS), 2012.

Paul de Kerret and David Gesbert, "MIMO interference alignment with incomplete CSIT", submitted to IEEE Trans. on Wireless Commun., Nov. 2012.

2.3.2 Other Contributions

A few other results are not mentioned in this thesis but have been also obtained in the course of the doctorate while dealing with the problems of interference management. They are briefly mentioned below.

- **Asymptotic performance of opportunistic distributed scheduling:** We investigate interference management through distributed scheduling in two typical scenarios of wireless communications. In the first one, all the users have the same average SNR, which models a dense network with homogeneous pathloss, while in the second the users are uniformly located around the TXs, which models a network for mobile communications. The asymptotic behavior of the average rate when the number of users becomes large is studied. It is shown that distributed opportunistic scheduling can efficiently manage interference and achieve close to the optimal performance without interference. However, the rate of convergence is calculated and shown to be extremely slow. These results have been published in

  Paul de Kerret, and David Gesbert, "The asymptotic limits of interference in multicell networks with channel aware scheduling", in Proc. IEEE International Workshop on Signal Processing Advances for Wireless Communications (SPAWC), 2011.

- **Sparse precoding in network MIMO channels:** Sharing the users data to all the cooperating TXs might create a too large burden on the backhaul architecture. Thus, we study how to optimally share the users data symbols in the case of joint precoding across distant TXs when there is a constraint on the total number of users data symbols shared in the backhaul. We develop a *sparse precoding* algorithm which aims at minimizing the number of non-zeros coefficients in the precoder (i.e. the number of users data symbols shared in the backhaul) while achieving given performance requirements. These results have been published in

  Rajeev Gangula, Paul de Kerret, and David Gesbert, "Optimized data symbol allocation in multicell MIMO channels", in proc. IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC), 2011.

  Paul de Kerret, and David Gesbert, "Sparse precoding in multicell MIMO systems", in proc. IEEE Wireless Communications and Networking Conference (WCNC), 2012.

  Paul de Kerret, and David Gesbert, "A practical precoding scheme for multicell MIMO channels with partial user's data sharing", in proc. IEEE Wireless Communications and Networking Conference Workshops (WCNCW), 2012.

- **Joint precoding with backhaul delays:** We study a transmit scenario where distant TXs cooperate by exchanging the locally available CSI via backhaul links introducing some delays. In the line of the recent works from [23], we investigate how to adapt the results presented in this work for the case of distributed CSIT to the scenarios where the TXs receive delayed CSIT. A scheme combining the approach for distributed CSIT [81] and for delayed CSIT [23] is derived and shown to achieve the optimal DoF. These results have been published in

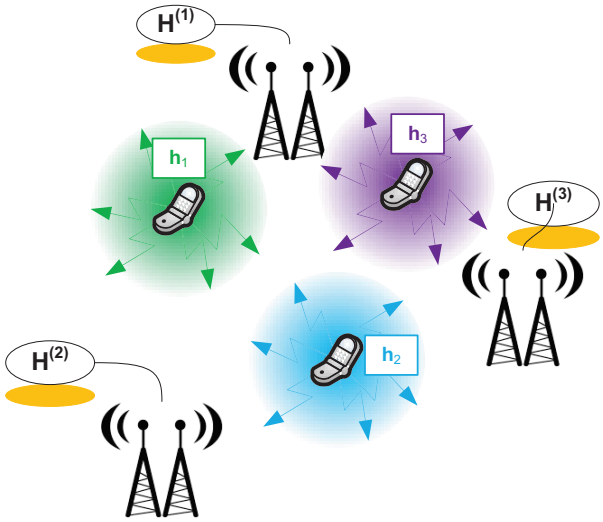
  Xinping Yi, Paul de Kerret, and David Gesbert, "The DoF of network MIMO with backhaul delays", in Proc. IEEE International Conference on Communications (ICC), 2013.

- **Precoding with delayed CSIT:** We consider this time a multiple-input single-output (MISO) BC with a single TX which receives a

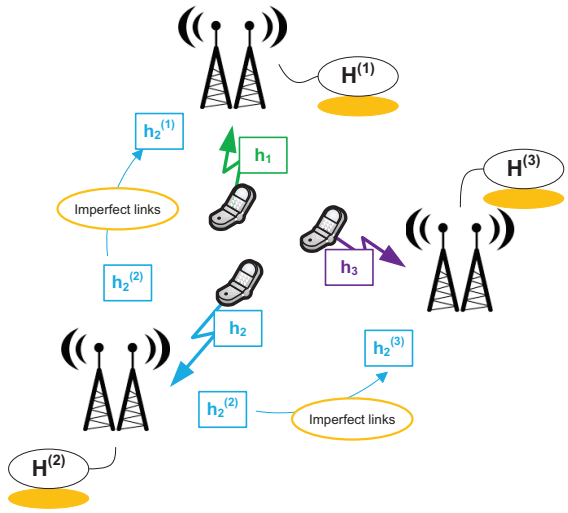
delayed channel estimate being correlated with the true channel state. With only two RXs, it has been shown in [24, 83] how to optimally exploit this CSI in terms of DoF. In particular, a combination of the methods in [23] and conventional ZF is used. We study how to generalize this approach to an arbitrary number of users. Although it was not possible to obtain the exact DoF characterization, a new precoding scheme outperforming schemes from the literature is developed and a new outerbound is derived to evaluate the performance of this precoding scheme. These results have been published in

Paul de Kerret, Xinping Yi, and David Gesbert, "On the degrees of freedom of the K-user time correlated broadcast channel with delayed CSIT", in proc. IEEE International Symposium on Information Theory (ISIT), 2013.

Paul de Kerret, Xinping Yi, and David Gesbert, "On the degrees of freedom of the K-user time correlated broadcast channel with delayed CSIT", extended journal version available under arxiv, 2013.



(a) Broadcast of the CSI



(b) CSI exchange between the TXs

Figure 2.1: In Figure a, the *broadcast* scenario is described: every RX directly transmits its CSI in a broadcast manner to all the TXs. In Figure b, we present the *sharing* scenario where every RX transmits its CSI to its serving TX which then forwards it to the other cooperating TXs. The channel from all the TXs to RX i is denoted by $\boldsymbol{h}_i$ while we use the superscript (j) to denote the estimate received at TX j.

Chapter 3

System Model and Problem Statement

3.1 Multi TXs Transmission

3.1.1 Received Signal

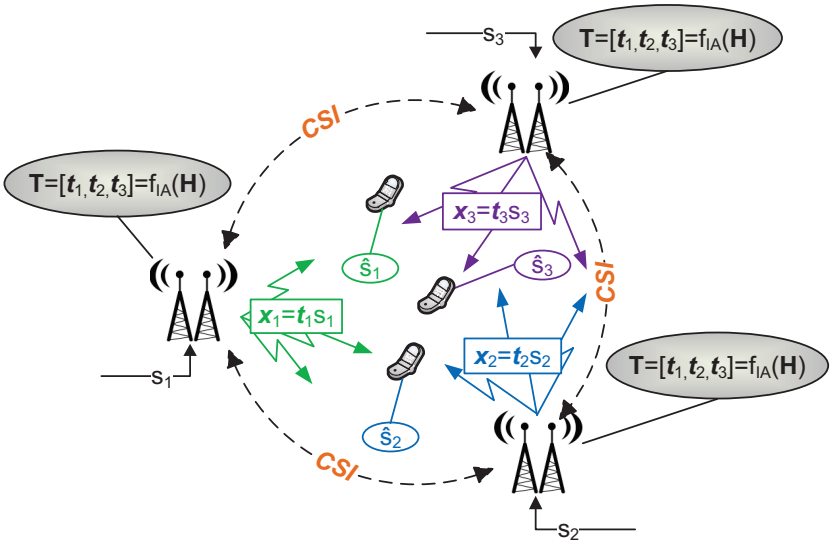
We study the transmission from K TXs to K RXs where the ith TX and the ith RX are equipped respectively with M_i and N_i antennas. We then denote the total number of RX antennas by

$$N_{\text{tot}} \triangleq \sum_{i=1}^{K} N_i \tag{3.1}$$

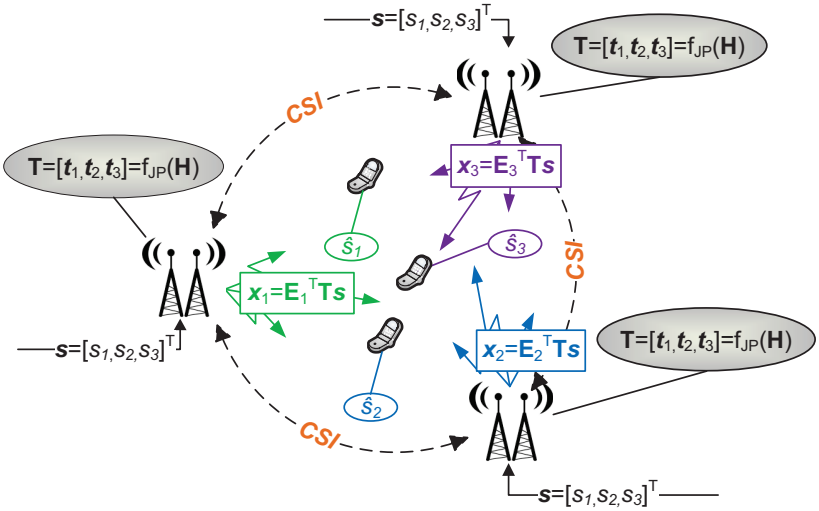
and the total number of TX antennas spread over the TXs by

$$M_{\text{tot}} \triangleq \sum_{i=1}^{K} M_i. \tag{3.2}$$

We further assume that the RXs have perfect CSI as we decide to focus primarily on the challenge of conveying CSI back to TXs through some form of feedback. We consider in this thesis that linear precoding and linear filtering are used and that the RXs treat interference as noise. We consider solely in this thesis the transmission of a single-stream to each user. These assumptions are designed to preserve clarity of exposition while preserving the essence of the novelty in our contributions. Extensions to multiple-streams are sometimes possible and will be discussed whenever relevant. The channel from the K TXs to the K RXs is represented by the channel



(a) Interference alignment



(b) Joint precoding

Figure 3.1: The description of a distributed IA algorithm is done in Figure a while Figure b represents the distributed precoding in a Network MIMO with user's data sharing. The matrix $\mathbf{E}_i^{\mathrm{T}}$ is a matrix which selects the rows of the multi-user precoder $\mathbf{T}$ which corresponds to the antennas at TX i.

matrix $\mathbf{H}^{\mathrm{H}} \in \mathbb{C}^{N_{\mathrm{tot}} \times M_{\mathrm{tot}}}$ where $\mathbf{H}_{ik}^{\mathrm{H}} \in \mathbb{C}^{N_i \times M_k}$ denotes the channel matrix from TX k to RX i and has all its elements i.i.d. as $\mathcal{CN}(0, \rho_{ik}^2)$ and independent of the other channel matrices. This corresponds to the conventional model in the literature for a Rayleigh fading environment. Note however that several results (the ones relative to the DoF) extend to any continuous distribution satisfying some mild constraints. The $\{\rho_{ik}^2\}_{i,k}$ represent the pathloss attenuation and allow to consider different networks geometries.

The transmission is then described as

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_K \end{bmatrix} = \mathbf{H}^{\mathrm{H}}\mathbf{x} + \boldsymbol{\eta} = \begin{bmatrix} \mathbf{H}_1^{\mathrm{H}}\mathbf{x} \\ \vdots \\ \mathbf{H}_K^{\mathrm{H}}\mathbf{x} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}_1 \\ \vdots \\ \boldsymbol{\eta}_K \end{bmatrix} \tag{3.3}$$

where $\boldsymbol{y}_i \in \mathbb{C}^{N_i \times 1}$ is the signal received at the i-th RX, $\mathbf{H}_i^{\mathrm{H}} \in \mathbb{C}^{N_i \times M_{\mathrm{tot}}}$ the channel from all TXs to the i-th RX, and $\boldsymbol{\eta} \triangleq [\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_K]^{\mathrm{T}} \in \mathbb{C}^{N_{\mathrm{tot}} \times 1}$ the normalized Gaussian noise with its elements i.i.d. as $\mathcal{CN}(0, 1)$. We denote the average per-TX transmit power by P and we also call P the average SNR.

The multi-TX transmitted signal $\mathbf{x} \in \mathbb{C}^{M_{\mathrm{tot}} \times 1}$ is obtained from the symbol vector $\boldsymbol{s} \triangleq [s_1, \ldots, s_K]^{\mathrm{T}} \in \mathbb{C}^{K \times 1}$ with its elements i.i.d. $\mathcal{CN}(0, 1)$ as

$$\mathbf{x} = \sqrt{P}\mathbf{U}\boldsymbol{s} \tag{3.4}$$

where the matrix $\mathbf{U} \in \mathbb{C}^{M_{\mathrm{tot}} \times K}$ is the normalized precoder matrix. Depending on how the users data symbols are shared to the TXs, it can either take any value respecting the power constraint or a block-diagonal form, as explained in the following. Note that we will sometimes consider also the total precoder $\mathbf{T} = \sqrt{P}\mathbf{U}$.

3.1.2 Precoding Schemes with Perfect CSIT

We present now the precoding schemes that we will consider in this thesis and which correspond to the most widely used transmission schemes. We present the main principles of the transmission schemes in the case where the CSI is *perfectly* known at each TX. A transmission with imperfect centralized CSIT is straightforwardly deduced from it as it suffices to replace the true channel matrix by the channel estimate. In contrast, the distributed CSIT scenario where the CSI is imperfectly shared between the TXs is completely different and calls for innovative designs. This is one of the central points of the thesis.

With Users Data Sharing: Joint Precoding (JP)

When all the users data symbols are available at each TX, there is no constraint on the structure of the precoder $\mathbf{U}$, at the exception of the normalization constraint. Hence, the TXs can apply a joint precoder (JP) to serve all the users collaboratively. Many precoding schemes exist, ZF [57,58], regularized ZF [20,84], sum rate maximizing [13], among others. Since we are interested in the problem of interference management, we will consider the (interference limited) high SNR regime. Hence, we assume that the TXs aim at completely removing the interference and use ZF precoding. We present here only the transmission with single-antenna RXs as it will be our focus when considering joint precoding across the TXs.

ZF is well known to achieve the maximal DoF in the MIMO BC with perfect CSIT [15,19]. Furthermore, considering limited feedback in the compound MIMO BC, it is revealed in [85] that no other precoding scheme can achieve the maximal DoF with a lower feedback scaling of the feedback rate. This confirms the efficiency of ZF in terms of DoF, even when confronted to imperfect CSI.

Remark 1. *Regularizing the matrix inversion by adding an identity, so called "regularized ZF" outperforms conventional ZF when confronted to imperfect CSIT. However, it does not reduce the interference but simply increases the strength of the desired signal by reducing the cost of the interference management [11, 20]. Intuitively, the TX does not waste its resource in managing the interference which, in any case, cannot be efficiently suppressed because of the CSI imperfections. Consequently, we will consider conventional ZF as this simplifies the analysis without changing the fundamental insights of interference management. This question will be further discussed in Chapter 4.* □

There are many possibilities of ZF precoding depending on the power normalization used. One of them, which will be considered in Chapter 5, is defined as $\mathbf{U}^\star \triangleq [\boldsymbol{u}_1^\star, \ldots, \boldsymbol{u}_K^\star] \in \mathbb{C}^{M_{\text{tot}} \times K}$ where

$$\boldsymbol{u}_i^\star \triangleq \left(\mathbf{I}_K - \mathbf{H}_i \left(\mathbf{H}_i^{\text{H}} \mathbf{H}_i\right)^{-1} \mathbf{H}_i^{\text{H}}\right) \boldsymbol{h}_i, \qquad \forall i \in \{1, \ldots, K\} \tag{3.5}$$

with

$$\mathbf{H}_i \triangleq \left[\boldsymbol{h}_1 \ \ldots \ \boldsymbol{h}_{i-1} \ \boldsymbol{h}_{i+1} \ \ldots \ \boldsymbol{h}_K\right], \qquad \forall i \in \{1, \ldots, K\}. \tag{3.6}$$

We use throughout this thesis the superscript $\star$ to denote the precoders obtained based on perfect CSIT.

Remark 2. *The ZF beamformer $\boldsymbol{u}_i^\star$ is not instantaneously normalized. Yet, it can easily be seen that* $\mathrm{E}[\|\boldsymbol{u}_i^\star\|^2] = 1, \forall i$ *and* $\mathrm{E}[\|\boldsymbol{e}_j^{\mathrm{H}}\mathbf{U}^\star\|^2] = 1, \forall j$*. This means that a per-user and a per-TX power constraint are fulfilled on average.* □

Without Users Data Sharing: Interference Alignment (IA)

When the users data symbols are not shared, the precoder $\mathbf{U}$ is restricted to a particular block-diagonal form. Indeed, the signal emitted by TX j is given by

$$\mathbf{x}_j = \sqrt{P}(\mathrm{E}_j)^{\mathrm{H}}\mathbf{U}\boldsymbol{s} \tag{3.7}$$

where $\mathrm{E}_j \in \mathbb{C}^{M_{\mathrm{tot}} \times M_j}$ is defined as

$$\mathrm{E}_j \triangleq \begin{bmatrix} \mathbf{0}_{\sum_{i=1}^{j-1} M_i \times M_j} \\ \mathbf{I}_{M_j \times M_j} \\ \mathbf{0}_{\sum_{i=j+1}^{K} M_i \times M_j} \end{bmatrix}. \tag{3.8}$$

Since TX j has only access to the data symbol s_j, this means that only the jth column of the matrix $(\mathrm{E}_j)^{\mathrm{H}}\mathbf{U}$ can be nonzero. This is the necessary condition so that the transmitted symbol at TX j only depends on the data symbol s_j. Note that we will slightly abuse notations by denoting by $\boldsymbol{u}_i$ the vector of size $M_i \times 1$ instead of the vector of size $M_{\mathrm{tot}} \times 1$. This means that we keep the same notation when considering the beamformer after the coefficients fixed to zero have been removed.

In that scenario, it is now well known that it is asymptotically optimal at high SNR to align interference in a restricted number of dimensions [50]. We say that IA is achieved if it is possible to transmit all streams interference-free. Denoting by $\boldsymbol{g}_i^{\mathrm{H}} \in \mathbb{C}^{1 \times N_i}$ the RX filter applied at RX i, this means that the RX filter $\boldsymbol{g}_i^{\mathrm{H}}$ should be able to ZF all the received interference. Mathematically, this is written as

$$\boldsymbol{g}_i^{\mathrm{H}}\mathbf{H}_{ij}^{\mathrm{H}}\boldsymbol{u}_j = 0, \qquad \forall j \neq i. \tag{3.9}$$

Our focus is not on the design of IA algorithms as many IA algorithms are already available in the literature (See [52, 86–90], among others). However, we present briefly the principle of the *minimum (min-) leakage algorithm* from [86] because this algorithm is the most simple one and contains the main principles which are used in most of the more complicated algorithms in the literature.

The min-leakage algorithm can be described in our setting as follows. The algorithm minimizes the sum of the interference power created at the RXs which is called I_{IA} and is equal to

$$I_{\mathrm{IA}} \triangleq \sum_{i=1}^{K} \sum_{k=1,k\neq i}^{K} |\boldsymbol{g}_i^{\mathrm{H}} \mathbf{H}_{ik}^{\mathrm{H}} \boldsymbol{u}_k|^2. \tag{3.10}$$

The algorithm is based on an alternating minimization in which the TX beamformers are first obtained from the RX beamformers as

$$\boldsymbol{u}_k = \mathrm{eig}_{\min} \left(\sum_{i=1,i\neq k}^{K} \mathbf{H}_{ik} \boldsymbol{g}_i \boldsymbol{g}_i^{\mathrm{H}} \mathbf{H}_{ik}^{\mathrm{H}} \right), \qquad \forall k \tag{3.11}$$

where $\mathrm{eig}_{\min}$ is the operator which returns the smallest eigenvector of the Hermitian matrix taken in argument. Similarly, the RX beamformers at all RXs are then obtained from the TX beamformers as

$$\boldsymbol{g}_k = \mathrm{eig}_{\min} \left(\sum_{i=1,i\neq k}^{K} \mathbf{H}_{ki}^{\mathrm{H}} \boldsymbol{u}_i \boldsymbol{u}_i^{\mathrm{H}} \mathbf{H}_{ki} \right), \qquad \forall k. \tag{3.12}$$

The TX and RX beamformers are updated iteratively until convergence to a local minimizer

We consider exclusively in this thesis MIMO IA for given channel realizations. In that case and considering only single-stream transmissions, the feasibility of IA depends only on the antenna configuration, i.e., whether there are enough antennas at the TXs and the RXs to ZF all interference [50]. The IA feasibility problem will be discussed in more details in Chapter 8.

3.2 Figures of Merit: Average Rate, DoF, Generalized DoF

3.2.1 Average Rate

Our main metric of interest is the average rate per user as it can be used to measure the average spectral efficiency of the transmission scheme. Since the data symbols are assumed to be i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$, the average rate of user i is equal to [6]

$$R_i \triangleq \mathrm{E}\left[\log_2 \left(1 + \frac{P|\boldsymbol{g}_i^{\mathrm{H}} \mathbf{H}_i^{\mathrm{H}} \boldsymbol{t}_i|^2}{1 + \sum_{j\neq i} P|\boldsymbol{g}_i^{\mathrm{H}} \mathbf{H}_i^{\mathrm{H}} \boldsymbol{t}_j|^2} \right) \right]. \tag{3.13}$$

Remark 3. *Note that without loss of generality (w.l.o.g.), the noise variance has been normalized to one by dividing by the noise variance both the numerator and the denominator in* (3.13) *which comes down to integrate the noise variance in the variance of the channel. The general expression in* (3.13) *will be rewritten in simplified forms when considering JP or IA.* □

3.2.2 Degrees-of-Freedom

As already mentioned earlier, our primary interest is on the high-SNR interference-limited regime in which the DoF is an interesting figure-of-merit. The DoF at user i, or prelog factor, is defined as [35]

$$\mathrm{DoF}_i \triangleq \lim_{P\to\infty} \frac{R_i}{\log_2(P)}. \tag{3.14}$$

Although an imperfect figure-of-merit, the DoF allows to obtain analytical results in complicated transmission scenarios. The insights obtained from the DoF analysis have already been the key to the development of many important new schemes, e.g., MIMO transmission [36], IA [37], delayed CSIT [38], etc... We will also show in this thesis how it sheds lights into the problem of interference management with distributed CSIT.

3.2.3 Generalized Number of Degrees-of-Freedom

Considering the limiting regime where the SNR becomes increasingly large has for consequence that the (finite) pathloss differences vanish in the analysis and do not impact the analysis. Indeed, any finite pathloss, even extremely large, represents a multiplicative term which does not affect the power exponent of the received signal. This is not problematic in a setting where all the wireless links are of the same order of magnitude. However, when this is not the case, a conventional DoF analysis will not take the pathloss geometry into account and will not provide accurate informations. We present now a toy example to emphasize our point.

Example 1. *Let us consider an IC with two TX/RX pairs and every node having a single-antenna. They interfere to each other via a channel of variance* $\alpha \in (0,1)$ *while the direct links have unit variance. A conventional DoF analysis (we use the term conventional to contrast with the generalized DoF analysis) states that the DoF is equal to* 0.5 *independently of the value of* α *[91]. However, if the TXs interfere with very low power, e.g.,* $\alpha = 10^{-12}$*, then the interference will be negligible for any realistic range of power used for the transmission. Thus, simulations would give a pre-log factor of the*

rate as a function of the SNR equal to 1 *for a wide range of practical SNRs. In that case, we can say that the DoF analysis does not fully represent the real behaviour of the average rate achieved in any practical transmission and does not provide meaningful enough insight into the actual transmission performance.*

In order to allow the rate analysis to depend on a richer set of parameters illustrating the impact of the network topology, we advocate the use of the *generalized* DoF introduced in [91] and used since in many other works [91–97]. It consists in using a model where the strength of any interference links is parameterized as a function of the operating SNR. This dependency of the pathloss with the SNR allows to take the network geometry into account in the analysis. The generalized DoF at RX i is then defined by

$$\mathrm{DoF}_i(\mathbf{\Gamma}) \triangleq \lim_{P\to\infty} \frac{R_i}{\log_2(P)} \quad \text{, subject to } \rho_{ij}^2 P = (\rho_{ii}^2 P)^{\{\mathbf{\Gamma}\}_{ij}}, \quad \forall i,j, \quad (3.15)$$

where the matrix $\mathbf{\Gamma} \in [-\infty, 1]^{K\times K}$ is called the *interference level matrix* and can be written as

$$\{\mathbf{\Gamma}\}_{ki} \triangleq \frac{\log_2(\rho_{ki}^2 P)}{\log_2(\rho_{ii}^2 P)}, \qquad \forall k,i. \quad (3.16)$$

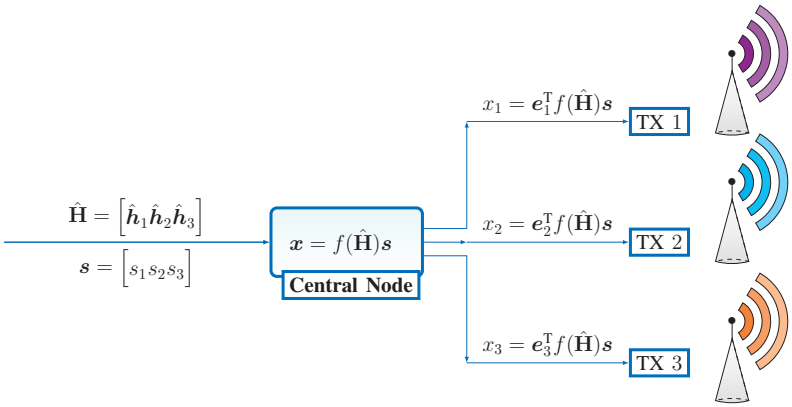
The interference level matrix models the geometry by representing the pathloss differences. The generalized DoF analysis provides an approximation of the capacity achieved in the original transmission setting. A common approach to ensure the accuracy of this approximation consists in upper-bounding the maximal difference between the approximated expression obtained via the generalized DoF approach and the true capacity [91].

3.3 The Distributed CSIT Configuration

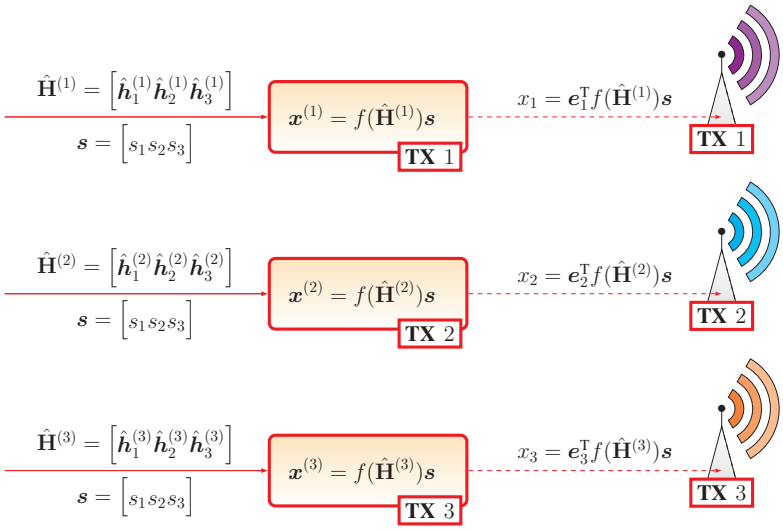
The key specificity of our work comes from taking into account the imperfect sharing of the CSI between the TXs. In this thesis, we refer to such a CSIT configuration as *distributed CSIT* configuration. In each chapter, further precisions will be provided to adapt the general distributed CSIT model to the specific transmission setting being studied. We emphasize the differences with the conventional centralized CSIT configurations in Fig. 3.2.

3.3.1 Distributed CSIT

In the distributed CSIT configuration, TX j receives an estimate $\hat{\mathbf{H}}^{(j)} \in \mathbb{C}^{N_{\mathrm{tot}}\times M_{\mathrm{tot}}}$ of the multi-user channel $\mathbf{H} \in \mathbb{C}^{N_{\mathrm{tot}}\times M_{\mathrm{tot}}}$. The estimate $\hat{\mathbf{H}}^{(j)}$ can



(a) Joint precoding with centralized CSIT



(b) Joint precoding with distributed CSIT

Figure 3.2: Joint precoding with centralized CSIT is shown in Figure a while Figure b represents the joint precoding in the distributed CSIT configuration where each TX receives its own multi-user channel estimate. The estimates $\hat{\boldsymbol{h}}_i$ denote the imperfect channel estimates in the centralized CSIT configuration.

a priori be arbitrarily chosen depending on the transmission scenarios that are considered. One example, which is studied in Chapter 4 and Chapter 5 consists in each element of $\hat{\mathbf{H}}^{(j)}$ being obtained from

$$\{\hat{\mathbf{H}}^{(j)}\}_{ik} = \sqrt{1-(\sigma_{ik}^{(j)})^2}\{\mathbf{H}\}_{ik} + \sigma_{ik}^{(j)}\{\boldsymbol{\Delta}\}_{ik}^{(j)} \tag{3.17}$$

where $\{\boldsymbol{\Delta}\}_{ik}^{(j)} \sim \mathcal{CN}(0,1)$ and $\sigma_{ik}^{(j)} \in (0,1)$ represents the quality of CSIT. Hence, the channel estimate at TX j is locally corrupted by an additive Gaussian noise whose variance varies for each channel element and for each TX.

Another distributed CSIT configuration that we will consider in Chapter 8 consists in letting each TX know perfectly a subset of the channel elements, with the subset of the channel elements being different for each TX. This CSIT configuration will then be denoted as *incomplete CSIT*.

3.3.2 Distributed Precoding

In this thesis, we will refer to distributed precoding for a parallel algorithm suited to the distributed CSIT configuration, which means capable of designing precoders at each of the TX separately, based solely on the CSI estimates available at that particular TX. This estimate is the result of a preamble where TXs have possibly exchanged some CSIT related data. The actual exchange mechanism is out of the scope here and left undetermined for the moment. Once the exchange took place, no more communications is allowed between TXs. This leads to TX j having the knowledge of the global multi-user channel estimate $(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}} \in \mathbb{C}^{N_{\mathrm{tot}} \times M_{\mathrm{tot}}}$. We introduce $(\hat{\mathbf{H}}_{ik}^{(j)})^{\mathrm{H}} \in \mathbb{C}^{N_i \times M_i}$, and $(\hat{\mathbf{H}}_{i}^{(j)})^{\mathrm{H}} \in \mathbb{C}^{N_i \times M_{\mathrm{tot}}}$ in a similar fashion as their counterparts for perfect CSIT.

Once the estimate $\hat{\mathbf{H}}^{(j)}$ has been obtained, TX j computes its transmit signal *without any additional communications with the other TXs*. This requirement follows from the delay introduced by the use of the backhaul. Hence, TX j designs its transmit coefficient $\mathbf{x}_j$ as a function of $\hat{\mathbf{H}}^{(j)}$.

Due to the assumption of distributed precoding, the actual precoder used for the transmission, which we denote by $\boldsymbol{u}^{\mathrm{DCSI}}$, is equal to

$$\boldsymbol{u}_i^{\mathrm{DCSI}} \triangleq \begin{bmatrix} \mathbf{E}_1^{\mathrm{H}} \boldsymbol{u}_i^{(1)} \\ \vdots \\ \mathbf{E}_K^{\mathrm{H}} \boldsymbol{u}_i^{(K)} \end{bmatrix}, \qquad \forall i \in \{1,\ldots,K\}. \tag{3.18}$$

with the matrix $\mathbf{E}_j^{\mathrm{H}}$ used to select the rows corresponding to the antennas at TX j and defined in (3.8). The superscript DCSI stands for *distributed CSI*.

Up to this point, the precoding scheme used to design the transmit coefficient forming the vector $\boldsymbol{u}^{\mathrm{DCSI}}$ has not been discussed. How to optimally design these transmit coefficients is an open question which will be investigated in the following.

Example 2. *One sub-optimal precoding scheme consists in applying a conventional precoding algorithm distributively at each TX. This precoding scheme does not take into account the fact that the other TXs do not share the same channel estimate and will be therefore denoted as the the* naive *approach. Let us consider for example that the users data symbols are shared to the K TXs and that each node is equipped with a single antenna. Naive ZF precoding consists in letting TX j compute the precoding matrix $\mathbf{U}^{(j)} = [\boldsymbol{u}_1^{(j)}, \ldots, \boldsymbol{u}_K^{(j)}] \in \mathbb{C}^{M_{\mathrm{tot}} \times K}$ where*

$$\boldsymbol{u}_i^{(j)} \triangleq \left(\mathbf{I}_K - \hat{\mathbf{H}}_i^{(j)} \left((\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}} \hat{\mathbf{H}}_i^{(j)} \right)^{-1} (\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}} \right) \hat{\boldsymbol{h}}_i^{(j)}, \qquad \forall i \in \{1, \ldots, K\} \tag{3.19}$$

where the matrix $\hat{\mathbf{H}}_i^{(j)}$ contains the channels from the interfered users and is defined as

$$\hat{\mathbf{H}}_i^{(j)} \triangleq \left[\hat{\boldsymbol{h}}_1^{(j)} \ \ldots \ \hat{\boldsymbol{h}}_{i-1}^{(j)} \ \hat{\boldsymbol{h}}_{i+1}^{(j)} \ \ldots \ \hat{\boldsymbol{h}}_K^{(j)} \right], \qquad \forall i \in \{1, \ldots, K\}. \tag{3.20}$$

Although TX j computes the full precoding matrix, only the jth row of the precoder is used by TX j because of the distributed precoding. Indeed, TX j transmits only $x_j = \sqrt{P} \boldsymbol{e}_j^{\mathrm{H}} \mathbf{U}^{(j)} \boldsymbol{s}$ such that this gives in total

$$\boldsymbol{u}_i^{\mathrm{DCSI}} \triangleq \begin{bmatrix} \boldsymbol{e}_1^{\mathrm{H}} \boldsymbol{u}_i^{(1)} \\ \vdots \\ \boldsymbol{e}_K^{\mathrm{H}} \boldsymbol{u}_i^{(K)} \end{bmatrix}, \qquad \forall i \in \{1, \ldots, K\}. \tag{3.21}$$

3.4 Optimal Precoding with Distributed CSIT: A Team Decision Problem

Our objective in this thesis is to study the maximization of one of the figure of merits introduced in Section 3.2 in the case of distributed CSIT. Considering for example the sum rate maximization, the optimization problem of

interest is then:

$$(\boldsymbol{u}_1^{\mathrm{DCSI}}, \ldots, \boldsymbol{u}_K^{\mathrm{DCSI}}) = \underset{(\boldsymbol{u}_1,\ldots,\boldsymbol{u}_K)}{\operatorname{argmax}} \sum_{k=1}^{K} R_k, \text{ s.to } \mathrm{E}[\|\boldsymbol{u}_i\|^2] = 1, \forall i. \tag{3.22}$$

Note that the jth coordinate of $\boldsymbol{u}_i$ is a function of $\hat{\mathbf{H}}^{(j)}$ since it is computed at TX j. This optimization problem is a difficult one and can be modeled as a *Team Decision* problem [98–100]. Indeed, the TXs aim at maximizing a common objective by choosing their transmission parameters on the basis of individual information (the channel estimates $\hat{\mathbf{H}}^{(j)}$). We will now reformulate the optimization problem (3.22) with the adequate formalism in order to highlight how it relates to the conventional team decision problems.

Let us denote by ω the sum rate function, and by $\gamma_j(\hat{\mathbf{H}}^{(j)})$ the signal transmitted by TX j based on its channel state information $\hat{\mathbf{H}}^{(j)}$. TX j can then be seen as player j while $\gamma_j(\hat{\mathbf{H}}^{(j)})$ becomes the "strategy", or "decision", of player j. Depending on the cooperation scenario (coordinated beamforming or joint precoding), the function $\gamma_j(\hat{\mathbf{H}}^{(j)})$ takes its value in a different space. Note that we can restrict ourselves to scalar "decisions", since a TX with multiple-antennas is simply modeled as two players having the same information. The objective to maximize is then represented as $\mathrm{E}[\omega(\gamma_1(\hat{\mathbf{H}}^{(1)}), \ldots, \gamma_K(\hat{\mathbf{H}}^{(K)}))]$.

Finding the optimal team decision strategy is a very difficult problem as it requires solving a distributed *functional optimization* problem. A first step towards the optimal solution is often obtained by writing the *person-by-person* optimal decision [98, 100]. It consists in writing explicitly the necessary condition that any optimal decision represents the optimal strategy given the decisions of the other players, i.e., one player cannot improve the expected payoff by changing only its strategy. Denoting by $(\gamma_1^\star, \ldots, \gamma_K^\star)$ one optimal decision, the necessary condition mentioned above is written mathematically as

$$\begin{aligned}\gamma_j^\star(\hat{\mathbf{H}}^{(j)}) = \underset{\gamma_j(\hat{\mathbf{H}}^{(j)})}{\operatorname{argmax}}\, \mathrm{E}[\omega(\gamma_1^\star(\hat{\mathbf{H}}^{(1)}), \ldots, \gamma_{j-1}^\star(\hat{\mathbf{H}}^{(j-1)}) \\ , \gamma_j(\hat{\mathbf{H}}^{(j)}), \gamma_{j+1}^\star(\hat{\mathbf{H}}^{(j+1)}), \ldots, \gamma_K^\star(\hat{\mathbf{H}}^{(K)}))], \qquad \forall j.\end{aligned} \tag{3.23}$$

3.5 Summary of Objectives

The considered sum rate optimization based on distributed CSIT has been rewritten as a Team Decision problem. Yet, Team Decision problems are generally very intricate and in most of the case, solving them remains an

open problem for mathematicians. Only some particular cases with simple objective functions (e.g. quadratic objectives) or with restricted decision spaces (e.g. few possible discrete choices) could be solved [99]. Hence, the results presented in this thesis do not rely on any result from the Team Decision literature but we exploit instead the particular properties of each considered optimization problem. However, keeping the Team Decision formulation in mind will prove helpful to obtain interesting insights and understand the fundamentals of the problem.

In the first part of this thesis, we aim at solving optimization problem (3.23) in the sense that we look for the optimal transmission strategies at the TXs. We provide new schemes outperforming the conventional ones from the literature and we also evaluate the performance of commonly (sub-optimal) used strategies for *given information structures*.

In the second part, we consider the other side of the TX cooperation with distributed CSIT, which is the optimization of the information structure. Specifically, we fix heuristically the TX strategies to commonly used transmission schemes from the literature and we study the optimal allocation of the available feedback and backhaul resources to the TXs. Practically, this means looking for the optimal design of the $\hat{\mathbf{H}}^{(j)}$ over a space of possible choices. We then show how solving this optimization problem leads to very efficient solutions which outperform strikingly transmissions relying on conventional CSI sharing strategies (e.g. sharing of the same CSI to all the cooperating TXs).

Part II

Precoding with Distributed CSIT

Chapter 4

DoF of ZF with Distributed CSIT

We study here the precoder design when the TXs are faced with distributed CSIT. We consider in this chapter that all the nodes are equipped with a single-antenna. We also focus on homogeneous pathloss configuration in which all the wireless links have the same variance

$$\rho_{ij}^2 = 1, \qquad \forall i, j. \tag{4.1}$$

Note that this assumption does not impact the DoF analysis as discussed in Section 3.2.

We aim at answering the following fundamental questions regarding the high SNR transmission with distributed CSIT:

"How does ZF perform in the distributed CSIT configuration?"

"Are the precoding schemes designed to be robust with respect to imperfect centralized CSIT efficient to combat CSIT discrepancies between the TXs?"

"How can we make the transmission more robust to CSI discrepancies between the TXs?"

4.1 Distributed CSIT Model

Our main focus in this work is on frequency division duplexing (FDD) systems where the CSI is fed back from the RXs to the TX. The feedback

channels being imperfect, and the resources limited, the CSI is quantized at the RX before being feedback. Hence, we start by studying the quantization scheme conventionally used in the literature and how they can be adapted to the distributed CSIT configuration. We will then use this analysis to obtain a simple modeling of the CSI imperfections.

4.1.1 Quantization for Distributed CSIT

The conventional quantization scheme in the literature is the Grassmannian quantization [16, 80]. It consists in selecting the quantized unit-norm channel $\hat{\boldsymbol{h}}$ from the codebook $\mathcal{W}$ as

$$\hat{\boldsymbol{h}} = \underset{\boldsymbol{w}\in\mathcal{W}}{\operatorname{argmax}} |\boldsymbol{h}^{\mathrm{H}}\boldsymbol{w}|. \tag{4.2}$$

Interestingly, this quantization scheme can be seen to be inadequate in the case of distributed CSIT: The objective which is maximized, is invariant by multiplication of the codeword by a unit-norm complex number. This means that the estimate received at TX j can be written as $\boldsymbol{h}^{(j)}\exp(\imath\theta_j)$ where θ_j is uniformly distributed over $[0, 2\pi)$. This phase invariance creates an ambiguity between the estimates which is very harmful for the distributed joint precoding.

As a consequence, the Grassmannian quantization is efficient only if all the TXs receive the estimates from the same codebook. This prohibits the sharing of estimates of heterogeneous qualities to the different TXs, which can be required when backhaul or feedback links of heterogeneous qualities are available.

Hence, it is necessary to use another quantization in the distributed CSIT configuration. An alternative scheme consists in selecting the quantized channel $\hat{\boldsymbol{h}}$ from the codebook $\mathcal{W}$ so as to verify

$$\hat{\boldsymbol{h}} = \underset{\boldsymbol{w}\in\mathcal{W}}{\operatorname{argmin}} \|\boldsymbol{w} - \boldsymbol{h}\|. \tag{4.3}$$

However, using directly (4.3) leads to lower performance as the phase of the channel also impacts the performance, in contrast to the Grassmannian quantization. To solve this problem, we multiply the unit-norm channel by a complex unit-norm number in order to let the first coefficient be real valued. The quantization can then be written as

$$\hat{\boldsymbol{h}} = \underset{\boldsymbol{w}\in\mathcal{W}}{\operatorname{argmin}} \left\| \boldsymbol{w} - \frac{\{\boldsymbol{h}\}_1^*}{|\{\boldsymbol{h}\}_1|}\frac{\boldsymbol{h}}{\|\boldsymbol{h}\|} \right\|. \tag{4.4}$$

The codebook $\mathcal{W}$ contains $L = 2^B$ codewords being normalized in the same way.

Random quantization has been considered in several works [15, 80] because it provides a lower bound for the performance and leads to analytical results. Furthermore, it is shown in [15] for the conventional MIMO BC that in the case of two antennas at the TX, no codebook can achieve a better DoF than the DoF achieved with RVQ. RVQ is also shown to be optimal for the point-to-point MIMO link as the number of antennas tends to infinity both at the TX and the RX [101].

Therefore, we provide in the following an analysis of the performance of the quantization scheme (4.4) with random codebooks.

Proposition 1. Considering the quantization scheme (4.4) with random codebooks $\mathcal{W}$ of size $L = 2^B$, it holds for large codebook size ($L >> 1$) that

$$\mathrm{E}_{\mathcal{W},\boldsymbol{h}}\left[\min_{\boldsymbol{w}\in\mathcal{W}}\left\|\boldsymbol{w}-\frac{\{\boldsymbol{h}\}_1^*}{|\{\boldsymbol{h}\}_1|}\frac{\boldsymbol{h}}{\|\boldsymbol{h}\|}\right\|^2\right] = O(2^{-\frac{B}{K-1}}). \tag{4.5}$$

and

$$\mathrm{E}_{\mathcal{W},\boldsymbol{h}}\left[\log_2\left(\min_{\boldsymbol{w}\in\mathcal{W}}\left\|\boldsymbol{w}-\frac{\{\boldsymbol{h}\}_1^*}{|\{\boldsymbol{h}\}_1|}\frac{\boldsymbol{h}}{\|\boldsymbol{h}\|}\right\|^2\right)\right] = \frac{B}{K-1}+O(1). \tag{4.6}$$

Proof. This proposition is a consequence of the results given in Appendix .2. □

We have provided only the main results relative to this random vector quantization scheme as we will not focus further in this thesis on the quantization schemes. We will instead use a simple model for the CSI imperfections in order to discuss the fundamental aspects of the transmission. Additional results on the quantization scheme can be found in Appendix .2.

4.1.2 Distributed CSIT Model

Following the above discussion of the quantization with distributed CSIT, the variance of the estimation error of $\boldsymbol{h}_i$ at TX j, which we have denoted by $(\sigma_i^{(j)})^2$ in Section 3.3, is set as $(\sigma_i^{(j)})^2 = 2^{-\frac{B_i^{(j)}}{K-1}}$. For the sake of analytical tractability, we model the quantization error at TX j such that

$$\hat{\boldsymbol{h}}_i^{(j)} = \sqrt{1-(\sigma_i^{(j)})^2}\boldsymbol{h}_i + \sigma_i^{(j)}\boldsymbol{\delta}_i^{(j)}, \qquad \forall i,j \in \{1,\ldots,K\}. \tag{4.7}$$

where $\boldsymbol{\delta}_i^{(j)} \in \mathbb{C}^{K\times 1}$ has i.i.d. $\mathcal{CN}(0,1)$ elements and is independent of $\boldsymbol{h}_i$. We assume furthermore that the quantization errors at the TXs are independent such that we have $\mathrm{E}[\boldsymbol{\delta}_i^{(j)}(\boldsymbol{\delta}_i^{(k)})^{\mathrm{H}}] = \delta_{jk}\mathbf{I}_K$. This corresponds to the assumption that the feedback channels to the different TXs are independent.

The assumption for the quantization error to be Gaussian distributed makes the calculation easier. However, it can be easily seen that our DoF analysis depends only on the variance of the quantization error as long as the distribution verifies some mild regularity constraints.

Remark 4. *A practical scenario where the CSIT is distributed arises when the CSI is broadcast from the RXs to the non-colocated TXs in an analog manner (analog feedback) [19, 61]. In fact, the digital quantization model used in this work can be seen as only a model for the CSIT error due to limited feedback and the results can be directly extended to the case of analog feedback. In this case, the number of feedback bits can be related to the average transmit power (or bandwidth) on the feedback channel.* □

In the conventional MIMO BC, it is shown in [15] that the number of quantization bits should scale indefinitely with the logarithm of the SNR in order to achieve a strictly positive DoF when using ZF precoding. Thus, we also focus on the *scaling in the logarithm of the SNR* of the number of quantization bits of all the channel estimates. Consequently, we introduce the *CSI scaling matrix* $\mathbf{A} \in \mathbb{R}^{K\times K}$ with its (i,j)-th element defined as

$$A_i^{(j)} \triangleq \lim_{P\to\infty} \frac{B_i^{(j)}}{B^\star}, \qquad \forall i,j, \tag{4.8}$$

where we have defined

$$B^\star \triangleq (K-1)\log_2(P). \tag{4.9}$$

The number of CSI feedback bits $B^\star$ corresponds to a CSIT error decreasing in $1/P$ and it will become clear in this chapter that this corresponds to the CSIT being perfect in terms of DoF. Hence, the CSIT scaling coefficient $A_i^{(j)}$ represents the percentage of the number of bits $B^\star$ used to quantize the channel of RX i (denoted by $\boldsymbol{h}_i^{\mathrm{H}} \in \mathbb{C}^{1\times K}$) at TX j. Note that $B_i^{(j)}$ is a design parameter, the limit in (4.8) always exists. In addition, we assume that the CSI scaling matrix $\mathbf{A}$ is known to all the TXs.

Remark 5. *We will always consider that $A_i^{(j)} \in (0,1]$. Indeed, using $A_i^{(j)} > 1$ means using more CSI feedback bits than $B^\star$, which does not bring any improvement in terms of DoF. It follows that in all the subsequent results,*

the scaling coefficients $A_i^{(j)}$ *should be replaced by* $\min(A_i^{(j)}, 1)$ *so as to be valid for arbitrary values of the CSIT scaling coefficients. In addition, we consider that* $A_i^{(j)} > 0$ *as we consider the high-precision quantization regime.* □

4.2 Review of the Results in the MIMO BC with Centralized CSIT

We recall in this section the main results from [15] on the DoF achieved with finite rate feedback for the conventional centralized CSIT configuration. This will be helpful to understand the differences between the centralized CSIT configuration and the distributed setting. Hence, we consider a conventional MIMO BC where M TXs are colocated and share the *same* channel estimate. For this setting, we need to use notations which are different from the ones previously introduced for the MIMO channel with distributed CSIT. We denote by $\hat{\boldsymbol{h}}_i$ the channel estimate of $\boldsymbol{h}_i$ obtained with B_i bits. Following [15], the channel estimate is obtained from

$$\hat{\boldsymbol{h}}_i = \underset{\boldsymbol{w} \in \mathcal{W}_i^{\mathrm{BC}}}{\operatorname{argmax}} |\boldsymbol{w}^{\mathrm{H}} \boldsymbol{h}_i|^2 \tag{4.10}$$

where $\mathcal{W}_i^{\mathrm{BC}}$ is a random codebook containing 2^{B_i} unit-norm vectors isotropically distributed in $\mathbb{C}^{K \times 1}$. We provide now the main result.

Theorem 1. [15] In the MISO BC with M antennas and centralized CSIT, if the channel estimate $\hat{\boldsymbol{h}}_i$ is obtained from the quantization scheme (4.10) with $B_i = A_i(M-1)\log_2(P)$ (and $A_i \in (0,1]$) the DoF achieved with ZF is given by

$$\mathrm{DoF}^{\mathrm{CCSI}} = \sum_{i=1}^{M} A_i. \tag{4.11}$$

This result was given in [15] for $A_i = A$, i.e., with the same quality for all the channel vectors, but the extension to different values of the CSIT scaling coefficients A_i follows directly from the proof in [15]. The extension to Theorem 1 has been also suggested in [102] where the same formula for the DoF is derived in the case where DPC is used instead of ZF.

We will now derive the equivalent result of Theorem 1 for the BC with distributed CSIT.

4.3 ZF in the Two-user BC with Distributed CSIT

As a starting point we consider the particular antenna configuration with only two users and only two antennas at the TX side. This setting is interesting for two main reasons. Firstly, the exposition is simpler in that case while most of the insights are the same as in the general case, and secondly it will become clear in the following that this scenario makes it possible to obtain stronger results.

In the conventional multiple-antenna BC with imperfect CSI, the DoF with ZF has been derived and shown to be defined by the CSI scaling. With distributed CSIT, the CSI scaling of each channel vector $\boldsymbol{h}_i$ is different at each TX. One central goal of our work consist in determining how the DoF expression with centralized CSIT generalized to the distributed CSIT configuration. This would then lead us to evaluate whether ZF is in that case a performing solution and if not, whether one can find better solutions.

4.3.1 Conventional Zero Forcing

With distributed CSIT, the conventional ZF precoder is made of the beamformer $\boldsymbol{t}_i^{\mathrm{cZF}} \triangleq [\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{cZF}(1)}, \boldsymbol{e}_2^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{cZF}(2)}]^{\mathrm{T}}$ to transmit s_i, with its elements defined in an intuitive way as

$$\boldsymbol{t}_i^{\mathrm{cZF}(j)} \triangleq \sqrt{\frac{P}{2}} \frac{\Pi^{\perp}_{\hat{\boldsymbol{h}}_i^{(j)}}\left(\hat{\boldsymbol{h}}_{\bar{i}}^{(j)}\right)}{\|\Pi^{\perp}_{\hat{\boldsymbol{h}}_{\bar{i}}^{(j)}}\left(\hat{\boldsymbol{h}}_i^{(j)}\right)\|}, \qquad j \in \{1,2\} \tag{4.12}$$

where we have denoted by $\bar{i}$ the complementary index of i, i.e., $\bar{i} \triangleq 1 + i \mod 2$. The interpretation behind conventional ZF is that each TX applies ZF using its own CSI, thus implicitly assuming that the other TX shares the same CSI estimate. Our first result given in the following theorem relates the DoF achieved with such a precoding strategy.

Theorem 2. Conventional ZF achieves the DoF

$$\mathrm{DoF}^{\mathrm{cZF}} = 2 \min_{i,j\in\{1,2\}} A_i^{(j)}. \tag{4.13}$$

Proof. A detailed proof is provided in Appendix .3. □

With distributed CSIT, the DoF is limited by the *worst* quality of the CSI across the channels to the RXs and across the TXs. Comparing this result with the DoF expression with centralized CSIT in Theorem 1, it is

striking that the DoF at *both* users is limited by the worst estimation error whether it is done relative to $\boldsymbol{h}_1$ or $\boldsymbol{h}_2$. This is contrast to the formula with centralized CSIT in Theorem 1, where the accuracy of the estimation of $\boldsymbol{h}_i$ impacts only the DoF of RX i.

Remark 6. *Note that when all the CSIT scaling coefficients are equal, the setting considered is still different from the conventional multiple-antenna BC. Indeed, the estimates at the different TXs have statistically the same accuracy since the CSIT scaling coefficients are equal, but the realizations of the quantization errors are still different.* □

It follows from Theorem 2 that the additional interference due to the CSI inconsistencies between the TXs do not lead to any loss in DoF compared to the conventional multiple-antenna BC if and only if the channel estimates are of the same quality.

4.3.2 Robust Zero Forcing

Robust precoding schemes have been derived in the literature either as statistical robust ZF precoders [20] or precoders optimizing the worst case performance [22] to reduce the harmful effect of the imperfect CSI. Since we consider the average sum rate, the most relevant approach is the statistical one.

Using this model, we can extend the approach from [20] and the beamformer transmitting symbol s_i at TX j is obtained from solving the following minimization:

$$\underset{\boldsymbol{t}_i}{\operatorname{argmin}} \, \mathrm{E}_{\boldsymbol{\Delta}^{(j)}}[\|\boldsymbol{e}_i - (\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\boldsymbol{t}_i\|^2], \quad \text{subject to } \|\boldsymbol{t}_i\|^2 = \frac{P}{K}. \tag{4.14}$$

Writing the Lagrangian of the minimization problem with the Lagrange variable λ for the power constraint and taking the derivative according to $\boldsymbol{t}_i^*$ yields the equation

$$\left(\mathbf{R}_{\boldsymbol{\Delta}}^{(j)} + \hat{\mathbf{H}}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}} + \lambda\mathbf{I}\right)\boldsymbol{t}_i - \hat{\mathbf{H}}^{(j)}\boldsymbol{e}_i = \mathbf{0} \tag{4.15}$$

where the covariance matrix of the estimation error at TX j is

$$\mathbf{R}_{\boldsymbol{\Delta}}^{(j)} \triangleq \mathrm{E}[\boldsymbol{\Delta}^{(j)}(\boldsymbol{\Delta}^{(j)})^{\mathrm{H}}] \tag{4.16}$$

$$= \mathrm{diag}([P^{-A_1^{(j)}}, P^{-A_2^{(j)}}]). \tag{4.17}$$

The factor λ improves the performance at intermediate SNR by striking a compromise between the orthogonality constraint and the power consumption but it cannot improve the DoF. Thus, we can let λ be equal to zero and normalize the beamformer to fulfill the power constraint. The ith robust ZF beamformer is denoted by $\boldsymbol{t}_i^{\mathrm{rZF}} \triangleq [\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{rZF}(1)}, \boldsymbol{e}_2^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{rZF}(2)}]^{\mathrm{T}}$ and $\forall j \in \{1,2\}$

$$\mathbf{t}_i^{\mathrm{rZF}(j)} \triangleq \sqrt{\frac{P}{K}} \frac{(\mathbf{R}_{\boldsymbol{\Delta}}^{(j)} + \hat{\mathbf{H}}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}})^{-1}\hat{\mathbf{H}}^{(j)}\boldsymbol{e}_i}{\left\|(\mathbf{R}_{\boldsymbol{\Delta}}^{(j)} + \hat{\mathbf{H}}^{(j)}(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}})^{-1}\hat{\mathbf{H}}^{(j)}\boldsymbol{e}_i\right\|}. \tag{4.18}$$

We then derive the DoF achieved by this robust precoder.

Proposition 2. The robust ZF precoder defined in (4.18) achieves the same DoF as conventional ZF.

Proof. Only a sketch of the proof is provided as the proof follows easily from the proof of Theorem 2. The lower bound can be obtained using the same approach as for conventional ZF. Indeed, a Taylor expansion can be applied over a space of probability close to one. From the distributed precoding assumption, every precoded coefficient has also its phase distributed uniformly at random and independently of the other coefficients, which concludes the proof. The extension of the proof of the outerbound is easily done with basic algebra following the same approach as for conventional ZF. □

Hence, even the existing designs of robust ZF precoders do not improve the DoF in the BC with distributed CSIT. Note that the extension of the definition of the statistical robust precoder as well as the extension of proposition 2 to the general setting with K users is trivial and will not be given explicitly.

4.3.3 Beacon Zero Forcing

Robust ZF schemes from the literature do not bring any DoFs improvement which leads to investigate other alternative schemes more adapted to the distributed CSIT assumption. As a result, we now propose a modification of the conventional ZF scheme which improves the DoF when the estimates for $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$ are of different qualities. We call it *Beacon ZF* (bZF) because it makes use of an arbitrary channel-independent vector known beforehand at both TXs (a *beacon* signal).

The ith beamformer is then $\boldsymbol{t}_i^{\mathrm{bZF}} \triangleq [\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{bZF}(1)}, \boldsymbol{e}_2^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{bZF}(2)}]^{\mathrm{T}}$, with its elements defined from

$$\boldsymbol{t}_i^{\mathrm{bZF}(j)} \triangleq \sqrt{\frac{P}{2}} \frac{\Pi^{\perp}_{\hat{\boldsymbol{h}}_i^{(j)}}(\boldsymbol{c}_i)}{\|\Pi^{\perp}_{\hat{\boldsymbol{h}}_i^{(j)}}(\boldsymbol{c}_i)\|} \tag{4.19}$$

where $\boldsymbol{c}_i$ is any non-zero vector chosen beforehand and known at the TXs. Due to the isotropy of the channel, the choice of $\boldsymbol{c}_i$ does not influence the performance of the precoder.

Corollary 1. The DoF achieved with beacon ZF is

$$\mathrm{DoF}^{\mathrm{bZF}} = \min_{j\in\{1,2\}} A_1^{(j)} + \min_{j\in\{1,2\}} A_2^{(j)}. \tag{4.20}$$

Proof. The DoF follows easily from Theorem 2. Indeed, when using beacon ZF, no error is induced by the projection of the direct channel which is replaced by a fixed given vector. In terms of DoF, there is no difference between projecting the direct channel or any given vector. Thus, it is possible to apply the formula for the DoF in Theorem 2 considering that the direct channel is perfectly known, which yields the result. □

The key idea behind beacon ZF is to reduce the impact of the differences in CSI quality by using only the CSI necessary to fulfill the orthogonality constraint. Thus, the direct channel, which does not change the DoF but only improves the finite SNR performance, is not used. It follows then that $\boldsymbol{t}_1^{\mathrm{bZF}}$ does no depend on the estimates of $\boldsymbol{h}_1$, and symmetrically $\boldsymbol{t}_2^{\mathrm{bZF}}$ does not depend on the estimates of $\boldsymbol{h}_2$.

4.3.4 Active-Passive Zero Forcing

Beacon ZF improves the DoF but it is still the worst CSI scaling across the TXs (although no longer across the RXs) which defines the DoF. To improve further the DoF, we propose a scheme called *Active-Passive Zero Forcing (AP ZF)*. Assuming w.l.o.g. that $A_i^{(2)} \geq A_i^{(1)}$, AP ZF consists in the precoder whose beamformer $\boldsymbol{t}_i^{\mathrm{APZF}}$ transmitting symbol s_i is given by

$$\boldsymbol{t}_i^{\mathrm{APZF}} \triangleq \sqrt{\frac{P}{2\log_2(P)}} \begin{bmatrix} 1 \\ -\frac{\{\hat{\boldsymbol{h}}_i^{(2)}\}_1}{\{\hat{\boldsymbol{h}}_i^{(2)}\}_2} \end{bmatrix} \tag{4.21}$$

$$= \sqrt{\frac{P(1+\nu_i^{(2)})}{2\log_2(P)}} \boldsymbol{u}_i^{\mathrm{APZF}} \tag{4.22}$$

where

$$\boldsymbol{u}_i^{\mathrm{APZF}} \triangleq \frac{\begin{bmatrix} 1 & \frac{-\{\hat{\boldsymbol{h}}_i^{(2)}\}_1}{\{\hat{\boldsymbol{h}}_i^{(2)}\}_2} \end{bmatrix}^{\mathrm{T}}}{\left\| \begin{bmatrix} 1 & \frac{-\{\hat{\boldsymbol{h}}_i^{(2)}\}_1}{\{\hat{\boldsymbol{h}}_i^{(2)}\}_2} \end{bmatrix}^{\mathrm{T}} \right\|} \tag{4.23}$$

and

$$\nu_i^{(2)} \triangleq \frac{|\{\hat{\boldsymbol{h}}_i^{(2)}\}_1|^2}{|\{\hat{\boldsymbol{h}}_i^{(2)}\}_2|^2}. \tag{4.24}$$

AP ZF is based on the idea that each beamforming vector has to fulfill only one orthogonality constraint so that only one free variable is necessary. Thus, one coefficient can be set to a constant while still fulfilling the ZF constraints. Intuitively, the only way to achieve the DoF stemming from the best CSI estimate is if TX 2 (which has the best knowledge of $\boldsymbol{h}_1$) can adapt to the coefficient transmitted at TX 1 to adjust its beamforming vector and improves the accuracy with which the interference are suppressed. This is possible only if TX 2 knows the transmit coefficient at TX 1.

Using this precoding scheme, the DoF is then given in the following proposition.

Proposition 3. Active-Passive ZF achieves the DoF:

$$\mathrm{DoF}^{\mathrm{APZF}} \geq \max_{j\in[1,2]} A_1^{(j)} + \max_{j\in[1,2]} A_2^{(j)}. \tag{4.25}$$

Proof. By symmetry, we consider w.l.o.g. the DoF at RX 1, and we assume that the beamformers $\boldsymbol{t}_1$ and $\boldsymbol{t}_2$ are given by (4.22). We still assume w.l.o.g. that $A_1^{(2)} \geq A_1^{(1)}$, i.e., TX 2 has the best CSI over $\boldsymbol{h}_1$. It follows easily from the definition that the DoF at RX 1 can be rewritten as

$$\mathrm{DoF}_1 = 1 - \lim_{P\to\infty} \frac{\mathrm{E}\left[\log_2(|\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{t}_2|^2)\right]}{\log_2(P)} \tag{4.26}$$

We now focus on the interference term:

$$|\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{t}_2|^2 = \frac{P}{2\log_2(P)} \left| \boldsymbol{h}_1^{\mathrm{H}} \begin{bmatrix} 1 \\ -\frac{\{\boldsymbol{h}_1^{(2)}\}_1}{\{\boldsymbol{h}_1^{(2)}\}_2} \end{bmatrix} \right|^2. \tag{4.27}$$

By construction, $\boldsymbol{t}_2$ is orthogonal to $\boldsymbol{h}_1^{(2)}$, so that

$$|\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{t}_2|^2 = \frac{P(1+\nu_2^{(2)})}{2\log_2(P)}\|\boldsymbol{h}_1\|^2 \left|\Pi^{\perp}_{\boldsymbol{h}_1^{(2)}}(\boldsymbol{h}_1)^{\mathrm{H}}\boldsymbol{u}_2\right|^2 \tag{4.28}$$

$$= \frac{P(1+\nu_2^{(2)})}{2\log_2(P)}\|\boldsymbol{h}_1\|^2\|\boldsymbol{\delta}_1^{(2)}\|^2 \tag{4.29}$$

Inserting (4.29) in the DoFs expression (4.26) and using Proposition 16 from Appendix .2 to evaluate the expectation with the quantization error, we obtain

$$\mathrm{DoF}_1 \geq \lim_{P\to\infty} \frac{B_1^{(2)}}{\log_2(P)} \tag{4.30}$$

$$= A_1^{(2)} \tag{4.31}$$

which is the best scaling across the TXs. □

Comparing the DoF achieved with AP ZF with the DoF achieved when both TXs share the estimate of a channel vector with the highest accuracy gives the following result.

Theorem 3. Active-Passive ZF achieves the same DoF in the 2-user BC with distributed CSIT as in the MIMO BC with centralized CSIT where both TXs share the estimates with the highest CSI accuracy.

Improved scheme at finite SNR AP ZF allows to recover the DoF which would have been achieved with the best CSI across the TXs. However, the choice of the coefficient used to transmit at TX 1 (with the lowest accuracy of the CSI) remains to be discussed. In fact, the beamformer can be multiplied arbitrarily by any unit-norm complex number without impacting the rate achieved so that only the power used at TX 1 needs to be decided. According to (4.22), the power used at TX 1 is set to $P/(2\log_2(P))$.

The normalization by $\log_2(P)$ is a consequence of the fading coefficient $\{\boldsymbol{h}_1\}_2$ which can have a very small amplitude. In this case it would be necessary for TX 2 to transmit with a very large power to fulfill the orthogonality constraint. To ensure that the interference are canceled for all channel realizations while respecting the power constraint, it is necessary to have the ratio between the power used at TX 1 and the sum power constraint tending to zero. The factor $\log_2(P)$ is used because it fulfills this property while not reducing the DoF due to the partial power consumption.

However, this comes at the cost of using only a small share of the available power, which is clearly inefficient and leads to a rate offset tending to minus infinity. To avoid this behavior, we propose that the TX with the worst CSI accuracy adapts its power consumption with respect to the channel realizations. In the following, we propose two possible solutions to improve the performance at finite SNR:

- Firstly, TX 1 can use its local CSI to normalize the beamformer which is then given by

$$\boldsymbol{t}_i^{\mathrm{APZF}} = \sqrt{\frac{P}{2}} \begin{bmatrix} \frac{1}{\sqrt{1+\nu_i^{(1)}}} \\ -\frac{\{\boldsymbol{h}_{\bar{i}}^{(2)}\}_1}{\sqrt{1+\nu_i^{(2)}}\{\boldsymbol{h}_{\bar{i}}^{(2)}\}_2} \end{bmatrix} \tag{4.32}$$

  with $\nu_i^{(j)} \triangleq |\{\boldsymbol{h}_{\bar{i}}^{(j)}\}_1|^2/|\{\boldsymbol{h}_{\bar{i}}^{(j)}\}_2|^2$, for $j = 1, 2$. This beamformer is not DoF maximizing because the local CSI is used at TX 1 so that TX 2 does not any longer have an exact knowledge of the coefficient used to transmit at TX 1. Consequently, beamformer $\boldsymbol{t}_i^{\mathrm{APZF}}$ is not any longer orthogonal to $\boldsymbol{h}_{\bar{i}}^{(2)}$. Yet, this solution achieves good performance at intermediate SNR.

- Another possibility is to assume that TX 1 receives the scalar $\nu_i^{(2)}$ (or ν_i) and use it to control its power. This means that TX 2 needs to share this scalar. This requires an additional feedback, but only a few bits are necessary to improve the performance at practical SNR.

4.4 ZF in the General K-Users BC with Distributed CSIT

In this section, we will show how the main results can be generalized to arbitrary number of users. The same approach as in the case $K = 2$ can be followed and we start by briefly generalizing to arbitrary number of users the previously described precoding schemes.

4.4.1 Conventional Zero Forcing

The conventional ZF precoder will be denoted as $\mathbf{T}^{\mathrm{cZF}} \triangleq [\boldsymbol{t}_1^{\mathrm{cZF}}, \ldots, \boldsymbol{t}_K^{\mathrm{cZF}}]$ with $\boldsymbol{t}_i^{\mathrm{cZF}} \triangleq [\boldsymbol{e}_1^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{cZF}(1)}, \boldsymbol{e}_2^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{cZF}(2)}, \ldots, \boldsymbol{e}_K^{\mathrm{T}}\boldsymbol{t}_i^{\mathrm{cZF}(K)}]^{\mathrm{T}}$ transmitting symbol s_i,

and the beamformer $\boldsymbol{t}_i^{\mathrm{cZF}(j)}$ computed at TX j to transmit symbol i given by

$$\boldsymbol{t}_i^{\mathrm{cZF}(j)} \triangleq \sqrt{\frac{P}{K}} \frac{\Pi^{\perp}_{\hat{\mathbf{H}}_i^{(j)}}(\hat{\boldsymbol{h}}_i^{(j)})}{\|\Pi^{\perp}_{\hat{\mathbf{H}}_i^{(j)}}(\hat{\boldsymbol{h}}_i^{(j)})\|} \tag{4.33}$$

with $\hat{\mathbf{H}}_i^{(j)} \triangleq [\hat{\boldsymbol{h}}_1^{(j)}, \ldots, \hat{\boldsymbol{h}}_{i-1}^{(j)}, \hat{\boldsymbol{h}}_{i+1}^{(j)}, \ldots, \hat{\boldsymbol{h}}_K^{(j)}]$.

We can then generalize the results from Theorem 2 to an arbitrary number of users.

Theorem 4. In the BC with distributed CSIT, the DoF achieved with conventional ZF is equal to

$$\mathrm{DoF}^{\mathrm{cZF}} = K \min_{i,j \in \{1,\ldots,K\}} A_i^{(j)}. \tag{4.34}$$

Proof. A detailed proof is provided in Appendix .3. □

In Theorem 4, we have shown that the results concerning conventional ZF can be exactly generalized and the DoF scales with the worst CSI accuracy across the TXs and the RXs. Indeed, the bad estimation of the channel to *one* user at *one* TX reduces the DoF of *all* the users. This is a very pessimistic result and represents a different behavior as the transmission with centralized CSIT.

Example 3. *The impact of the CSIT discrepancies between the TXs can be observed clearly by the analysis of the particular CSIT configuration where* $A_i^{(j)} = A_i$ *for all* j*. In that case, the DoF achieved with distributed CSIT is given in* (4.34) *by*

$$\mathrm{DoF}^{\mathrm{cZF}} = K \min_i A_i. \tag{4.35}$$

In contrast, considering the same CSIT scaling coefficients (which is possible since the quality is the same across the TXs) with centralized CSIT, the DoF expression is given in (4.11) *and the DoF is equal to*

$$\mathrm{DoF}^{\mathrm{CCSI}} = \sum_{i=1}^{K} A_i. \tag{4.36}$$

The comparison of the two DoF expressions shows a very interesting consequence of the CSIT discrepancies: The DoF achieved at one RX depends on the quality of all *the channel estimates.*

4.4.2 Beacon Zero Forcing

The beacon ZF precoder is denoted as $\mathbf{T}^{\text{bZF}} \triangleq [\boldsymbol{t}_1^{\text{bZF}}, \boldsymbol{t}_2^{\text{bZF}} \dots, \boldsymbol{t}_K^{\text{bZF}}]$ with the beamformer $\boldsymbol{t}_i^{\text{bZF}} \triangleq [\boldsymbol{e}_1^{\text{T}} \boldsymbol{t}_i^{\text{bZF}(1)}, \boldsymbol{e}_2^{\text{T}} \boldsymbol{t}_i^{\text{bZF}(2)}, \dots, \boldsymbol{e}_K^{\text{T}} \boldsymbol{t}_i^{\text{bZF}(K)}]^{\text{T}}$ transmitting symbol s_i. The beamformer $\boldsymbol{t}_i^{\text{bZF}(j)}$ computed at TX j to transmit symbol s_i is given by

$$\boldsymbol{t}_i^{\text{bZF}(j)} \triangleq \sqrt{\frac{P}{K}} \frac{\Pi^{\perp}_{\hat{\mathbf{H}}_i^{(j)}}(\boldsymbol{c}_i)}{\|\Pi^{\perp}_{\hat{\mathbf{H}}_i^{(j)}}(\boldsymbol{c}_i)\|} \tag{4.37}$$

where $\boldsymbol{c}_i$ is any non-zero vector chosen beforehand and known at all TXs.

Proposition 4. The DoF achieved with beacon ZF is equal to

$$\text{DoF}^{\text{bZF}} = \sum_{k=1}^{K} \min_{\substack{i \in \{1,\dots,K\}, \\ i \neq k}} \min_{\substack{\ell,j \in \{1,\dots,K\}, \\ \ell \neq i}} A_\ell^{(j)}. \tag{4.38}$$

Proof. To derive the DoF at a RX k, we need to compute the scaling of the interference at RX k stemming from the transmission to the $K-1$ other RXs. In the proof of Theorem 4, it is in fact the scaling of the interference resulting from the transmission of one stream which is calculated. To obtain the DoF at one RX, the scaling of the interference resulting from the transmission of each of the $K-1$ interfering streams needs to be computed. This is represented by the first summation over i. Determining the interference leaked by the transmission of symbol s_i using beacon ZF leads to the second minimum in the formula. □

We have derived the DoF for beacon ZF, but we will show in the following corollary that beacon ZF is only attractive in terms of DoF in the two-user case.

Corollary 2. For $K \geq 3$, beacon ZF achieves the same DoF as conventional ZF.

Proof. The result is easily obtained by studying the effect of the two successive minimums in (4.38). □

4.4.3 Active-Passive Zero Forcing

The generalization of AP ZF is intuitive and consists simply, for the computation of each beamforming vector, in letting one TX arbitrarily fix its

precoding coefficient while the other TXs adapt to this coefficient. Nevertheless, it requires the introduction of a few more notations.

We define the ordered set $\mathbb{S} \triangleq \{n_1, \ldots, n_K\}$ as the set whose i-th element corresponds to the indice of the TX with fixed coefficient when transmitting the symbol s_i (passive TX for s_i). We then introduce the (column) channel vector from TX ℓ to all the RXs except the i-th RX:

$$\boldsymbol{g}_i^{(j)}(\ell) \triangleq [\{(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\}_{1,\ell}, \ldots, \{(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\}_{i-1,\ell}, \{(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\}_{i+1,\ell}, \ldots, \{(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\}_{K,\ell}]^{\mathrm{T}}. \tag{4.39}$$

Using the previous definition, we can then define

$$(\hat{\mathbf{H}}_i^{(j)}(n_i))^{\mathrm{H}} \triangleq [\boldsymbol{g}_i^{(j)}(1), \ldots, \boldsymbol{g}_i^{(j)}(n_i-1), \boldsymbol{g}_i^{(j)}(n_i+1), \ldots, \boldsymbol{g}_i^{(j)}(K)] \tag{4.40}$$

which represents the estimate at TX j of the multi-user channel from all the TXs except TX n_i to all the RXs except RX i.

For a given set $\mathbb{S}$, we write $\mathbf{T}^{\mathrm{APZF}}(\mathbb{S}) \triangleq [\boldsymbol{t}_1^{\mathrm{APZF}}(n_1), \ldots, \boldsymbol{t}_K^{\mathrm{APZF}}(n_K)]$ where the beamformer $\boldsymbol{t}_i^{\mathrm{APZF}}(n_i) \triangleq [\boldsymbol{e}_1^{\mathrm{T}} \boldsymbol{t}_i^{\mathrm{APZF}(1)}(n_i), \ldots, \boldsymbol{e}_K^{\mathrm{T}} \boldsymbol{t}_i^{\mathrm{APZF}(K)}(n_i)]^{\mathrm{T}}$ transmits symbol s_i. The beamformer $\boldsymbol{t}_i^{\mathrm{APZF}(j)}(n_i)$ computed at TX j to transmit symbol s_i is given by

$$\boldsymbol{t}_i^{\mathrm{APZF}(j)}(n_i) \triangleq \sqrt{\frac{P}{K \log_2(P)}} \boldsymbol{u}_i^{\mathrm{APZF}(j)}(n_i) \tag{4.41}$$

where we have defined

$$\begin{aligned} \boldsymbol{u}_i^{\mathrm{APZF}(j)}(n_i) \triangleq [\, & \check{u}_{1i}^{\mathrm{APZF}(j)}(n_i), \ldots, \check{u}_{n_i-1,i}^{\mathrm{APZF}(j)}(n_i), 1 \\ & , \check{u}_{n_i,i}^{\mathrm{APZF}(j)}(n_i), \ldots, \check{u}_{K-1,i}^{\mathrm{APZF}(j)}(n_i)]^{\mathrm{T}} \end{aligned} \tag{4.42}$$

with $\check{\boldsymbol{u}}_i^{\mathrm{APZF}(j)}(n_i) \triangleq \left[\check{u}_{1i}^{\mathrm{APZF}(j)}(n_i), \ldots, \check{u}_{K-1,i}^{\mathrm{APZF}(j)}(n_i)\right]^{\mathrm{T}} \in \mathbb{C}^{K-1}$ and

$$\check{\boldsymbol{u}}_i^{\mathrm{APZF}(j)}(n_i) \triangleq \frac{-\left((\hat{\mathbf{H}}_i^{(j)}(n_i))^{\mathrm{H}}\right)^{-1} \boldsymbol{g}_i^{(j)}(n_i)}{\sqrt{1 + \|\left((\hat{\mathbf{H}}_i^{(j)}(n_i))^{\mathrm{H}}\right)^{-1} \boldsymbol{g}_i^{(j)}(n_i)\|^2}}. \tag{4.43}$$

Even though the notations are quite heavy, the intuition behind the construction of the precoder is exactly the same as for the two-user case. TX n_i is the *passive* TX and transmits with a fixed coefficient $\sqrt{P/K \log_2(P)}$ while the other *active* TXs then choose their coefficients in order to ZF the interference. This is obtained by setting their coefficients so as to fulfill (4.41). The notational complexity comes only from the fact that we need to introduce a "reduced" channel without the direct channel as well as without the channel from the *passive* TX.

Proposition 5. Active-Passive ZF with the set $\mathbb{S} = \{n_1, \ldots, n_K\}$ achieves the DoF

$$\mathrm{DoF}^{\mathrm{APZF}}(\mathbb{S}) = \sum_{k=1}^{K} \min_{\substack{i \in \{1,\ldots,K\}, \\ i \neq k}} \min_{\substack{\ell, j \in \{1,\ldots,K\}, \\ \ell \neq i, j \neq n_i}} A_\ell^{(j)}. \tag{4.44}$$

Proof. Due to the symmetry between the RXs, we will show the result only for the DoF at RX k. Let assume that AP ZF is used with the set $\mathbb{S}$. To obtain the DoF, we need to derive the scaling of the interference at RX i when all streams are transmitted using AP ZF. The first minimum of the DoFs formula follows from the summation over all the $K-1$ interfering streams. It remains then to determine the scaling of the interference resulting from the transmission of one given data symbol.

TX j computes the beamformer $\boldsymbol{t}_\ell^{\mathrm{APZF}(j)}(n_\ell)$ according to (4.41). This formula is similar to the one for conventional ZF so that the scaling of the remaining interference power can be derived with a proof very akin to that of Theorem 4 which is omitted to avoid repetitions. Thus, the interference received at RX k due to the transmission of symbol s_i corresponds to the second minimum of the DoF formula. This expression follows from the fact that the CSI at TX n_ℓ and the CSI on the direct channel $\boldsymbol{h}_\ell$ are not used to design the beamformer transmitting s_ℓ. □

The DoF given in Proposition 5 is given by two successive minimizations. This is similar to beacon ZF at the difference that the index of one TX is not taken into account in the second minimization. This leads then to a larger DoF. The formula for the DoF depends on the set $\mathbb{S}$ but we will show that the optimal set is easily derived when the number of users is larger than 4.

Corollary 3. For $K \geq 4$ users, it is optimal in terms of DoF to choose all the indices in $\mathbb{S}$ to be equal. Therefore, it is optimal to choose n_i as the indice of the minimum over all the CSIT scaling coefficients, and the DoF reads as

$$\mathrm{DoF}^{\mathrm{APZF}} = K \min_{\substack{i, j \in \{1,\ldots,K\}, \\ j \neq \mathrm{argmin}_k \min_\ell A_\ell^{(k)}}} A_i^{(j)}. \tag{4.45}$$

Proof. Similar to the proof of the corollary for Beacon ZF, the proof follows by studying the effect of the two successive minimums and for $K \geq 4$, it has for consequence that it is optimal to choose $\forall i, j, n_i = n_j$. □

Exactly as in the two-user case, AP ZF leads to an improvement in DoF but this comes at the cost of an unbounded negative rate offset. To improve on this feature, the percentage of the available power which is consumed by the TXs needs to be increased. The sames solutions as described for the two-user case in Subsection 4.3.4 can be applied, i.e., either a heuristic power control or the transmission of a scalar to control the power. Note that the scalar can be transmitted by any of the other $K-1$ TXs and that one scalar needs to be transmitted for each stream. We refer to Subsection 4.3.4 for more details.

4.4.4 Discussion of the Results

Altogether, we have shown in this section that the results for the two-user case given in Section 4.3 could be generalized to an arbitrary number of users. However, the results suggest in all cases a fundamental lack of robustness of the performance as we increase the number of users. Indeed, with conventional ZF, a single inaccurate channel estimate can reduce the DoF of all the users while the novel precoding schemes proposed can only cope with a few channel estimates being of insufficient quality. This shows the need for other methods to make the transmission more robust to imperfect distributed CSI when more than two-user are present.

4.5 Precoding Using Hierarchical Quantization

In view of the rather pessimistic results in the previous section, we propose now an alternative method to make the transmission more robust to the CSI discrepancies. It consists in modifying the CSI quantization and using a Hierarchical Quantization (HQ) scheme to encode the CSI [75, 103]. This HQ is in fact very well known in the Rate Distortion theory and called "multi-level" quantization, or "successive refinement" [104]. We recall briefly the main principle of this quantization scheme and we particularize it in the case of interest which is random HQ.

4.5.1 Hierarchical Quantization

Hierarchical quantization (or multi-resolution quantization) is a quantization scheme in which the information is encoded so that the original message can be decoded up to a number of bits depending on the quality of the feedback channel. The better the channel is, the more bits can be decoded. Thus, if one entity receives a codeword with a higher accuracy than another entity,

and has the knowledge of the feedback qualities, it also knows what has been decoded at the other entity. Conversely, if one entity can detect the feedback information at a given resolution level but knows that another entity can decode the same information at a higher resolution level, it can use its individual decoded codeword to form a limited set of guesses around it as to which higher resolution codeword may have been detected at the other TX.

In our setting, it means that each TX can decode the CSI feedback up to a certain number of bits depending on the quality of the feedback link. If TX j_1 receives a CSI of better quality than another TX j_2, it can decode more bits from the CSI and can get the knowledge of the CSI at TX j_2 with less decoded bits. Note that this implies that two TXs with the same CSI quality have the *same* codebook and thus exactly the same realization for the channel estimation error. This is in contrast to what has been considered in the previous sections.

We wish to continue using the properties of RVQ so that we need to design *hierarchical random codebooks*, i.e., codebooks fulfilling the properties of both kinds of codebooks. Since this is not the main focus of the work, we just briefly describe a possible method to construct such codebooks and the quantization scheme associated.

We start by considering a random codebook of size corresponding to the best accuracy, say $2^{\ell_{\max}}$. This random codebook is then divided into two random codebooks containing each half the elements. This process is then applied on the two smaller codebooks obtained until having $2^{\ell_{\max}}$ codebooks of one element. In each of the sub-codebooks of different sizes created, we pick randomly one elements to be the *representative* of this codebook.

Once the quantized vector maximizing the figure of merit has been chosen among the $2^{\ell_{\max}}$ vectors, the encoding can be easily done. The chosen vector belongs to one set of each size and the encoding bits are used to select among the two possible choices, the set to which the quantized vector belongs.

The decoding step works as follows. The first bit denotes one of the two codebooks of size $2^{\ell_{\max}-1}$, the second bit denotes one of the two codebooks of size $2^{\ell_{\max}-2}$ inside this codebook, and so on, until the last bit is decoded. Once this is done, the codeword decoded is chosen to be the *representative* codeword of the obtained codebook.

It is then easily verified that the proposed quantization scheme has the hierarchical properties desired.

4.5.2 Conventional Zero Forcing with Hierarchical Quantization

In the previous sections, we have shown that the quality of the estimation of one channel $\boldsymbol{h}_i$ to one given RX had an impact on the DoF achieved *at all RXs*. This is a surprising property which follows from the joint precoding where the consistency between the transmissions of the different TXs is critical. We will show how the hierarchical quantization described above can be used to avoid this very inefficient property.

In the following, we will consider a particularly simple use of hierarchical quantization consisting in letting all the TXs designing the beamforming vector use only the part of the CSI which is common to all the TXs, and simply "forget" about the more accurate CSI knowledge. We then obtain a CSI configuration where all the TXs share the same CSI and the DoF can be obtained from Theorem 1.

Theorem 5. The DoF achieved using Conventional ZF with hierarchical quantization is

$$\mathrm{DoF}^{\mathrm{cZF}} = \sum_{i=1}^{K} \min_{j\in\{1,\ldots,K\}} A_i^{(j)}. \tag{4.46}$$

Using HQ as described, i.e., using only the estimate of a channel vector $\boldsymbol{h}_i$ common to all the TXs, follows from the observation that the worst estimation error of $\boldsymbol{h}_i$ limits in any case the DoF at RX i. Thus, using only the common part of the estimate of $\boldsymbol{h}_i$ does not reduce the DoF at RX $\boldsymbol{h}_i$. Yet, it leads to an improved consistency between the beamformers computed at the TXs. This has for consequence that the error in the estimate of the channel $\boldsymbol{h}_i$ only impacts the DoF at RX i and not at the other RXs.

Remark 7. *Note that the proposed scheme using HQ is very simple and more gains could certainly be obtained with a more sophisticated use of the additional CSI knowledge available at some TXs.* □

4.5.3 Active-Passive Zero Forcing with Hierarchical Quantization

Hierarchical quantization is used for AP ZF in the same way as for Conventional ZF. This consists in using the CSI which is common to all the active TXs considered in the definition of the beamformer in (4.41).

Proposition 6. The DoF achieved using Active-Passive ZF with Hierarchical Quantization and the set $\mathbb{S}$ is

$$\mathrm{DoF}^{\mathrm{APZF}}(\mathbb{S}) = \sum_{k=1}^{K} \min_{\substack{i\in\{1,\dots,K\},\\ i\neq k}} \min_{\substack{j\in\{1,\dots,K\},\\ j\neq n_i}} A_k^{(j)}. \tag{4.47}$$

The two successive minimums come from the fact that it is not the same TX which is *passive* for the different streams. It is clear from (4.47) that it is optimal to choose all the n_i to be equal for $K \geq 3$. However, the indice of the optimal passive TX, which we denote by n_{HQ}, is now different from the case without HQ. It is easily obtained by looking for the passive TX bringing the largest improvement in DoF:

$$n_{\mathrm{HQ}} \triangleq \operatorname*{argmax}_{n\in\{1,\dots,K\}} \sum_{k=1}^{K} \min_{\substack{j\in\{1,\dots,K\},\\ j\neq n}} A_k^{(j)}. \tag{4.48}$$

The maximum DoF using AP-ZF with HQ follows then directly.

Proposition 7. For $K \geq 3$, it is optimal to choose the passive TX to be TX j with $j = n_{\mathrm{HQ}}$ defined in (4.48), for all the data streams. The DoF achieved with Active-Passive ZF based on Hierarchical Quantization is then equal to

$$\mathrm{DoF}^{\mathrm{APZF}} = \sum_{i=1}^{K} \min_{\substack{j\in\{1,\dots,K\},\\ j\neq n_{\mathrm{HC}}}} A_i^{(j)}. \tag{4.49}$$

4.6 Simulations

4.6.1 In the Two-User Case

We consider two models for the imperfect channel CSI, a statistical model and RVQ as described in Section 4.1.2.

In the statistical model, the quantization error is modeled by adding a Gaussian i.i.d. quantization noise to the channel with the covariance matrix at TX j equal to $\mathrm{diag}([P^{-A_1^{(j)}}, P^{-A_2^{(j)}}])$. This corresponds to the scaling in P of the variance provided in Proposition 1. The averaging is then done over 10000 realizations.

In the RVQ, we consider a given number of feedback bits and we average over 100 random codebooks and 1000 channel realizations. In the simulations, we consider the following precoders: ZF with perfect CSI, conventional ZF [cf. (4.12)], Beacon ZF [cf. (4.19)], and Active-Passive ZF [cf. (4.22)] with heuristic power control and with 3-bits power control.

In Fig. 4.1, we consider the statistical model with the CSI scaling matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0.5 \\ 0 & 0.7 \end{bmatrix}. \tag{4.50}$$

We remind the reader that the first column corresponds to the CSIT scaling coefficients at TX 1 and the second to the CSIT scaling coefficient of TX 2. To emphasize the DoF (i.e., the slope of the curve in the figure), we let the SNR grow large. As expected theoretically, conventional ZF scales with the worst accuracy and saturates at high SNR, while Beacon ZF has a positive slope and Active-Passive ZF performs closer to perfect ZF with a slope only slightly smaller than the optimal one.

In Fig. 4.2, we plot the sum rate achieved using RVQ with the CSI feedback

$$\mathbf{B} = \begin{bmatrix} 6 & 3 \\ 3 & 6 \end{bmatrix}. \tag{4.51}$$

The matrix $\mathbf{B}$ contains the number of feedback bits used, with its (i,j)th element being equal to $B_i^{(j)}$ the number of bits used to quantize channel $\boldsymbol{h}_i$ at TX j. From the theoretical analysis, the DoF should be equal to zero for all the precoding schemes since the number of feedback bits used does not increase with the SNR. This is confirmed by the saturation of the sum rate as the SNR increases. Yet, the saturation occurs at a higher SNR for Beacon ZF compared to conventional ZF, and at an even higher SNR for Active-Passive ZF. This translates into an improvement of the sum rate at intermediate SNR.

4.6.2 With Arbitrary Number of Users

For the simulations with $K \geq 3$ users, only the statistical model described in the previous paragraph for the two-user case is considered due to the complexity of RVQ. To model easily the use of hierarchical quantization, we simply consider that a TX has the knowledge of the channel estimate at another TX if this TX receives a feedback concerning this channel vector with a lower CSI scaling coefficient. Since we have derived that Beacon ZF [Cf. (4.37)] does not bring any improvement in DoF for $K \geq 3$, we

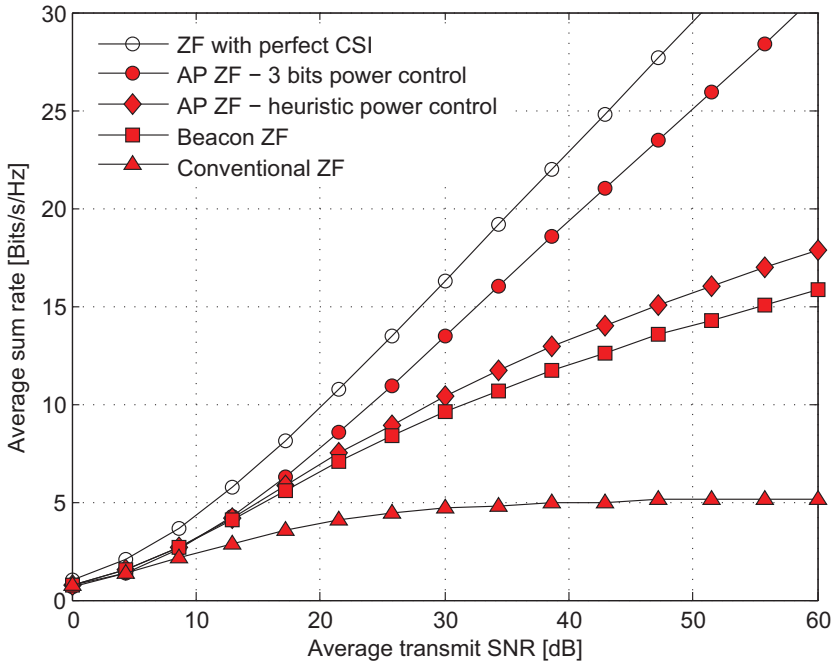


Figure 4.1: Sum rate in terms of the SNR with a statistical modeling of the error from RVQ using $[A_1^{(1)}, A_1^{(2)}] = [1, 0.5]$ and $[A_2^{(1)}, A_2^{(2)}] = [0, 0.7]$.

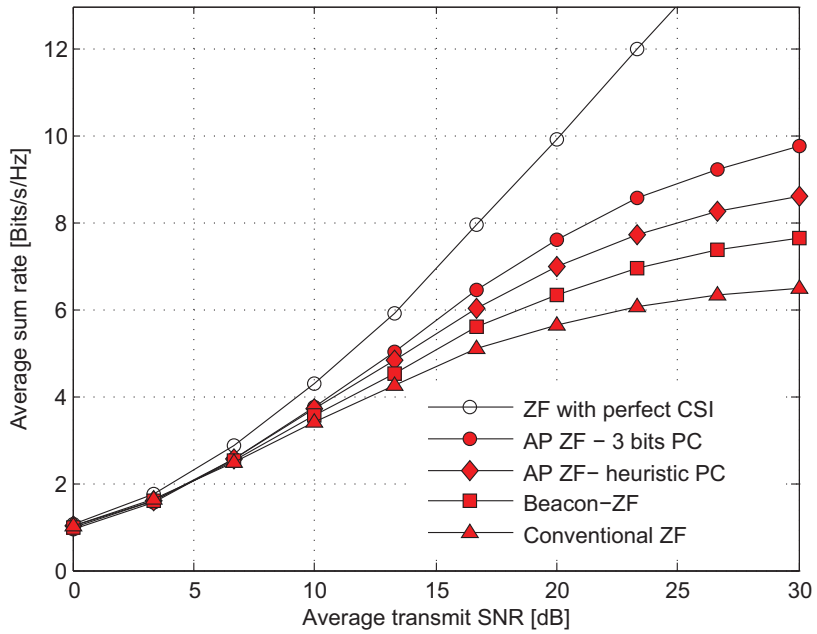


Figure 4.2: Sum rate in terms of the SNR with RVQ using $[B_1^{(1)}, B_1^{(2)}] = [6, 3]$ and $[B_2^{(1)}, B_2^{(2)}] = [3, 6]$.

Precoding Scheme	DoF
Conventional ZF	0
Active-Passive ZF	2.1
Conventional ZF with HQ	5.3
Active-Passive ZF with HQ	6.3

Table 4.1: DoF achieved by different precoding schemes for the CSIT scaling coefficients given in (4.52)

will consider in the figures only conventional ZF [Cf. (4.33)] and Active-Passive ZF [Cf. (4.41)] where the transmission of 3-bits to the *passive* TX is allowed for every beamforming vector. For both precoding schemes, we will furthermore consider both the case of hierarchical quantization with random codebooks and conventional RVQ.

.

We consider the performance achieved with an arbitrary chosen CSI scaling matrix to verify that the precoding schemes behave as expected. Thus, we consider $K = 7$ users and we set all the elements of the CSI scaling matrix $\mathbf{A}$ equal to 1 at the exception of two coefficients corresponding to different TXs and RXs set to 0 and 0.3, respectively. Hence, the CSI scaling matrix is equal to

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0.3 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}. \tag{4.52}$$

. The DoFs achieved with the different precoding schemes for this CSIT scaling matrix are given in Table 4.1.

The average sum rate achieved for this setting is shown in terms of the SNR in Fig. 4.3. We can observe that the schemes using HQ achieve a much larger DoF (i.e., slope in terms of the SNR) which is in agreement with the theoretical expressions in Table 4.1. Furthermore, the increase in DoF translates to better performance at intermediate SNRs.

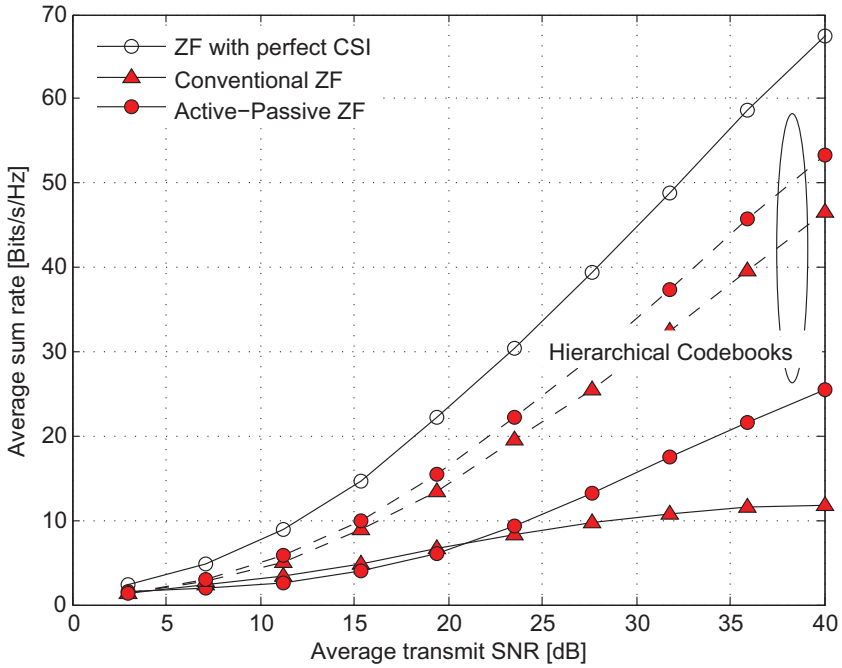


Figure 4.3: Sum rate achieved for the arbitrarily chosen CSI scaling configuration $\mathbf{A}$ given in (4.52).

4.7 Conclusion

We have shown that conventional ZF precoding applied without taking into account the CSI discrepancies achieves far from the maximal DoF and is limited by the worst accuracy of the CSI over the whole multi-user channel. This is particularly striking as the bad estimate of the channel to one given user at a unique TX reduces the DoF of all the users. This represents a different behavior from the transmission with centralized CSIT. In the particular case with only two users, we have provided a precoding scheme achieving the DoF corresponding to the most accurate CSI across the TXs, which is a strong improvement compared to conventional ZF. With an arbitrary number of users, the DoF achieved by conventional ZF has been derived and precoding schemes to improve over this DoF value have been provided. Interestingly, it has been shown how using codebooks with a hierarchical structure to quantize the CSI leads to a significant DoF improvement thanks to the higher degree of consistency obtained between the transmit coefficients.

Chapter 5

Rate Loss of ZF With Distributed CSIT

In the previous chapter, we have studied the impact of the CSIT distributedness over the DoF achieved. The DoF is a very interesting figure of merit which for example has helped us understand the impact of CSIT discrepancies. However, it provides only a rough understanding of the transmission and a more accurate evaluation of the performance is necessary. In [15], a relation between the number of bits used for the CSI feedback and the average rate achieved is provided. This result is very useful as it provide design guidelines which can be used by engineers to design feedback channels in practical networks. This work aims at generalizing to the case of distributed CSIT the finite rate feedback study done by Jindal in the centralized CSIT configuration [15].

Considering the same transmission setting as in the previous section, we go beyond the DoF analysis to study the *rate loss* between the transmission with perfect CSIT and the transmission with limited feedback. Quantifying the rate loss provides a more precise understanding of the transmission and is more relevant for practical system design as it allows to guaranty a minimal performance achieved with the limited feedback.

5.1 CSIT Configuration and Precoding Schemes

We keep in this chapter the same CSIT model as in the previous chapter. We also study for the sake of comparison the transmission with (imperfect) centralized CSIT. However, we will consider a different ZF scheme. This is done as a mean to discuss the different possible ZF precoders. Hence,

we study the ZF precoder $\mathbf{U}^\star \triangleq [\boldsymbol{u}_1^\star, \ldots, \boldsymbol{u}_K^\star]$ which, in the case of perfect CSIT, is defined from

$$\boldsymbol{u}_i^\star \triangleq \left(\mathbf{I}_K - \mathbf{H}_i \left(\mathbf{H}_i^{\mathrm{H}} \mathbf{H}_i\right)^{-1} \mathbf{H}_i^{\mathrm{H}}\right) \boldsymbol{h}_i, \qquad \forall i \in \{1, \ldots, K\} \tag{5.1}$$

where we have defined

$$\mathbf{H}_i \triangleq \left[\boldsymbol{h}_1 \ \ldots \ \boldsymbol{h}_{i-1} \ \boldsymbol{h}_{i+1} \ \ldots \ \boldsymbol{h}_K\right], \qquad \forall i \in \{1, \ldots, K\}. \tag{5.2}$$

Remark 8. *The difference with the ZF precoder used in Chapter 4 comes from the fact that there is no normalization of the beamformer. This means that the power constraint is only fulfilled in average over the channel realizations. However, this scheme is interesting as the power constraint cannot –in general– be perfectly fulfilled when considering distributed CSIT.* □

5.1.1 ZF with Centralized CSIT

In the case of centralized CSIT, the same estimate is received at all the TXs. We then denote the full multi-user channel estimate by $\hat{\mathbf{H}} = [\hat{\boldsymbol{h}}_1, \ldots, \hat{\boldsymbol{h}}_K]$. We consider the same CSIT model as for the distributed CSIT such that

$$\hat{\boldsymbol{h}}_i = \sqrt{1 - \sigma_i^2} \boldsymbol{h}_i + \sigma_i \boldsymbol{\delta}_i, \qquad \forall i \in \{1, \ldots, K\} \tag{5.3}$$

with $\sigma_i^2 = 2^{-\frac{B_i}{K-1}}$ and B_i being the number of bits used for the quantization of $\boldsymbol{h}_i$ and δ_i being i.i.d. $\mathcal{N}_{\mathbb{C}}(0, 1)$. With centralized CSIT, the ZF precoder $\mathbf{U}^{\mathrm{CCSI}} \triangleq [\boldsymbol{u}_1^{\mathrm{CCSI}}, \ldots, \boldsymbol{u}_K^{\mathrm{CCSI}}]$ is then given by

$$\boldsymbol{u}_i^{\mathrm{CCSI}} \triangleq \left(\mathbf{I}_K - \hat{\mathbf{H}}_i \left(\hat{\mathbf{H}}_i^{\mathrm{H}} \hat{\mathbf{H}}_i\right)^{-1} \hat{\mathbf{H}}_i^{\mathrm{H}}\right) \hat{\boldsymbol{h}}_i, \qquad \forall i \in \{1, \ldots, K\} \tag{5.4}$$

and

$$\hat{\mathbf{H}}_i \triangleq \left[\hat{\boldsymbol{h}}_1 \ \ldots \ \hat{\boldsymbol{h}}_{i-1} \ \hat{\boldsymbol{h}}_{i+1} \ \ldots \ \hat{\boldsymbol{h}}_K.\right], \qquad \forall i \in \{1, \ldots, K\}. \tag{5.5}$$

5.1.2 ZF with Distributed CSI

Considering distributed CSIT, the ZF precoder $\mathbf{U}^{\mathrm{DCSI}} \triangleq [\boldsymbol{u}_1^{\mathrm{DCSI}}, \ldots, \boldsymbol{u}_K^{\mathrm{DCSI}}]$ is then given by

$$\{\boldsymbol{u}_i^{\mathrm{DCSI}}\}_j \triangleq \{\boldsymbol{u}_i^{(j)}\}_j, \qquad \forall i, j \in \{1, \ldots, K\} \tag{5.6}$$

with

$$\boldsymbol{u}_i^{(j)} \triangleq \left(\mathbf{I}_K - \hat{\mathbf{H}}_i^{(j)} \left((\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}} \hat{\mathbf{H}}_i^{(j)}\right)^{-1} (\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}\right) \hat{\boldsymbol{h}}_i^{(j)}, \qquad \forall i, j \in \{1, \ldots, K\} \tag{5.7}$$

with

$$\hat{\mathbf{H}}_i^{(j)} \triangleq \left[\hat{\boldsymbol{h}}_1^{(j)} \;\; \ldots \;\; \hat{\boldsymbol{h}}_{i-1}^{(j)} \;\; \hat{\boldsymbol{h}}_{i+1}^{(j)} \;\; \ldots \;\; \hat{\boldsymbol{h}}_K^{(j)}\right], \qquad \forall i, j \in \{1, \ldots, K\}. \tag{5.8}$$

5.2 Broadcast Channel with Centralized CSIT

We study in this section the feedback design in the conventional CSIT scenario of centralized CSIT. The scaling of variance of the estimation error (or equivalently the scaling of the number of feedback bits) in terms of the SNR P is a well known result [15, 19], which has had a strong impact on practical systems. Yet, it is obtained with a different ZF precoder and a different quantization scheme than considered here such that we have to prove that these results remain valid in the system model used here. Furthermore, we provide additional insights by studying $\mathrm{E}[|\boldsymbol{h}_i^{\mathrm{H}} \boldsymbol{u}_j^{\mathrm{CCSI}}|^2]$ for $i \neq j$, which represents the average leaked interference at RX i resulting from the transmission to RX j.

5.2.1 Rate Loss Analysis

Our focus in this work is on the interference management via ZF precoding such that the amount of interference leaked to one user represents a crucial metric. This analysis is a preliminary step for the analysis of the rate loss but also provides an interesting insight.

Proposition 8. In the BC with centralized CSIT, it holds that

$$E[|\boldsymbol{h}_j^{\mathrm{H}} \boldsymbol{u}_i^{\mathrm{CCSI}}|^2] = \sigma_j^2 + o(\sigma_j^2), \qquad \forall i \neq j \in \{1, \ldots, K\}. \tag{5.9}$$

Proof. A detailed proof is provided in Appendix .6. □

It is interesting to observe that the formula obtained depends only on the indice of the channel but not on the indice of the beamformer. Furthermore, the amount of interference leaked depends only on the accuracy with which the channel $\boldsymbol{h}_i$ is known but not on the CSI relative to the other channels, which could even not be known at all.

The following upper bound for the rate loss follows easily from Proposition 8.

Theorem 6. In the BC with centralized CSIT, the rate loss at user i, which we denote by $\Delta_{\mathrm{R},i}^{\mathrm{CCSI}}$, can be upperbounded at high SNR as

$$\Delta_{\mathrm{R},i}^{\mathrm{CCSI}} \leq \log_2\left(1+(K-1)P\sigma_i^2\right)+o(1). \tag{5.10}$$

Proof. Starting from the definitions of the rate loss, we write

$$\begin{aligned}
\Delta_{\mathrm{R},i}^{\mathrm{CCSI}} &\triangleq \mathrm{E}\left[\log_2(1+P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star}|^2)\right] \\
&\quad - \mathrm{E}\left[\log_2\left(1+\frac{P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\mathrm{CCSI}}|^2}{1+\sum_{j=1,j\neq i}^{K}P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{CCSI}}|^2}\right)\right] && (5.11)\\
&= \mathrm{E}\left[\log_2(1+P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star}|^2)\right] - \mathrm{E}\left[\log_2\left(1+\sum_{j=1}^{K}P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{CCSI}}|^2\right)\right] \\
&\quad + \mathrm{E}\left[\log_2\left(1+\sum_{j=1,j\neq i}^{K}P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{CCSI}}|^2\right)\right] && (5.12)\\
&\overset{(a)}{\leq} \mathrm{E}\left[\log_2\left(1+\sum_{j=1,j\neq i}^{K}P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{CCSI}}|^2\right)\right] && (5.13)
\end{aligned}$$

with (a) due to the fact that $\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star}$ and $\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\mathrm{CCSI}}$ are random variables with the same distribution[1]. Using Jensen's inequality for concave functions, we can then write

$$\Delta_{\mathrm{R},i}^{\mathrm{CCSI}} \leq \log_2\left(1+(K-1)P\mathrm{E}\left[|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{CCSI}}|^2\right]\right). \tag{5.14}$$

The proof concludes by using Proposition 8 to compute the expectation inside (5.14). □

5.2.2 Feedback Design

In Section 4.1.2, a digital quantization scheme is described with the variance of the estimation error given by $\sigma_i^2 = 2^{-\frac{B_i}{K-1}}$. Considering that a maximal number of bits B can be allocated freely among all the users, it is then relevant to determine how the feedback bits should be allocated in order to minimize the total rate loss. Relaxing the constraint that the number of bits

[1] This follows from the fact that there are as many TXs as RXs such that the nullspace of $\mathbf{H}_i$ is of dimension one.

has to be an integer, we can use the upperbound for the rate loss given in Theorem 6 to write the optimization problem as follows:

$$\min_{\{B_i\}_i} \log_2\left(1+(K-1)P2^{-\frac{B_i}{K}}\right) \text{ , s. to. } \sum_{i=1}^{K} B_i = B_{\text{sum}}. \tag{5.15}$$

It is the minimization of a sum of convex functions subject to a linear constraint and the optimal solution is well known to be [105]

$$B_i = \frac{B_{\text{sum}}}{K}, \qquad \forall i. \tag{5.16}$$

Thus, we consider in the following the case where the feedback is allocated uniformly to all the users such that $\sigma_i^2 = 2^{-\frac{B}{K-1}}$ for some given B. The following theorem provides then an upperbound for the rate loss as a function of the number of feedback bits available B.

Theorem 7. In the BC with centralized CSIT with $\sigma_i^2 = 2^{-\frac{B}{K-1}}, \forall i$, the rate loss $\Delta_{\text{R},i}^{\text{CCSI}}$ is upper bounded at high SNR by $\log_2(1+b)+o(1)$ bits if $B \geq B^{\text{CCSI}}$ with

$$B^{\text{CCSI}} \triangleq (K-1)\log_2\left(\frac{(K-1)P}{b}\right). \tag{5.17}$$

Proof. Inserting $\sigma_i^2 = 2^{-\frac{B}{K-1}}$ inside (5.10) and writing the number of bits B as $B = (K-1)\alpha\log_2(\beta P)$ for $\alpha > 0, \beta > 0$, gives $\sigma^2 = (\beta P)^{-\alpha}$ which gives

$$\Delta_{\text{R},i}^{\text{CCSI}} \leq \log_2\left(1+(K-1)\beta^{-\alpha}(P)^{1-\alpha}+o(P^{1-\alpha})\right)+o(1). \tag{5.18}$$

The scaling in the SNR P of the rate loss is given in (5.18) to be equal to $\max(1-\alpha,0)$ such that the DoF achieved with limited feedback is $1-\max(1-\alpha,0) = \min(\alpha,1)$. Hence, we choose $\alpha = 1$ to achieve the maximal DoF. Setting the rate loss $\Delta_{\text{R},i}^{\text{CCSI}}$ lower or equal to $\log_2(1+b)$ and solving for β gives the result of the theorem. □

We can see that the expression obtained is very similar to the expression obtained in [15] which would gives with our notations $(K-1)\log_2(KP/b)$. The only difference comes from the replacement of K by $K-1$ inside the logarithm, which follows from the different distribution of the CSI errors.

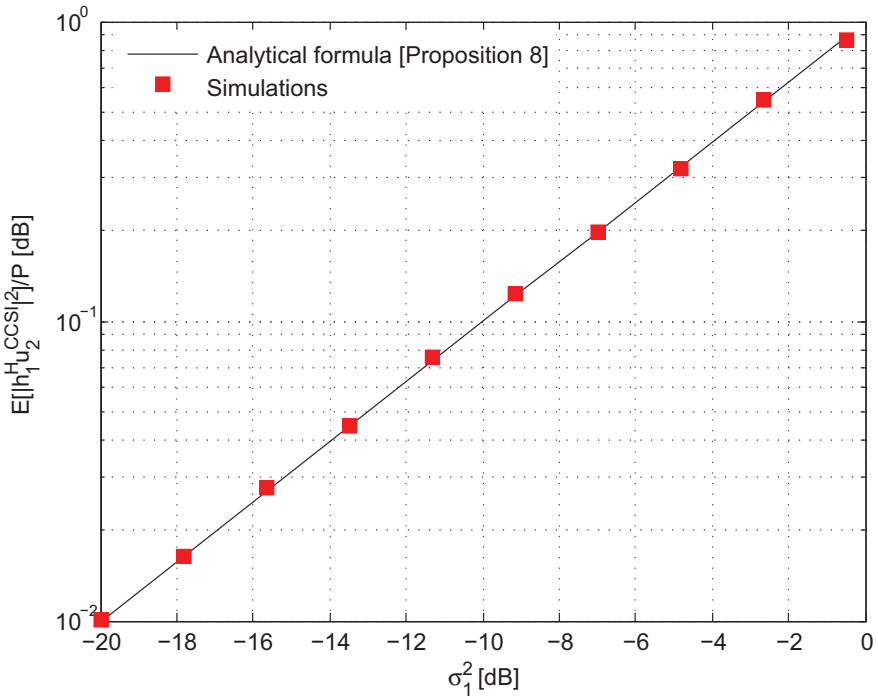


Figure 5.1: Average interference power $\mathrm{E}[|\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{u}_2^{\mathrm{CCSI}}|^2]$ after normalization by the SNR P as a function of the variance σ_1^2

.

5.2.3 Simulation Results

We will now verify by simulations for a BC with centralized CSIT and $K = 5$ the analytical results provided above. In a first step, we let variances of the CSIT errors be given by

$$\sigma_1^2 = \frac{1}{\sqrt{P}}, \qquad \sigma_i^2 = \frac{1}{P}, \forall i \neq 1. \tag{5.19}$$

With these parameters, we show in Fig. 5.1 the average interference power normalized by the average SNR P, $\mathrm{E}[|\boldsymbol{h}_1^{\mathrm{H}}\boldsymbol{u}_2^{\mathrm{CCSI}}|^2]$ as a function of the variance σ_1^2. We can verify the perfect match between the analytical expression and the simulations.

We show then the average rate per user achieved in that scenario in Fig. 5.2. We can see that the analytical expression is indeed a lower bound

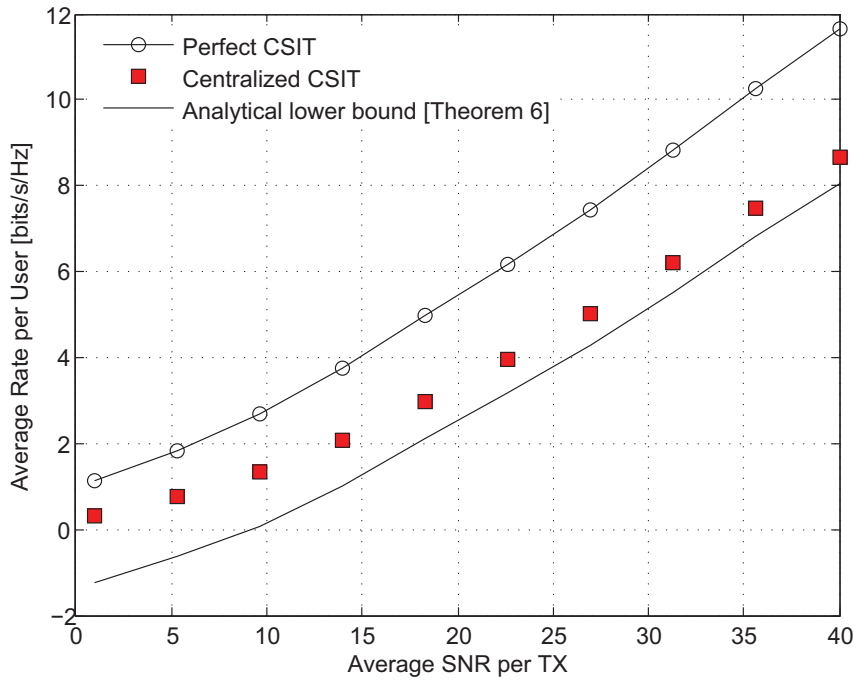


Figure 5.2: Average rate per user versus the average SNR P with centralized CSIT.

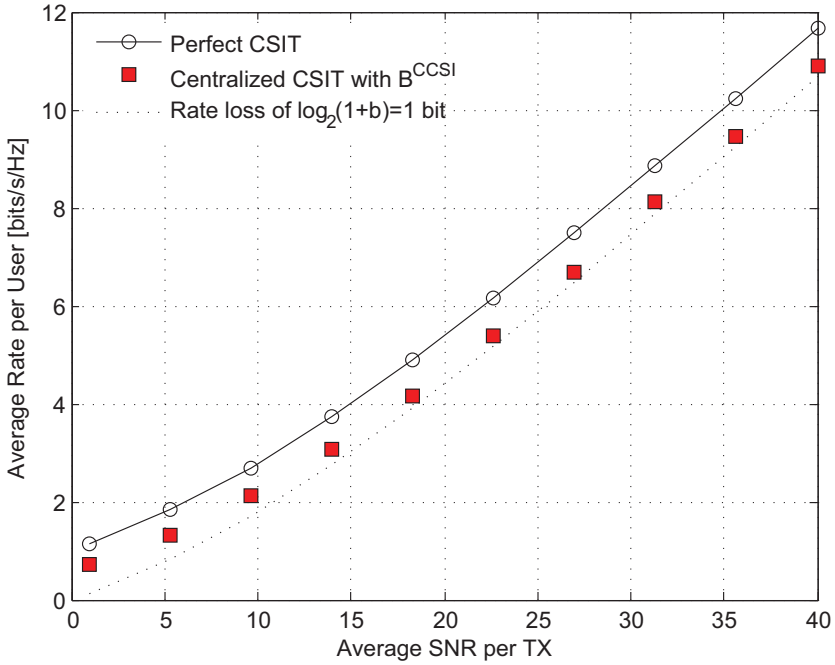


Figure 5.3: Average rate per user versus the average SNR P with centralized CSIT considering digital feedback.

for the rate achieved with imperfect centralized CSIT, and that the gap remains relatively small, particularly at high SNR.

Finally, we consider that the digital quantization scheme described earlier is used. Hence, we show in Fig. 5.3 the average rate per user achieved when using a number of quantization bits equal to B^{CCSI} as defined in (5.17) with $b = 1$. We compare it to the average rate obtained using perfect CSIT and to the maximal tolerated rate loss of $\log_2(1 + b) = 1$ bit. As expected, the rate loss remains below the threshold value.

5.3 Broadcast Channel with Distributed Limited CSI

We now turn to the analysis of the transmission with distributed CSIT, which represents the core of our work. It is intuitive that the distributedness of the CSIT leads to an increased amount of leaked interference since the precoding coefficients are less coordinated. However, this degradation has not yet been quantified and the feedback requirements with distributed CSIT are unknown. Our main goal is henceforth to see if the analysis carried out above for the centralized case can be adapted to the distributed CSIT scenario.

5.3.1 Rate Loss Analysis

In contrast to the analysis of the centralized CSIT case, we do not study directly the leaked interference $|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{DCSI}}|^2$ but we focus instead on the norm of the difference between the ZF precoder computed and the ZF precoder based on perfect CSIT $\|\boldsymbol{u}_j^{\mathrm{DCSI}} - \boldsymbol{u}_j^{\star}\|^2$. This is due to the fact that the interference $|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{DCSI}}|^2$ has a very complicated distribution in the case of precoding with distributed CSIT, which has prevented us for studying it directly.

Proposition 9. In the BC with distributed CSIT, it holds with probability one that

$$\boldsymbol{u}_i^{(j)} = \boldsymbol{u}_i^{\star} + \boldsymbol{a}_i^{(j)} + o(\max_q \sigma_q^{(j)}), \qquad \forall i,j \in \{1,\dots,K\}. \tag{5.20}$$

with

$$\mathrm{E}[|\boldsymbol{e}_p^{\mathrm{H}}\boldsymbol{a}_i^{(j)}|^2] = \frac{2\sum_{k=1,k\neq i}^{K}(\sigma_k^{(j)})^2 + (\sigma_i^{(j)})^2}{K}, \qquad \forall i,j,p. \tag{5.21}$$

Proof. A detailed proof is provided in Appendix .4. □

As intuitively expected from the symmetry, the average error done on the precoding does not depend on the row (i.e., on p in (5.21)). In contrast, the dependency on i is rather unexpected, since the ith beamformer is less dependent on the accuracy of $\boldsymbol{h}_i$ than on the accuracy of the other channels.

Proposition 9 provides the first expression linking the accuracy (in the sense of the MSE with the precoder based on perfect CSIT) with which a precoder is computed to the quality of the CSIT available.

Remark 9. *We can note that Proposition 9 remains clearly valid in the case of centralized CSIT. In contrast, Proposition 8 does not a priori holds in the case of distributed CSIT. It is the particular relation between the elements of the beamformer* $\boldsymbol{u}_i^{\mathrm{CCSI}}$ *with centralized CSIT which makes that the stronger result given in Proposition 8 holds.* □

Building upon this proposition, we can then quantify the rate loss due to the distributed CSIT as it is done in the following theorem.

Theorem 8. In the BC with distributed CSIT, if all the variances $(\sigma_i^{(j)})^2$ decrease polynomially with the SNR P, the rate loss $\Delta_{\mathrm{R},i}^{\mathrm{DCSI}}$ at user i can be upperbounded at high SNR as

$$\Delta_{\mathrm{R},i}^{\mathrm{DCSI}} \leq \mu(K)+\log_2\Bigg(1+P\sum_{j=1}^{K}\left((2K-3)\sum_{k=1,k\neq i}^{K}(\sigma_k^{(j)})^2+2(K-1)(\sigma_i^{(j)})^2\right)\Bigg)+o(1) \tag{5.22}$$

with the function $\mu(x)$ defined for $x > 0$ as

$$\mu(x) \triangleq \log_2\left(3+2\log(x)\right). \tag{5.23}$$

Proof. A detailed proof is provided in Appendix .5. □

By increasing by 1 the coefficient $2K-3$ to $2(K-1)$, the bound is made only slightly looser and rewritten in the following simpler form.

Corollary 4. In the BC with distributed CSIT, if all the variances $(\sigma_i^{(j)})^2$ decrease polynomially with the SNR P, the rate loss $\Delta_{\mathrm{R},i}^{\mathrm{DCSI}}$ at user i can be upperbounded at high SNR as

$$\Delta_{\mathrm{R},i}^{\mathrm{DCSI}} \leq \mu(K) + \log_2\left(1+\frac{2P(K-1)}{K}\sum_{j=1}^{K}\sum_{k=1}^{K}(\sigma_k^{(j)})^2\right)+o(1). \tag{5.24}$$

The expression in (5.24) is interesting because it shows in a very simple way how the accuracy of all channel elements impact the derived expression. This is in strong contrast with the centralized case where only the quality of $\boldsymbol{h}_i$ has an impact on the rate of user i.

The additive term $\mu(K)$ is believed to result only from our difficulty to bound the rate loss when the CSIT is distributed. However, this term increases doubly logarithmic with the number of users K such that it is not very significant as the number of users K increases. Still, we define $\Gamma_{\approx,i}^{\mathrm{DCSI}}$

as

$$\Gamma_{\approx,i}^{\mathrm{DCSI}} \triangleq \log_2\left(1+\frac{2P(K-1)}{K}\sum_{j=1}^{K}\sum_{k=1}^{K}(\sigma_k^{(j)})^2\right) \tag{5.25}$$

and we will use $\Gamma_{\mathrm{R},i}^{\mathrm{DCSI}}$ as an approximate expression for the rate loss to obtain an approximate feedback rate. Note that due to this conjecture, the feedback rate obtained from (5.25) will not guaranty that the rate loss remains below the given threshold and will have to use (5.24) to obtain a formal guaranty. Nevertheless, we will verify by simulations our conjecture that removing $\mu(K)$ leads to a good approximation of the true rate loss.

5.3.2 Feedback Design

The dependency between the upperbound for the rate loss obtained in (5.24) and the CSIT accuracy is very simple since it only depends on the sum of all the variances. Hence, we can very easily find the bit allocation to minimize this upperbound for the rate loss. Using the relation between the variance of the estimation error and the number of feedback bits, this gives

$$\min_{\{B_i^{(j)}\}_{i,j}} \sum_{i,j} 2^{-\frac{B_i^{(j)}}{K-1}}, \quad \text{s. to} \sum_{i,j} B_i^{(j)} \leq B. \tag{5.26}$$

This optimization problem is the minimization of a convex function subject to a linear constraint and it easily solved to obtain

$$B_i^{(j)} = \frac{B}{K^2}, \qquad \forall i,j. \tag{5.27}$$

Thus, it is optimal in the homogeneous setting to allocate uniformly the bits to the channel vectors and to the TXs. A critical guideline for practical design is then finding what is the minimal value for B to ensure a maximal rate loss. Such a guideline is provided in the following theorem.

Theorem 9. In the BC with distributed CSIT and $(\sigma_i^{(j)})^2 = 2^{-\frac{B}{K-1}}, \forall i,j$, the rate loss is upper bounded by $\log_2(1+b)+o(1)$ bits if $b > 2+2\log(K)$ and $B \geq B^{\mathrm{DCSI}}$ with

$$B^{\mathrm{DCSI}} = (K-1)\log_2\left(\frac{(2K-1)(K-1)P}{b}\right)+(K-1)\log_2\left(\frac{b(3+2\log(K))}{b-2-2\log(K)}\right). \tag{5.28}$$

Proof. Similarly to the proof of Theorem 7, we use the parametrization $B = \beta P^{\alpha}$ with $\beta > 0, \alpha > 0$ inside (5.22). Setting the right hand side of (5.24) lower or equal than $\log_2(1+b)$ gives

$$\mu(K) + \log_2\left(1 + P(2K-1)(K-1)\beta^{-\alpha}P^{1-\alpha}\right) \leq \log_2(1+b). \tag{5.29}$$

Achieving the maximal DoF requires $\alpha = 1$, and solving for β then gives the desired expression. □

In Theorem 9, we provide the desired feedback rate. Considering K larger, this feedback rate is roughly $K\log_2(K)$ larger than the feedback rate obtained in the centralized case, implying that the difference between the two expressions becomes increasingly large as the number of users K increases.

Considering the approximate expression $\Gamma^{\mathrm{DCSI}}_{\approx,i}$, we obtain $B^{\mathrm{DCSI}}_{\approx}$ given by

$$B^{\mathrm{DCSI}}_{\approx} = (K-1)\log_2\left(\frac{(2K-1)(K-1)P}{b}\right). \tag{5.30}$$

5.3.3 Simulation Results

We verify now by simulations in the BC with distributed CSIT and $K = 5$ users the theoretical results stated above. We start by discussing Proposition 9 and Theorem 8 for the following CSIT configuration:

$$(\sigma_i^{(1)})^2 = P^{-\frac{2}{3}}, \qquad \forall i \tag{5.31}$$

$$(\sigma_i^{(j)})^2 = P^{-1}, \qquad \forall i, j \neq 1. \tag{5.32}$$

In this scenario we show in Fig. 5.4 the average difference $\mathrm{E}[|\boldsymbol{e}_1^{\mathrm{T}}(\boldsymbol{u}_2^{\mathrm{DCSI}} - \boldsymbol{u}_2^{\star})|^2] = \mathrm{E}[|\boldsymbol{e}_1^{\mathrm{T}}(\boldsymbol{u}_2^{(1)} - \boldsymbol{u}_2^{\star})|^2]$ and we can verify the very good match between the simulations and the analytical results.

We then turn to Fig. 5.5 where the average rate per user is shown as a function of the SNR. We compare the simulation results to the lower bound given in Theorem 8 and the approximated expression in (5.25). We can see that the lower bound shows a large gap compared to the simulations results while the approximated expression presents also a gap, but smaller.

Finally, we show in Fig. 5.6, the average rate achieved when using the digital quantization model with different feedback rate. In particular, we show that using B^{CCSI} does not allow to bound the rate loss under the desired threshold. On the opposite, using the approximated feedback rate $B^{\mathrm{DCSI}}_{\approx}$ is

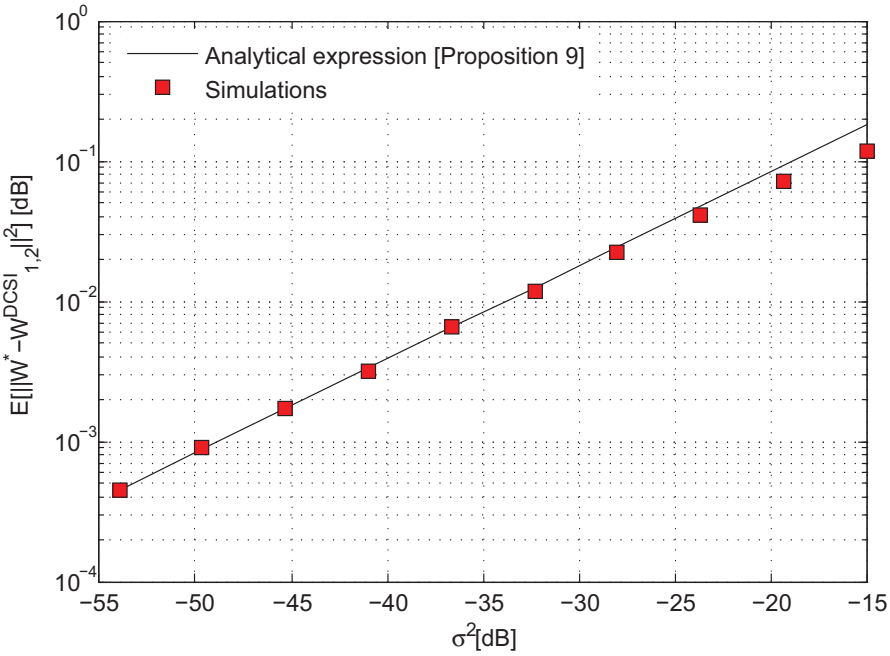


Figure 5.4: Average difference $\mathrm{E}[|\boldsymbol{e}_1^{\mathrm{T}}(\boldsymbol{u}_2^{(1)} - \boldsymbol{u}_2^{\star})|^2]$ as a function of the CSI quality $(\sigma_i^{(1)})^2 = P^{-\frac{2}{3}}$.

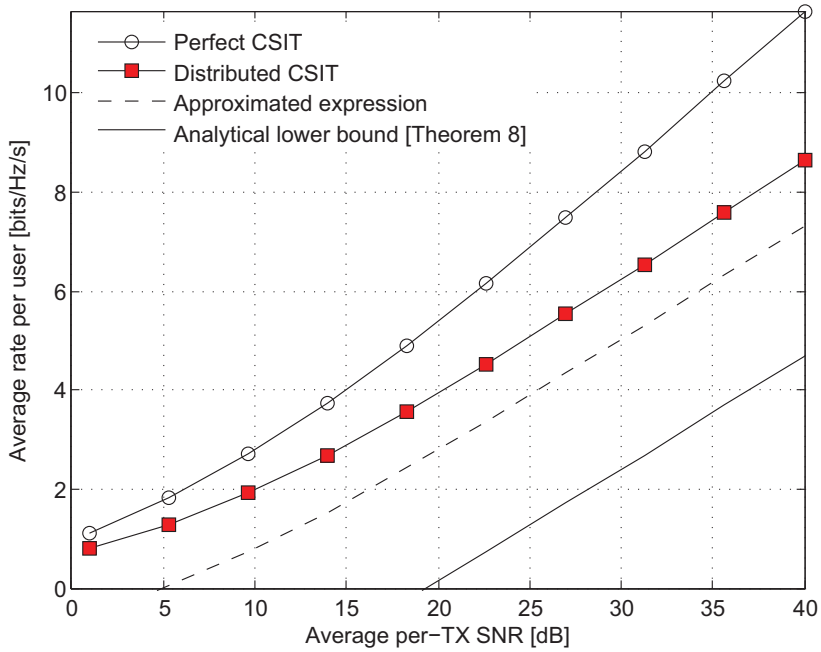


Figure 5.5: Average rate per user versus the average SNR P with distributed CSIT.

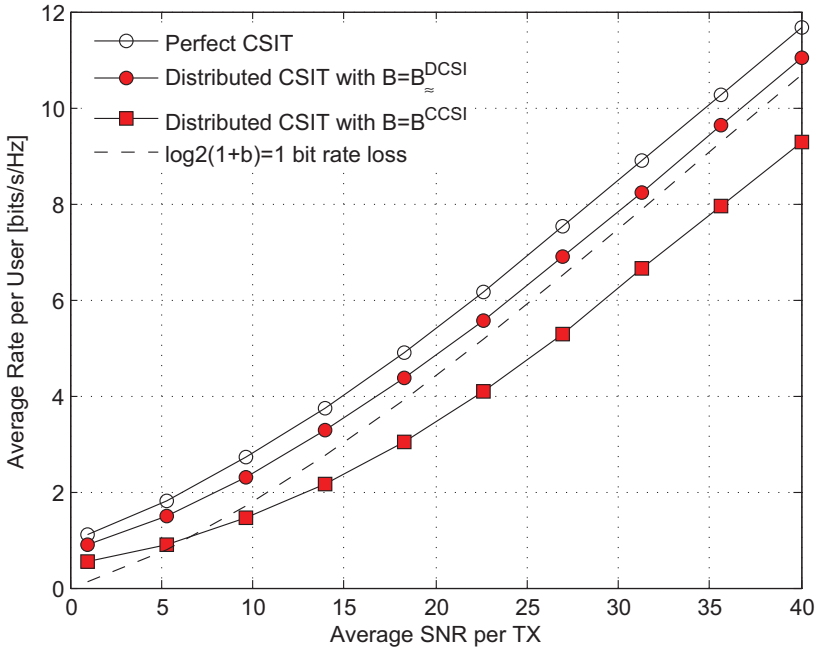


Figure 5.6: Average rate per user versus the average SNR P with distributed CSIT and digital quantization.

sufficient for restraining the rate loss to be smaller than the desired threshold. Note that because of the different scaling in the number of users, the loss achieved using B^{CCSI} increases with the number of users K.

5.4 Conclusion

To study the impact of the resulting CSI discrepancies at the different TXs, we have discussed how it compares to the conventional *centralized CSIT* configuration where all the TXs share the same channel estimate. We have first studied the dependency between the accuracy with which a precoder is computed at a TX and the accuracy with which the ZF precoder is computed. Furthermore, we have derived a sufficient feedback rate to ensure that the rate loss compared to the transmission with perfect CSIT remains below a threshold value. Interestingly, the expression obtained in Theorem 9 can be seen to increase more quickly with the number of users K than its counterpart in the BC with *centralized CSIT* in Theorem 7. Hence, not taking into account the CSI distributedness in the feedback design leads to important performance degradations, especially as the total number of users served increases. We have provided here an upper-bound for the rate loss and the derivation of a lower-bound is the focus of undergoing research. However, the statistical distribution of the interference with distributed CSIT makes this problem especially challenging. The extension to scenarios with a different pathloss between every TX and every RX represents also an interesting research problem. Finally, we have studied the performance of ZF and it is believed that using robust (regularized) ZF precoding would help reduce the impact of the CSI discrepancies. Hence, studying how regularization can help reduce the cost of CSIT distributedness will be the focus of future research.

Chapter 6

DoF of IA with Distributed CSIT

We have studied in the previous chapters the impact of distributed CSIT when joint precoding is applied at all the TXs, and we now turn to the analysis of the settings where the users data symbols are not shared between the TXs. We consider in particular that IA schemes for MIMO statics ICs are to be used [49].

Note that for this section only, multiple-streams transmission are considered such that the precoding matrix $\mathbf{T}_i$ is of size $M_i \times d_i$ where d_i is the number of independent data symbols sent to user i.

6.1 Distributed CSIT and Distributed Precoding

We introduce further in this chapter the normalized estimate for the channel between TX k and RX i, denoted by $\tilde{\mathbf{H}}_{ik}^{\mathrm{H}}$ and equal to

$$\tilde{\mathbf{H}}_{ik}^{\mathrm{H}} \triangleq \frac{\mathbf{H}_{ik}^{\mathrm{H}}}{\|\mathbf{H}_{ik}\|_{\mathrm{F}}}. \tag{6.1}$$

We define $\tilde{\mathbf{H}}_{i,k}^{(j)}$ in a similar way from $\hat{\mathbf{H}}_{i,k}^{(j)}$. As in the previous chapters, TX j computes its precoder based only on $\tilde{\mathbf{H}}^{(j)}$. In fact, TX j computes then the full IA transmission strategy (in terms of the precoders and receive filters $\mathbf{U}_k^{(j)}, k = 1 \dots K$ and $\mathbf{G}_i^{(j)}, i = 1 \dots K$) based on $\tilde{\mathbf{H}}^{(j)}$ such as to fulfill

$$(\mathbf{G}_i^{(j)})^{\mathrm{H}} (\tilde{\mathbf{H}}_{i,k}^{(j)})^{\mathrm{H}} \mathbf{U}_k^{(j)} = \mathbf{0}_{d_i \times d_j} \quad \forall k \neq i \tag{6.2}$$

where $\mathbf{U}_k^{(j)}$ is the unitary precoder designed to be used by TX k and $\mathbf{G}_i^{(j)}$ is the receive filter assumed at RX i. Due to the distributed precoding assumption, the precoder $\mathbf{U}_j^{(j)}$ is used for the actual transmission at TX j, while the $\mathbf{U}_i^{(j)}, i \neq j$ are discarded. In total, this gives

$$\mathbf{U}_j = \mathbf{U}_j^{(j)}, \qquad \forall j. \tag{6.3}$$

This distributed CSIT setting is depicted and compared to the centralized CSIT configuration in Fig. 6.1.

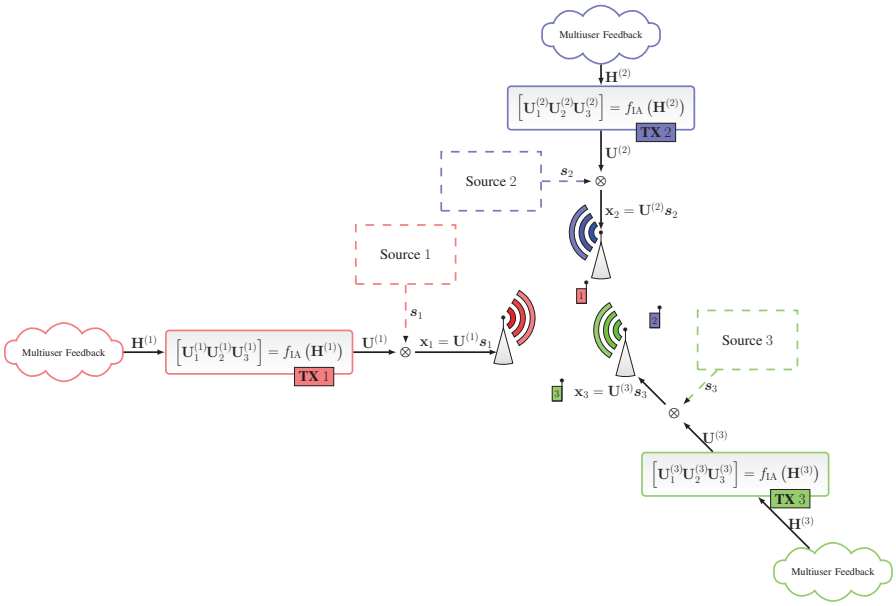


Figure 6.1: Symbolic representation of IA with precoding and distributed CSIT.

Let us assume that $\tilde{\mathbf{H}}_{i,k}^{(j)}$ results from the quantization of $\tilde{\mathbf{H}}_{i,k}$, using a quantization scheme using $B_{i,k}^{(j)}$ bits according to

$$\tilde{\mathbf{H}}_{i,k}^{(j)} = \underset{\mathrm{vect}(\mathbf{W}) \in \mathcal{W}_{i,k}^{(j)}}{\operatorname{argmin}} \left\| \tilde{\mathbf{H}}_{i,k} - \mathbf{W} \right\|_{\mathrm{F}}, \qquad \forall k, i, j, \tag{6.4}$$

where $\mathcal{W}_{i,k}^{(j)}$ contains $2^{B_{i,k}^{(j)}}$ vectors of size $\mathbb{C}^{N_i M_k}$ isotropically distributed over the unit-sphere and rotated to have their first element real-valued. We

further define

$$(\sigma_{i,k}^{(j)})^2 \triangleq \mathrm{E}_{\mathcal{H},\mathcal{W}}\left[\left\|\tilde{\mathbf{H}}_{i,k}^{(j)} - \tilde{\mathbf{H}}_{i,k}\right\|_{\mathrm{F}}^2\right], \qquad \mathbf{N}_{i,k}^{(j)} \triangleq \frac{\tilde{\mathbf{H}}_{i,k}^{(j)} - \tilde{\mathbf{H}}_{i,k}}{\sigma_{i,k}^{(j)}}, \tag{6.5}$$

where $\mathrm{E}_{\mathcal{W}}[\cdot]$ denotes the expectation over the random codebooks. It then gives

$$\tilde{\mathbf{H}}_{i,k}^{(j)} = \tilde{\mathbf{H}}_{i,k} + \sigma_{i,k}^{(j)}\mathbf{N}_{i,k}^{(j)}. \tag{6.6}$$

Since there is no confusion possible we use the short notation $\mathrm{E}[\cdot]$ instead of $\mathrm{E}_{\mathcal{H},\mathcal{W}}[\cdot]$.

Using the results in Section 4.1.2, the variance of the estimation error can then be related to the number of quantization bits as follows.

Proposition 10 ([80, Theorem 2]). When the size $L_{i,k}^{(j)} = 2^{B_{i,k}^{(j)}}$ of the random codebook is sufficiently large, it then holds that

$$(\sigma_{i,k}^{(j)})^2 = C_{i,k}^{(j)} 2^{-B_{i,k}^{(j)}/(N_i M_k - 1)} \tag{6.7}$$

for some constant $C_{i,k}^{(j)} > 0$.

Similarly to Chapter 4 and Chapter 5, we define the *CSIT scaling coefficients* $A_{i,k}^{(j)}$ as

$$A_{i,k}^{(j)} \triangleq \lim_{P\to\infty} \frac{B_{i,k}^{(j)}}{B_{i,k}^{\star}}, \forall k, i, \tag{6.8}$$

where we have defined

$$B_{i,k}^{\star} \triangleq (N_i M_k - 1)\log_2(P). \tag{6.9}$$

The pre-log coefficient $N_i M_k - 1$ corresponds to the number of channel coefficients to feedback after normalization of the channel matrix. $B_{i,k}^{\star}$ is the number of bits which corresponds to a quantization error decreasing as P^{-1}, which is essentially perfect in terms of DoF [15, 81]. Hence, $A_{i,k}^{(j)}$ can be seen as the fraction of the feedback requirements to achieve the maximal DoF.

Remark 10. *Note however that the CSIT scaling coefficients introduced here do not exactly correspond to the ones defined for joint precoding previously. Indeed, we consider here a different CSIT scaling coefficient for each MIMO wireless channel between one TX and one RX.* □

6.2 DoF Analysis with Static Coefficients and Distributed CSI

Let us now focus on the situation where every TX designs its precoder based on a different multi-user channel estimate. Hence, the precoding matrices used for the transmission do not form exactly an IA solution for any imperfect estimate of the multi-user channel. This is in contrast to the centralized case studied in [59, 60]. Hence, the analysis done in these works does not hold in the setting considered here and a new approach is required.

The analysis of this situation is complicated by the fact that the function that gives the precoders as a function of the channel coefficients can not be assumed to be continuous. This can be seen by considering that there are in general multiple solutions to the IA equations [106], while iterative algorithms, such as the iterative leakage minimization from [86], converge to one of the IA solutions. So far this convergence is not fully understood, and it can not be ruled out that a small change in the CSI (as in the case in the distributed CSI considered here) leads to a convergence to completely different solutions across the users.

Furthermore, the channel estimates at the different TXs are potentially of different accuracies such that it is not clear which accuracy dictates the DoF. Answering this question is the main goal of this work.

6.2.1 Sufficient Condition for an Arbitrary IA Scheme

Let us denote by $\mathbf{U}_i^\star$ and $\mathbf{G}_i^\star$ the precoder and the RX filter at TX i and RX i, respectively, when perfect CSIT is available at the TXs for *an arbitrary IA scheme*, i.e., verifying $(\mathbf{G}_i^\star)^{\mathrm{H}}\mathbf{H}_{ij}^{\mathrm{H}}\mathbf{U}_j^\star = \mathbf{0}_{d_i \times d_j}, \forall i \neq j$. We further define

$$\boldsymbol{\Delta}\mathbf{U}_i^{(j)} \triangleq \mathbf{U}_i^{(j)} - \mathbf{U}_i^\star, \qquad \forall i, j. \tag{6.10}$$

We now characterize the DoF achieved as a function of the precoder accuracy.

Proposition 11. In the IC with distributed CSIT as described in Section 6.1, if the CSIT is such that

$$\mathrm{E}[\|\boldsymbol{\Delta}\mathbf{U}_j^{(j)}\|_{\mathrm{F}}^2] \doteq P^{-\beta_j}, \qquad \forall j, \tag{6.11}$$

with $\beta_j \in [0, 1]$, then

$$\mathrm{DoF}_i \geq d_i \min_{j \neq i} \beta_j, \qquad \forall i \tag{6.12}$$

and where we have used $x \doteq y$ to represent the exponential equality in the SNR P, i.e., $\lim_{P\to\infty} \log_2(x)/\log_2(P) = \lim_{P\to\infty} \log_2(y)/\log_2(P)$.

Proof. Since we want to derive a lower bound for the DoF, we can choose $\mathbf{G}_k = \mathbf{G}_k^\star, \forall k$. Following a classical derivation [15, 19], we can write

$$R_i \geq R_i^\star - \mathrm{E}\left[\log_2\left|\mathbf{I}_{d_i} + P\sum_{j=1,j\neq i}^{K} (\mathbf{G}_i^\star)^{\mathrm{H}}\mathbf{H}_{i,j}^{\mathrm{H}}\mathbf{U}_j^{(j)}(\mathbf{U}_j^{(j)})^{\mathrm{H}}\mathbf{H}_{i,j}\mathbf{G}_i^\star\right|\right] \quad (6.13)$$

where we have defined

$$R_i^\star \triangleq \mathrm{E}\left[\log_2\left|\mathbf{I}_{d_i} + P(\mathbf{G}_i^\star)^{\mathrm{H}}\mathbf{H}_{i,i}^{\mathrm{H}}\mathbf{U}_i^\star(\mathbf{U}_i^\star)^{\mathrm{H}}\mathbf{H}_{i,i}(\mathbf{G}_i^\star)\right|\right]. \quad (6.14)$$

It is easily seen that $R_i^\star \doteq d_i \log_2(P)$, such that it remains to study the second term of (6.13), which we denote by $\mathcal{I}_i$. Since $(\mathbf{G}_i^\star)^{\mathrm{H}}\mathbf{H}_{i,j}^{\mathrm{H}}\mathbf{U}_j^\star = \mathbf{0}_{d_i\times d_j}$ for $i \neq j$, it holds that

$$\mathcal{I}_i = \mathrm{E}\left[\log_2\left|\mathbf{I}_{d_i} + P\sum_{j=1,j\neq i}^{K} (\mathbf{G}_i^\star)^{\mathrm{H}}\mathbf{H}_{i,j}^{\mathrm{H}}\boldsymbol{\Delta}\mathbf{U}_j^{(j)}(\boldsymbol{\Delta}\mathbf{U}_j^{(j)})^{\mathrm{H}}\mathbf{H}_{i,j}\mathbf{G}_i^\star\right|\right]. \quad (6.15)$$

Since $\|\mathbf{G}_i^\star\|_{\mathrm{F}}^2 = 1$, we can upper bound the interference to write

$$\begin{aligned}
\mathcal{I}_i &\leq \mathrm{E}\left[\log_2\left|\mathbf{I}_{d_i} + \left(P\sum_{j=1,j\neq i}^{K}\|\mathbf{H}_{i,j}\|_{\mathrm{F}}^2\|\boldsymbol{\Delta}\mathbf{U}_j^{(j)}\|_{\mathrm{F}}^2\right)\mathbf{I}_{d_i}\right|\right] \\
&\overset{(a)}{\leq} d_i\left(\mathrm{E}\left[\log_2\left(1+P\sum_{j=1,j\neq i}^{K}\|\mathbf{H}_{i,j}\|_{\mathrm{F}}^2\right)\right] + \mathrm{E}\left[\log_2\left(1+P\sum_{j=1,i\neq j}^{K}\|\boldsymbol{\Delta}\mathbf{U}_j^{(j)}\|_{\mathrm{F}}^2\right)\right]\right) \\
&\overset{(b)}{\leq} d_i\left(\mathrm{E}\left[\log_2\left(1+P\sum_{j=1,j\neq i}^{K}\|\mathbf{H}_{i,j}\|_{\mathrm{F}}^2\right)\right] + \log_2\left(1+P\sum_{j=1,i\neq j}^{K}\mathrm{E}\left[\|\boldsymbol{\Delta}\mathbf{U}_j^{(j)}\|_{\mathrm{F}}^2\right]\right)\right)
\end{aligned} \quad (6.16)$$

where inequality (a) can be seen to hold since only positive terms have been added and we have used Jensen's inequality to obtain inequality (b). Using that $\mathrm{E}[\|\boldsymbol{\Delta}\mathbf{U}_j^{(j)}\|_{\mathrm{F}}^2] \doteq P^{-\beta_j}$, we can write that

$$\sum_{j=1,j\neq i}^{K}\mathrm{E}\left[\|\boldsymbol{\Delta}\mathbf{U}_j^{(j)}\|_{\mathrm{F}}^2\right] \doteq P^{-\min_{j\neq i}\beta_j}. \quad (6.17)$$

Inserting (6.17) inside (6.16) and (6.13) gives

$$R_i \dot{\geq} d_i \left(\log_2(P) - \log_2(1 + PP^{-\min_{j\neq i}\beta_j}) \right) \tag{6.18}$$

$$\dot{\geq} d_i (\min_{j\neq i} \beta_j) \log_2(P), \tag{6.19}$$

which concludes the proof. □

Proposition 11 provides some insights into the performance by relating the accuracy with which the precoder is computed to the achieved DoF. However, the accuracy of the precoder design is difficult to relate to the accuracy of the CSIT. Indeed, this relation is dependent on the precoding method used and some precoding schemes are more or less robust to imperfections in the CSIT. Furthermore, obtaining the relation between the CSIT quality and the accuracy of the precoding is especially difficult in the case of iterative IA algorithms.

Indeed, in contrast to the conventional centralized CSIT configuration studied in [59, 60, 78, 107], it is not possible to study solely the IA alignment obtained at the end of the precoding scheme. The precoders $\{\mathbf{U}_j^{(j)}\}_j$ do not form together (a priori) an alignment solution for any of the multi-user channel estimates available at the TXs. Hence, the structure of the IA algorithm has to be studied to observe what is the impact of the CSIT imperfection over the precoding at each TX.

6.2.2 DoF Analysis in the 3-user Square MIMO IC

Perfect CSIT Solution We consider now a 3-user IC with $M_i = M, N_i = N, \forall i$ and $d_i = d, \forall i$. We also assume for the description of the IA scheme that perfect CSIT is available such that we denote the precoder used at TX j by $\mathbf{U}_j^\star$. Since we consider the tightly-feasible case [108], we have $M = N = 2d$. In that case, the IA constraints can be written as [54]

$$\begin{aligned} \text{span}\left(\tilde{\mathbf{H}}_{3,1}^{\text{H}}\mathbf{U}_1^\star\right) &= \text{span}\left(\tilde{\mathbf{H}}_{3,2}^{\text{H}}\mathbf{U}_2^\star\right), \\ \text{span}\left(\tilde{\mathbf{H}}_{1,2}^{\text{H}}\mathbf{U}_2^\star\right) &= \text{span}\left(\tilde{\mathbf{H}}_{1,3}^{\text{H}}\mathbf{U}_3^\star\right), \\ \text{span}\left(\tilde{\mathbf{H}}_{2,3}^{\text{H}}\mathbf{U}_3^\star\right) &= \text{span}\left(\tilde{\mathbf{H}}_{2,1}^{\text{H}}\mathbf{U}_1^\star\right). \end{aligned} \tag{6.20}$$

In particular, this system of equations can be easily seen to be fulfilled if the precoders verify

$$\begin{aligned}\mathbf{U}_1^\star \mathbf{\Lambda}_1 &= (\tilde{\mathbf{H}}_{3,1}^{\mathrm{H}})^{-1}\tilde{\mathbf{H}}_{3,2}^{\mathrm{H}}(\tilde{\mathbf{H}}_{1,2}^{\mathrm{H}})^{-1}\tilde{\mathbf{H}}_{1,3}^{\mathrm{H}}(\tilde{\mathbf{H}}_{2,3}^{\mathrm{H}})^{-1}\tilde{\mathbf{H}}_{2,1}^{\mathrm{H}}\mathbf{U}_1^\star \\ \mathbf{U}_3^\star &= (\tilde{\mathbf{H}}_{2,3}^{\mathrm{H}})^{-1}\tilde{\mathbf{H}}_{2,1}^{\mathrm{H}}\mathbf{U}_1^\star \\ \mathbf{U}_2^\star &= (\tilde{\mathbf{H}}_{1,2}^{\mathrm{H}})^{-1}\tilde{\mathbf{H}}_{1,3}^{\mathrm{H}}\mathbf{U}_3^\star\end{aligned} \tag{6.21}$$

for some diagonal matrix $\mathbf{\Lambda}_1$. We also define for clarity the matrix $\mathbf{Y}^\star$ equal to

$$\mathbf{Y}^\star \triangleq (\tilde{\mathbf{H}}_{3,1}^{\mathrm{H}})^{-1}\tilde{\mathbf{H}}_{3,2}^{\mathrm{H}}(\tilde{\mathbf{H}}_{1,2}^{\mathrm{H}})^{-1}(\tilde{\mathbf{H}}_{1,3}^{\mathrm{H}})(\tilde{\mathbf{H}}_{2,3}^{\mathrm{H}})^{-1}\tilde{\mathbf{H}}_{2,1}^{\mathrm{H}}. \tag{6.22}$$

The system of equations (6.21) is then fulfilled by setting

$$\begin{aligned}\mathbf{U}_1^\star &= \frac{1}{\sqrt{d}}\mathrm{EVD}(\mathbf{Y}^\star)\left[\boldsymbol{e}_1,\ldots,\boldsymbol{e}_d\right] \\ \mathbf{U}_3^\star &= \frac{1}{\|(\tilde{\mathbf{H}}_{2,3}^{\mathrm{H}})^{-1}\tilde{\mathbf{H}}_{2,1}^{\mathrm{H}}\mathbf{U}_1^\star\|_{\mathrm{F}}}(\tilde{\mathbf{H}}_{2,3}^{\mathrm{H}})^{-1}\tilde{\mathbf{H}}_{2,1}^{\mathrm{H}}\mathbf{U}_1^\star \\ \mathbf{U}_2^\star &= \frac{1}{\|(\tilde{\mathbf{H}}_{1,2}^{\mathrm{H}})^{-1}\tilde{\mathbf{H}}_{1,3}^{\mathrm{H}}\mathbf{U}_3^\star\|_{\mathrm{F}}}(\tilde{\mathbf{H}}_{1,2}^{\mathrm{H}})^{-1}\tilde{\mathbf{H}}_{1,3}^{\mathrm{H}}\mathbf{U}_3^\star\end{aligned} \tag{6.23}$$

where we denote by $\mathrm{EVD}(\mathbf{A})$ the eigenvector basis of the Hermitian matrix $\mathbf{A}$.

Distributed CSIT Solution With distributed CSIT, TX j computes using its channel estimate $\tilde{\mathbf{H}}^{(j)}$ the matrix

$$\mathbf{Y}^{(j)} = ((\tilde{\mathbf{H}}_{3,1}^{(j)})^{\mathrm{H}})^{-1}((\tilde{\mathbf{H}}_{3,2}^{(j)})^{\mathrm{H}}((\tilde{\mathbf{H}}_{1,2}^{(j)})^{\mathrm{H}})^{-1}(\tilde{\mathbf{H}}_{1,3}^{(j)}))^{\mathrm{H}}((\tilde{\mathbf{H}}_{2,3}^{(j)})^{\mathrm{H}})^{-1}(\tilde{\mathbf{H}}_{2,1}^{(j)}))^{\mathrm{H}}. \tag{6.24}$$

The precoding matrices are then obtained from

$$\begin{aligned}\mathbf{U}_1^{(j)} &= \frac{1}{\sqrt{d}}\mathrm{EVD}(\mathbf{Y}^{(j)})\left[\boldsymbol{e}_1,\ldots,\boldsymbol{e}_d\right] \\ \mathbf{U}_3^{(j)} &= \frac{1}{\|((\tilde{\mathbf{H}}_{2,3}^{(j)})^{\mathrm{H}})^{-1}(\tilde{\mathbf{H}}_{2,1}^{(j)})^{\mathrm{H}}\mathbf{U}_1^{(j)}\|_{\mathrm{F}}}((\tilde{\mathbf{H}}_{2,3}^{(j)})^{\mathrm{H}})^{-1}(\tilde{\mathbf{H}}_{2,1}^{(j)})^{\mathrm{H}}\mathbf{U}_1^{(j)} \\ \mathbf{U}_2^{(j)} &= \frac{1}{\|((\tilde{\mathbf{H}}_{1,2}^{(j)})^{\mathrm{H}})^{-1}(\tilde{\mathbf{H}}_{1,3}^{(j)})^{\mathrm{H}}\mathbf{U}_3^{(j)}\|_{\mathrm{F}}}((\tilde{\mathbf{H}}_{1,2}^{(j)})^{\mathrm{H}})^{-1}(\tilde{\mathbf{H}}_{1,3}^{(j)})^{\mathrm{H}}\mathbf{U}_3^{(j)}.\end{aligned} \tag{6.25}$$

Note that only $\mathbf{U}_j^{(j)}$ is effectively used for the transmission due to the distributed precoding assumption.

In that case, we can give the following result on the DoF achieved.

Theorem 10. Using the 3-User IA scheme described above with distributed CSIT, the DoF achieved at user i is denoted by $\mathrm{DoF}_i^{\mathrm{DCSI}}$ and verifies

$$\mathrm{DoF}_i^{\mathrm{DCSI}} \geq d \min_{j \neq i} \min_{k,\ell,k\neq\ell} A_{k,\ell}^{(j)}. \tag{6.26}$$

Proof. The main idea of the proof is to consider only the rate achieved over the channel realizations which are "well enough" conditioned. Over these channel realizations, the precoding is robust enough to the errors in the CSIT. Due to the continuous distribution of the channel matrices, the probability of the "badly conditioned" channel realizations is small enough such that the loss due to removing these channel realizations can be made arbitrarily small.

We consider hereafter that $\forall k, \ell, j, A_{k,\ell}^{(j)} > 0$ since the result is otherwise trivial. We also consider without loss of generality the precoding at TX j. For a given $\varepsilon > 0$, we define the following channel subsets:

$$\mathcal{X}^{\varepsilon} \triangleq \{\tilde{\mathbf{H}} | \forall i, k,\ \lambda_{\min}(\tilde{\mathbf{H}}_{i,k}) \geq \varepsilon\} \tag{6.27}$$

$$\mathcal{Y}^{\varepsilon} \triangleq \{\tilde{\mathbf{H}} | \forall i \neq j,\ |\lambda_i(\mathbf{Y}^{\star}) - \lambda_j(\mathbf{Y}^{\star})| \geq \varepsilon\} \tag{6.28}$$

and

$$\mathcal{H}^{\varepsilon} \triangleq \mathcal{X}^{\varepsilon} \bigcap \mathcal{Y}^{\varepsilon}. \tag{6.29}$$

Since we aim at deriving a lower bound for the DoF (and the rate is nonnegative), we can consider only the rate achieved for the channel realizations belonging to $\mathcal{H}^{\varepsilon}$. From (6.13), we can then write

$$\begin{aligned} R_i \geq \mathrm{E}_{\mathcal{H}^{\varepsilon}} \left[\log_2 \left| \mathbf{I}_{d_i} + P (\mathbf{G}_i^{\star})^{\mathrm{H}} \mathbf{H}_{i,i}^{\mathrm{H}} \mathbf{U}_i^{(i)} (\mathbf{U}_i^{(i)})^{\mathrm{H}} \mathbf{H}_{i,i} (\mathbf{G}_i^{\star}) \right| \right] \\ - \mathrm{E}_{\mathcal{H}^{\varepsilon}} \left[\log_2 \left| \mathbf{I}_{d_i} + P \sum_{j=1, j\neq i}^{K} (\mathbf{G}_i^{\star})^{\mathrm{H}} \mathbf{H}_{i,j}^{\mathrm{H}} \mathbf{U}_j^{(j)} (\mathbf{U}_j^{(j)}) \mathbf{H}_{i,j} \mathbf{G}_i^{\star} \right| \right]. \end{aligned} \tag{6.30}$$

It can be easily seen from the continuous distribution of the channel matrices that $\forall \eta > 0, \exists \varepsilon > 0, \Pr(\mathcal{H}^{\varepsilon}) \geq 1 - \eta$. Hence, it follows that

$$\mathrm{E}_{\mathcal{H}^{\varepsilon}} \left[\log_2 \left| \mathbf{I}_{d_i} + P (\mathbf{G}_i^{\star})^{\mathrm{H}} \mathbf{H}_{i,i}^{\mathrm{H}} \mathbf{U}_i^{(i)} (\mathbf{U}_i^{(i)})^{\mathrm{H}} \mathbf{H}_{i,i} (\mathbf{G}_i^{\star}) \right| \right] \dot{\geq} (1 - \eta) d_i \log_2(P). \tag{6.31}$$

We now need to upper bound the second term of (6.30) which we denote by $\mathcal{J}_i^{\varepsilon}$. We can then proceed similarly to (6.16) to write

$$\mathcal{J}_i^{\varepsilon} \leq d_i \Bigg(\mathrm{E}_{\mathcal{H}^{\varepsilon}} \Bigg[\log_2 \left(1 + \sum_{j=1, j\neq i}^{K} \|\mathbf{H}_{i,j}\|_{\mathrm{F}}^2 \right) \Bigg] + \log_2 \left(1 + P \sum_{j=1, j\neq i}^{K} \mathrm{E}_{\mathcal{H}^{\varepsilon}} \left[\|\boldsymbol{\Delta}\mathbf{U}_j^{(j)}\|_{\mathrm{F}}^2 \right] \right) \Bigg) \tag{6.32}$$

$$\dot{\leq} \, d_i \Bigg(\log_2 \left(1 + P \sum_{j=1, j\neq i}^{K} \mathrm{E}_{\mathcal{H}^{\varepsilon}} \left[\|\boldsymbol{\Delta}\mathbf{U}_j^{(j)}\|_{\mathrm{F}}^2 \right] \right) \Bigg). \tag{6.33}$$

It remains then only to compute $\mathrm{E}_{\mathcal{H}^{\varepsilon}}[\|\boldsymbol{\Delta}\mathbf{U}_j^{(j)}\|_{\mathrm{F}}^2]$. Le us now consider the error due to the imperfect CSIT at TX j on one of the matrix inversion required to compute $\mathbf{Y}^{(j)}$ in (6.24). We start by introducing $\boldsymbol{\Delta}_{i,k}^{(j)}$ to represent the error done in computing the channel inverse:

$$\boldsymbol{\Delta}_{i,k}^{(j)} \triangleq \frac{1}{\sigma_{i,k}^{(j)}} \left(\left((\tilde{\mathbf{H}}_{i,k}^{(j)})^{\mathrm{H}} \right)^{-1} - ((\tilde{\mathbf{H}}_{i,k})^{\mathrm{H}})^{-1} \right), \qquad \forall i,k. \tag{6.34}$$

Using the resolvent equality recalled in Lemma 2 in Appendix .1, we can write

$$\boldsymbol{\Delta}_{i,k}^{(j)} = -((\tilde{\mathbf{H}}_{i,k})^{\mathrm{H}})^{-1} (\mathbf{N}_{i,k}^{(j)})^{\mathrm{H}} ((\tilde{\mathbf{H}}_{i,k})^{\mathrm{H}})^{-1} + \sigma_{i,k}^{(j)} \boldsymbol{\Theta}_{i,k}^{(j)}, \qquad \forall i,k, \tag{6.35}$$

where we have defined

$$\boldsymbol{\Theta}_{i,k}^{(j)} \triangleq ((\tilde{\mathbf{H}}_{i,k}^{(j)})^{\mathrm{H}})^{-1} (\mathbf{N}_{i,k}^{(j)}))^{\mathrm{H}} ((\tilde{\mathbf{H}}_{i,k})^{\mathrm{H}})^{-1} (\mathbf{N}_{i,k}^{(j)})^{\mathrm{H}} (\tilde{\mathbf{H}}_{i,k})^{\mathrm{H}})^{-1}, \qquad \forall i,k. \tag{6.36}$$

We can then use the properties of the Frobenius norm to obtain the upper bound

$$\|\boldsymbol{\Delta}_{i,k}^{(j)}\|_{\mathrm{F}} \leq \|\mathbf{N}_{i,k}^{(j)}\|_{\mathrm{F}} \|\tilde{\mathbf{H}}_{i,k}^{-1}\|_{\mathrm{F}}^2 + \sigma_{i,k}^{(j)} \|\boldsymbol{\Theta}_{i,k}^{(j)}\|_{\mathrm{F}}. \tag{6.37}$$

Taking the expectation, we have then

$$\mathrm{E}_{\mathcal{H}^{\varepsilon}}[\|\boldsymbol{\Delta}_{i,k}^{(j)}\|_{\mathrm{F}}^2] \leq \mathrm{E}_{\mathcal{H}^{\varepsilon}} \left[\left(\|\mathbf{N}_{i,k}^{(j)}\|_{\mathrm{F}} \|\tilde{\mathbf{H}}_{i,k}^{-1}\|_{\mathrm{F}}^2 + \sigma_{i,k}^{(j)} \|\boldsymbol{\Theta}_{i,k}^{(j)}\|_{\mathrm{F}} \right)^2 \right] \tag{6.38}$$

The expectation in (6.38) exists and is finite because $\mathbf{H} \in \mathcal{H}^{\varepsilon}$ such that the channel matrix $\tilde{\mathbf{H}}_{i,k}$ (and $\tilde{\mathbf{H}}_{i,k}^{(j)}$) has its eigenvalues bounded away from zero. We have therefore obtained

$$\mathrm{E}_{\mathcal{H}^{\varepsilon}}[\|\boldsymbol{\Delta}_{i,k}^{(j)}\|_{\mathrm{F}}^2] \dot{\leq} 1. \tag{6.39}$$

We can then write

$$\mathbf{Y}^{(j)} = (\tilde{\mathbf{H}}_{3,1}^{-1} + \sigma_{3,1}^{(j)}\mathbf{\Delta}_{3,1}^{(j)})^{\mathrm{H}}(\tilde{\mathbf{H}}_{3,2} + \sigma_{3,2}^{(j)}\mathbf{N}_{3,2}^{(j)})^{\mathrm{H}}(\tilde{\mathbf{H}}_{1,2}^{-1} + \sigma_{1,2}^{(j)}\mathbf{\Delta}_{1,2}^{(j)})^{\mathrm{H}} \\ (\tilde{\mathbf{H}}_{1,3} + \sigma_{1,3}^{(j)}\mathbf{N}_{1,3}^{(j)})^{\mathrm{H}}(\tilde{\mathbf{H}}_{2,3}^{-1} + \sigma_{2,3}^{(j)}\mathbf{\Delta}_{2,3}^{(j)})^{\mathrm{H}}(\tilde{\mathbf{H}}_{2,1} + \sigma_{2,1}^{(j)}\mathbf{N}_{2,1}^{(j)})^{\mathrm{H}}. \tag{6.40}$$

The relation (6.39) holds for every matrix inversion in (6.31) such that putting all the errors terms together, we can write from (6.40) that

$$\mathrm{E}_{\mathcal{H}^{\varepsilon}}[\|\mathbf{Y}^{(j)} - \mathbf{Y}^{\star}\|_{\mathrm{F}}^2] \dot{\leq} \max_{\ell \neq k}(\sigma_{\ell k}^{(j)})^2. \tag{6.41}$$

Since $\mathbf{H} \in \mathcal{H}^{\varepsilon}$, all the eigenvalues of $\mathbf{Y}^{\star}$ (and $\mathbf{Y}^{(j)}$) are different and the matrices $\mathbf{Y}^{\star}$ and $\mathbf{Y}^{(j)}$ are diagonalizable. Let $\mathbf{Y}^{\star} = \mathbf{V}^{\star}\mathbf{\Lambda}(\mathbf{V}^{\star})^{\mathrm{H}}$, with $\mathbf{V}^{\star} \in \mathbb{C}^{M\times M}$ and $\mathbf{\Lambda}^{\star} \in \mathbb{C}^{M\times M}$, and $\mathbf{Y}^{(j)} = \mathbf{V}^{(j)}\mathbf{\Lambda}^{(j)}(\mathbf{V}^{(j)})^{\mathrm{H}}$, with $\mathbf{V}^{(j)} \in \mathbb{C}^{M\times M}$ and $\mathbf{\Lambda}^{(j)} \in \mathbb{C}^{M\times M}$, be the spectral decomposition of $\mathbf{Y}^{\star}$ and $\mathbf{Y}^{(j)}$, respectively. Applying Theorem 2.1 from [109] to $\mathbf{Y}^{\star}$ and $\mathbf{Y}^{(j)}$ and taking the expectation we can show that for some constant $\gamma^{(j)} > 0$ independent of the SNR P,

$$\mathrm{E}_{\mathcal{H}^{\varepsilon}}[\|\mathbf{V}^{(j)} - \mathbf{V}^{\star}\|_{\mathrm{F}}^2] \leq \gamma^{(j)}\mathrm{E}_{\mathcal{H}^{\varepsilon}}[\|\mathbf{Y}^{(j)} - \mathbf{Y}^{\star}\|_{\mathrm{F}}^2] \tag{6.42}$$

$$\dot{\leq} \max_{\ell \neq k}(\sigma_{\ell,k}^{(j)})^2 \tag{6.43}$$

$$\doteq P^{-\min_{\ell\neq k} A_{\ell,k}^{(j)}}. \tag{6.44}$$

Let us denote by $\bar{\mathbf{V}}^{(j)}$ and $\bar{\mathbf{V}}^{\star}$ the matrices made of the first d columns of $\mathbf{V}^{(j)}$ and $\mathbf{V}^{\star}$, respectively. The precoding scheme is such that $\mathbf{U}_1^{\star} = \bar{\mathbf{V}}^{\star}$ and $\mathbf{U}_1^{(1)} = \bar{\mathbf{V}}^{(1)}$. Hence,

$$\mathrm{E}_{\mathcal{H}^{\varepsilon}}[\|\mathbf{\Delta}\mathbf{U}_1^{(1)}\|_{\mathrm{F}}^2] \dot{\leq} P^{-\min_{\ell\neq k} A_{\ell,k}^{(1)}}. \tag{6.45}$$

The relation (6.45) is easily extended to the other precoders $\mathbf{U}_2^{(2)}$ and $\mathbf{U}_3^{(3)}$ to obtain that

$$\sum_{j=1,j\neq i}^{3} \mathrm{E}_{\mathcal{H}^{\varepsilon}}[\|\mathbf{\Delta}\mathbf{U}_j^{(j)}\|_{\mathrm{F}}^2] \dot{\leq} \sum_{j=1,j\neq i}^{3} P^{-\min_{\ell\neq k} A_{\ell,k}^{(j)}} \tag{6.46}$$

$$\dot{\leq} P^{-\min_{j\neq i}\min_{\ell\neq k} A_{\ell,k}^{(j)}}. \tag{6.47}$$

Coming back to (6.30), this gives

$$R_i \dot{\geq} d_i \left((1-\eta)\log_2(P) - \log_2(1 + PP^{-\min_{j\neq i}\min_{\ell\neq k} A_{\ell,k}^{(j)}} \right) \tag{6.48}$$

$$\dot{\geq} d_i \left(\min_{j\neq i}\min_{\ell\neq k} A_{\ell,k}^{(j)} - \eta \right) \log_2(P). \tag{6.49}$$

Choosing η arbitrarily small concludes the proof. □

We have shown that for the 3-user IA closed-form alignment scheme, the achieved DoF is larger than the worst accuracy of the channel estimates across the TXs. Note that this lower bound is in fact conjectured to be tight.

Interestingly, the lower bound at RX j is limited by the accuracy of the estimates relative to the channels of all the *other* RXs. This result is in strong contrast with the centralized setting where the DoF of user i depends *solely* on the accuracy with which the channel matrices from the TXs to RX i are feedback.

6.3 Simulations

In this section, we validate by Monte-Carlo simulations the results in the 3-user square IC channel studied in Subsection 6.2.2. We consider $M = N = 4$ and $d = 2$ and we average the performance over 10000 realizations of a Rayleigh fading channel. We consider the distributed CSIT configuration described in Section 6.1. The quantization error is modeled using (6.6) with $(\sigma_{i,k}^{(j)})^2 = 2^{-B_{i,k}^{(j)}/(N_i M_j - 1)}$ and $\mathbf{N}_{i,k}^{(j)}$ having its elements i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$. We choose the CSIT scaling coefficients as

$$\forall (i,k,j) \in \{1,2,3\}^3 \setminus \{(3,2,2),(3,2,3)\},\ A_{i,k}^{(j)} = 1, A_{3,2}^{(2)} = 0.5,\ A_{3,2}^{(3)} = 0. \tag{6.50}$$

Following Theorem 10, we have for the CSIT configuration described in (6.50) that $\mathrm{DoF}_1 \geq 0$, $\mathrm{DoF}_2 \geq 0$, and $\mathrm{DoF}_3 \geq 0.5d = 1$. The average rate achieved is shown for each user in Fig. 6.2. For comparison, we have also simulated the average rate per-user achieved based on perfect CSIT and with distributed CSIT when the CSIT scaling coefficients are set equal to 1 for every TX ($\forall i,k,j, A_{i,k}^{(j)} = 1$). It can then be verified that having all CSIT scaling coefficients equal to one allows to achieve the maximal DoF.

With the CSIT configuration described in (6.50), the slope of the rate of user 3 decreases as the SNR increases, revealing a very slow convergence

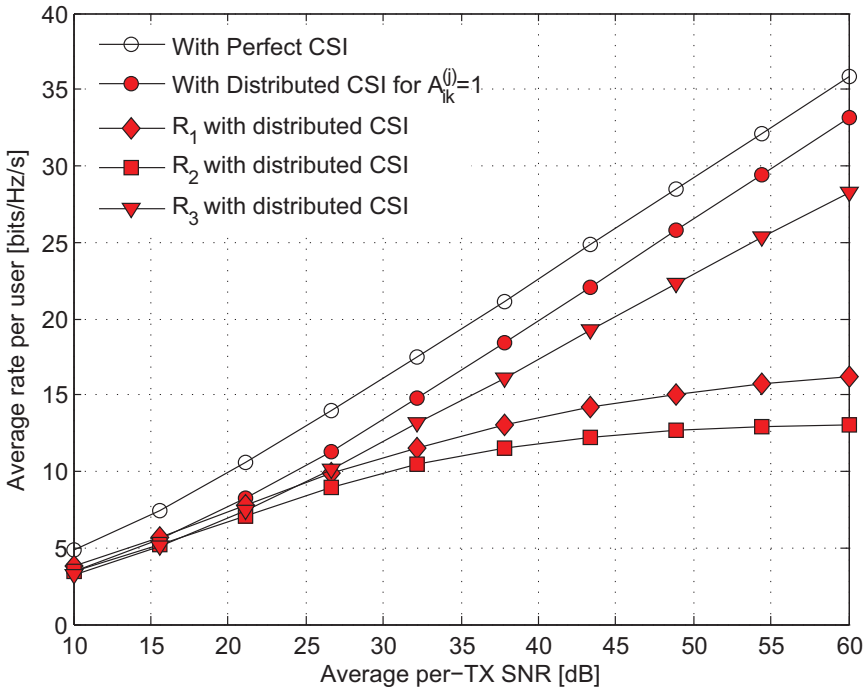


Figure 6.2: Average rate per user in the square setting $M = N = 4$ with $d = 2$ for the CSIT scaling coefficients given in (6.50).

to the DoF. This makes it difficult to accurately observe the DoF achieved. Yet, it can be seen that having only $A^{(3)}_{3,2}$ equal to zero leads already to the saturation of the rates of RX 1 and RX 2 (i.e., their DoF is equal to 0), which tends to confirm our conjecture.

6.4 Conclusion and Outlook

We have shown that the DoF that can be achieved with distributed CSIT in the static 3-user MIMO square IC is at least equal to the DoF achieved with the worst accuracy taken across the TXs *and* across the interfering links. We conjecture further that this represents exactly the DoF achieved. This result is in strong contrast with the *centralized CSIT* configuration for which

it was shown that the DoF achieved at RX i is solely limited by the quality of its *own* feedback. This property has already been observed in Chapter 4 when joint precoding is used, and appears as the cost of distributed CSIT.

This particular antenna configuration has been considered both because it is believed to be a simple, yet practically relevant configuration, and because the knowledge of a closed-form precoding formula is necessary for our analysis. In fact, our approach is expected to easily extend to numerous scenarios where a closed form expression exists for the IA precoding, under the condition that the precoding scheme is "robust" enough to the quantization errors, e.g., it consists of matrix inversions or matrix multiplications where the matrices have their elements distributed according to a continuous distribution. This in particular the case of the original time-alignment IA scheme from [37, 110]. Hence, our results can be trivially extended to this setting.

Obtaining the DoF achieved with an iterative IA algorithm like the min-leakage algorithm or the max-SINR algorithm [53, 86] is a challenging open problem which will be investigated in subsequent works. As a prerequisite step, it requires deriving some basic properties of the IA algorithm, such as convergence properties, which have remained out of reach until now. Furthermore, it will be shown in Chapter 8 that heterogeneous antenna configurations can be exploited to achieve IA even when some of the TXs do not have any CSIT. Such behavior should be taken into account when analyzing the feedback requirements and complicate further the analysis of iterative IA algorithms.

Part III

Spatial Allocation of Feedback Resources in Cooperative Networks

Chapter 7

Distance-based CSIT Allocation for Network MIMO

In the previous part, we have dealt with the design of the precoders for given CSIT configurations. We now approach the problem of the TX cooperation with distributed CSIT under a different angle: We consider from now on that the precoding schemes are fixed, and we discuss what is the necessary CSIT at each TX to achieve required performances. More specifically, our goal is to find the most parcimonious CSIT allocation strategy which allows to achieve close to the performance obtained with perfect CSIT. Finding such a CSIT allocation would open the door to potential reductions of the CSIT requirements at the cost of acceptable (negligible?) performance degradations. It will become clear in this part that significant CSIT reductions are in fact possible at virtually no cost.

7.1 System Setting and Problem Statement

7.1.1 Distributed CSI at the TXs

The joint precoder is implemented distributively at the TXs with each TX relying solely on its own estimate of the channel matrix in order to compute its transmit coefficient, without any exchange of information with the other TXs. To model the imperfect CSIT, the channel estimate at each TX is assumed to be obtained from a limited rate digital feedback scheme. Consequently, we introduce the following definitions.

Definition 1 (Distributed Finite-Rate CSIT). We represent a *CSIT allocation* by the collections of matrices $\{\mathbf{B}^{(j)}\}_{j=1}^{K}$ where $\mathbf{B}^{(j)} \in \mathbb{R}_{+}^{K\times K}$ denotes the CSIT allocation at TX j. Hence, TX j receives the multi-user channel estimate $\hat{\mathbf{H}}^{(j)}$ defined from

$$\hat{H}_{ki}^{(j)} = H_{ki} + \rho_{ki}\sqrt{2^{-B_{ki}^{(j)}}}\Delta H_{ki}^{(j)}, \qquad \forall i,k \tag{7.1}$$

where $\Delta H_{ki}^{(j)} \sim \mathcal{N}(0,1)$ and the $\Delta H_{ki}^{(j)}$ are mutually independent and independent of the channel. The $\Delta H_{ki}^{(j)}$ are then regrouped to form the matrix $\mathbf{\Delta H}^{(j)}$.

Remark 11. *The reasons for modeling the imperfect CSIT via* (7.1) *are as follows. First, it is well known from rate-distortion theory that the minimal distortion when quantizing a standard Gaussian source using B bits is equal to 2^{-B} [6, Theorem* 13.3.3*] while this distortion value is also achieved up to a multiplicative constant using for example the Lloyd algorithm [111] or even scalar quantization. Thus, the decay in 2^{-B} as the number of quantizations bits increases, represents a reasonable model.*

Furthermore, only the asymptotic behavior exponentially in the SNR is of interest in this work such that the distribution of the CSIT error does not matter here. We have chosen a Gaussian distribution but any other distribution fulfilling some mild regularity constraints could be chosen. □

It is now clear from the previous chapters that the pre-log factor of the number of feedback bits represents an appropriate measure at high SNR of the amount of CSIT required. Thus, we define the *size* of a CSIT allocation as follows.

Definition 2 (Size of a CSIT allocation). The size $\mathrm{s}(\bullet)$ of a CSIT allocation $\mathbf{B}^{(j)}$ at TX j is defined as

$$\mathrm{s}(\mathbf{B}^{(j)}) \triangleq \lim_{P\to\infty} \frac{\sum_{i,k} B_{ki}^{(j)}}{\log_2(P)} \tag{7.2}$$

such that the total size of a CSIT allocation $\{\mathbf{B}^{(j)}\}_{j=1}^{K}$ is

$$\mathrm{s}(\{\mathbf{B}^{(j)}\}_{j=1}^{K}) \triangleq \sum_{j=1}^{K} \mathrm{s}(\mathbf{B}^{(j)}) \tag{7.3}$$

$$= \lim_{P\to\infty} \frac{\sum_{i,j,k} B_{ki}^{(j)}}{\log_2(P)}. \tag{7.4}$$

Remark 12. *We consider here a digital quantization of the channel vectors but the results can be easily translated to a setting where analog feedback is used since digital quantization is simply used as a way to quantify the variance of the CSIT errors [19, 112]. Furthermore, only CSI requirements at the TXs are investigated, and different scenarios can be envisaged for the sharing of the channel estimates (e.g., direct broadcasting from the RXs to all the TXs, sharing through a backhaul, ...) [113].* □

7.1.2 Distributed Precoding

We focus here on the CSI dissemination problem under a conventional precoding framework. Hence, we assume that the sub-optimal ZF precoder is used. Based on its own channel estimate $\hat{\mathbf{H}}^{(j)}$, TX j computes then the ZF beamforming vector $\boldsymbol{t}_i^{(j)}$ to transmit symbol s_i such that

$$\boldsymbol{t}_i^{(j)} \triangleq \sqrt{P}\frac{\left(\hat{\mathbf{H}}^{(j)}\right)^{-1}\boldsymbol{e}_i}{\|\left(\hat{\mathbf{H}}^{(j)}\right)^{-1}\boldsymbol{e}_i\|}, \qquad \forall i \in \{1,\ldots,K\}. \tag{7.5}$$

Although a given TX j may compute the whole precoding matrix $\mathbf{T}^{(j)}$, only the j-th row is of practical interest. Indeed, TX j transmits solely $x_j = \boldsymbol{e}_j^{\mathrm{H}}\mathbf{T}^{(j)}\boldsymbol{s}$. The effective multi-user precoder $\mathbf{T}$ verifies then

$$\mathbf{x} = \mathbf{T}\boldsymbol{s} = \begin{bmatrix} \boldsymbol{e}_1^{\mathrm{H}}\mathbf{T}^{(1)} \\ \boldsymbol{e}_2^{\mathrm{H}}\mathbf{T}^{(2)} \\ \vdots \\ \boldsymbol{e}_K^{\mathrm{H}}\mathbf{T}^{(K)} \end{bmatrix}\boldsymbol{s}. \tag{7.6}$$

Finally, we denote by $\mathbf{T}^{\star} = [\boldsymbol{t}_1^{\star},\ldots,\boldsymbol{t}_K^{\star}]$ the precoder obtained with perfect CSI at all TXs. It is hence equal to

$$\boldsymbol{t}_i^{\star} \triangleq \sqrt{P}\frac{(\mathbf{H})^{-1}\boldsymbol{e}_i}{\|(\mathbf{H})^{-1}\boldsymbol{e}_i\|}, \qquad \forall i \in \{1,\ldots,K\}. \tag{7.7}$$

7.1.3 Optimization of the CSIT Allocation

Optimizing directly the allocation of the number of bits at finite SNR represents a challenging problem which gives little hope for analytical results. Instead, we will try to identify one CSIT allocation solution achieving the same DoF as under the fully shared CSIT setting.

Definition 3. We define the set of DoF-achieving CSIT allocations $\mathbb{B}_{\mathrm{DoF}}(\mathbf{\Gamma})$ as

$$\mathbb{B}_{\mathrm{DoF}}(\mathbf{\Gamma}) \triangleq \{\{\mathbf{B}^{(j)}\}_{j=1}^{K} | \forall i, \mathrm{DoF}_i(\{\mathbf{B}^{(j)}\}_{j=1}^{K}, \mathbf{\Gamma}) = 1\}. \tag{7.8}$$

Hence, an interesting problem consists in finding the minimal CSIT allocation (where minimality refers to the size in Definition 2) which achieves the maximal generalized DoF at every user:

$$\text{minimize}\, \mathrm{s}\left(\{\mathbf{B}^{(j)}\}_{j=1}^{K}\right), \text{ subject to } \{\mathbf{B}^{(j)}\}_{j=1}^{K} \in \mathbb{B}_{\mathrm{DoF}}(\mathbf{\Gamma}). \tag{7.9}$$

We focus here on an "achievability" result, by exhibiting a CSIT allocation that achieves the maximal DoF while having a much lower size than the conventional (uniform) CSIT allocation. Furthermore, the proposed "achievable scheme" will prove to have very good properties which distinguish it from other solutions in the literature (e.g., clustering) and makes it practically interesting. The problem of finding a minimal-size allocation policy while guaranteeing full DoF (i.e. DoF equal to the perfect CSIT case) is an interesting problem, but an extreme challenging one, which, to our best knowledge, remains open.

7.2 Preliminary Results

7.2.1 A Sufficient Criterion

As a preliminary step, we derive a simple sufficient criterion on the precoder for achieving the maximal DoF.

Proposition 12. The maximal DoF is achieved by using the precoder $\mathbf{T}$ if the CSIT allocation $\{\mathbf{B}^{(j)}\}_{j=1}^{K}$ is such that

$$\mathrm{E}\left[\|\mathbf{T} - \mathbf{T}^{\star}\|_{\mathrm{F}}^{2}\right] \doteq P^{0} \tag{7.10}$$

where we have used the exponential equality $f(P) \doteq P^{b}$ as in Chapter 6 to denote $\lim_{P\to\infty} \frac{\log(f(P))}{\log(P)} = b$ [1] and $\mathbf{T}^{\star}$ has been defined previously as the precoder based on perfect CSIT.

Proof. We start by defining the rate difference $\Delta_{\mathrm{R},i}$ between the rate of user i based on perfect CSI and the rate achieved with limited feedback. As

in [15, 19], we can then write

$$\Delta_{\mathrm{R},i} \triangleq \mathrm{E}\big[\log_2(1+|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_i^{\star}|^2)\big] - \mathrm{E}\left[\log_2\left(1+\frac{|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_i|^2}{1+\sum_{j\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_j|^2}\right)\right] \tag{7.11}$$

$$= \mathrm{E}\left[\log_2\left(\frac{1+|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_i^{\star}|^2}{1+\sum_{j=1}^{K}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_j|^2}\right)\right] + \mathrm{E}\left[\log_2\left(1+\sum_{j\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_j|^2\right)\right] \tag{7.12}$$

$$\doteq \mathrm{E}\left[\log_2\left(1+\sum_{j\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{t}_j|^2\right)\right]. \tag{7.13}$$

We further obtain

$$\Delta_{\mathrm{R},i} \dot{\leq} \mathrm{E}\left[\log_2\left(1+\sum_{j\neq i}|\boldsymbol{h}_i^{\mathrm{H}}(\boldsymbol{t}_j^{\star}+(\boldsymbol{t}_j-\boldsymbol{t}_j^{\star}))|^2\right)\right] \tag{7.14}$$

$$= \mathrm{E}\left[\log_2\left(1+\sum_{j\neq i}|\boldsymbol{h}_i^{\mathrm{H}}(\boldsymbol{t}_j-\boldsymbol{t}_j^{\star})|^2\right)\right]. \tag{7.15}$$

We can then easily upper-bound (7.15) to write

$$\Delta_{\mathrm{R},i} \dot{\leq} \mathrm{E}\left[\log_2\left(1+\|\boldsymbol{h}_i\|^2\sum_{j\neq i}\|\boldsymbol{t}_j-\boldsymbol{t}_j^{\star}\|^2\right)\right] \tag{7.16}$$

$$= \mathrm{E}\left[\log_2\left(1+\|\mathbf{T}-\mathbf{T}^{\star}\|_{\mathrm{F}}^2\right)\right] + \mathrm{E}\left[\log_2\left(1+\|\boldsymbol{h}_i\|^2\right)\right] \tag{7.17}$$

$$\dot{\leq} \log_2\left(\mathrm{E}\left[\|\mathbf{T}-\mathbf{T}^{\star}\|_{\mathrm{F}}^2\right]\right). \tag{7.18}$$

The maximal DoF is achieved if the rate difference $\Delta_{\mathrm{R},i}/\log_2(P)$ tends to zero as the SNR increases, which is the result of the proposition. □

The condition obtained above is very intuitive and will be used in the remaining of this work. However, Proposition 12 does not solve the main question, which is to determine what kind of CSIT allocation allows to achieve (7.10). This question is central to this work and will be tackled in Section 7.3.

7.2.2 The Conventional CSIT Allocation is DoF Achieving

The term "conventional" hereby corresponds to conveying to each TX the CSI relative to the full multi-user channel, enabling all the TXs to do the

same processing and compute a common precoder, such that $\mathbf{T}^{(j)} = \hat{\mathbf{T}}, \forall j$. Hence, the condition of Proposition 12 can be rewritten as

$$\mathrm{E}\left[\left\|\mathbf{T}^{(j)} - \mathbf{T}^{\star}\right\|_{\mathrm{F}}^{2}\right] \doteq P^{0}, \qquad \forall j \in \{1, \ldots, K\}. \tag{7.19}$$

Based on this, the following result is obtained.

Proposition 13. The following "conventional" CSIT allocation $\{\mathbf{B}^{\mathrm{conv},(j)}\}_{j=1}^{K}$ defined as

$$\{\mathbf{B}^{\mathrm{conv},(j)}\}_{ki} \triangleq [\lceil \log_2(P\rho_{ki}^2) \rceil]^{+}, \qquad \forall k, i, j \tag{7.20}$$

$$= [\lceil \{\mathbf{\Gamma}\}_{ki} \log_2(P) \rceil]^{+} \tag{7.21}$$

is DoF achieving, i.e., $\{\mathbf{B}^{\mathrm{conv},(j)}\}_{j=1}^{K} \in \mathbb{B}_{\mathrm{DoF}}$.

Proof. We consider without loss of generality the precoding at TX j. Using the CSIT allocation in (7.21), it holds that

$$\rho_{ki}\sqrt{2^{-B_{ki}^{(j)}}} = \sqrt{\frac{1}{P}} \tag{7.22}$$

such that $\mathbf{H}^{(j)} = \mathbf{H} + \sqrt{\frac{1}{P}}\mathbf{\Delta H}^{(j)}$. We start by recalling the well known resolvent equality. Using the resolvent equality given in Lemma 2 in Appendix .1 two times successively, we can then write

$$\left(\hat{\mathbf{H}}^{(j)}\right)^{-1} - \mathbf{H}^{-1} \tag{7.23}$$

$$= \left(\mathbf{H} + \sqrt{P^{-1}}\mathbf{\Delta H}^{(j)}\right)^{-1} - \mathbf{H}^{-1} \tag{7.24}$$

$$= \mathbf{H}^{-1}(-\sqrt{P^{-1}}\mathbf{\Delta H}^{(j)}) \left(\mathbf{H} + \sqrt{P^{-1}}\mathbf{\Delta H}^{(j)}\right)^{-1} \tag{7.25}$$

$$= \mathbf{H}^{-1}(-\sqrt{P^{-1}}\mathbf{\Delta H}^{(j)})\mathbf{H}^{-1} + \mathbf{H}^{-1}(-\sqrt{P^{-1}}\mathbf{\Delta H}^{(j)}) \left((\hat{\mathbf{H}}^{(j)})^{-1} - \mathbf{H}^{-1}\right) \tag{7.26}$$

$$= -\sqrt{P^{-1}}\mathbf{H}^{-1}\mathbf{\Delta H}^{(j)}\mathbf{H}^{-1} + P^{-1}\mathbf{H}^{-1}\mathbf{\Delta H}^{(j)}\mathbf{H}^{-1}\mathbf{\Delta H}^{(j)}(\hat{\mathbf{H}}^{(j)})^{-1}. \tag{7.27}$$

We can then use the triangular inequality to obtain the upperbound

$$\| \left(\hat{\mathbf{H}}^{(j)}\right)^{-1} - \mathbf{H}^{-1}\|_{\mathrm{F}} \tag{7.28}$$

$$\leq \sqrt{P^{-1}}\|\mathbf{H}^{-1}\|_{\mathrm{F}}^{2}\|\mathbf{\Delta H}^{(j)}\|_{\mathrm{F}} + P^{-1}\|(\hat{\mathbf{H}}^{(j)})^{-1}\|_{\mathrm{F}}\|\mathbf{H}^{-1}\|_{\mathrm{F}}^{2}\|\mathbf{\Delta H}^{(j)}\|_{\mathrm{F}}^{2} \tag{7.29}$$

$$\doteq \sqrt{P^{-1}}\|\mathbf{H}^{-1}\|_{\mathrm{F}}^{2}\|\mathbf{\Delta H}^{(j)}\|_{\mathrm{F}}. \tag{7.30}$$

Because of our assumption of outer-diagonal decay, the channel is well conditioned and all the expectations in (7.30) are finite. Taking the square and the expectation, we obtain then

$$\mathrm{E}[\| \left(\hat{\mathbf{H}}^{(j)}\right)^{-1} - \mathbf{H}^{-1}\|_{\mathrm{F}}^2] \dot{\leq} P^{-1}. \tag{7.31}$$

After normalization, it follows easily from (7.31) that the sufficient condition in Proposition 12 is fulfilled, which concludes the proof. □

This CSIT allocation provides to each TX the K channel vectors relative to the K RXs. This means that each TX requires a number of channel estimates growing unbounded with K. This represents a serious issue in large/dense networks which prompts designers, in practice, to restrict cooperation to small cooperations clusters.

7.2.3 CSIT Allocation with Distributed Precoding

We now turn our attention to the derivation of a more efficient CSIT allocation strategy. A crucial observation is that each TX does not need to compute accurately the *full* precoder. Indeed, the sufficient criterion (7.10) can be written in the distributed CSI setting as

$$\mathrm{E}\left[\left\|\boldsymbol{e}_j^{\mathrm{H}}(\mathbf{T}^{(j)} - \mathbf{T}^{\star})\right\|^2\right] \doteq P^0, \qquad \forall j \in \{1, \ldots, K\}. \tag{7.32}$$

Intuition has it that a channel coefficient relative to a TX/RX pair which interferes little with TX/RX j has little impact on the jth precoding row and hence does not need to be known accurately at TX j. What follows is a quantitative assessment of this intuition.

7.3 Distance-Based CSIT Allocation

We consider in this section a two dimensional network consisting of K TX/RX pairs. The geometry of the network is abstracted as follows. Let us assume that there is a given map from the user's indices $i \in \{1, \ldots, K\}$ to some coordinates $(x_i, y_i) \in \mathbb{R}^2$. TX/RX pair i is then assumed to be located at the position (x_i, y_i). We denote by $\mathrm{d}(i,k)$ the Euclidian distance over $\mathbb{R}^2$. The pathloss between TX i and RX k is then given by

$$\rho_{k,i}^2 = (\mu^2)^{\mathrm{d}(i,k)}, \qquad \forall i, k \in \{1, \ldots, K\} \tag{7.33}$$

for a given $\mu^2 < 1$. The interference level matrix $\mathbf{\Gamma}$ defined in (3.16) is then equal to

$$\{\mathbf{\Gamma}\}_{ki} = \frac{\log((\mu^2)^{\mathrm{d}(i,k)}P)}{\log(P)}, \qquad \forall k, i \tag{7.34}$$

$$= 1 + (\gamma - 1)\,\mathrm{d}(k, i) \tag{7.35}$$

where we have introduced

$$\gamma \triangleq \frac{\log(\mu^2 P)}{\log(P)}. \tag{7.36}$$

7.3.1 Distance-based CSIT Allocation

With the notations set above, the conventional CSIT allocation given in (7.21) can be rewritten as

$$\{\mathbf{B}^{\mathrm{conv},(j)}\}_{ki} = \lceil [1 + (\gamma - 1)\,\mathrm{d}(k, i)]^+ \log_2(P) \rceil, \qquad \forall k, i, j. \tag{7.37}$$

We can now state our first main result.

Theorem 11. Let us define the CSIT allocation $\{\mathbf{B}^{\mathrm{dist},(j)}\}_{j=1}^{K}$ such that

$$\{\mathbf{B}^{\mathrm{dist},(j)}\}_{ki} \triangleq \lceil [1 + (\gamma - 1)(\mathrm{d}(j, k) + \mathrm{d}(k, i))]^+ \log_2(P) \rceil, \qquad \forall k, i, j. \tag{7.38}$$

Then $\mathbf{B}^{\mathrm{dist}} \in \mathbb{B}_{\mathrm{DoF}}$.

Proof. Let us focus without loss of generality on the CSIT allocation at TX j. Following the sufficient condition in Proposition 12, we want to find a CSIT allocation such that

$$\mathrm{E}[|\boldsymbol{e}_j^{\mathrm{H}}\mathbf{H}^{-1}\boldsymbol{e}_i - \boldsymbol{e}_j^{\mathrm{H}}\left(\hat{\mathbf{H}}^{(j)}\right)^{-1}\boldsymbol{e}_i|^2] \dot{\leq} P^{-1}, \qquad \forall i \in \{1, \ldots, K\}. \tag{7.39}$$

Indeed, once (7.39) is fulfilled, the the sufficient condition in Proposition 12 follows easily.

Let us first define the matrix $\mathbf{D} \triangleq \mathrm{diag}(\mathbf{H})$. We also define for this proof the normalized channel coefficient $\tilde{H}_{ki} \triangleq H_{ki}/\rho_{ki}$. It then holds that $\tilde{H}_{ki} \sim \mathcal{N}_{\mathbb{C}}(0, 1)$. Because of the outer-diagonal attenuation by $\mu^{\min_{i,j}\mathrm{d}(i,j)}$, we have that

$$\lim_{n\to\infty} (\mathbf{I}_K - \mathbf{D}^{-1}\mathbf{H})^n = \mathbf{0}_K. \tag{7.40}$$

Hence, we can proceed by writing the channel inverse using the series von Neumann as

$$\boldsymbol{e}_j^{\mathrm{H}}\mathbf{H}^{-1}\boldsymbol{e}_i = \boldsymbol{e}_j^{\mathrm{H}}\sum_{n=0}^{\infty}(\mathbf{D}^{-1}(\mathbf{D}-\mathbf{H}))^n\mathbf{D}^{-1}\boldsymbol{e}_i, \qquad \forall i,j \tag{7.41}$$

$$= \sum_{n=0}^{\infty} C_n^{ji} \tag{7.42}$$

where we have defined

$$C_n^{ji} \triangleq \boldsymbol{e}_j^{\mathrm{H}}(\mathbf{D}^{-1}(\mathbf{D}-\mathbf{H}))^n\mathbf{D}^{-1}\boldsymbol{e}_i, \qquad \forall i,j,n. \tag{7.43}$$

It can be seen from (7.43) that C_n^{ji} verifies

$$\mathrm{E}[|C_n^{ji}|^2] \dot{\leq} (\mu^{2\min_{i,j}\mathrm{d}(i,j)})^n \tag{7.44}$$

$$= P^{n\left(\frac{\log(\mu^{2\min_{i,j}\mathrm{d}(i,j)}P)-\log(P)}{\log(P)}\right)} \tag{7.45}$$

$$= P^{(\gamma_{\min}-1)n} \tag{7.46}$$

where we have defined

$$\gamma_{\min} \triangleq \frac{\log(\mu^{2\min_{i,j}\mathrm{d}(i,j)}P)}{\log(P)}. \tag{7.47}$$

Hence, the infinite summation can be truncated to a finite summation up to $n_0 \triangleq \lceil 1/(1-\gamma_{\min})\rceil$ without impacting the DoF. We further introduce

$$\mathbf{D}^{(j)} \triangleq \mathrm{diag}(\hat{\mathbf{H}}^{(j)}), \qquad \forall j, \tag{7.48}$$

$$C_n^{ji,(j)} \triangleq \boldsymbol{e}_j^{\mathrm{H}}((\mathbf{D}^{(j)})^{-1}(\mathbf{D}^{(j)}-\hat{\mathbf{H}}^{(j)}))^n(\mathbf{D}^{(j)})^{-1}\boldsymbol{e}_i, \qquad \forall i,j,n. \tag{7.49}$$

We can then write

$$\mathrm{E}[|\boldsymbol{e}_j^{\mathrm{H}}\mathbf{H}^{-1}\boldsymbol{e}_i - \boldsymbol{e}_j^{\mathrm{H}}\left(\hat{\mathbf{H}}^{(j)}\right)^{-1}\boldsymbol{e}_i|^2] \doteq \mathrm{E}\left[\left|\sum_{n=1}^{n_0} C_n^{ji} - C_n^{ji,(j)}\right|^2\right] \tag{7.50}$$

$$\leq \mathrm{E}\left[\left(\sum_{n=1}^{n_0}|C_n^{ji} - C_n^{ji,(j)}|\right)^2\right] \tag{7.51}$$

$$\dot{\leq} \sum_{n=1}^{n_0}\mathrm{E}[|C_n^{ji} - C_n^{ji,(j)}|^2] \tag{7.52}$$

where we have used iteratively that $(a+b)^2 \leq 2(a^2+b^2), \forall a, b \in \mathbb{R}^2$ to obtain the last inequality (and the multiplicative constants could be removed because of the exponential inequality). We now look for a CSIT allocation $\mathbf{B}^{(j)}$ ensuring that

$$\mathrm{E}[|C_n^{ji} - C_n^{ji,(j)}|^2] \dot{\leq} P^{-1}, \qquad \forall i, j, n. \tag{7.53}$$

In particular, let us write the first coefficients

$$\mathrm{E}[|C_0^{ji} - C_0^{ji,(j)}|^2] = \mathrm{E}\left[\left|\frac{\boldsymbol{e}_j^{\mathrm{H}}\boldsymbol{e}_i}{H_{ii}} - \frac{\boldsymbol{e}_j^{\mathrm{H}}\boldsymbol{e}_i}{\hat{H}_{ii}^{(j)}}\right|^2\right] \tag{7.54}$$

$$= \mathrm{E}\left[\left|\frac{\sqrt{2^{-B_{ii}^{(j)}}}\Delta H_{ii}^{(j)}}{H_{ii}\hat{H}_{ii}^{(j)}}\right|^2\right] \delta_{ji} \tag{7.55}$$

$$\doteq 2^{-B_{ii}^{(j)}} \delta_{ji}. \tag{7.56}$$

Thus, we set

$$B_{jj}^{(j)} = \lceil \log_2(P) \rceil. \tag{7.57}$$

This ensure to fulfill (7.53) for $n = 0$. The error done over H_{jj} becomes then negligible in terms of DoF (i.e., in terms of exponential equality). For $n \geq 1$, it holds that $C_n^{jj} = 0$ such that we assume that $i \neq j$ in the following,

$$\mathrm{E}[|C_1^{ji} - C_1^{ji,(j)}|^2] = \mathrm{E}\left[\left|\frac{\{\mathbf{D}-\mathbf{H}\}_{j,i}}{H_{jj}H_{ii}} - \frac{\{\mathbf{D}^{(j)}-\hat{\mathbf{H}}^{(j)}\}_{j,i}}{\hat{H}_{jj}^{(j)}\hat{H}_{ii}^{(j)}}\right|^2\right] \tag{7.58}$$

$$\doteq \mathrm{E}\left[\left|\frac{\hat{H}_{i,i}^{(j)}\tilde{H}_{j,i} - H_{i,i}\tilde{H}_{j,i}^{(j)}}{H_{jj}H_{ii}\hat{H}_{ii}^{(j)}}\right|^2\right] (\mu^2)^{\mathrm{d}(i,j)} \tag{7.59}$$

$$\doteq (2^{-B_{ii}^{(j)}} + 2^{-B_{ji}^{(j)}}) P^{(\gamma-1)\,\mathrm{d}(i,j)} \tag{7.60}$$

Setting

$$B_{ii}^{(j)} = \lceil [1 + (\gamma-1)\,\mathrm{d}(i,j)]^+ \log_2(P) \rceil, \qquad \forall i \neq j \tag{7.61}$$

$$B_{ji}^{(j)} = \lceil [1 + (\gamma-1)\,\mathrm{d}(i,j)]^+ \log_2(P) \rceil, \qquad \forall i \neq j \tag{7.62}$$

ensures to fulfill (7.53) for all streams i for $n = 1$. Going further, we consider then C_2^{ji},

$$\mathrm{E}[|C_2^{ji} - C_2^{ji,(j)}|^2] \tag{7.63}$$

$$= \mathrm{E}[|\boldsymbol{e}_j^{\mathrm{H}}(\mathbf{D}^{-1}(\mathbf{D}-\mathbf{H}))^2\mathbf{D}^{-1}\boldsymbol{e}_i - \boldsymbol{e}_j^{\mathrm{H}}((\mathbf{D}^{(j)})^{-1}(\mathbf{D}^{(j)} - \hat{\mathbf{H}}^{(j)}))^2(\mathbf{D}^{(j)})^{-1}\boldsymbol{e}_i|^2] \tag{7.64}$$

$$= \mathrm{E}[|\sum_{k=1}^{K} \boldsymbol{e}_j^{\mathrm{H}}\mathbf{D}^{-2}(\mathbf{D}-\mathbf{H})\boldsymbol{e}_k\boldsymbol{e}_k^{\mathrm{H}}(\mathbf{D}-\mathbf{H})\mathbf{D}^{-1}\boldsymbol{e}_i - \boldsymbol{e}_j^{\mathrm{H}}(\mathbf{D}^{(j)})^{-2}(\mathbf{D}^{(j)} - \hat{\mathbf{H}}^{(j)})\boldsymbol{e}_k \cdot \boldsymbol{e}_k^{\mathrm{H}}(\mathbf{D}^{(j)} - \hat{\mathbf{H}}^{(j)})(\mathbf{D}^{(j)})^{-1}\boldsymbol{e}_i|^2] \tag{7.65}$$

$$= \mathrm{E}[|\sum_{k=1,k\neq i,k\neq j}^{K} \frac{1}{H_{jj}^2 H_{ii}}\tilde{H}_{jk}\tilde{H}_{ki} - \frac{1}{(\hat{H}_{jj}^{(j)})^2\hat{H}_{ii}^{(j)}}\tilde{H}_{jk}^{(j)}\tilde{H}_{ki}^{(j)}|^2](\mu^2)^{\mathrm{d}(j,k)+\mathrm{d}(k,i)} \tag{7.66}$$

$$\dot{\leq} \mathrm{E}[|\sum_{k=1,k\neq i,k\neq j}^{K} \tilde{H}_{jk}\tilde{H}_{ki} - \tilde{H}_{jk}^{(j)}\tilde{H}_{ki}^{(j)}|^2]P^{(\gamma-1)(\mathrm{d}(j,k)+\mathrm{d}(k,i))} \tag{7.67}$$

$$\doteq \sum_{k=1,k\neq i,k\neq j}^{K} (2^{-B_{jk}^{(j)}} + 2^{-B_{ki}^{(j)}})P^{(\gamma-1)(\mathrm{d}(j,k)+\mathrm{d}(k,i))} \tag{7.68}$$

$$\doteq \sum_{k=1,k\neq i,k\neq j}^{K} 2^{-B_{ki}^{(j)}}P^{(\gamma-1)(\mathrm{d}(j,k)+\mathrm{d}(k,i))} \tag{7.69}$$

where we could remove $2^{-B_{jk}^{(j)}}$ because of (7.62). Setting

$$B_{ki}^{(j)} = \lceil [1 + (\gamma-1)(\mathrm{d}(j,k) + \mathrm{d}(k,i))]^+ \log_2(P)\rceil, \qquad \forall k \neq i, k \neq j \tag{7.70}$$

allows to fulfill (7.53) for $n = 2$. Going to arbitrary value of n, we write

$$\mathrm{E}[|C_n^{ji} - C_n^{ji,(j)}|^2] = \mathrm{E}\left[\left|\sum_{k_1\neq j}^{K}\sum_{k_2\neq k_1}^{K}\cdots\sum_{k_{n-1}\neq k_{n-2},k_{n-1}\neq i}^{K}\left(\frac{\tilde{H}_{j,k_1}\tilde{H}_{k_1,k_2}\ldots,\tilde{H}_{k_{n-1},i}}{H_{jj}^n H_{ii}} - \frac{\tilde{H}_{j,k_1}^{(j)}\tilde{H}_{k_1,k_2}^{(j)}\ldots,\tilde{H}_{k_{n-1},i}^{(j)}}{(\hat{H}_{jj}^{(j)})^n\hat{H}_{ii}^{(j)}}\right)\right|^2\right](\mu^2)^{\mathrm{d}(j,k_1)+\mathrm{d}(k_1,k_2)+\ldots+\mathrm{d}(k_{n-1},i)}. \tag{7.71}$$

Yet, it follows from the distance properties of d (triangular inequality, positivity) that the exponents in μ will always be larger than the ones obtained

for $C_0^{ji}, C_1^{ji}, C_2^{ji}$. Hence, setting the $B_{ki}^{(j)}$ as in (7.57), (7.61),(7.62) and (7.70) ensures that (7.53) is fulfilled for all n and for all i. Inserting this result in (7.52) concludes the proof. □

Letting γ tend to one inside (7.38), the network geometry becomes homogeneous with all the links having the same variance, asymptotically in the SNR. There is then no attenuation of the interference due to the pathloss and the distance-based CSIT allocation converges as expected to the conventional CSIT allocation given in (7.37). More generally, the distance-based CSIT allocation exploits the fact that two users being further away that a given distance do not impact the design of precoding coefficients for each other. This can be related to the the space of infinite (either polynomially or exponentially) decaying matrices being closed under inversion [114, 115].

Remark 13. *The result easily extends to the case of multiple TX/RX pairs located at the same position. This models then a TX with multiple-antennas serving multiple single-antenna RXs.* □

7.3.2 Scaling Properties of the Distance-based CSIT Allocation

We consider in this section the scaling behaviour as the number of TX/RX pairs increases. It is then differentiated in the literature between so-called *dense* networks and *extended* networks [116, 117]. In the first model, the size of the network remains constant and the density (number of TX/RX pairs/m^2) increases, while in the second the density of the network remains constant as the number of TX/RX pairs increases. Our analysis being on large networks, we consider the extended model and we assume that the density of TX/RX pairs remains constant, i.e., that the size of the network increases with the number of TX/RX pairs.

The critical question that we want to tackle here is to determine how the size of the CSIT allocation increases with the number of TX/RX pairs. Indeed, this scaling represents an important figure-of-merit to evaluate the feasibility of large cooperation areas.

Corollary 5. In the distance-based CSIT allocation, TX j does not need to receive any CSIT relative to the ith RX if

$$\mathrm{d}(i,j) > d_0 \tag{7.72}$$

with d_0 defined as

$$d_0 \triangleq \frac{1}{1-\gamma}. \tag{7.73}$$

In particular, the size of the distance-based CSIT allocation at any TX remains bounded in the extended model as the number of TX/RX pairs K increases:

$$\mathrm{s}(\mathbf{B}^{\mathrm{dist},(j)}) = O(1) \text{ as } K \text{ grows}, \qquad \forall j \in \{1, \ldots, K\}. \tag{7.74}$$

Proof. This result follows directly by observing that $\{\mathbf{B}^{\mathrm{dist},(j)}\}_{ki}$ is equal to zero if $\mathrm{d}(k,j) > d_0$. From the assumption of finite density, there are only a finite number of RXs fulfilling this condition such that the size of the CSIT allocation at TX j remains bounded as K increases. □

This result is in stark contrast with the conventional CSIT allocation where the size $\mathrm{s}(\mathbf{B}^{\mathrm{conv},(j)})$ scales linearly with K. This corollary confirms the intuition that a CSIT-exchange restricted to a finite neighborhood is sufficient to achieve global coordination.

We have considered so far a scenario with global sharing of the user's data symbols. The result above leads to ask ourselves whether the sharing of the user's data could also be reduced to a local neighborhood without reduction of the DoF achieved.

Corollary 6. Let us denote by $\mathcal{K}_j$ the set of user's data symbols being shared to TX j. It is sufficient for achieving the maximal DoF that $s_i \in \mathcal{K}_j$ if

$$\mathrm{d}(i,j) < d_0. \tag{7.75}$$

In the extended model, it follows that

$$\forall j, \ |\mathcal{K}_j| = O(1), \text{ as } K \text{ grows}. \tag{7.76}$$

Proof. From the proof of Theorem 11, it is known that

$$\mathrm{E}[|\boldsymbol{e}_j^{\mathrm{H}} \mathbf{H}^{-1} \boldsymbol{e}_i|^2] \dot{\leq} (\mu^2)^{\mathrm{d}(i,j)}. \tag{7.77}$$

Hence, if (7.72) is fulfilled, $(\mu^2)^{\mathrm{d}(i,j)} P = P^{1+(\gamma-1)\,\mathrm{d}(i,j)} \dot{\leq} 1$ and setting $\{\mathbf{T}\}_{ji} = 0$ does not impact the DoF. Thus, it is not necessary in terms of DoF that TX j participates to the transmission of the ith stream, and hence it is not necessary that TX j receives data symbol s_i. □

The operational meaning of the above result is that no exchange of information (CSI or data symbol) is necessary between two TXs i and k if they verify that $\mathrm{d}(i,k) > d_0$. Intuitively, d_0 is the size of the neighborhood inside which the cooperation should occur. Altogether, the distance-based CSIT allocation along with the matching users data sharing provides an attractive alternative to clustering. The difference being that the hard-boundaries of the cluster are replaced by a smooth decrease of the level of cooperation.

7.4 Simulations

We verify now by simulations that the maximal DoF per user is achieved by the distance based CSIT allocation. At the same time, we compare the distance based CSIT allocation to the CSI disseminations commonly used, i.e., uniform CSIT allocation and clustering.

We consider a channel model with $\gamma = 0.6$ and we use Monte-Carlo averaging over 1000 channel realizations. We study first a network with a regular geometry where $K = 36$ TX/RX pairs are placed at the integer values inside a square of dimensions 6×6. We show in Fig. 7.1 the average rate achieved with different CSIT allocation policies. Specifically, the distance-based CSIT allocation in (7.38) is compared to two alternative CSIT allocations, being the uniform CSIT allocation $\{\mathbf{B}^{\mathrm{unif},(j)}\}_{j=1}^{K}$ where the bits are allocated *uniformly* to the TXs, and the clustering one $\{\mathbf{B}^{\mathrm{cluster},(j)}\}_{j=1}^{K}$ in which (non-overlapping) *regular clustering of size* 4 is used. Both CSIT allocations are chosen to have the same size as the distance-based one:

$$\mathrm{s}\left(\{\mathbf{B}^{\mathrm{unif},(j)}\}_{j=1}^{K}\right) = \mathrm{s}\left(\{\mathbf{B}^{\mathrm{cluster},(j)}\}_{j=1}^{K}\right) = \mathrm{s}\left(\{\mathbf{B}^{(\mathrm{dist},(j)}\}_{j=1}^{K}\right). \quad (7.78)$$

With these parameters, the size of the distance based CSIT allocation is only equal to 6.5% of the size of the conventional CSIT allocation. Nevertheless, it can be observed to achieve the maximal generalized DoF while the clustering solution has a smaller slope, yet larger than the uniform CSIT allocation. The distance-based CSIT allocation suffers from a strong negative rate offset. This offset is a consequence of our analysis being limited to the high SNR regime. Indeed, using ZF with many users is very inefficient at intermediate SNR, particularly in a network with strong pathloss. Furthermore, the number of TX/RX pairs K which is here relatively large, has not been taken into account. Hence, this strong negative rate offset can be easily reduced by optimizing the precoding scheme and the CSIT allocation at finite SNR. The key element being that the distance-based CSIT allocation does not present the usual limitations of clustering, i.e., edge-interference and large scaling behaviour as the size of the cluster increases.

Finally, we show in Fig. 7.2 the average rate per user in a network made of $K = 15$ TX/RX pairs being located uniformly at random over a square of dimensions 6×6 for $\gamma = 0.7$. We compare the average rate achieved with the distance-based CSIT allocation to the average rate obtained if we use the distance-based allocation given in Theorem 11, but with $1+\alpha(\gamma-1)(\mathrm{d}(j,k)+\mathrm{d}(k,i))$ for a given parameter $\alpha > 0$ instead of $1 + (\gamma - 1)(\mathrm{d}(j,k) + \mathrm{d}(k,i))$. This allows to observe the impact of reducing ($\alpha > 1$) or increasing ($\alpha < 1$) the CSIT compared to the distance-based CSIT allocation.

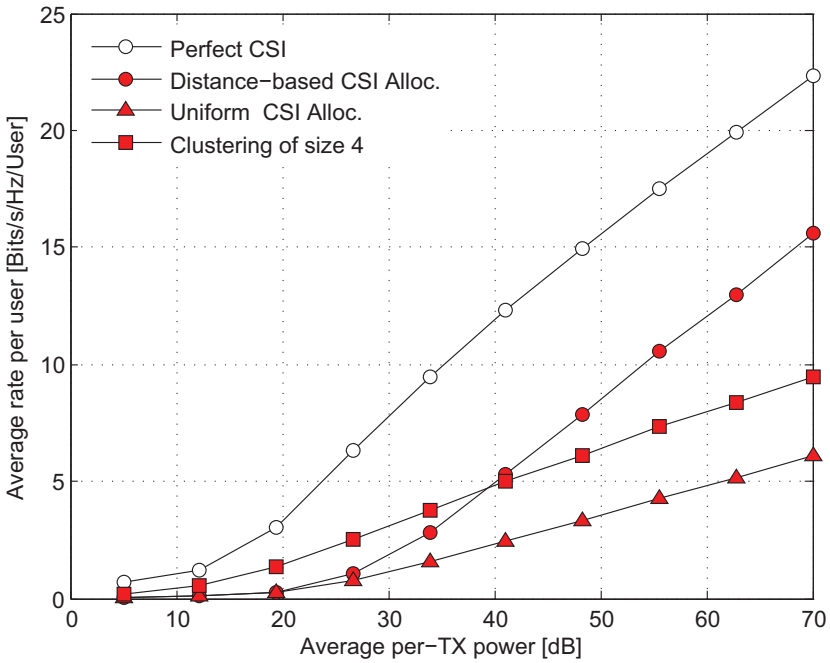


Figure 7.1: Average rate per user as a function of the SNR P for $K = 36$ and $\gamma = 0.6$. The TX/RX pairs are positioned at the integer values inside a square of dimensions 6×6. The 3 limited feedback CSIT allocations used have the same size which is equal to 6.5% of the size of the conventional CSIT allocation in (7.37).

We can observe that reducing the CSIT allocation leads to reducing the slope, i.e., the DoF, while using more feedback bits leads to a vanishing rate offset. This is in agreement with our theoretical result that the distance-based CSIT allocation leads to a finite rate offset.

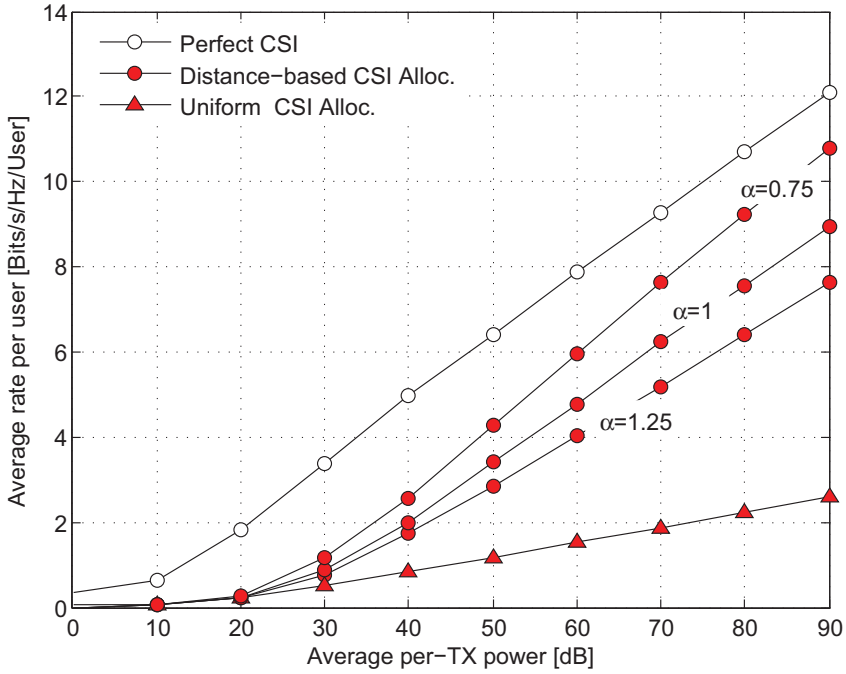


Figure 7.2: Average rate per user as a function of the SNR P for $K = 15$ and $\gamma = 0.7$. The TX/RX pairs are placed uniformly at random over a square of dimensions 6×6. The CSIT allocation with $\alpha = 1.25$, $\alpha = 1$, and $\alpha = 1.25$, use respectively 10%, 17%, and 30% of the number of bits relative to the conventional CSIT allocation.

7.5 Conclusion

We have discussed the problem of optimizing the CSIT dissemination in a network MIMO scenario. In particular, following a generalized DoF analysis,

we have exhibited a CSIT allocation which allows to achieve the optimal generalized DoF while restricting the cooperation to a local scale. This behavior is critical for the cooperation of a large number of TXs to be possible in a realistic network where the backhaul links are imperfect and of finite capacity. The proposed CSIT allocation can be seen as an alternative to clustering where the hard boundaries of the cluster are replaced by a smooth decrease of the cooperation strength.

In this chapter, we have then shown how optimizing the CSI dissemination allows to achieve the required performance with a CSIT dissemination strategy being much more parcimonious than the conventional one. Interestingly, this new approach allows to overcome limitations of the TX cooperation which seemed fundamental but were in fact only a consequence of the particular CSIT dissemination strategy chosen. Studying the CSIT dissemination in other application scenarios will be the topic of future research and is expected to lead to further savings and new interesting behaviours.

Chapter 8

Interference Alignment with Incomplete CSIT

In the previous chapter, we have shown how to exploit the pathloss attenuation to reduce the CSIT requirements. The approach could not easily be extended to IA because of the different kind of coordination required between the TXs (the dimensions where the interference are "aligned" have to be jointly chosen). However, we will now show that there is another dimension that can be exploited in the case of MIMO IA.

Let us consider as toy example a K-user IC with all TXs and all RXs having respectively M and N antennas. It is shown in [50] that IA is feasible if and only if $M + N \geq K + 1$ for single-stream transmissions. This result is obtained with the assumption of perfect CSIT at all RXs. However, if $M = 1$ and $N = K$, it is clear that no CSI is necessary at the TXs since the RXs have enough antennas to ZF the $K - 1$ dimensions of interference. Similarly, if $M = K$ and $N = 1$, each TX can apply conventional ZF to emit no interference to the other RXs. It is then not necessary to provide each TX with the CSI relative to the full multi-user channel but solely with the "local" CSI relative to this particular TX.

Those particular examples highlight the fact that IA can be achieved in some cases without the requirement for each TX to receive the CSI relative to the full multi-user channel. This is very interesting practically as it means that it is possible to reduce the amount of CSI shared in the backhaul at the cost of *no performance reduction*. We investigate in the following what is the minimal – in the sense of a metric which will be introduced in the following – CSIT allocation which is necessary in order to achieve IA.

Practically, this means revisiting the so-called "feasibility problem" for

MIMO IA, which consists in determining for given antenna configurations whether IA is feasible or not [49,50,118,119], under the scope of incomplete CSIT at the TXs.

8.1 Incomplete CSIT Configuration and Problem Statement

We consider in this section a particular case of distributed CSIT which we call *incomplete CSIT*. In this model, a TX has either perfect knowledge of a channel coefficient or no information at all on that element. We represent the CSIT structure at TX j by the *CSIT matrix* $\mathbf{F}^{(j)} \in \{0,1\}^{N_{\text{tot}} \times M_{\text{tot}}}$ such that $\{\mathbf{F}^{(j)}\}_{ik} = 1$ if $\{\mathbf{H}^{\text{H}}\}_{ik}$ is known at TX j, and 0 otherwise. Denoting by $\hat{\mathbf{H}}^{(j)}$ the available CSI at TX j, we obtain

$$(\hat{\mathbf{H}}^{(j)})^{\text{H}} = \mathbf{F}^{(j)} \odot \mathbf{H}^{\text{H}} \tag{8.1}$$

with $\odot$ denoting the element-wise (or Hadamart) product. We define the CSIT allocation $\mathcal{F}$ as the set of CSI representations available at all TXs:

$$\mathcal{F} = \{\mathbf{F}^{(j)} | \mathbf{F}^{(j)} \in \{0,1\}^{N_{\text{tot}} \times M_{\text{tot}}}, j = 1, \ldots, K\} \tag{8.2}$$

and we define the space $\mathbb{F}$ containing all the possible CSIT allocations. We can then define the *size* of an incomplete CSIT allocation as follows.

Definition 4. The size of a CSIT allocation $\mathcal{F}$, denoted by $\text{s}(\mathcal{F})$, is equal to the overall number of complex channel coefficients fed back to the TXs. Thus,

$$\text{s}(\mathcal{F}) \triangleq \sum_{j=1}^{K} \|\mathbf{F}^{(j)}\|_{\text{F}}^2. \tag{8.3}$$

Remark 14. *This size definition can be linked to the size used in Chapter 7. Indeed, considering a quantization of the channel elements with* $\log_2(P)$ *bits per-element, the size defined in* (8.3) *is then equal to the total pre-log factor of the number of CSI bits exchanged in the backhaul. In fact, the size in* (8.3) *can be seen as the size introduced in Definition 2 for the particular case where the number of quantization bits used for a channel element can only be* $\log_2(P)$ *or* 0. □

To check whether IA feasibility is preserved with a given CSIT allocation, we introduce the function f_{feas} which takes as argument a CSIT allocation $\mathcal{F}$ and an antenna configuration $\prod_{k=1}^{K}(N_k, M_k)$ and returns 1 if IA is feasible

with these parameters and 0 otherwise. Note that this means that there exists one algorithm achieving IA with this CSIT allocation but it does not precise the algorithm. We also define the set $\mathbb{F}_{\text{feas}}$ containing all the CSIT allocations for which IA is feasible. Hence,

$$\mathbb{F}_{\text{feas}} \triangleq \{\mathcal{F} | \mathcal{F} \in \mathbb{F}, f_{\text{feas}}\left(\mathcal{F}, \prod_{k=1}^{K} (N_k, M_k)\right) = 1\}. \tag{8.4}$$

Only the interfering channel matrices $\mathbf{H}_{ij}^{\text{H}}$ with $i \neq j$ are required to fulfill the IA constraints, and not the direct channel matrices $\mathbf{H}_{jj}^{\text{H}}$. Thus, from a DoF point of view, we can always skip the direct channel matrices $\mathbf{H}_{jj}^{\text{H}}$ in the feedback, which leads to the following definition.

Definition 5. A *complete* CSIT allocation, denoted by $\mathcal{F}_{\text{comp}}$, is defined by the knowledge of all the interfering channel matrices $\mathbf{H}_{ij}^{\text{H}}$ with $i \neq j$ at all TXs. Thus, the size of a complete CSIT allocation is

$$\text{s}\left(\mathcal{F}_{\text{comp}}\right) = K\left(N_{\text{tot}} M_{\text{tot}} - \sum_{i=1}^{K} N_i M_i\right). \tag{8.5}$$

A CSIT allocation with a size smaller than $\text{s}(\mathcal{F}_{\text{comp}})$ is said to be *strictly incomplete*.

At this stage, a natural question is to ask what is the most incomplete CSIT allocation which preserves the feasibility of IA, i.e., to find

$$\mathcal{F}_{\min} = \underset{\mathcal{F} \in \mathbb{F}_{\text{feas}}}{\text{argmin}}\, \text{s}(\mathcal{F}). \tag{8.6}$$

Note that we limit here our study to the IA feasible settings, i.e., such that $\mathcal{F}_{\text{comp}} \in \mathbb{F}_{\text{feas}}$.

8.2 Feasibility Results

8.2.1 Results from the Literature

We start by recalling some results from the literature on IA feasibility in a conventional IC with full CSIT sharing for the case of single stream transmission. In [49], the notion of *proper* antenna configurations in introduced. An IC is said to be proper if and only if the number of variables in the RX and TX beamformers involved in any set of IA constraints is larger than

the number of scalar equations. Following [49], let us denote by E_{ij} the IA equation for the jth stream at RX i (See Subsection 3.1.2)

$$\boldsymbol{g}_i^{\mathrm{H}}\mathbf{H}_{ij}^{\mathrm{H}}\boldsymbol{u}_j = 0 \qquad (E_{ij})$$

and by $\mathrm{var}(\mathrm{E}_{ij})$ the set of free variables involved in this equation. It holds then

$$|\,\mathrm{var}(\mathrm{E}_{ij})| = N_i - 1 + M_j - 1. \tag{8.7}$$

A system is said to be proper if and only if

$$\forall \mathcal{I} \subseteq \mathcal{J}, |\mathcal{I}| \leq |\bigcup_{(i,j)\in\mathcal{I}} \mathrm{var}(\mathrm{E}_{ij})| \tag{8.8}$$

where $\mathcal{J} \triangleq \{(i,j)|1 \leq i,j \leq K, i \neq j\}$ and $\mathcal{I}$ is an arbitrary subset of $\mathcal{J}$. In the homogeneous $(N,M)^K$ IC, this condition can be reduced to $M+N \geq K+1$. The following result has been later obtained in [50] and is restated here for convenience.

Theorem 12 ([50]). IA is feasible in the $\prod_{k=1}^{K}(N_k, M_k)$ IC if and only if the antenna configuration is proper, i.e., if (8.8) is verified.

Hence, we can use here the condition (8.8) to determine the feasibility of IA with complete CSIT sharing.

8.2.2 Tightly-feasible and Super-feasible Settings

Whether the total number of variables is strictly larger than the number of equations will be shown to impact significantly the CSIT needed. Hence, we introduce the following definitions.

Definition 6. An IC setting is called *tightly-feasible* if this IC is feasible and removing a single antenna at any TX or RX renders IA unfeasible. Equivalently, an IC is tightly-feasible if and only if it is *feasible* and

$$\sum_{i=1}^{K} N_i + M_i = K(K+1). \tag{8.9}$$

The characterization follows directly from (8.8) applied with the set $\mathcal{I} = \mathcal{J}$.

Definition 7. A feasible setting which does not verify the tightly-feasible condition is said to be *super-feasible*. Equivalently, a super-feasible setting is a feasible setting such that

$$\sum_{i=1}^{K} N_i + M_i > K(K+1). \tag{8.10}$$

8.2.3 New Formulation of the Feasibility Results

Condition (8.8) requires verifying a number of conditions increasing exponentially with the size of the network. As a preliminary step, we show that condition (8.8) can be simplified to obtain the following condition.

Theorem 13. IA is feasible in the $\prod_{k=1}^{K}(N_k, M_k)$ IC if and only if, for any TX subset $\mathcal{S}_{\mathrm{TX}}$ and any RX subset $\mathcal{S}_{\mathrm{RX}}$, it holds that

$$\forall \mathcal{S}_{\mathrm{TX}}, \mathcal{S}_{\mathrm{RX}} \subseteq \mathcal{K}, \quad \mathcal{N}_{\mathrm{var}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}) \geq \mathcal{N}_{\mathrm{eq}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}) \tag{8.11}$$

where $\mathcal{N}_{\mathrm{var}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}})$ and $\mathcal{N}_{\mathrm{eq}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}})$ are respectively the number of variables and the number of equations stemming from the subset of RXs $\mathcal{S}_{\mathrm{RX}}$ and the subset of TXs $\mathcal{S}_{\mathrm{TX}}$ and are mathematically defined as

$$\begin{aligned} \mathcal{N}_{\mathrm{var}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}) &\triangleq \sum_{i \in \mathcal{S}_{\mathrm{RX}}} N_i - 1 + \sum_{i \in \mathcal{S}_{\mathrm{TX}}} M_i - 1, \\ \mathcal{N}_{\mathrm{eq}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}) &\triangleq \sum_{k \in \mathcal{S}_{\mathrm{TX}}} \sum_{j \in \mathcal{S}_{\mathrm{RX}}, j \neq k} 1. \end{aligned} \tag{8.12}$$

Proof. For $\mathcal{I} \subseteq \mathcal{J} = \{(i,j) | 1 \leq i, j \leq K, i \neq j\}$, we define the sets

$$\mathcal{S}_{\mathrm{TX}}(\mathcal{I}) \triangleq \{j | \exists k', (k', j) \in \mathcal{I}\}, \quad \mathcal{S}_{\mathrm{RX}}(\mathcal{I}) \triangleq \{k | \exists j', (k, j') \in \mathcal{I}\}. \tag{8.13}$$

Hence, $\mathcal{S}_{\mathrm{RX}}(\mathcal{I})$ and $\mathcal{S}_{\mathrm{TX}}(\mathcal{I})$ contain respectively the set of RXs and the set of TXs appearing in at least one equation of the set of equations $\mathcal{I}$. With these notations, equation (8.8) can be rewritten as

$$\forall \mathcal{I} \subseteq \mathcal{J}, \quad |\mathcal{I}| \leq \sum_{k \in \mathcal{S}_{\mathrm{TX}}(\mathcal{I})} (M_k - 1) + \sum_{j \in \mathcal{S}_{\mathrm{RX}}(\mathcal{I})} (N_j - 1). \tag{8.14}$$

Adding equations to $\mathcal{I}$ without increasing $\mathcal{S}_{\mathrm{RX}}(\mathcal{I})$ or $\mathcal{S}_{\mathrm{TX}}(\mathcal{I})$ makes condition (8.14) tighter. Hence, it is only necessary to verify (8.14) for the sets of equations made of all the equations generated by the RXs in $\mathcal{S}_{\mathrm{RX}}(\mathcal{I})$ and the TXs in $\mathcal{S}_{\mathrm{TX}}(\mathcal{I})$. □

Remark 15. *Using Theorem* 1*, it is necessary to try all the subsets $\mathcal{I}$ included in $\mathcal{J}$. The set $\mathcal{J}$ can be seen to be of cardinal $K(K-1)$, such that there are $2^{K(K-1)}$ subsets to test. In contrast, Theorem* 2 *requires "only" to try all the sub-ICs. This means choosing a number of RXs between* 1 *and K and a number of TXs between* 1 *and K, which gives in total 2^{2K} possibilities.* □

We can note that the number of conditions to test in Theorem 13 is still exponential. However, the main interest of this result does not lie in the cardinality reduction. It comes from the fact that the TXs and the RXs can then be ordered to reduce the complexity from an exponential number of possibilities to a polynomial one. In fact, one contribution of this work is a simple and intuitive algorithm for testing the feasibility of IA with single streams. Since this algorithm is obtained after very simple modifications of our CSIT allocation algorithm (which will be described later on), it is not described here in more details. However, detailed description of the IA feasibility test with linear complexity can be found online in [120] along with the MATLAB code of the test.

The criterion (3.15) provides also an interesting insight into IA feasibility: The feasibility of IA in the full IC is verified by analyzing the feasibility of IA in all the *sub-ICs* included in the full IC.

Note that the sub-IC obtained after selection of the RXs inside $\mathcal{S}_{\mathrm{RX}}$ and the TXs inside $\mathcal{S}_{\mathrm{TX}}$ is not a conventional IC due to the fact that the TXs and the RXs are not necessarily *paired*. To model this scenario, we introduce the notion of *generalized IC*.

8.2.4 Generalized interference channels

We refer to an IC in which at least one TX or RX does not have its paired RX or TX included in the IC as a *generalized IC*. We represent this by writing a "*" instead of the number of antennas of the paired RX or TX. For example, an IC containing TXs 1, 2, and 3 but only RXs 1, 2, and 4 (with all the TXs and the RXs having two antennas) is denoted by $(2,2).(2,2).(*,2).(2,*)$. The IA feasibility criterion (3.15) is trivially extended to generalized ICs by verifying the condition for all the sets of TXs and RXs included in the generalized IC.

8.3 IA with Incomplete CSIT for Tightly-Feasible Channels

8.3.1 General Criterion

Parameterization of the CSIT allocation In order to write concisely our results, we need to introduce a last notation. With simple words, we define the matrix $\mathbf{F}_{\mathcal{S}_{\mathrm{RX}},\mathcal{S}_{\mathrm{TX}}}$, where $\mathcal{S}_{\mathrm{RX}}$ is a set of RXs and $\mathcal{S}_{\mathrm{TX}}$ a set of TXs, such that $\mathbf{F}_{\mathcal{S}_{\mathrm{RX}},\mathcal{S}_{\mathrm{TX}}} \odot \mathbf{H}$ contains all the channel coefficients relative to the

generalized sub-IC formed by the set of RXs $\mathcal{S}_{\mathrm{RX}}$ and the set of TXs $\mathcal{S}_{\mathrm{TX}}$, at the exception of the direct channel matrices $\mathbf{H}_{jj}, \forall j$.

Mathematically, this means that the matrix $\mathbf{F}_{\mathcal{S}_{\mathrm{RX}},\mathcal{S}_{\mathrm{TX}}}$ of size $N_{\mathrm{tot}} \times M_{\mathrm{tot}}$ has its only nonzero elements chosen to verify

$$\forall x \neq y, x \in \mathcal{S}_{\mathrm{RX}}, y \in \mathcal{S}_{\mathrm{TX}}, \quad \left(\mathbf{E}_{\mathrm{RX}}^{x}\right)^{\mathrm{T}} \mathbf{F}_{\mathcal{S}_{\mathrm{RX}},\mathcal{S}_{\mathrm{TX}}} \mathbf{E}_{\mathrm{TX}}^{y} = \left(\mathbf{E}_{\mathrm{RX}}^{x}\right)^{\mathrm{T}} \mathbf{1}_{N_{\mathrm{tot}} \times M_{\mathrm{tot}}} \mathbf{E}_{\mathrm{TX}}^{y} \tag{8.15}$$

with

$$\mathbf{E}_{\mathrm{TX}}^{n} \triangleq \left[\mathbf{0}_{M_n \times \sum_{k=1}^{n-1} M_k}, \mathbf{I}_{M_n}, \mathbf{0}_{M_n \times \sum_{k=n+1}^{K} M_k}\right]^{\mathrm{T}} \tag{8.16}$$

and the matrix $\mathbf{E}_{\mathrm{RX}}^{n}$ defined similarly, solely with N_i replacing M_i.

Main theorem We can now state one of our main results.

Theorem 14. In a tightly-feasible $\prod_{k=1}^{K}(N_k, M_k)$ IC, let us assume that there exists a tightly-feasible generalized sub-IC formed by the set of TXs $\mathcal{S}_{\mathrm{TX}}$ and the set of RXs $\mathcal{S}_{\mathrm{RX}}$, i.e.,

$$\mathcal{N}_{\mathrm{var}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}) = \mathcal{N}_{\mathrm{eq}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}). \tag{8.17}$$

Then the incomplete CSIT allocation $\mathcal{F} = \{\mathbf{F}^{(j)} | j = 1, \ldots, K\}$ preserves IA feasibility, i.e., $\mathcal{F} \in \mathbb{F}_{\mathrm{feas}}$ if

$$\forall j \in \mathcal{S}_{\mathrm{TX}}, \mathbf{F}^{(j)} = \mathbf{F}_{\mathcal{S}_{\mathrm{RX}},\mathcal{S}_{\mathrm{TX}}}, \qquad \forall j \notin \mathcal{S}_{\mathrm{TX}}, \mathbf{F}^{(j)} = \mathbf{F}_{\mathcal{K},\mathcal{K}} = \mathbf{1}_{N_{\mathrm{tot}} \times M_{\mathrm{tot}}}. \tag{8.18}$$

Proof. We have by assumption that

$$\mathcal{N}_{\mathrm{var}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}) = \mathcal{N}_{\mathrm{eq}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}). \tag{8.19}$$

This setting being tightly-feasible, it is possible to align interference inside this sub-IC. In the following we assume that the beamformers of the TXs and the RXs inside this sub-IC have been fixed and fulfill all IA constaints inside this sub-IC. I0n terms of IA feasibility, this is equivalent to replacing the TXs and the RXs inside this sub-IC by non-interfering single-antenna nodes. Indeed, these nodes do not have any ZF capabilities left but do not create any IA constraints among themselves.

We will now show that IA remains feasible in the IC obtained once these beamformers have been fixed. Since the initial IC was feasible, it holds for any subset of TX $\mathcal{S}'_{\mathrm{TX}}$ and any subset of RX $\mathcal{S}'_{\mathrm{RX}}$ that

$$\mathcal{N}_{\mathrm{var}}(\mathcal{S}'_{\mathrm{RX}}, \mathcal{S}'_{\mathrm{TX}}) \geq \mathcal{N}_{\mathrm{eq}}(\mathcal{S}'_{\mathrm{RX}}, \mathcal{S}'_{\mathrm{TX}}). \tag{8.20}$$

A fortiori, it holds for $\mathcal{S}_{\mathrm{RX}} \cap \mathcal{S}'_{\mathrm{RX}} = \emptyset$ and $\mathcal{S}_{\mathrm{TX}} \cap \mathcal{S}'_{\mathrm{TX}} = \emptyset$ that

$$\mathcal{N}_{\mathrm{var}}(\mathcal{S}_{\mathrm{RX}} \cup \mathcal{S}'_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}} \cup \mathcal{S}'_{\mathrm{TX}}) \geq \mathcal{N}_{\mathrm{eq}}(\mathcal{S}_{\mathrm{RX}} \cup \mathcal{S}'_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}} \cup \mathcal{S}'_{\mathrm{TX}}). \tag{8.21}$$

It follows from the definition of $\mathcal{N}_{\mathrm{var}}$ and $\mathcal{N}_{\mathrm{eq}}$ in (8.12), that $\forall \mathcal{A}, \mathcal{A}', \mathcal{B}, \mathcal{B}' \subset \mathcal{K}$, with $\mathcal{A}' \cap \mathcal{A} = \emptyset$ and $\mathcal{B}' \cap \mathcal{B} = \emptyset$, it holds that

$$\begin{aligned}\mathcal{N}_{\mathrm{var}}(\mathcal{A} \cup \mathcal{A}', \mathcal{B} \cup \mathcal{B}') &= \mathcal{N}_{\mathrm{var}}(\mathcal{A}, \mathcal{B}) + \mathcal{N}_{\mathrm{var}}(\mathcal{A}', \mathcal{B}') \\ \mathcal{N}_{\mathrm{eq}}(\mathcal{A} \cup \mathcal{A}', \mathcal{B} \cup \mathcal{B}') &= \mathcal{N}_{\mathrm{eq}}(\mathcal{A}, \mathcal{B}) + \mathcal{N}_{\mathrm{eq}}(\mathcal{A}', \mathcal{B}) + \mathcal{N}_{\mathrm{eq}}(\mathcal{A}, \mathcal{B}') + \mathcal{N}_{\mathrm{eq}}(\mathcal{A}', \mathcal{B}').\end{aligned} \tag{8.22}$$

Applying the relations in (8.22) to rewrite (8.21) and using also (8.20) gives

$$\mathcal{N}_{\mathrm{var}}(\mathcal{S}'_{\mathrm{RX}}, \mathcal{S}'_{\mathrm{TX}}) \geq \mathcal{N}_{\mathrm{eq}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}'_{\mathrm{TX}}) + \mathcal{N}_{\mathrm{eq}}(\mathcal{S}'_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}) + \mathcal{N}_{\mathrm{eq}}(\mathcal{S}'_{\mathrm{RX}}, \mathcal{S}'_{\mathrm{TX}}). \tag{8.23}$$

The relation (8.23) describes exactly all the feasibility conditions in the IC obtained once the beamformers inside the sub-IC containing the RXs in $\mathcal{S}_{\mathrm{RX}}$ and the TXs in $\mathcal{S}_{\mathrm{TX}}$ have been fixed. This shows that IA remains feasible and concludes the proof. □

Hence, the existence of a tightly-feasible generalized sub-IC strictly included in the full IC yields the existence of a strictly incomplete CSIT allocation preserving IA feasibility.

Remark 16. *From the iterative use of this property, we will show in the following that a reduced CSIT allocation ensuring IA feasibility is obtained if each TX receives the CSIT relative to the smallest tightly-feasible IC to which it belongs.* □

Applying Theorem 14 in an homogeneous setting gives a more pessimistic result.

Corollary 7. In the tightly-feasible setting $(N, M)^K$ (i.e., $M + N = K + 1$) with $M \neq 1$ and $M \neq K$, there exists no generalized tightly-feasible sub-IC which is strictly included in the full IC. Hence, the previous sufficient condition leads to no CSIT reduction.

Proof. The proof follows easily by evaluating (8.17) in an homogeneous setting and is omitted for brevity. □

If the full IC is tightly-feasible, a strictly smaller IC can be tightly-feasible only by exploiting the heterogeneity in the antenna configuration. Hence, the sufficient condition given in Theorem 14 cannot be fulfilled in any tightly-feasible homogeneous IC with $M \neq 1$ and $M \neq K$.

8.3.2 CSIT Allocation Algorithm

We now describe an incomplete CSIT allocation policy based on Theorem 14. The corresponding problem of designing an algorithm achieving IA based on the incomplete CSIT allocation is tackled in Subsection 8.3.3.

The CSIT allocation algorithm takes as input the antenna configuration $\prod_{k=1}^{K}(N_k, M_k)$ and returns as output the incomplete CSIT allocation

$$\mathcal{F} = \{\mathbf{F}^{(j)} | j = 1, \ldots, K\}$$

such that

$$\mathbf{F}^{(j)} = \mathbf{F}_{\mathcal{S}_{\mathrm{RX}}^{(j)}, \mathcal{S}_{\mathrm{TX}}^{(j)}}, \qquad \forall j. \tag{8.24}$$

Let us consider w.l.o.g. the problem of allocating the CSI to TX j.

Initialization: We first define an initial pair of sets $\mathcal{S} \triangleq (\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}})$ initialized such that

$$\mathcal{S} = (\emptyset, \{j\}). \tag{8.25}$$

The remaining TXs (without considering TX j) are ordered by increasing number of antennas, i.e., with the permutation σ_{TX} verifying

$$M_{\sigma_{\mathrm{TX}}(i)} \leq M_{\sigma_{\mathrm{TX}}(i+1)}, \qquad \forall i = \{1, \ldots, K-2\} \tag{8.26}$$

and symmetrically, the RXs are ordered by increasing number of antennas, i.e., with the permutation σ_{RX} verifying

$$N_{\sigma_{\mathrm{RX}}(i)} \leq N_{\sigma_{\mathrm{RX}}(i+1)}, \qquad \forall i = \{1, \ldots, K-1\}. \tag{8.27}$$

In case of equality, we order the TXs to ensure that

$$(M_{\sigma_{\mathrm{TX}}(i)} = M_{\sigma_{\mathrm{TX}}(i+1)}) \Rightarrow N_{\sigma_{\mathrm{TX}}(i)} \geq N_{\sigma_{\mathrm{TX}}(i+1)}, \qquad \forall i = \{1, \ldots, K-1\}. \tag{8.28}$$

Similarly, the RX ordering is modified to ensure that

$$(N_{\sigma_{\mathrm{RX}}(i)} = N_{\sigma_{\mathrm{RX}}(i+1)}) \Rightarrow M_{\sigma_{\mathrm{RX}}(i)} \geq M_{\sigma_{\mathrm{RX}}(i+1)}, \qquad \forall i = \{1, \ldots, K-1\}. \tag{8.29}$$

In case both the two TXs and their matched RXs have the same number of antennas, the RX ordering σ_{RX} is modified to ensure that the RXs are ordered in the opposite of the TXs, i.e.,

$$\begin{aligned} &\left(M_{\sigma_{\mathrm{TX}}(i)} = M_{\sigma_{\mathrm{TX}}(i+1)}, N_{\sigma_{\mathrm{TX}}(i)} = N_{\sigma_{\mathrm{TX}}(i+1)}\right) \\ &\qquad \Rightarrow \left(\sigma_{\mathrm{RX}}^{-1}(\sigma_{\mathrm{TX}}(i+1)) < \sigma_{\mathrm{RX}}^{-1}(\sigma_{\mathrm{TX}}(i))\right), \quad \forall i. \end{aligned} \tag{8.30}$$

Remark 17. *This ensures that selecting the TXs and the RXs respectively according to σ_{TX} and σ_{RX}, non-matched TXs and RXs will be selected first for equal number of antennas.* □

Update at step n: Let us assume that we are given the pair of sets $\mathcal{S} = (\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}})$.

1. If equation (8.17) is verified with the sets $\mathcal{S}_{\mathrm{RX}}$ and $\mathcal{S}_{\mathrm{TX}}$, the algorithm has reached its end. We set $\mathcal{S}_{\mathrm{RX}}^{(j)} = \mathcal{S}_{\mathrm{RX}}$, $\mathcal{S}_{\mathrm{TX}}^{(j)} = \mathcal{S}_{\mathrm{TX}}$ and
$$\mathbf{F}^{(j)} = \mathbf{F}_{\mathcal{S}_{\mathrm{RX}}^{(j)}, \mathcal{S}_{\mathrm{TX}}^{(j)}}. \tag{8.31}$$

2. If equation (8.17) does not hold, we verify whether adding the next RX adds more equations than variables, i.e.,
$$\begin{aligned}\mathcal{N}_{\mathrm{var}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}) - \mathcal{N}_{\mathrm{eq}}(\mathcal{S}_{\mathrm{RX}}, \mathcal{S}_{\mathrm{TX}}) \geq \mathcal{N}_{\mathrm{var}}(\{\mathcal{S}_{\mathrm{RX}}, \sigma_{\mathrm{RX}}(|\mathcal{S}_{\mathrm{RX}}|+1)\}, \mathcal{S}_{\mathrm{TX}}) \\ - \mathcal{N}_{\mathrm{eq}}(\{\mathcal{S}_{\mathrm{RX}}, \sigma_{\mathrm{RX}}(|\mathcal{S}_{\mathrm{RX}}|+1)\}, \mathcal{S}_{\mathrm{TX}})\end{aligned} \tag{8.32}$$
   - If (8.32) is verified, we set
   $$\mathcal{S}_{\mathrm{RX}} = \{\mathcal{S}_{\mathrm{RX}}, \sigma_{\mathrm{RX}}(|\mathcal{S}_{\mathrm{RX}}|+1)\} \tag{8.33}$$
   and we start over at step $n+1$.
   - If (8.32) is not verified, then
     - If $|\mathcal{S}_{\mathrm{TX}}| < K$, we increase the set of TXs as
     $$\mathcal{S}_{\mathrm{TX}} = \{\mathcal{S}_{\mathrm{TX}}, \sigma_{\mathrm{TX}}(|\mathcal{S}_{\mathrm{TX}}|+1)\} \tag{8.34}$$
     and we start over at step $n+1$.
     - If $|\mathcal{S}_{\mathrm{TX}}| = K$, then the algorithm has reached its end and we set $\mathcal{S}_{\mathrm{RX}}^{(j)} = \mathcal{S}_{\mathrm{RX}}$ and $\mathcal{S}_{\mathrm{TX}}^{(j)} = \mathcal{S}_{\mathrm{TX}}$ and
     $$\mathbf{F}^{(j)} = \mathbf{F}_{\mathcal{S}_{\mathrm{RX}}^{(j)}, \mathcal{S}_{\mathrm{TX}}^{(j)}}. \tag{8.35}$$

8.3.3 IA Algorithm for Incomplete CSIT Allocation

We consider now the CSIT allocation $\mathcal{F}$ to be given and we describe a novel IA algorithm which achieves IA using an adequate incomplete CSIT allocation. The algorithm runs in a distributed fashion at each TX. This IA algorithm, which we denote by f_{IA}, takes as input the antenna configuration,

the CSIT allocation policy, and the channel coefficients known at the TX, and returns the beamformer at TX j. Thus, we can write at TX j

$$\boldsymbol{t}_j = f_{\text{IA}}\Big(\prod_{k=1}^{K}(N_k, M_k), \mathcal{F}, \hat{\mathbf{H}}^{(j)}\Big). \tag{8.36}$$

IA algorithm for the effective channel We start by introducing an IA algorithm f_{eff} which will be a building block for our algorithm. It consists in running an IA algorithm over the *effective* channel obtained once a fraction of the TX beamformers have been fixed. Hence, taking as input the set containing the fixed beamformers $\mathcal{B}_{\text{TX}}^{\text{fix}}$ and a channel matrix $\mathbf{G}$, it returns as output the set of beamformers $\mathcal{B}_{\text{TX}}$ obtained after having run a conventional IA algorithm from the literature over this effective channel. Note that since the TX beamformers inside $\mathcal{B}_{\text{TX}}^{\text{fix}}$ are not modified, it holds that $\mathcal{B}_{\text{TX}}^{\text{fix}} \subset \mathcal{B}_{\text{TX}}$. We can then write

$$\mathcal{B}_{\text{TX}} = f_{\text{eff}}\big(\mathbf{G}, \mathcal{B}_{\text{TX}}^{\text{fix}}\big). \tag{8.37}$$

A number of IA algorithms can be run over the effective channel, and we will use the most simple IA algorithm called the *min-leakage* algorithm [86]. The main steps of the min-leakage algorithm have been recalled in Chapter 3. Our IA algorithm is obtained from the min-leakage algorithm after two simple modifications of the update formulas [Cf. equations (3.11) and (3.12)]:

- The update of the beamformers is done by considering all the interfering links and not by summing from 1 to K because we consider here *generalized* ICs.
- The TX beamformers contained in $\mathcal{B}_{\text{TX}}^{\text{fix}}$ are kept unchanged.

Precoding with incomplete CSIT Let us consider now the precoding at TX j with the CSIT allocation $\hat{\mathbf{H}}^{(j)} = \mathbf{F}_{\mathcal{S}_{\text{RX}}^{(j)}, \mathcal{S}_{\text{TX}}^{(j)}} \odot \mathbf{H}$. We define now in a recursive manner the precoding algorithm f_{IA} introduced in (8.36).

We start by defining the set $\mathcal{F}_j$ containing all the TXs whose CSIT allocations are strictly included in the CSIT known at TX j. Hence the set $\mathcal{F}_j$ is defined as

$$\mathcal{F}_j \triangleq \{k | \mathcal{S}_{\text{RX}}^{(k)} \subset \mathcal{S}_{\text{RX}}^{(j)}, \mathcal{S}_{\text{TX}}^{(k)} \subset \mathcal{S}_{\text{TX}}^{(j)}\}. \tag{8.38}$$

The beamformer $\boldsymbol{t}_j$ is then obtained from

$$\boldsymbol{t}_j = f_{\text{eff}}\big(\tilde{\mathbf{H}}^{(j)}, \{\boldsymbol{t}_k\}_{k \in \mathcal{F}_j}\big) \tag{8.39}$$

where $\tilde{\mathbf{H}}^{(j)}$ is the submatrix of $\hat{\mathbf{H}}^{(j)}$ containing only the columns and rows which are nonzero, and the beamformers $\{\boldsymbol{t}_k\}_{k\in\mathcal{F}_j}$ are obtained from

$$\boldsymbol{t}_k = f_{\text{IA}}\Big(\prod_{k=1}^{K}(N_k, M_k), \mathcal{F}, \hat{\mathbf{H}}^{(k)}\Big), \qquad \forall k \in \mathcal{F}_j. \tag{8.40}$$

Note that if $\mathcal{F}_j = \emptyset$, the beamformer $\boldsymbol{t}_j$ is simply obtained from $\boldsymbol{t}_j = f_{\text{eff}}(\tilde{\mathbf{H}}^{(j)}, \emptyset)$.

Achievability of interference alignment We have described a precoding algorithm but it remains to prove that IA is indeed achieved.

Theorem 15. The CSIT allocation policy $\mathcal{F}$ obtained with the incomplete CSIT allocation algorithm preserves IA feasibility: $\mathcal{F} \in \mathbb{F}_{\text{feas}}$: IA is achieved by the IA algorithm described above.

Proof. Let us consider w.l.o.g. the precoding at TX j. By construction, TX j is allocated with the CSI relative to the IC formed by the sets $(\mathcal{S}_{\text{RX}}^{(j)}, \mathcal{S}_{\text{TX}}^{(j)})$, which is *tightly-feasible*. We have shown in the proof of Theorem 14 that setting the beamformers in a tightly-feasible sub-IC to align interference in this sub-IC, does not reduce the feasibility of IA in the full IC. Thus, if all the TXs included in $\mathcal{S}_{\text{TX}}^{(j)}$ would design jointly their beamformers with the other TXs adapting to these TX beamformers, IA feasibility would then be preserved. Yet, all the TXs in $\mathcal{S}_{\text{TX}}^{(j)}$ do not necessarily share the same CSIT and thereby cannot necessarily design jointly the beamformers. Thus, it remains to prove that all the TXs included in $\mathcal{S}_{\text{TX}}^{(j)}$ design their beamformers in such a way that IA is achieved inside this sub-IC.

By inspection of the CSIT allocation algorithm, the CSIT allocations of all the TXs contained in $\mathcal{S}_{\text{TX}}^{(j)}$ are included in the CSIT of TX j. Thus, TX j can compute the beamformers of these TXs following the IA algorithm for incomplete CSIT exactly as it is done at these TXs. This is exactly what is done in our IA algorithm described in Subsection 8.3.3. This ensures the coherency between the beamformers of all the TXs in $\mathcal{S}_{\text{TX}}^{(j)}$ so that IA is achieved. □

Remark 18. *The RXs in $\mathcal{S}_{\text{RX}}^{(j)}$ and the TXs in $\mathcal{S}_{\text{TX}}^{(j)}$, as returned by the CSIT allocation algorithm, form together the smallest tightly-feasible setting containing TX j. If the algorithm is initialized with $\mathcal{S}_{\text{TX}}^{(j)} = \emptyset$ instead of $\mathcal{S}_{\text{TX}}^{(j)} = \{j\}$, the smallest tightly-feasible generalized IC is obtained. Hence, this algorithm can also be used to verify the IA feasibility of an antenna configuration.* □

We will now discuss an example illustrating the operational meaning of our approach.

8.3.4 Example of Tightly-feasible Configuration

We consider the IC formed by the antenna configuration

$$\prod_{k=1}^{K}(N_k, M_k) = (2,3).(2,4).(3,5).(3,2).(4,2). \tag{8.41}$$

The CSIT allocation algorithm presented in Subsection 8.3.2 returns

$$\begin{aligned}\mathcal{F} = \Big\{&\mathbf{F}^{(1)} = \mathbf{F}_{\{1,2,3\},\{4,5,1\}}, \mathbf{F}^{(2)} = \mathbf{F}_{\{1,2,3,4\},\{1,2,4,5\}},\\ &\mathbf{F}^{(3)} = \mathbf{F}_{\{1,2,3,4,5\},\{1,2,3,4,5\}}, \mathbf{F}^{(4)} = \mathbf{F}_{\{1,2\},\{4,5\}}, \mathbf{F}^{(5)} = \mathbf{F}_{\{1,2\},\{4,5\}}\Big\}\end{aligned} \tag{8.42}$$

which indicates for example that TX 4 receives the CSI relative to the *generalized* IC formed by TX 4 and TX 5 and RX 1 and RX 2. The size of the incomplete CSIT allocation obtained is equal to 346 while the complete CSIT allocation has a size of 905.

TX 4 and TX 5 have only the CSI required to align their interference at RX 1 and RX 2, which is in fact the first step of the IA algorithm. Once this is done, TX 1 can design its beamformer to align its interference on the interference subspace created by TX 4 and TX 5 at RX 2 and RX 3. Proceeding further, TX 2 aligns its interference on the interference subspace spanned at RX 1, RX 3, and RX 4 by the previous TX beamformers. At this step, all the interference subspaces have been generated so that TX 3 can use its 5 antennas to align its interference at all the RXs.

As described in Subsection 8.3.2, each TX computes the beamformers of the TXs having a CSIT allocation included in its own CSIT before computing its own TX beamformer. For example, all the TXs start here by computing the beamformers of TX 4 and TX 5.

To verify the achievement of IA, we introduce a symbolic representation of IA in Fig. 8.1. We represent the dimensions available at RX i by an array of N_i boxes. The first box on the right represents the dimension taken by the signal while the other boxes represent the dimensions left free for the interferences. For each RX, another box indicates if a TX precodes its signal so as to align with the interference subspace, thus creating no additional dimension of interference. If this is not the case, the stream transmitted by this TX creates a dimension of interference at the RX considered. When

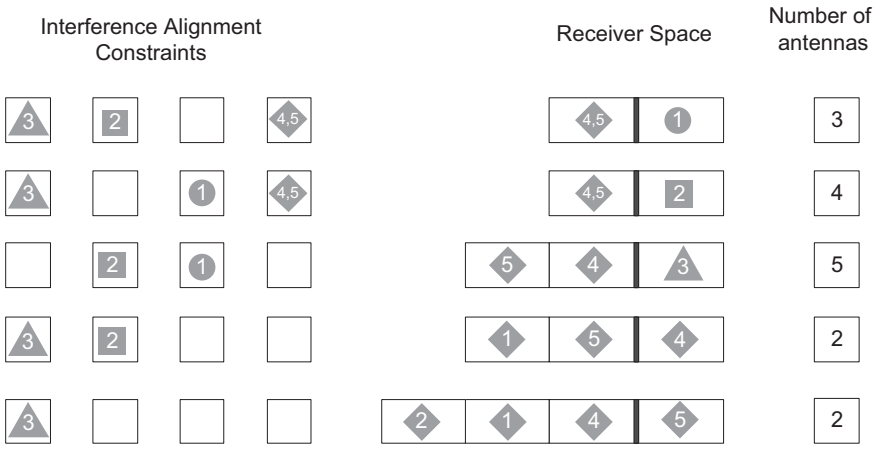


Figure 8.1: Symbolic Representation of the IA algorithm with incomplete CSIT for the tightly-feasible IC $(2,3).(2,4).(3,5).(3,2).(4,2)$.

the TX beamformers are not obtained by ZF but via iterations of an IA algorithm, it is not clear which TX generates the interference. We represent this by writing the indices of the interferers both in the IA box and in the RX box.

In this symbolic representation, a precoding scheme achieves IA if and only if the number of interfering dimensions at a RX does not exceed the number of boxes (dimensions) available at the RX while ensuring that each TX fulfills a number of IA constraint attainable with its antenna configurations (i.e. at most $M-1$ IA constraints if this TX has M antennas). In the case of iterative IA algorithm, this interpretation still holds, it is only not known at which RX the IA constraint is fulfilled.

We can observe in Fig. 8.1 that for all j, TX j aligns its interference at $M_j - 1$ RXs, and for all i, the interference subspace at RX i spans $N_i - 1$ dimensions. One can also notice that the setting is indeed tightly-feasible since removing an antenna at any TX or RX makes IA unfeasible.

The intuition behind the IA algorithm for incomplete CSIT is to break the IA into successive steps. The successive precoding steps are symbolized by the different columns on the left of the symbolic representation.

8.4 Interference Alignment with Incomplete CSIT for Super-Feasible Channels

The previous section indicates how CSIT savings can be obtained for tightly-feasible scenarios. When additional antennas are available, the intuition goes that further CSIT savings should be possible at no cost in terms of IA feasibility. We now investigate this question.

A distinct feature of super-feasible settings is that there must exist a corresponding tightly-feasible setting that can be obtained by keeping all TXs and RXs and simply ignoring certain antennas among the overall antenna set. Clearly, there are generally multiple ways for arriving at a tightly-feasible setting from a super-feasible one. Depending on the choice of which antennas are ignored in the initial super-feasible setting, the obtained tightly-feasible will exhibit particular CSIT requirements.

As a consequence, instead of considering directly (8.6), we consider the following optimization problem :

$$\begin{aligned} \mathcal{F} = \underset{\mathcal{F}\in\mathbb{F}}{\operatorname{argmin}} \min_{\prod_{k=1}^{K}(N_k',M_k')} \mathrm{s}(\mathcal{F}) \quad & \text{s.t. } f_{\text{feas}}\left(\mathcal{F}, \prod_{k=1}^{K}(N_k', M_k')\right) = 1 \\ & \text{s.t. } \sum_{i=1}^{K} M_i' + N_i' = (K+1)K \\ & \text{s.t. } 1 \leq M_i' \leq M_i \text{ and } 1 \leq N_i' \leq N_i. \end{aligned} \tag{8.43}$$

The problem of finding the minimal CSIT allocation has been reduced to finding the tightly-feasible setting (containing all the users) included in the full super-feasible setting which requires the smallest CSIT allocation. Since a CSIT allocation algorithm has been derived for tightly-feasible settings, it remains only to determine which RXs or TXs should not fully exploit their antennas to ZF interference dimensions, i.e., where some antennas should be "removed" in terms of IA feasibility.

Remark 19. *Practically, the antennas are not removed but some precoding dimensions are used for another purpose than aligning interference inside the IC (e.g., reducing interference to other RXs, increasing signal power, diversity, etc...). As an example, we will now show how it can be used to increase the received signal power. Intuitively, we select the0 precoding subspace of dimension n with $n < M_i$ which provides the largest received power to the RX. As a consequence, the quality of the direct channel is improved. Let us write the singular value decomposition of $\mathbf{H}_{ii} \in \mathbb{C}^{N_i \times M_i}$ as $\mathbf{H}_{ii} = \mathbf{U}_i \mathbf{\Sigma}_i \mathbf{V}_i^{\mathrm{H}}$*

with $\mathbf{V}_i = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{M_i}] \in \mathbb{C}^{M_i \times M_i}$ *and* $\mathbf{U}_i = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{N_i}] \in \mathbb{C}^{N_i \times N_i}$ *being two unitary matrices and* $\mathbf{\Sigma}_i = \mathrm{diag}(\sigma_1, \ldots, \sigma_{\min(M_i,N_i)}, 0, \ldots, 0)$. *We set* $\boldsymbol{t}_i = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n]\boldsymbol{t}'_i$ *with* $\boldsymbol{t}'_i \in \mathbb{C}^{n \times 1}$ *such that the dimension of the precoding subspace is reduced from* M_i *to* n. *However, the vectors* $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ *span the subspace of dimension* n *with the largest power. Altogether, the number of dimensions available for ZF precoding is reduced by one, which is equivalent in terms of IA feasibility to removing one antenna, while the quality of the direct channel is improved.* [1] □

This problem is still combinatorial in the total number of TXs and RXs which makes exhaustive search only practical for small settings. As a consequence, we provide in the following a CSIT allocation policy exploiting heuristically the additional antennas available to reduce the size of the CSIT allocation. The heuristic behind the algorithm comes from the insight gained in the analysis of tightly-feasible settings that the more heterogeneous is the antenna configuration, the smaller is the size of the CSIT allocation.

8.4.1 CSIT Allocation Algorithm

We consider in the following an heterogeneous IC and we denote by S the total number of additional antennas in the sense that S is defined as

$$S \triangleq \sum_{i=1}^{K} M_i + N_i - (K+1)K. \tag{8.44}$$

The following algorithm will provide the pair of sets $\mathcal{S}^{\mathrm{NT}} = (\mathcal{S}^{\mathrm{NT}}_{\mathrm{RX}}, \mathcal{S}^{\mathrm{NT}}_{\mathrm{TX}})$ containing respectively the RXs and the TXs where the additional antennas should be "removed". Once these antennas have been removed, the incomplete CSIT allocation policy for tightly-feasible can be applied to obtain the incomplete CSIT allocation. Note that we need to ensure that IA feasibility is preserved by the removing of the antennas.

Initialization: We define inside the algorithm a *virtual antenna configuration* $\prod_{i=1}^{K}(N_i^{\mathrm{v}}, M_i^{\mathrm{v}})$ which we initialize with the true antenna configuration $N_i^{\mathrm{v}} = N_i, M_i^{\mathrm{v}} = M_i$. We then initialize the two sets that will be given as output $\mathcal{S}^{\mathrm{NT}}_{\mathrm{RX}} = \emptyset, \mathcal{S}^{\mathrm{NT}}_{\mathrm{TX}} = \emptyset$.

Step n**:** We start by ordering the TXs inside $\prod_{i=1}^{K}(N_i^{\mathrm{v}}, M_i^{\mathrm{v}})$ by increasing number of antennas with the permutation $\sigma^{\mathrm{v}}_{\mathrm{TX}}$: $\forall i, M^{\mathrm{v}}_{\sigma^{\mathrm{v}}_{\mathrm{TX}}(i)} \leq M^{\mathrm{v}}_{\sigma^{\mathrm{v}}_{\mathrm{TX}}(i+1)}$. The RX ordering $\sigma^{\mathrm{v}}_{\mathrm{RX}}$ is defined similarly. Note that a node with $*$ antennas is considered as being an infinite number of antennas.

[1] Note that this step can be applied similarly on the RX side and that this process on the TX side requires the CSI relative to the direct channel.

In case of equality, the permutation is modified such that if $M^{\mathrm{v}}_{\sigma^{\mathrm{v}}_{\mathrm{TX}}(i)} = M^{\mathrm{v}}_{\sigma^{\mathrm{v}}_{\mathrm{TX}}(i+1)}$, then

$$(M^{\mathrm{v}}_{\sigma^{\mathrm{v}}_{\mathrm{TX}}(i)} = M^{\mathrm{v}}_{\sigma^{\mathrm{v}}_{\mathrm{TX}}(i+1)}) \Rightarrow N^{\mathrm{v}}_{\sigma^{\mathrm{v}}_{\mathrm{TX}}(i)} \geq N^{\mathrm{v}}_{\sigma^{\mathrm{v}}_{\mathrm{TX}}(i+1)}, \qquad \forall i. \tag{8.45}$$

We apply the same process symmetrically for the permutation $\sigma^{\mathrm{v}}_{\mathrm{RX}}$. Finally, if both RXs and TXs have the same number of antennas, the RX ordering is made opposite to the TX ordering:

$$\begin{aligned}\left(M^{\mathrm{v}}_{\sigma^{\mathrm{v}}_{\mathrm{TX}}(i)} = M^{\mathrm{v}}_{\sigma^{\mathrm{v}}_{\mathrm{TX}}(i+1)}, N^{\mathrm{v}}_{\sigma^{\mathrm{v}}_{\mathrm{TX}}(i)} = N^{\mathrm{v}}_{\sigma^{\mathrm{v}}_{\mathrm{TX}}(i+1)}\right) \\ \Rightarrow \left((\sigma^{\mathrm{v}}_{\mathrm{RX}})^{-1}(\sigma^{\mathrm{v}}_{\mathrm{TX}}(i+1)) < (\sigma^{\mathrm{v}}_{\mathrm{RX}})^{-1}(\sigma^{\mathrm{v}}_{\mathrm{TX}}(i))\right), \quad \forall i\end{aligned} \tag{8.46}$$

We define also the number of RXs and the number of TXs actually in the generalized IC $\prod_{i=1}^{K}(N^{\mathrm{v}}_i, M^{\mathrm{v}}_i)$ by respectively $K^{\mathrm{v}}_{\mathrm{RX}}$ and $K^{\mathrm{v}}_{\mathrm{TX}}$.

1. 1.1. We now define a set $\mathcal{S}^{\mathrm{v}} \triangleq (\mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}, \mathcal{S}^{\mathrm{v}}_{\mathrm{TX}})$. If $K^{\mathrm{v}}_{\mathrm{RX}} > 0$, we start by setting $\mathcal{S}^{\mathrm{v}} = (\{\sigma^{\mathrm{v}}_{\mathrm{RX}}(1)\}, \emptyset)$. If (8.17) is not fulfilled with this choice of $\mathcal{S}^{\mathrm{v}}$ and $K^{\mathrm{v}}_{\mathrm{TX}} > 0$, we reinitialize it with $\mathcal{S}^{\mathrm{v}} = (\emptyset, \{\sigma^{\mathrm{v}}_{\mathrm{TX}}(1)\})$.

   1.2. If the equality is reached in (8.17) with $\mathcal{S}^{\mathrm{v}}$, the sub-IC obtained is tightly-feasible and we update the antenna configuration to the one of the effective IC once IA is fulfilled in this sub-IC:

$$\begin{aligned}&\forall i \in \mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}, N^{\mathrm{v}}_i = *, \qquad \forall i \in \mathcal{S}^{\mathrm{v}}_{\mathrm{TX}}, M^{\mathrm{v}}_i = *.\\ &\forall i \notin \mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}, \begin{cases} N^{\mathrm{v}}_i = N^{\mathrm{v}}_i - |\mathcal{S}^{\mathrm{v}}_{\mathrm{TX}}|, & \text{if } i \notin \mathcal{S}^{\mathrm{v}}_{\mathrm{TX}}.\\ N^{\mathrm{v}}_i = N^{\mathrm{v}}_i - (|\mathcal{S}^{\mathrm{v}}_{\mathrm{TX}}| - 1) & \text{if } i \in \mathcal{S}^{\mathrm{v}}_{\mathrm{TX}}. \end{cases}\\ &\forall i \notin \mathcal{S}^{\mathrm{TX}}_{\mathrm{v}}, \begin{cases} M^{\mathrm{v}}_i = M^{\mathrm{v}}_i - |\mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}|, & \text{if } i \notin \mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}.\\ M^{\mathrm{v}}_i = M^{\mathrm{v}}_i - (|\mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}| - 1) & \text{if } i \in \mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}. \end{cases}\end{aligned} \tag{8.47}$$

   We then start over at step $n + 1$.

   1.3. If equation (8.17) is not verified.

   - If $|\mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}| < K^{\mathrm{v}}_{\mathrm{RX}}$, we verify whether adding the next RX adds more equations than variables, i.e.,

$$\begin{aligned}\mathcal{N}_{\mathrm{var}}(\mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}, \mathcal{S}^{\mathrm{v}}_{\mathrm{TX}}) - \mathcal{N}_{\mathrm{eq}}(\mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}, \mathcal{S}^{\mathrm{v}}_{\mathrm{TX}}) \\ \geq \mathcal{N}_{\mathrm{var}}(\{\mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}, \sigma^{\mathrm{v}}_{\mathrm{RX}}(|\mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}| + 1)\}, \mathcal{S}^{\mathrm{v}}_{\mathrm{TX}}) \\ - \mathcal{N}_{\mathrm{eq}}(\{\mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}, \sigma^{\mathrm{v}}_{\mathrm{RX}}(|\mathcal{S}^{\mathrm{v}}_{\mathrm{RX}}| + 1)\}, \mathcal{S}^{\mathrm{v}}_{\mathrm{TX}})\end{aligned} \tag{8.48}$$

If (8.48) is verified, we set

$$\mathcal{S}_{\mathrm{RX}}^{\mathrm{v}} = \{\mathcal{S}_{\mathrm{RX}}^{\mathrm{v}}, \sigma_{\mathrm{RX}}^{\mathrm{v}}(|\mathcal{S}_{\mathrm{RX}}^{\mathrm{v}}| + 1)\} \tag{8.49}$$

and we start over at step 1.*b*).

- If $|\mathcal{S}_{\mathrm{TX}}^{\mathrm{v}}| < K_{\mathrm{TX}}^{\mathrm{v}}$, we increase the set of TXs as

$$\mathcal{S}_{\mathrm{TX}}^{\mathrm{v}} = \{\mathcal{S}_{\mathrm{TX}}^{\mathrm{v}}, \sigma_{\mathrm{TX}}^{\mathrm{v}}(|\mathcal{S}_{\mathrm{TX}}^{\mathrm{v}}| + 1)\} \tag{8.50}$$

and we start over at step 1.*b*).

2. Otherwise, there is no tightly-feasible sub-IC and *removing an antenna cannot render IA unfeasible.*

- If $K_{\mathrm{TX}}^{\mathrm{v}} > 0$, we compute

$$M_{\sigma_{\mathrm{TX}}^{\mathrm{v}}(1)}^{\mathrm{v}} = M_{\sigma_{\mathrm{TX}}^{\mathrm{v}}(1)}^{\mathrm{v}} - 1, \quad \mathcal{S}_{\mathrm{TX}}^{\mathrm{NT}} = \{\mathcal{S}_{\mathrm{TX}}^{\mathrm{NT}}, \sigma_{\mathrm{TX}}^{\mathrm{v}}(1)\} \tag{8.51}$$

and we start at step $n + 1$.

- If $K_{\mathrm{TX}}^{\mathrm{v}} = 0$ but $K_{\mathrm{RX}}^{\mathrm{v}} > 0$, we set

$$N_{\sigma_{\mathrm{RX}}^{\mathrm{v}}(1)}^{\mathrm{v}} = N_{\sigma_{\mathrm{RX}}^{\mathrm{v}}(1)}^{\mathrm{v}} - 1, \quad \mathcal{S}_{\mathrm{RX}}^{\mathrm{NT}} = \{\mathcal{S}_{\mathrm{RX}}^{\mathrm{NT}}, \sigma_{\mathrm{RX}}^{\mathrm{v}}(1)\} \tag{8.52}$$

and we start at step $n + 1$.

- If $K_{\mathrm{v}}^{\mathrm{TX}} = 0$ and $K_{\mathrm{v}}^{\mathrm{RX}} = 0$, the algorithm has reached its end.

Description of the Algorithm Each iteration step is divided into two procedures marked with the 1) and the 2), respectively.

The first process consists in finding all the generalized tightly-feasible sub-ICs. The tightly-feasible sub-ICs are obtained following similar steps as in the algorithm in Subsection 8.3.2. It consists in the gradual increase of the set of TXs and the set of RXs so as to always obtain the "most tight" sub-ICs. For each of these sets, equation (8.17) is tested to verify whether the setting is tightly-feasible. This is carried out in steps 1.*c*) and 1.*d*).

Once a tightly-feasible subset is found, the TX beamformers in this sub-IC are computed and the channel matrix is replaced by the effective channel matrix. This is done by the intermediate of the *virtual antenna configuration* which represents the number of free variables remaining in the effective channel obtained. This process corresponds to step 1.*b*).

At the end of procedure 1), the virtual antenna configuration obtained does not contain any tightly-feasible sub-ICs. This is critical for the second step because it means that reducing the number of variables by one *cannot*

lead to the violation of (8.11) in Theorem 13. As a consequence, IA feasibility is preserved by removing one antenna in procedure 2). Any antenna can be removed and the policy chosen in the algorithm consists in removing the antenna at the TX with the smallest number of antennas if there is at least one TX left [Cf. equation (8.51)] and otherwise at the RX with the smallest number of antennas [Cf. equation (8.52)].

Remark 20. *Our algorithm relies on the property shown during the analysis of the tightly-feasible case that setting the TX beamformers inside a tightly-feasible sub-IC to fulfill IA solely inside this sub-IC does not reduce the feasibility of IA in the full IC.* □

8.4.2 Toy-Example of the Incomplete CSIT-Algorithm in Super-Feasible Settings

Let us consider as a toy-example the super-feasible IC $(2,2).(3,2).(2,3)$ containing two additional antennas since $\sum_i N_i + M_i - K(K+1) = 2$. We will now go through the steps of our CSIT allocation algorithm for non-tightly feasible ICs.

- $n = 1$: In phase 1), the algorithm starts by verifying whether there is any tightly-feasible set. This is not the case here such that phase 2) begins and one antenna is removed at TX 1. The virtual IC obtained is then $(2,1).(3,2).(2,3)$.

- $n = 2$: It is again verified in phase 1) whether there is any tightly-feasible set. Since TX 1 has only one antenna, it forms by itself a tightly-feasible set, so that its TX beamformer can be fixed and the antenna configuration is replaced by the virtual antenna configuration $(2,*).(2,2).(1,3)$. RX 3 has then only one antenna, so that we can obtain the virtual IC $(2,*).(2,1).(*,3)$. Once more, the same procedure applies to TX 2 to obtain $(1,*).(2,*).(*,3)$ and then again to RX 1 to get $(*,*).(2,*).(*,2)$. Finally, there is no tightly-feasible set so that phase 1) ends and phase 2) begins. Consequently, one antenna is removed at TX 3 to obtain the IC $(*,*).(2,*).(*,1)$.

- Step 3: TX 3 has one antenna left so that the IC $(*,*).(1,*).(*,*)$ is obtained. The same is done for RX 2 to obtain the IC $(*,*).(*,*).(*,*)$. Both the TX set and the RX set are empty so that the stopping criteria is reached and the algorithm returns the set containing the indices of the "removed antennas" $\mathcal{S}_{\text{TX}}^{\text{NT}} = \{1,3\}$ and $\mathcal{S}_{\text{RX}}^{\text{NT}} = \emptyset$.

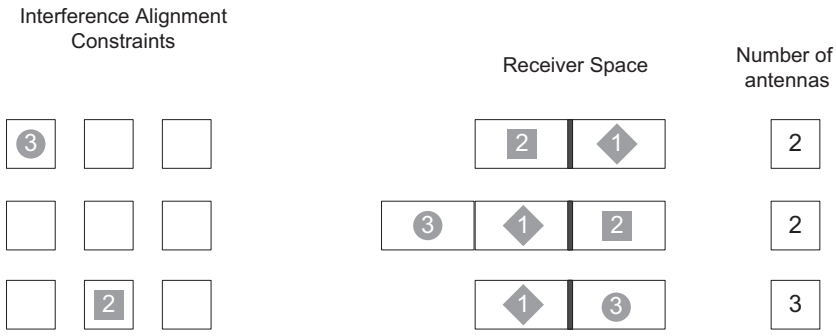


Figure 8.2: Symbolic representation of the IA algorithm with incomplete CSIT for the super-feasible IC $(2,2).(3,2).(2,3)$.

The CSIT allocation algorithm leads to remove the antennas at TX 1 and TX 3 to obtain the IC $(2,1).(3,2).(2,2)$. This setting being tightly-feasible, we can run the CSIT allocation for tightly-feasible ICs described in Subsection 8.3.2 which returns the CSIT allocation

$$\mathcal{F} = \{\mathbf{F}^{(1)} = \mathbf{F}_{\emptyset,\emptyset}, \mathbf{F}^{(2)} = \mathbf{F}_{\{3\},\{1,2\}}, \mathbf{F}^{(3)} = \mathbf{F}_{\{1,3\},\{1,2,3\}}\}. \tag{8.53}$$

The size of the CSIT allocation in (8.53) is equal to 20 while the complete CSIT allocation has a size of 99. Thus, the additional antennas have been used to reduce the feedback size by practically a factor of 4. The IA algorithm for incomplete CSIT sharing which follows from this CSIT allocation is symbolically represented in Fig. 8.2. It can be seen that TX 1 fulfills no ZF constraint and that both TX 2 and TX 3 fulfill each one ZF constraint.

8.5 Simulations

8.5.1 Tightly-Feasible Setting

We start by verifying by simulations that IA is indeed achieved by our new IA algorithm. We consider for the simulations the $(2,3).(2,4).(3,5).(3,2).(4,2)$ IC which has been studied in the example in Subsection 8.3.4. This example has been chosen to illustrate our approach, but the CSIT reduction brought by our approach being different for each antenna configuration, it is necessary to consider the average reduction over all the antenna configurations. This will be done in the following subsection.

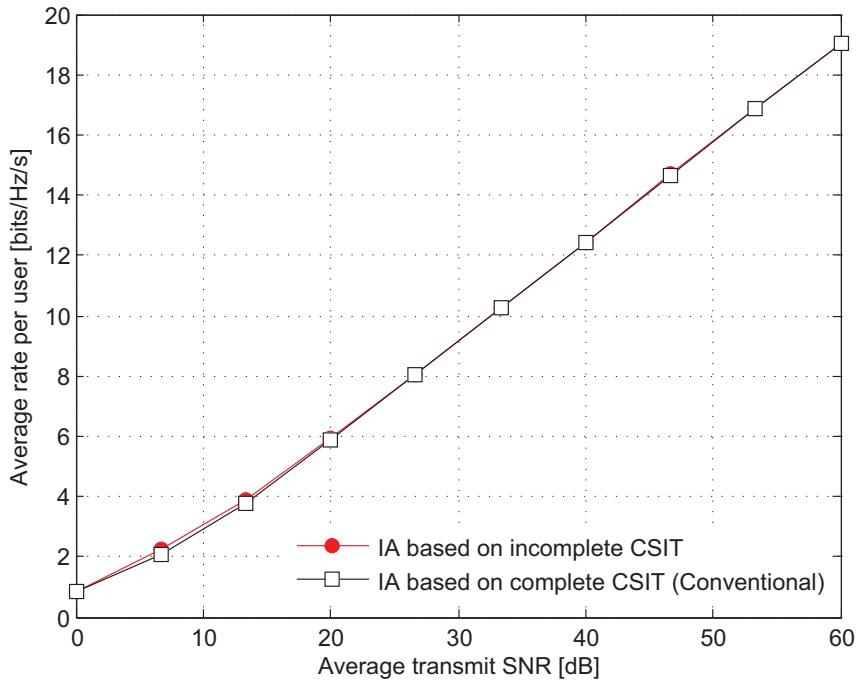


Figure 8.3: Average rate per user in terms of the normalized TX power for the tightly-feasible IC $(2,3).(2,4).(3,5).(3,2).(4,2)$ with the size of the incomplete CSIT allocation being only equal to 40% of the size of the complete CSIT allocation.

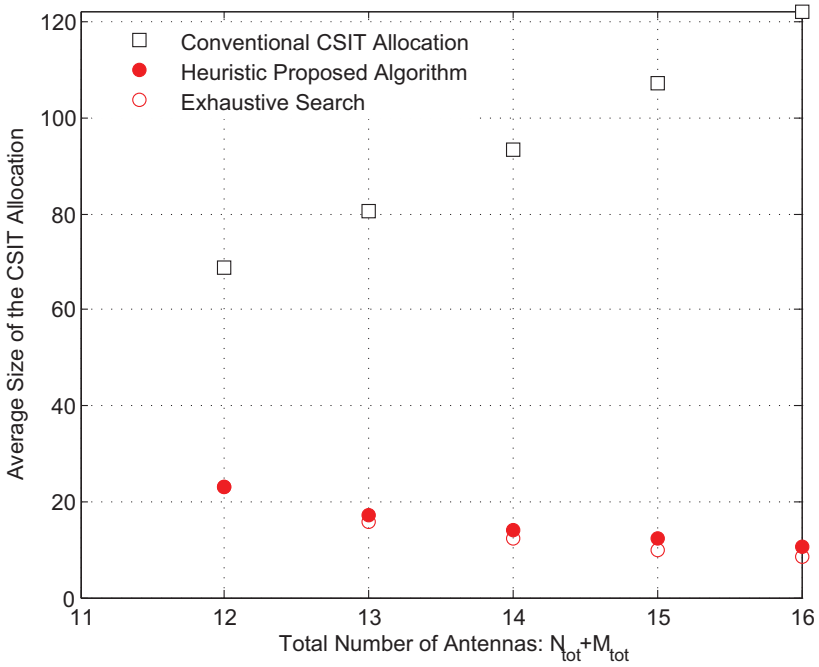


Figure 8.4: Average CSIT allocation size in terms of the number of antennas distributed across the TXs and the RXs for $K = 3$ users.

We show in Fig. 8.3 the average rate per user achieved in terms of the SNR. We compare then our IA algorithm based on incomplete CSIT to the min-leakage IA algorithm based on complete CSIT. Our algorithm achieves virtually the same performance as the min-leakage algorithm. Hence, the reduction of 60% of the feedback size (Cf. Subsection 8.3.4) comes for "free", making it especially interesting in practice.

8.5.2 Performance Evaluation of the CSIT allocation Algorithm

We will now evaluate the feedback reduction obtained with our CSIT allocation policy in super-feasible settings. Since this gain depends on the antenna configuration, we show in Fig. 8.4 the average size of the CSIT allocation for $K = 3$ users when the antennas are allocated uniformly at random to

the TXs and the RXs. Note that the antenna configurations obtained can be unfeasible. When this is the case, we redistribute the antennas until a feasible antenna configuration is obtained.

We average over 1000 antenna configurations and the proposed heuristic CSIT allocation policy is compared with the exhaustive search. The exhaustive search consists in testing all the possibilities for removing the additional antennas[2]. For reference, we also show the average size of the complete (conventional) CSIT allocation. We consider only $K = 3$ users because of the exponential complexity of the exhaustive search.

If the aggregate number of antennas is strictly smaller than $K(K+1) = 12$, IA is clearly not feasible. If more than 12 antennas are available, each additional antenna is exploited by the heuristic algorithm to reduce the size of the CSIT allocation. This algorithm brings a reduction of the CSIT size which is only slightly smaller than the reduction brought by exhaustive search, but has a polynomial complexity.

8.6 Discussion

IA feasibility is studied in the literature under the assumption of full CSIT sharing. In contrast, we have investigated here the relation between IA feasibility and CSIT allocation. Specifically, we have shown how IA can be achieved in some configurations without the need for full CSIT sharing. When extra-antennas are available, the existence of a trade-off between the number of antennas available and the CSIT sharing requirements is shown. Our approach brings a significant reduction of the feedback size while introducing no losses in terms of DoF compared to the conventional IA algorithm with full CSIT sharing.

Optimizing directly the CSIT allocation, it was possible to reduce the total CSIT requirements by providing each TX solely with the CSI which is necessary for its role in the transmission scheme. Providing each TX solely with the CSI it really needs appears then as having a strong potential to reduce the CSIT requirements and make TX cooperation more practical.

[2]Note that a true exhaustive search through all the possible CSIT allocations (i.e., coming back to the original optimization problem (3.19)) is too complex even for trivial antenna configurations.

Chapter 9

Conclusion

The focus of this thesis has been the analysis of the multi-antenna cooperation schemes in the case of CSI being imperfectly shared between the TXs. We have shown how the consideration of CSI discrepancies between the cooperating TXs significantly impacts the transmission as it can no longer be modeled as a conventional optimization problem but it becomes instead a Team Decision problem where distant TXs aim at achieving cooperation without sharing the same information.

Considering the cooperation of distant TXs with individual information is a very general problem which can be tackled under many approaches. We have provided in the first part of this thesis some results which shed some lights on the impact of distributed CSIT over the precoder design. The main message is that distributed CSIT impacts considerably the precoding step and that omitting to consider the CSI discrepancies leads to a large tribute in terms of performance in practical settings. Beyond the performance evaluation of the conventional precoding schemes which can be used to dimension the feedback schemes or predict the performance, novel precoding paradigms have been provided. The design of precoders being robust to CSI discrepancies remains however open in many aspects and will be further investigated in the future.

The focus of the second part of this thesis has been the optimization of the spatial allocation of the CSIT. Providing each TX with a different CSI which corresponds to its actual need has been shown to be the key to significant reductions of the CSIT requirements at virtually no cost. Although many practical aspects have to be considered in order to translate the promised theoretical gains into improvements in realistic networks, this approach appears to be very promising and to have a strong potential for

improvements. The optimization has been described in two scenarios, exploiting either the antenna configuration or the pathloss attenuations, but the idea of adapting the CSIT allocation to the environment is very general and it is believed that CSIT reductions could be achieved in many other transmission scenarios. We have provided the first results emphasizing the interest of optimizing the CSIT spatial allocation, but there are a number of open questions that should be solved and that could lead to further results. For example, proving the optimality of our novel scheme –or finding the optimal one– is a challenging research problem.

The initial Team Decision problem formulated is a very intricate problem which can be tackled under many angles and approaches. We have studied some aspects of the problem using some tools, but there are many more possibilities, depending on the modelization of the distributed CSIT, the communication capabilities of the TX, the channel model, etc... Solving the Team Decision problem formulated in its full generality represents a real challenge for the interested researchers.

Appendices

.1 Useful Lemmas

Lemma 1 ([24], Appendix A). Let θ be a random variable uniformly distributed in $[0, 2\pi)$. Then, we have

$$\mathrm{E}_\theta[\log(|B + Aexp(j\theta)|^2)] = \log(\max\{|A|^2, |B|^2\}) \tag{1}$$

Lemma 2 (Resolvent identity). Let $\mathbf{A} \in \mathbb{C}^{n\times n}$ and $\mathbf{B} \in \mathbb{C}^{n\times n}$ be two invertible matrices, it then holds that

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

Proof. This is a well known result, whose (simple) proof can be found for example in [121, Lemma 6.1]. □

Lemma 3. Let $X_i, i = 1, \ldots, n$, be n real random variables, and $\rho_i, i = 1, \ldots, n$, be n positive real numbers. Then

$$\Pr\left(\prod_{i=1}^{n} X_i > \prod_{i=1}^{n} \rho_i\right) \leq \sum_{i=1}^{n} \Pr(X_i > \rho_i) \tag{i}$$

and

$$\Pr\left(\sum_{i=1}^{n} X_i > \sum_{i=1}^{n} \rho_i\right) \leq \sum_{i=1}^{n} \Pr(X_i > \rho_i). \tag{ii}$$

Proof. We prove the lemma for $n = 2$. The generalization to arbitrary n follows by direct induction.

- We start by proving (i). From the conditional expectation property, we write

$$\begin{aligned}
&\Pr(X_1X_2 > \rho_1\rho_2) \\
&= \Pr\left(X_2 > \rho_2\right) \Pr\left(X_1 > \frac{\rho_1\rho_2}{X_2}\middle| X_2 > \rho_2\right) \\
&\qquad + \Pr\left(X_2 \leq \rho_2\right) \Pr\left(X_1 > \frac{\rho_1\rho_2}{X_2}\middle| X_2 \leq \rho_2\right) &(3)\\
&\leq \Pr\left(X_2 > \rho_2\right) + \Pr\left(X_1 > \frac{\rho_1\rho_2}{X_2}\middle| X_2 \leq \rho_2\right) &(4)\\
&\overset{(a)}{\leq} \Pr(X_2 > \rho_2) + \Pr(X_1 > \rho_1) &(5)
\end{aligned}$$

  where inequality (a) is verified because $\rho_1\rho_2/X_2 \geq \rho_1$ when $X_2 \leq \rho_2$.

- We now turn to the proof of (ii). Using once more the conditional expectation property, we write

$$\begin{aligned}
&\Pr(X_1 + X_2 > \rho_1 + \rho_2) && (6)\\
&= \Pr\left(X_2 > \rho_2\right) \Pr\left(X_1 > \rho_1 + \rho_2 - X_2 | X_2 > \rho_2\right) \\
&\qquad + \Pr\left(X_2 \leq \rho_2\right) \Pr\left(X_1 > \rho_1 + \rho_2 - X_2 | X_2 \leq \rho_2\right) && (7)\\
&\leq \Pr\left(X_2 > \rho_2\right) + \Pr\left(X_1 > \rho_1 + \rho_2 - X_2 | X_2 \leq \rho_2\right) && (8)\\
&\overset{(a)}{\leq} \Pr\left(X_2 > \rho_2\right) + \Pr\left(X_1 > \rho_1\right) && (9)
\end{aligned}$$

where inequality (a) holds true because $\rho_2 - X_2 \geq 0$ given that $X_2 \leq \rho_2$.

□

Lemma 4 (Lemma 2, [122]). Let $Y = \sum_{i=1}^{n} X_i^2$ with X_i i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$. Then it holds for $x \geq n$ that

$$\Pr(Y > x) \leq e^{-\frac{x}{2} + \frac{n}{2}\log(\frac{ex}{n})}. \tag{10}$$

Lemma 5. Let $X_i, i = 1, \ldots, n$ be n i.i.d. Chi-square random variable with 2 degrees-of-freedom. It then holds

$$\mathrm{E}[\max_{i=1,\ldots,n} X_i] = 2 \sum_{m=1}^{n} \frac{1}{m}. \tag{11}$$

Proof. The distribution of X_i is known to verify

$$\Pr(X_i < x) = 1 - e^{-\frac{x}{2}}. \tag{12}$$

Using that $X_i \geq 0$, we can further write

$$\mathrm{E}[\max_{i=1,\ldots,n} X_i] = \int_0^\infty \Pr(\max_{i=1,\ldots,n} X_i \geq x)\mathrm{d}x \tag{13}$$

$$= \int_0^\infty 1 - \Pr(\max_{i=1,\ldots,n} X_i < x)\mathrm{d}x \tag{14}$$

$$= \int_0^\infty 1 - (\Pr(X_i < x))^n \mathrm{d}x \tag{15}$$

$$= \int_0^\infty 1 - (1 - e^{-\frac{x}{2}})^n \mathrm{d}x \tag{16}$$

$$\overset{(a)}{=} \sum_{k=1}^{n} \binom{n}{k} (-1)^k \int_0^\infty e^{-\frac{kx}{2}} \mathrm{d}x \tag{17}$$

$$= 2\sum_{k=1}^{n} \binom{n}{k} (-1)^{k+1} \frac{1}{k} \tag{18}$$

$$\overset{(b)}{=} 2\sum_{m=1}^{n} \frac{1}{m} \tag{19}$$

where (a) is obtained after using the binomial expansion [123, Equation 0.111] and (b) after using [123, Equation 0.155]

$$\sum_{k=1}^{n} \binom{n}{k} (-1)^{k+1} \frac{1}{k} = \sum_{m=1}^{n} \frac{1}{m}. \tag{20}$$

□

Lemma 6. Let $n \in \mathbb{N}^*$, it then holds

$$\sum_{k=2}^{n} \frac{1}{k} \leq \log(n) \leq \sum_{k=1}^{n-1} \frac{1}{k} \tag{21}$$

Proof. The proof is based on the logarithm property that

$$\log(x) = \int_1^x \frac{\mathrm{d}t}{t}. \tag{22}$$

We have then for $n \in \mathbb{N}^*$,

$$\int_1^n \frac{\mathrm{d}t}{\lceil t \rceil} \leq \int_1^n \frac{\mathrm{d}t}{t} \leq \int_1^n \frac{\mathrm{d}t}{\lfloor t \rfloor} \tag{23}$$

which can be rewritten as

$$\sum_{k=2}^{n} \frac{1}{k} \leq \log(n) \leq \sum_{k=1}^{n-1} \frac{1}{k}. \tag{24}$$

□

Lemma 7. Let us consider a random variable $X \in \mathbb{R}^+$ and let $\mathcal{O}$ be a subset of $\mathbb{R}^+$. It then holds

$$\mathrm{E}_{\mathcal{O}}[X] \leq \frac{\mathrm{E}[X]}{\Pr(\mathcal{O})}. \tag{25}$$

Proof. Using the conditional expectation property, we write

$$\mathrm{E}[X] = \Pr(\mathcal{O})\mathrm{E}_{\mathcal{O}}[X] + \Pr(\bar{\mathcal{O}})\mathrm{E}_{\bar{\mathcal{O}}}[X] \tag{26}$$

where $\bar{\mathcal{O}} \triangleq \mathcal{Q} \setminus \mathcal{O}$. It then easily follows

$$\mathrm{E}_{\mathcal{O}}[X] = \frac{\mathrm{E}[X] - \Pr(\bar{\mathcal{O}})\mathrm{E}_{\bar{\mathcal{O}}}[X]}{\Pr(\mathcal{O})} \tag{27}$$

$$\leq \frac{\mathrm{E}[X]}{\Pr(\mathcal{O})}. \tag{28}$$

□

.2 Analysis of the Random Vector Quantization Scheme

We consider the random vector quantization of the unit-norm complex vector $\tilde{\boldsymbol{h}} \in \mathbb{C}^K$ over a codebook $\mathcal{W}$ where both the channel to quantize and the elements of the codebook are multiplied by a unit-norm complex number (i.e., are rotated in the complex space) so as to let the first element of the vector be real valued. Mathematically, the quantized estimate $\hat{\boldsymbol{h}}$ is obtained from

$$\hat{\boldsymbol{h}} = \underset{\boldsymbol{w} \in \mathcal{W}}{\operatorname{argmin}} \left\| \frac{\{\boldsymbol{w}\}_1^*}{|\{\boldsymbol{w}\}_1|} \tilde{\boldsymbol{w}} - \frac{\{\boldsymbol{h}\}_1^*}{|\{\boldsymbol{h}\}_1|} \tilde{\boldsymbol{h}} \right\| \tag{1}$$

where $\mathcal{W}$ is a codebook made of vectors isotropically distributed over the unit sphere. The multiplication by the unit-norm complex number is done in order to optimize the performance of the quantization. Since the norm is conserved when considering the canonical isomorphism from $\mathbb{C}^K$ to $\mathbb{R}^{2K}$, we can consider for the quantization the vectors as elements of $\mathbb{R}^{2K}$ made of the stacked real and imaginary parts of the original vector.

Because of the rotation, the first element of the vector has its complex value which can be set to 0. It is hence only necessary to consider the space $\mathbb{R}^{2K-1}$. Thus, a vector $\boldsymbol{u} = [u_1, u_2, \ldots, u_K]^{\mathrm{T}} \in \mathbb{C}^K$ with its first coefficient real valued is represented in $\mathbb{R}^{2K-1}$ by $\boldsymbol{u}_{\mathbb{R}^{2K-1}}$ defined as

$$\boldsymbol{u}_{\mathbb{R}^{2K-1}} \triangleq \big[\mathrm{Re}(u_1) \quad \mathrm{Re}(u_2) \quad \mathrm{Im}(u_2) \quad \mathrm{Re}(u_3) \quad \ldots \quad \mathrm{Im}(u_K)\big]^{\mathrm{T}}. \tag{2}$$

We can then define the angle between $\boldsymbol{u}_{\mathbb{R}^{2K-1}}$ and $\boldsymbol{v}_{\mathbb{R}^{2K-1}}$ in $\mathbb{R}^{2K-1}$ as

$$\angle(\boldsymbol{u}_{\mathbb{R}^{2K-1}}, \boldsymbol{v}_{\mathbb{R}^{2K-1}}) \triangleq \arccos\left(\frac{|\boldsymbol{u}_{\mathbb{R}^{2K-1}}^{\mathrm{T}} \boldsymbol{v}_{\mathbb{R}^{2K-1}}|}{\|\boldsymbol{u}_{\mathbb{R}^{2K-1}}\| \|\boldsymbol{v}_{\mathbb{R}^{2K-1}}\|}\right). \tag{3}$$

Using the conservation of the norm by the canonical isomorphism, the quantization in (1) is rewritten as

$$\hat{\boldsymbol{h}}_{\mathbb{R}^{2K-1}} = \underset{\boldsymbol{w}_{\mathbb{R}^{2K-1}} \in \mathcal{W}_{\mathbb{R}^{2K-1}}}{\operatorname{argmin}} \|\boldsymbol{w}_{\mathbb{R}^{2K-1}} - \tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}\|^2 \tag{4}$$

$$= \underset{\boldsymbol{w}_{\mathbb{R}^{2K-1}} \in \mathcal{W}_{\mathbb{R}^{2K-1}}}{\operatorname{argmin}} (2 - 2\boldsymbol{w}_{\mathbb{R}^{2K-1}}^{\mathrm{T}} \tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}) \tag{5}$$

where the codebook $\mathcal{W}_{\mathbb{R}^{2K-1}}$ is the counterpart of $\mathcal{W}$ in $\mathbb{R}^{2K-1}$. We can see from (5) that the quantization scheme aims at maximizing $\boldsymbol{w}_{\mathbb{R}^{2K-1}}^{\mathrm{T}} \tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}$. This figure of merit can be linked to the commonly used chordal distance $\mathrm{d}_{\mathrm{ch}}(\bullet)$ which is defined for two vectors as [80]

$$\mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}, \boldsymbol{w}_{\mathbb{R}^{2K-1}}) \triangleq \sqrt{\sin^2(\angle(\boldsymbol{w}_{\mathbb{R}^{2K-1}}, \tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}))} \tag{6}$$

$$= \sqrt{1 - |\boldsymbol{w}_{\mathbb{R}^{2K-1}}^{\mathrm{T}} \tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}|^2}. \tag{7}$$

Thus, minimizing the chordal distance is equivalent to the maximization of the objective $|\boldsymbol{w}_{\mathbb{R}^{2K-1}}^{\mathrm{T}} \tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}|^2$. This is then equivalent to the quantization scheme (3) if the half-space where $\tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}$ belongs is known. This requires solely one additional bit. Since we are interested in the scaling of the number of bits, this will not make any difference. Consequently, we will study in the following the quantization scheme based on the minimization of the chordal distance

$$\hat{\boldsymbol{h}}_{\mathbb{R}^{2K-1}} = \underset{\boldsymbol{w}_{\mathbb{R}^{2K-1}} \in \mathcal{W}_{\mathbb{R}^{2K-1}}}{\operatorname{argmin}} \mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}}_{\mathbb{R}^{2K-1}}, \boldsymbol{w}_{\mathbb{R}^{2K-1}}). \tag{8}$$

This quantization scheme corresponds to the conventional quantization over the Grassmannian manifold of dimensions $(1, 2K-1)$ in the field $\mathbb{R}$ (i.e., on the unitary ball in $\mathbb{R}^{2K-1}$). This quantization scheme is studied (in a much

more general form) in [80] and we start by recalling some results adapted to our notations. We then derive some new properties which will be needed in the derivations. [1]

We denote by $\mathrm{F}(x) \triangleq \Pr\{\mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}}, \boldsymbol{w}) \leq x\}$ the cumulative distribution function (CDF) of $\mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}}, \boldsymbol{w}) = \sin^2(\angle(\tilde{\boldsymbol{h}}, \boldsymbol{w}))$ where $\boldsymbol{w} \in \mathbb{R}^{2K-1}$ is an element of a random codebook.

Proposition 14 ([80], Corollary 2). The CDF $\mathrm{F}(x)$ verifies that for all $x \leq 1$

$$c_{2K-1} x^{K-1} \leq \mathrm{F}(x) \leq c_{2K-1} x^{K-1} (1-x)^{\frac{-1}{2}} \tag{9}$$

where $c_{2K-1} \triangleq \Gamma(K-1/2)/(\Gamma(K)\Gamma(1/2))$.

Proposition 15 ([80], Theorem 2). As the size $L = 2^B$ of the random codebook increases, it then holds

$$\begin{aligned} \frac{2K-1}{2K+1} c_{2K-1}^{-1/(K-1)} 2^{-B/(K-1)} &\leq \mathrm{E}_{\mathcal{W},\tilde{\boldsymbol{h}}}[\min_{\boldsymbol{w}\in\mathcal{W}} \mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}}, \boldsymbol{w})] + o(1) \\ &\leq \frac{\Gamma(\frac{1}{K-1})}{K-1} c_{2K-1}^{-1/(K-1)} 2^{-B/(K-1)}. \end{aligned} \tag{10}$$

Proposition 16. When the size $L = 2^B$ of the random codebook is sufficiently large, the expectation of the logarithm of the quantization error is bounded as

$$\begin{aligned} \frac{B + \log_2(c_{2K-1})}{(K-1)} &\leq \mathrm{E}_{\mathcal{W},\tilde{\boldsymbol{h}}} \left[-\log_2 \left(\min_{\boldsymbol{w}\in\mathcal{W}} \mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}}, \boldsymbol{w}) \right) \right] + o(1) \\ &\leq \frac{B + \log_2(c_{2K-1}) + \log_2(e)}{(K-1)}. \end{aligned} \tag{11}$$

Proof. **Upper Bound:** The derivation of an upper bound follows the same idea as the proof in Appendix B of [80] which derives an upper bound for Proposition 15. We start by recalling a Lemma from [80] which follows easily from the definition but is helpful.

Lemma 8 ([80], Lemma 3). The empirical distribution function minimizing the distorsion for a given $L = 2^B$ is

$$\mathrm{F}^*_{\mathcal{W}^*}(x) = \begin{cases} 0 & \text{if } x < 0 \\ L\mathrm{F}(x) & \text{if } 0 \leq x \leq x^* \\ 1 & \text{if } x > x^* \end{cases} \tag{12}$$

[1] We will do the abuse of notation consisting in removing the subscript $\mathbb{R}^{2K-1}$ in the derivations but it will be clear that any mention of an angle will refer to the angle defined in $\mathbb{R}^{2K-1}$.

where x^* satisfies $L\mathrm{F}(x^*)=1$ and $\mathrm{F}(x) \triangleq \Pr\{\mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}}, \boldsymbol{w})| \leq x\}$.

Note that Lemma 8 corresponds to the optimal codebook minimizing the average distance and leads thusly to a lower bound for the distorsion. We can then write

$$\mathrm{E}_{\mathcal{W},\tilde{\boldsymbol{h}}}\left[-\log\left(\min_{\boldsymbol{w}\in\mathcal{W}} \mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}}, \boldsymbol{w})\right)\right] = \int_0^{\infty} \Pr\{-\log\left(\min_{\boldsymbol{w}\in\mathcal{W}} \mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}}, \boldsymbol{w})\right) \geq z\}\mathrm{d}z \tag{13}$$

$$= \int_0^{\infty} \Pr\{\min_{\boldsymbol{w}\in\mathcal{W}} \mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}}, \boldsymbol{w}) \leq e^{-z}\}\mathrm{d}z \tag{14}$$

$$\leq \int_0^{-\log(x^*)} \mathrm{d}z + \int_{-\log(x^*)}^{-\infty} L\Pr\{\mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}}, \boldsymbol{w}) \leq e^{-z}\}\mathrm{d}z \tag{15}$$

where (13) is obtained by exploiting the fact that the term in the expectation is a positive random variable and (15) follows from the previous lemma since the CDF obtained with the optimal codebook dominates the CDF obtained with any other codebook of the same size.

Following the same approach as the proof in Appendix B of [80], we define $\mathrm{F}_0(x) \triangleq c_{2K-1}x^{K-1}$ and x_0 so that $L\mathrm{F}_0(x_0) = 1$. Let also define $\mathrm{F}_{\mathrm{ub}}(x) \triangleq c_{2K-1}x^{K-1}(1-x)^{-1/2}$ and x_{ub} so that $L\mathrm{F}_{\mathrm{ub}}(x_{\mathrm{ub}}) = 1$. Finally, we define $\mathrm{F}_{\mathrm{ubub}}(x) \triangleq c_{2K-1}x^{K-1}(1-x_0)^{-1/2}$ and x_{ubub} so that $L\mathrm{F}_{\mathrm{ubub}}(x_{\mathrm{ubub}}) = 1$.

It holds by construction that $x_{\mathrm{ub}} \leq x^* \leq x_0$ since we know from Proposition 14 that $\mathrm{F}_0(x) \leq \mathrm{F}(x) \leq \mathrm{F}_{\mathrm{ub}}(x)$. Clearly $(1-x)^{-1/2} \leq (1-x_0)^{-1/2}$ for $x \in [0, x_0]$ so that $\mathrm{F}_{\mathrm{ub}}(x) \leq \mathrm{F}_{\mathrm{ubub}}(x)$ for $x \in [0, x_0]$, which finally implies $x_{\mathrm{ubub}} \leq x_{\mathrm{ub}}$. We can then use these relations to derive an upper bound for (15).

$$\mathrm{E}_{\mathcal{W},\tilde{\boldsymbol{h}}}\left[-\log\left(\min_{\boldsymbol{w}\in\mathcal{W}} \mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}}, \boldsymbol{w})\right)\right]$$

$$\leq \int_0^{-\log(x^*)} \mathrm{d}_{\mathrm{ch}}z + \int_{-\log(x^*)}^{-\infty} L\mathrm{F}(e^{-z})\mathrm{d}z \tag{16}$$

$$\leq \int_0^{-\log(x_{\mathrm{ubub}})} \mathrm{d}z + \int_{-\log(x_0)}^{\infty} L\mathrm{F}(e^{-z})\mathrm{d}z \tag{17}$$

$$\leq \int_0^{-\log(x_{\mathrm{ubub}})} \mathrm{d}z + \int_{-\log(x_0)}^{\infty} L\mathrm{F}_{\mathrm{ubub}}(e^{-z})\mathrm{d}z. \tag{18}$$

Equation (17) follows from $x_{\mathrm{ubub}} \leq x^* \leq x_0$ and (18) follows from the fact that $\mathrm{F}_{\mathrm{ub}}(x) \leq \mathrm{F}_{\mathrm{ubub}}(x)$ for $x \in [0, x_0]$. We now replace $\mathrm{F}_{\mathrm{ubub}}(x)$, x_{ubub},

and x_0 by their expressions to evaluate the integral.

$$\begin{aligned}
&\mathrm{E}_{\mathcal{W},\tilde{\boldsymbol{h}}}\left[-\log\left(\min_{\boldsymbol{w}\in\mathcal{W}} \mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}},\boldsymbol{w})\right)\right] \\
&\leq -\tfrac{1}{K-1}\log\left(\tfrac{(1-x_0)^{1/2}}{Lc_{2K-1}}\right)+\tfrac{Lc_{2K-1}}{(1-x_0)^{1/2}}\int_{-\log(x_0)}^{\infty} e^{-z(K-1)}\mathrm{d}z \qquad (19)\\
&= -\frac{1}{K-1}\log\left(\frac{(1-x_0)^{1/2}}{Lc_{2K-1}}\right)+\frac{1}{(1-x_0)^{1/2}(K-1)} \qquad (20)\\
&= \frac{1}{K-1}\left(\log\left(Lc_{2K-1}\right)+1\right)+o(1) \qquad (21)
\end{aligned}$$

as L increases. Dividing by $\log(2)$ yields the final upper bound.

Lower Bound: We start from the lower bound for the CDF given in Proposition 14. It has a form very similar to the CDF for the quantization of a complex vector in the unit-ball in $\mathbb{C}^K$ which is usually used for multiple-antenna BC. Hence, we adapt the approach of the proof of Lemma 3 in [15] to the current setting.

From the lower bound in Proposition 14, we write

$$\Pr\{\min_{\boldsymbol{w}\in\mathcal{W}}\left(\mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}},\boldsymbol{w})\right)\leq z\}\geq 1-(1-c_{2K-1}x^{(K-1)})^L. \qquad (22)$$

A lower bound for the expectation of the logarithm can then be calculated as follows.

$$\begin{aligned}
\mathrm{E}_{\mathcal{W},\tilde{\boldsymbol{h}}}\left[-\log\left(\min_{\boldsymbol{w}\in\mathcal{W}} \mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}},\boldsymbol{w})\right)\right] &= \int_0^{\infty}\Pr\{\min_{\boldsymbol{w}\in\mathcal{W}} \mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}},\boldsymbol{w})\leq e^{-z}\}\mathrm{d}z \qquad (23)\\
&\geq \int_0^{\infty} 1-(1-c_{2K-1}e^{-z(K-1)})^L\mathrm{d}z \qquad (24)\\
&= \int_0^{\infty} 1-\sum_{k=0}^{L}\binom{L}{k}(-1)^k c_{2K-1}^k e^{-z(K-1)k}\mathrm{d}z \qquad (25)\\
&= \frac{1}{K-1}\sum_{k=1}^{L}\binom{L}{k}(-1)^{k+1}\frac{c_{2K-1}^k}{k} \qquad (26)\\
&= \frac{1}{K-1}f(L) \qquad (27)
\end{aligned}$$

where we have defined $f(p)\triangleq\sum_{k=1}^{p}\binom{p}{k}(-1)^{k+1}\frac{c_{2K-1}^k}{k}$ for $p\in\mathbb{N}$. To compute the value of $f(L)$, we will use the following relation given in [123, Section

0.155].

$$\sum_{k=0}^{n} \binom{n}{k} \frac{\alpha^{k+1}}{k+1} = \frac{(\alpha+1)^{n+1}-1}{n+1}. \tag{28}$$

We now rewrite $f(L)$ in order to be able to apply (28)

$$\begin{aligned}
f(L) &\triangleq \sum_{k=1}^{L} \binom{L}{k} (-1)^{k+1} \frac{c_{2K-1}^{k}}{k} && (29)\\
&= \sum_{k=1}^{L} \left[\binom{L-1}{k-1} + \binom{L-1}{k} \right] (-1)^{k+1} \frac{c_{2K-1}^{k}}{k} && (30)\\
&= \sum_{k'=0}^{L-1} \binom{L-1}{k'} (-1)^{k'+2} \frac{c_{2K-1}^{k'+1}}{k'+1} + f(L-1) && (31)\\
&= -\frac{(-c_{2K-1}+1)^{L}-1}{L} + f(L-1) && (32)\\
&= \sum_{p=2}^{L} \frac{1-(-c_{2K-1}+1)^{p}}{p} + f(1) && (33)\\
&= \sum_{p=1}^{L} \frac{1}{p} - \sum_{p=1}^{L} \frac{1-(-c_{2K-1}+1)^{p}}{p}. && (34)
\end{aligned}$$

where we have used (28) to obtain (31) and we have iteratively expressed $f(x)$ in terms of $f(x-1)$ to write (33). Furthermore we have the two following relations:

$$\log(L) \leq \sum_{p=1}^{L} \frac{1}{p} \leq \log(L) + 1, \tag{35}$$

$$\log(1-x) = -\sum_{n=1}^{\infty} \frac{x^{n}}{n} \text{ , for } -1 < x < 1. \tag{36}$$

Inserting the expression derived for $f(x)$ inside (27) and using the two

bounds provided above, we can obtain the final lower bound as

$$\begin{aligned}
&\mathrm{E}_{\mathcal{W},\tilde{\boldsymbol{h}}}\left[-\log_2\left(\min_{\boldsymbol{w}\in\mathcal{W}} \mathrm{d}_{\mathrm{ch}}(\tilde{\boldsymbol{h}},\boldsymbol{w})\right)\right] \\
&\geq \frac{1}{(K-1)\log(2)}\sum_{p=1}^{L}\frac{1}{p} - \frac{1}{(K-1)\log(2)}\sum_{p=1}^{L}\frac{(1-c_{2K-1})^p}{p} \qquad (37)\\
&\geq \frac{\log_2(L)}{(K-1)} - \frac{1}{(K-1)\log(2)}\sum_{p=1}^{\infty}\frac{(1-c_{2K-1})^p}{p} \qquad (38)\\
&= \frac{\log_2(L)+\log_2(c_{2K-1})}{(K-1)} \qquad (39)
\end{aligned}$$

where we have used that the constant c_{2K-1} is smaller than one to apply (36) and obtain the term $\log_2(c_{2K-1})$. □

.3 Proof of Theorem 4

.3.1 Preliminaries

We start by proving two lemmas which will form the basis of the following proof. Throughout this proof, we use a Taylor approximation of the beamformer by considering the quantization error to be small. This approximation will be detailed in the following. For the sake of clarity, we consider that the perfect ZF is normalized by $\|(\hat{\mathbf{H}}^{(j)})^{-1}\boldsymbol{e}_k\|$ instead of $\|\mathbf{H}^{-1}\boldsymbol{e}_k\|$. This does not have any impact since it can be shown with a simple first order approximation that

$$\frac{1}{\|\mathbf{H}^{-1}\boldsymbol{e}_k\|} = \frac{1-(\|(\mathbf{H})^{-1}\boldsymbol{e}_k\|-\|(\hat{\mathbf{H}}^{(j)})^{-1}\boldsymbol{e}_k\|)+o(\|(\mathbf{H})^{-1}\boldsymbol{e}_k\|-\|(\hat{\mathbf{H}}^{(j)})^{-1}\boldsymbol{e}_k\|)}{\|(\hat{\mathbf{H}}^{(j)})^{-1}\boldsymbol{e}_k\|}. \qquad (1)$$

Inserting this approximation in the Taylor expansion leads to an additional error term in the direction $\boldsymbol{u}_k^{\star}$ which is orthogonal to the channel vector. Hence, we will omit for the sake of clarity to mention this aspect in the following.

Furthermore, we denote for simplicity by σ^2 the largest variance of the quantization error, i.e., $\sigma^2 \triangleq P^{-\min_{i,j} A_i^{(j)}}$.

Lemma 9. In the BC with distributed CSIT introduced previously (with positive CSIT scaling coefficients),

$$\Delta R_i = \mathrm{E}[\log_2(1+\sum_{k\neq i} P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_k^{\mathrm{DCSI}}|^2)] + O(1) \qquad (2)$$

with the vector $\boldsymbol{a}_k^{\mathrm{DCSI}}$ defined as

$$\{\boldsymbol{a}_k^{\mathrm{DCSI}}\}_j = \{\boldsymbol{u}_k^\star\}_j + \left\{\mathbf{H}^{-1}\boldsymbol{\Delta}^{(j)}\frac{\mathbf{H}^{-1}\boldsymbol{e}_k}{\|\mathbf{H}^{-1}\boldsymbol{e}_k\|}\right\}_j, \qquad \forall j \in \{1,\ldots,K\}. \tag{3}$$

Proof. We will denote for clarity by $\Delta R_i^{\mathrm{Tay}}$ the expectation in the RHS of (2) From the conditional expectation property, we can write

$$\Delta R_i = \Pr(\Omega)\mathrm{E}_\Omega[\log_2(1+\sum_{k\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_k^{\mathrm{DCSI}}|^2)] + \Pr(\bar{\Omega})\mathrm{E}_{\bar{\Omega}}[\log_2(1+\sum_{k\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_k^{\mathrm{DCSI}}|^2)] \tag{4}$$

where we have defined the set Ω as

$$\Omega \triangleq \left\{\left(\mathbf{H}, \{\boldsymbol{\Delta}^{(j)}\}_j\right) \middle| \|\boldsymbol{u}_i^{(j)} - (\boldsymbol{u}_i^\star + \boldsymbol{a}_i^{(j)})\| \leq C\sigma^{5/4}, \forall i,j \in \{1,\ldots,K\}\right\}. \tag{5}$$

Hence, the set Ω contains channel realizations where $\boldsymbol{a}_i^{(j)}$ (the first order Taylor approximation) is an "accurate" approximation. We will show what we mean by "accurate" in the following, as it means that considering $\boldsymbol{a}_i^{\mathrm{DCSI}}$ instead of $\boldsymbol{u}_i^{\mathrm{DCSI}}$ leads to a rate difference in the order of $o(1)$. We start by discussing the upperbound part of the equality before turning to the lower bound one.

- Using Jensen's inequality, we write

$$\begin{aligned}
\Delta R_i &\leq \Pr(\Omega)\mathrm{E}_\Omega\left[\log_2\left(1+P\sum_{k\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_k^{\mathrm{DCSI}}|^2\right)\right] \\
&\quad + \Pr(\bar{\Omega})\log_2\left(1+P\mathrm{E}_{\bar{\Omega}}\left[\sum_{k\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_k^{\mathrm{DCSI}}|^2\right]\right) &(6)\\
&\overset{(a)}{\leq} \Pr(\Omega)\mathrm{E}_\Omega\left[\log_2\left(1+P\sum_{k\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_k^{\mathrm{DCSI}}|^2+o(\sigma^2P)\right)\right] \\
&\quad + \Pr(\bar{\Omega})\log_2\left(1+P\frac{\mathrm{E}[\sum_{k\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_k^{\mathrm{DCSI}}|^2]}{\Pr(\bar{\Omega})}\right) &(7)\\
&\leq \Delta R_i^{\mathrm{Tay}} + \Pr(\bar{\Omega})\log_2\left(1+P\frac{\mathrm{E}[\sum_{k\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_k^{\mathrm{DCSI}}|^2]}{\Pr(\bar{\Omega})}\right)+o(1) &(8)\\
&\overset{(b)}{\leq} \Delta R_i^{\mathrm{Tay}} + \Pr(\bar{\Omega})\log_2\left(1+P\frac{K^2}{\Pr(\bar{\Omega})}\right)+o(1) &(9)
\end{aligned}$$

where inequality (a) follows from the definition of the set Ω and inequality (b) from Lemma 7 in Appendix .1.

- Turning to the lower bound, we write

$$\Delta R_i \geq \Pr(\Omega)\mathrm{E}_{\Omega}\left[\log_2\left(1+P\sum_{k\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_k^{\mathrm{DCSI}}|^2\right)\right] \tag{10}$$

$$\geq \Delta R_i^{\mathrm{Tay}} - \Pr(\bar{\Omega})\mathrm{E}_{\bar{\Omega}}\left[\log_2\left(1+P\sum_{k\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_k^{\mathrm{DCSI}}|^2\right)\right] \tag{11}$$

$$\geq \Delta R_i^{\mathrm{Tay}} - \Pr(\bar{\Omega})\sum_{k\neq i}\mathrm{E}_{\bar{\Omega}}\left[\log_2\left(1+P\|\boldsymbol{h}_i\|^2|\|\boldsymbol{a}_k^{\mathrm{DCSI}}\|^2\right)\right]. \tag{12}$$

Using the definition of $\boldsymbol{a}_k^{\mathrm{DCSI}}$ in (3), we can write that

$$\|\boldsymbol{a}_k^{\mathrm{DCSI}}\|^2 = \sum_{j=1}^{K}|\{\boldsymbol{a}_k^{(j)}\}_{\ell}|^2 \tag{13}$$

$$\leq \|\mathbf{H}^{-1}\|_{\mathrm{F}}^2\sum_{j=1}^{K}\|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}}^2. \tag{14}$$

Inserting (18) and using one more time Lemma 7 in Appendix .1, we can write

$$\Delta R_i \geq \Delta R_i^{\mathrm{Tay}} - \Pr(\bar{\Omega})\sum_{k\neq i}\mathrm{E}_{\bar{\Omega}}\left[\log_2\left(P\sum_{\ell=1}^{K}|\{\boldsymbol{a}_k^{\mathrm{DCSI}}\}_{\ell}|^2\right)\right] + O(1) \tag{15}$$

$$\geq \Delta R_i^{\mathrm{Tay}} - \Pr(\bar{\Omega})\sum_{k\neq i}\log_2\left(P\sum_{\ell=1}^{K}\mathrm{E}_{\bar{\Omega}}[\sum_{j=1}^{K}\|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}}^2]\right) + O(1) \tag{16}$$

$$\geq \Delta R_i^{\mathrm{Tay}} - \Pr(\bar{\Omega})\sum_{k\neq i}\log_2\left(P\sum_{j=1}^{K}\frac{\mathrm{E}[\|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}}^2]}{\Pr(\bar{\Omega})}\right) + O(1) \tag{17}$$

$$\dot{\geq}\ \Delta R_i^{\mathrm{Tay}} - \Pr(\bar{\Omega})\sum_{k\neq i}\log_2(P\frac{P^{-\min_{i,j}A_i^{(j)}}}{\Pr(\bar{\Omega})}). \tag{18}$$

For both cases, it remains to compute the probability of the space Ω. Using the resolvent inequality two times, we can write

$$\begin{aligned}\mathbf{H}^{-1} - (\hat{\mathbf{H}}^{(j)})^{-1} &= -(\hat{\mathbf{H}}^{(j)})^{-1}\boldsymbol{\Delta}^{(j)}\mathbf{H}^{-1} && (19)\\ &= -\mathbf{H}^{-1}\boldsymbol{\Delta}^{(j)}\mathbf{H}^{-1} - ((\hat{\mathbf{H}}^{(j)})^{-1} - \mathbf{H}^{-1})\boldsymbol{\Delta}^{(j)}\mathbf{H}^{-1} && (20)\\ &= -\mathbf{H}^{-1}\boldsymbol{\Delta}^{(j)}\mathbf{H}^{-1} + (\hat{\mathbf{H}}^{(j)})^{-1}\boldsymbol{\Delta}^{(j)}\mathbf{H}^{-1}\boldsymbol{\Delta}^{(j)}\mathbf{H}^{-1}. && (21)\end{aligned}$$

It then follows that

$$\begin{aligned}\boldsymbol{u}_k^{(j)} &= \boldsymbol{u}_k^{\star} + \mathbf{H}^{-1}\boldsymbol{\Delta}^{(j)}\frac{\mathbf{H}^{-1}\boldsymbol{e}_k}{\|\mathbf{H}^{-1}\boldsymbol{e}_k\|} - (\hat{\mathbf{H}}^{(j)})^{-1}\boldsymbol{\Delta}^{(j)}\mathbf{H}^{-1}\boldsymbol{\Delta}^{(j)}\frac{\mathbf{H}^{-1}\boldsymbol{e}_k}{\|\mathbf{H}^{-1}\boldsymbol{e}_k\|} && (22)\\ &= \boldsymbol{u}_k^{\star} + \mathbf{H}^{-1}\boldsymbol{\Delta}^{(j)}\frac{\mathbf{H}^{-1}\boldsymbol{e}_k}{\|\mathbf{H}^{-1}\boldsymbol{e}_k\|} - \boldsymbol{\varepsilon}_i^{(j)} && (23)\end{aligned}$$

where we have introduced

$$\boldsymbol{\varepsilon}_i^{(j)} \triangleq (\hat{\mathbf{H}}^{(j)})^{-1}\boldsymbol{\Delta}^{(j)}\mathbf{H}^{-1}\boldsymbol{\Delta}^{(j)}\frac{\mathbf{H}^{-1}\boldsymbol{e}_k}{\|\mathbf{H}^{-1}\boldsymbol{e}_k\|}. \quad (24)$$

The definition of the set Ω can be rewritten with the new notation as

$$\Omega \triangleq \left\{ \left(\mathbf{H}, \{\boldsymbol{\Delta}^{(j)}\}_j\right) \middle| \|\boldsymbol{\varepsilon}_i^{(j)}\| \leq C\sigma^{5/4}, \forall i, j \in \{1, \ldots, K\} \right\}. \quad (25)$$

It can easily be shown that

$$\begin{aligned}\|\boldsymbol{\varepsilon}_i^{(j)}\| &\leq \|(\hat{\mathbf{H}}^{(j)})^{-1}\|_{\mathrm{F}}\|\mathbf{H}^{-1}\|_{\mathrm{F}}\|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}}^2 && (26)\\ &\leq \sqrt{\frac{K}{\lambda_{\min}((\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)})}\frac{K}{\lambda_{\min}(\mathbf{H}^{\mathrm{H}}\mathbf{H})}}\|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}}^2. && (27)\end{aligned}$$

We can then write

$$\begin{aligned}&\Pr\left(\|\boldsymbol{\varepsilon}_i^{(j)}\| > K^3\sigma^{5/4}\right)\\ &= \Pr\left((\sqrt{\lambda_{\min}((\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}))^{-1}(\lambda_{\min}(\mathbf{H}^{\mathrm{H}}\mathbf{H}))^{-1}}\|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}}^2 > \sigma^{5/4}\right) && (28)\\ &\overset{(a)}{\leq} \Pr\left(\lambda_{\min}((\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}) < \sigma^{1/2}\right) + \Pr\left(\lambda_{\min}(\mathbf{H}^{\mathrm{H}}\mathbf{H}) < \sigma^{1/2}\right)\\ &\qquad + \Pr\left(\|\boldsymbol{\Delta}^{(j)}\|_{\mathrm{F}}^2 > \sigma^{7/4}\right). && (29)\end{aligned}$$

where inequality (a) follows from Lemma 3 in Appendix .1.

1. The matrix $(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}$ is a Wishart matrix of size $K-1$ with K degrees-of-freedom from which the probability distribution of the minimal eigenvalue is known from [124, Theorem 5.4] to be exponential and given by

$$\Pr\left(\lambda_{\min}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i) \leq x\right) = 1 - \exp\left(-\frac{(K-1)}{2}x\right). \tag{30}$$

   Hence, with $x = \sigma^{\frac{1}{4}}$ and σ being small, it gives after a Taylor expansion

$$\Pr\left(\lambda_{\min}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i) < \sigma^{\frac{4}{3}}\right) = \frac{K-1}{2}\sigma^{\frac{1}{4}} + o(\sigma^{\frac{1}{4}}). \tag{31}$$

2. The matrix $\hat{\mathbf{H}}^{(j)}$ has the same distribution as $\mathbf{H}_i$ such that we can use the same upperbound with the probability distribution of $(\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}$.

3. The matrix $\mathbf{\Delta}^{(j)}$ has its elements i.i.d. complex Gaussian with the variance smaller than σ^2. After normalizing by σ^2, the squared Frobenius norm can be upperbounded by a Chi-square random variable with $(K-1)K$ degrees-of-freedom. The tail of the probability distribution can be upperbounded with the help of Lemma 4 in Appendix .1 with $n = K(K-1)$ and $x = \sigma^{-1/4}$. For σ sufficiently small, this gives

$$\Pr\left(\left\|\mathbf{\Delta}_i^{(j)}\right\|_{\mathrm{F}}^2 > \sigma^{7/4}\right) \leq \exp\left(-\sigma^{-1/4}/2 + K(K-1)\log\left(\frac{e\sigma^{-1/4}}{K(K-1)}\right)\right) \tag{32}$$

   which decrease exponentially fast with σ.

In total this gives

$$\Pr\left(\|\boldsymbol{\varepsilon}_i^{(j)}\| > K^3\sigma^{5/4}\right) \leq (K-1)\sigma^{\frac{1}{4}} + o(\sigma^{\frac{1}{4}}). \tag{33}$$

A lower bound for the probability of Ω is easily obtained by writing

$$\begin{aligned}
&\Pr\left(\|\boldsymbol{\varepsilon}_i^{(j)}\| > K^3\sigma^{5/4}\right) \\
&= \Pr\left((\sqrt{\lambda_{\min}((\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}))^{-1}(\lambda_{\min}(\mathbf{H}^{\mathrm{H}}\mathbf{H}))^{-1}}\|\mathbf{\Delta}^{(j)}\|_{\mathrm{F}}^2 > \sigma^{-3/4}\right) &(34)\\
&\geq \Pr\left(\sqrt{(\lambda_{\min}((\hat{\mathbf{H}}^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}^{(j)}))^{-1}} > \sigma^{-3/4} | \lambda_{\min}(\mathbf{H}^{\mathrm{H}}\mathbf{H}))^{-1} \geq 1, \|\mathbf{\Delta}^{(j)}\|_{\mathrm{F}}^2 \geq 1\right). &(35)
\end{aligned}$$

Inserting these two bounds of $\Pr(\Omega)$ inside the rate expressions concludes the proof. □

Lemma 10. Considering the BC with distributed CSIT introduced previously, it holds

$$\mathrm{E}[\log_2(|\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{a}_i^{(j)}\|^2)] \doteq -\min_{\ell} A_i^{(j)} \log_2(P), \qquad \forall i,j,k. \tag{36}$$

Proof. Our first step is to prove that the expected value does not depend on the value of k (similarly, it can be shown to be independent on the value of i). To show this result, we will use the permutation matrix $\boldsymbol{\Pi}_1 \in \{0,1\}^{K\times K}$ (which permutes the columns when multiplied from the right of the matrix, see [125] for more results). We can write that

$$\begin{aligned}
\boldsymbol{e}_k^{\mathrm{H}}\mathbf{H}^{-1}\boldsymbol{\Delta}\mathbf{H}^{-1}\boldsymbol{e}_i &= (\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{\Pi}_1)\boldsymbol{\Pi}_1^{-1}\mathbf{H}^{-1}\boldsymbol{\Delta}\boldsymbol{\Pi}_1(\boldsymbol{\Pi}_1^{-1}\mathbf{H}^{-1})\boldsymbol{e}_i && (37)\\
&= (\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{\Pi}_1)(\mathbf{H}\boldsymbol{\Pi}_1)^{-1}\boldsymbol{\Delta}\boldsymbol{\Pi}_1(\mathbf{H}\boldsymbol{\Pi}_1)^{-1}\boldsymbol{e}_i && (38)\\
&= \boldsymbol{e}_{k'}^{\mathrm{H}}\mathbf{H}'^{-1}\boldsymbol{\Delta}'\mathbf{H}'^{-1}\boldsymbol{e}_i && (39)\\
& && (40)
\end{aligned}$$

where $\mathbf{H}'$ has the same distribution as $\mathbf{H}$ and $\boldsymbol{\Delta}'$ has the same distribution as $\boldsymbol{\Delta}$. It follows from this symmetry property that

$$\mathrm{E}[\log_2(|\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{a}_i^{(j)}\|^2)] = \mathrm{E}[\log_2\left(\frac{\|\boldsymbol{a}_i^{(j)}\|^2}{K}\right)]. \tag{41}$$

We prove first that the RHS is a lower bound and then that it is also an upper bound. Using basic algebra, we can show that

$$\begin{aligned}
\|\boldsymbol{a}_i^{(j)}\|^2 &\geq \left|\frac{\boldsymbol{h}_\ell^{\mathrm{H}}}{\|\boldsymbol{h}_\ell\|}\boldsymbol{a}_i^{(j)}\right|^2 && (42)\\
&= \frac{1}{\|\boldsymbol{h}_\ell\|^2}\frac{1}{\|\mathbf{H}^{-1}\boldsymbol{e}_i\|^2}(\sigma_\ell^{(j)})^2\left|(\boldsymbol{\delta}^{(j)})^{\mathrm{H}}\mathbf{H}^{-1}\boldsymbol{e}_i\right|^2. && (43)
\end{aligned}$$

The two vectors forming the inner product are independent and $\boldsymbol{\delta}_i^{(j)}$ is isotropically distributed such that the inner product is a Beta random variable whose expected logarithm is finite. Furthermore, we can write this relation for any channel vector $\boldsymbol{h}_\ell$, which concludes the proof of the lower bound. The upper bound follows easily using the same properties of the norm as in the previous derivations. □

We have now all the elements necessary to derive easily the main theorem.

.3.2 Proof of the Theorem

Proof of Theorem 4. It is now clear that the DoF is determined by the rate loss expressed as

$$\Delta R_i = \mathrm{E}\left[\log_2\left(1+\sum_{j\neq i} P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{DCSI}}|^2\right)\right]. \tag{44}$$

Since all the CSIT scaling coefficient are positive, we can use Lemma 9 to write that

$$\Delta R_i = \mathrm{E}\left[\log_2\left(1+\sum_{j\neq i} P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_j^{\mathrm{DCSI}}|^2\right)\right] + O(1). \tag{45}$$

We now prove the results by deriving first an upper bound and then a matching lower bound.

1. Focusing on the first expectation, we can write

$$\begin{aligned}
\Delta R_i &\leq \mathrm{E}\left[\log_2\left(1+P\sum_{j\neq i}\|\boldsymbol{h}_i\|^2\|\boldsymbol{a}_j^{\mathrm{DCSI}}\|^2\right)\right] + O(1) && (46)\\
&\doteq \mathrm{E}\left[\log_2\left(\sum_{j\neq i} P\|\boldsymbol{a}_j^{(\mathrm{DCSI})}\|^2\right)\right] && (47)\\
&\doteq \mathrm{E}\left[\log_2\left(\sum_{j\neq i}\sum_{k=1}^{K} P|\{\boldsymbol{a}_j^{(k)}\}_k|^2\right)\right] && (48)\\
&\leq \mathrm{E}\left[\log_2\left(\sum_{j\neq i}\sum_{k=1}^{K} P\|\boldsymbol{a}_j^{(k)}\|^2\right)\right] && (49)\\
&\doteq \log_2\left(P\max_{k,j}(\sigma_k^{(j)})^2\right) && (50)\\
& && (51)
\end{aligned}$$

where the last equality follows from Lemma 10 and the fact that $\mathrm{E}[\log_2(\|\mathbf{H}^{-1}\|_{\mathrm{F}}^2)]$ is finite [124].

2. Turning to the lower bound, we write

$$\Delta R_i = \mathrm{E}\left[\log_2\left(1 + P\sum_{k\neq i}|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_j^{\mathrm{DCSI}}|^2\right)\right] + O(1) \tag{52}$$

$$\geq \mathrm{E}\left[\log_2\left(1 + P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_\ell^{\mathrm{DCSI}}|^2\right)\right] + O(1) \tag{53}$$

$$\geq \mathrm{E}\left[\log_2\left(P|\sum_{j=1}^{K}\{\mathbf{H}\}_{ij}\{\boldsymbol{a}_\ell^{(j)}\}_j|^2\right)\right] + O(1). \tag{54}$$

for any given index $\ell \neq i$. From the definition of $\{\boldsymbol{a}_\ell^{(j)}\}_j$, every element of that vector has its phase uniformly distributed and independent of the phase of the other elements of the beamformer. This comes from the distribution of the estimation error being invariant by multiplication by any unit norm complex number. Hence, we can rewrite (56) as

$$\Delta R_i \geq \mathrm{E}\left[\log_2\left(\left|\sum_{j=1}^{K}\{\mathbf{H}\}_{ij}\{\boldsymbol{a}_\ell^{(j)}\}_j \exp(\imath\theta_j)\right|^2\right)\right] + O(1) \tag{55}$$

with the θ_j being i.i.d. over $[0, 2\pi)$. Applying Lemma 1 iteratively, yields

$$\mathrm{E}[\log_2(|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_\ell^{\mathrm{DCSI}}|^2)] = \mathrm{E}\left[\log_2\left(\max_k\left(\left|\{\mathbf{H}\}_{ij}\{\boldsymbol{a}_\ell^{(j)}\}_j\right|^2\right)\right)\right]. \tag{56}$$

The proof concludes by using Lemma 10. □

.4 Proof of Proposition 9

The decomposition in (5.20) is obtained after a Taylor expansion with the CSIT qualities σ_i considered to be small, i.e., $o(1)$. The derivation of this expression is detailed in Appendix .5.1. We consider for ease of notation and without loss of generality that $i = K$. From (15) and using in addition that $\boldsymbol{h}_i$, $\boldsymbol{\delta}_i^{(j)}$ and $\mathbf{H}_i$ are independent, we obtain

$$\begin{aligned}
&\mathrm{E}[|\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{a}_i^{(j)}|^2] = \\
&+ \mathrm{E}[\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}(\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)})^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{e}_k] + (\sigma_i^{(j)})^2\mathrm{E}[\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{e}_k] \\
&+ \mathrm{E}[\boldsymbol{e}_k^{\mathrm{H}}\mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}(\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)})^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\boldsymbol{e}_k] \qquad (1)
\end{aligned}$$

We compute now each term of (1) separately.

1. Starting with the first term, we have

$$\begin{aligned}
&\mathrm{E}[\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}(\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)})^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{e}_k] \\
&= \sum_{\ell,\ell'}\mathrm{E}[\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)}\boldsymbol{e}_\ell\boldsymbol{e}_\ell^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\boldsymbol{e}_{\ell'}\boldsymbol{e}_{\ell'}^{\mathrm{H}}(\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)})^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{e}_k] \qquad (2)\\
&\overset{(a)}{=} \sum_{\ell,\ell'}(\sigma_\ell^{(j)})(\sigma_{\ell'}^{(j)})\mathrm{E}_{\mathbf{H}_i}[\boldsymbol{e}_\ell^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\boldsymbol{e}_{\ell'}\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp} \\
&\qquad\qquad \cdot \mathrm{E}_{\boldsymbol{\Delta}_i^{(j)}}[\boldsymbol{\Delta}_i^{(j)}\boldsymbol{e}_\ell\boldsymbol{e}_{\ell'}^{\mathrm{H}}(\boldsymbol{\Delta}_i^{(j)})^{\mathrm{H}}]\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{e}_k] \qquad (3)\\
&\overset{(b)}{=} \sum_{\ell}(\sigma_\ell^{(j)})^2\mathrm{E}[\boldsymbol{e}_\ell^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\boldsymbol{e}_\ell\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{e}_k] \qquad (4)
\end{aligned}$$

where the two first equalities are obtained via basic algebra, equality (a) follows from the independence of the channel and the estimation error, and equality (b) from the estimation errors being uncorrelated.

To compute the expectation, we will exploit the fact that the channel elements are i.i.d.. Hence, let us consider two permutation matrices $\boldsymbol{\Pi}_1 \in \mathbb{N}^{(K-1)\times(K-1)}$ and $\boldsymbol{\Pi}_2 \in \mathbb{N}^{K\times K}$. We also denote by π_1 and π_2 the permutation functions associated with $\boldsymbol{\Pi}_1$ and $\boldsymbol{\Pi}_2$, respectively. Note that we will exploit in the following that the inverse of a permutation matrix is its transpose, i.e., $\boldsymbol{\Pi}_i^{\mathrm{T}} = \boldsymbol{\Pi}_i^{-1}$ for $i = 1, 2$ [125]. It follows that

$$\begin{aligned}
&\mathrm{E}[\boldsymbol{e}_\ell^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\boldsymbol{e}_\ell \cdot \boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{e}_k] \\
&\overset{(a)}{=} \mathrm{E}[\boldsymbol{e}_\ell^{\mathrm{H}}\boldsymbol{\Pi}_1\boldsymbol{\Pi}_1^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\boldsymbol{\Pi}_2^{\mathrm{H}}\boldsymbol{\Pi}_2\mathbf{H}_i\right)^{-1}\boldsymbol{\Pi}_1\boldsymbol{\Pi}_1^{\mathrm{H}}\boldsymbol{e}_\ell \\
&\qquad \cdot \boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{\Pi}_2^{\mathrm{H}}\boldsymbol{\Pi}_2\left(\mathbf{I} - \mathbf{H}_i\boldsymbol{\Pi}_1\boldsymbol{\Pi}_1^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\boldsymbol{\Pi}_2^{\mathrm{H}}\boldsymbol{\Pi}_2\mathbf{H}_i\right)^{-1}\boldsymbol{\Pi}_1\boldsymbol{\Pi}_1^{\mathrm{H}}\mathbf{H}_i^{\mathrm{H}}\right)\boldsymbol{\Pi}_2^{\mathrm{H}}\boldsymbol{\Pi}_2\boldsymbol{e}_k] \qquad (5)\\
&= \mathrm{E}[(\boldsymbol{\Pi}_1^{\mathrm{H}}\boldsymbol{e}_\ell)^{\mathrm{H}}\left((\boldsymbol{\Pi}_2\mathbf{H}_i\boldsymbol{\Pi}_1)^{\mathrm{H}}\boldsymbol{\Pi}_2\mathbf{H}_i\boldsymbol{\Pi}_1\right)^{-1}\boldsymbol{\Pi}_1^{\mathrm{H}}\boldsymbol{e}_\ell(\boldsymbol{\Pi}_2\boldsymbol{e}_k)^{\mathrm{H}} \\
&\qquad \cdot \left(\mathbf{I} - \boldsymbol{\Pi}_2\mathbf{H}_i\boldsymbol{\Pi}_1\left((\boldsymbol{\Pi}_2\mathbf{H}_i\boldsymbol{\Pi}_1)^{\mathrm{H}}\boldsymbol{\Pi}_2\mathbf{H}_i\boldsymbol{\Pi}_1\right)^{-1}(\boldsymbol{\Pi}_2\mathbf{H}_i\boldsymbol{\Pi}_1)^{\mathrm{H}}\right)\boldsymbol{\Pi}_2\boldsymbol{e}_k] \qquad (6)\\
&\overset{(b)}{=} \mathrm{E}[\boldsymbol{e}_\ell'^{\mathrm{H}}\left(\mathbf{H}_i'^{\mathrm{H}}\mathbf{H}_i'\right)^{-1}\boldsymbol{e}_\ell' \cdot \boldsymbol{e}_k'^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i'}^{\perp}\boldsymbol{e}_k'] \qquad (7)\\
&\overset{(c)}{=} \mathrm{E}[(\boldsymbol{e}_\ell'^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\boldsymbol{e}_\ell' \cdot \boldsymbol{e}_k'^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{e}_k'] \qquad (8)
\end{aligned}$$

where equality (a) follows simply from the insertion of the permutation matrices in the inverse, equality (b) is obtained after defining $\mathbf{H}_i' \triangleq$

$\mathbf{\Pi}_2\mathbf{H}_i\mathbf{\Pi}_1$, $\ell' = \pi_1^{-1}(\ell)$ and $k' = \pi_2(k)$. Equality (c) holds because the channel elements are i.i.d. such that the distribution of $\mathbf{H}'_i$ is the same as the distribution of $\mathbf{H}_i$.

The permutations π_1 and π_2 can be chosen arbitrarily which implies that the expectation in (8) is independent of k and ℓ. Hence, we can write

$$\sum_{\ell=1}^{K-1} (\sigma_\ell^{(j)})^2 \mathrm{E}\left[\frac{1}{K-1}\mathrm{tr}\left(\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\right)\frac{1}{K}\mathrm{tr}\left(\mathbf{\Pi}_{\mathbf{H}_i}^{\perp}\right)\right] \tag{9}$$

$$\overset{(a)}{=} \frac{1}{(K-1)K}\sum_{\ell=1}^{K-1} (\sigma_\ell^{(j)})^2 \mathrm{E}\left[\mathrm{tr}\left(\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\right)\right] \tag{10}$$

$$\overset{(b)}{=} \frac{1}{K}\sum_{\ell=1}^{K-1} (\sigma_\ell^{(j)})^2 \tag{11}$$

with equality (a) obtained because the matrix in the trace is the matrix of a projector over an hyperplane and equality (b) obtained from $\mathrm{E}\left[\mathrm{tr}\left(\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\right)\right] = K-1$ [126, Lemma 2.10].

2. The second term of (1) is easily computed from the results obtained above to be equal to

$$(\sigma_i^{(j)})^2 \mathrm{E}[\boldsymbol{e}_k^{\mathrm{H}}\mathbf{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{e}_k] = (\sigma_i^{(j)})^2 \mathrm{E}\left[\frac{1}{K}\mathrm{tr}\left(\mathbf{\Pi}_{\mathbf{H}_i}^{\perp}\right)\right] \tag{12}$$

$$= \frac{(\sigma_i^{(j)})^2}{K} \tag{13}$$

3. The third term of (1) is computed using the same properties as for the

previous terms, as described below.

$$
\begin{aligned}
&\mathrm{E}[\boldsymbol{e}_k^{\mathrm{H}}\mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}(\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)})^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\boldsymbol{e}_k] \\
&= \sum_{\ell=1}^{K}\sum_{\ell'=1}^{K}\mathrm{E}\Big[\boldsymbol{e}_k^{\mathrm{H}}\mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\boldsymbol{e}_\ell\boldsymbol{e}_\ell^{\mathrm{H}}(\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)})^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)}\boldsymbol{e}_{\ell'} \\
&\qquad\qquad \cdot \boldsymbol{e}_{\ell'}^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\boldsymbol{e}_k\Big] \qquad (14) \\
&= \sum_{\ell=1}^{K}\sum_{\ell'=1}^{K}\sigma_\ell^{(j)}\sigma_{\ell'}^{(j)}\mathrm{E}_{\mathbf{H}_i}[\boldsymbol{e}_k^{\mathrm{H}}\mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\boldsymbol{e}_\ell \\
&\qquad\qquad \cdot \mathrm{tr}\left(\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\mathrm{E}_{\boldsymbol{\Delta}_i^{(j)}}[\boldsymbol{\Delta}_i^{(j)}\boldsymbol{e}_{\ell'}\boldsymbol{e}_\ell^{\mathrm{H}}(\boldsymbol{\Delta}_i^{(j)})^{\mathrm{H}}]\right)\boldsymbol{e}_{\ell'}^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\boldsymbol{e}_k] \\
&\qquad\qquad (15) \\
&= \sum_{\ell=1}^{K}(\sigma_\ell^{(j)})^2\mathrm{E}[\boldsymbol{e}_k^{\mathrm{H}}\mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\boldsymbol{e}_\ell\cdot\boldsymbol{e}_\ell^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\boldsymbol{e}_k] \qquad (16)
\end{aligned}
$$

Similarly to the derivation done for the first term, we introduce the permutation matrices $\boldsymbol{\Pi}_1$ and $\boldsymbol{\Pi}_2$ to exploit the invariance of the distribution of $\mathbf{H}_i$. Hence,

$$
\begin{aligned}
&\mathrm{E}[\boldsymbol{e}_k^{\mathrm{H}}\mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\boldsymbol{e}_\ell\cdot\boldsymbol{e}_\ell^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\boldsymbol{e}_k] \\
&= \mathrm{E}[\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{\Pi}_2^{\mathrm{H}}\boldsymbol{\Pi}_2\mathbf{H}_i\boldsymbol{\Pi}_1\boldsymbol{\Pi}_1^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\boldsymbol{\Pi}_2^{\mathrm{H}}\boldsymbol{\Pi}_2\mathbf{H}_i\right)^{-1}\boldsymbol{\Pi}_1\boldsymbol{\Pi}_1^{\mathrm{H}}\boldsymbol{e}_\ell \\
&\qquad \cdot \boldsymbol{e}_\ell^{\mathrm{H}}\boldsymbol{\Pi}_1\boldsymbol{\Pi}_1^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\boldsymbol{\Pi}_2^{\mathrm{H}}\boldsymbol{\Pi}_2\mathbf{H}_i\right)^{-1}\boldsymbol{\Pi}_1\boldsymbol{\Pi}_1^{\mathrm{H}}\mathbf{H}_i^{\mathrm{H}}\boldsymbol{\Pi}_2^{\mathrm{H}}\boldsymbol{\Pi}_2\boldsymbol{e}_k] \qquad (17) \\
&= \mathrm{E}[\boldsymbol{e}_{k'}^{\mathrm{H}}\mathbf{H}_i'\left(\mathbf{H}_i'^{\mathrm{H}}\mathbf{H}_i'\right)^{-1}\boldsymbol{e}_{\ell'}\cdot\boldsymbol{e}_{\ell'}^{\mathrm{H}}\left(\mathbf{H}_i'^{\mathrm{H}}\mathbf{H}_i'\right)^{-1}\mathbf{H}_i'^{\mathrm{H}}\boldsymbol{e}_k'] \qquad (18)
\end{aligned}
$$

where we have defined $\mathbf{H}_i' \triangleq \boldsymbol{\Pi}_2\mathbf{H}_i\boldsymbol{\Pi}_1$, $\ell' = \pi_1^{-1}(\ell)$ and $k' = \pi_2(k)$ in the same way as in the computation of the first term. We can choose the permutations arbitrarily, implying that the expectation is equal

for all values of ℓ an k. Thus, we can continue from (16) to write

$$\sum_{\ell=1}^{K}(\sigma_\ell^{(j)})^2\mathrm{E}[\boldsymbol{e}_k^{\mathrm{H}}\mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\boldsymbol{e}_\ell\cdot\boldsymbol{e}_\ell^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\boldsymbol{e}_k] \tag{19}$$

$$=\sum_{\ell=1}^{K}\frac{(\sigma_\ell^{(j)})^2}{K(K-1)}\mathrm{E}\left[\sum_{\ell'=1}^{K-1}\sum_{k'=1}^{K}\boldsymbol{e}_{k'}^{\mathrm{H}}\mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\boldsymbol{e}_{\ell'}\boldsymbol{e}_{\ell'}^{\mathrm{H}}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\boldsymbol{e}_{k'}\right] \tag{20}$$

$$=\sum_{\ell=1}^{K}(\sigma_\ell^{(j)})^2\frac{1}{K(K-1)}\mathrm{E}\left[\mathrm{tr}\left(\mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-2}\mathbf{H}_i^{\mathrm{H}}\right)\right] \tag{21}$$

$$=\frac{1}{K}\sum_{\ell=1}^{K-1}(\sigma_\ell^{(j)})^2. \tag{22}$$

Putting the three terms together concludes the proof.

.5 Proof of Theorem 8

.5.1 Preliminaries

We prove in this appendix two side results that we will use in the main proofs. Firstly, we will show that our results remain unchanged if we consider a quantization model where

$$\hat{\boldsymbol{h}}_i^{(j)}=\boldsymbol{h}_i+\sigma_i^{(j)}\boldsymbol{\delta}_i^{(j)} \tag{1}$$

instead of the model given in (3.15). Note that this step is only helpful for the sake of clarity as using (1) leads to shorter expressions.

We then study how the precoding at a given TX depends on the quality of its CSIT. This analysis being a bit lengthy, we write it here to avoid breaking the flow of the main proof.

Simplification of the CSIT model

We start by introducing the matrix $\boldsymbol{\Omega}^{(j)}$ as

$$\boldsymbol{\Omega}^{(j)}\triangleq\sqrt{\mathbf{I}_K-\left(\boldsymbol{\Sigma}^{(j)}\right)^2}. \tag{2}$$

and $\boldsymbol{\Omega}_i^{(j)}$ the $(K-1)\times(K-1)$ matrix equal to $\boldsymbol{\Omega}^{(j)}$ once the ith row and the ith columns have been removed.

We can then define the beamformers $\tilde{\boldsymbol{u}}_i^{(j)}$ such that $i, j \in \{1, \ldots, K\}$,

$$\tilde{\boldsymbol{u}}_i^{(j)} \triangleq \left(\mathbf{I}_K - \hat{\mathbf{H}}_i^{(j)} \boldsymbol{\Omega}_i^{(j)} \left((\hat{\mathbf{H}}_i^{(j)} \boldsymbol{\Omega}_i^{(j)})^{\mathrm{H}} \hat{\mathbf{H}}_i^{(j)} \boldsymbol{\Omega}_i^{(j)} \right)^{-1} (\hat{\mathbf{H}}_i^{(j)} \boldsymbol{\Omega}_i^{(j)})^{\mathrm{H}} \right) \frac{\hat{\boldsymbol{h}}_i^{(j)}}{\sqrt{1 - (\sigma_i^{(j)})^2}}. \tag{3}$$

After simplification, we have that

$$\boldsymbol{u}_i^{(j)} = \frac{1}{\sqrt{1 - (\sigma_i^{(j)})^2}} \tilde{\boldsymbol{u}}_i^{(j)}, \qquad \forall i, j \in \{1, \ldots, K\}. \tag{4}$$

We can then directly write that

$$|\boldsymbol{h}_i^{\mathrm{H}} \boldsymbol{u}_k^{\mathrm{DCSI}}|^2 = |\boldsymbol{h}_i^{\mathrm{H}} \tilde{\boldsymbol{u}}_k^{\mathrm{DCSI}}|^2 (1 + o(\sigma_i^{(j)})), \qquad \forall i, k \in \{1, \ldots, K\}. \tag{5}$$

The multiplication by $1 + o(\sigma_i^{(j)})$ does not change the results given in the rest of these works such that we can use $\hat{\boldsymbol{h}}_i^{(j)} / \sqrt{1 - (\sigma_i^{(j)})^2}$ instead of $\hat{\boldsymbol{h}}_i^{(j)}$ without any change on our results. Furthermore, it holds that

$$\frac{\hat{\boldsymbol{h}}_i^{(j)}}{\sqrt{1 - (\sigma_i^{(j)})^2}} = \boldsymbol{h}_i + \frac{\sigma_i^{(j)}}{\sqrt{1 - (\sigma_i^{(j)})^2}} \boldsymbol{\delta}_i^{(j)} \tag{6}$$

$$\stackrel{(a)}{=} \boldsymbol{h}_i + \sigma_i^{(j)} \left(1 + \frac{1}{2} (\sigma_i^{(j)})^2 + o((\sigma_i^{(j)})^2) \right) \boldsymbol{\delta}_i^{(j)} \tag{7}$$

$$= \boldsymbol{h}_i + \sigma_i'^{(j)} \boldsymbol{\delta}_i^{(j)} \tag{8}$$

where we have used a Taylor expansion to obtain (a) and we have defined $\sigma_i'^{(j)} \triangleq \sigma_i^{(j)} + x_i^{(j)}$ with $x_i^{(j)} = O((\sigma_i^{(j)})^3)$. Once more the additional term obtained is negligible as the accuracy of the CSIT increases and thus does not impact our approach.

Putting the two results together, we have shown that we can use (1) to model the CSIT errors in the following proofs.

Precoding at TX j

As a preliminary step, we discuss the precoding at TX j based on the imperfect channel estimate $\hat{\mathbf{H}}^{(j)}$. We consider without loss of generality the design of the ith beamformer $\boldsymbol{u}_i^{(j)}$. We define

$$\boldsymbol{\Delta}_i^{(j)} \triangleq \left[\boldsymbol{\delta}_1^{(j)}, \ldots, \boldsymbol{\delta}_{i-1}^{(j)}, \boldsymbol{\delta}_{i+1}^{(j)}, \ldots, \boldsymbol{\delta}_K^{(j)} \right], \tag{9}$$

$$\boldsymbol{\Sigma}_i^{(j)} \triangleq \mathrm{diag}\left(\left[\sigma_1^{(j)}, \ldots, \sigma_{i-1}^{(j)}, \sigma_{i+1}^{(j)}, \ldots, \sigma_K^{(j)} \right] \right) \tag{10}$$

such that $\hat{\mathbf{H}}_i^{(j)} = \mathbf{H}_i + \boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)}$. Note that the random matrices $\boldsymbol{\Delta}_i^{(j)}$ and $\mathbf{H}_i$ are independent. We further define for ease of notation

$$\boldsymbol{\Phi}_i^{(j)} \triangleq (\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_i^{(j)} - \mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i. \tag{11}$$

Applying two times successively the resolvent inequality in Lemma 2 with $\mathbf{A} = (\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_i^{(j)}$ and $\mathbf{B} = \mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i$, we get

$$\begin{aligned}((\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_i^{(j)})^{-1} &= (\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1} - \\ &\quad ((\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_i^{(j)})^{-1}\left[(\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_i^{(j)} - \mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right](\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1} \quad (12)\\ &= (\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1} - (\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\boldsymbol{\Phi}_i^{(j)}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1} + \boldsymbol{\Theta}_i^{(j)} \quad (13)\end{aligned}$$

where we have defined

$$\boldsymbol{\Theta}_i^{(j)} \triangleq ((\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_i^{(j)})^{-1}\boldsymbol{\Phi}_i^{(j)}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\boldsymbol{\Phi}_i^{(j)}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}. \tag{14}$$

Inserting (13) in the definition of $\boldsymbol{u}_i^{(j)}$ in (3.17) and using $\hat{\mathbf{H}}_i^{(j)} = \mathbf{H}_i + \boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)}$, we can write

$$\boldsymbol{u}_i^{(j)} = \boldsymbol{u}_i^{\star} + \boldsymbol{a}_i^{(j)} + \boldsymbol{\Xi}_i^{(j)}\boldsymbol{h}_i^{(j)} + \sigma_i^{(j)}\boldsymbol{\Omega}_i^{(j)}\boldsymbol{\delta}_i^{(j)} \tag{15}$$

where we have introduced

$$\begin{aligned}\boldsymbol{a}_i^{(j)} \triangleq\ & \boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\boldsymbol{h}_i - \sigma_i^{(j)}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{\delta}_i^{(j)} \\ & + \mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}(\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)})^{\mathrm{H}}\boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\boldsymbol{h}_i\end{aligned} \tag{16}$$

as well as

$$\begin{aligned}\boldsymbol{\Xi}_i^{(j)} \triangleq\ & (\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}\boldsymbol{\Theta}_i^{(j)}\hat{\mathbf{H}}_i^{(j)} + (\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)})^{\mathrm{H}}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)} \\ & - (\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)})^{\mathrm{H}}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\boldsymbol{\Phi}_i^{(j)}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\hat{\mathbf{H}}_i^{(j)} \\ & - (\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\boldsymbol{\Phi}_i^{(j)}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)} \\ & - (\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)}(\boldsymbol{\Delta}_i^{(j)}\boldsymbol{\Sigma}_i^{(j)})^{\mathrm{H}}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\hat{\mathbf{H}}_i^{(j)}\end{aligned} \tag{17}$$

and finally

$$\boldsymbol{\Omega}_i^{(j)} \triangleq \left[\boldsymbol{\Pi}_{\hat{\mathbf{H}}_i^{(j)}}^{\perp} - \boldsymbol{\Pi}_{\mathbf{H}_i}^{\perp}\right]. \tag{18}$$

Finally, we regroup the two last terms of (15) to define

$$\boldsymbol{\varepsilon}_i^{(j)} \triangleq \boldsymbol{\Xi}_i^{(j)}\boldsymbol{h}_i^{(j)} + \sigma_i^{(j)}\boldsymbol{\Omega}_i^{(j)}\boldsymbol{\delta}_i^{(j)}, \qquad \forall i, j \in \{1, \ldots, K\}. \tag{19}$$

.5.2 Sketch of the proof

During all this proof, we will use the short-hand notation σ to denote $\max_{i,j} \sigma_i^{(j)}$. The proof consists in considering first the channel realizations and the channel quantization errors belonging to the subspace $\mathcal{A}_\sigma$ defined as

$$\mathcal{A}_\sigma \triangleq \left\{ \left(\mathbf{H}, \left(\boldsymbol{\Delta}_i^{(j)}, \boldsymbol{\delta}_i^{(j)} \right)_{i,j} \right) \middle| \|\boldsymbol{\varepsilon}_i^{(j)}\| \leq C(\sigma)\sigma^{1+\frac{1}{8}}, \forall i,j \in \{1,\ldots,K\} \right\} \quad (20)$$

with $\boldsymbol{\varepsilon}_i^{(j)}$ defined in (19) and with $C(\sigma)$ being a function verifying that $\lim_{\sigma\to 0} C(\sigma) = 0$. By construction of the set $\mathcal{A}_\sigma$, it holds over $\mathcal{A}_\sigma$ that (15) can be written as

$$\boldsymbol{u}_i^{(j)} = \boldsymbol{u}_i^\star + \boldsymbol{a}_i^{(j)} + o(\sigma), \qquad \forall i,j \in \{1,\ldots,K\}. \quad (21)$$

We will proceed by computing an upperbound on the average rate loss over the subspace $\mathcal{A}_\sigma$ and over its complementary subspace $\bar{\mathcal{A}}_\sigma \triangleq \mathcal{H} \setminus \mathcal{A}_\sigma$. Hence, we write the rate loss $\Delta_{\mathrm{R},i}^{\mathrm{DCSI}}$ as

$$\Delta_{\mathrm{R},i}^{\mathrm{DCSI}} = \Pr(\mathcal{A}_\sigma)\Delta_{\mathrm{R},i|_{\mathcal{A}_\sigma}}^{\mathrm{DCSI}} + \Pr(\bar{\mathcal{A}}_\sigma)\Delta_{\mathrm{R},i|_{\bar{\mathcal{A}}_\sigma}}^{\mathrm{DCSI}} \quad (22)$$

where $\Delta_{\mathrm{R},i|_{\mathcal{A}_\sigma}}^{\mathrm{DCSI}}$ and $\Delta_{\mathrm{R},i|_{\bar{\mathcal{A}}_\sigma}}^{\mathrm{DCSI}}$ denote respectively the expected rate loss over $\mathcal{A}_\sigma$ and $\bar{\mathcal{A}}_\sigma$.

Specifically, the proof will be divided into four steps:

1. We show that
$$\Pr(\mathcal{A}_\sigma) = (1 - O(\sigma^{\frac{1}{4}})). \quad (23)$$

2. We compute the expected rate loss over the subspace $\mathcal{A}_\sigma$.

3. We provide a (trivial) upperbound for the expected rate loss over the subspace $\bar{\mathcal{A}}_\sigma$.

4. We put the results together to obtain an upperbound for the expected rate loss using the conditional expectation formula (22).

In the following, we describe each step in full details.

.5.3 Probability of $\mathcal{A}_\sigma$

We prove here that there is a function $C(\sigma)$ with $\lim_{\sigma\to 0} C(\sigma) = 0$ such that $\forall i, j$

$$1 - \Pr(\mathcal{A}_\sigma) = \Pr\left(\|\boldsymbol{\varepsilon}_i^{(j)}\| > C(\sigma)\sigma^{1+\frac{1}{8}}\right) = O\left(\sigma^{\frac{1}{4}}\right). \tag{24}$$

Using that $\|\mathbf{A}\boldsymbol{x}\| \leq \|\mathbf{A}\|_{\mathrm{F}}\|\boldsymbol{x}\|$, we can easily upperbound $\|\boldsymbol{\varepsilon}_i^{(j)}\|$ as

$$\|\boldsymbol{\varepsilon}_i^{(j)}\| \leq \|\boldsymbol{\Xi}_i^{(j)}\|_{\mathrm{F}}\|\boldsymbol{h}_i^{(j)}\| + \sigma\|\boldsymbol{\Omega}_i^{(j)}\|_{\mathrm{F}}\|\boldsymbol{\delta}_i^{(j)}\|. \tag{25}$$

Hence, we have

$$\Pr\left(\|\boldsymbol{\varepsilon}_i^{(j)}\| > x\right) \leq \Pr\left(\|\boldsymbol{\Xi}_i^{(j)}\|_{\mathrm{F}}\|\boldsymbol{h}_i^{(j)}\| + \sigma\|\boldsymbol{\Omega}_i^{(j)}\|_{\mathrm{F}}\|\boldsymbol{\delta}_i^{(j)}\| > x\right) \tag{26}$$

$$\overset{(a)}{\leq} \Pr\left(\|\boldsymbol{\Xi}_i^{(j)}\|_{\mathrm{F}}\|\boldsymbol{h}_i^{(j)}\| > \frac{x}{2}\right) + \Pr\left(\sigma\|\boldsymbol{\Omega}_i^{(j)}\|_{\mathrm{F}}\|\boldsymbol{\delta}_i^{(j)}\| > \frac{x}{2}\right). \tag{27}$$

where inequality (a) follows from (ii) in Lemma 3, Appendix .1. Focusing on the first term of (25), it follows from the definition of $\boldsymbol{\Xi}_i^{(j)}$ in (17) that

$$\begin{aligned}
&\|\boldsymbol{\Xi}_i^{(j)}\|_{\mathrm{F}}\|\boldsymbol{h}_i^{(j)}\| \\
&\leq \|\hat{\mathbf{H}}_i^{(j)}\|_{\mathrm{F}}^2\|((\mathbf{H}_i^{(j)})^{\mathrm{H}}\mathbf{H}_i^{(j)})^{-1}\|_{\mathrm{F}}\|(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\|_{\mathrm{F}}^2\|\boldsymbol{\Phi}^{(j)}\|_{\mathrm{F}}^2\|\mathbf{H}_i\| \\
&\quad + 2\|\mathbf{H}_i\|_{\mathrm{F}}\|\boldsymbol{\Delta}_i^{(j)}\|_{\mathrm{F}}\|\boldsymbol{\Sigma}_i^{(j)}\|_{\mathrm{F}}\|\hat{\mathbf{H}}_i^{(j)}\|_{\mathrm{F}}\|(\mathbf{H}_i^{(j)})^{\mathrm{H}}\mathbf{H}_i^{(j)})^{-1}\|_{\mathrm{F}}^2\|\boldsymbol{\Phi}^{(j)}\|_{\mathrm{F}} \\
&\quad + \|\boldsymbol{\Delta}_i^{(j)}\|_{\mathrm{F}}^2\|\boldsymbol{\Sigma}_i^{(j)}\|_{\mathrm{F}}^2\|(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\|_{\mathrm{F}}^2\|\mathbf{H}_i\|_{\mathrm{F}}^3 \\
&\quad + \|\mathbf{H}_i\|_{\mathrm{F}}\|\boldsymbol{\Delta}_i^{(j)}\|_{\mathrm{F}}^2\|\boldsymbol{\Sigma}_i^{(j)}\|_{\mathrm{F}}^2\|(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\|_{\mathrm{F}}.
\end{aligned} \tag{28}$$

Since $\sigma = \max_{i,j}\sigma_i^{(j)}$, it is clear that $\forall i, j, \|\boldsymbol{\Sigma}_i^{(j)}\|_{\mathrm{F}} \leq K\sigma$ and (ii) from Lemma 3 in Appendix .1 to write

$$\begin{aligned}
&\Pr\left(\|\boldsymbol{\Xi}_i^{(j)}\|_{\mathrm{F}}\|\boldsymbol{h}_i^{(j)}\| > \frac{x}{2}\right) \\
&\leq \Pr\left(\|\hat{\mathbf{H}}_i^{(j)}\|_{\mathrm{F}}^2\|\boldsymbol{\Phi}^{(j)}\|_{\mathrm{F}}^2\|\mathbf{H}_i\|\|((\mathbf{H}_i^{(j)})^{\mathrm{H}}\mathbf{H}_i^{(j)})^{-1}\|_{\mathrm{F}}\|(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\|_{\mathrm{F}}^2 > \frac{x}{8}\right) \\
&\quad + \Pr\left(2K\sigma\|\mathbf{H}_i\|_{\mathrm{F}}\|\boldsymbol{\Delta}_i^{(j)}\|_{\mathrm{F}}\|\hat{\mathbf{H}}_i^{(j)}\|_{\mathrm{F}}\|(\mathbf{H}_i^{(j)})^{\mathrm{H}}\mathbf{H}_i^{(j)})^{-1}\|_{\mathrm{F}}^2\|\boldsymbol{\Phi}^{(j)}\|_{\mathrm{F}} > \frac{x}{8}\right) \\
&\quad + \Pr\left((K\sigma)^2\|\boldsymbol{\Delta}_i^{(j)}\|_{\mathrm{F}}^2\|(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\|_{\mathrm{F}}^2\|\mathbf{H}_i\|_{\mathrm{F}}^3 > \frac{x}{8}\right) \\
&\quad + \Pr\left((K\sigma)^2\|\mathbf{H}_i\|_{\mathrm{F}}\|\boldsymbol{\Delta}_i^{(j)}\|_{\mathrm{F}}^2\|(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)^{-1}\|_{\mathrm{F}} > \frac{x}{8}\right).
\end{aligned} \tag{29}$$

In fact, it will become clear that the first term is the most critical one. Hence, we start by studying this term and we define $\mathcal{I}$ to denote it:

$$\mathcal{I} \triangleq \|\hat{\mathbf{H}}_i^{(j)}\|_{\mathrm{F}}^2 \|\mathbf{\Phi}^{(j)}\|_{\mathrm{F}}^2 \|\mathbf{H}_i\| \|((\mathbf{H}_i^{(j)})^{\mathrm{H}} \mathbf{H}_i^{(j)})^{-1}\|_{\mathrm{F}} \|(\mathbf{H}_i^{\mathrm{H}} \mathbf{H}_i)^{-1}\|_{\mathrm{F}}^2. \tag{30}$$

We derive now an upperbound for $\mathcal{I}$ by upper-bounding each factor separately.

1. The Frobenius norm of $\mathbf{\Phi}^{(j)}$ can be upperbounded as follows.

$$\|\mathbf{\Phi}^{(j)}\|_{\mathrm{F}} = \left\| (\mathbf{\Delta}_i^{(j)} \mathbf{\Sigma}_i^{(j)})^{\mathrm{H}} \mathbf{H}_i + \mathbf{H}_i^{\mathrm{H}} \mathbf{\Delta}_i^{(j)} \mathbf{\Sigma}_i^{(j)} + (\mathbf{\Delta}_i^{(j)} \mathbf{\Sigma}_i^{(j)})^{\mathrm{H}} \mathbf{\Delta}_i^{(j)} \mathbf{\Sigma}_i^{(j)} \right\|_{\mathrm{F}} \tag{31}$$

$$\leq K\sigma \left\| (\mathbf{\Delta}_i^{(j)})^{\mathrm{H}} \mathbf{H}_i + \mathbf{H}_i^{\mathrm{H}} \mathbf{\Delta}_i^{(j)} + K\sigma (\mathbf{\Delta}_i^{(j)})^{\mathrm{H}} \mathbf{\Delta}_i^{(j)} \right\|_{\mathrm{F}} \tag{32}$$

$$\overset{(a)}{<} 2K\sigma \left\| \frac{(\mathbf{H}_i + \mathbf{\Delta}_i^{(j)})^{\mathrm{H}}}{\sqrt{2}} \frac{(\mathbf{H}_i + \mathbf{\Delta}_i^{(j)})}{\sqrt{2}} \right\|_{\mathrm{F}} \tag{33}$$

$$\leq 2K\sigma \left\| \frac{(\mathbf{H}_i + \mathbf{\Delta}_i^{(j)})^{\mathrm{H}}}{\sqrt{2}} \right\|_{\mathrm{F}} \left\| \frac{(\mathbf{H}_i + \mathbf{\Delta}_i^{(j)})}{\sqrt{2}} \right\|_{\mathrm{F}} \tag{34}$$

$$= 2K\sigma \left\| \frac{(\mathbf{H}_i + \mathbf{\Delta}_i^{(j)})}{\sqrt{2}} \right\|_{\mathrm{F}}^2 \tag{35}$$

   where we have used that σ is very small compared to one to obtain inequality (a).

2. The Frobenius norm of the matrix $(\mathbf{H}_i^{\mathrm{H}} \mathbf{H}_i)^{-1}$ verifies [125]

$$\|(\mathbf{H}_i^{\mathrm{H}} \mathbf{H}_i)^{-1}\|_{\mathrm{F}} \leq \frac{\sqrt{K}}{\lambda_{\min}(\mathbf{H}_i^{\mathrm{H}} \mathbf{H}_i)}. \tag{36}$$

3. The matrix $\hat{\mathbf{H}}_i^{(j)}$ has asymptotically the same distribution as $\mathbf{H}_i$ as the accuracy of the CSIT increases such that (36) also holds for the matrix $((\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}} \hat{\mathbf{H}}_i^{(j)})^{-1}$.

In total, we have obtained

$$\mathcal{I} \leq K^3 \sqrt{K} \frac{\|\hat{\mathbf{H}}_i^{(j)}\|^2 (2\sigma \| \frac{(\mathbf{H}_i + \mathbf{\Delta}^{(j)})}{\sqrt{2}} \|_{\mathrm{F}}^2)^2 \|\mathbf{H}_i\|}{\lambda_{\min}((\mathbf{H}_i^{(j)})^{\mathrm{H}} \mathbf{H}_i^{(j)}) \lambda_{\min}^2 (\mathbf{H}_i^{\mathrm{H}} \mathbf{H}_i)}. \tag{37}$$

Using (i) from Lemma 3 in Appendix .1 , we now upperbound the tail of the probability distribution of $\mathcal{I}$.

$$\begin{aligned}
&\Pr\left(\mathcal{I} > 4K^3\sqrt{K}\sigma^{\frac{5}{4}}(\log(\sigma^{-4}))^4\right) \\
&\overset{(a)}{\leq} \Pr\left(\frac{\|\hat{\mathbf{H}}_i^{(j)}\|^2\|\frac{(\mathbf{H}_i+\mathbf{\Delta}^{(j)})}{\sqrt{2}}\|_{\mathrm{F}}^4\|\mathbf{H}_i\|^2}{\lambda_{\min}((\mathbf{H}_i^{(j)})^{\mathrm{H}}\mathbf{H}_i^{(j)})\lambda_{\min}^2(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i)} > \sigma^{-\frac{3}{4}}(\log(\sigma^{-4}))^4\right) \qquad (38) \\
&\overset{(b)}{\leq} \Pr\left(\left\|\hat{\mathbf{H}}_i^{(j)}\right\|_{\mathrm{F}}^2 > \log(\sigma^{-4})\right) + 2\Pr\left(\left\|\frac{(\mathbf{H}_i+\mathbf{\Delta}^{(j)})}{\sqrt{2}}\right\|_{\mathrm{F}}^2 > \log(\sigma^{-4})\right) \\
&\quad + \Pr\left(\|\mathbf{H}_i\|_{\mathrm{F}}^2 > \log(\sigma^{-4})\right) + \Pr\left(\lambda_{\min}((\mathbf{H}_i^{(j)})^{\mathrm{H}}\mathbf{H}_i^{(j)}) < \sigma^{\frac{1}{4}}\right) \\
&\quad + 2\Pr\left(\lambda_{\min}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i) < \sigma^{\frac{1}{4}}\right) \qquad (39)
\end{aligned}$$

where (a) follows from (37) and from $\|\mathbf{H}_i\|_{\mathrm{F}} \leq \|\mathbf{H}_i\|_{\mathrm{F}}^2$ when $\|\mathbf{H}_i\|_{\mathrm{F}} \geq 1$ and (b) from (i) in Lemma 3 in Appendix .1. We continue by upper-bounding separately each term in (39).

1. The matrix $\hat{\mathbf{H}}_i^{(j)}$ has its elements i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$[2]. Hence, the squared Frobenius norm is a Chi-square random variable with $(K-1)K$ degrees-of-freedom. The tail of the probability distribution can be upper-bounded with the help of Lemma 4 in Appendix .1 with $n = K(K-1)$ and $x = \log(\sigma^{-4})$. For σ sufficiently small, this gives

$$\Pr\left(\left\|\hat{\mathbf{H}}_i^{(j)}\right\|_{\mathrm{F}}^2 > \log(\sigma^{-4})\right) \leq e^{\frac{K(K-1)}{2}\log\left(\frac{e}{K(K-1)}\right)}(\log(\sigma^{-4}))^{\frac{K(K-1)}{2}}\sigma^2. \qquad (40)$$

2. Both the matrix $(\mathbf{H}_i + \mathbf{\Delta}_i^{(j)})/(\sqrt{2})$ and the matrix $\mathbf{H}_i$ have their elements i.i.d. $\mathcal{N}_{\mathbb{C}}(0,1)$ such that we can use the same upperbound.

3. The matrix $\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i$ is a Wishart matrix of size $K-1$ with K degrees-of-freedom from which the probability distribution of the minimal eigenvalue is known from [124, Theorem 5.4] to be exponential and given by

$$\Pr\left(\lambda_{\min}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i) \leq x\right) = 1 - \exp\left(-\frac{(K-1)}{2}x\right). \qquad (41)$$

[2]The distribution is in fact $\mathcal{N}_{\mathbb{C}}(0,1)$ because of we use (1), but we do not consider this detail for the sake of clarity.

Hence, with $x = \sigma^{\frac{1}{4}}$ and σ being small, it gives after a Taylor expansion

$$\Pr\left(\lambda_{\min}(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i) < \sigma^{\frac{1}{4}}\right) = \frac{K-1}{2}\sigma^{\frac{1}{4}} + o(\sigma^{\frac{1}{4}}). \tag{42}$$

4. The matrix $\hat{\mathbf{H}}_i^{(j)}$ has the same distribution as $\mathbf{H}_i$ such that we can use the same upperbound with the probability distribution of $(\hat{\mathbf{H}}_i^{(j)})^{\mathrm{H}}\hat{\mathbf{H}}_i^{(j)}$.

Inserting these last four results in (39) gives

$$\begin{aligned}\Pr\left(\mathcal{I} > 4K^3\sqrt{K}\sigma^{\frac{5}{4}}(\log(\sigma^{-4}))^4\right) &\leq 4e^{\frac{K(K-1)}{2}\log\left(\frac{e}{K(K-1)}\right)}(\log(\sigma^{-4}))^{\frac{K(K-1)}{2}}\sigma^2 \\ &\quad + 2\frac{K-1}{2}\sigma^{\frac{1}{4}} + o(\sigma^{\frac{1}{4}}) \qquad (43)\\ &= O(\sigma^{\frac{1}{4}}). \qquad (44)\end{aligned}$$

Coming back to (29), it remains to derive an upperbound for the three other terms in the RHS. The calculation carried out for the first term can be adapted to the three other terms without any difficulty. Furthermore, it can easily be seen from the above calculation that we have considered the largest term such that the other probabilities are in fact smaller. Thus, it holds that

$$\Pr\left(\|\boldsymbol{\Xi}_i^{(j)}\|_{\mathrm{F}}\|\boldsymbol{h}_i\| > 16K^3\sqrt{K}\sigma^{\frac{5}{4}}(\log(\sigma^{-4}))^4\right) = O(\sigma^{\frac{1}{4}}). \tag{45}$$

Our goal is to derive an upperbound for (27) and we still have to study the second term in (27). This is done by writing

$$\begin{aligned}&\|\sigma\boldsymbol{\Omega}_i^{(j)}\boldsymbol{\delta}_i^{(j)}\|_{\mathrm{F}} \\ &\leq K\sigma\|\left(\mathbf{I}_K - \mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\right)\boldsymbol{\Delta}_i^{(j)}\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\|_{\mathrm{F}}\|\boldsymbol{\Delta}_i^{(j)}\|_{\mathrm{F}} \\ &+ K\sigma\|\boldsymbol{\Xi}_i^{(j)}\|_{\mathrm{F}}\|\boldsymbol{\Delta}_i^{(j)}\|_{\mathrm{F}} \\ &\quad + K\sigma\|\mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}(\boldsymbol{\Delta}_i^{(j)})^{\mathrm{H}}\left(\mathbf{I}_K - \mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\right)\|_{\mathrm{F}}\|\boldsymbol{\Delta}_i^{(j)}\|_{\mathrm{F}} \qquad (46)\\ &\leq 2K\sigma\|\boldsymbol{\Delta}_i^{(j)}\|_{\mathrm{F}}^2\|\mathbf{H}_i\|_{\mathrm{F}}\|\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\|_{\mathrm{F}} + K\sigma\|\boldsymbol{\Xi}_i^{(j)}\|_{\mathrm{F}}\|\boldsymbol{\Delta}_i^{(j)}\|_{\mathrm{F}} \qquad (47)\end{aligned}$$

Similarly, we easily obtain that the RHS of (47) has a smaller probability than $\mathcal{I}$ to be asymptotically large so that we have

$$\Pr\left(\sigma\|\boldsymbol{\Omega}_i^{(j)}\|_{\mathrm{F}}\|\boldsymbol{\delta}_i^{(j)}\| > 16K^3\sqrt{K}\sigma^{\frac{5}{4}}(\log(\sigma^{-4}))^4\right) = O(\sigma^{\frac{1}{4}}). \tag{48}$$

We can then use (48) and (45) in (27) to obtain that

$$\Pr\left(\|\boldsymbol{\varepsilon}_i^{(j)}\| > 32K^3\sqrt{K}\sigma^{\frac{5}{4}}(\log(\sigma^{-4}))^4\right) = O(\sigma^{\frac{1}{4}}). \tag{49}$$

Applying the above calculation for every i and every j, we have then shown that

$$\Pr\left(\|\boldsymbol{\varepsilon}_i^{(j)}\| > C(\sigma)\sigma^{1+\frac{1}{8}}\right) = O\left(\sigma^{\frac{1}{4}}\right) \tag{50}$$

for

$$C(\sigma) \triangleq 32K^3\sqrt{K}\sigma^{\frac{1}{8}}(\log(\sigma^{-4}))^4. \tag{51}$$

.5.4 Upperbound for the rate loss over $\mathcal{A}_\sigma$

By construction, it holds over the subspace $\mathcal{A}_\sigma$ that

$$\boldsymbol{u}_i^{\mathrm{DCSI}} = \boldsymbol{u}_i^\star + \boldsymbol{a}_i^{\mathrm{DCSI}} + \boldsymbol{\varepsilon}_i^{\mathrm{DCSI}}, \qquad i,j \in \{1,\ldots,K\} \tag{52}$$

where we have defined

$$\boldsymbol{a}_i^{\mathrm{DCSI}} \triangleq \begin{bmatrix} \boldsymbol{e}_1^{\mathrm{H}}\boldsymbol{a}_i^{(1)} \\ \vdots \\ \boldsymbol{e}_K^{\mathrm{H}}\boldsymbol{a}_i^{(K)} \end{bmatrix}, \qquad \boldsymbol{\varepsilon}_i^{\mathrm{DCSI}} \triangleq \begin{bmatrix} \boldsymbol{e}_1^{\mathrm{H}}\boldsymbol{\varepsilon}_i^{(1)} \\ \vdots \\ \boldsymbol{e}_K^{\mathrm{H}}\boldsymbol{\varepsilon}_i^{(K)} \end{bmatrix}, \qquad i \in \{1,\ldots,K\} \tag{53}$$

such that $\boldsymbol{\varepsilon}_i^{\mathrm{DCSI}} = o(\sigma)$.

We will use this property to upperbound the rate loss. We start from the definition of the rate loss and proceed as follows.

$$\begin{aligned}
\Delta_{\mathrm{R},i|\mathcal{A}_\sigma}^{\mathrm{DCSI}} &\triangleq \mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2(1+P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^\star|^2)\right] \\
&\quad - \mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2\left(1+\frac{P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\mathrm{DCSI}}|^2}{1+\sum_{j=1,j\neq i}^K P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{DCSI}}|^2}\right)\right] \qquad (54)\\
&= \mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2(1+P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^\star|^2)\right] - \mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2\left(1+\sum_{j=1}^K P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{DCSI}}|^2\right)\right] \\
&\quad + \mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2\left(1+\sum_{j=1,j\neq i}^K P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{DCSI}}|^2\right)\right] \qquad (55)\\
&\leq \mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2(1+P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^\star|^2)\right] - \mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2\left(1+P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\mathrm{DCSI}}|^2\right)\right] \\
&\quad + \mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2\left(1+\sum_{j=1,j\neq i}^K P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{DCSI}}|^2\right)\right]. \qquad (56)
\end{aligned}$$

Focusing on the inner product $|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\mathrm{DCSI}}|^2$, we can write

$$
\begin{aligned}
&|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\mathrm{DCSI}}|^2 \\
&= \boldsymbol{h}_i^{\mathrm{H}}(\boldsymbol{u}_i^{\star} + \boldsymbol{a}_i^{\mathrm{DCSI}} + \boldsymbol{\varepsilon}_i^{\mathrm{DCSI}})(\boldsymbol{u}_i^{\star} + \boldsymbol{a}_i^{\mathrm{DCSI}} + \boldsymbol{\varepsilon}_i^{\mathrm{DCSI}})^{\mathrm{H}}\boldsymbol{h}_i && (57)\\
&= |\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star}|^2 + (\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star})((\boldsymbol{a}_i^{\mathrm{DCSI}})^{\mathrm{H}}\boldsymbol{h}_i) + (\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_i^{\mathrm{DCSI}})((\boldsymbol{u}_i^{\star})^{\mathrm{H}}\boldsymbol{h}_i) + |\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_i^{\mathrm{DCSI}}|^2 \\
&\qquad + (\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{\varepsilon}_i^{\mathrm{DCSI}})((\boldsymbol{u}_i^{\mathrm{DCSI}})^{\mathrm{H}}\boldsymbol{h}_i) + (\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\mathrm{DCSI}})((\boldsymbol{\varepsilon}_i^{\mathrm{DCSI}})^{\mathrm{H}}\boldsymbol{h}_i) && (58)\\
&\overset{(a)}{=} |\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star}|^2 + O(\sigma) && (59)
\end{aligned}
$$

where (a) follows from the definition of $\boldsymbol{a}_i^{\mathrm{DCSI}}$ in (53) and from $\boldsymbol{\varepsilon}_i^{\mathrm{DCSI}} = o(\sigma)$. Using (59), we can show that

$$
\begin{aligned}
&\mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2(1 + P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star}|^2)\right] - \mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2\left(1 + P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\mathrm{DCSI}}|^2\right)\right] && (60)\\
&= -\mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2\left(1 + \frac{O(P\sigma)}{1 + P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star}|^2}\right)\right] && (61)\\
&= o(1) && (62)
\end{aligned}
$$

when σ tends to zero. We have then obtained

$$
\Delta_{\mathrm{R},i|\mathcal{A}_\sigma}^{\mathrm{DCSI}} \leq \mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2\left(1 + \sum_{j=1, j\neq i}^{K} P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{DCSI}}|^2\right)\right] + o(1). \tag{63}
$$

Note that we will omit for ease of notation to write explicitly the $o(1)$ in the rest of the proof as our calculation is always done up to a $o(1)$ and removing it leads to no confusion.

Inserting the Taylor approximation of $\boldsymbol{u}_i^{\mathrm{DCSI}}$ from (52) in (63), we write

$$\begin{aligned}
&\Delta_{\mathrm{R},i|\sigma}^{\mathrm{DCSI}} \\
&\leq \mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2\left(1+\sum_{j=1,j\neq i}^{K} P|\boldsymbol{h}_i^{\mathrm{H}}(\boldsymbol{u}_j^{\star}+\boldsymbol{a}_j^{\mathrm{DCSI}}+\boldsymbol{\varepsilon}_j^{\mathrm{DCSI}})|^2\right)\right] && (64)\\
&\overset{(a)}{=} \mathrm{E}_{\mathcal{A}_\sigma}\left[\log_2\left(1+\sum_{j=1,j\neq i}^{K} P|\boldsymbol{h}_i^{\mathrm{H}}(\boldsymbol{a}_j^{\mathrm{DCSI}}+\boldsymbol{\varepsilon}_j^{\mathrm{DCSI}})|^2\right)\right] && (65)\\
&= \mathrm{E}_{\mathcal{A}_\sigma}\Bigg[\log_2\Big(1+\sum_{j=1,j\neq i}^{K} P\big(|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_j^{\mathrm{DCSI}}|^2+(\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_j^{\mathrm{DCSI}})(\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{\varepsilon}_j^{\mathrm{DCSI}})^{\mathrm{H}} \\
&\qquad\qquad + (\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{\varepsilon}_j^{\mathrm{DCSI}})(\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_j^{\mathrm{DCSI}})^{\mathrm{H}}+|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{\varepsilon}_j^{\mathrm{DCSI}}|^2\big)\Big)\Bigg] && (66)\\
&\overset{(b)}{\leq} \mathrm{E}_{\mathcal{A}_\sigma}\Big[\log_2\big(1+\sum_{j=1,j\neq i}^{K} P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_j^{\mathrm{DCSI}}|^2+o(\sigma^2P)\big)\Big] && (67)
\end{aligned}$$

where (a) is obtained because $\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\star}=0$ for $j\neq i$, (b) is obtained because $\|\boldsymbol{a}_j^{\mathrm{DCSI}}\|=O(\sigma)$ and $\|\boldsymbol{\varepsilon}_i^{\mathrm{DCSI}}\|=o(\sigma)$ over $\mathcal{A}_\sigma$.

We can then rewrite the upperbound in (67) to obtain

$$\begin{aligned}
&\Delta_{\mathrm{R},i|\sigma}^{\mathrm{DCSI}} \\
&\leq \mathrm{E}_{\mathcal{A}_\sigma}\Bigg[\log_2\Bigg(1+\max_{\ell}|\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2\Bigg(\sum_{j=1,j\neq i}^{K} P\frac{|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_j^{\mathrm{DCSI}}|^2}{\max_\ell|\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2}+o(\sigma^2P)\Bigg)\Bigg)\Bigg] && (68)\\
&\overset{(a)}{\leq} \mathrm{E}_{\mathcal{A}_\sigma}\Big[\log_2\big(1+\max_{\ell}|\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2\big)\Big] \\
&\quad + \mathrm{E}_{\mathcal{A}_\sigma}\Big[\log_2\big(1+\sum_{j=1,j\neq i}^{K} P\frac{|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_j^{\mathrm{DCSI}}|^2}{\max_\ell|\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2}+o(\sigma^2P)\big)\Big] && (69)\\
&\overset{(b)}{\leq} \log_2\Big(1+\mathrm{E}_{\mathcal{A}_\sigma}\Big[\max_{\ell}|\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2\Big]\Big) \\
&\quad + \log_2\Bigg(1+P\sum_{j=1,j\neq i}^{K}\mathrm{E}_{\mathcal{A}_\sigma}\Bigg[\frac{|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_j^{\mathrm{DCSI}}|^2}{\max_\ell|\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2}\Bigg]+o(\sigma^2P)\Bigg) && (70)
\end{aligned}$$

where (a) is obtained by using that $\log(1+ab)\leq\log(1+a)+\log(1+b)$ for $a\geq 0$ and $b\geq 0$, and (b) follows from Jensen's inequality. To compute

the expectation over $\mathcal{A}_\sigma$, we resort to Lemma 7 in Appendix .1 to obtain

$$\Delta_{\mathrm{R},i|_\sigma}^{\mathrm{DCSI}} \leq \log_2\left(1+\frac{\mathrm{E}\left[\max_\ell |\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2\right]}{\Pr(\mathcal{A}_\sigma)}\right)$$
$$+\log_2\left(1+P\sum_{j=1,j\neq i}^{K}\frac{\mathrm{E}\left[\frac{|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_j^{\mathrm{DCSI}}|^2}{\max_\ell |\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2}\right]}{\Pr(\mathcal{A}_\sigma)}+o(\sigma^2P)\right) \tag{71}$$
$$\leq \log_2\left(1+\mathrm{E}\left[\max_\ell |\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2\right](1-o(1))\right)$$
$$+\log_2\left(1+P\sum_{j=1,j\neq i}^{K}\mathrm{E}\left[\frac{|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_j^{\mathrm{DCSI}}|^2}{\max_\ell |\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2}\right](1-o(1))+o(\sigma^2P)\right). \tag{72}$$

where we have used that $\Pr(\mathcal{A}_\sigma) = 1 - o(1)$ to obtain the last inequality. This can easily be obtained from Appendix .5.3 and is further discussed in Appendix .5.5. It remains then to compute the two expectations. Focusing first on the second expectation, we can write

$$\mathrm{E}\left[\frac{|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{a}_j^{\mathrm{DCSI}}|^2}{\max_\ell |\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2}\right]$$
$$=\sum_{k=1}^{K}\sum_{k'=1}^{K}\mathrm{E}\left[\frac{\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_k\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{a}_j^{(k)}(\boldsymbol{a}_j^{(k')})^{\mathrm{H}}\boldsymbol{e}_{k'}\boldsymbol{e}_{k'}^{\mathrm{H}}\mathbf{H}\boldsymbol{e}_i}{\max_\ell |\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2}\right] \tag{73}$$
$$\overset{(a)}{=}\sum_{k=1}^{K}\sum_{k'=1}^{K}\mathrm{E}_{\mathbf{H}_i}\left[\frac{\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_k\boldsymbol{e}_k^{\mathrm{H}}\mathrm{E}_{\{\boldsymbol{\Delta}_i^{(j)}\}_{i,j}}\left[\boldsymbol{a}_j^{(k)}(\boldsymbol{a}_j^{(k')})^{\mathrm{H}}\right]\boldsymbol{e}_{k'}\boldsymbol{e}_{k'}^{\mathrm{H}}\mathbf{H}\boldsymbol{e}_i}{\max_\ell |\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2}\right] \tag{74}$$
$$\overset{(b)}{=}\sum_{k=1}^{K}\mathrm{E}_{\mathbf{H}_i}\left[\frac{|\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_k|^2\boldsymbol{e}_k^{\mathrm{H}}\mathrm{E}_{\{\boldsymbol{\Delta}_i^{(j)}\}_{i,j}}\left[\boldsymbol{a}_j^{(k)}(\boldsymbol{a}_j^{(k)})^{\mathrm{H}}\right]\boldsymbol{e}_k}{\max_\ell |\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2}\right] \tag{75}$$
$$\overset{(c)}{\leq}\mathrm{E}\left[\sum_{k=1}^{K}|\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{a}_j^{(k)}|^2\right] \tag{76}$$

where (a) follows from the independence between the quantization errors and the channel, (b) from $\mathrm{E}_{\{\boldsymbol{\Delta}_i^{(j)}\}_{i,j}}[\boldsymbol{a}_j^{(k)}(\boldsymbol{a}_j^{(k')})^{\mathrm{H}}] = \delta_{kk'}$, (c) is obtained by taking the maximum over the $|\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_k|^2$.

We now turn to the first expectation in (72). It is the expectation of the maximum of K independent Chi-square random variable with 2 degrees-of-

freedom and is computed via Lemmas 5 and Lemma 6 in Appendix .1. We have then obtained

$$\mathrm{E}\left[\max_{\ell}|\boldsymbol{e}_i^{\mathrm{H}}\mathbf{H}^{\mathrm{H}}\boldsymbol{e}_\ell|^2\right] = 2\sum_{m=1}^{K}\frac{1}{m} \tag{77}$$

$$\leq 2(1+\log(K)). \tag{78}$$

Using (76) and (78) inside (72) gives

$$\begin{aligned}
&\Delta_{\mathrm{R},i|\sigma}^{\mathrm{DCSI}} \\
&\leq \log_2(3+2\log(K)) \\
&\quad + \log_2\left(1+P\sum_{j=1,j\neq i}^{K}\mathrm{E}\left[\sum_{k=1}^{K}|\boldsymbol{e}_k^{\mathrm{H}}\boldsymbol{a}_j^{(k)}|^2\right](1+o(1))\right) && (79)\\
&\overset{(a)}{\leq} \log_2(3+2\log(K))+ \\
&\quad \log_2\left(1+P\sum_{j=1,j\neq i}^{K}\sum_{k=1}^{K}\frac{2\sum_{\ell=1,\ell\neq j}^{K}(\sigma_\ell^{(k)})^2+(\sigma_j^{(k)})^2}{K}(1+o(1))\right) && (80)\\
&= \log_2(3+2\log(K))+ \\
&\quad \log_2\left(1+\frac{P}{K}\sum_{k=1}^{K}\sum_{j=1,j\neq i}^{K}\left(2\sum_{\ell=1,\ell\neq j}^{K}(\sigma_\ell^{(k)})^2+(\sigma_j^{(k)})^2\right)(1+o(1))\right) && (81)
\end{aligned}$$

where (a) is obtained after using Proposition 9. By inspection of the successive summations in (82), it easily follows that

$$\begin{aligned}
\Delta_{\mathrm{R},i|\sigma}^{\mathrm{DCSI}} \leq \mu(K)+ \\
\log_2\left(1+\frac{P}{K}\sum_{k=1}^{K}\left((2K-3)\sum_{p=1,p\neq i}^{K}(\sigma_\ell^{(k)})^2+2(K-1)(\sigma_p^{(k)})^2\right)(1+o(1))\right)
\end{aligned} \tag{82}$$

with $\mu(K)$ defined in (5.23).

.5.5 Upperbound for the Rate Loss over $\bar{\mathcal{A}}_\sigma$

Due to the quickly vanishing probability of the set $\bar{\mathcal{A}}_\sigma$, a loose upperbound for the rate loss is sufficient, and is obtained as follows:

$$\begin{aligned}
\Delta^{\mathrm{DCSI}}_{\mathrm{R},i|\bar{\mathcal{A}}_\sigma} &\triangleq \mathrm{E}_{\bar{\mathcal{A}}_\sigma}\left[\log_2(1+P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star}|^2)\right] \\
&\quad - \mathrm{E}_{\bar{\mathcal{A}}_\sigma}\left[\log_2\left(1+\frac{P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\mathrm{DCSI}}|^2}{1+\sum_{j=1,j\neq i}^{K}P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_j^{\mathrm{DCSI}}|^2}\right)\right] && (83)\\
&\leq \mathrm{E}_{\bar{\mathcal{A}}_\sigma}\left[\log_2(1+P|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star}|^2)\right] && (84)\\
&\leq \log_2\left(1+P\mathrm{E}_{\bar{\mathcal{A}}_\sigma}\left[|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star}|^2\right]\right) && (85)\\
&\overset{(a)}{\leq} \log_2\left(1+P\frac{\mathrm{E}\left[|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star}|^2\right]}{\Pr(\bar{\mathcal{A}}_\sigma)}\right) && (86)\\
&\overset{(b)}{\leq} \log_2\left(1+P\frac{\mathrm{E}\left[|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star}|^2\right]}{\sigma^2}\right) && (87)
\end{aligned}$$

where (a) follows from Lemma 7 in Appendix .1, inequality (b) is obtained from using (41) with $x=\sigma^2$. Indeed, it can easily be shown that $\Pr(\bar{\mathcal{A}}_\sigma)\geq\sigma^2$. Finally, the expectation of the inner-product is computed as follows

$$\begin{aligned}
\mathrm{E}\left[|\boldsymbol{h}_i^{\mathrm{H}}\boldsymbol{u}_i^{\star}|^2\right] &= \mathrm{E}\left[|\boldsymbol{h}_i^{\mathrm{H}}\left(\mathbf{I}_K-\mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\right)\boldsymbol{h}_i|^2\right] && (88)\\
&= \mathrm{E}\left[\|\left(\mathbf{I}_K-\mathbf{H}_i\left(\mathbf{H}_i^{\mathrm{H}}\mathbf{H}_i\right)^{-1}\mathbf{H}_i^{\mathrm{H}}\right)\boldsymbol{h}_i\|^2\right] && (89)\\
&\overset{(a)}{=} 1 && (90)
\end{aligned}$$

where (a) is obtained after basic manipulations exploiting that $\boldsymbol{h}_i$ and $\mathbf{H}_i$ are independent.

.5.6 Total Expected Rate Loss

Putting all the results together, we can write the expected rate loss as

$$\Delta_{\mathrm{R},i}^{\mathrm{DCSI}} = \Pr(\mathcal{A}_\sigma)\Delta_{\mathrm{R},i|_{\mathcal{A}_\sigma}}^{\mathrm{DCSI}} + \Pr(\bar{\mathcal{A}}_\sigma)\Delta_{\mathrm{R},i|_{\bar{\mathcal{A}}_\sigma}}^{\mathrm{DCSI}} \tag{91}$$

$$\leq \mu(K)+ \log_2\left(1+\frac{P}{K}\sum_{k=1}^{K}\left((2K-3)\sum_{p=1,p\neq i}^{K}(\sigma_\ell^{(k)})^2+2(K-1)(\sigma_i^{(k)})^2\right)+o(P\sigma^2)\right) +O(\sigma^{\frac{1}{4}}\log_2(P\sigma^{-2})) \tag{92}$$

$$\overset{(a)}{=} \mu(K)+ \log_2\left(1+\frac{P}{K}\sum_{k=1}^{K}\left((2K-3)\sum_{p=1,p\neq i}^{K}(\sigma_\ell^{(k)})^2+2(K-1)(\sigma_i^{(k)})^2\right)\right)+o(1) \tag{93}$$

with equality a holding true because of the assumption that all the $\sigma_i^{(j)}$ decrease polynomially with the SNR P. This concludes the proof of Theorem 8.

.6 Proof of Proposition 8

We consider for the sake of clarity and without loss of generality, the inner product between the beamformer $\boldsymbol{u}_K^{\mathrm{CCSI}}$ with the jth channel vector $\boldsymbol{h}_j$ for $j < K$. We define the matrix $\boldsymbol{\Sigma}_K \triangleq \mathrm{diag}([\sigma_1,\ldots,\sigma_{K-1}])$ and the matrix $\boldsymbol{\Delta}_K \in \mathbb{C}^{K\times K-1}$ as

$$\boldsymbol{\Delta}_K \triangleq \left[\boldsymbol{\delta}_1,\ldots,\boldsymbol{\delta}_{K-1}\right]. \tag{94}$$

It then holds

$$\hat{\mathbf{H}}_K = \mathbf{H}_K + \boldsymbol{\Delta}_K\boldsymbol{\Sigma}_K. \tag{95}$$

Let us introduce further the matrix $\mathbf{M}_{\boldsymbol{\delta}_j} \in \mathbb{C}^{K\times K-1}$ as

$$\mathbf{M}_{\boldsymbol{\delta}_j} \triangleq \begin{bmatrix}\mathbf{0} & \ldots & \mathbf{0} & \boldsymbol{\delta}_j & \mathbf{0} & \ldots & \mathbf{0}\end{bmatrix} \tag{96}$$

with $\boldsymbol{\delta}_j$ forming the jth column of $\mathbf{M}_{\boldsymbol{\delta}_j}$. Finally, we define $\hat{\mathbf{G}}_K \in \mathbb{C}^{K\times K-1}$ such that

$$\hat{\mathbf{G}}_K \triangleq \hat{\mathbf{H}}_K - \mathbf{M}_{\boldsymbol{\delta}_j}\boldsymbol{\Sigma}_K \tag{97}$$

$$= \mathbf{H}_K + \boldsymbol{\Delta}_K\,\mathrm{diag}([\sigma_1,\ldots,\sigma_{j-1},0,\sigma_{j+1},\ldots,\sigma_{K-1}]). \tag{98}$$

Intuitively, $\hat{\mathbf{G}}_K$ is the estimate of $\hat{\mathbf{H}}_K$ without the error done on $\boldsymbol{h}_j$. By construction, the matrix $\hat{\mathbf{G}}_K$ is independent of $\boldsymbol{\delta}_K$.

Using the definitions introduced above, we can then write for $j < K$,

$$|\boldsymbol{h}_j^{\mathrm{H}} \boldsymbol{u}_K^{\mathrm{CCSI}}|^2 = |(\hat{\boldsymbol{h}}_j - \sigma_j \boldsymbol{\delta}_j)^{\mathrm{H}} \boldsymbol{u}_K^{\mathrm{CCSI}}|^2 \tag{99}$$

$$= \sigma_j^2 |\boldsymbol{\delta}_j^{\mathrm{H}} \boldsymbol{u}_K^{\mathrm{CCSI}}|^2 \tag{100}$$

since $\hat{\boldsymbol{h}}_j^{\mathrm{H}} \boldsymbol{u}_K^{\mathrm{CCSI}} = 0$ for $j < K$. We then rewrite the beamformer $\boldsymbol{u}_K^{\mathrm{CCSI}}$ as

$$\boldsymbol{u}_K^{\mathrm{CCSI}} = \Bigg(\mathbf{I}_K - (\hat{\mathbf{G}}_K + \sigma_j \mathbf{M}_{\boldsymbol{\delta}_j}) \Big((\hat{\mathbf{G}}_K + \sigma_j \mathbf{M}_{\boldsymbol{\delta}_j})^{\mathrm{H}} (\hat{\mathbf{G}}_K + \sigma_j \mathbf{M}_{\boldsymbol{\delta}_j}) \Big)^{-1} \cdot (\hat{\mathbf{G}}_K + \sigma_j \mathbf{M}_{\boldsymbol{\delta}_j})^{\mathrm{H}} \Big) \hat{\boldsymbol{h}}_K \tag{101}$$

$$= \hat{\boldsymbol{v}}_K + \boldsymbol{\varepsilon}_K \tag{102}$$

where we have defined

$$\hat{\boldsymbol{v}}_K \triangleq \left(\mathbf{I}_K - \hat{\mathbf{G}}_K \left(\hat{\mathbf{G}}_K^{\mathrm{H}} \hat{\mathbf{G}}_K \right)^{-1} \hat{\mathbf{G}}_K^{\mathrm{H}} \right) \hat{\boldsymbol{h}}_K \tag{103}$$

and

$$\boldsymbol{\varepsilon}_K \triangleq \boldsymbol{u}_K^{\mathrm{CCSI}} - \hat{\boldsymbol{v}}_K. \tag{104}$$

This gives

$$|\boldsymbol{h}_j^{\mathrm{H}} \boldsymbol{u}_K^{\mathrm{CCSI}}|^2 = \sigma_j^2 |\boldsymbol{\delta}_j^{\mathrm{H}} (\hat{\boldsymbol{v}}_K + \boldsymbol{\varepsilon}_K)|^2 \tag{105}$$

$$= \sigma_j^2 \left(|\boldsymbol{\delta}_j^{\mathrm{H}} \hat{\boldsymbol{v}}_K|^2 + \left(\boldsymbol{\delta}_j^{\mathrm{H}} \hat{\boldsymbol{v}}_K\right) \left(\boldsymbol{\delta}_j^{\mathrm{H}} \boldsymbol{\varepsilon}_K\right)^{\mathrm{H}} + \left(\boldsymbol{\delta}_j^{\mathrm{H}} \boldsymbol{\varepsilon}_K\right) \left(\boldsymbol{\delta}_j^{\mathrm{H}} \hat{\boldsymbol{v}}_K\right)^{\mathrm{H}} + |\boldsymbol{\delta}_j^{\mathrm{H}} \boldsymbol{\varepsilon}_K|^2 \right). \tag{106}$$

Taking the expectation yields

$$\mathrm{E}\left[|\boldsymbol{\delta}_j^{\mathrm{H}} \hat{\boldsymbol{v}}_K|^2 \right] = \mathrm{E}\left[\|\boldsymbol{\delta}_j\|^2 \right] \mathrm{E}\left[\|\hat{\boldsymbol{v}}_K\|^2 \right] \mathrm{E}\left[\left| \frac{\boldsymbol{\delta}_j^{\mathrm{H}} \hat{\boldsymbol{v}}_K}{\|\boldsymbol{\delta}_j\| \|\hat{\boldsymbol{v}}_K\|} \right|^2 \right] \tag{107}$$

$$= K(1 + o(1)) \mathrm{E}\left[\left| \frac{\boldsymbol{\delta}_j^{\mathrm{H}} \hat{\boldsymbol{v}}_K}{\|\boldsymbol{\delta}_j\| \|\hat{\boldsymbol{v}}_K\|} \right|^2 \right]. \tag{108}$$

The expectation of the norms have been easily obtained and the remaining expectation is the expectation of a Beta $(1, K-1)$ random variable because

the two vectors forming the inner product are independent, and $\boldsymbol{\delta}_j$ is isotropically distributed in the space. Thus, the expectation in (108) is equal to $1/K$ which gives in total

$$\mathrm{E}\left[|\boldsymbol{\delta}_j^{\mathrm{H}}\hat{\boldsymbol{v}}_K|^2\right] = (1 + o(1)) \tag{109}$$

All the expectations considered can be easily shown to exist such that we can take let σ_j tends to zero inside the expectation. Hence, if we write

$$\begin{aligned}&\mathrm{E}[|\boldsymbol{h}_j^{\mathrm{H}}\hat{\boldsymbol{u}}_K|^2] - \sigma_j^2\mathrm{E}[|\boldsymbol{\delta}_j^{\mathrm{H}}\hat{\boldsymbol{v}}_K|^2] \\ &= \sigma_j^2\mathrm{E}\left[\left(\boldsymbol{\delta}_j^{\mathrm{H}}\hat{\boldsymbol{v}}_K\right)\left(\boldsymbol{\delta}_j^{\mathrm{H}}\boldsymbol{\varepsilon}_K\right)^{\mathrm{H}} + \left(\boldsymbol{\delta}_j^{\mathrm{H}}\boldsymbol{\varepsilon}_K\right)\left(\boldsymbol{\delta}_j^{\mathrm{H}}\hat{\boldsymbol{v}}_K\right)^{\mathrm{H}} + |\boldsymbol{\delta}_j^{\mathrm{H}}\boldsymbol{\varepsilon}_K|^2\right],\end{aligned} \tag{110}$$

then the expectation in the RHS tends to zero as σ_j tends to zero meaning that the RHS is a $o(\sigma_j^2)$.

We have thus obtained that

$$\begin{aligned}\mathrm{E}[|\boldsymbol{h}_j^{\mathrm{H}}\hat{\boldsymbol{u}}_K|^2] &= \sigma_j^2\mathrm{E}[|\boldsymbol{\delta}_j^{\mathrm{H}}\hat{\boldsymbol{v}}_K|^2] + o(\sigma_j^2) &(111)\\ &= \sigma_j^2 + o(\sigma_j^2) &(112)\end{aligned}$$

which concludes the proof.

Bibliography

[1] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inf. Theo.*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[2] (2013, Oct.) Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012-2017. [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html

[3] 3GPP, "3gpp tr 36.932 "scenarios and requirements for small cell enhancements for e-utra and e-utran (release 12)"," December 2012.

[4] 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.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[5] K. Marton, "A coding theorem for the discrete memoryless broadcast channel," *IEEE Trans. Inf. Theo.*, vol. 25, no. 3, pp. 306–311, 1979.

[6] T. Cover and A. Thomas, *Elements of information theory*. Wiley-Interscience, Jul. 2006.

[7] G. Caire and S. Shamai (Shitz), "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, 2003.

[8] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. on Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[9] M. H. M. Costa, "Writing on dirty paper (Corresp.)," *IEEE Trans. Inf. Theo.*, vol. 29, no. 3, pp. 439–441, 1983.

[10] J. Lee and N. Jindal, "High SNR analysis for MIMO broadcast channels: Dirty paper coding Versus linear precoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, Dec. 2007.

[11] S. Wagner, R. Couillet, M. Debbah, and D. Slock, "Large system analysis of linear precoding in correlated MISO broadcast channels under limited feedback," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4509–4537, July 2012.

[12] D. P. Palomar, J. M. Cioffi, and M.-A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, 2003.

[13] 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. on Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, 2008.

[14] S. A. Jafar and A. J. Goldsmith, "Isotropic fading vector broadcast channels: The scalar upper bound and loss in degrees of freedom," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 848–857, Mar. 2005.

[15] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, Nov. 2006.

[16] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.

[17] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channels with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, 2005.

[18] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1478–1491, Sep. 2007.

[19] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, Jun. 2010.

[20] M. B. Shenouda and T. N. Davidson, "Robust linear precoding for uncertain MISO broadcast channels," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2006.

[21] ——, "Convex conic formulations of robust downlink precoder designs with quality of service Constraints," *IEEE Journal of Sel. Topics in Sign. Process.*, vol. 1, no. 4, pp. 714–724, 2007.

[22] N. Vucic, H. Boche, and S. Shi, "Robust transceiver optimization in downlink multiuser MIMO systems," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3576–3587, Sep. 2009.

[23] M. A. Maddah-Ali and D. N. C. Tse, "Completely stale transmitter channel state information is still very useful," in *Proc. Allerton Conference on Communication, Control, and Computing (Allerton)*, 2010.

[24] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, "Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315–328, Jan. 2013.

[25] R. Tandon, S. A. Jafar, S. Shamai (Shitz), and H. V. Poor, "On the synergistic benefits of alternating CSIT for the MISO BC," *IEEE Trans. Inf. Theory*, vol. 59, no. 7, pp. 4106–4128, 2013.

[26] J. Chen, S. Yang, and P. Elia, "On the fundamental feedback-vs-performance tradeoff over the MISO-BC with imperfect and delayed CSIT," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2013.

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

[28] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, 2013.

[29] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 264–273, 2013.

[30] D. J. Love, J. Choi, and P. Bidigare, "A closed-loop training approach for massive MIMO beamforming systems," in *Proc. Conference on Information Sciences and Systems (CISS)*, 2013.

[31] S. H. Ali and V. C. M. Leung, "Dynamic frequency allocation in fractional frequency reused OFDMA networks," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 8, pp. 4286–4295, 2009.

[32] C. H. Y. Xiang, J. Luo, "Inter-cell interference mitigation through flexible resource reuse in OFDMA based communication networks," in *Proc. 13th European Wireless Conference (EW)*, 2007.

[33] Y. Huang and B. D. Rao, "An analytical framework for heterogeneous partial feedback design in heterogeneous multicell OFDMA networks," *IEEE Trans. Signal Process.*, vol. 61, no. 3, pp. 753–769, 2013.

[34] ——, "Performance analysis of heterogeneous feedback design in an OFDMA downlink with partial and imperfect feedback," *IEEE Trans. Signal Process.*, vol. 61, no. 4, pp. 1033–1046, 2013.

[35] D. Tse and P. Viswanath, *Funadamentals of Wireless Communications*. Cambridge University Press, 2005.

[36] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transaction on Communications*, vol. 10, pp. 585–595, 1999.

[37] V. R. Cadambe and S. A. Jafar, "Interference alignment and degrees of freedom of the K-user interference channel," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3425–3441, Aug. 2008.

[38] M. Maddah-Ali, A. Motahari, and A. Khandani, "Communication over MIMO X channels: Interference alignment, decomposition, and performance analysis," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, pp. 3457–3470, Aug. 2008.

[39] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. on Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, 2010.

[40] R. Zakhour and D. Gesbert, "Distributed multicell-MISO precoding using the layered virtual SINR framework," *IEEE Trans. Wireless Commun.*, vol. 9, no. 8, Aug. 2010.

[41] 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.

[42] R. Zhang and S. Cui, "Cooperative interference management with MISO beamforming," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5450–5458, 2010.

[43] B. Özbek and D. Le Ruyet, "Adaptive limited feedback for intercell interference cancelation in cooperative downlink multicell networks," in *Proc. IEEE International Symposium on Wireless Communication Systems (ISWCS)*, 2010.

[44] H. A. A. Saleh and S. D. Blostein, "Single-cell vs. multicell MIMO downlink signalling strategies with imperfect CSI," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2010.

[45] W. W. L. Ho, T. Q. S. Quek, S. Sun, and R. W. Heath, "Decentralized precoding for multicell MIMO downlink," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1798–1809, Jun. 2011.

[46] R. Bhagavatula and R. W. Heath, "Adaptive limited feedback for sum-rate maximizing beamforming in cooperative multicell systems," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 800–811, Feb. 2011.

[47] A. Tajer and X. Wang, "Information exchange limits in cooperative MIMO networks," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2927–2942, Jun. 2011.

[48] B. Khoshnevis, W. Yu, and Y. Lostanlen, "Two-stage channel feedback for beamforming and scheduling in network MIMO systems," in *Proc. International Conference on Communications (ICC)*, 2012.

[49] C. Yetis, T. Gou, S. A. Jafar, and A. H. Kayran, "On feasibility of interference alignment in MIMO interference networks," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4771–4782, Sep. 2010.

[50] M. Razaviyayn, G. Lyubeznik, and Z.-Q. Luo, "On the degrees of freedom achievable through interference alignment in a MIMO interference channel," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 812–821, Feb. 2012.

[51] 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.*, vol. 57, no. 6, pp. 3309–3322, Jun. 2011.

[52] D. Schmidt, C. Shi, R. Berry, M. L. Honig, and W. Utschick, "Minimum mean squared error interference alignment," in *Proc. IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC)*, 2009.

[53] S. Peters and R. Heath, "Cooperative algorithms for MIMO interference channels," *IEEE Trans. Veh. Technol.*, vol. 60, no. 1, Jan. 2011.

[54] G. Bresler, D. Cartwright, and D. N. C. Tse, "Geometry of the 3-User MIMO interference channel," in *Proc. Allerton Conference on Communication, Control, and Computing (Allerton)*, 2011.

[55] M. K. Karakayali, G. J. Foschini, and R. A. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Communications*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[56] O. Somekh, O. Simeone, Y. Bar-Ness, and A. M. Haimovich, "Distributed multi-cell zero-forcing beamforming in cellular downlink channels," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2006.

[57] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO Channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[58] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, 2008.

[59] J. Thukral and H. Boelcskei, "Interference alignment with limited feedback," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2009.

[60] R. Krishnamachari and M. Varanasi, "Interference alignment under limited feedback for MIMO interference channels," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2010.

[61] 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.

[62] O. E. Ayach, S. W. Peters, and R. W. Heath, "The practical challenges of interference alignment," *IEEE Wireless Commun. Mag.*, vol. 1, no. 20, pp. 35–42, Feb. 2013.

[63] 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.

[64] X. Rao, L. Ruan, and V. K. N. Lau, "CSI feedback reduction for MIMO interference alignment," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4428–4437, Sept. 2013.

[65] S. A. Jafar, "Topological interference management through index coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 529–568, 2014.

[66] B. L. Ng, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theo.*, vol. 54, no. 12, pp. 5491–5499, 2008.

[67] A. Papadogiannis, D. Gesbert, and E. Hardouin, "A dynamic clustering approach in wireless networks with multi-cell cooperative processing," in *Proc. International Conference on Communications (ICC)*, 2008.

[68] A. Papadogiannis and G. C. Alexandropoulos, "The value of dynamic clustering of base stations for future wireless networks," in *Proc. IEEE International Conference on Fuzzy Systems (FUZZ)*, 2010.

[69] J. Gong, S. Zhou, Z. Niu, L. Geng, and M. Zheng, "Joint scheduling and dynamic clustering in downlink cellular networks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2011.

[70] A. Giovanidis, J. Krolikowski, and S. Brueck, "A 0-1 program to form minimum cost clusters in the downlink of cooperating base stations," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, 2012.

[71] I. Bergel, D. Yellin, and S. Shamai (Shitz), "Linear precoding bounds for Wyner-type cellular networks with limited base-station cooperation and dynamic clustering," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3714–3725, Jul. 2012.

[72] H. Huh, A. M. Tulino, and G. Caire, "Network MIMO with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2911–2934, May 2012.

[73] A. Lozano, R. W. Heath, and J. G. Andrews, "Fundamental limits of cooperation," *Information Theory, IEEE Transactions on*, vol. 59, no. 9, pp. 5213–5226, 2013.

[74] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS long term evolution: From theory to practice*, 2nd ed. Wiley, 2011.

[75] R. Zakhour and D. Gesbert, "Team decision for the cooperative MIMO channel with imperfect CSIT sharing," in *Proc. Information Theory and Applications Workshop (ITA)*, 2010.

[76] D. t. h. Rao, "A matter of perspective: Reliable communication and coping with interference with only local views," Ph.D. dissertation, Rice University, 2012.

[77] 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.

[78] 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.

[79] ——, "Precoder optimization with local and shared CSI on the K-user MIMO interference channel," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2013.

[80] W. Dai, Y. Liu, and B. Rider, "Quantization bounds on Grassmann manifolds and applications to MIMO communications," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1108–1123, Mar. 2008.

[81] 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.

[82] P. E. Santacruz, V. Aggarwal, and A. Sabharwal, "Beyond interference avoidance: Distributed sub-network scheduling in wireless networks with local views," in *in Proc. IEEE International Conference on Computer Communication (INFOCOM)*, 2013.

[83] T. Gou and S. Jafar, "Optimal use of current and outdated channel state information: Degrees of freedom of the MISO BC with mixed

CSIT," *IEEE Communications Letters*, vol. 16, no. 7, pp. 1084–1087, Jul. 2012.

[84] 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.

[85] G. Caire, N. Jindal, and S. Shamai (Shitz), "On the required accuracy of transmitter channel state information in multiple antenna broadcast channels," in *Proc. IEEE Asilomar Conference on Signals, Systems and Computers (ACSSC)*, 2007.

[86] K. Gomadam, V. R. Cadambe, and S. A. Jafar, "Approaching the capacity of wireless networks through distributed interference alignment," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2008.

[87] K. R. Kumar, "An iterative algorithm for joint signal and interference alignment," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2010.

[88] I. Santamaría, Ó. González, R. W. Heath, and S. W. Peters, "Maximum sum-rate interference alignment algorithms for MIMO channels," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2010.

[89] D. Papailiopoulos and A. Dimakis, "Interference alignment as a rank constrained rank minimization," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2010.

[90] S. W. Peters and R. W. Heath, "User partitioning for less overhead in MIMO interference channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 592–603, Feb. 2012.

[91] R. Etkin, D. Tse, and H. Wang, "Gaussian interference channel capacity to within one bit," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5534–5562, Dec. 2008.

[92] D. Tuninetti, "On interFerence channel with generalized feedback (IFC-GF)," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2007.

[93] W. Wu, S. Vishwanath, and A. Arapostathis, "Capacity of a class of cognitive radio channels: Interference channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, Nov. 2007.

[94] S. A. Jafar and S. Vishwanath, "Generalized degrees of freedom of the symmetric Gaussian K user interference channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3297–3303, Jul. 2010.

[95] P. Mohapatra and C. Murthy, "Inner bound on the GDOF of the K-user MIMO Gaussian symmetric interference channel," *IEEE Trans. Commun.*, vol. PP, no. 99, pp. 1–10, 2012.

[96] C. S. Vaze and M. K. Varanasi, "The degree-of-freedom regions of MIMO broadcast, interference, and cognitive radio channels with no CSIT," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5354–5374, Aug. 2012.

[97] A. Chaaban and A. Sezgin, "On the generalized degrees of freedom of the Gaussian interference relay channel," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4432–4461, Jul. 2012.

[98] R. Radner, "Team decision problems," *The Annals of Mathematical Statistics*, 1962.

[99] J. Marschak and R. Radner, *Economic theory of teams*. Yale University Press, New Haven and London, Feb. 1972.

[100] Y. C. Ho, "Team decision theory and information structures," *Proceedings of the IEEE*, vol. 68, no. 6, pp. 644–654, 1980.

[101] W. Santipach and M. L. Honig, "Capacity of a multiple-antenna fading channel with a quantized precoding matrix," *IEEE Trans. on Inf. Theory*, vol. 55, no. 3, pp. 1218–1234, Mar. 2009.

[102] C. S. Vaze and M. K. Varanasi, "CSI feedback scaling rate vs multiplexing gain tradeoff for DPC-based transmission in the Gaussian MIMO broadcast channell," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2010.

[103] F. Boccardi, H. Huang, and A. Alexiou, "Hierarchical quantization and its application to multiuser eigenmode transmissions for MIMO broadcast channels with limited feedback," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2007.

[104] C. T. K. Ng, D. Gunduz, A. J. Goldsmith, and E. Erkip, "Distortion minimization in Gaussian layered broadcast coding with successive Refinement," *IEEE Trans. Inf. Theo.*, vol. 55, no. 11, pp. 5074–5086, 2009.

[105] W. Xu, C. Zhao, and Z. Ding, "Optimisation of limited feedback design for heterogeneous users in multi-antenna downlinks," *IET Communications*, vol. 3, no. 11, pp. 1724–1735, 2009.

[106] Óscar González, C. Beltrán, and I. Santamaría, "On the number of interference alignment solutions for the K-User MIMO channel with constant coefficients," 2013, submitted to IEEE Trans. Inf. Theory. [Online]. Available: http://arxiv.org/abs/1301.6196

[107] M. Rezaee and M. Guillaud, "Interference alignment with quantized Grassmannian feedback in the K-user MIMO interference channel," 2013, submitted to IEEE Trans. Inf. Theory. [Online]. Available: http://arxiv.org/abs/1207.6902

[108] P. de Kerret and D. Gesbert, "Interference alignment with incomplete CSIT sharing," 2012, submitted to IEEE Trans. Wireless Commun., Nov. 2012. [Online]. Available: http://arxiv.org/pdf/1211.5380v1.pdf

[109] W. X. Xiao Chen, Wen Li, "Perturbation analysis of the eigenvector matrix and singular vector matrices," *Taiwanese Journal of Mathematics*, vol. 16, no. 1, pp. 179–194, Feb. 2012.

[110] T. Gou and S. Jafar, "Degrees of freedom of the K user M x N MIMO interference channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 12, pp. 6040–6057, Dec. 2010.

[111] B. Girod. (2013, Aug.) Lecture on image and video compression (EE398A). [Online]. Available: http://www.stanford.edu/class/ee398a/handouts/lectures/05-Quantization.pdf

[112] D. Samardzija and N. Mandayam, "Unquantized and uncoded channel state information feedback in multiple-antenna multiuser systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1335–1345, Jul. 2006.

[113] P. de Kerret and D. Gesbert, "CSI sharing strategies for transmitter cooperation in wireless networks," *IEEE Wireless Commun. Mag.*, vol. 20, no. 1, pp. 43–49, Feb. 2013.

[114] S. Jaffard, "Decay rates for inverses of band matrices," *Gauthiers-Villars, Annales de l'I.H.P., section C*, vol. 7, no. 5, 1990.

[115] K. Groechenig and M. Leinert, "Symmetry and inverse-closedness of matrix algebras and functional calculus for infinite matrices," *Trans. of the American Mathematical Society*, Jan 2006.

[116] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communication: scaling laws and optimal operation," *IEEE Trans. Inf. Theo.*, vol. 50, no. 5, pp. 748–767, May 2004.

[117] A. Ozgur, O. Leveque, and D. N. C. Tse, "Hierarchical cooperation achieves optimal capacity scaling in Ad Hoc networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3549–3572, Oct. 2007.

[118] O. Gonzalez, I. Santamaria, and C. Beltran, "A general test to check the feasibility of linear interference alignment," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2012, pp. 2481–2485.

[119] L. Ruan, V. K. N. Lau, and M. Z. Win, "The feasibility conditions for interference alignment in MIMO networks," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2066–2077, 2013.

[120] P. de Kerret and D. Gesbert. (2013, Jun.) IA feasibility test and IA with incomplete CSIT sharing. [Online]. Available: https://sites.google.com/site/pdekerret/home

[121] R. Couillet and M. Debbah, *Random matrix methods for wireless Communications*. Cambridge University Press, 2011.

[122] T. Inglot and T. Ledwina, "Asymptotic optimality of new adaptive test in regression model," *Annales de l'institut Henri Poincarï¿½ (B) Probabilitï¿½s et Statistiques*, vol. 42, no. 5, pp. 579–590, 2006. [Online]. Available: http://eudml.org/doc/77909.

[123] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, 7th ed. Elsevier/Academic Press, Amsterdam, 2007.

[124] A. Edelman, "Eigenvalues and condition numbers of random matrices," Ph.D. dissertation, MIT, 1989.

[125] C. D. Meyer, Ed., *Matrix analysis and applied linear algebra*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2000.

[126] A. Tulino and S. Verdu, *Random matrix theory and wireless communications*. Now Publisher Inc., 2004.